

SOCIAL MEDIA AND INNOVATION ECOSYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

Sanjay K. Arora

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Public Policy

Georgia Institute of Technology
May 2016

Copyright © 2016 by Sanjay K. Arora

SOCIAL MEDIA AND INNOVATION ECOSYSTEMS

Approved by:

Dr. Philip Shapira, Advisor
School of Public Policy
Georgia Institute of Technology

Dr. Jan Youtie
Enterprise Innovation Institute
Georgia Institute of Technology

Dr. Diana Hicks
School of Public Policy
Georgia Institute of Technology

Dr. Julia Melkers
School of Public Policy
Georgia Institute of Technology

Dr. Eric Gilbert
School of Interactive Computing
Georgia Institute of Technology

Date Approved: January 12, 2016


~~~~  
To my friends, family, and spiritual community, who mean the world to me.  
To a vision of a better world, which keeps me inspired, curious, and restless.  
And to today, for which I am grateful and content.

~~~~

ACKNOWLEDGEMENTS

I would like to thank my dissertation chair, Philip Shapira, and other committee members, Jan Youtie, Diana Hicks, Eric Gilbert, and Julia Melkers for their time and commitment in making this research a reality. Their input and encouragement has been priceless. I would also like to convey my gratitude to the Program in Science, Technology, and Innovation Policy, led by Philip Shapira, Jan Youtie, Alan Porter and Juan Rogers, who have provided the type of apprenticeship that any aspiring PhD student could want, namely the freedom to explore new ideas and the guidance and support to grow as a scholar. The School of Public Policy's faculty, staff, and students; the Enterprise Innovation Institute; the Ivan Allen College of Liberal Arts; and broader Georgia Tech community have also been instrumental in preparing the foundations of this work. I should also thank Twitter and the interviewee respondents, who have freely provided me with two robust data sets. EY and IBM, my two links to the corporate world throughout this five-year endeavor, encouraged my continuing education and provided me the time and space to initiate and complete this work.

This research was supported by the US National Science Foundation through the Center for Nanotechnology in Society (Award No. 0937591). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xii
NOMENCLATURE	xiv
SUMMARY	xv
Chapter 1: INTRODUCTION.....	1
1.1. Definitions	6
1.2. Innovation Models and Implications for Studying Networking and Communication	8
1.3. Research Question and Method.....	12
1.4. Contributions	17
1.5. Dissertation organization.....	20
Chapter 2: LITERATURE REVIEW.....	21
2.1. Open innovation and ICTs.....	23
2.2. Social capital and information and knowledge as resources	30
2.3. Ecosystem actors	40
2.4. Exploratory Propositions	53
Chapter 3: RESEARCH CONTEXT	67
3.1. Twitter	67

3.2. Graphene.....	70
Chapter 4: DATA AND METHOD	79
4.1. Data sources and sampling strategy.....	80
4.2. Quantitative method	81
4.2.1. Visualization.....	83
4.2.2. Variables.....	85
4.2.3. Model selection	93
4.3. Qualitative method	96
4.4. Summary.....	99
Chapter 5: NETWORK ORGANIZATION AND ANALYSIS.....	103
5.1. Select Firm Ego Networks.....	103
5.1.1. @ZyvexTech.....	104
5.1.2. @TeamGraphene.....	110
5.1.3. Cross Case Analysis.....	114
5.2. Global Patterns and Community Structure.....	116
5.2.1. Combined Friend Network.....	116
5.2.2. Combined Follower Network.....	130
5.2.3. Comparison	133
5.3. Summary and Synthesis.....	135
Chapter 6: QUANTITATIVE MODELING.....	138

6.1. Modeling Issues and an Alternative Approach.....	140
6.2. ERGM Example Run.....	144
6.3. Testing the Relationship Between Actor Type and Network Structure	152
6.4. The Mediating Role of Information Distance.....	163
6.4.1. Topic Modeling Descriptive Results.....	164
6.4.2. Testing the Relationship.....	168
6.4.3. Post-Hoc Analysis	172
6.5. Summary.....	179
Chapter 7: QUALITATIVE FINDINGS	182
7.1. Sample Overview and Usage Patterns.....	183
7.2. Awareness.....	190
7.3. Problem solving.....	193
7.4. Transitioning from Public to Private Interaction.....	195
7.5. Other Innovation Outcomes.....	198
7.5.1. Sales and Marketing	199
7.5.2. Jobs.....	202
7.5.3. Impacts on the Broader Scientific Community	203
7.6. Summary and Synthesis.....	205
Chapter 8: DISCUSSION AND CONCLUSION.....	211
8.1. Summary and Integration of Findings	215

8.2. A Communication-Engagement Model to Predict Innovation Outcomes	221
8.3. Theoretical Implications	227
8.3.1. Integrating Two Research Streams through Narrative Construction	228
8.3.2. Social Capital and Information Value	233
8.4. Innovation Policy and Management Implications	235
8.5. Limitations	242
8.5.1. Alternative Explanations	243
8.5.2. Methodological Reflections	245
8.5.3. Usage Dynamism	249
8.6. Future Work	250
8.6.1 Temporality	251
8.6.2. Science Maps	254
8.7. Concluding Thoughts	257
APPENDIX	260
REFERENCES	269

LIST OF TABLES

Table 2.1: Ecosystem actor types, information needs, and expected behaviors	42
Table 3.1: Comparison of random tweets vs. graphene tweets by content category	77
Table 3.2: Illustrative tweets by five content categories.....	77
Table 4.1: Variable operationalizations for the first research question	86
Table 4.2: Terms to identify user categories on Twitter.....	87
Table 4.3: Interview protocol.....	101
Table 5.1: Non-random selection of tweets issued by Zyvex Technologies	109
Table 5.2: Non-random selection of tweets issued by Graphene Frontiers	115
Table 5.3: List of graphene firms in the sample and their corresponding cluster numbers	118
Table 5.4: Number of users by actor type in the combined friend network	124
Table 5.5: Mixing matrix for combined friend network.....	129
Table 5.6: Number of users by actor type in the combined follower and friend networks	134
Table 6.1: Descriptive statistics and correlations for continuous variables.....	144
Table 6.2: Basic ERGM model.....	146
Table 6.3: Adding a binary attribute for whether actors are graphene firms	147
Table 6.4: Adding a term for mutuality	148
Table 6.5: Actor mixing parameters with usage and network structural controls	155
Table 6.6: Summary of s.s. sample coefficients	156
Table 6.7: Summary of s.s. sample coefficients	159

Table 6.8: Comparison of top three topics trained on sample tweets from C3-945 for three select users	166
Table 6.9: Comparison of top three topics trained on 100,000 random graphene tweets for three select users	167
Table 6.10: Model results adding information distance variables	169
Table 6.11: Adding an information distance term for differences in graphene tweeting	177
Table 7.1: Interviewee profile and usage characteristics	185
Table 7.2: Word table summarizing qualitative findings.....	186
Table 7.3: Usage and outcome evidence by respondent	208
Table 8.1: Summary of propositions, findings, and remarks	214
Table 8.2: A set of temporality-driven propositions for future work	253
Table A.0.1: Coding Outsourcing Job Description.....	264
Table A.0.2: Log-likelihood estimates.....	267
Table A.0.3: Log-likelihood estimates.....	268

LIST OF FIGURES

Figure 2.1: Conceptual framework	23
Figure 2.2: Conceptual model.....	54
Figure 4.1: Platelet diagram.....	90
Figure 4.2: A logic model.	101
Figure 5.1: Zyvex Technology’s friend network.	106
Figure 5.2: Graphene Frontier’s friend network.	112
Figure 5.3: The combined firm friend network.	120
Figure 5.4: Median number of followers by actor type in combined friend network.....	125
Figure 5.5: Median number of friends by actor type in combined friend network.....	126
Figure 5.6: Mean number of graphene tweets by actor type in combined friend network.	126
Figure 5.7: Median normalized tweet activity by actor type in the combined friend network.	127
Figure 5.8: The combined firm follower network.	131
Figure 6.1: C3 community from the combined friend network.....	143
Figure 6.2: Goodness of fit diagnostics for in-degree distribution, 0-25.....	150
Figure 6.3: Goodness of fit diagnostics for out-degree distribution, 0-25.....	151
Figure 6.4: Path diagram showing direct and indirect effects	175
Figure 8.1: A social media communication-engagement model building on Lin’s (2001) model of social capital development.....	226
Figure 8.2: Narrative co-creation process.....	232
Figure A.1: The linear model of innovation.	260

Figure A.2: Modeling a theory of social capital.	261
Figure A.3: Social capital and the development of intellectual capital.	262
Figure A.4: A structural model of network emergence.	263

NOMENCLATURE

ST&I	Science, Technology, and Innovation
NIS	National Innovation System
RIS	Regional Innovation System
SMEs	Small and Medium-Sized Enterprises
Darwinian Sea	A series of challenges, including resource constraints and information asymmetries, that high-technology entrepreneurs face
GPT	General Purpose Technology
ICT	Information and Communications Technology
API	Application Programming Interface
ERGM (a.k.a. p^* models)	Exponential Random Graph Model
Actor	An individual or organization in the ecosystem network, also known as a user in an online network
Linkage	A relationship between any two actors, also known as a tie or edge
Community	A grouping of densely connected actors, also referred to as a cluster
Twitter	Social media platform allowing users to tweet (author) and read 140 characters of text
Friend	An actor whom a user follows
Follower	An actor following a user
Timeline	All the tweets from a user's following network placed in a single "feed"
Handles	A username preceded by an "@" sign
Hashtag	Topical metadata identified through the "#" symbol
Mentioning	Placing another user's handle in the body of a tweet
Direct messaging	Tweeting to private inboxes
Retweet	Syndicating an existing tweet to a user's followers

SUMMARY

The innovation ecosystem's online presence continues to grow with the emergence and maturation of ICT-based platforms. With these new channels, a diversity of actors, including firms, scientists, universities, media entities, and individuals, interact to satisfy their information needs and to access and mobilize *network-based* resources. This research is among a growing number of social science studies examining the advent of social media and its influence on the innovation process, asking, "How do different types of actors use social media to form network linkages, and what kinds of innovation outcomes will result?" The purpose of this work is (1) to explore whether established theories relevant to science, technology, and innovation (ST&I) inquiry perform as expected in social media domains, and (2) to examine some of the consequences of ICT-enabled innovation. To accomplish this task, I rely on an approach that considers both institutional factors (top-down and as prescribed by the broader innovation literature) and emergent phenomena (bottom-up interaction based on micro-level network behaviors). Pulling from three literature streams, I focus my theoretical development on open innovation, social capital, and the broader innovation system landscape to explain how individuals access information and resources online. In the process, I take into account that social media acts as a conduit for communication, often times in platform-specific ways.

To study this complex network activity, I turn to Twitter, the popular microblogging service, and focus on the case of graphene, a novel nanotechnology material consisting of a two-dimensional sheet of carbon atoms. Twitter is one of the world's most often-used social networks, boasting over 500 million users (200+ million

active). Graphene, on the other hand, is a relatively well-bounded area of scientific inquiry with ongoing, concurrent scientific and commercialization activity.

Twitter and graphene can be viewed as “radical” innovations (i.e., new technologies advancing new markets) in ostensibly divergent domains (i.e., ICTs and material science), yet both are approximately the same age with Twitter having been founded in 2006 and graphene’s first major breakthrough occurring in 2004. This work provides a lens into the graphene online innovation ecosystem which has developed in tandem with Twitter’s distinctive brand of social media and notable impacts on popular culture.

To position the value of social media as a novel method for innovation studies, I compare tweets from the Twitter public timeline with a random sample of graphene tweets. The sample of graphene tweets shows lower frequencies of tweets containing mundane musings and conversational discourse, thus indicating an overall higher level of professionalized communication than the strictly random sample. Twitter, I conclude, can act as a legitimate forum for analyzing ecosystem discourse and studying innovation outcomes.

This research contains three propositions guiding the exploratory empirical results. The first two propositions address how actors follow one another on Twitter to generate network linkages. The third proposition posits that beneficial innovation outcomes result as a consequence of social media participation. The primary sample dataset derives from 34 graphene firms’ friend and followers relationships captured in early 2014. Nine interview transcripts supply qualitative data.

The results show that network formation on Twitter is not random and that actors choose whom to follow by mixing within and across actor affiliation types. For instance, all user types are likely to follow media entities, a group of actors traditionally not considered important in innovation processes. Furthermore, while nanotechnology firms in general are likely to follow other types of actors, these other actors are generally not likely to follow nanotechnology firms. This implies that nanotechnology firms use Twitter to gain access to information and resources, but they face an up-hill battle. Consequently, the online ecosystem may look remarkably similar to the prevailing offline “Darwinian Sea”, where high-tech SMEs work to enhance their reputations, compete for scarce resources, and struggle with market positioning and management *savoir-faire* to commercialize new inventions.

A series of network visualizations reveal that users agglomerate in communities; these communities exhibit greater density than the larger ecosystem network and a propensity to congeal in topically focused ways. That is, each community indicates a coherent topical focus, suggesting that graphene firms follow specific sets of users in ways that support their information and resource needs. For example, based on ego-network visualizations, some firms show a propensity to access regional communities (i.e., the actors comprising the “regional innovation system”), in addition to a globally disperse set of ecosystem actors (and their communities). This finding underlines the importance of local contexts in high-technology entrepreneurship, even as captured on a globally far-reaching virtual platform such as Twitter. At the micro-level, an unstructured text mining approach to operationalizing and computing information distance shows that increasing amounts of topical distance between any two users

decreases the likelihood of a tie existing. In sum, users not only follow actors because of their revealed identities but also because of the information value those actors provide. Moreover, even whilst maintaining topical focus, actor diversity in some communities is the norm, not the exception.

Are innovation outcomes more likely to occur in strategically-developed and information-rich social media networks? Drawing on different sources of “behavioral additionality” – or changes in behaviors as a result of social media participation – I identify *ex-ante* several such plausible outcomes, which could include increased awareness, improved problem solving ability, community and brand development, enhanced participation, and greater sales. The qualitative results show that social media participation results in increased awareness of graphene and related ecosystem topics, but engagement is a key tactical maneuver that actors pursue, often in varying ways, to access and mobilize other resources. The findings also indicate that problem solving and deep contextual learning do not systematically occur on Twitter.

Taken together, the findings reveal that identity and information help inform the following decisions of users. In addition, the online innovation ecosystem on Twitter is both a broadcast communication channel where actors share and consume messages *en masse* and where strategic forms of engagement allow ecosystem actors to transfer resources (beyond information). A theoretical implication of this synthesis suggest that, through engagement, some actors seek to influence the development of various narratives emerging in the graphene innovation ecosystem; that is, they shape the contour of discourse to build a wider following and to position themselves in the center of their respective communities.

Policy implications are targeted at intermediary institutions and scientists, while management implications focus on high-technology SMEs. For example, intermediaries and scientists as a whole appear to interact in insular rather than boundary-spanning ways: Scientists follow other scientists but not nanotechnology firms; nanotechnology firms follow intermediaries without receiving the same level of interest in return. If the online ecosystem indeed conveys the same set of information asymmetries characterizing the offline “Darwinian Sea”, the first step in promoting more balanced interactions is to encourage a more fluid set of bi-directional following relationships. While difficult to manifest online in self-organizing networks, policies could facilitate the transition of offline casual introductions (e.g., at conferences or forums) into online following relationships. In theory, while such a maneuver guarantees access to network relationships, it may not necessarily result in resource mobilization. Yet, according to social capital theory, access is a necessary condition for mobilization to occur.

In terms of management implications, firms should focus their efforts on identifying gaps in their social media presence by examining topical holes in their networks and/or observing the following relationships of their peers and competitors. There is also an opportunity for firms to look further upstream or downstream in the value chain to gain a better perspective of their current position and anticipated placement in the ecosystem. For instance, commensurate with exploration and search, graphene firms manufacturing the material as an intermediate input could follow end-user product firms to learn more about those downstream industries.

This work concludes with a discussion of limitations and opportunities for future work. Limitations include alternative theories to explaining social media participation

and engagement, methodological issues, and the continuing evolution of social media platforms and usage patterns. Future work is considered to address the temporal nature of network construction and topical growth (or constriction), as well as the ability to map areas of science and technology through social media data.

In summary, this dissertation reveals that the (graphene) innovation ecosystem has an active and growing online footprint and that participation is more diverse than other models of innovation would predict. Yet this diversity exists within the confines of a professionalized and on-point discourse, as mediated by dense communities with which actors may engage to access information and resources. Further, through its research design, this work contributes to the emerging research stream known as computational social science, which uses ICT-based datasets to study human interaction and communication at multiple levels of analysis. As ICT platforms increasingly mediate innovation processes and influence the socio-cultural environment in which innovation occurs, a computational approach to studying high-technology entrepreneurship acknowledges much of the extant ecosystem complexity whilst enabling novelty and ingenuity in research design and method. Such a development will not only help scholars better characterize and understand how innovation in the 21st century will unfold but also act to advance a *contemporary* and *interdisciplinary* social science.

CHAPTER 1: INTRODUCTION

Today's innovation landscape looks considerably different than it did in much of the second half of the 20th century, where large firms dominated the industrial R&D landscape (Fagerberg, 2004). Knowledge, for one, is becoming increasingly specialized while at the same time becoming increasingly amenable to abstraction (A. Arora & Gambardella, 1994). As a result, organizations can focus on “core competencies” and rely on external, often globally dispersed channels of know-how to link together complimentary assets across a value chain (Teece, 1986). This type of vertical disintegration gives rise not only to niche markets but also embedded networks where actors transact for goods, services, infrastructure and intangible assets such as knowledge and intellectual property (Granovetter, 1985; Powell & Snellman, 2004). The dispersion of innovation-related behavior across a variety of actors, including large and small firms, intermediaries, universities, and public labs is a hallmark of a “systems” view of innovation (Sharif, 2006). More recently, an even broader framework for inclusivity, one that accounts for the media and public, has risen in the form of an “innovation ecosystem”.

An innovation ecosystem is a “dynamic system of interconnected institutions, persons, and policies that are necessary to propel technological and economic development” (President's Council of Advisors on Science and Technology, 2008, p. 1). Implicit in this definition is the underpinning role of social networks which bind actors together via similar interests, shared challenges, competition, and technological progress. Innovation ecosystems evolve in response to inputs from participants, who complement one another through symbiosis and learning, and as a response to exogenous stimuli

(Thomas & Autio, 2012). At the heart of the ecosystem metaphor is the ability to communicate and interact across organizational lines, and in this regard, ICTs have taken a prominent role in the development of the innovation ecosystem's online footprint. It's not that ICTs *are* the innovation ecosystem, but rather ICTs facilitate novel ways of structuring, governing, funding, and profiting from the innovation process itself.

For example, consider the proliferation of online platforms that address information asymmetries, uncertainty, risk, and cost in entrepreneurial markets: In the crowd-funding space, sites including Kickstarter, Indiegogo, and Qwirky allow individuals to finance projects, many of which relate to novel science and engineering applications and which would otherwise not attract sufficient visibility and support to implement. Open innovation sites including IdeaExchange, OpenIDEO, and Innocentive outsource ideation and problem solving to a community of interested, sometimes paid parties. Cloud technologies simplify and reduce an organization's cost-commitment to in-house computing infrastructure and ICTs to meet elastic demand. In sum, these high-fidelity platforms may facilitate the construction of participants' value propositions, provide a framework to adjudicate the merit of ideas, and guide norms of interaction between members. Through these platforms, actors access resources and information which can accelerate and expand the scope of innovation outcomes (Ashurst, Freer, Ekdahl, & Gibbons, 2012).

Though communication with and financing from customers, the quality of an idea may improve with each prototype until it graduates into a bonafide product or service (Lynn, Morone, & Paulson, 1996). In some cases, business model innovation also occurs via a "lean", hypothesis-driven approach to maximizing resources while avoiding

common pitfalls in the startup's growth cycle (Ries, 2011). In this paradigm of entrepreneurial activity, the ecosystem metaphor lends itself as the setting against which value chains are navigated and negotiated. Add to this "value capture" dynamic the prospect of disruptive innovation (Bower & Christensen, 1995) stemming from emerging science-based technologies such as synthetic biology, additive (3-D) manufacturing, and nanotechnologies,¹ and it becomes evident that value capture is a function of both business model innovation and sound technology development (Chesbrough, 2010): This is because even good ideas and technologies must be paired with effective business models to achieve desirable economic outcomes. From this perspective, an entrepreneur's ability to access customers in diverse markets and in their day-to-day social environments becomes increasingly important, e.g. to devise, monitor, and learn from interactions that explore different product, service, and business model possibilities.

Operating adroitly in the ecosystem environ suggests an ability to leverage ICTs to get closer to end-user markets, attract resources, and even to collaborate. Indeed, it is here that the literature on open innovation refers to ICT-mediated networks as enabling inbound (and to a lesser degree outbound) knowledge flows (Chesbrough, 2006; Michelino, Caputo, Cammarano, & Lamberti, 2014; Ooms, Bell, & Kok, 2015; van de Vrande, de Jong, Vanhaverbeke, & de Rochemont, 2009). However, in contrast to platforms designed for specific functional needs such as ideation or financing, some ICTs facilitate communication and networking in broader terms. These online channels, which include household names such as Facebook, Twitter, and LinkedIn, provide general

¹ Futurists, and research institutes continue to conjecture on the potential of various emerging technologies. For example, see Manyika et al. (2013) for a survey from McKinsey Global Institute outlining several potentially disruptive innovations, which include nanotechnology, synthetic biology, and additive manufacturing.

forums for innovation discourse that may intersect with other, unrelated talking points and/or spheres of knowledge.² These sites add a critical dimension to participation in the ecosystem because they attract users who would not otherwise visit a specific site to observe or contribute to the ecosystem. Thus, multi-purpose social media platforms act as an entryway into the ecosystem; that is, these platforms boast usage from a cross-section of actor types, even as entrants into the ecosystem setting self-select in.

To-date, there has been a dearth of scholarship examining ecosystem network formation and the potential impacts of social media usage on innovation outcomes, particularly in high-technology entrepreneurial domains. And yet, because of the increase in usage and availability of data, the empirical landscape is ripe for such scholarly work. This research addresses this gap by examining the innovation ecosystem in the online realm, asking *how do different types of actors use social media to form network linkages, and what kinds of innovation outcomes will result?* It posits that users choose their relationships based on revealed professional affiliation, as well as the degree of information value subjectively perceived as being accessible through such relationships. In addition, this research argues that as ecosystem communications go online, and as network relationships become easier to form, innovation outcomes manifest in terms of broader awareness, deeper knowledge building and learning capacity, and the acquisition and transfer of resources and opportunities that may not otherwise be accessible without such ICT platforms. Also foundational to this study is a significant explorative component which assesses the merit of social media data and

² For example, Runge et al. (2013) study approximately 500,000 tweets through content analysis and find “that incidental exposure to nanotechnology seems plausible for Twitter users” when non-ecosystem stumble upon nanotechnology tweets by accident, e.g., when searching for “iPod nano”.

examines visualizations presenting the ecosystem's building blocks, one network at a time.

The research design is constructed in two phases: (1) a quantitative portion which assesses the nature of existing network linkages as a function of actor interests and goals and information type and content, and (2) a second segment which relies primarily on qualitative methods and uses a case study approach to explain innovation outcomes that arise as a result of social media usage and network formation. The empirical context is Twitter, the popular microblogging platform, while the innovation case study setting is graphene, a novel nanotechnology material consisting of a single layer of carbon atoms. Graphene, while being very strong, exhibiting high thermal and electrical conductivity, and providing the material-base for a host of applications, is also characterized by significant risk and uncertainty. The focus on graphene is appropriate because of the high visibility of the technology as well as its age: Graphene is a 21st century invention, and Twitter is a 21st century innovation. Consequently, this research studies one highly dynamic invention with numerous innovative possibilities (i.e., graphene) through another highly dynamic information and communication technology (ICT) platform (i.e., Twitter).

This introductory chapter aims to define the ecosystem concept in further detail (Section 1.1), explain where the ecosystem metaphor sits in the broader innovation literature and discuss subsequent implications for studying communication and networking (Section 1.2), lay out the research question and method (Section 1.3), and summarize key contributions (Sections 1.4). The chapter concludes with an overview of the dissertation's organization (Section 1.5).

1.1. Definitions

According to the President's Council of Advisors on Science and Technology (PCAST) (2008), the innovation ecosystem is a “dynamic system of interconnected institutions, persons, and policies that are necessary to propel technological and economic development” (pg. 1). In their 2008 report focusing on university-private sector research partnerships, PCAST identifies five broad areas to improve the vitality of the US innovation ecosystem: basic research and innovation, the economic regulatory environment, open innovation, connection points between partners, and finally measurement of innovation. These themes speak to the networked nature of innovation in the 21st century, as well as how to measure the process by which innovation occurs.

Thomas and Autio (2012) offer a similar definition, albeit one that emphasizes the firm as a central actor: The innovation ecosystem is “a network of interconnected organizations, organized around a focal firm or a platform and incorporating both production and use-side participants” (p. 2). The ecosystem, according to these authors, includes entities involved in the focal firm's value chain, as well as its customers. The combination of production and use side-participants distinguishes the ecosystem construct from other research on innovation networks, which typically address one side or the other (i.e., the value chain or customers). Prahalad and Ramaswamy (2003) argue that marrying these supplier and customer networks facilitates innovation processes focused not only on products and services but also on the co-creation of value through customer *experiences*.

As applied to geographically bound “innovation districts”, the innovation ecosystem consists of three primary components, according to the Brookings Institution:

[First,] economic assets are the firms, institutions, and organizations that drive, cultivate, or support an innovation-rich environment. [Second,] physical assets are the public and privately owned spaces—buildings, open spaces, streets, and other infrastructure—designed and organized to stimulate new and higher levels of connectivity, collaboration, and innovation. Lastly, networking assets are the relationships between actors—such as between individuals, firms, and institutions—that have the potential to generate, sharpen, and/or accelerate the advancement of *ideas*. These assets, taken together, create an *innovation ecosystem—the synergistic relationship between people, firms, and place that facilitates idea generation and advances commercialization* [emphasis added]. (Katz, Vey, & Wanger, 2015)

I argue that the innovation ecosystem is a type of complex system that can enable interactions across multiple, sometimes competing value chains (c.f. Simon, 1996). A complex system (1) consists of diverse and dynamic individual components, i.e., actors (2) responds to stimuli, (3) is organized into hierarchies, (4) and maintains a structure that can determine different ways of evolving (Neal, Smith, & McCormick, 2008). Some complex systems exhibit *emergence*, which refers to “a set of arguments that higher-level phenomena appear to exhibit properties that are not revealed at lower levels” (Monge & Contractor, 2003, p. 11). In contrast to top-down and control-centered theories that stress stasis and equilibrium, complex systems theory explicitly addresses dynamism and change. In terms of the innovation ecosystem construct, complex systems offer value by encouraging flexibility, transaction cost efficiencies, and resource exchange (Thomas & Autio, 2012).

Complex systems are fundamentally networks with complex topologies (Barabási, 1999). That is, emergence in complex systems is a result of the activities of semi-autonomous agents who act interdependently to form local networks, which then give rise to meso-level communities and an overarching global structure; networks first develop at the local level, with structure “emerging” upward to influence the construction of

communities and the global structure (Monge & Contractor, 2003). The Internet seems particularly amenable to promoting dynamic, self-organizing networks because of its ability to effectively link people, organizations, and knowledge (Wellman, 2001). Well-known publicly accessible online platforms for social networking include Facebook, Twitter, Meetup, and LinkedIn, and a host of other sites, some of which are specific to science and technology (e.g., ResearchGate and Engineering Exchange).

The innovation literature assumes productive connections between agents as critical to innovation outcomes, and scholars have produced numerous empirical studies substantiating this claim. However, research is more likely to examine formal networks even though informal linkages may be significantly more pervasive and consequential (C. Freeman, 1991; also see Blau, 1955). An explanation of this trend is the inherent difficulty in tracing informal connections as they mature and wane. As communication and social networks increasingly move to online environments, however, additional data sources become available by which to test propositions regarding micro-level interactions that embody the informal social networks underlying the innovation ecosystem. It is also becomes possible to visualize and comprehend how these micro-level linkages scale up into communities, and how these communities constitute smaller components of the larger ecosystem.

1.2. Innovation Models and Implications for Studying Networking and Communication

The notion of an innovation ecosystem stands in contrast to other models of innovation, including the linear model and institutional approaches such as the national innovation system and triple helix. The highly stylized, yet widely deployed linear model

may be traced to V. Bush's (1945) manifesto on the state of national science policy in the US (Godin, 2006). The linear model positions innovation as a sequence of temporally distinct phases in which investments in inputs result in a series of outputs (Neal et al., 2008): If basic research provides the bedrock for fundamental understanding, then applied research and development builds on prior knowledge for the purpose of commercialization and productization (Figure A.1). The linear model persists as a dominant archetype of innovation not because of its accuracy of how innovation actually progresses but rather because of its close association with innovation measurement (Godin, 2006). For example, scholars will often rely on publications as a measure of basic and applied research and patenting activity as a proxy for invention and development.

Yet the process of successfully monetizing nascent ideas presents technological, economic, *and* social challenges (Stinchcombe, 1990); that is, an invention is not an innovation without broader societal and market support. In this light, innovation and high-technology entrepreneurship may be viewed as a continuous search for relevant and effective ideas, information, and resources (Nelson & Winter, 1982). These ingredients are not always readily available to the people and organizations that have the potential to exploit them. This may result from information asymmetries, uncertainty or the friction not just of distance but also of different knowledge communities (Monge & Contractor, 2003). Thus, while the linear model suggests a straightforward unidirectional flow of information from upstream R&D to downstream market activities, it is not an empirically valid construct.

In response to the linear model's deficiencies, scholars in the 1980s turned to institutional frameworks (Etzkowitz & Leydesdorff, 2000; Sharif, 2006). Freeman (1987) defines the *national innovation system* as "the network of institutions in the public and private sectors whose activities and interactions initiate, import, modify and diffuse new technologies." Lundvall (2007) contends that firms constitute the core of an innovation system, but these firms interact with a variety of other institutions including other firms, universities and national labs, the prevailing intellectual property regime, labor markets, and the venture capital industry. To Lundvall, the innovation system is an analytical framework that explains how knowledge and learning evolves. In particular, learning through tacit knowledge sharing and recombination is the result of interaction (i.e., in network settings). A second institutional approach, the *triple helix model*, adds to this mix of "interactive" relationships between the governments, industry, and academic sectors an emphasis on evolving internal transformations within each institution. For example, the university continues to take a more central role in the knowledge-economy, as it acts as an education, research, and technology transfer hub (Godin & Gingras, 2000; Youtie & Shapira, 2008).

Both the national innovation system and triple helix model suffer from notable limitations, however. The national innovation system literature lacks conceptual clarity, a standard set of tactics to operationalize key concepts, and prescriptive guidance for the policy community (Sharif, 2006). The triple helix model, on the other hand, lacks explanatory power when comparing and contrasting across national boundaries (Shinn, 2002). Furthermore, I contend that both frameworks fail to sufficiently acknowledge the role of culture as a key ingredient in innovation. Wallner and Menrad (2011) argue that

culture embodies “the beliefs, values, and attitudes of a social system” (p. 4) and cannot be created; rather culture *emerges*. Moreover, the authors maintain that innovations are the byproduct of *innovativeness*, a cultural construct which can be enhanced with a diversity and tolerance of ideas. The ecosystem metaphor is amenable to studying a culture of innovativeness, to the extent that such a construct can be measured by the “emergence” characteristic of complex networks and information and discourse being communicated across individual relationships.

Implicit in the linear and systems models is the directionality and type of communication occurring among actors. In the linear model, the flow of information is largely directed downstream from basic and applied researchers to technology development and product teams. Based on the model (Figure A.1 in the Appendix), communication appears formalized and rigidly segregated between adjacent actors; for example, basic researchers communicate with applied researchers but not with product development teams. The innovation systems and triple helix literature assumes a series of bidirectional communication channels, and this communication need not be formal when individuals and institutions facilitate the transfer of tacit knowledge in informal settings (Saxenian, 1996). However, communication occurs with a select number of participants who often represent formal institutions. The concept of an innovation ecosystem, in contrast, suggests not only maintaining bidirectional flows but also casting a broader net in terms of actor participation. Thus, because communication can occur between any individual or institution and in any direction, the ecosystem construct is more amenable to empirically observing a wider array of informal networking and interaction than the institutional frameworks.

To briefly illustrate the value of the innovation ecosystem as a rhetorical device, consider the following set of actors and their (stylized) interests: the high-technology entrepreneur traverses a perilous “Darwinian Sea” in his search for funding and reputation, in his pursuit to find the right balance between managerial *savoir faire* and scientific expertise, and in his need to seek first mover advantages and/or protect intellectual property (Auersald, 2007; Auerswald & Branscomb, 2003). The venture capitalist, on the other hand, invests in a number of promising private firms with the hope that a few will sufficiently develop to warrant a liquidity event through a merger or acquisition or initial public offering (IPO). University scientists look for inspiration and commercialization opportunities in the “real-world” around them, while institutions foster a productive research and teaching environment from which to enhance their reputation (Stinchcombe, 1990). Students; media professionals; professionals including accountants, lawyers, and consultants; government agencies; and customers and suppliers also seek an interest in developing and shaping an innovation ecosystem.

In sum, the ecosystem approach explicitly recognizes a diversity of actors, some of whom represent formal institutions and others who do not (e.g., users with no revealed identity or professional affiliation). Moreover, the innovation ecosystem is more than just a “fourth helix” (c.f. Leydesdorff, 2012) (re-)introducing the public into the innovation terrain; it is a complex system that acknowledges both top-down structural and bottom-up emergent phenomena.

1.3. Research Question and Method

This research investigates the role of online networks in innovation ecosystems, investigating the extent to which social media contributes to ecosystem organizing

activity and beneficial innovation outcomes. In particular, it asks “*How and why do actors in innovation ecosystems use social media to communicate, and what are the impacts of social media on innovation?*” The purpose of this work is (1) to explore whether established theories relevant to science, technology, and innovation (ST&I) inquiry perform as expected in social media domains, and (2) to examine some of the consequences of ICT-enabled innovation. To accomplish this task, I rely on an approach that considers both *institutional factors* (top-down and as prescribed by the broader innovation literature) and *emergent phenomena* (bottom-up interaction based on micro-level network behaviors).

At its heart, this research is an exploratory study that attempts to quantify the *emerging* role of social media in innovation processes in *emerging* science-based technologies. The “how” aspect of the research question is tested through propositions that consider actor role identity, information as content, and the ensuing theoretical motivations for “why” individual actors would construct network ties with other actors. The recurring theoretical justification primarily draws on the benefits of ecosystem diversity, in terms of both actor diversity and information distance: Different types of actors have different information and resource needs, and thus, the empirical results *should* putatively show this type of variety in terms of inter-actor class linkages and underlying information content.

Beyond the propositions lie additional contextual insights which better position the formal results through a very pragmatic lens: Since social media is relatively novel both as a communication and social networking device, as well as a data source for research purposes, I consider descriptive aspects of the research setting to better support

the more sophisticated analysis. For example, I address foundational questions that the curious reader may have, e.g., “What does the innovation ecosystem even look like? How do individual actor networks compare with one another? And are actors actually communicating anything worthwhile, and if so, how might this differ from non-innovation ecosystem domains?” While these are not my primary research questions, I attempt to answer them *en route* for a better contextual understanding of the online environment and data.

As an applied social science study, this work also provides an outcomes-based assessment of social media usage in the innovation ecosystem. Social capital theory posits that networks are formed to access and transfer resources, resulting in returns (i.e., beneficial gains) to actors (Lin, 2001). While traditional ST&I indicators include outputs such as published manuscripts, patent applications and issuances, and firm survival, this study, given the nature of the data source, assumes a different set of plausible beneficial outcomes. For example, social media actors may realize improved environmental awareness; information transfer, learning, and problem solving; community/brand development; employment opportunities; and customer and revenue growth a result of participation. Exactly how resources and reputation are mobilized in the online innovation ecosystem is an uncertain process which this research seeks to elucidate.

I study ecosystem activity within one particular realm of social media – and within one particular realm of science-based technology development. The social media platform of interest is Twitter, the microblogging service. Twitter allows its users to “tweet” up to 140 characters; while this content may incite personal “blabber”, tweeting is often strategically aimed with the knowledge of a larger public audience in mind

(Marwick & Boyd, 2010). Twitter supports conversations, information and url sharing, and live news reporting in addition to personal status updates (Dann, 2010). The social networking component on Twitter allows users to follow one another, and tweets are curated on a user's timeline via the user's unique network lens. Thus, the information component on Twitter is highly tailored based on each user's following choices and interests. It is also possible to search for users and tweets thereby facilitating exploration of the larger Twitter universe. Trending topics alert users to content which has already attracted a relatively large set of tweeters and readers and which may continue to attract even more attention. From this perspective, Twitter also acts as an online mass-communication channel.

The technological domain is graphene, a highly touted nanotechnology material consisting of a one-layer-thick sheet of carbon atoms. Although in an early stage of development, graphene is currently being incorporated into improvements to existing applications such as inks, composites, and RFID (Shapira, Youtie, & Arora, 2012). In the future, experts expect more discontinuous advances from graphene-enabled products such as batteries, displays, hydrogen storage, and silicon-based transistors (Geim & Novoselov, 2007; Van Noorden, 2011). Yet, uncertainty about commercialization of graphene-enabled products is consequential on a number of important fronts (Segal, 2009; Van Noorden, 2011). For instance, current costs of production are prohibitively high and technical obstacles impede theoretically plausible performance characteristics, an especially vexing problem in transistors. In addition, graphene belongs to a broader class of carbon-based nanomaterials, which present a series of environmental, health, and

safety (EHS) concerns (Sweet & Strohm, 2006; Youtie, Porter, Shapira, Tang, & Benn, 2011).

The case study context is appropriate given that graphene may enable a host of incremental and disruptive innovations; graphene is also a “hype” technology with relatively broad interest from both the popular and S&T press, precisely because it is a material with a fast-growing research profile and commercialization trajectory. Actors operating in the online graphene innovation ecosystem are subject to and exert both influences (i.e., they participate in the hype and in narratives shaping the realistic potential of graphene). Consequently, this research examines both the broadcast communication component of Twitter as it relates to graphene’s “hype”, in conjunction with Twitter’s more subtle networking and personalized information sharing services. In addition, both Twitter and graphene are approximately the same age with Twitter being founded in 2006 and graphene having been first isolated in 2004. Twitter is an appropriate platform for examining the innovation ecosystem online because it is widely considered as one of the more active social media platforms. Furthermore, much of its data is freely available via publicly accessible APIs.

The research method uses secondary data from Twitter and primary data in the form of qualitative interview transcripts. The quantitative method specifies an exponential random graph model (ERGM), a type of network regression approach allowing parameter estimations for exogenous attributes at the actor, edge, and network levels. Furthermore, via simulations, the ERGM handles endogenous network parameters capturing emergent network behavior (e.g., reciprocity). The quantitative method assesses propositions concerning emergent networking behavior. In comparison,

the qualitative analysis treats each interview as a case study in a multi-case setting to address potential innovation-related outcomes of social media usage.

1.4. Contributions

The long tail of the innovation ecosystem suggests that we often are not in an immediate position to know its positive or negative effects until sometime in the future. We need more empirical research, case studies, and comparative analysis to better inform policymakers regarding what actually shapes innovation cycles. Innovation needs to be a real conversation starter, not just a dire warning that a particular initiative will have adverse consequences for our economy and our ability to enjoy the comforts of contemporary life. (Brotman, 2014)

This excerpt from the Brookings Institution calls for new research to better understand the dynamics of the innovation ecosystem to inform policymakers “regarding what actually shapes innovation cycles”. In response to this appeal, I contend that few studies examine the intersection of high-technology entrepreneurship *and* social media in the context of innovation ecosystems. This research intends to fill that gap from both a theoretical and methods standpoint. Consequently, it is of significance to scholars and policymakers alike on a number of different dimensions.

From a policy perspective, Auersald (2007) argues that policymakers should better understand the knowledge, incentives, and constraints that actors face within the domain of high-technology entrepreneurship. By illuminating (1) problem areas caused by a dearth of information and (2) market failures caused by underinvestment in public goods, policies can stimulate economic development in strategic sectors; i.e., enhanced ability to manage economic outcomes may result in increased economic performance. Economic development via increased levels of entrepreneurial activity is a particularly attractive talking point for policymakers, since small firms are a leading source of

employment in the US, contributing up to 60-80% of new jobs over the past decade (Wessner, 2007). Audretsch and Beckman (2007) contend that policy interventions intending to develop entrepreneurial activity not only support existing business and invest in the creation of knowledge but also cultivate a culture of *commercializing knowledge*. Because knowledge and entrepreneurial activity can be thought of as a public good (Audretsch, Grilo, & Thurik, 2007; Audretsch, Weigand, & Weigand, 2002; von Hippel, 1994), the policy intent to commercialize knowledge can be thought of as synonymous with the “strategic management of economies” (Audretsch & Beckman, 2007, p. 42). By studying self-organizing communication patterns in innovation ecosystems, scholars may offer prescriptive advice to policymakers regarding the effectiveness and implementation of policies that compliment and/or direct online activity.

Pulling from three literature streams, I focus my theoretical development on open innovation and ICTs, social capital, and the broader innovation literature to explain how individuals access information and resources online. The open innovation literature contrasts closed modes of R&D production occurring solely within an organization (i.e., the “vertical integration” of R&D) to an open stance whereby firms source knowledge and products from outside their boundaries and export knowledge to external consuming entities (Chesbrough, 2006). While not a “novel” theoretical framework by any means (Lichtenthaler, 2011), open innovation is a convenient label for synthesizing related work on “outsourcing, networks, core competencies, collaboration, and the internet” (Huizingh, 2011, p. 3). Few studies have examined SME orientation towards open innovation (van de Vrande et al., 2009), and to this author’s knowledge, no existing work investigates how SMEs use ICTs to further innovation outcomes. In addition, while ample research

studies the content shared and diffused on social media, no systematic evaluation has been conducted to empirically explore Twitter “as a science communication platform from an audience perspective” (Brossard, 2013, p. 14100). In this research, the audience consists of not only the public but also the same set of actors (firms, universities, etc.) that seek to influence and/or grow the ecosystem.

Researchers have long demonstrated that innovation networks are an important determinant of firm performance and capabilities; yet the use of networks and the benefits they confer are contingent on a firm’s stage of development (Rothaermel, Agung, & Jiang, 2007). Implicit in this research is a focus on early activities in the entrepreneurial process: Two critical phases, opportunity framing and pre-organization, are characterized in part by talent shortfalls, low thresholds for risk and uncertainty, incomplete recognition of personal limitations, insufficient access to social capital, and a lack of funding and tangible resources (Vohora, Wright, & Lockett, 2004). “Not much has been written about what entrepreneurs can do to increase social capital or about how social capital can be exploited for new venture creation and development” (Simoni & Labory, 2007, p. 110). Accordingly, this study examines the consequences of social media participation on social capital. It does this by simultaneously considering multiple dimensions of network phenomena (e.g., node attributes and structural features), an area of “future work” recently identified by social and communication network scholars (Monge & Contractor, 2003; Rivera, Soderstrom, & Uzzi, 2010).

I also contribute to the literature on innovation systems and innovative and behavioral additionality. As described by Hobday (2005), recent alternatives to the linear model include interactive, integrated, and networked models, which portray recurring

(yet still stylized) interactions among institutions, markets, and firms. However, this progression of models, from sequential to interactive, lacks empirical evidence. Because it makes “visible the invisible” (Miller, 2011), social media sheds light on a broad cross-section of ecosystem actors, including the media and lay public, in ways that were previously unattainable. By collecting granular communication content and tie data from the individual user, I am able to examine the nature and quality of assumed feedback loops depicted in recent, non-sequential models of innovation. Furthermore, this work draws on the behavioral additionality (BA) literature, which distinguishes between short-term outcomes of innovation policy interventions and longer-term increases in system capacity to support innovation activity, e.g., through sustained learning and collaboration (Georghiou, 2002; Gok & Edler, 2012). I extend BA by applying its theoretical and methodological tenets on an empirical study that transcends firm-level phenomena and that does not rely on top-down, government intervention.

1.5. Dissertation organization

Chapter 2 contains the literature review, and the following short chapter sets the research context by further introducing Twitter and graphene as the case study context. Chapter 4 lays out the research design, data, and method. Chapters 5, 6, and 7 convey the results: Chapter 5 descriptively overviews the sample’s network topographies through visualizations and interpretation. Chapter 6 and 7 present the quantitative (regression) and qualitative results, respectively. Chapter 8 comprises a detailed discussion of the findings, implications, and potential future work; it also concludes with some final thoughts.

CHAPTER 2: LITERATURE REVIEW

The innovation ecosystem is a type of complex network with multiple levels of organizing principles (c.f. Monge & Contractor, 2003). Micro-level relationships are the foundation of this self-organizing system: Individuals generate directed network linkages through following requests. These linkages, when rolled-up to a meso-level, may depict properties not readily discernible at lower levels (e.g., the emergence of communities). Taken together, the set of micro-level linkages and meso-level communities produce a macro-level view of the (online) innovation ecosystem. This research primarily theorizes on the constitutive elements comprising micro-level behaviors, though the results show evidence of ecosystem organizing principles at all three layers of analysis.

This literature review explores concepts and prior empirical work on open innovation, social capital theory, and different types of relevant actor classes currently on social media. The purpose of this framework is to set-up a discussion on why openness matters and how ecosystem actors may leverage social networks to mobilize information and knowledge flows to their advantage. Since open innovation is a framework and not a causal theory (Lichtenthaler, 2011), it directs the review only in terms of its ability to frame the use of ICTs as being important in producing beneficial outcomes. I subsequently turn to social capital theory to consider how micro-level interactions may confer information and knowledge gains.

The relationship between open innovation and social capital theory unfolds as follows (Figure 2.1): On one side (top path in Figure 2.1), the literature on open innovation considers the import of inter-actor interfaces directing innovative activity in a larger ecosystem environment. ICTs facilitate the implementation of an open innovation

strategy with certain beneficial outcomes resulting as a consequence of usage. On the other side (lower path in Figure 2.1), networks enable the organization and coordination of open innovation activities, often through ICT platforms such as online social networks and social media. When actors participate on these platforms, information and knowledge is shared and developed through recombination, albeit in platform specific ways. This process too leads to beneficial innovation outcomes. In essence, Figure 2.1 deconstructs the open innovation framework into a series of network based concepts stressing the importance of network linkages, information and knowledge transfer, and online ICTs. Note that the open innovation literature does not necessarily invoke network scholarship or ICTs, e.g., as in the case when IP is transacted on the market.

The literature review culminates in three informal propositions or areas of inquiry, the first of which posits that online social networks facilitate a type of network topography reflecting strategic interaction within and across actor types. This anticipated behavior is commensurate with innovation management and policy theory describing how different actors are likely to connect with one another based on their information and resource needs. However, online social networks are not only about exogenous actor attributes; they also incorporate a set of usage factors that may influence how users interact. To this end, I consider in the second proposition how information distance impacts the relationship between actor type and network structure. Finally, in the third proposition, I take into account how social media can further a broad set of innovation outcomes relevant to emerging technology commercialization.

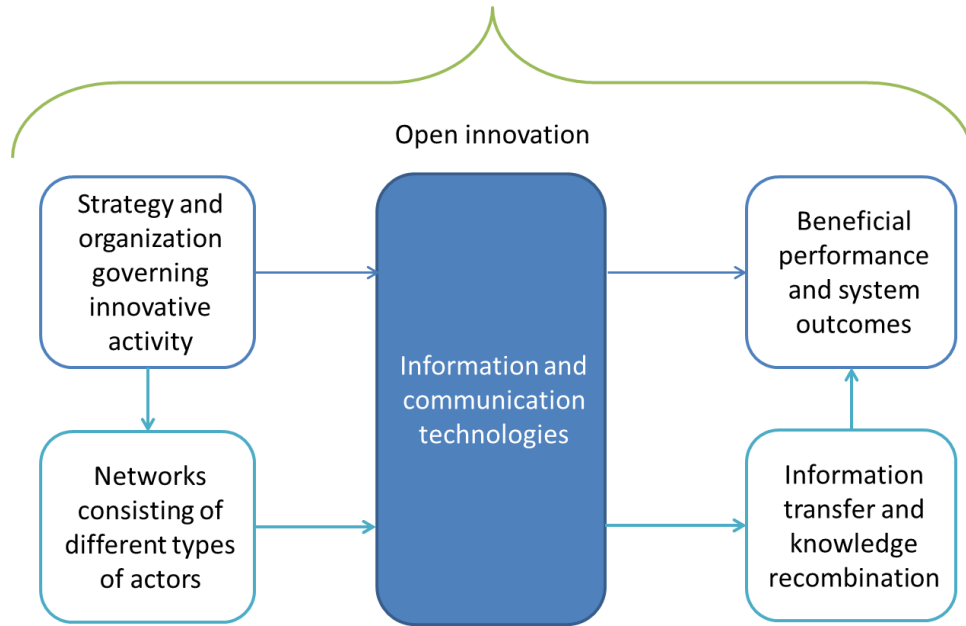


Figure 2.1: Conceptual framework

Section 2.1 covers the relationship between open innovation, ICTs and beneficial innovation outcomes. Section 2.2 goes one layer deeper to describe how networks facilitate the development of social capital, which allows actors to effectively transfer and recombine information and knowledge within and across communities. Section 2.3 reviews several classes of innovation actors expected to participate in the online ecosystem along with their notional information and resource needs. Section 2.4 develops three propositions for empirical testing.

2.1. Open innovation and ICTs

In the innovation management and policy literature, the importance of networks has emerged as a broad and growing research field over the past 25 years (Ahuja, Lampert, & Tandon, 2008). Only in the last ten years, however, has open innovation gained traction among scholars as a unifying framework to organize research around the

sourcing and sharing of ideas and resources (Dahlander & Gann, 2010). Inherent in the open innovation literature is an emphasis on networks and information and communication technologies (ICTs) as foundational mechanisms allowing firms to extend their reach into the larger “innovation ecosystem”. While “openness” is certainly not new as an organizational concept (Scott & Davis, 2006), some of the outcomes assumed as important in the open innovation literature are (Chesbrough, 2006). For example, participation by a variety of actor types is considered as an important indicator of the evolution of innovation activity from in-house R&D dominated by corporate labs to a more permeable interface between firms and their many influencers.

Chesbrough (2003) defines open innovation as “a paradigm that assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as firms look to advance their technology” (p. XXIV). Whereas the prior “closed” paradigm of innovation drew primarily on in-house R&D for new product development, open innovation assumes permeable firm boundaries such that (a) external ideas inform internal R&D practices, (b) internal R&D may be strategically shared with outsiders, and (c) new products can be sourced internally or externally (i.e., via merger or acquisition) or spun-out in new ventures.

In addition to networks, the open innovation research agenda considers other contemporary changes in the socioeconomic landscape, including how new communication technologies enable novel forms of collaboration and how globalization, competition, and the atomization of knowledge encourage specialization and vertical disintegration. Nonetheless, many innovation scholars question the utility of open innovation as a valid theoretical construct, criticizing it as “old wine in new bottles”

because it simply revisits prior work such as absorptive capacity and lead-user innovation (Lichtenthaler, 2011; c.f. Cohen & Levinthal, 1990 and von Hippel, 1988).

Indeed, part of the allure of open innovation as a field of scholarship is that it is a convenient label for synthesizing related work on “outsourcing, networks, core competencies, collaboration, and the internet” (Huizingh, 2011, p. 3). Chesbrough (2006) acknowledges that open innovation is closely linked to the broader R&D literature but argues that the paradigm offers several new contributions as a research stream. First, scholarship on open innovation directly addresses the question of how much external sourcing of ideas and resources is optimal vis-à-vis internal processes. Second, the business model takes center stage here; in the previous paradigm, outputs from in-house R&D fed directly into new product development, whereas in open innovation, alternative business models include licensing, spin-outs, and acquisitions. Third, as knowledge becomes increasingly diffuse, less costly to access, and of higher quality, firms realize efficiencies in working and learning symbiotically with external entities. Fourth, with vertical disintegration and a greater intensity of interaction across network linkages, intermediaries play a more visible and important role in brokering connections and facilitating transactions. Finally, as with the approach to studying innovation ecosystems, research on open innovation requires a new set of metrics that address the peculiarities of the framework.

Open innovation research may incorporate varying units of analysis, including the firm, organization, dyad, interorganizational network, or encompassing innovation system (Vanhaverbeke, 2006). Recent work by Cheng and Huizingh (2014) shows that open innovation activities positively affects firm performance on four dimensions, in

order of increasing impact: new product/service innovation, financial performance, new product/service success, and customer performance (i.e., customer satisfaction and loyalty). The authors explain this finding by noting that broad searches are especially effective in producing out-of-the-box thinking and therefore in generating non-incremental innovations. In addition, this study tests for the moderating effect of entrepreneurial orientation, which captures the extent to which a firm is proactive in its competitive approach to market, industry, and governmental dynamism. The results show that firms with high levels of entrepreneurial orientation and open innovation activities are more likely to experience positive performance outcomes vis-à-vis firms with lower levels of either measure, suggesting that “opening up” is most valuable when a firm subscribes to a responsive and nimble strategic disposition.

While business models define the mechanism of “value capture” (Teece, 2010), the use of these models in a heavily networked context necessarily impacts and requires input from other actors (Thomas & Autio, 2012; Vanhaverbeke, 2006). This is particularly important for the open innovation framework, which incorporates a wider variety of actors than more conventional models of innovation (e.g., the linear, innovations systems, or triple helix models). For instance, Lee et al. (2010) provide case study evidence that intermediaries are becoming increasingly important in SME networks to facilitate faster commercialization times, to encourage cooperation among SMEs (instead of with large firms, where power asymmetries may disadvantage the SME), and to allow SMEs to focus on building core competencies. The current research views network interactions in an even broader way by considering ecosystem relationships

among not only SMEs, established companies, and intermediaries but also scientists, the media, financial firms, and the general public.

The question of how new actors emerge on ICT platforms is addressed by Vaast et al. (2013). They argue that in innovation contexts new actors are a source of disruption, particularly in light of their ability to shape discourse around a specific set of technologies. In their study of Web 2.0 innovation, the authors find that new actors provide additional insight and perspectives that may conflict with traditional actor viewpoints: Novel contributions are mediated by the collective's within-group's ability to develop a unique, legitimate, and sustained identity. Identity "coalescence", however, is challenged by within-group fragmentation (where for example disparate sub-groups compete to set the larger collective's agenda) and dispersion (when the line separating the new actor type from other established actors blurs). Vaast et al. note limitations of their work as applying to only Web 2.0 innovation, and they suggest that scholars study new forms of ICT-mediated discourse and actor participation in other "professionalized fields."³

Vaast et al. (2013) write, "Discursive practices are intricately tied to the technical features of the new media, and different new media are likely to result in different discursive practices" (pp. 1087). Subsequently, the discourse that develops on micro-blogs such as Twitter is likely to be much different than the discourse facilitated by (standard) blogs. Moreover, as new media changes in its technical design and user adoption, the new actor category also evolves. In other words, ecosystem discourse shifts

³ This research studies Web 2.0 enabled innovation on social media in the professionalized field of nanotechnology R&D.

not only with substantive content changes and fluid participation but also with shifts in the underlying ICT platform.

In general, ICTs play an important role in the open innovation framework as enabling the types of interactions that result in network-driven, innovation outcomes.

Brynjolffson (2010) remarks:

Information technology is... a catalyst for complementary changes: It's what economists call a "general purpose technology" that sets off waves of complementary innovations in things like business processes, new ways of reaching customers, new ways of connecting to suppliers, internal organization to the company. These complementary changes are often 10 times as large as the size of the initial investments in the IT itself and have profound and long-lasting effects on our ability to create goods and services.

But there's a factor that has not been studied very much and, frankly, is not very well understood. And that is the possibility that IT can change the innovation process itself [sic] (Brynjolffson, 2010).

To understand how IT influences the innovation process itself, we must first consider the role of ICTs and knowledge in the "digital era". Stone (2004) argues that ICTs and knowledge embodied in human capital characterize production: Profits are often a function of intellectual property and other intangible assets, and ICTs both coordinate internal firm activities and provide an interface between the firm and its external environment. As such, today's firms value employee knowledge in three forms, including 1) technical expertise; 2) familiarity with firm business processes and operations; and 3) the ability to recognize and discard ineffectual approaches to problem solving, product development, etc. The cognitive processing capabilities of a firm's employees are likely to hinge on the quality and diversity of information sources (Cohen & Levinthal, 1990). This is because learning about product development opportunities,

consumer preferences, technological performance, and industry adoption rates all impact organizational decision making (Loch & Huberman, 1999; Lynn et al., 1996).

Recent work shows that ICT adoption furthers open innovation outcomes, but that selection processes determine ICT adoption patterns: Saldanha and Krishnan (2012) find that firm size and degree of knowledge intensity by industry positively predict Web 2.0 technology adoption: Larger firms face greater coordination challenges, which Web 2.0 technologies help address, than smaller firms, and firms in knowledge-intensive industries benefit from codifying tacit knowledge and promoting collaboration, both of which are also enabled by Web 2.0 platforms. However, the mean firm size in Saldanha and Krishnan's sample approaches 5-10 thousand employees, leaving much to be explained with respect to small firm adoption.

Other empirical research reveals a broad set of open innovation outcomes as a result of ICT adoption. Dodgson et al. (2006) coin the term "innovation technologies" to describe how Procter & Gamble uses tools to facilitate communication across knowledge communities internal to the firm and to find, explore, and test ideas (e.g., via simulation and modeling, data mining, and virtual prototyping). The result is that by 2004 P&G was able to source up to 35% of its innovations externally in part due to innovation technology adoption (as well as other organizational changes), an increase from a historical rate of ~20%. Using a case study approach to studying entrepreneurship on social media, Fischer and Reuber (2011) report that Twitter, through its social network capabilities, provides entrepreneurs a tangible source for identifying and developing critical resources, relationships, and business opportunities. In their case study research of two large, high-technology multinationals, Ooms et al. (2015) find that social media

improves inbound knowledge flows into an organization by increasing availability and access to diverse users and communities.

In sum, open innovation is an emerging field of research that offers a unifying framework for understanding inbound and outbound flows of resources and knowledge that traverse through networks to produce innovation outcomes. ICTs facilitate these flows. However, few studies examine the intersection of ICTs, open innovation, and SME-based networks, and moreover, what work does exist focuses primarily on the firm as a unit of analysis (e.g., c.f. Ooms et al., 2015; van de Vrande et al., 2009). This focus helps researchers isolate the effects of open innovation on firm outcomes (e.g., performance) but does little to account for how and why other actors come into connection with each other. As a result, while the locus of open innovation necessarily resides outside any given firm, scholarship has often overlooked the “community” as a unit of analysis (West & Lakhani, 2008).

Because open innovation is a framework and not a theory per se (Lichtenthaler, 2011), a host of other literature fills in theoretical gaps to explain micro-level foundations. To this end, the next section reviews background concepts on social capital to explain how networks facilitate the exchange of knowledge and resources both in dyadic and in community contexts.

2.2. Social capital and information and knowledge as resources

This section and the next section lay the foundations of a micro-level theory of social capital by considering why individuals create linkages with others and by outlining some key benefits with respect to information and resource access and mobilization.

Underlying this theoretical development is the assumption that, compared to other forms

of organization such as markets and hierarchies, networks provide a (more) efficient medium by which intangible assets, such as information and knowledge, can be transferred (Powell, 1990). A detailed discussion on actor class motivations and behaviors is deferred until the next section, but here I examine how structural positions in networks lead to distinct types of outcomes in terms of information access and resource mobilization. The analysis primarily draws on the economic sociology literature, but it also acknowledges that ICT platform-specific design criteria influence how social capital accrues to individuals in online settings.

Social capital is often contrasted with two other types of capital, physical and human capital (Coleman, 1988; Kadushin, 2011; Lin, 2001). Physical capital conveys surplus and investment properties: Investments in physical capital (e.g., machinery) yield profits earned on transforming inputs to outputs through production processes (Lin, 2001 citing Marx). Human capital, on the other hand, constitutes investments in individual competencies, such as knowledge, skills, and aptitude, as positive predictors of individual, organizational, and societal outcomes (Nordhaug, 1993). Human capital is particularly salient in high-technology domains where increasing levels of knowledge specialization require longer periods of education and larger team sizes (B. F. Jones, 2009). Additionally, information asymmetries and competing incentives between business and technology managers require scientists and entrepreneurs to traverse multiple landscapes (Auersald, 2007), suggesting the import of human capital in the form of generalized business and people skills (R. A. Baron & Markman, 2003).

Social capital differs from other types of capital in that it is primarily accessed and conferred through network relationships (Coleman, 1988; Lin, 2001; Monge &

Contractor, 2003). Lin (2001) defines social capital as the “resources embedded in a social structure which are accessed and/or mobilized in purposive actions” (p. 35). Three components constitute Lin’s framework: embeddedness, capitalization, and effects (outcomes) (Figure A.2 in the Appendix). Embeddedness includes collective assets (such as trust, norms, etc.) and structural and positional variations, both of which influence capitalization, or accessibility of network locations and resources. Accessibility enables mobilization – i.e., the use of contacts and their resources – and both accessibility and mobilization result in *instrumental* and *expressive* outcomes. Instrumental effects include resources the individual does not already have, while expressive effects concern resources that the individual would like to maintain and preserve. This causal approach implies that inequalities in embeddedness levels lead to variations in capitalization and thus diverging outcomes. Rarely do scholars measure investments and actual resources, however; the theory is usually tested with *potential* access to resources and perceived advantages as outcomes (Kadushin, 2011).

Resource dependency and exchange theories help explain the creation and maintenance of network ties over time. *Exchange theory* suggests that actors exchange resources because of resource needs and largess; furthermore, they enter into trade relationships to maximize *power* derived from their network position and to minimize their dependency on (and exclusion from) trade partners (Monge & Contractor, 2003). *Resource dependency theory*, conversely, posits that actors are more likely to structure their ties in networks to guard against environmental uncertainty (Street & Cameron, 2007). This occurs through two mechanisms: network extension, where organizations seek to develop new ties, and network consolidation, where past (positive) interactions

predict future relationships (Monge & Contractor, 2003). Recurring communication positively impacts likelihood of persistent links, even after the value of resource exchange declines.

Two types of *resources* available via network interactions include information and knowledge (Phelps, Heidl, & Wadhwa, 2012). Some scholars treat information and knowledge synonymously, yet there are important differences between these two concepts. Information communicates data, or empirically observed (subjective or objective) facts, in a coherent manner (Aamodt & Nygård, 1995); information is data *with meaning*, encapsulated in a message, and therefore alters perceptions about the world and one's knowledge. High quality information informs decision making by acting as a catalyst in problem solving; people search selectively and incompletely for information to determine alternatives to and possible consequences of their decisions (Choo, 2006; B. D. Jones, 2003; Newman, 2010; Simon, 1997). Information is particularly important in science-based domains because of the evolving nature of the technology, because of information asymmetries between different types of actors (e.g., the entrepreneur and scientist), and because of uncertain market environments (Audretsch et al., 2007).

Knowledge, on the other hand, is considered as the tacit, applied form of information, e.g., in problem solving contexts (Zins, 2007). In addition, knowledge is often conceived of as an output of learning and informs expectations about the world (Choo, 2006). When communicated from one party to another, information leaves traces behind for empirical observation (e.g., through artifacts such as email and news media). While this may also be true for knowledge, highly complex information stored as know-

how is difficult to describe and transfer, suggesting a less discernible and codifiable imprint than information per se (von Hippel, 1994).

If we can think of the innovation ecosystem's online footprint as a marketplace for ideas, the ability to access, recombine, and create new knowledge is an important outcome of social capital development. Nahapiet and Ghoshal's (1998) work offers a theoretical lens from which to view this process (Figure A.3 in the Appendix). The authors distinguish between three types of social capital, structural, relational, and cognitive. Structural aspects of social capital relate to network tie formation and whole network configuration, whereas relational capital concerns trust, norms, obligations, and identification (e.g., within an encompassing homogenous group). Cognitive capital captures the extent to which two or more parties share a common code, language, and narrative. While social, relational, and cognitive capital interact with one another, they also establish a framework for the creation of *intellectual capital through the combination and exchange of existing knowledge*. Comparing Lin's path diagram to Nahapiet and Ghoshal's, we see that the construction of intellectual capital through social capital is a more specific case of Lin's generalized framework. With the development of intellectual capital (i.e., knowledge), organizations are able to develop other sources of competitive advantage, assuming that knowledge is the basis for other innovation outcomes, such as patenting and new product development (Kogut & Zander, 1992).

Once network ties develop at the micro-level, meso-level structural phenomena materialize at an aggregate level. Two camps emerge here, distinguishing between social capital accruing at the individual and network levels (Kadushin, 2011). The first camp argues that social capital develops primarily through dense and repeated connections.

Coleman (1988, 1993) and Putnam (1995) discuss social capital in terms of the development of common norms and trust which facilitate low-cost monitoring and robust civic engagement. In particular, Coleman's (1988) view of social capital emphasizes the benefits of an ego's placement in a densely connected group of actors (i.e., alters). Dense clusters result from high levels of triadic *closure*, wherein ties from alter to alter supplement ties emanating from ego; that is, if B is connected to A and C, then A is also connected to C. Closure is assumed to result in lower costs of monitoring and sanctioning because opportunistic behavior is 1) less likely to go unnoticed given network density; 2) more likely to be punished by immediate actors; and 3) less likely to occur because of shared norms. Because everyone knows everyone else, closure contributes to reputation building (Coleman, 1988).

Social capital as closure appears in the economic sociology literature as "embedded networks", which present an economic context for transacting within the boundaries of *well-established relationships* (Granovetter, 1985). Uzzi's (1996, 1997) qualitative and quantitative studies of New York's fashion district show that embedded networks exhibit three important characteristics: (1) an ability to develop trust beyond contractual obligations. Individuals can fall back on heuristics, thereby saving resources and time; (2) fine grained information transfer, which conveys bundles of information beyond price alone through a chunking mechanism. The information is detailed, tacit and holistic (rather than divisible) in terms of conceptual substance; and (3) joint problem solving, which allows for both dynamic *and* ad-hoc problem solving. Uzzi contends that people search deeply within relationships rather than broadly across different relationships.

The second camp of theorists considers social capital primarily as a theory of self-interest (Lin, 2001; Monge & Contractor, 2003). In contrast to emphasizing closure and overall network density, Granovetter's (1973) seminal work stresses the benefits of weak ties that connect densely connected clusters of actors. Diverse information traverses across these links and enables individuals to access resources that would otherwise be unavailable. Social capital as a theory of self-interest often adopts brokerage and structural holes to explain how benefits accrue at the individual level. A broker in essence connects two different (i.e., non-redundant) groups that would otherwise not be connected without the bridging relationship. A structural hole exists when B is connected to A and C, but A and C are not tied.⁴ In this example, B is a broker between A and C and may exploit the structural hole to access and filter information by sharing and withholding information in potentially strategic ways. In addition, B is also the first to know new content originating from either A or C and therefore may enjoy referrals and high visibility among other actors (e.g., by being sought out for information before others are approached) (Burt, 1992). Burt (2004) finds that individuals spanning structural holes are more likely to generate and be rewarded for subjectively perceived "good ideas" in part because they bridge distinct specialty areas. Yet, it is unclear whether actors always choose to manipulate information flows in structural holes when given the opportunity (Monge & Contractor, 2003).

While brokering often results in returns to the individual in terms of an information or reputational advantage, the organization may benefit from employees'

⁴ In structural hole theory, A and C belong to distinct clusters. Multiplexity – multiple relational dimensions – in networks suggests multiple types of relationships between actors. In this context, the network is viewed only in one dimension.

boundary spanning behaviors, too. In Cohen and Levinthal's (1990) seminal paper, the concept of absorptive capacity refers to a firm's ability to exploit prior held knowledge via its recognition, assimilation, and application of new knowledge. Sources of absorptive capacity include (1) cultivating knowledge and learning in an-house manner, such as through R&D investments and capabilities, (2) by hiring expertise from outside the firm, and (3) *via communication systems and networks of organizations and individuals*. The appropriate balance of one source vis-à-vis another is determined by how tacit and firm-specific the knowledge is: Gilsing et al. (2008) find that cognitive and technological distance between two actors is curvilinear in relationship to the successful exploitation of the network linkage (i.e., too much or too little distance is counterproductive).

Continuing this line of reasoning, if online communication provides codified, generalized information, it seems likely that communication with external entities may provide a source of valuable environmental information. And by definition, if ecosystem communications occur in one substantive ST&I domain, then cognitive distance may never reach unproductive limits.

Both types of social capital (i.e., relational and structural) may exist within and across communities in a network, sometimes simultaneously. Burt (2001) argues that actors within a "group" (i.e., community) can cohesively create dense linkages internally, thereby facilitating closure, trust, shared norms, as well as enabling a capacity to self-monitor and implement sanctions for deviant behavior. Linkages across groups, in contrast, create structural holes and the information advantage characterized most aptly by Granovetter (1973) and Burt (1992). Burt (2001) shows evidence supporting better performance as a result of fewer network constraints (i.e., fewer dense linkages), but he

goes on to theorize that within group closure in conjunction with boundary spanning ties may result in unparalleled performance outcomes. This claim supports Hansen's (1999) results showing weak ties accelerate search processes but impede the complex transfer of information; organizations may wish to design internal structures that facilitate both search and transfer processes simultaneously depending on the complexity of the knowledge involved.

I contend that online platforms alter the dynamics of social capital formation and maintenance vis-à-vis offline networks. From a purely structural standpoint, closure resulting in dense communities and brokerage across weak ties resulting in structural holes are possible in the online ecosystem. However, outcomes realized through different structural arrangements may depend on the platform's design. That is, on online platforms, it does not necessarily follow that closure and dense communities result in high levels of trust, reputation building, monitoring, and sanctioning; it also does not follow that brokerage and structural holes result in high levels of social capital via reputational and gatekeeping advantages. For example, Ganley and Lampe (2009) study social capital on Slashdot, a technology centric news and commentary site and find that closure ("network constraint") predicts higher levels of "karma", or a social capital measure of a user's contributions to the site. To the surprise of the authors, however, between-ness negatively predicts karma.

On Twitter, a platform with significant character constraints, strong ties via closure may never result in the complex transfer of knowledge, as Uzzi (1996, 1997) and Hansen (1999) might predict. In addition, as Kielser et al. (2012) note, anonymity, a proliferation of pseudonyms, and a subsequent inability to sanction, may erode trust and a

sense of shared norms, impeding any sense of embeddedness. Nevertheless, the platform may facilitate the development of weak ties and structural holes, which could facilitate the search for information across diverse topical areas.⁵

Recent work shows that Web 2.0 technologies and social media can indeed facilitate the discovery of useful information. Ashurst et al. (2012) find that strategic use of Web 2.0 platforms, including social media, allow firms to effectively *explore* and/or *exploit* resources and business opportunities for product or process innovation. Exploration refers to the discovery of new concepts, innovations, and business opportunities outside of established organizational routines, whereas exploitation is the process of achieving operational efficiencies of known certainties that are, for example, most likely to maintain short-term revenue streams and profitability (March, 1991). In a study of Twitter in Korea, Choi et al. (2011) compare four innovation-oriented communities with four non-innovation communities and find that members of technology-based ecosystems are more open, less geographically defined, and share more awareness of issues than their counterparts.

In sum, the literature on social capital asserts that network structure and relational assets matter for the mobilization of resources, which include information and knowledge. Network structures form over time in such a way that enhance the informational advantage of actors, in terms of novelty and diversity of information, as well as an ability to synthesize “good” ideas (Aral & Walker, 2012; Burt, 2004). Yet, relational elements such as reputation, trust, and shared norms facilitate the transfer of complex knowledge that information exchanges alone cannot (Hansen, 1999; Uzzi, 1996,

⁵ Refer to Chapter 3 for a more detailed discussion of Twitter’s design and usage features.

1997). That said, in public social media contexts where participation is inexpensive and therefore fluid – and where reputation, trust, and shared norms can be challenging to normalize – it is unclear whether actors experience any returns other than improved awareness and an enhanced ability to explore the environment.

2.3. Ecosystem actors

The innovation ecosystem consists of a broad cross-section of actors who, while acknowledged in the open innovation literature, are rarely studied at the same time. A possible explanation of this limited empirical research stems from the magnitude of actors' diverse interests and motivations and the difficulty that arises with articulating a set of testable propositions by which theory can be explored, refuted, or confirmed.

The full range of ecosystem actors includes a set of *primary* actors as identified by the innovation literature (e.g., Corsaro, Cantù, & Tunisini, 2012; Leydesdorff & Meyer, 2006; Lundvall, 2007; Neal et al., 2008): established (larger) firms, entrepreneurs and high-technology SMEs, financial professionals, scientists, and intermediaries. Cooke (2001) also identifies a cadre of supporting agents (e.g., lawyers, consultants, and accountants) that facilitate the “exploitation and commercialization of scientific findings” (pg. 962). Additionally, because social media is inherently a communications channel with strong media representation, I add to this list the media entity, which may be traditional or new in its choice of channel and generalist or specialist in focus. Also included within the scope of ecosystem actors is “other users”, which may include hobbyists and individuals not ostensibly affiliated with an institution.

Under ideal conditions, each actor in the ecosystem is conceived of as interacting symbiotically with other actors in terms of (a) specialization, (b) complementarities

within that specialization, and (c) the ability to co-evolve based on industry and technological progress (Thomas & Autio, 2012). Each of these actor types are described below and are compared and contrasted across two dimensions, by (1) anticipated information needs and motivations, as informed by theory and empirical research, and (2) expected information behavior with respect to social media usage. Table 2.1 summarizes this orientation for each actor type.

Established firms and entrepreneurs. Small firms differ from large firms in several ways, and these differences influence small firms' orientation towards network participation, open innovation, and social media usage. Most notably, large firms enjoy wide and deep resource bases, thus making the commercialization of innovations operationally easier than for the resource strapped SMEs (Narula, 2004; Rothwell, 1989). In contrast to the large firm, the entrepreneur is a risk seeker and investor of time and capital: In high-technology domains in particular, entrepreneurs seek to exploit inefficiencies in transaction costs; they evaluate opportunities and “internalize externalities [in order to] solve complex coordination problems” (Auersald, 2007, p. 24). In terms of competencies, evolutionary scholars assume that large firms are more adept at “scaling up” and achieving mass economies of scale than their small firm counterparts (Nelson & Winter, 1982), but smaller firms are more agile and responsive to changing environmental conditions (A. Arora & Gambardella, 1994). This is what Rothwell (1989) terms the SME *behavioral* advantage (as opposed to the large firm's *material* advantages).

Table 2.1: Ecosystem actor types, information needs, and expected behaviors

Actor type	Information needs and motivations	Expected information seeking behavior in the online innovation ecosystem
Established large firms	Brand management, ideation, monitoring and/or leading of the ecosystem environment	Network with customers, suppliers, and employees. Consume and produce information
Entrepreneurs and SMEs	Reduce uncertainty introduced by information asymmetries, environmental dynamism, and unclear goals	Network widely across actor groups on social media to access diverse information and complimentary assets. May or may not be a producer of information
Supporting actors	Seek to understand the market and find a niche for their services	Aggressively network with potential clients (e.g., large firms and entrepreneurs and SMEs). Primarily produces but also consumes some information
Finance professionals	Discovery of emerging technologies, product and market potential, and promising start-ups	Network with media entities and thought-leaders to stay abreast of technological and market developments; network and engage with promising start-ups
Scientists	Access literature, disseminate findings, participate in online meetings, discover peers	Network primarily with other scientists, media entities, and intermediaries. Produce and consume of information
Intermediaries	Build entrepreneurial capacity, act as clearinghouses for information, increase participation, and confer reputational benefits	Network with other ecosystem actor types as soft brokers (i.e., not economic brokers). Produce and consume information
Media actors	Work under time constraints to access sources and prepare a narrative of ST&I events	Network with other ecosystem actors to prepare and diffuse storylines. Primarily produce but also consume information
Unaffiliated “other” actors	Unknown personal and/or professional interests and motivations	Unknown

Resource and capability constraints lead small firms to seek complementarities in network settings, and this trend is increasing with globalization (Autio, 1997; Lee et al., 2010; Narula, 2004). In open innovation contexts, SMEs appear more likely to view external sourcing through networks as being more important in the (later) commercialization stages of product development than during the initial R&D stages because of the ability to access in-house intellectual property (Lee et al., 2010). In other words, complementarities are lacking in the later stages of the commercialization process, including large scale production, marketing, and distribution, and therefore network participation is focused on addressing these deficiencies. Using survey data, van de Vrande et al. (2009) provide evidence that SMEs are more likely to employ networks for open innovation purposes when those practices require less commitment: “The more popular practices like customer involvement and external networking are informal, unstructured practices which do not require substantial investments. IP licensing, venturing and external participation on the contrary, require financial investments, formalized contracts and a structured innovation portfolio approach to manage the risks” (p. 434).

Aslesen and Freel (2012) report that more than 50% of firms in their sample engage in inbound open innovation through the Internet, though high-tech firms in particular are more likely to resort to other channels of information including suppliers, customers, universities, and technology centers. (Recall that inbound open innovation refers to the sourcing of new ideas and resources.) The authors suggest that the Internet is less amenable to conveying tacit knowledge, and therefore, the Internet is more appropriate as an information source for low-tech firms. Based on this review of the

literature, it appears external sourcing of R&D on social media is unlikely due to the public nature of the Internet, limitations of online communication in transferring tacit knowledge, and the broader trend for SMEs to seek later stage and lower commitment open innovation activities.

According to organizational theorists, uncertainty is a primary driver of information seeking behavior (Case, 2012; Choo, 2006), and I contend that social media may be helpful to SMEs in this broader context. Firstly, information asymmetries contribute to uncertainty when high-tech entrepreneurs suffer from incomplete information about their customers, suppliers, and funders (Audretsch et al., 2007). Environmental dynamism is another source of uncertainty and may stem from an evolving industry structure, technological landscape, and market demand (Hough & White, 2003; Schilke, 2013). Finally, entrepreneurs in particular may lack clarity about their goals and instead elect to mobilize their resources (including network based resources) to *effectuate* an outcome (Sarasvathy, 2001). Outcomes here are not optimized but are rather selected as a byproduct of the mobilization process.

Yet, high quality information does inform better decision making (Newman, 2010). Uncertainty generally rises and then falls during the information seeking and search process, and as understanding and meaning develop, confidence, interest, and motivation to act increase accordingly.⁶ In comparison to scientists and engineers, however, managers must respond to pressing problems – i.e., their timeframe for responding is much shorter – and they rarely have an opportunity to read in detail (Case, 2012). Social media may then act as a way for managers to reduce information

⁶ This is the stylized application of information as knowledge argument introduced above in Section 2.2.

asymmetries, monitor environmental dynamism, and articulate their goals with the expectation of iterating in the effectuation process. Implicit in this supposition is a propensity for managers and small firms to network broadly across groups to access information through a social network structure⁷, without having to invest considerable time in any of these activities.

In contrast to small firms, larger firms may be more likely to use social media in early stage idea generation. Dahl et al. (2011) report that firms employ social media to broaden the scope of participation; for example, businesses may direct ideas generated by consumers into R&D product funnels. In addition, MNCs are particularly likely to operate multiple accounts on the same platform (e.g., Twitter) for separate business units, business functions, or geographic locales to manage and market their brand and to facilitate customer interaction from an operational perspective (Burton & Soboleva, 2011). It may also be the case that large firms embed themselves into an online ecosystem in much the same way as they do in (offline) innovation systems, e.g., where they may act as a central, coordinating force in a regional supply chain (Christopherson & Clark, 2007a, 2007b; Sternberg & Tamásy, 1999). In this type of network setting, large MNEs maintain a disproportionate share of power vis-à-vis smaller firms.

While social media is most visible in B2C channels, it is also evolving in B2B markets; challenges in the B2B arena include the perceived irrelevance of social media, unfamiliarity with the technology, and a lack of causal understanding as to how social media can directly support brand development (Michaelidou, Siamagka, & Christodoulides, 2011). Based on a survey of 351 company executives in eight countries,

⁷ Refer to the discussion in Section 2.2 relating social capital development to information and knowledge development advantages

Baird and Parasnis (2011) find that 79% of companies have a presence on social networking sites, 55% on media sharing sites, and 52% on microblogging sites.

Supporting actors. SMEs and entrepreneurs rely on the services of supporting actors, such as consultants, designers, software engineers, lawyers, and accountants, to access complimentary assets not available within the firm (Muller & Zenker, 2001). Rather than being influenced by highly uncertain environments, I contend that these supporting actors work to decrease uncertainty⁸ by for example helping SMEs better navigate intellectual property regimes, accounting rules, and the broader ICT landscape. It is due to this difference in motivation that supporting agents are introduced as a separate actor type than high-technology SMEs and large firms, respectively.

The practitioner literature has tracked the rise of the social media consultant, a profit seeking individual (or small firm) offering services to help clients develop an “open” and “transparent” online presence (Samuel, 2009). A typical argument for a client’s social media participation is to “get closer to the customer” and to better initiate the front-end of the sales cycle via lead identification and relationship management (Giamanco & Gregoire, 2012; Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). According to a 2013 Entrepreneur article, for the social media consultant, one way to demonstrate marketing know-how and effectiveness is to maintain a successful personal social media presence, e.g., by sharing valuable content and responding promptly to inquiries (Shandrow, 2013). Another qualification of a competent social media

⁸ Muller and Zenker (2001) present a conceptual framework wherein knowledge-intensive based services (KIBS) firms codify and apply knowledge in manufacturing SME contexts: KIBS firms problem solve on behalf of their clients and in the process generate interaction-based knowledge outputs. Working across different clients promotes the recombination of knowledge and can for example guide client decision-making based on industry best practices.

consultant is to understand a prospective client's market. Thus, since consultants (and other service providers) are in the business of attracting and maintaining clients, I expect these actors to aggressively network with ecosystem firms in such a way as to favorably market their competencies *for the purpose of enhancing their own sales cycles*.

Financiers. Limited research exists on the information needs and behaviors of professionals in financial institutions: through a case study design, Miranda and Tarapanoff (2008) show organizations continuously gauge risks and valuations of internally controlled assets and liabilities vis-à-vis similar external market assessments. Information technologies facilitate the routinization of some monitoring tasks and consequently direct the attention of skilled individuals to problem areas on an as-needed basis. Huvila (2010) examines another sample of (corporate) financial professionals and finds that selection of information sources is a function of perceived effectiveness. Information sources are diverse and include commissioned reports, market information, general and financial statistics, newspapers, conferences, alerts, and websites.

Organizations and individuals seeking to *invest* private funds into new venture creation use several decision making heuristics. Angel investors, for instance, are often themselves wealthy entrepreneurs or experienced business professionals, who invest in promising ideas with expectations for moderate returns in a 5-10 year time horizon (Amatucci & Sohl, 2007). Venture capitalists, on the other hand, may expect a 40% (or more) rate of return and significant control over a firm's operations, management team, and strategy. Research shows that venture capitalists routinely fund a low percentage (e.g., ~2%) of all requests received and consider a venture's management team, market, product, and financial potential when deciding which firms to invest in (Petty & Gruber,

2011). In addition, internal factors, such as the current portfolio of investments, also provide a salient input into the decision making process. The decision of whether to invest in a particular company is therefore the outcome of a complex set of selection rules. Yet the pool of potential candidates for investment can be thought of as an exploration and discovery-driven process which may in part be facilitated by open innovation technologies. With respect to product and market potential, venture capitalists and other financial professionals may monitor social media to assess a technology or company's promise or even to solicit pitches for funding.⁹

Scientists. When not teaching, writing or reviewing proposals and manuscripts, and serving in a professional capacity, scientists focus on producing research that conforms to a generally accepted scientific method. Intuitively, the process by which new scientific achievement advances may not be amenable to 140 or fewer characters or of text in Twitter. In addition, increasing pressure to publish before the competition and to commercialize research outputs limit certain types of information and data sharing (Neal et al., 2008; Shibayama, Walsh, & Baba, 2012; Van Noorden, 2014). Scientists' information needs lie at the frontier of knowledge, and conferences, databases and journal literature, and interpersonal relationships all constitute conventional sources of information. Bichteler and Ward (1989, cited in Case, 2012) find that a sample of 56 geoscientists spent an average of four hours searching for information, and that interpersonal sources of information were most often used followed by the journal literature.

⁹ Venture capitalists are online; see <http://blog.hubspot.com/blog/tabid/6307/bid/9273/30-Most-Influential-Venture-Capitalists-on-Twitter-VC-Remix.aspx> for a list of influential Twitter venture capitalists.

Turning to peers as a primary source of new information can also be seen in the context of modern ICTs, which enable interpersonal communication largely through social networks. As Bonetta (2009) reports, by 2009, some early adopting scientists were using Twitter to publicize their work, access relevant literature, disseminate findings, and generally promote public engagement in a spirit of “open” science (c.f. Bonte, 2011). By 2011, some scientists were participating in online Twitter meetings with agendas ranging from new work to operational aspects ongoing research (Reich, 2011), while others were using the platform to comment on papers post-publication (Mandavilli, 2011).

A 2014 survey published in *Nature* indicates that scientists are more likely to use social media to discover peers, post work content, and follow discussions than to discover jobs, track online metrics, and contact peers (Van Noorden, 2014). In addition, scientists use social media in ways that differ sharply by platform. For example, while almost 50% of the 3,000 respondents (not including researchers in the social sciences and humanities) report using ResearchGate, a social networking platform exclusively for researchers, less than 15% use Twitter. Some platforms promote certain types of behavior, thus suggesting self-selection into one mode of usage vs. another. For instance, “half of the Twitteratti said that they use it to follow discussions on research related issues, and 40% said that it is a medium for ‘commenting on research that is relevant to my field’ (compared with 15% on ResearchGate)” (Van Noorden, 2014, p. 127).¹⁰

This review of the literature suggests scientists’ usage of social media is largely confined to interacting with other scientists, media entities, and intermediary institutions

¹⁰ This survey does not address how scientists interact on social media to further innovative outcomes (e.g., how they interact with high-technology firms), however.

(e.g., professional associations) to access information and to expand the scope of their networks. That is, although the ecosystem metaphor emphasizes interconnectedness among different actor types, there is scarce evidence that scientists as a whole are actively observing or engaging with downstream commercialization of research by firms, as revealed through their social media activity.

New and traditional media players. The production of news content necessarily relies on information seeking behaviors: Journalists seek information from a broad array of sources to contextualize events and to devise and communicate the “news” in a narrative format (Case, 2012). However, unlike scientists who have the luxury of contemplating the implications of events on theory and research, journalists deploy pre-conceived frames of reference, which facilitate timely reporting but sometimes at the expense of objectivity. The transition to online news production has resulted in several key developments, including (1) faster news cycles threatening methodical and credible reporting; (2) a blurred distinction between the gatekeeping role of conventional journalism and user-generated content; and (3) a “convergence” across media types, e.g., print, webs, blogs, and social (Kolodzy, Grant, DeMars, & Wilkinson, 2014; Mitchelstein & Boczkowski, 2009).

In the online realm, media entities release new information, amplify trends, and direct traffic from social media platforms to their websites via hyperlinks. Presumably many specialized sources rely on an open-access, advertising business model to generate revenues (c.f. Gallagher, Auger, & BarNir, 2001 for other relevant business models). Increased visibility on social media sites generates traffic, which in turn increases the likelihood of a user clicking on a sponsor’s advertising, thus producing revenue for the

online site. Colliander and Dahlen (2011) note that online publicity can result in noteworthy gains for firms via increased sales and enhanced reputation. Consequently, a trade journalist or media entity promulgating scientific achievements and leading-edge industry research and development may be perceived as an authoritative source of information.

For the purposes of this work, a media entity can be designated as “traditional”, (e.g., *The New York Times*, *BBC*, and *Science Friday*) maintaining print, radio, and/or television outlets, or “new” with a primarily online presence (e.g., *Graphene Times*). Media entities can also be designated as generalist or specialized in focus, with a possible correlation between traditional sources and general content and between new media sources and more specialized content. The basis of this claim lies within the proliferation of new media entities serving the “long tail” of readers’ information needs and interests (Carpenter, 2010): With the falling cost of publication, more esoteric information sources satisfy the information preferences of an increasingly fragmented audience. Large, established, and generalist media outlets may enjoy larger readership audiences than their newer and more specialized counterparts, and controlling for this size and reach differential is an important component in assessing ecosystem dynamics.

Intermediaries. Nationally or regionally-based intermediaries, such as economic development agencies, incubators, and public-private partnerships, aspire to build entrepreneurial capacity, diffuse knowledge and transfer technology, broker relationships, and increase participation (Howells, 2006). Intermediaries may also mitigate transaction costs and information asymmetries among various ecosystem stakeholders (Cooke, 2002; Diaz-Puente et al. 2009). Furthermore, by symbolically and financially awarding firms

for their ingenuity and/or market success, intermediaries confer reputational effects on individuals and organizations who may not otherwise be differentiated in a crowded milieu of actors. Katzy et al. (2013) find that in open innovation contexts, some intermediaries go beyond passive facilitation to act as strategic deal-brokers by increasing the number of potential partners in a network and by matching select partners to meet client objectives. The economic utility of deal-making is due in part to information asymmetries that exist between network participants and concerns of moral hazard stemming from revealing sensitive intellectual property for assessment purposes (Lee et al., 2010).

The social media presence of intermediaries may relate to promoting open innovation outcomes. For example, the mission of the United States Economic Development Administration is to “lead the federal economic development agenda by promoting innovation and competitiveness, preparing American regions for growth and success in the worldwide economy” (EDA, n.d.). In practice, the US Federal government sponsors challenge.gov, a site that offers prizes to creative problem-solving entrepreneurs. Government entities participating in challenge.gov include the National Science Foundation, Department of Energy, Department of Commerce, Department of Health and Human Services, and the White House. Each challenge features Twitter and Facebook widgets allowing users to “share the challenge” with their private and/or public networks.

At the regional level, economic development agencies leverage social media to stimulate the local innovation ecosystem. For example, Invest Atlanta in 2012 hosted startupatlanta.com, which offered observers an opportunity to “like” local entrepreneurs’

videos. The winner – the firm with the most Facebook “likes” – won ten thousand dollars. However, it remains to be seen whether these government uses of social media engage the public and result in better entrepreneurial outcomes. Government bodies recognizing the benefits of social media must also acknowledge considerable challenges: For example, agencies must grapple with the novelty of data sources, a lack of methodological know-how, and risks related to privacy, trust, and shifting perceptions of authority (Leavey, 2013).

Given the objective to mediate in networked environments, intermediaries are expected to maintain linkages with all types of ecosystem actors to lower costs of transacting, reduce information asymmetries, and broker relationships. While its social media presence will not likely result in newfound economic deals per se, an intermediary may be able to facilitate the cohesion of the ecosystem by actively or passively “introducing” an array of users who would not otherwise know of each other’s existence.

2.4. Exploratory Propositions

The exploratory propositions are guided by a simple conceptual model (Figure 2.2), which shows information distance mediating the relationship between actor needs and goals and network formation. As network structure develops, so too does social capital, which facilitates innovation outcomes. The conceptual model is in part inspired in part by a structural framework emphasizing internal and external rules of engagement (Figure A.4 in the Appendix) (Whitbred, Fonti, Steglich, & Contractor, 2011). Frames of identity established via traditional roles – e.g., the entrepreneur, venture capitalist, scientist or inventor, etc. – offer a set of salient *external* (i.e., exogenous) structural rules that guide network interaction. According to Whitbred et al. (2011), “external structural

rules are factors exogenous to a network and are based on theoretical mechanisms, which previous studies have identified as influencing communication behavior on how the communication network emerges” (pg. 408). As observed above, this literature review draws primarily from research on innovation management and policy to develop expected modes of interaction conforming to external structural rules.

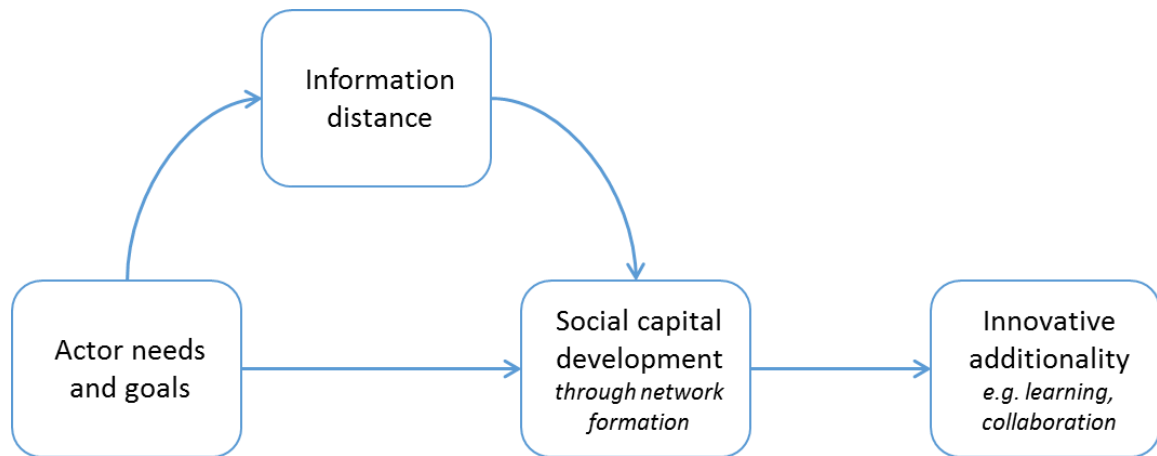


Figure 2.2: Conceptual model

In contrast to external motivations, a set of *internal* structural rules shapes the development of the network in a real-time and dynamic manner (Whitbred et al., 2011). In particular, internal rules develop as a byproduct of communication. Outcomes after each time period of analysis result in one of four mutually exclusive states: New ties form, old ties break, disconnected actors remain detached, or previous channels of communication remain intact. Internal structures are influenced by elements of intimacy and influence, e.g., reciprocity, transitivity, brokerage, trust, and the ability to meaningfully convey status and reputation through identity. Here, although controlling for some of these factors, I depart from Whitbred’s et al.’s framework to explain network

formation in terms of information distance, as subjectively discerned by actors considering the initiation of a network linkage.

Actor needs and goals. Much of the innovation literature views relationships between traditional innovation actors – i.e., firms, scientists and their research institutions, and government – as key to developing innovative capacity (Leydesdorff & Meyer, 2006; Nelson & Rosenberg, 1993). This perspective is an institutional one that values formal linkages between established entities. For instance, basic and applied research formulated in the university environment informs development and commercialization efforts by firms, hence the nation-state’s role in funding such research to promote economic growth and long-term comparative advantage (Neal et al., 2008). In emerging science-based industries, the pace of technological change is rapid and knowledge evolves quickly, suggesting the need for even more integrative coupling between types of actors (e.g., universities and firms) to improve learning (Lynn et al., 1996; Owen-Smith & Powell, 2004; Powell, White, Koput, & Owen-Smith, 2005).

As discussed above, new ways of thinking about “open innovation” suggest a variety of actors are important in the development of a sustainable and competitive innovative platform (Chesbrough, 2003, 2006). Lead-user innovation, for example, emphasizes the role of savvy users that modify existing products and processes to build next generation models and production techniques (von Hippel, 1988). In more radical innovation contexts, open innovation might include design and idea competitions, as well as informal networks that can easily arrange and reconfigure knowledge outputs to produce innovation outcomes. In some sense, the broader the input from a cross-section of actors, the more open the innovation process is.

The first proposition suggests non-random interactions between classes of ecosystem actors; i.e., actor interests and goals shape network formation in strategic ways, as discussed in Section 2.3. Small innovative firms, for example, reach out to other users to access critical information and resources (Hoang & Antoncic, 2003); they may “broker” distinct communities while at the same time embedding themselves in those communities to increase social capital (Simoni & Labory, 2007). Firms may also follow their competitors to track the evolving marketplace (Baker, 1984). Intermediaries, on the other hand, may be less likely to follow one another and more likely to follow incubated firms. Media entities, conversely, might follow a broad cross section of actors in their efforts to diffuse a broad set of ecosystem relevant news.

Table 2.1 (third column) provides additional detail around expected networking behaviors (as a function of their information needs) for each actor type. I anticipate large firms will follow customers, suppliers, and employees, while small nanotechnology firms will network widely across all actor groups on social media to access diverse information and complimentary assets. Because of their continuous search for new business, supporting actors may aggressively network with potential clients (e.g., large firms and entrepreneurs and SMEs). Financial professionals, in contrast, may follow media entities and thought-leaders to stay abreast of technological and market developments and engage with promising start-ups. I envisage scientists primarily following other scientists, media entities, and intermediaries to interact and learn within their respective field(s) of interest. Intermediaries may network with other ecosystem actor types, especially firms, as soft brokers (i.e., not economic brokers). Finally, media

entities are expected to network with an array of other ecosystem actors to prepare and diffuse storylines.

Rather than document as hypotheses these anticipated following relationships between actor types (and within an actor type), it may instead be easier to concentrate on the larger implications for the innovation literature. For example, small firms, researchers and their universities, and government bodies could be more likely to follow each other in a structure that supports a traditional innovation system or triple helix model (Leydesdorff & Meyer, 2006; Nelson & Rosenberg, 1993). An alternative viewpoint, in line with the open innovation framework, might be that traditional actors and firms in particular are more likely to follow non-traditional actors (e.g., media entities) to better access evolving research, product, and especially market-related trends. In this potential outcome case, actors use social media to explore the broader socio-technical landscape. This could occur in many different ways, e.g., when firms follow media entities that diffuse new research breakthroughs or when firms follow finance-related users who discuss market trends and/or offer equity funding.

This discussion on actor dissimilarity can be linked to the emerging literature on heterophily as a source of social capital. Whereas *homophily* suggests that individuals of similar backgrounds are more likely to congregate and converse (Monge & Contractor, 2003; Rivera et al., 2010; Rogers, 2003), heterophily offers an alternative social selection explanation: People of assorted professions offer distinct types of resource bases and are thus more likely to synergistically connect with diverse consumers of those resources. In their review of the literature on *heterophily*, Rivera et al. (2010) report that relationships based on attribute *dissimilarity* are likely to appear in challenging, complex, and

collaborative tasks such as scientific research and new product development. If homophily generally encourages feelings of safety and familiarity within the collective, heterophily drives efficiency producing and individualized behavior (Kadushin, 2011, pt. 22%).¹¹ In sum, diverse networks offer a wider range of interests and information than more homogenous networks (Kadushin, 2011).

The low cost of participating on social media should reveal a diverse set of inter-actor category linkages, assuming that indeed users value a diversity of opinion, resources, and information from these linkages. This argument presumes that users make *a priori* value judgements about the revealed identity of other users *and then proceed to follow across actor types* to access diverse network resources.

PIa: Actors choose whom to follow by mixing across affiliation types (i.e., via heterophilous relationships).

Homophily, or similarity in actor attributes, may also inform network structure in certain cases. First, high-technology firms could be more likely to initiate network linkages with one other than non-technology firm actors for competitive and/or knowledge building purposes: Each high-tech firm may wish to stay abreast of product, technology, and marketing updates materializing from other firms in the same industry. Observation of the market does not help a firm anticipate the demand curve per se, as Harrison White (1981) explains: While neoclassical theory posits that in perfectly competitive markets, firms will produce to the point that marginal cost equals marginal revenue (or market price if the firm is a price taker), White argues that producers watch

¹¹ It is worth noting that such “social selection” processes explain how actor attributes determine network structure; it is distinct from contagion and social influence models that seek to explain how network structure affects actor characteristics (e.g., how close proximity may lead to the spread of infectious diseases) (c.f. Robins, Lewis, & Wang, 2012).

each other within a market and base their actions accordingly. White's model presumes that quality is not determined by the firm but rather by consumers, who assign different valuations to the firm's product. Because valuations and cost structures differ, firms operate in ecological niches such that "watching" other firms (and possibly their respective network interactions) can become one way of generating a firm-specific market schedule of revenue opportunity as a function of output. As a second example of homophilous relations, scientists too may benefit from within-group interaction if it furthers their careers and helps coordinate ongoing work (see Section 2.3 for additional context).

P1b: Actors choose whom to follow by matching on affiliation type (i.e., via homophilous relationships).

Both P1a and P1b assume professional identity matters in the development of network linkages. If actor affiliation is not important in making following decisions, then homophilous matching within actor category and heterophilous mixing across actor categories will not constitute discernible patterns within the ecosystem's structural fabric. In this case, other explanations, such as information value, may be consequential.

Information value. Even when individuals weigh the consequences of different actions, they are unable to accurately predict outcomes, attach value to choices, and exhaustively consider all possible alternatives (Simon, 1997). This is a particularly salient issue in entrepreneurship, where cognitive biases often lead the entrepreneur to overestimate the merit of his idea (De Carolis & Saporito, 2006). However, as Simon (1997) argues, if the individual is incapable of being entirely rational, the organization, as

a set of bureaucratic processes and norms, may guide choice in rational or efficient ways. In a similar vein, innovation technologies – i.e., innovation that is organized around ICT platforms – may act to guide choice in rational or efficient ways, for example by surfacing important technological and market trends and providing a forum around which to discuss such issues.

As opposed to offline social networks and even many online varieties, information on Twitter is largely public. This means that it is relatively easy for individuals to survey the information landscape and follow other users to subscribe to a distinctive set of content. Moreover, if after a certain period of time, individuals feel that the information is no longer relevant, they may unfollow users. Indeed, Fischer and Reuber (2011) provide case-study evidence that entrepreneurs on Twitter build friend networks to learn and gain new insights. On the other hand, if a following relationship becomes stale and unhelpful in this respect, the entrepreneur will “unfollow” the user in order to save time and reduce cognitive burden. So, in some cases, where social media connections persist and offer valuable sources of information, social capital develops; in other cases, the relationship is too short lived to provide any meaningful value. Nonetheless, as theorized by Nahapiet and Ghoshal (1998), access to parties is the first step in developing exchanging, recombining, and creating knowledge, or intellectual capital. In other words, network structure is a prerequisite to the acquisition of information and potential exchange and recombination of knowledge: if such a relationship does not exist, resource transfer is impossible to achieve.

“Novel information is thought to be valuable because of its local scarcity” (Aral & Van Alstyne, 2011, p. 92): Note here that actors subjectively assess the merit of novel

information. Assuming “performance is better when communication structure matches the information processing requirements of a task” (Brass, Galaskiewicz, Greve, & Tsai, 2004, p. 799), it is likely that new ties develop in response to potential sources of new and valuable information. This is essentially Granovetter’s (1973)’s argument of the strength of weak ties turned on its head: Instead of weak ties providing novel information, it is the opportunity to access novel information that results in tie formation. Furthermore, highly valuable (and subjectively perceived) novel information may provide ecosystem actors with an opportunity to explore other scarce resources and business opportunities either online or offline. At the same time, knowledge acquisition is not entirely predicated on dialogue or interaction. Desrochers (2001), for example, argues that resource and knowledge combination often occurs through observation.

If ecosystem actors value certain types of information, then complementarities among actor interests and goals alone may not predict existing and new ecosystem linkages: although certain types of information content may be highly correlated with a specific actor type (e.g., researchers tweet mostly about research-based topics), this may not be always be the case. For example, publishers of academic journals may tweet highly scientific content, and individual users not ostensibly tied to a finance organization may tweet actively about the stock market. This decoupling between information content and actor type suggests a potentially unconventional network typology where ecosystems actors follow a diverse set of other users in order to access and process information. For example, because of the underlying uncertainty and rapid pace of technological and market change, actors may be more likely to value information stemming from research, application/product, and market-based topical areas than other types of content, which

could be attuned to less relevant topics that commonly emerge on Twitter (e.g., political events). Consequently, actors that provide relevant information in research, application/product, and market-based topical areas may be more likely to be followed, regardless of actor type.

So, even if some users are more likely to follow other actors in strategic ways in accordance with *a priori* held beliefs about how ecosystem actors link to one another, information distance (which measures novelty) may also be likely to account for the creation of network ties. This is due to the subjective perception of obtaining novel (and therefore valuable) information as information distance increases. Additionally information distance may better explain whom users choose to follow than actor mixing and matching alone.¹²

P2a: Actors choose whom to follow based on the perceived novelty of information accessible through network linkages.

P2b: Information distance explains the following decisions of users better than actor affiliation mixing and matching alone can.

Innovative additionality. I argue above that network formation is determined by actors' information needs, interests, and goals, as well as by information distance. Network formation, in turn, develops social capital (Figure 2.2). Returns to social capital may include wealth, power, reputation, and physical and mental well-being, though in this research context, potential outputs also include effects such as patenting and publishing. At first glance, the link between social media and innovation outputs appears tenuous. For example, there are too many confounding factors to suggest that social

¹² This proposition can be further formalized and tested through mediation. A mediating relationship implies that the direct effect of an independent variable (in this case actor interests and goals) on the dependent variable (network formation) can be explained in part by a third variable (information distance).

media is useful in explaining publishing or patenting output. Still, the growing adoption of social media platforms indicates less tangible (but equally important) benefits to be gained. Here I focus on building innovative capacity and pay special attention to information as an important resource that contributes to broader awareness, knowledge building, and other potential open innovation outcomes.

Much of the work on innovation links various inputs to outputs and outcomes, with the open innovation framework providing yet another set of inputs. Traditional inputs include science and technology workforce characteristics, R&D expenditure and subsidies, and resource-based views at the network, firm, or individual levels (Ahuja et al., 2008; Barney, 1991; David, Hall, & Toole, 2000; Felin & Hesterly, 2007; R. B. Freeman, 2011; Powell, Owen-Smith, & Smith-Doerr, 2011). Open innovation inputs include measures such as R&D expenditures across the supply chain, investments in outside firms, and customer and non-R&D employment involvement (Chesbrough, 2006; van de Vrande et al., 2009). Common innovation outputs include counts of patents, publications, prizes, educational degrees awarded, and employment opportunities (National Research Council, 1997). Outcomes are often less tangible than outputs and might include knowledge flows, collaboration, revenue growth, and productivity (NRC, 1997; Jaffe, 2011; K. Smith, 2004).

If one adopts a market failure approach to innovation policy, then a straightforward rationale regarding intervention ensues: the policy instrument addresses the under-provision of innovative activity occurring as a result of appropriation concerns and/or information asymmetries (Auersald, 2007). It is here that many “hyper-rational” innovation system models fail to adequately capture firm inexperience and trial-and-error

approaches to problem resolution (Hobday, 2005). In contrast, in evolutionary economic theory, systemic failure results from an inability to transition from one *underlying* knowledge structure to the next; markets exist to address exchange only for operational purposes (Bleda & del R o, 2013). In evolutionary economics, knowledge evolves over time at the individual and organizational levels, knowledge facilitates change, and change is essential to economic growth (Nelson & Winter, 1982).¹³ Consequently, instead of relying solely on an input-output or input-outcome view of the world, the literature on behavioral additionality (BA) tends also to focus on aspects of knowledge acquisition and learning (Clarysse, Wright, & Mustar, 2009; Gok & Edler, 2012). For example, Autio et al. (2008) distinguish between first-order (R&D input and output effects) and second-order (knowledge-spillover) additionalities and find that community development leads to firm-level technical, business, and marketing learning.

Similar to policy evaluation, BA addresses impact by carefully identifying the value-add of the intervention in the presence of the counterfactual (i.e., if nothing happened) (Georghiou, 2002; also see K. B. Smith & Larimer, 2009). This approach is valuable because it can explain the mechanisms that link inputs to outcomes which affect system capacity, e.g., through additional levels of firm collaboration (Polt & Streicher, 2005) and broader participation from non-traditional actors. However, as recent work notes, BA suffers from “conceptual fuzziness” across different types of additionality, imperfect operationalization of core variables, and a lack of longitudinal, process-informed empirical data by which to test causal hypotheses (Autio et al., 2008; IDEA Consult & Falk, 2009; Gok & Edler, 2012). Moreover, my use of BA departs from prior

¹³ As a normative issue, growth is good and failure is not.

scholarship in that (1) the firm is not the only actor under investigation, and (2) the government is not the sole source of the intervention.

In the third and final proposition, I summarize this discussion by considering how network linkages may contribute to increased innovative capacity: By developing strategic social media networks in terms of actor composition and information content, individuals (and their organizations) may develop broader awareness of ecosystem activity. Acting “strategically” conveys a set of engagement principles wherein actors may diversify their networks and latch-on to communities in such a way as to make their contributions to the ecosystem salient and timely. In return, these actors experience broader awareness, in tandem with learning and knowledge building, which may result in innovation outcomes, such as better product development processes, strategic alliances, funding opportunities, and community development.¹⁴ In short, strategically-developed and information-rich social media networks, when coupled with an internal capacity to learn, may result in innovation-related outcomes.

P3: Innovation outcomes are likely to occur in strategically-developed and information-rich social media networks.

In conclusion, this literature review describes and synthesizes three distinct research streams on open innovation, social capital, and various ecosystem actors. Prior work examines one or two of these literatures in isolation, but this research requires an integrative stance because it seeks to explain a contemporary phenomenon characterized by an encompassing view of ecosystem interactions. Social media directs participation and behavior in platform-specific ways, and to that end, each section in the

¹⁴ Note that while awareness and knowledge building are two outcomes, they are intermediate effects which facilitate the development of more tangible outcomes.

literature review examines recent developments in ICTs, particularly in social media, where relevant for additional insight. The mediating role of social media in innovation studies presents a number opportunities and challenges, as discussed in more detail in the following two chapters on research context and data and methods.

CHAPTER 3: RESEARCH CONTEXT

The empirical context is Twitter, the popular microblogging platform, while the innovation case study setting is graphene, a novel nanotechnology material consisting of a single layer of carbon atoms. The focus on graphene is appropriate because of the high visibility of the technology as well as its age: Graphene is a 21st century invention, and furthermore, Twitter is a 21st century innovation. Consequently, this research studies one highly dynamic invention (graphene) with numerous innovative possibilities through another highly dynamic information and communication technology (ICT) platform (Twitter). Section 3.1 covers Twitter from a usage perspective, while Section 3.2 discusses graphene's technical characteristics and potential socioeconomic impact.

3.1. Twitter

Social media and online social networks can be understood as a type of communication network, which consists of “patterns of contact that are created by the flow of messages among communicators through time and space” (Monge & Contractor, 2003, p. 3). Of course, each social platform designs its communication network in idiosyncratic ways. Some are inherently more private; some stress certain types of content; and some operate in relatively narrow contexts, while others host a variety of interaction. Twitter is a social networking site and microblogging service that allows users to share “tweets” – 140 characters of text – to a public audience while “following” other users. While its user base is not as large as Facebook, which has over one billion accounts, Twitter far exceeds other types of *public* online forums in terms of user population. For instance, by 2000, Usenet had 8.1 million participants posting on 80

thousand topics, constituting 151 million messages (Wellman, 2001). Today, Twitter has amassed over 200 million active user accounts worldwide, is available in over 20 languages, and handles 340 million tweets and over 1.6 billion search queries daily.¹⁵ On November 7, 2013, the company went public on the New York Stock Exchange with an initial valuation of \$18 billion.¹⁶

Tweets are shared with a user's followers by default and may diffuse throughout the network through "retweeting". In 2009, only about 8% of all Twitter users set their accounts to private; consequently, most user content is in the public domain (Sharma, Ghosh, Benevenuto, Ganguly, & Gummadi, 2012). A directed tweet is either a "mention" or a "reply" and occurs when users incorporate an "at" (@) sign in front of another user's screen name (Page, 2012). Not surprisingly, tweets are more likely to be retweeted when a message is not directed to a specific user (i.e., it is not conversational in nature) and when the subject content is general rather than specific (Naveed, Gottron, Kunegis, & Alhadi, 2011). To facilitate topical searching, users tag subject keywords within a tweet's content through hashtags (#).

On Twitter, one's audience may consist of friends, professional contacts, fans, the larger public, or no one in particular; some users tweet to themselves (Marwick & Boyd, 2010). Because the exact readership of public tweets is not known *a priori*, users will often envisage an audience to deal with this problem of *context collapse*; i.e., when complex and multi-faceted roles encountered in offline modes of interaction are collapsed and made more abstract in online communities (Xia, Huang, Duan, & Whinston, 2007).

¹⁵ See <http://en.wikipedia.org/wiki/Twitter>. Of course, this different in usage may also be explained a massive rise in Internet usage over the same time period.

¹⁶ See <http://finance.yahoo.com/news/twitter-raises--xx-billion-in-ipo-190114410.html>

Marwick and boyd (2010) find that users, whether tweeting on behalf of an organization or for personal gain, balance expectations of authenticity with a propensity to fashion identities and disclose information strategically. However, this calculus is not a static one because users routinely monitor reaction to their tweets and adjust accordingly. To maintain a certain level of decorum or to adhere to professional norms, users will engage in self-censorship. Furthermore, to achieve authenticity, users will straddle a balance between the strategic management of a perceived, yet nonetheless well-defined audience by sharing personal stories and quaint information.

Social media users, including those on Twitter, may self-organize into innovation-based communities (Hautz, Hutter, Fuller, Matzler, & Rieger, 2010). In contrast to “communities of practice”, which feature significant face-to-face interaction, online communities may best be characterized as “networks of practice”, which transcend geographic boundaries and involve considerable virtual communication (North & Smallbone, 2000; van den Hoof, Huysman, & Agterberg, 2007). A network of practice is not ad hoc, however: individuals still participate because their work tasks (or hobbies) benefit from transfers of knowledge and information. In this way, actors are at least loosely dependent on one another. Because of low costs of communication and due to technological innovations, participation in online networks of practice intensifies as a result of feedback among actors (Xia et al., 2007). As noted, networks of practice may operate on a variety of social platforms, including Twitter, but the nature of the platform dictates the boundaries of communication. For example, because Twitter limits text to 140 characters, and because most accounts are public, sensitive discourse (e.g.,

concerning product ordering, services, partnerships, customer feedback, etc.) is likely to be taken offline or into private online forums.

Even while some types of communication are unlikely to flourish on Twitter, compared to email, still the most dominant information sharing medium (Bernstein, Marcus, Karger, & Miller, 2010), Web 2.0 technologies achieve greater levels of interaction. For instance, social media exhibits high levels of social presence and media richness, both of which contribute to the diffusion of timely and useful information (Kaplan & Haenlein, 2010). High levels of social presence denote intimate and synchronous channels of communication, where synchronicity refers to short-time lags in between bursts of communication. High media richness indicates a medium through which communication resolves ambiguity and uncertainty.

3.2. Graphene

Graphene consists of a one layer thick sheet of carbon atoms arranged in a hexagonal lattice pattern. The material has been around for over twenty years, but only since 2004 has it emerged as one of nanotechnology's most promising materials:

Graphene is incredibly strong for being so light weight; it exhibits high thermal and electrical conductivity and is suitable for a variety of applications, including transistors, diodes, batteries, light bulbs, water purification, heaters, inks, composites, sensors, and optoelectronics (Chen, Liu, & Lin, 2013; Geim & Novoselov, 2007; Stoye, 2015).

However, the technology is in its early stage of development, and uncertainty is high with respect to commercial opportunities (Segal, 2009; Van Noorden, 2011). Today, incremental consumer-grade products are currently on the market, ranging from a tennis

racket made with graphene composite materials to graphene-enabled electrically conductive ink.

In their study of graphene patenting and scholarly publishing, Shapira et al. (2012) find that graphene may be experiencing a period of intense, concurrent research and commercialization activity. Their empirical evidence suggests a highly dynamic environment in which corporate and academic researchers contribute to frequent and interdependent breakthroughs. On the patenting end of the spectrum, both large and small firms are active in the area, with large multinationals occupying the top four positions in terms of number of patents applied for and/or granted. These corporations include Samsung (US), Sandisk (US), Teijin (JPN), and Fujitsu with at least 17 patents each. Further down the list of top ten corporate patentees are two US SMEs, Vorbeck and Nanotek Instruments. Arora et al. (2013), however, find that in their sample of twenty graphene SMEs based in the US, UK, and China, only three are listed as patent assignees (including Vorbeck and Nanotek Instruments). Because small firms are less likely to patent than larger firms (Brouwer, 1999), Arora et al. (2013) turn to unstructured data from SME websites to characterize innovative activity. They find that while some graphene SMEs are application and end-market oriented, many others are focused on producing and selling graphene as an intermediate input for further experimentation or downstream productization.

Technology enthusiasts, publication outlets, and market forecasters have termed graphene a “hype” technology (or more perhaps appropriately, a hype material) (c.f. Stoye, 2015; Van Noorden, 2014). Formally defined, the “Hype Cycle” is a Gartner conceptual model illustrated on two axes, one for visibility (y-axis) and another for

maturity (x-axis).¹⁷ The Hype Cycle is entered via a substantial technology trigger, e.g., graphene's isolation by Geim and Novoselov in 2004. Subsequent phases of the cycle include a period of "inflated expectations", followed by a trough of disillusionment, slope of enlightenment, and plateau of productivity. In brief, the visibility of hyped technologies eventually stabilize as they mature and as expectations align with R&D and product development accomplishments.

The hype surrounding graphene may be well-founded. As a material in the broader nanotechnology space, graphene may be considered a general purpose technology (GPT) (c.f. Youtie, Iacopetta, & Graham, 2007). GPTs are characterized by their pervasiveness in an array of ostensibly unrelated markets, with potentially transformative effects on total factor productivity, gross domestic product, real wages, and profitability (Helpman & Trajtenberg, 1994). Novel science-based materials such as graphene provide an intermediate input into more complex applications, whose profitability and ultimate diffusion may depend on the degree of successful integration of the nanomaterial into those final products. For example, graphene is a potential replacement for silicon-based transistors, but scientists and engineers must first understand how to properly contain voltage levels so that on-off gating (to support digital signaling) may occur (Van Noorden, 2011). If or when this breakthrough occurs, graphene enabled transistors could overcome the inherent size and processing limitations of conventional silicon-based semiconductors (otherwise known as Moore's Law).

Ott et al. (2009) argue that improvements in GPTs alter the trajectories of downstream (using) industries. In addition, the peculiar nature of adoption patterns of the

¹⁷ See <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>.

GPT downstream affects the nature of R&D at the “origins of the value chain”. For example, scientists may learn to manufacture graphene in large quantities to satisfy the unique requirements of transistor production; these manufacturing advances may then enhance other unrelated applications of graphene. The bidirectional process of upstream innovations magnifying downstream R&D efficiency is an example of innovation complementarities (Bresnahan & Trajtenberg, 1995). Because the application context is evolving concurrently with the underlying technology, the extent to which a GPT diffuses is a result of both R&D and socioeconomic factors.

In the early phases of a technology’s lifecycle, a firm may choose to vertically integrate as a way to address uncertainty and complexity, but as the technology matures and standardized architectures emerge, it becomes more efficient to streamline componentization and interaction between ecosystem actors through vertical *disintegration* across the value chain (Chesbrough, 2003). In cases where the GPT experiences ongoing improvements *while* being transacted arm’s length in markets, coordination between both ends of the value chain become increasingly complicated due to information asymmetries (Bresnahan & Trajtenberg, 1995). Yet the long-term benefits of coordinating the diffusion of the GPT outweigh short-term costs (Helpman & Trajtenberg, 1994).

While uncertainty is high, different narratives appear to project the emerging technology’s potential. Geels and Smit (2000) view the nascent process of technology development as a “social function of promises and expectations” in two primary stages. The first stage is where opportunities are signaled, where promises are made regarding scenarios that feature the transformative technology’s impact, and where an agenda is

formed to guide the technology's R&D. The second stage includes formal specifications of requirements and a space to refine the technology's scope and utility during implementation. While the first stage is often characterized by hype and unrealistic expectations, Geels and Smit counter that signaling opportunities and constructing scenarios – even when clearly hyperboles – act as important catalysts in attracting attention and resources.

Since the performance of discontinuous innovations initially lags behind that of the incumbent technology, but because the radical innovation's long-term potential is so promising (Bower & Christensen, 1995), new technologies

... cannot immediately compete in the market. They first need to be nurtured and further developed. Developers of new technologies, therefore, try to create a “protected space” in which they can improve their technologies, hopefully increasing performance characteristics. In quasi-evolutionary theories of technological developments, these protected spaces are called “niches”. A niche consist of a network of actors (e.g., funding organisations, technology developers) that share a belief in the future of a new technology and are willing to invest time and money in its further development (Geels & Smit, 2000, p. 880).

In sum, the focus on a network of actors promoting a protected space for the incubation of a technology may be heightened by the pervasive nature of a GPT. In this case, graphene is a novel intermediate input with potential for many different final product applications; hence the protected space is one that may benefit from broad exploration of uses. Furthermore, assuming small firms are unable or unwilling to vertically integrate, networks consisting of many actors form because of the various interdependencies GPTs facilitate across the still undeveloped value chain (Ott et al., 2009). So, on one hand, Geels and Smit argue that a network of actors work together to develop a protective technological niche. On the other hand, Ott et al. argue that the GPT

encourages network formation through value chain interdependencies. Participation on social media may achieve both goals simultaneously; that is, social media may offer a protected space in which participating actors are able to construct communities and engage in discourse around the emerging technology to strengthen the value proposition of using the technology in various end-user product applications. Therefore, the focus on graphene provides a suitable case study context because of the evolving nature of the GPT and the need to develop a protective space consisting of diverse actors representing various interests across the value chain.

There are three other reasons why graphene is a suitable case-study technology for this research. First, Twitter *and* graphene represent two classes of GPTs – ICTs and nanotechnology, respectively – and are about the same “age”, Twitter having been founded in 2006 and graphene’s major breakthrough occurring in 2004. How Twitter facilitates the development of graphene is an interesting question because of the dynamism of both technologies: While graphene evolves in its technology profile, so too does the social media platform that may facilitate graphene’s commercial potential.

Second, out of a list of approximately 100 graphene SMEs worldwide, over 30 maintain Twitter accounts. While some of these firms appear more active than others (e.g., in terms of frequency of tweeting), recent activity suggests that the platform is being used for more than marketing. For example, Lomiko Metals, a Canadian firm, and Graphene Laboratories, based in New York, announced a strategic alliance in early 2012.¹⁸ In this arrangement, Graphene Labs will source graphite, an input to the production process for graphene, from Lomiko. This partnership is perhaps nothing out

¹⁸ See <http://finance.yahoo.com/news/lomiko-provides-strategic-alliance-graphene-192925633.html> and <http://investorintel.com/graphite-graphene-intel/lomiko-secures-graphene-deal-through-twitter/>

of the ordinary, except that the two companies found each other on Twitter through the use of info-graphics and tweeting. This anecdotal evidence offers some motivation that graphene commercialization is being furthered by the online ecosystem.

The third rationale is a practical one: The term “graphene” exists as a standalone term with no dominant synonyms; this makes searching in Twitter (and other data sources) relatively straightforward. While such a benefit may not seem substantial at first, Twitter’s data access limitations make this an important consideration.

In an oft-cited analysis of 2000 random tweets, Pear Analytics (2009) classifies Twitter content into six categories, (Table 3.1). In Table 3.1, most of the random tweets belong to one of two categories: “pointless babble” (40.55%) and “conversational” (37.55%). Table 3.1 also contains the results of coding 1000 random graphene tweets into the same typology. Although no distinction is made between self-promotion, pass along, and news types, most of the tweets (92.60%) contain urls, retweets, or some type of brand promotion. At the same time, very few of the tweets in the graphene sample can be characterized as pointless babble or conversational. This preliminary analysis reveals that tweeting in the graphene ecosystem consists of a high level of informational content vis-à-vis the random sample. However, it is not clear whether the difference in content is an outcome of changes in platform-wide trends (i.e., that tweeting in general has become more informational and less conversational since 2009) or whether some topics on Twitter in general attract a more professional and formal type of discourse.

Table 3.2 contains a (non-random) sample of illustrative graphene tweets by five categories, as defined by Dann (2010). Here we see evidence of users exploring future applications, framing the technology in terms of its GPT potential, and monitoring startup

company news. While this sample provides some evidence of the types of tweets being authored, it does not speak to Marwick and boyd’s assertion that users are likely to straddle different content categories while maintaining their online presence.

Table 3.1: Comparison of random tweets vs. graphene tweets by content category

Content Type	Description	General Sample [1]	Graphene Sample [2]
<i>Spam</i>	Tweets with nonsensical or misleading agenda	5.85%	0%
<i>Self-promotion</i>	Tweets issued to promote a brand, product, or service	5.85%	92.60%
<i>Pass-along</i>	Retweets shared via the RT tag	8.70%	
<i>News</i>	Tweets issued by well-known media conglomerates	3.60%	
<i>Pointless babble</i>	Mundane tweets with no clear utility to readers	40.55%	2.60%
<i>Conversational</i>	Tweets directed to a particular user with an “@” sign	37.55%	4.80%

Sources: [1] Pear Analytics (2009), n=2000 tweets; [2] Random sample of graphene tweets from early 2013, n=1000 tweets.

Table 3.2: Illustrative tweets by five content categories

Type	Tweet
conversational	@calestous Hm. Graphene wedding rings?
news	Ron Dennis tells us that Graphene and SAP HANA are the technologies of the future. HANA rocks on Cisco UCS servers #g8ic #ciscouki
pass along - no url	RT @AngelVentures: Philadelphia TechBreakfast Thurs June 27 featuring Graphene Frontiers, Osmosis, EventCatalyst, OpiaTalk, SurveySnap. htt
pass along – url	RT @NatureNews: Graphene knock-offs probe ultrafast electronics http://t.co/DIVP7iAjqq
phatic	After my study on the hydrophobicity of reduced graphene oxide, it's on to sonoluminescence of imploding bubbles :3 #nerdtm

Sources: Non-random sample of tweets captured in early 2013; typology from Dann (2010). Notes: Conversational tweets are directed to other users via the “@” sign. News tweets, in this typology, do not contain urls, whereas pass along tweets may contain urls or retweet previously authored content. The phatic category is akin to Pear Analytic’s (2009) pointless babble class, although in this example, the user indicates something

relevant to his or her work in the laboratory which may be of interest to others from a professional standpoint.

CHAPTER 4: DATA AND METHOD

This chapter covers the research design in terms of data and method. It lays out an operational plan to answer the research question, “How do different types of actors use social media to form network linkages, and what kinds of innovation outcomes will result?” The method takes advantage of secondary data accessible through Twitter, as well as primary data collected via a series of interviews.

The method connects differing units of analysis, beginning with graphene firm ego networks and agglomerated “combined” following and friend networks. The combined networks can be visualized and segmented into communities of users around specific topic areas. The method also addresses the process by which network connections unfold at the micro level, i.e., between any two users in the sample. Here, the unit of analysis is the relationship between one Twitter user and another. Finally, I present an interview protocol that addresses how individual actors accrue returns from their participation on social media. With the interviews, I attempt to isolate and connect variations in usage to variations in innovation outcomes.

This chapter is organized as follows. First, I describe the data sources and sampling strategy for the quantitative and qualitative components of the research design (Section 4.1). Next, I describe my approach to quantitative analysis (Section 4.2), which consists of a network visualization component, a detailed overview of variable operationalization, and a summary of the focal statistical model (exponential random graph models, or ERGMs). Being under 25 years old – and certainly evolving considerably within the last ten years alone – ERGMs are a relatively new addition to the social science research toolkit, and hence particular attention is given to the way in which

this family of models compares to more traditional and well known regression techniques such as logit. The third section (Section 4.3) outlines the qualitative analysis plan for the interview data and also presents a logic model highlighting the larger innovation context in which this research is situated.

4.1. Data sources and sampling strategy

Twitter offers two free APIs (application programming interfaces), a REST API and Streaming API. The Streaming API pushes a limited set of real-time tweets to subscribed users. This research study employs the REST API, a *pull* mechanism wherein a client specifies his search query and retrieves data from Twitter at some moment in time; in terms of “graphene” tweets, results may be collated to produce a longitudinal dataset of tweets.¹⁹ A random sample of one hundred thousand of these graphene tweets are used in this study. Other Twitter APIs exist to collect user and network data. Twitter is accessed through a custom code base, which in turn makes extensive use of a Java library, Twitter4J, to access Twitter data.²⁰ Once retrieved, this data is stored into MySQL, an open-source relational database. A series of Java programs access this data and prepares it for multivariate statistical analysis and network visualization.

The quantitative component of the study requires no *ex-ante* approach for random sampling due to the ability to sufficiently retrieve and process all tweets in the target time frame; i.e., I follow a dense or saturation sample, completely “enumerating” the population of graphene tweets and associated users (Marsden, 1990).²¹ One important

¹⁹ To get a better understanding of the API, including search parameters and returned data, see <http://dev.twitter.com>.

²⁰ See <http://twitter4j.org/en/>.

²¹ Not all tweets are returned via the Search API, which is “focused on relevance and not completeness”. Although I collect all graphene tweets returned by the API on an hourly basis, the sampling strategy suffers from this limitation. See <https://dev.twitter.com/docs/using-search>.

caveat is the omission of some celebrity users in the combined firm-centric innovation ecosystem network, as described in more detail below.

To assess P3, and to help interpret the findings from the other two propositions, I conducted interviews with nine ecosystem actors. The interviews address issues related to awareness, problem solving, innovation additionality, and counterfactual evidence as highlighted in the interview protocol (see Table 4.3 below). The nonprobability sampling strategy combines purposive and quota approaches, accounting for both a diversity of role types (e.g., graphene firms, financial professionals, scientists, etc.) and activity measures (e.g., number of followers and frequency of communication). A quota is set to ensure a range of participation and perspectives and purposive sampling is achieved by selecting users with compelling organizational affiliations and tweet content. Prospective respondents were contacted through Twitter private messaging and/or email. All interviews were conducted by telephone or Skype.

4.2. Quantitative method

P1 and P2 are evaluated primarily through quantitative means via tweet, network, and user data, available via Twitter's suite of free APIs. At the heart of this exploration are cross-sectional, firm-centric following and friend networks. To produce these networks, which consist of many types of actors, three main steps were taken, beginning with data capture and concluding with network build-out:

1. I began with 37 graphene firms with Twitter accounts. From this number, I subtracted three firms with very large networks (over 5,000 followers) and another firm maintaining a restricted (i.e., non-public) account.
2. Collected all follower and friend relationships of these firms (i.e., the alters).

3. For each firm alter, captured and stored friend and follower relationships for the given alter to another alter, assuming that both alters have a tie to at least one of the graphene firms. If the alter-of- an-alter does not maintain a link to a graphene firm, exclude this relationship from the sample. This policy limits the size of the firm-centric network to a reasonable size (e.g., thousands of nodes vs. millions of nodes). These data were collected in early 2014.

Two important filters were applied in the data capture and build-out processes. First, if an alter had greater than 100,000 followers, a manual review process ensued to verify a direct science, technology, or innovation orientation of that user. For example, the alter could not be just any media entity; the user with over 100,000 followers would have to be focused on business, science, or technology news. Similarly, if the user was a large firm, it would have to speak to a specific ST&I audience (e.g., Amazon was excluded due to its propensity to tweet about daily deals, while Intel was included because of its emphasis on technology-focused tweets). This first filter avoided 457 popular celebrities and mass-media outlets with millions of followers, thereby saving weeks of data capture time. The second filter omitted inactive Twitter users with no active tweets since the beginning of 2012. In total, 2,356 users were excluded from the follower network and 2,559 users from the friend network.

The ego networks were thus combined and produced one large follower and one large friend graph. The combined friend network consists of 8,621 actors and 737,360 edges, while the combined follower network consists of 6,584 actors and 297,040 edges. The number of overlapping users between the two graphs stands at 3,383; that is, there are 3,383 common users that the firms follow and that in turn follow the firm. A note on

terminology: I use actors, nodes, and sometimes vertices interchangeably in the coming chapters. Ego refers to a focal node, whereas the term alters refers to his connections. In addition, edges, links, and relationships are used synonymously.

The ego networks are supplemented by user data – e.g., the number of tweets issued by a user, the account’s age, number of friends and followers, etc. – as well as “user timelines”. Each user timeline consists of up to 200 tweets, some of which relate to graphene and nanotechnology and many others that do not. In sum, the quantitative portion of this research study relies on four types of data from the Twitter API, including a) a random sample of one hundred thousand graphene tweets, b) user attributes for each active, non-celebrity user in the combined follower and friend networks, c) a sample of up to 200 tweets from each of these network user’s timelines, and d) the relationships that constitute the underlying network structures.

4.2.1. Visualization

In this research, the purpose of network visualization is to facilitate the qualitative exploration of the social media dataset. As revealed in the forthcoming chapters, social media interaction with respect to meso and macro network structure emerges through micro level linkages, which I have argued develop because of information needs and resource seeking behavior. How this process unfolds in a social media context is a relatively new stream of inquiry that benefits from illustrations showing the boundaries of differing communities that organize around topical areas. Moreover, as we will see, the communities differ in terms of actor constitution: Some types of actors are more likely than others to find themselves in certain clusters of activity. Thus, in addition to enabling exploration of the dataset, visualization also provides preliminary evidence with respect

to the study's first proposition (P1a and P1b), which suggests actors develop following relationships in non-random ways within and across actor affiliation types.

I sample and visualize both ego networks and global ecosystem networks. Visualization and community detection is performed in Gephi (Bastian, Heymann, & Jacomy, 2009) using the Force Atlas mapping and modularity algorithms. Force directed graph algorithms use the structure of the network to produce visually appealing diagrams where edges are less likely to cross (Kobourov, 2012). Modularity or community detection, on other hand, is particularly helpful in complex networks where both organization and randomness co-exist; that is, modularity is one way of isolating tightly-knitted groups of actors for further analysis (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008).

After producing the visualizations in Gephi, I implement pairwise mutual information (PMI) for all terms found in tweets sampled from the graphene firm combined follower network, as segregated by communities identified via the modularity detection algorithm. I repeat this process for the combined friend network, as well. The purpose of this analytical exercise is to identify which communities tweet about which topics.

PMI is calculated for a pair of outcomes x and y derived from discrete random variables; in this case, X represents a community (number), and Y is a set of terms appearing across all tweets in the combined follower or friend network. PMI is computed by first creating a contingency table of term frequencies to community numbers. For example, if nodes A and B belong to one community ($y = C1$) and nodes C, D, and E to other communities, I calculate the pairwise mutual information of a given term x

appearing in C1 vis-à-vis the other communities in the network. Nodes A and B may tweet nanotechnology keywords with greater frequency in comparison to nodes C, D and E. Consequently, the results of PMI in this simplified example will note these differences. Accordingly, I report the top ten terms with the highest PMI scores per community. Words with high scores are more likely to appear in the focal community vis-à-vis any other community.

4.2.2. Variables

Actor attributes affect network structure through social selection processes. In exponential random graph models (ERGMs), actor (nodal) attributes are considered exogenous such that these characteristics are immutable or change so slowly as to be considered fixed. Exogenous variables are not influenced by the dependent variable (network ties) and are also independent of endogenous dependencies within the network structure. Social selection, simply put, occurs when actors “select one another as network partners, depending on the attributes that they have” (Robins & Daraganova, 2013). In contrast, social influence occurs when the presence of a network tie alters the characteristics of an actor (Kadushin, 2011). The foundation of this study focuses on social selection processes, and the variables below reflect this. Exogenous variables include actor type and several user controls. Endogenous variables include information content and type, as well as network structural controls. All variables are summarized in Table 4.1. The primary dependent variable is *tie existence* between one ecosystem actor and another, and this variable is always measured in cross-sectional format.

Table 4.1: Variable operationalizations for the first research question

Research Question and Propositions	Attribute	Type	Operationalization	Source	Expected Sign
Research Question 1: How do different types of actors use social media to form network linkages?					
<p><i>P1a: Actors choose whom to follow by mixing across affiliation types (i.e. via heterophilous relationships).</i></p> <p><i>P1b: Actors choose whom to follow by matching on affiliation type (i.e. via homophilous relationships).</i></p>	IV: mix.actor_type	Actor	Examine user profiles and lists to semi-automatically code user types (see Table 4.3) as nanotechnology firm, other firm, support firm, finance, media, intermediary, scientists, and unknown	Twitter hand coding and API	+
	IV: match.actor_type	Actor	Examine user profiles and lists to semi-automatically code user type (see Table 4.3). Matching only occurs across the nanotechnology firm, intermediary, and scientist categories	Twitter hand-coding and API	+
	DV: directed network tie	Edge	Coded 1 if a tie exists between actors (i,j), 0 otherwise	Twitter API	
<p><i>P2a: Actors choose whom to follow based on the perceived novelty of information accessible through network linkages.</i></p> <p><i>P2b: Information distance explains the following decisions of users better than actor affiliation mixing and matching alone can.</i></p>	IV: dpq	Edge	Topic modeling to determine different content areas; the distance measure $D(p,q)$ to compute the information distance between both users' topic probability vectors	Twitter API	+
	DV: Presence of network tie		See above		
<i>Controls</i>	User status (logged)	Actor	Total number of tweets issued by the user normalized by days since account creation	Twitter API	
	Account age	Actor	Number of days since account creation	Twitter API	
	Followers (logged)	Actor	Number of user's followers	Twitter API	
	Friends (logged)	Actor	Number of user's friends	Twitter API	
	Mutuality	Edge	Whether a tie exists between (j,i)	Twitter API	

Notes: IV = independent variable; DV=dependent variable; n.p. indicates no prediction.

Table 4.2: Terms to identify user categories on Twitter

Category	Firm	Finance	Media	Scientist	Intermediary
Terms	Entrepreneur, founder, manager, management, executive, startup, business, company	Venture capital, vc, banker, investor, financier, trader	Advertiser, media, magazine, association, news, editor, writer, author, bot, marketing, blogger, blogspot, publisher	Scientist, expert, inventor, academic researcher, professor, faculty, graduate student, postgrad, postdoc, physicist, chemist, biologist, phd	Incubator, accelerator, community, association, club, institute, intermediary, public interest organization, military, army, government, city, town, municipality, state, province, technology transfer, tech transfer, science park, national lab, university, school, college

Independent variables. I identify two sets of explanatory variables. First, *user type* consists of seven categorical variables, *graphene firm*, *other firm*, *support firm*, *finance*, *media*, *scientist*, *intermediary*, and *unknown* actors that cannot be readily classified in the above six categories (e.g., science and technology enthusiasts without a professional or organizational affiliation). This set of variables was first coded in a semi-automated way using public list and user profile data to classify users before a manual validation step. For the semi-automated step, this study follows Sharma et al. (2012), who identify and compare user role metadata through three complimentary means. First, the authors extract key data from self-disclosed, yet unverified, user profiles. Second, top words from tweets posted by the user may suggest a genuine list of key terms

characterizing the user's personal and professional interests. Finally, the authors present a novel method of top words tagged by other user's lists. With lists, users may categorize their contacts into logical categories and curate these groups with meaningful metadata. In this study, only keywords associated with user profiles and lists were used to pre-classify users into one of the above actor type categories. See Table 4.2 for a list of various keywords that identify each type of user category.

The literature review covers each of these ecosystem actor types besides "unknown" actors, but a clarification should be made at this point distinguishing firms into the three distinct categories (nanotechnology firms, other firms, and support firms). Nanotechnology firms are small-to-medium sized high-technology companies dedicated to the commercialization of nanomaterials and nanotechnology enabled products. Other firms may also be focused in high-technology areas but exhibit a broader portfolio of product lines; these firms may also span industries not specifically attuned to nanotechnology (e.g., 3d printing and additive manufacturing). Support firms, lastly, represent the digital marketing efforts of social media experts and other types of consultants and professionals providing soft support services to entrepreneurs. The motivation for classifying firms into three different categories lies with the potentially divergent information and resource seeking needs of each group. For example, nanotechnology-focused SMEs may have a much different profile than support firms in terms of connecting and engaging with scientists. In addition, nanotechnology firms may be more likely to connect with intermediaries than larger, more diversified firms with access to in-house complimentary assets.

Sharma et al. (2012) find that users with few followers are unlikely to be listed more than ten times, and indeed this is where the semi-automated approach fails. Rather than relying on more sophisticated machine learning approaches, I instead verified and hand-coded a total of 11,822 unique users across the friend and follower networks following the semi-automated classification scheme. Some of this work (coding for approximately 3,000 users) was outsourced to an offshore resource in India. A random sample of 582 users was coded concurrently by the author and the offshore resource with a reliability rate of 62.2%. Because of this low inter-coder reliability, all offshore coded records were again verified by the author to maintain consistency across the entire dataset. The Appendix (Table A.1) contains documentation of the coding protocol provided by the author to the offshore resource.

Once the actors are coded into their respective categories, specifying an ERGM with mixing and matching terms invokes a process by which pairs of relationships are automatically constructed by the model. These pairs appear as mixing and matching variables to isolate the effects of heterophily (mixing) and homophily (matching) as two important social selection effects. Mixing variables excluded from the model represent cases with unconventional pairings not necessarily relevant to the innovation context.²² For example, whether unknown users follow media users is of little relevance for the present study. The reference group is also determined by a set of salient technical

²² This includes relationships from intermediaries to unknown users, media entities to unknown users, media entities to other firms, unknown users to intermediaries, unknown users to media entities, unknown users to other unknown users, unknown users to other firms, other firms to intermediaries, other firms to media entities, other firms to unknown users, other firms to other firms, scientists to other users, and scientists to other scientists.

limitations of the selected ERGM implementation, as discussed in greater depth in Chapter 6.

The second explanatory variable of interest is information distance between any two actors. Topics are discovered using a type of content analysis called topic modeling. One recently devised technique is latent Dirichlet allocation (LDA), a generative probabilistic model where each document in a corpus is a random mixture over latent topics, and each latent topic is characterized by a probabilistic distribution over observed words (Blei, Ng, & Jordan, 2003; Blei, 2012). Through topic modeling, it is possible to distinguish between the types of content that actors generate (or retweet) without having a human coder involved in the process. As shown in the platelet diagram (Figure 4.1), LDA is conceptualized as observing N words in one of M documents. Each latent topic z is a bag of w words which are distributed over each document according to a probability distribution θ . α and β are Dirichlet priors on the topic and word distributions, respectively, and act as smoothing parameters. See Blei et al. (2003) for a more formal discussion and relevant derivations.

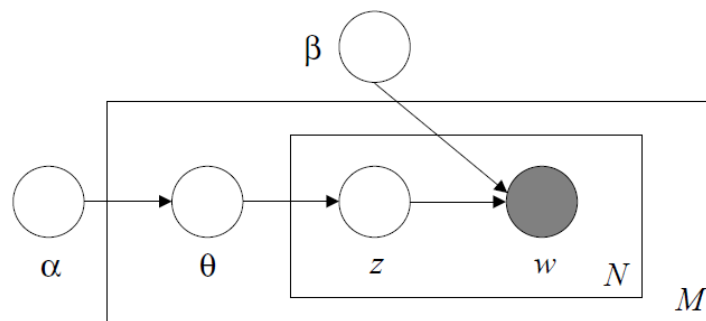


Figure 4.1: Platelet diagram. *Source:* Blei et al., 2003

To compare information distance between one actor and the next, I train LDA in two separate ways. First, I use the sample of one hundred thousand graphene tweets as

training material, and second, I train the model on a combined sample representing up to 200 tweets authored by users in the focal network. These distinct approaches build two different LDA models which differ in their orientation toward S&T and nanotechnology related concepts (found by training LDA on the set of graphene tweets) versus a broader cross-section of topics (found by training LDA on the network users' tweets). Next, I collate network tweet content on a user-by-user basis to produce a corpus of user "documents". These documents are then subject to an LDA inference process to determine the underlying topical focus of each user's tweets.

As described by Giffiths and Steyvers (2007), one commonly used measure of information distance is the asymmetric measure, dpq. Other related measures include KL divergence and JS divergence. The variable dpq is defined as follows:

$$D(p, q) = \sum_i^K p_i * \log_2\left(\frac{p_i}{q_i}\right)$$

where i iterates over K topics, p is the probability vector of topics to the first user's corpus of tweets, and q is the probability vector of topics to the second user's corpus of tweets. The asymmetrical nature of the measure is a natural operational fit because, on Twitter, the first user seeks to follow a second user based on the value that the second user offers. Theoretically speaking, I anticipate that the further away the second user is from the first, the more likely the first user is to follow that second user.

Control variables. Blanchard and Markus (2007) characterize three social factors that vary from one online environment to another. First, users project their *identity* and interpret others' identities based on disclosed profiles and usernames. Second, *influence* as measured by frequency of communication presents some users with the ability to shape social norms and sanction behavior that falls outside of established or informal

rules. Third, *intimacy* denotes a type of content and quality of interaction which results in increased levels of trust and advice. Intimate conversation establishes shared meanings and a common perception among participants of genuine engagement in the online community.

Context cues include salient information about the communication exchange; e.g., the geographic locales of the participants, their positions within a formal organizational hierarchy, and other situational elements, such as age, gender, relationship with others, topic, and social norms (Sproull & Kiesler, 1986). This discussion suggests that Twitter heightens some social context cues over others. For instance, Twitter displays the number of friends the user follows, as well as the number of following users. Frequency of activity is also revealed in the number of tweets a user issues. Both of these metrics convey social context, such as position and *influence* in the network. As another example, many users disclose personal profile information as a way to convey *identity*, social status, and geographic location. (Indeed, identity is captured directly as the actor type explanatory variable.) In contrast, level of *intimacy* is likely dependent on the specific content issued between the sender and receiver and the strength of their relationship (Gilbert & Karahalios, 2009); thus it is not considered here in greater depth because the phenomenon to model is at the network relationship level *a priori* to any other communication occurring.

In sum, a set of control variables relates to actor attributes. This group of variables covers four different aspects of a single user's influence; these include *user statuses*, the number of tweets issued by a given user normalized by the number of days the account has been active; *account age*, the number of days since account creation;

followers, the number of users following ego; and *friends*, the number of users that ego follows.

Network structural variables. A final set of variables will be included to control for the likelihood of certain links to occur endogenously based on network structure. For example, the literature has found that in certain high risk situations unbalanced triads (i.e., two ties existing between three actors) are likely to close such that all three actors become connected, e.g., through a transitive relationship (Kadushin, 2011). Similarly, depending on social norms which vary from one context to the next, network ties may be likely to be reciprocated.

4.2.3. Model selection

The unit of analysis for both P1 and P2 is the network tie. For quantitative modeling, I turn to exponential random graph modeling (ERGMs) using cross-sectional firm-centric network data. Also known as p^* models, ERGMs test whether the observed network is more or less likely to occur given basic assumptions regarding how relationships come into being (Monge & Contractor, 2003). Here, the sample space is the total number of possible configurations of network relationships given the size of the network and the number of observed edges.

A brief note on terminology: sample regression coefficients are synonymous with the term “parameters”, while the right hand side variables are known as statistics. In ERGMs, statistics are count values that capture the number of times a configuration occurs in a network. For example, a network that exhibits very little reciprocity will maintain a small statistic for the mutual term because that configuration is rarely

observed. The parameter, in contrast to the statistic, relates how important the configuration is in predicting whether a tie exists.

A critical feature of an ERGM is the distinction between endogenous and exogenous statistics. Exogenous attributes exist outside of the network context and are thus independent of how network linkages form; for example being a certain type of actor (venture capitalist vs. graphene firm) is not likely influenced by network structure. On the other hand, endogenous variables assume some type of interdependence between the focal statistic and the presence or absence of a network tie. For instance, the likelihood of higher organizing network phenomena such as triangle formation, triadic balancing, or reciprocity likely determine whether a specific directed tie between any two actors (i,j) exists. In other words, endogenous statistics capture organizing principles of network development between any two actors embedded in a larger set of nearby relationships. Incorporating endogenous structural parameters into an ERGM specification is important because actor attribute effects depend on a correctly identified model. Without such parameters, an ERGM may be underspecified in ways similar to other regression models suffering from omitted variable bias (Lusher & Robbins, 2013).

The dependent variable in most ERGMs is binary, and in very simple cases without endogenous variables, the ERGM devolves into a logistic regression model where the likelihood of a tie existing is approximated via a logistic (s-curve) function (Koskinen & Daraganova, 2013). The special case of endogenous statistics demands a unique approach to parameter estimation, however: ERGMs employ Markov Chain Monte Carlo (MCMC) simulation to produce a number of simulated networks to compare against the observed network. Yet, even in the more complicated case concerning

network structural variables, ERGMs may still be interpreted as conditional logits. Here, the statistic is known as the “change statistic” defined by two functions, $\delta_{ij, k}^+(x)$, which represent both the count of the specified configurations in the network as well as the change in count values given the presence or omission of a tie (i,j) holding the rest of the network constant. The general form of the model is specified as:

$$\Pr(X = x|\theta) \equiv P_{\theta}(x) = \frac{1}{\kappa(\theta)} \exp\{\theta_1 z_1(x) + \theta_2 z_2(x) + \dots + \theta_p z_p(x)\}$$

where $z_k(x)$ is a function that corresponds to a change statistic $\delta_{ij, k}^+(x) = z_k(\Delta_{i,j}^+x) - z_k(\Delta_{i,j}^-x)$ where $\Delta_{i,j}^+x$ is a matrix x where a specific directed linkage (i,j) equals 1, and where $\Delta_{i,j}^-x$ is a matrix x where (i,j) equals 0. Large positive (negative) parameter θ values show that the theoretical phenomena of interest occurred more often (less often) than we would expect in random configurations. $\kappa(\theta)$ is a normalizing term that accounts for all possible network configurations in the sample space.

As Monge and Contractor (2003) explain in a very simple example, relationships may develop according to a uniform probability distribution where there is a 50/50 chance that a link appears from one disconnected node to another. A more sophisticated way to model network link creation is through a Bernoulli distribution where we identify p (the probability of success) and n (the number of trials or possible links). Because networks develop in non-random ways, however, we can instead rely on certain observed traits of the network (e.g., number of total links) to limit (i.e., condition) the sample space, as well as posit certain causal forces at play that direct network growth in one way or another.

Monge and Contractor (2003) recommend specifying ERGMs via log-linear estimation of the dependent variable (probability of a tie occurring). Assuming a logistic regression-like form, the following parameters are estimated in the full model, *where* P is the probability of an observed edge; α is an edge parameter akin to a constant; X_1 is a vector of binary actor “mixing” and “matching” types in categorical variable format, x_2 is a continuous variable capturing information distance between two actors (i,j), X_3 is a vector of actor attributes, and X_4 is a vector of endogenous network processes (such as reciprocity). Partial models will isolate the effects of control variables and the mediating effect of information type and content.

$$\ln\left(\frac{P}{1-P}\right) = \ln(e^Z), \quad Z = \alpha + \beta_1 X_1 + b_2 x_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

In sum, ERGMs introduce three particularly important contributions to social network analysis (Robins, Pattison, Kalish, & Lusher, 2007). Firstly, one may learn about the distribution and nature of outcomes through parameter estimation and statistical inference; *it is not a purely descriptive approach to network outcomes*. Secondly, these models expose how “localized” social processes and structures impact global network patterns. Thirdly, while not explored in this research, ERGMs can be used to understand network evolution over time.

4.3. Qualitative method

As described by Luker (2008) and Babbie (2004), qualitative research often emphasizes a process of pattern discovery. However, qualitative modes of observation and analysis need not be confined to inductive modes of theory building. Depending on the approach, qualitative analysis offers several appealing characteristics for deductive

work (Mahoney & Gertz, 2006). Firstly, qualitative research may appeal to logic and causal modeling through Boolean models and set theory. Secondly, while maintaining sufficient levels of internal validity and consistency, qualitative research often explains individual outcomes more thoroughly than quantitative analysis alone. Lastly, and perhaps most importantly, since qualitative research does not assume equifinality, or one causal path to a particular outcome, other explanatory factors can organically surface even while the main causal relationships are tested.

Each interview will be treated as a case study in a multi-case setting. Robert Yin (2003) observes that the case study is often used to explain the “how and why” behind certain real-world phenomena when the boundaries between the phenomenon and context are not entirely clear. Moreover, the case study method is appropriate when there are more variables than (easily accessible) data points, when there are multiple sources of evidence, when the analyst is required to triangulate findings, and when prior theories can be used to guide the data collection and analysis effort. When employed in conjunction with other data sources, in this case Twitter and websites, the case method can also illuminate threats to internal and external validity.

To contextualize social media usage within the larger set of ecosystem activities, inputs, outputs, and outcomes, I constructed a logic model which, although not explicitly reflected in the interview protocol, nonetheless orients the analysis around a series of salient and possibly conflating factors (Figure 4.2). For example, the utility of social media as a novel communication platform facilitating innovative additionality could be related to the ability of users to transfer their networks between online and offline modes (or from offline to online). Furthermore, there are many types of potential outputs and

outcomes that could result social media usage, but some are more likely than others. For instance, new patents as an output of social media usage seems improbable, while learning and awareness of relevant science breakthroughs, new products, and environmental, health, and safety risks seem more likely. The interview protocol directly addresses some of the more likely possibilities (e.g., awareness and learning), but the logic model as an organizing conceptual framework includes many more short-term and long-term outcomes.

In line with the explanatory nature of P3, the interview protocol outlines question components that tie Twitter usage, network formation, and social capital to outcome variables of interest (Table 4.3). The interview protocol contains 10 multi-faceted questions to span approximately 30-60 minutes of engagement time. Five pilot interviews were conducted in the spring of 2014 to gauge whether the protocol elicited appropriate feedback to sufficiently assess P3. Some of the questions in the second section appeared redundant, and consequently the protocol was revised to gain a broader understanding of various outcomes vis-à-vis different contextual influences. For instance, instead of asking, “Could you tell me about a time when your usage of Twitter helped you achieve something?” in the pilot interview protocol, I revised this question to probe more specifically, “Could you tell me about a time when your usage of Twitter helped you modify or create a new product, process, or service?” Of course, depending on the respondent, this and other similar types of questions are not so relevant (e.g., as in the case of a scientist versus a firm, which may have a tangible product, process or service to sell or maintain). So, while the interview protocol appears quite structured in its current form, it was necessary to tailor questions to each type of actor. To this end, I

prepared for each interview and developed some custom questions depending on what I learned from the user's Twitter profile, website, LinkedIn account, and general web searches.

For each interview, responses were transcribed during the interview, and after the interview, clarifications and initial reactions were noted. To facilitate analysis, the data were organized by theme (e.g., usage patterns and problem solving outcomes vs. awareness building outcomes) in a spreadsheet with rows as cases, columns as operationalized constructs, and cell values as findings. Interpretation of the results is based on patterns across a majority of the case interviews. Deviations that reveal a wide array of usage patterns are noted on a case-by-case basis.

4.4. Summary

This chapter presents the data and method employed in this research study. Most of the data for the quantitative analysis (relating to P1 and P2) come directly from Twitter, though there is a fair amount of coding done to classify actors into one user category or another. In contrast, the qualitative portion of the method draws heavily on interview data to test whether social media contributes to beneficial outcomes.

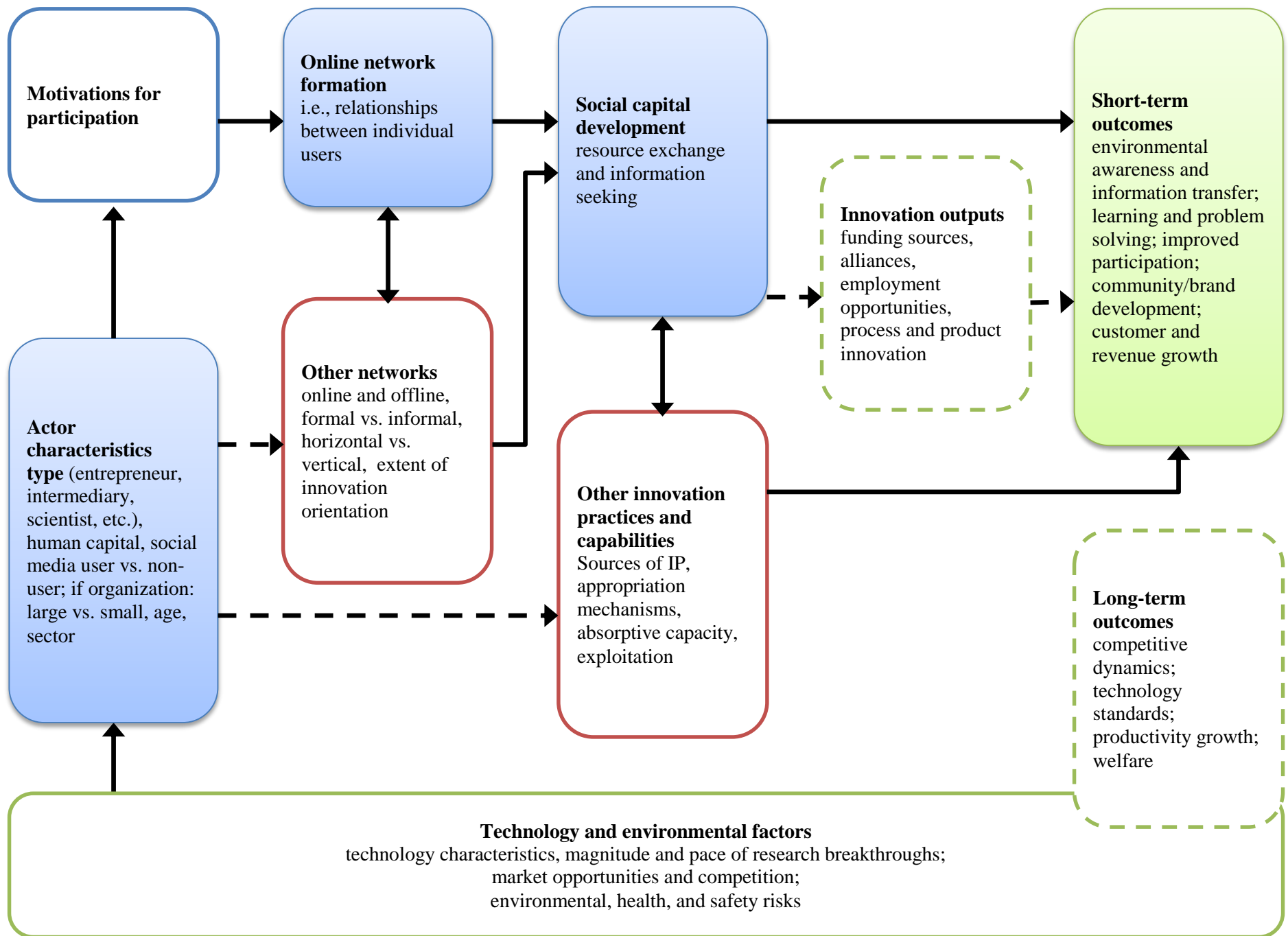


Figure 4.2: A logic model. *Notes:* The dashed boxes are not thought to result directly from social media usage; shaded boxes are explicitly incorporated in the causal model (Figure 2.2). The other solid line boxes are possible confounds that position social media usage in a broader innovation context.

Table 4.3: Interview protocol

The first set of questions asks some general questions about how you use Twitter.

1. Tell me about one of your early uses of Twitter and how your use of Twitter has changed since then?
2. How did you learn about the best way to use Twitter? What unwritten norms and rules for using Twitter are there?
3. How do you decide what to send via Twitter? What criteria do you use?
Probe: How would you compare your use of Twitter with other social media or communication platforms? e.g., LinkedIn, Facebook, and email
Probe: Could you describe the tradeoffs you make when communicating publicly vs. privately?
Probe: How do you orient your content to attract certain followers?
4. How do you decide who to follow? What criteria do you use?
Clarification: For graphene and related science or technology topics, why do you follow other Twitter users? What do you hope to get from these relationships?
Probe: To what extent do these relationships exist outside of Twitter, including in other social media and offline?
5. How do your Twitter exchanges differ between those with a business affiliation versus those who are individuals? What kinds of benefits do you get from those affiliated with businesses?

Now, I am going to ask a series of questions about how Twitter might impact innovation outcomes. I might ask some follow-up questions to obtain more detail. Feel free to elaborate whenever needed.

6. Please describe how Twitter usage has improved your awareness of issues related to graphene or nanotechnology.
Probe: Graphene technological characteristics (e.g., physical structure, properties, methods of production, and form) and research breakthroughs?
Probe: Graphene-enabled applications and market opportunities (circuits, transistors, biodevices, solar cells and batteries)?
Probe: Environmental, health and safety risks?
7. Can you describe a situation where you were able to generate a new idea or solve a problem as a result of your Twitter usage?
Clarification: Does this relate to the way you approach graphene research or commercialization? If so, to what extent is this a result of participating in an online social media community?

8. Can you describe a situation where you obtained any new resources – which could include revenue, reputational benefits, or other business opportunities – as a result of your Twitter usage?
Probe: If yes, what role did your network play in such an outcome?
9. Could you tell me about a time when your usage of Twitter helped you modify or create a new product, process, or service?
Probe: In terms of applications for nanotechnology and graphene research, does Twitter offer a space to talk about potential end-use products? If so, how?
10. In general, can you say that you've changed something about your approach to graphene research or commercialization as a result of your participation on Twitter?
Clarification: For example, based on the hype around a new application that received attention on Twitter or other social media platforms.
Probe: If yes, please can you elaborate.
Probe: If no, can you think of a situation where you would change your approach based on Twitter usage?

The qualitative method allows for a multiple case study analysis that considers an array of explanations considering the relationship between social media usage and innovation outcomes: The data collection effort is semi-structured, and as a consequence, the analysis that follows does not necessarily assume a social capital, resource exchange, and information seeking view of network structure and development. The logic model incorporates a broad array of factors that qualify the varying impacts of social media usage on outputs and outcomes, and moreover, I modify the interview protocol for each type of user (on a case by case basis) to identify salient behaviors and usage consequences that fall outside of the proposed theoretical framework.

CHAPTER 5: NETWORK ORGANIZATION AND ANALYSIS

The question of whether social media facilitates the development of innovation ecosystems relies in part on the extent to which actors tap diverse actors and communities to further their innovation goals. This is because the innovation ecosystem construct is an amalgamation of micro-level, self-organizing network behavior that aggregates into larger meso-network phenomena. This chapter examines from a descriptive perspective how this process unfolds. First, I offer a case comparison between two graphene firms' ego networks to demonstrate that social media participation reflects firm-specific resource needs and market orientation. These ego networks convey how actors are able to develop a variety of connections that are often not captured and evaluated simultaneously in innovation research. Second, I build on these firm ego networks to portray a series of *combined* network visualizations and related descriptive statistics to familiarize the reader with the global community structure of the firm-centric ecosystem. In particular, I contrast the combined friend network with the combined follower network to illustrate the asymmetric nature of the Twitter social network. The results show that graphene firms follow a distinct set of users embedded in larger communities to strategically access content (and potentially resources); at the same time, assorted communities of users follow graphene firms to presumably stay abreast of graphene and nanotechnology developments.

5.1. Select Firm Ego Networks

This section provides exploratory case study evidence to explain how two select graphene firms manage their online networks. The firms were not necessarily chosen to

select as broad a cross-section as possible (e.g., network size, tweet frequency, and account age) but are rather used to illustrate the types of networks that emerge in a few years of social media usage. Indeed, very large networks were avoided for this analysis because fewer actor-nodes convey greater clarity in the diagrams.

The purpose of this case study approach is to provide evidence that different firms generate different kinds of networks based on their information and resource needs; that is, it argues that social media networks accurately reveal how high technology firms navigate and develop ecosystem connections to further their innovation agenda. This behavior consequently engages various types of actors in the ecosystem network. All directed ties in this section show friend (i.e., following) relationships, which indicate firm agency in network development as opposed to the ecosystem's interest in the firm (the follower relationship). In addition to visualizing the firm friend networks, descriptive data and tweet excerpts are also provided to expose how Twitter contributes to a firm's online presence and interactions.

5.1.1. @ZyvexTech

Established in 1997, Zyvex Technologies is an Ohio-based firm which offers a suite of nanomaterial products and services. In 2014, OCSiAl formally acquired Zyvex Technologies, a firm with a “graphene nanotube” product on the market.²³ Zyvex created a Twitter account in January 2010 and as of July 2014 follows 188 users (Figure 5.1). In terms of actor composition, Zyvex Technologies' friend network consists of 1.82% finance users, 26.06% intermediaries, 35.15% media entities, 6.67% nanotechnology

²³ <http://www.zyvextech.com/about/>. OCSiAl is not included in the larger sample of graphene companies. As of July 2014, OCSiAl maintains a Twitter account.

firms, 7.27% unknown users, 16.36% other firms, 1.82% scientists, and 4.85% support firms. Nodes are shaded by their actor class affiliation; for example intermediaries are green, nanotechnology firms are red, and other firms are blue. Betweenness values, which represent the extent to which a user “bridges” other sets of disconnected users, determine node size. For instance, Alcoa, the multi-national aluminum producer, bridges all four communities in terms of friend relationships while the media entity, nanocomposites, bridges the nanotechnology and composites clusters.

Figure 5.1 shows four distinct communities of friends centered on Columbus, OH; the defense industry; composites; and nanotechnology. A fifth community of sparsely connected actors exists in the center of the diagram. Users in this group are fairly diverse in terms of geography and sector. For example, @mcuban is the Twitter handle of sports mogul and business tycoon Mark Cuban. The Kauffman Foundation, social media evangelists (e.g., kelleejohnson), and intermediary event organizers (e.g., chicagoideas) are among the other users in this unlabeled community.

Each of the other four communities exhibits a mix of intermediaries, firms, and specialized media entities. For example, the nanotechnology community contains several media and intermediary actors, including @nanofutures (an initiative of the European Union tweeting about developments in nanotechnology, including technology roadmaps), @nanowerk (general nanotechnology news), @NNInanonews (the US National Nanotechnology Initiative’s Twitter handle), and @rusnano_en (the Russian Federation’s nanotechnology investment fund). Also found in this cluster are two nanotechnology firms, @CambridgeNano (a Harvard nanotechnology spinout) and @slebrid8 (Serge Lebid, Executive Vice President at NanoSpire, a nanotechnology IP holding company).

Similarly, the composites cluster contains a blend of intermediaries and media entities (e.g., @plasticnews, @CompositesWrld, and @CompositesOz, a composites and fiberglass Australian industry association), as well as a number of established multinationals, including Owens Corning and DuPont.

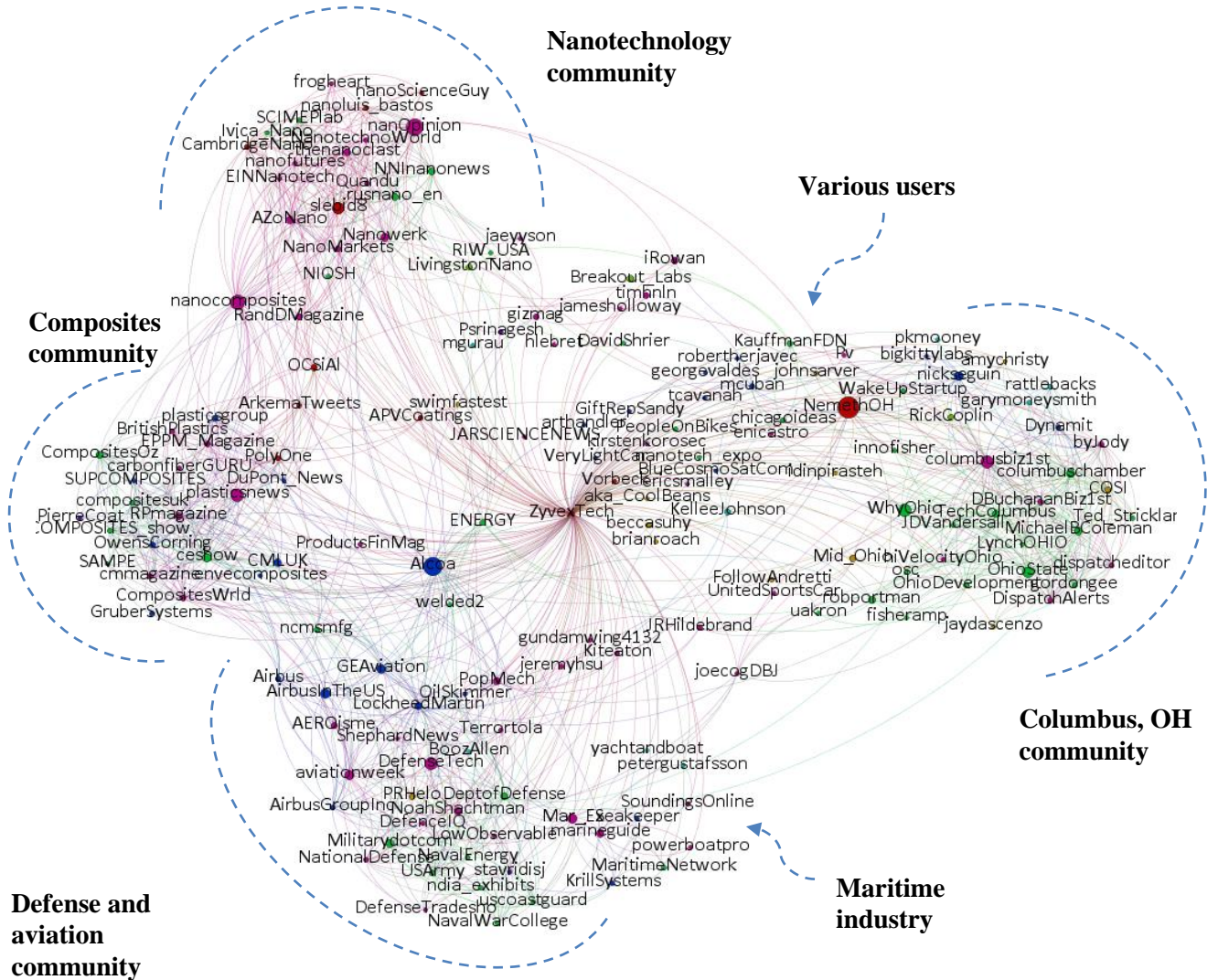


Figure 5.1: Zyvex Technology’s friend network. *Source:* Twitter; network consists of 165 users and 1,301 directed friend ties; data collected in early 2014. *Notes:* 23 additional users are not depicted because they have not actively tweeted since early 2013 or they maintain very large follower networks, thus precluding efficient data capture.

The defense and aviation industry community reveals a list of industry specific intermediaries, media entities, and firms. Firms include well-established names Airbus, Lockheed Martin, Booz Allen Hamilton, and GE, while the list of media entities is specific to the industry, e.g., @aviationweek and @DefenceIQ. Intermediary users include the US Army, the Naval War College, the US Department of Defense (DOD), and Military.com, which offers both news and membership services to active services members, their families, and veterans. While the nanotechnology and composites clusters reveal Zyvex Technology's interest in the "upstream" R&D market in which the firm operates, the defense and aviation industry community shows potential users or sponsors of its technologies: For example, in 2012 Zyvex entered into a cooperative agreement with Airbus to provide the aerospace giant with advanced materials for its next generation aircrafts.²⁴ In addition, Zyvex has received over \$404,000 and almost two million dollars in SBIR Phase I and II awards, respectively, by NASA, the Department of Energy, and DOD/DARPA.²⁵ The bottom right corner of this community contains a small grouping of maritime industry users. In 2010, Zyvex Technologies piloted an "unmanned service vessel" called Piranha, which at the time was the "largest structure ever built with nano-enhanced carbon fiber".²⁶

The fourth community is one that is specific to the firm's home city and state of Columbus, Ohio. Zyvex Technologies moved from Dallas, Texas, to Columbus in 2008 with the encouragement of Ohio intermediaries PolymerOhio and the Center for

²⁴ See <http://www.zyvextech.com/news/2012/01/airbus-announces-rd-consortium-with-national-composite-center>

²⁵ See <https://www.sbir.gov/sbirsearch/detail/355021>.

²⁶ See <http://www.zyvextech.com/news/2010/10/zyvex-technologies-reveals-the-piranha-unmanned-surface-vessel>

Multifunctional Polymer Nanomaterials and Devices, a state backed consortium of business and research activity to promote the commercialization of promising polymer materials.²⁷ While Zyvex Technologies (surprisingly) does not follow @PolymerOhio, the firm maintains linkages to several other Ohio-based institutions including The Ohio State University, The University of Akron, the Columbus Chamber of Commerce, and TechColumbus, a local startup accelerator providing venture capital funding. According to Zyvex's CEO, the state's assistance programs helped the firm develop its supply chain and marketing capabilities. Indeed, the Columbus, Ohio community shows some evidence of support firms, e.g., with Zyvex following @garymoneysmith, a web development and digital marketing specialist in nearby Illinois, and @rattlebacks, a marketing agency based in Columbus.

Overall, these four communities show a mix of local, national, and global following connections that link Zyvex Technologies to the broader innovation ecosystem. The communities span the value chain from upstream nanotechnology and composites research, development, and commercialization to a set of downstream industry users and funding sponsors. While the value chain shows a global orientation, location matters too with respect to embeddedness in the region. Contrary to many qualitative studies on innovation networks that sample based on a specific research question that limits the diversity of the resulting formal networks, Figure 5.1 simultaneously shows a range of informal network linkages that reflect the “real-world” diversity and range in this firm's innovation network. So, while technology firms and universities are important to Zyvex Technologies to follow, so too are international research initiatives, local economic

²⁷ See <http://polymerohio.org/news/news-archive/2008-news/zyvex-performance-materials-inc/>

development institutions, specialized media outlets, museums (e.g., COSI), marketing firms, politicians (e.g., Columbus mayor Michael Coleman), and even some actors with undisclosed professional affiliations. As a relatively mature nanotechnology firm with a well-developed R&D pipeline, product portfolio, and reputation, however, Zyvex seeks limited connections with scientists and financial institutions.

Table 5.1 shows a non-random sample of tweets authored by Zyvex Technologies since 2010. These tweets reveal how Zyvex provides information value to its followers. The selected tweets revolve around press releases (e.g., product announcements), an acquisition, a visit from the state’s governor, and interaction with conference participants and the broader media. Taken together, these excerpts show that Zyvex Technologies tweets to better position its in-house R&D product development capabilities to a potential set of end-users (i.e., B2B marketing). In addition, the firm appears to connect with and refer to other users on social media as a larger networking strategy that increases its ecosystem visibility while at the same time allowing it to monitor its image in the media.

Table 5.1: Non-random selection of tweets issued by Zyvex Technologies

Month of authorship	Tweet content
April 2010	Ohio Governor Strickland with ZPM staff - Thanks for visiting @Ted_Strickland #nanotech @techcolumbus http://twitpic.com/1gc3gl
December 2011	@DavidShrier Great to meet you at @LivSec conference - we'll reach out via email...
July 2012	Note to the editors at Plastics Technology: We dont make MWCNT - we just make them useful. http://bit.ly/N31OzQ
June 2013	We are excited to be a part of @OCSiAl and to welcome the next generation of #composites
January 2014	Zyvex Technologies Announces Second-Generation #Nanocomposite: Arovex® HT... http://www.zyvextech.com/news/

Source: Collected from Twitter in early 2014

5.1.2. @TeamGraphene

Graphene Frontiers is a University of Pennsylvania spin-out established in 2011. The company has maintained a Twitter account since 2012 and follows 359 users as of July 2014 (Figure 5.2). Unlike Zyvex Technologies, Graphene Frontiers is in the early stage of its product development lifecycle with a focus on producing graphene via a propriety approach to chemical vapor deposition (a bottom-up technique relying on chemical synthesis via a substrate). The firm expects scalable production processes to “facilitate market disruption within the biosensor, desalinization, and electronics industries, among others.”²⁸ Yet, according to its friend network in Figure 5.2, we see few signs of any downstream industry or application focus. Indeed, Graphene Frontiers appears highly embedded in the Philadelphia start-up scene; the firm also maintains a number of connections to users involved with nanotechnology and/or graphene.

In essence, the founders of Graphene Frontiers are trained scientists with intellectual property needing sufficient resources for implementation. Accordingly, the firm’s friend network clearly shows a large community of venture capital institutions including Sequoia Capital, Greylock Partners, and Google Ventures. Whether the firm’s Twitter usage actually helped it navigate and eventually secure venture capital is unknown, though Crunchbase, a freely accessible online database of venture capital transactions, shows one round of Series B investment in Graphene Frontiers totaling \$1.6MM by Trimaran Capital Partners in July 2014.²⁹ Trimaran Capital Partners does not maintain a Twitter account, making the case for a direct relationship between social media usage and resource acquisition problematic. Still, this evidence lends preliminary

²⁸ See <http://graphenefrontiers.com/aboutgraphene.html>

²⁹ See <http://www.crunchbase.com/organization/graphene-frontiers>

support that social media usage accurately measures a firm's innovation-related activity and can potentially contribute to positive outcomes.

The venture capital community in Figure 5.2 is noticeably homogenous in terms of actor type: only finance actors appear with very few media entities intermixed. The nanotechnology community also appears dominated by one actor class, namely specialized media outlets such as @grapheneinfo, @nanowerk, and @nanoreport. On closer inspection, however, the periphery of the community contains other nanotechnology firms including California Nanotechnologies and Oxford Nanopore Technologies. The bottom right corner of this community contains several graphene specific media entities and graphene firms (e.g., Abacus Orange, Graphene Labs, Max Materials, and Blue Stone Global Technologies), while the bottom left corner reveals several prominent US national laboratories (e.g., Lawrence Livermore and Oak Ridge). The media entities in the center of the community are an artifact of the visualization approach which places the most well-connected actors in the center of the cluster. So, media entities in this community are very likely to follow one another, and they are also likely to attract following relationships from other actor classes too.

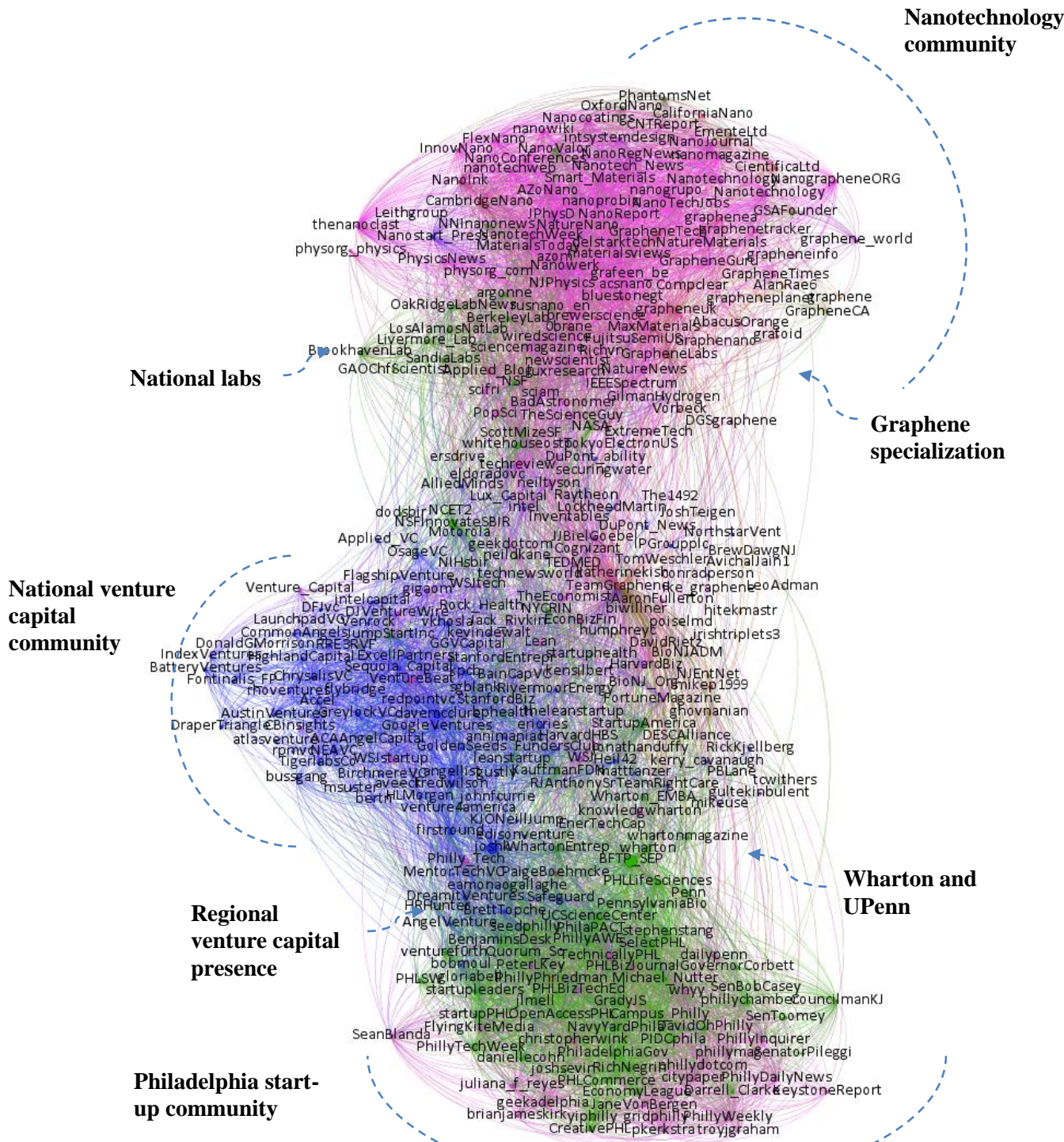


Figure 5.2: Graphene Frontier’s friend network. *Source:* Twitter; network consists of 328 users and 8,515 directed friend ties; data collected in early 2014. *Notes:* 31

additional users are not depicted because they have not actively tweeted since early 2013 or they maintain very large follower networks, thus precluding efficient data capture

The Philadelphia start-up community contains a section of densely connected intermediaries, such as PHL Life Sciences (an organization promoting life sciences in the region), the City of Philadelphia, various Twitter handles representing the Wharton School and University of Pennsylvania, and the Science Center, the oldest urban research park in the US. (Graphene Frontiers is an incubated company in the Science Center.³⁰) On the bottom of this community lies several local media entities including the historic Philadelphia Inquirer and the Geekadelphia technology blog, which covers technology trends, arts, and culture in the greater Philadelphia area. In contrast, within this community on the upper left corner – adjacent to the large venture capital community with a national presence – is a finance group of users with a regional or local focus. For instance, DreamIt Ventures is an accelerator with a presence in three northeast American cities, including Philadelphia, and provides \$25,000 in seed funding to participating companies.³¹

In sum, these three communities convey different levels of focus for Graphene Frontiers, reflecting the company's diverse resource and information needs. On one hand, the venture capital community signals a clear interest in obtaining investment funds. On the other hand, the nanotechnology community shows a set of users involved in nanotechnology R&D and news reporting, as well as a set of users specializing in graphene. The Philadelphia community, while highly connected from within, also shows some evidence of differentiation with some actors identifying purely as intermediaries,

³⁰ See <https://www.sciencecenter.org/companies/port-business-incubator-companies>

³¹ See <http://www.dreamitventures.com/about/program/about-dreamit-ventures/>

while others take on more of a financing or news reporting role. While Graphene Frontiers is unlikely to determine each following decision according to a grand strategic plan, these results do show a type of rational approach to determining which users are worthwhile to follow – and which ones are not. For instance, like Zyvex Technologies, Graphene Frontiers does not follow many scientists, perhaps indicating that intellectual property is not a significant resource constraint or that scientists are not a likely target customer for the firm’s products. The end result is the same, namely a social media network mirroring a socioeconomic intent.

Table 5.2 conveys a list of non-random, select tweets authored by Graphene Frontiers since 2010. Although not included in detail, the firm’s complete corpus of tweets reveals a detailed chronology of the company’s growth from founding, conference events, and two funding rounds. In addition, the tweets in Table 5.2 show a certain level of engagement within the graphene innovation ecosystem, e.g., as exemplified by the compliments to Dr. Elena Polyakova, CEO of Graphene Laboratories, and University of California Riverside and University of Manchester researchers, who recently achieved an experimental breakthrough that increased the conductivity of copper by “sandwiching” in graphene.

5.1.3. Cross Case Analysis

These two case studies show that graphene firms on Twitter develop networks in idiosyncratic ways depending on their market orientation and resource needs. In both cases, the firms are deeply connected to their regional innovation systems, which may provide a set of hard and soft resources, including access to human capital, advisory services, office space, machinery, and reputational benefits (Cooke, 2001). Both firms

also show connectivity to users in nanotechnology R&D and news clusters. However, while Zyvex appears to look down market via its friend network (e.g., to potential customers), Graphene Frontiers follows the venture capital industry to explore potential sources of funding. Chapter 7 provides additional qualitative evidence examining innovation outcomes resulting from social media usage.

Table 5.2: Non-random selection of tweets issued by Graphene Frontiers

Month of authorship	Tweet content
July 2012	We're pleased to share that Graphene Frontiers was announced as winner of 2012 Princeton Business Plan Competition #princetonreunions
September 2013	Excited to start roll-to-roll #graphene production! Graphene Frontiers Awarded \$744k @NSF Grant! http://buff.ly/1eLobMs
November 2013	Come meet us at the @IDTechEx #GrapheneLive trade show
December 2013	Congratulations, Elena! @GrapheneLabs partners with @StonyBrookU and Lomiko Metals on graphene supercapacitor project http://buff.ly/193Txd3
March 2014	Good work, @UCRiverside & @UoMNews: Graphene-copper sandwich enhances heat conduction http://buff.ly/1IWRVsh

Source: Collected from Twitter in early 2014

The first set of propositions examine the extent to which actors mix across or within categories to generate ecosystem diversity. Both ego networks (Figures 5.1 and 5.2) show distinct clusters and yet varying levels of actor mixing within each cluster. This is perhaps most easily seen in the finance cluster in Figure 5.2, which reveals mostly finance users intermixed with a few media entities (e.g., The Wall Street Journal’s entrepreneurship blog). Likewise, the Philadelphia start-up community consists of mostly intermediaries while the nanotechnology R&D cluster contains mostly media users. Zyvex Material’s ego network, while smaller than that of Graphene Frontiers, qualitatively displays more mixing. Consider the defense industry cluster, for example, which contains sets of “other firms” such as Lockheed Martin and Booz Allen Hamilton;

media entities covering the defense industry; and US Department of Defense agencies such as the US Army. The propensity of actors to mix within and across classes is taken up quantitatively in the following chapter.

5.2. Global Patterns and Community Structure

This section shows through network visualizations how the graphene firms' friend and follower ego networks scale when combined together. The purpose is to depict the sampled ecosystem as a whole and then identify meso-level communities for subsequent quantitative analysis. The first part of this section presents a detailed descriptive summary of the combined *friend* network (i.e., users that the graphene firms follow); the second part briefly overviews the combined *follower* network (i.e., users that follow the graphene firms). A comparison of the two networks ensues to highlight the asymmetric quality of social networking on Twitter.

5.2.1. Combined Friend Network

The combined friend network consists of the combined ego networks of 37 graphene firms on Twitter; however only users that tweeted at least once in the prior 12 calendar months (i.e., from February 2013 to February 2014) were retained for visualization and analysis purposes. This filter resulted in three fewer firms with inactive accounts. In addition, one firm maintains a private account, and its tweets and ego network relations are unavailable via the Twitter API. The final number of firm ego networks available stands at 33, with 34 graphene firms represented in the sample overall.

Descriptive statistics for the 34 graphene firms along with their respective communities are found in Table 5.3. By the first quarter of 2014, the average graphene firm had been on Twitter for just over 882 days, with the youngest having an account for just under six months and the oldest having been on Twitter for just under six years. The average firm tweeted .25 times per day (median: 0.04) with the least and most active firms tweeting zero and 2.73 times per day, respectively. The median graphene firm tweets very little about graphene (two times in total). As discussed in Chapter 4, these graphene firms are grouped together with other nanotechnology firms for subsequent analysis, unless otherwise noted.

The combined friend network consists of 8,621 active users (i.e., users that these 33 firms follow) and 737,360 edges (Figure 5.3). The overall graph density, which denotes the number of observed edges in relation to the number of possible edges³², is 1%. The average degree is 86.89. Both the in-degree and out-degree centrality distributions, which reveal the number of inbound and outbound connections by user, respectively, are highly skewed with a few users exhibiting very high in-degree distributions and a few users with very high out-degree distributions. This skewness of data is consistent with comparable measures from other types of social graphs, including co-citation, small world, and hyperlink networks (De Bellis, 2009; Kadushin, 2011).

³² The number of possible edges is $n * (n-1)$ where n is the number of nodes in the network. For the friend network, this product is 74,313,020.

Table 5.3: List of graphene firms in the sample and their corresponding cluster numbers

Screen Name	Location	Account Created Date	Normalized Status Count	Account Age	Followers Count	Friends Count	Graphene Tweets	Cluster
AbacusOrange	Nuneaton, UK	October 23, 2013	0.64	217	34	171	15	C0
CrayoNanoAS	Trondheim, Norway	August 30, 2012	0.02	636	58	5	0	C0
grapheneuk	Sheffield, UK	November 1, 2011	0.11	939	494	0	66	C0
RedexNano	Ghaziabad, India	December 17, 2011	0.00	893	6	47	0	C0
SolagraTech	Novata, California	October 31, 2013	0.00	209	1	7	1	C0
Targray	Kirkland, Quebec	January 6, 2012	0.02	873	76	167	0	C0
Vorbeck	Jessup, Maryland	October 27, 2011	0.04	944	133	0	20	C0
Vulvox	Syosset, New York	May 2, 2010	0.01	1487	23	28	0	C0
lunainnovations	Roanoke, Virginia	June 16, 2008	0.06	2172	280	85	0	C1
ZyvexTech	Columbus, Ohio	January 15, 2010	0.14	1594	262	187	2	C1
Graphene3D	Calverton, NY	July 31, 2013	0.39	301	246	760	175	C2
2DTECH	Manchester, UK	January 31, 2013	0.01	482	10	0	21	C3
AgSterilized	New Delhi, India	December 12, 2013	0.01	167	2	147	0	C3
Appliednanotech	Austin, Texas	September 14, 2009	0.00	1717	66	0	0	C3
bluestonegt	Wappingers Falls, NY	July 10, 2012	0.02	687	77	308	7	C3
DGSgraphene	Redcar, UK	April 30, 2012	0.03	758	38	38	5	C3
Directa_Plus	Lomazzo, Italy	February 4, 2013	0.00	478	2	10	0	C3
Grafentek	Istanbul, Turkey	June 10, 2013	0.02	352	7	74	2	C3
grafoid	Ottawa, Canada	November 22, 2011	0.06	918	243	41	9	C3
graphenea	Gipuzkoa, Spain	May 25, 2011	0.24	1099	818	221	233	C3
GrapheneLabs	Calverton, New York	October 2, 2010	1.18	1334	1604	1941	774	C3
grapheneplat	The Woodlands, TX	August 7, 2013	0.02	294	14	10	2	C3
GrapheneTech	Novato, California	October 13, 2011	0.46	958	443	0	31	C3
HarperIntl	Buffalo, New York	December 17, 2010	0.16	1258	187	0	12	C3
MaxMaterials	Calverton, New York	November 15, 2012	0.64	559	375	1127	148	C3
MedNanoTech	Dallas, Texas	January 6, 2013	0.41	507	36	87	2	C3
NanoIntegriss	Skokie, Illinois	September 27, 2010	0.05	1339	47	47	0	C3

Screen Name	Location	Account Created Date	Normalized Status Count	Account Age	Followers Count	Friends Count	Graphene Tweets	Cluster
TiSapphire	San Diego, California	July 6, 2009	0.71	1787	847	1988	0	C3
apaulgill	Vancouver, British Columbia	April 29, 2009	2.73	1855	1265	1649	2000	C4
Graphenano	Murcia, Spain	January 23, 2012	0.18	856	201	0	21	C5
graphenstone	Seville, Spain	October 14, 2013	0.00	226	3	16	0	C5
SiNodeSystems	Evanston, IL	October 30, 2012	0.02	575	66	55	0	C6
TeamGraphene	Philadelphia, Pennsylvania	January 13, 2012	0.01	866	136	0	72	C6
XolveInc	Middleton, Wisconsin	August 2, 2012	0.00	664	3	14	0	C6

Source: Twitter, n=34 firms collected in early 2014

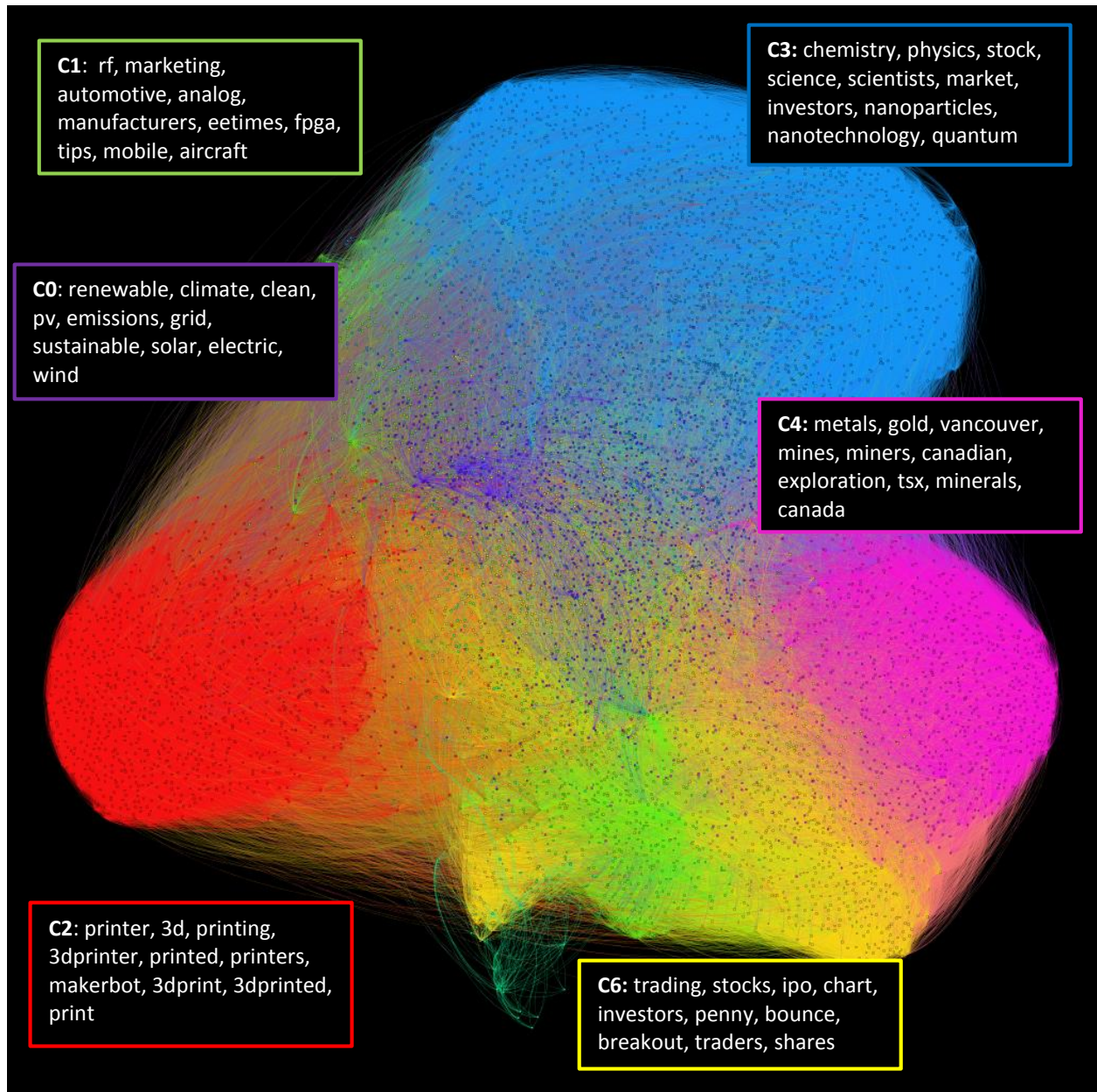


Figure 5.3: The combined firm friend network. *Source:* Twitter; network consists of 8,621 active users and 737,360 edges collected in early 2014. *Notes:* Each community (Cx) is shaded in a different color with results from PMI run on a sample of up to 200 timeline tweets for each user in the cluster. Visualization in Gephi. C5 not visible.

In Figure 5.3, the network contains seven communities, as identified by Gephi's modularity detection algorithm. Each community is shaded in a different color with

results from pairwise mutual information (PMI) run on a sample of up to 200 timeline tweets for each user in the cluster. The following text describes each community:

1. C0 (1013 users, or 11.75% of the network), distributed in the center of the graph, spans the other communities in terms of structure; it focuses on renewable energy topics.
2. C1, the dispersed light green cluster consisting of 1045 users (or 12.12% of the network), contains a mix of terms that loosely convey a manufacturing, industrial, or application focus. This notion is partly confirmed when analyzing PMI results from the same community analysis, but using user description content instead of timeline tweets; the following terms materialize: *composites, semiconductor, electronics, manufacturing, marketing, science, chemistry, aerospace, and mining*.
3. C2, covering 1089 users or 12.63% of the network, represents a large 3d printing community. One of the graphene firms, Graphene 3D, is firmly embedded in this cluster.
4. C3, the largest community with 2942 users (34.13%), conveys a general nanotechnology and science orientation. To confirm this inference, results from PMI user description data reveals the terms: *chemistry, 3d, nanotechnology, university, physics, research, printing, phd, student, and stock*. Indeed, many of these terms offer a coherent message that C3 is comprised of researchers, perhaps based at universities, undertaking natural science research. (See Table 5.3 for additional confirming evidence.)

5. C4 with 898 users (10.42%) is a mining and mineral resource cluster, in which a Vancouver-based graphite mining firm moving upstream to graphene production is a member.
6. C5 is a Spanish speaking community consisting of only 133 actors and is not visible on the graph.
7. C6, consisting of 1501 users (17.41%), is a financial markets community.

The communities differ not only by content but also by actor composition. Table 5.3 contains a descriptive analysis juxtaposing actor type by community. Here we see that many communities attract certain types of actors. For example, a majority of finance users in the combined friend network can be found in cluster 6 (financial markets). Over 46% of all intermediaries, almost 75% of nanotechnology firms, over 45% of unknown users, and over 90% of all scientists can be found in cluster 3. Recall that C3 appears to be focused on nanotechnology research based on the PMI results. Support firms, while broadly distributed appear most often in C1 and C3, while other firms are most likely to be found in C2. (C2 contains many startup 3d printing firms.) Finally, media actors appear evenly distributed throughout the communities, with representation as a percent of all media actors ranging approximately from 10-25% in any given cluster. Indeed, the media category represents the greatest number of users by far (33.50%) overall, followed by other firms (17.70%), unknown users (i.e., non-identifiable users at 13.50%), intermediaries (15.70%), scientists (7.60%), support firms (5.60%), finance (3.80%), and nanotechnology firms (2.50%). Because users in this network are only included if they are followed by one of the 33 graphene firms, this finding suggests that many of the firms in the sample use social media for information purposes from news outlets, bloggers, and

market analysts (all belonging to the media category). As suggested above, most graphene firms can be found in C0 (10 firms) and C3 (17 firms).³³

Each actor type exhibits a unique set of profile characteristics (Figures 5.4 – 5.7). Because many typical user-level Twitter variables are highly skewed, I present the median values for number of followers (Figure 5.4), number of friends (Figure 5.5), and normalized status count (i.e., the average number of tweets authored per day since account has been active, Figure 5.7). However, because the median number of graphene tweets per actor type is zero, I show mean values for this measure by actor type (Figure 5.6).

Media entities, followed by intermediaries and financial organizations, attract the greatest number of followers with median values of 3,083, 2,528.50, and 2,456.50 respectively. Support firms and other firms attract a median of 1,357 and 862 followers, respectively. At the lowest end of the spectrum are scientists (360), nanotechnology firms (201), and unknown (not readily identifiable) actors (176.50). In terms of friends, or users that a given user follows, support firms lead with a median value of 1000. They are followed by media entities (530), other firms (496.50), financial organizations (451.50), intermediaries (442.50), scientists (273), unknown actors (253), and nanotechnology firms (146).

³³ This is an important observation. Refer to the Section 6.1 for additional details.

Table 5.4: Number of users by actor type in the combined friend network

Actor type and community	Number of actors (in community)	Percent of actor (in community)	Actor type and community	Number of actors (in community)	Percent of actor (in community)
finance	330	3.83%	other firm	1526	17.70%
0	6	1.82%	0	185	12.12%
1	17	5.15%	1	254	16.64%
2	2	0.61%	2	528	34.60%
3	16	4.85%	3	203	13.30%
4	74	22.42%	4	199	13.04%
6	215	65.15%	6	137	8.98%
intermediary	1354	15.71%	scientist	652	7.56%
0	290	21.42%	0	14	2.15%
1	126	9.31%	1	12	1.84%
2	107	7.90%	2	13	1.99%
3	627	46.31%	3	593	90.95%
4	41	3.03%	4	3	0.46%
6	142	10.49%	6	17	2.61%
media	2893	33.56%	support firm	483	5.60%
0	359	12.41%	0	39	8.07%
1	349	12.06%	1	142	29.40%
2	309	10.68%	2	23	4.76%
3	739	25.54%	3	75	15.53%
4	388	13.41%	4	85	17.60%
6	687	23.75%	6	113	23.40%
nano firm	217	2.52%	unknown	1166	13.53%
0	16	7.37%	0	104	8.92%
1	22	10.14%	1	123	10.55%
2	5	2.30%	2	102	8.75%
3	161	74.19%	3	528	45.28%
4	2	0.92%	4	106	9.09%
6	6	2.76%	6	184	15.78%

Source: Twitter, n=8,621 users in the combined friend network, collected in early 2014

On average, nanotechnology firms tweet the most about graphene with a mean of over 30 tweets per user. (The mean is not normalized by account age; these tweets were captured during a 1.5 year time span from late 2012 until early 2014). Much further behind are media entities (7.27); financial firms, support firms, and intermediaries (each type with between 3 and 4 graphene tweets per user on average), scientists (2.49), unknown actors (1.87), and other firms (1.10). In contrast, normalized tweet activity per user by actor type shows media entities as the most prolific authors with a median of 1.71 tweets per day, followed by support firms (0.95), financial organizations (0.91), intermediaries (0.89), scientists (0.68), other firms (0.49), other actors (0.39), and nano firms (0.18).

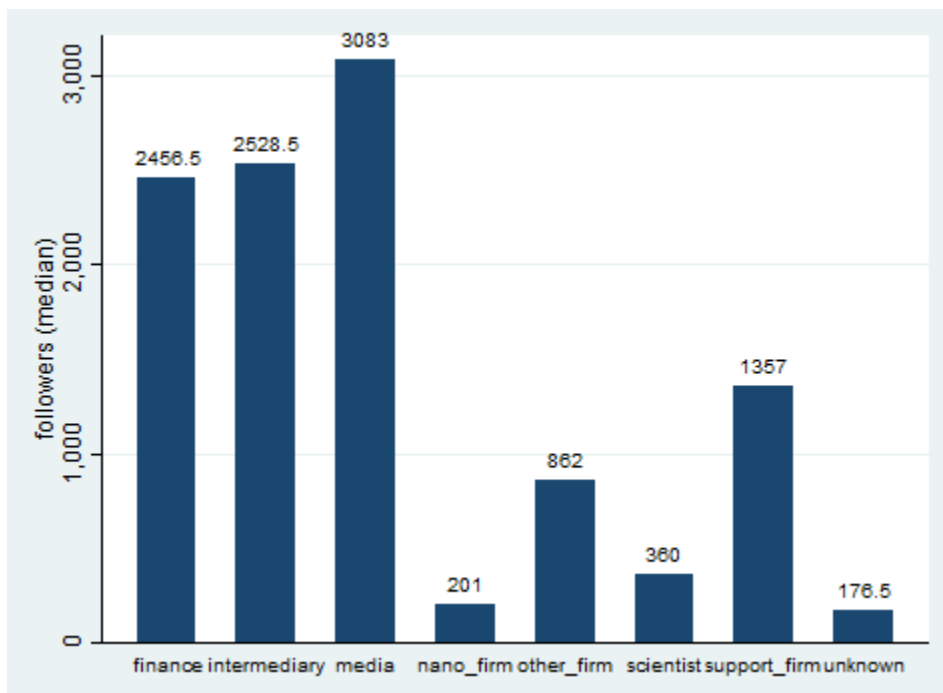


Figure 5.4: Median number of followers by actor type in combined friend network.

Source: Twitter, n=8,621 users collected in early 2014

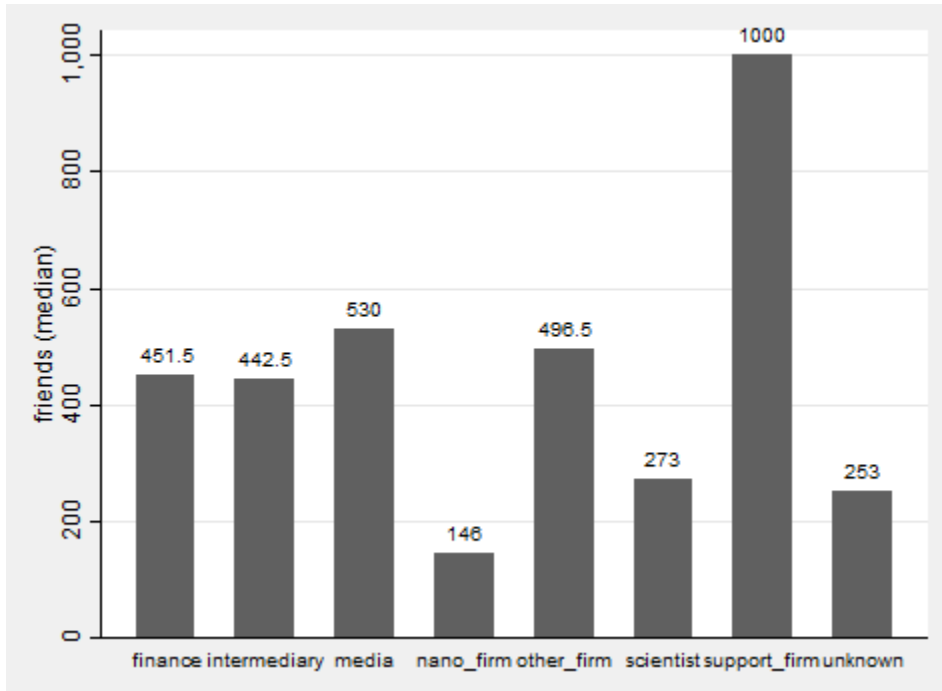


Figure 5.5: Median number of friends by actor type in combined friend network.
Source: Twitter, n=8,621 users collected in early 2014

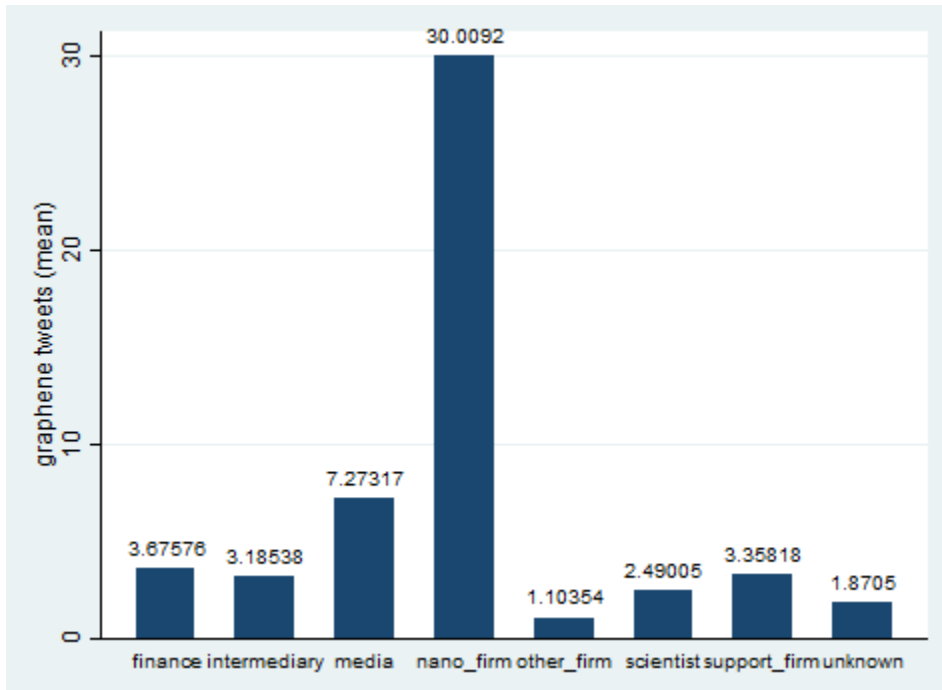


Figure 5.6: Mean number of graphene tweets by actor type in combined friend network.
Source: Twitter, n=8,621 users collected in early 2014

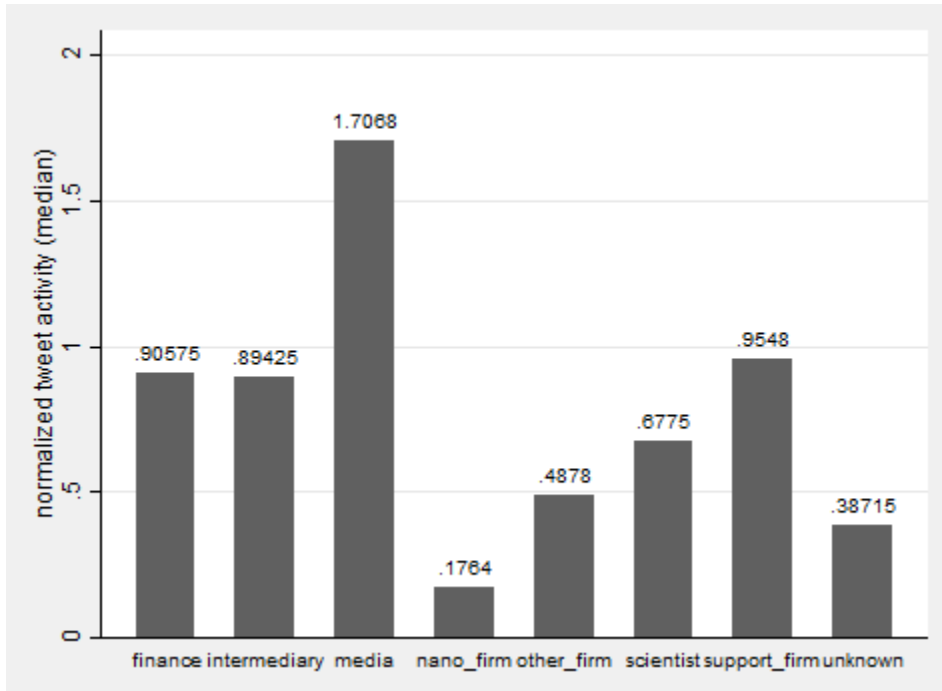


Figure 5.7: Median normalized tweet activity by actor type in the combined friend network. *Source:* Twitter, n=8,621 users collected in early 2014

Taken together, these descriptive statistics by actor type reveal three noteworthy trends: First, nanotechnology firms are among the users with the smallest networks; these firms are not usually as active as other actor type accounts, but at the same time, they are more likely to tweet about graphene than other types of users. Second, media entities have the largest follower networks and second largest friend network and are the most active tweeters. Third, unknown actors and scientists are among the two actor types with the smallest follower and friend network sizes (behind nanotechnology firms). Because unknown actors do not have or choose not to disclose a professional affiliation, they do not appear to attract much interest from the ecosystem, as evidenced by their relatively small friend networks. In contrast, the small size of scientist networks may come as a surprise. This could be due to many scientists using their accounts for personal reasons,

thus attracting only personal connections, or it could be due to highly specialized communication that does not easily appeal to a wide audience. Although the results cannot distinguish between these competing explanations, the ERGM analysis found in the next chapter revisits this topic with additional scrutiny.

The number of edges in the combined friend network stands at 737,360. The “mixing matrix” in Table 5.4 shows the percent of all ties from a given actor type category to another actor type category. The darker shading indicates higher percentages. These data provide descriptive evidence from which to evaluate P1a and P1b, which examine the degree of mixing within and across actor categories. Immediately visible is the high propensity for each type of actor category to initiate linkages with media entities; in aggregate, the percent of ties emanating from each of the eight categories to the media class ranges from 42% - 52%. The intermediary class of users is second-most likely (after media) to attract followers from a broad cross-section of actor types: besides the financial user class, at least 10% of outbound following relationships are directed to intermediary users with nanotechnology firms and scientists in particular directing 20% of their ties to intermediary users. At first glance, this finding appears commensurate with existing theoretical and empirical work that places intermediaries at the center of open innovation networks (e.g., see Lee et al., 2010). However, recent work by Lovejoy et al. (2012) assessing the Twitter usage of 73 non-profits reports that many of these organizations use Twitter not as a tool for engagement but rather as a medium for one-way mass communication. From this perspective, intermediaries may act more like media entities than mediators. In addition, each class of users besides scientists directs at least 10% of their following relationships to other firms. With the exception of scientists,

nanotechnology firms direct a greater percentage of their overall following ties to scientists than any other class of users, suggesting that scientists on social media attract the attention of high-technology firms also on social media.

Also visible is the relatively strong tendency for actors to follow similar actors in the same class; that is, homophily as a social selection process is evident here. 15% of all friend ties from finance users are directed to other finance users. This percentage is 31% for intermediaries, 33% for other firms, 21% for scientists, and 11% for support firms. In sum, the mixing matrix reveals homophily as a social selection process to follow others in the same actor category. In addition, this descriptive data show significant signs of diversity in ecosystem connections in terms of following media and intermediary actors. However, for some of the other inter-category relationships, the evidence is too ambiguous to make a final determination and requires the accuracy of statistical ERGMs (Chapter 7).

Table 5.5: Mixing matrix for combined friend network

From	To								
	finance	inter- mediary	media	nano firm	unknown	other firm	scientist	support firm	total
finance	15%	8%	52%	1%	3%	13%	1%	7%	100%
intermediary	2%	31%	42%	2%	2%	13%	5%	3%	100%
media	4%	13%	49%	2%	3%	19%	5%	5%	100%
nano firm	2%	20%	43%	7%	5%	12%	7%	4%	100%
unknown	4%	16%	50%	2%	3%	17%	3%	4%	100%
other firm	3%	11%	44%	1%	3%	33%	2%	4%	100%
scientist	1%	20%	47%	3%	2%	5%	21%	2%	100%
support firm	7%	11%	46%	2%	4%	18%	2%	11%	100%

Source: Twitter, n=737,360 edges in the combined friend network, collected in early 2014

5.2.2. Combined Follower Network

Having described the friend (i.e., following) network in some detail, I now turn to a brief overview of the combined follower network (Figure 5.8). In contrast to the combined friend network, the combined follower network contains 6,584 actors and 297,040 directed ties. Whereas the combined friend network consists of users that the 33 graphene firms follow, the follower network contains all the users that have chosen (as of early 2014) to follow the 33 graphene firms. Of these 6,584 actors, roughly half (3,384) exist in the combined friend network, demonstrating that in aggregate graphene firms follow about half of the users that follow them.

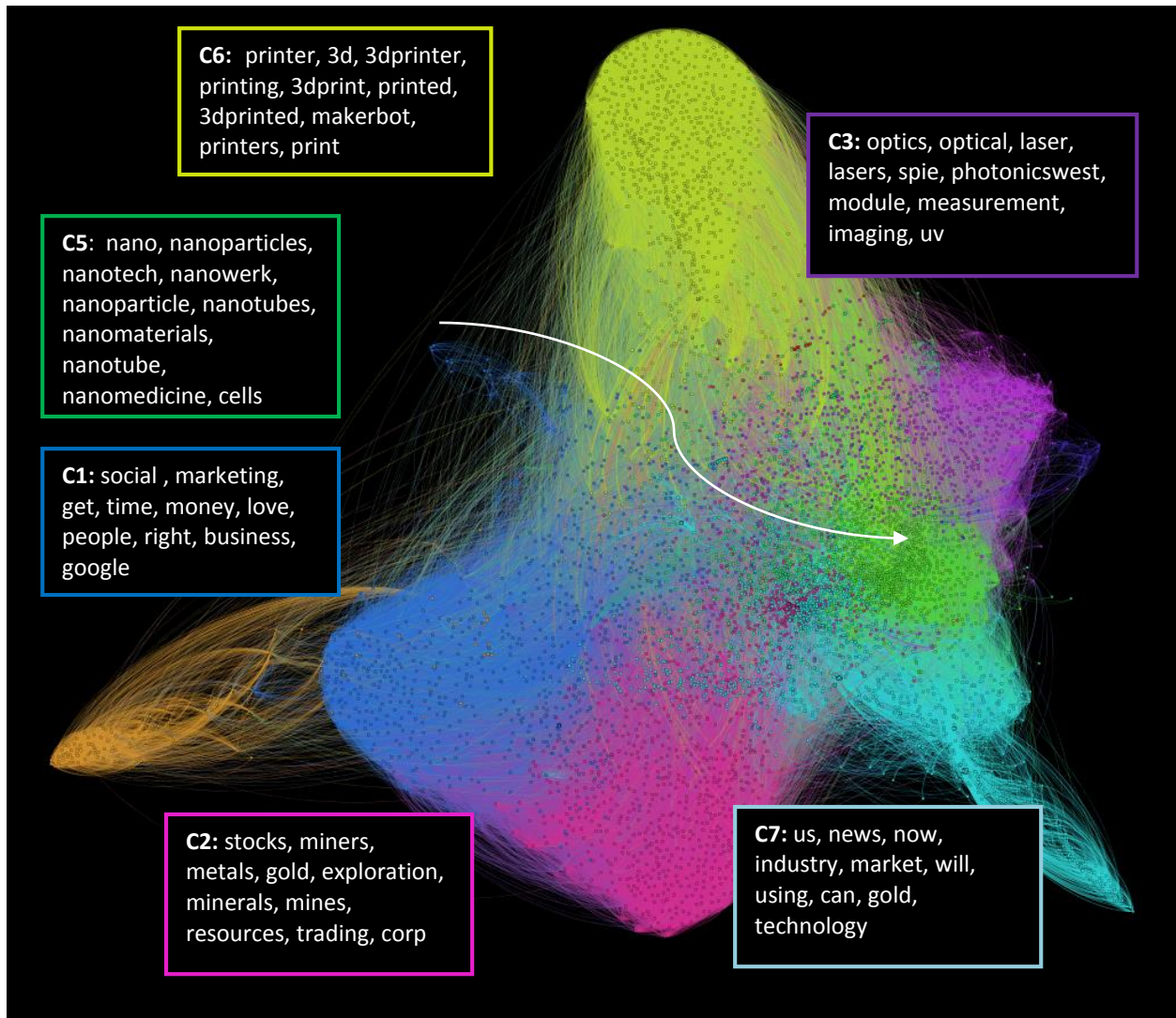


Figure 5.8: The combined firm follower network. *Source:* Twitter, network consists of 6,584 active users and 297,040 edges collected in early 2014. *Notes:* Each community (Cx) is shaded in a different color with results from PMI run on a sample of up to 200 timeline tweets for each user in the cluster. Visualization in Gephi. C0, C4, C8, and C9 not included in analysis because of insufficient size, primarily non-English language content, and/or incoherent PMI results.

Ten communities, also identified by Gephi’s modularity detection algorithm, exist in the combined follower network (Figure 5.8). That is, there are three additional communities found in the combined follower network (10) than in the combined friend network (7). Three of these communities are very small, consisting of fewer than 5% of the users in total. These clusters are not described in detail, though based on the PMI

results, one appears focused on a specific geographic region in Virginia, USA, while the other two are non-English tweeting communities of users. A fourth cluster also is not assessed, although it is larger than the previously mentioned three with 9% of all users: This community appears in Figure 5.8 in the center with nodes shaded in light green. However, the PMI results do not convey a meaningful focus in orientation (i.e., the keywords are not coherently oriented around a specific topic).

Excluding these four communities, six remain viable for analysis:

1. C1 consists of 785 users or 12% of the combined follower network's users and appears to focus on social marketing and search engine optimization.
2. C2 contains 1084 users or 16% of the network's users and represents users with mining interests (similar to C4 in Figure 5.3).
3. C3 represents 778 users (12% of the full network) and reflects an optical and laser focus, indicating that users working or interested in one potential downstream application follow one or more of the graphene firms.
4. C5 (785 users or 12% of the network) is a general nanotechnology research cluster with one keyword containing the name of a leading news portal (nanowerk). To ensure the PMI results on C5 using sample tweets is not biased by this or any other subset of users, I reviewed the PMI key terms from the user description data. The identified terms include *nano*, *nanotech*, *research*, *science*, *3d*, *nanomedicine*, *nanoscience*, *materials*, *chemistry*, and *rd*. Thus, this cluster appears research focused in a similar way to C3 in Figure 5.3, though C5 here is only 27% of the size of C3 there.

5. C6 (762 users or 12% of the network) is a community focused on 3d printing, analogous to C2 in Figure 5.3.
6. Finally, C7 (1638 users or 25% of the network) extends from the lower right hand corner of the graph to its center. The key terms suggest a market and industry forecasting/news orientation, though additional focus is not easily obtained through a review of the PMI results based on user descriptions.

5.2.3. Comparison

The actor type composition of the two friend and follower social graphs look remarkably similar, though not exactly the same, in terms of the percentage of ties from one actor class to another (Table 5.5): the percent representation of nanotechnology firms, other firms, support firms, financial entities, and scientists is almost the same between the two networks. However, representation by actor type differs from the firm friend network in three specific ways. First, media actors are much less likely to be included in the follower network (20%) than in the friend (34%) network. This is likely due to the fact that media types attract a large number of followers themselves, but many accounts, especially those of the large established outlets, do not follow many other users in turn. Second, intermediaries have a much stronger presence in the friend network (16%) than the follower network (8%), suggesting an asymmetric level of interest in the intermediaries by the graphene firms; that is, intermediaries are less likely to follow graphene firms than graphene firms are likely to follow intermediaries. The third difference concerns the “unknown”, non-identifiable user category: Unidentifiable users constitute 33% of the combined follower network and only 14% of the combined friend

network. This finding can be accounted for by the role that identity disclosure plays online. If accounts in general do not explicitly reveal a professional association, it may be more difficult to gauge the benefit of initiating a network linkage. Thus, since this “unknown” account type is not as relevant to ecosystem dynamics than the other actor categories (e.g., scientists, small firms, intermediaries), users are less likely to begin following “unknown” individuals.

Table 5.6: Number of users by actor type in the combined follower and friend networks

actor type	finance	inter- mediary	media	nano firm	unknown	other firm	scientist	support firm	total
<i>follower (n)</i>	166	556	1335	196	2201	1180	403	547	6584
<i>network (%)</i>	3%	8%	20%	3%	33%	18%	6%	8%	100%
<i>friend (n)</i>	330	1354	2892	217	1166	1526	653	483	8621
<i>network (%)</i>	4%	16%	34%	3%	14%	18%	8%	6%	100%

Source: Twitter; 8,621 users in the combined friend network, and 6,584 users in the combined follower network; data collected in early 2014

The combined follower network is less dense than the combined friend network (0.007), meaning that this graph has fewer ties than the larger friend network even after controlling for size differences. This may appear surprising at first because the literature on social networks finds that smaller networks are often denser than larger networks because communication is easier to maintain, and it is less cumbersome to develop cognitive models of who-knows-whom (Baker, 1984; Kadushin, 2011). However, as noted above, in comparison to the combined friend network, the combined follower network contains many more “unknown” (identity-less) users who are much less likely to attract following relationships than media entities or intermediaries, thus accounting for the additional sparsity.

In sum, the combined follower network (i.e., users that follow the 33 graphene firms) is noticeably more fragmented than the combined friend network. While some of the communities overlap in focus, we see differentiation not only in the type of users in each network but also the content that these different communities share. For example, an energy community found in the middle of the combined follower network (Figure 5.3) appears to bridge the other clusters. This energy community is not immediately visible in the combined friend network (Figure 5.8). In addition, the communities in Figure 5.8 appear less well integrated overall. For instance, the optical and laser focused community exists on the periphery of the graph, as do some of the non-labeled clusters. The result in the combined follower network is a fragmented set of communities that follow one or more graphene firms in a certain context (e.g., market forecasting, regional economic development, specific applications or industries). In contrast, the graphene firms follow users which in the aggregate generate in the combined friend network a set of more general industry, application, and research specific communities.

5.3. Summary and Synthesis

I conclude this chapter with a brief summary and synthesis of four key results. First, in the combined friend network, we see that most users are coded as media entities and that most users follow media entities (and to a lesser extent intermediaries). This finding suggests that Twitter is first and foremost a media platform, albeit one that facilitates interaction and the diffusion of specialized information. For instance, many of the media entities are non-traditional in the sense that they focus on new media (e.g., blogs) in specialized subject areas. This can be easily seen in the types of media entities

that exist within the communities found in Zyvex Technologies and Graphene Frontier's ego networks in Section 5.1.

An additional consideration is an alternative mode of interacting that taps non-media relationships across and within communities, as seen in the mixing matrix (Table 5.4). From this standpoint, the communities in the combined friend network show signs of heterophily (e.g., the research cluster consisting of scientists, other firms, intermediaries, nanotechnology firms, other users) and homophily (e.g., the financial markets cluster that contains a majority of the finance users), thus offering some preliminary support for both P1a and P1b. In sum, while media entities are at the center of each community within the combined friend and follower networks, thus indicating their importance in each topical cluster, there appears to be other types of network linkages that also are important to ecosystem network structure.

Second, each actor class maintains a unique set of usage characteristics. For example, graphene firms tweet about graphene more than any other actor type, but these firms are overall much less active than the other actor types in terms of tweet frequency and network size. Third, graphene firms follow many more media and intermediary types than non-identifiable users, but the reverse is true for users following graphene firms. This suggests that Twitter as a platform for information sharing is asymmetric in nature: on local and global levels, the value of information is contingent on the receiver's preferences such that following and friend networks diverge in substantive ways. Indeed, the overlap in actor composition between the two networks shows just 3,383 common users, indicating that graphene firms do not always reciprocate the following relationship.

This topic is returned to in greater depth in Chapter 7, which provides qualitative evidence of usage patterns.

Finally, the cross case comparison of two ego networks reveals that graphene firms develop their Twitter networks to align with high-level business development and information needs, e.g., in the value chain or for venture capital. Thus, this evidence supports the notion that studying innovation on social media can be a fruitful exercise in terms of advancing our understanding how firms within the same industry and product space connect with different communities to further firm survival and growth. This topic is also explored in greater detail in Chapter 7. First, however, I turn to quantitatively testing how innovation ecosystem networks develop on social media in Chapter 6.

CHAPTER 6: QUANTITATIVE MODELING

The network analyses in the last chapter reveal the building blocks of the graphene innovation ecosystem through the lens of firm-centric ego networks. In particular, the case studies show how two graphene firms approach their friend network to tap into distinct communities. When these relationships aggregate into a *combined* network, it becomes possible to discern some of the factors shaping interaction. For example, in the combined friend network, media entities permeate all communities whereas finance users agglomerate in a separate community. Further, the nanotechnology R&D cluster contains most of the scientists and nanotechnology firms and just under half of all intermediaries. Viewing actor linkages within each community, as well as topical differences across communities, conveys preliminary evidence supporting propositions that Twitter users engage in non-random relationships with an intent to access diverse information.

This chapter presents results from the exponential random graph models, which explore through quantitative means the first two sets of propositions, P1 and P2: P1 speaks to the following choices of users either within and across classes of professional affiliation,³⁴ thereby examining two competing (but not mutually exclusive) explanations of social selection processes contributing to network structure:

P1a: Actors choose whom to follow by mixing across affiliation types (i.e., via heterophilous relationships).

P1b: Actors choose whom to follow by matching on affiliation type (i.e., via homophilous relationships).

³⁴ Note that terms related to network edges, including network structure, linkages, ties, relationships, and connections are used interchangeably in this chapter, while terms related to a graph's vertices, including users, actors, and nodes, are also used synonymously.

The second set of propositions addresses the relationship between information distance and two actors' likelihood of connecting. In aggregate, this line of reasoning posits that R&D and entrepreneurial communities develop around disparities in topical content such that users can access and later recombine diverse knowledge for intellectual capital development (Nahapiet & Ghoshal, 1998). P2a positively associates information novelty (i.e., information distance) and the presence of following relationships, while P2b assesses whether information distance explains the following decisions of users better than actor affiliation mixing and matching alone.

P2a: Actors choose whom to follow based on the perceived novelty of information accessible through network linkages.

P2b: Information distance explains the following decisions of users better than actor affiliation mixing and matching alone can.

Although descriptive data from Chapter 5 offers cursory evidence regarding the relationship between actor type and local, community, and global network structure, the findings below speak directly to micro-level trends. As reviewed in Chapter 4, the ERGM is a statistical model that estimates parameters (i.e., sample regression coefficients) by isolating change statistics given the existing network structure. ERGMs allow the analyst to interpret the model's results in terms of the log-odds of an additional tie appearing. This chapter begins by introducing a subset of the combined friend network on which the above propositions are explored (Section 6.1), followed by a set of simplified models that guide the reader through the interpretation of ERGM results (Section 6.2). Section 6.3 presents results from ERGMs examining P1a and P1b, while Section 6.4 addresses P2a and P2b. I conclude with a brief summary of the main

findings (Section 6.5), which partially substantiate both P1a and P1b but offer no evidence supporting P2a and P2b.

6.1. Modeling Issues and an Alternative Approach

The modeling approach as described in Chapter 4 can be applied to the entire ecosystem sample (either the combined friend or following networks); however none of the several model specifications converge when using Markov Chain Monte Carlo (MCMC) simulation. This may be due to the presence of several different communities, some of which are relatively isolated from one another, in both graphs. Also problematic is the number of overall users in the network, with 5,000-10,000 users considered a large sized network for ERGM purposes (Goodreau, 2007). In any case, when running the ERGM in simulation mode with network structural parameters (e.g., gwesp, triangle, and mutual) using either the combined friend or follower networks, the models do not converge such that the log-likelihood improves and eventually stabilizes in any reasonable number of iterations (e.g., $n=100$). Therefore any results obtained from modeling the entire combined friend or follower networks are subject to intractable concerns about validity.³⁵

³⁵ In addition, as I learned through numerous modeling iterations over the course of six months, the implementation of the `ergm` library in the R package `statnet`, as of version 2014.2.0, does not handle large networks well. In particular, the `ergm` library is unable to model continuous nodal covariates in simulation mode (with other endogenous network structural controls); it responds with an error that indicates too little available memory despite adequate physical stores of RAM. More importantly, the library is also unable to handle continuous attributes at the dyadic level in large networks, in or outside of simulation mode. This second problem was a major impediment to the successful execution of the research design because the information distance explanatory variable used to evaluate P2a and P2b requires pair-wise continuous measures at the dyadic level. These limitations were brought to the attention of the `statnet` development team but unfortunately were not attended to in a timely fashion despite a handful of follow-up requests. One response indicated that the first error noted above was a systematic problem in the code and would take some time to triage and rectify. Regrettably, the second error was not addressed in a meaningful way by any member of the developer team. Eventually, however, I realized that `ergm` operated as specified and without major issues using smaller network sizes (e.g., fewer than 1000 nodes).

As a consequence of these modeling difficulties and yet in light of the promise of using the ERGM approach with smaller networks, I turn to one particular community in the combined friend network to quantitatively explore P1 and P2.³⁶ This community is C3 as shown in Figure 5.3, and it consists of over 46% of all intermediaries, almost 75% of nanotechnology firms, greater than 45% of other users, and more than 90% of all scientists in the combined firm friend network. In addition, in terms of topical content based on the PMI results, this community appears to be focused on nanotechnology R&D activity, broadly construed (e.g., in terms of news coverage or university support). Thus, both the actor composition and content focus of this cluster make it the most appropriate community in the combined friend network on which to explore the innovation ecosystem concepts underlying P1 and P2. C3 contains 2,942 actors in total and 205,287 directed edges; it has a network density of 2.4%. Recall from Chapter 5 that network density describes the ratio of actual ties to all possible ties; higher levels of network density convey greater levels of information sharing, diffusion, and general cohesiveness (Kadushin, 2011).

To further reduce the number of actors in C3 to fewer than 1000 nodes, I retain only those users that have authored a tweet containing the term “graphene” in an 18 month time span from the beginning of 2013 to mid-2014, as well as all graphene firms in the cluster. (Five of the seventeen graphene firms show no graphene tweets on record, yet they are included in the sample to maintain the network as one large connected

³⁶ Recall that the combined friend network is sampled by identifying the users that graphene firms follow. Using the friend relation as the basis for quantitative modeling necessarily omits the follower relationship; that is, the subsequent analyses look only at one type of relationship in the Twitter social graph. However, evaluating the following relationship allows us to clearly discuss the set of relationships as an articulation of graphene firms’ user agency. This is not the case with the follower relationships, where seemingly anyone can follow one of the sampled graphene firms.

component.) This results in a single sub-component of C3 containing 945 actors and 57,935 edges. I refer to this graph as C3-945. Contrary to initial expectations, this network is even denser than C3 at 6.5%. Thus, it appears that users tweeting about graphene in C3-945 are better connected as a whole than users in C3. C3-945 is comprised of 35.56% media entities, 22.86% intermediaries, 18.52% scientists, 8.89% other unidentifiable users, 7.3% nanotechnology firms, 3.6% other firms, 2.54% support firms, and 0.74% financial institutions. Because other firm, support firm, and finance users represent such small percentages of C3-945, I combine these users together into the “other firm” category for all model specifications below.

Instead of coloring nodes by community number as done in Chapter 5, I instead shade them by their actor affiliation type (see caption of Figure 6.1.) As mentioned in Chapter 4, Force Atlas positions nodes on a graph to minimize the distance between high-connectivity actors and to avoid overlapping edges. The graph shows some degree of actor mixing and also a fair amount of homophily. For example, the upper right corner shows a section dominated by connected scientists, while the bottom part reveals a number of intermediaries. Most of the nanotechnology and graphene firms can be found in the center and left parts of the graph. Media entities are broadly distributed throughout C3-945, as are many of the other firms and “unknown” actor types.

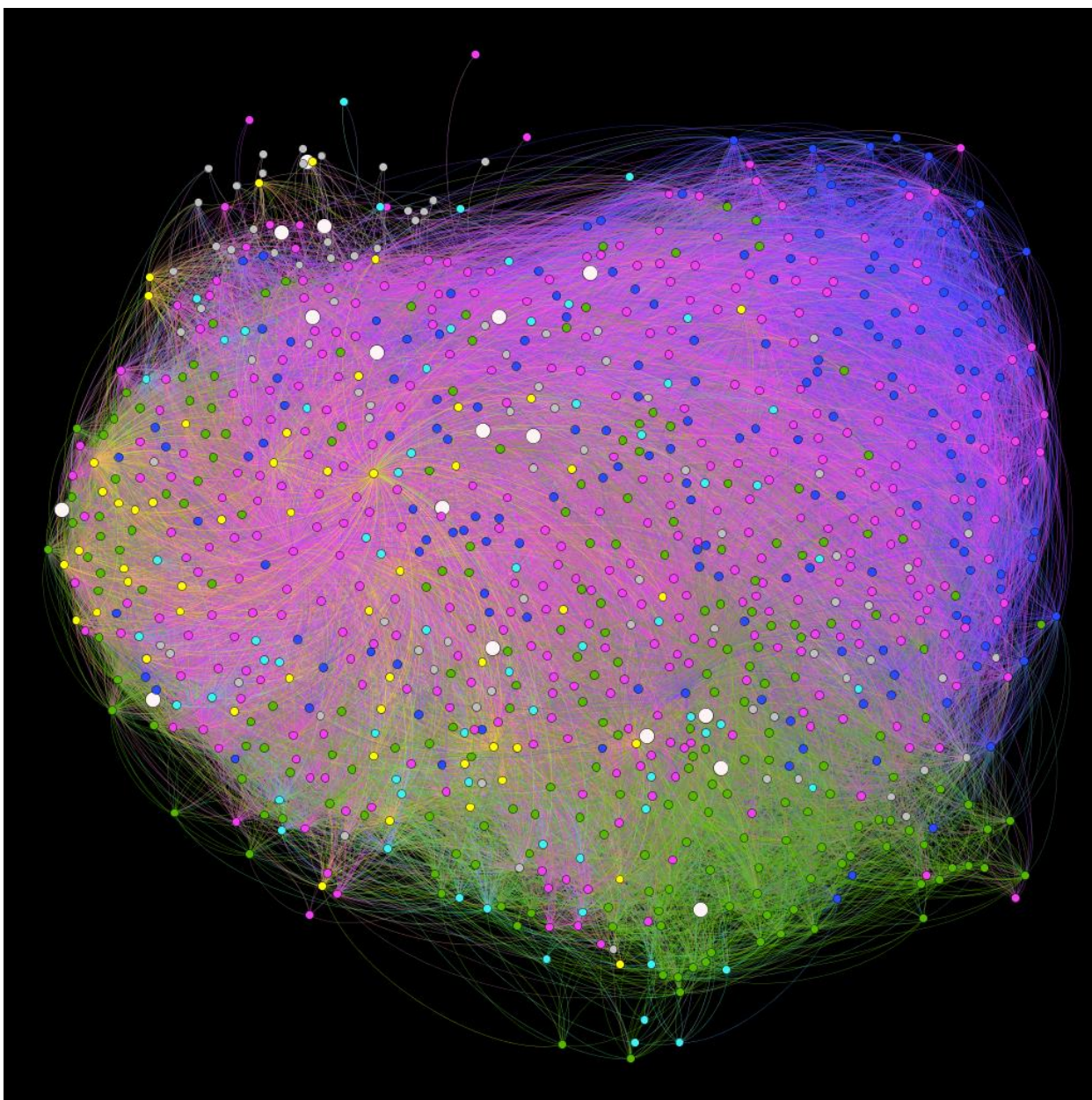


Figure 6.1: C3 community from the combined friend network. *Source:* Twitter; network consists of 2,942 users and 205,287 directed ties; data collected in early 2014. *Notes:* Seventeen of the 33 graphene firms are found in C3; they are shaded in white and sized larger than the other nodes. 144 other nanotechnology related firms are yellow and appear smaller in size. Other actors include 593 scientists (dark blue), 528 “unknown” users (gray), 739 media entities (magenta), 627 intermediaries (green), and 294 other firms (cyan), which include financial intuitions and support firms.

Descriptive statistics and correlation matrices for continuous actor attributes

variables are given in Table 6.1. All variables except *account_age* are log transformed

due to the skewed nature of the distributions. Most variables exhibit statistically significant positive correlations; for example, normalized tweeting activity is positively correlated with number of followers, number of friends, number of graphene tweets, and account age. Number of followers is positively correlated with number of friends and account age, and number of friends is positively correlated with account age. These positive correlations suggest that larger network sizes are often associated with more frequent usage. Finally, number of graphene tweets is negatively correlated with account age, indicating that newer users in C3-945 are among the most active tweeters in this domain.

Table 6.1: Descriptive statistics and correlations for continuous variables

variable	Mean	S.d.	Min	Max	(1)	(2)	(3)	(4)	(5)
(1) <i>ln_norm_status_cnt</i>	-0.26	1.79	-9.21	5.24	1				
(2) <i>ln_followers_count</i>	6.56	2.24	0	14.01	0.64*	1			
(3) <i>ln_friends_count</i>	5.76	1.61	0	9.73	0.37*	0.34*	1		
(4) <i>ln_graphene_tweets</i>	2.02	1.31	0	7.97	0.13*	-0.01	0.04	1	
(5) <i>account_age</i>	1374.12	557.86	123	2695	0.33*	0.56*	0.13*	-0.09*	1

Source: Twitter, n=945 users collected in early 2014. *Notes:* Descriptive statistics and correlations for continuous actor attributes in C3-945. * Denotes significance at $\alpha=.01$.

6.2. ERGM Example Run

While ERGMs are becoming more common in the toolkits of social scientists, they are not yet broadly used (Goodreau, 2007). The purpose of this section is to provide a detailed example demonstrating how this statistical approach differs from logit and probit regression, where the dependent variable also takes the form of 1 or 0. The difference between ERGM and logit or probit models, as noted in Chapter 4, is the

assumption of independence across observations (in logit or probit) or dependence across observations (in ERGMs). ERGMs are specifically tailored for social network analysis in terms of accounting for endogenous network activity. Endogenous variables (e.g., mutuality, triangles, and higher order organizing phenomenon such as k-stars) capture the process by which certain network links develop in response to the existence of other network connections (Lusher & Robbins, 2013). When ERGMs do not include endogenous network structural controls, they become very similar to logit specifications (Robins & Lusher, 2013). This is because in social selection models with network structure as the dependent variable, all nodal and/or dyadic attribute covariates are considered exogenous and thus independent of one another. For example, being a graphene firm is not likely to be influenced by whether or not a user is in contact with other graphene firms on Twitter, and thus this variable (*graphene_firm*) is exogenous.

I present the full output of three related models to show how ERGMs work with and without endogenous structural variables (Tables 6.2-6.4). Table 6.2 is the simplest model possible: The edges parameter is analogous to an intercept (Robins, Lewis, & Wang, 2012), and the negative statistic is often reported as the log-odds of any tie occurring. In this case, the log-odds is -2.667, and the corresponding probability of any tie occurring in the sample is $\frac{e^{(-2.667)}}{1+e^{-2.667}} = .065$. Recall that the density of C3-945 is 6.5%. Thus, an ERGM model with no other attributes gives the log-odds of the edges parameter that corresponds with the graph's density. The caption of Table 6.2 contains both Akaike information criterion (AIC) and Bayesian information criterion (BIC) measures, which report model quality as a tradeoff between number of parameters (parsimony) and explanatory power. Smaller values are better.

Table 6.2: Basic ERGM model on C3-945

	Estimate	Std. Error	MCMC %	p-value
<i>edges</i>	-2.667	0.004	NA	0.00

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Note:* AIC=428841, BIC=428853.

Table 6.3 includes one exogenous binary variable that captures whether two graphene firms are more or less likely to maintain a directed following tie (hence the variable prefix “nodematch”). More specifically, it tests the proposition that graphene firms are more likely to follow each other in C3-945 than other types of non-graphene users. Because ERGMs report parameters (i.e., regression coefficients) on change statistics assuming the addition of a tie conditional on all other ties, sets of model attributes can be interpreted by isolating the context in which a given tie appears. For instance, if a tie links a graphene firm to a non-graphene firm, its log-odds is -2.979 (4.8%). However, if the tie occurs between two graphene firms, then its log-odds is $-2.979 + 1.130 = -1.849$ (13.6%). Finally, if the tie occurs between two non-graphene firm users, its log-odds is -2.658 (6.5%). This finding reveals that graphene firms are more likely to follow one another (13.6%) than non-graphene firm actors are likely to follow each other (6.5%). Additionally, graphene firms are more likely to follow each other than they are non-graphene firms (4.8%). . Thus, this model provides partial support for P1b, which suggests that actors choose whom to follow based on homophilous social selection processes.³⁷

³⁷ Note that there are no mixing terms across actor categories (e.g., between scientists and intermediaries), so this model says nothing about P1a.

Table 6.3: Adding a binary attribute for whether actors are graphene firms

	Estimate	Std. Error	MCMC %	p-value
<i>edges</i>	-2.979	0.026	NA	0.00
<i>nodematch.graphene_firm.0</i>	0.321	0.027	NA	0.00
<i>nodematch.graphene_firm.1</i>	1.130	0.179	NA	0.00

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Notes:* AIC=428667, BIC= 428702. In contrast to the baseline ERGM specification containing just edges, the AIC and BIC improve in this model.

What happens when we take into account the propensity of Twitter users to reciprocate following relationships? In the social network literature, mutuality in directed ties is considered an endogenous network feature (Kadushin, 2011). Endogenous network parameters cannot be estimated without simulation; thus we see the “MCMC (Markov chain Monte Carlo) percentage” column populated with numeric values (Table 6.4). Furthermore, p-values are not the only criterion by which to judge a model’s accuracy in simulation mode. We also need to ensure the model converges; i.e., that it shows stability in log-likelihood improvements after each round of maximum likelihood estimation. If the model fails to converge (or even run), this indicates that the ERGM cannot fit the observed network data; it is *degenerate*. Model degeneracy does not signify weaknesses in the MCMC estimation procedure per se but rather with a given model specification. “The solution is to specify a better-fitting model for the data, but this is less straightforward for networks than for other statistical contexts. In ERG modeling, a

misspecified model can fail to converge, yielding no parameter estimates to guide model diagnosis or respecification” (Goodreau, Kitts, & Morris, 2009, p. 110).³⁸

The model depicted in Table 6.4 was run in a control environment with up to 100 iterations, though it converged well before then. Here, we see that the log-odds of a following relationship appearing without a mutual tie between graphene firms and non-graphene firm users is -3.337 (3.4%). Contrast this to a tie between two graphene firms (i,j) where a mutual following relationship already exists between (j,i); the log-odds is 0.103 or a probability of 52.6%. This suggests that once graphene firms find out about each other online (i.e., not necessarily that the firms exist, but that the firms have a Twitter account), they are quite likely to follow one another. The AIC and BIC improve considerably over the model specification in Table 6.3. This interpretation offers even stronger support for P1b.

Table 6.4: Adding a term for mutuality

	Estimate	Std. Error	MCMC %	p-value
<i>edges</i>	-3.337	0.025	0.00	0.00
<i>nodematch.graphene_firm.0</i>	0.241	0.025	0.00	0.00
<i>nodematch.graphene_firm.1</i>	0.821	0.174	0.00	0.00
<i>mutual</i>	2.619	0.015	0.00	0.00

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Note:* AIC= 403038, BIC: 403085.

³⁸ Recall that a lack of convergence with the combined follower and friend networks led to the eventual selection of C3-945 as the network to model.

To test the validity of the model, goodness of fit and model degeneracy analyses can be applied. For illustrative purposes, I review these techniques here in some detail in order to briefly summarize robustness checks on the main findings later below.

A correctly specified ERGM that converges on an observed network produces a probability distribution over similar networks of the same size. Sampling from this distribution should ideally produce networks that are related to the observed network. The purpose of goodness of fit measures, therefore, is to examine whether features show comparable frequency distributions between the simulated networks and observed network; it is a validity check of sorts, often applied to non-explicitly modeled attributes (Goodreau, 2007). Since the ERGM accounts for micro-level network activity which aggregates to higher level meso and global structural characteristics, consistency across the simulated and observed networks at these higher levels provide verifiable signs of model robustness. For instance, suppose that in-degree and out-degree are not included as parameters in the ERGM. Yet, in the network literature, distributions of centrality scores are important in characterizing overall network topography (Goodreau et al., 2009). Examining goodness of fit using centrality scores consequently offers one way to assess whether the simulated networks produced by the ERGM “look like” the observed network.

Figures 6.2 and 6.3 illustrate the in-degree and out-degree sample distributions of the simulations produced by the ERGM in Table 6.4 vis-à-vis the observed in-degree and out-degree values from C3-945. Figure 6.2 shows a black trend line that reflects the proportion of nodes in C3-945 exhibiting in-degrees from 0-25 in the observed network. The box plots in the background reveal the first and third quartiles of the simulated

network in-degree distributions. The trend line falls within an acceptable range for in-degrees beginning at 3. Figure 6.3 shows the analogous diagram for out-degree. Here the deviation between the observed data and simulated networks appear most significant at out-degree 0.³⁹

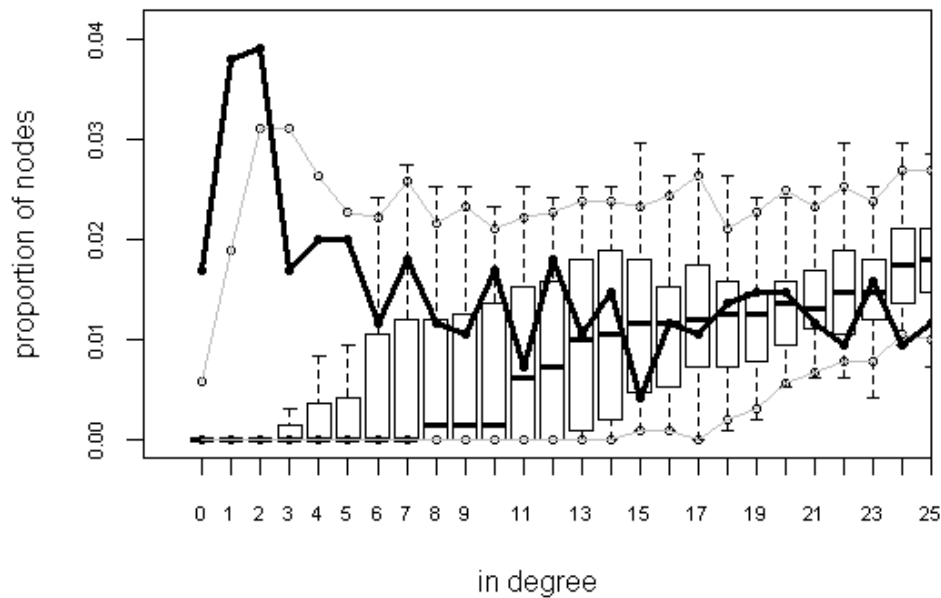


Figure 6.2: Goodness of fit diagnostics for in-degree distribution, 0-25.
Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Note:* Diagnostics run on ERGM reported in Table 6.4.

³⁹ Because the ERGM does a poor job of predicting a relatively large number of nodes with few in-degree or out-degrees, it makes sense to attempt to fit a new model specification controlling for these factors, e.g., adding parameters for in-degrees 0, 1, and 2, as well as out-degrees 0, 1, and 2. However, in several specifications of in and out degree combinations, the resulting models failed to converge.

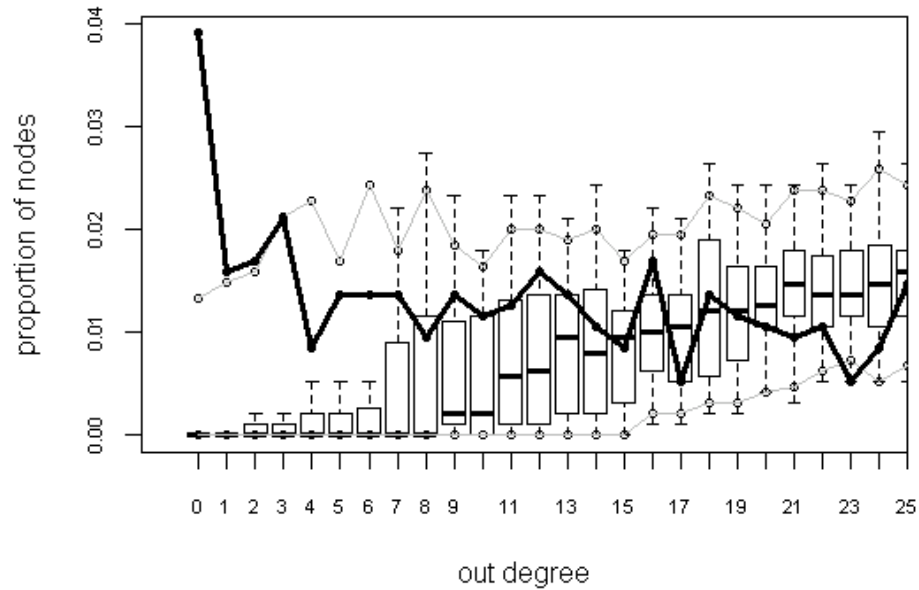


Figure 6.3: Goodness of fit diagnostics for out-degree distribution, 0-25.
Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Note:* Diagnostics run on ERGM reported in Table 6.4.

The next step in assessing model robustness is a series of tests to determine whether the model is *degenerate*. Although the model presented in Table 6.4 converged, it is possible that individual sample statistics significantly differ from the observed values. Indeed, this is the case with the mutual parameter, which shows a moderately significant p-value at 0.049 in Table 6.4. Other signs of degeneracy at the individual variable level include (a) tests for auto-correlation across sample statistics taken in intervals from the simulated networks and (b) Geweke statistics that test whether means from two parts of the Markov chain (at 0.1 and 0.5 “windows”) are from the same distribution. The p-values for both the auto-correlation and Geweke statistics are not significant in the model specification found in Table 6.4.

The simple models presented in this section compare network relationships between graphene firms and non-graphene firm users. In the following sections, I turn to broader classes of ecosystem actors; i.e., nanotechnology firms, of which graphene firms are a part; scientists; media entities; other firms; unknown users with no readily identifiable professional affiliation; and intermediaries.

6.3. Testing the Relationship Between Actor Type and Network Structure

This section quantitatively explores P1a and P1b, which propose that ecosystem actors on social media engage in social selection processes guided by heterophily and homophily, respectively. There are six actor classes in all, and therefore a total of 36 different types of “mixing” relationships. To limit the quantity of pair-wise variables to a manageable number, only select actor mixing variables are included in the model; that is, I include mixing variables when there is a theoretical or methodological motivation for doing so. For instance, because C3-945 contains many nanotechnology firms and because innovation necessarily requires the involvement of high-technology firms, I specify terms for all following relationships from other classes of actors to nanotechnology firms, and from nanotechnology firms to other actors. As another example, I specify mixing variables between intermediaries and media entities and between scientists and media entities due to the likelihood that the media acts as a valuable source of S&T news, even for experts and reputable institutions. Mixing variables (beginning with *mix.actor_type*) examine heterophily when two actors are of dissimilar classes (e.g., scientist and firm), and matching terms (beginning with *match.actor_type*) assess homophily when the actors are within the same class.

Mixing variables excluded from the model represent cases with unconventional pairings not necessarily relevant to the innovation context.⁴⁰ For example, whether unknown users follow media users is of little relevance for the present study. Furthermore, limiting the number of actor mixing variables also helps with model convergence. This particular combination of actor mixing variables initially converges in a reasonable number of iterations (<20) in MCMC mode, a boon that becomes increasingly important with the addition of continuous control and explanatory variables in later models, which require up to 100+ iterations over 24+ hours to estimate.⁴¹

To reiterate, the reference group with respect to mixing variables is the set of ties between different actor categories *not explicitly modeled*. To make the presentation of findings easier to digest, phrases such as “less likely” and “more likely” refer to comparisons with this reference group. Out of 57,935 following relationships in C3-945, 37,185 are reflected in the mixing variables while 20,750 are in the reference group.

Because the number of observations is so large ($945 * 945 = 893,025$), only variables significant at the $p < .01$ level are inferred as statistically significant in the larger population of Twitter users. In general, large datasets produce narrow standard errors which give rise to statistical significance even in cases where the magnitude of the parameter’s change on the dependent variable is very small (Wooldridge, 2003). Thus setting a relatively low tolerance level for Type I errors (e.g., $\alpha = .01$) makes it more

⁴⁰ This includes relationships from intermediaries to unknown users; from media entities to media entities, unknown users, and other firms; from unknown users to intermediaries, media entities, other unknown users, and other firms; other firms to intermediaries, media entities, unknown users, and other firms; and scientists to unknown users and to other firms.

⁴¹ On a machine with 4 2.70GHz 8 core processors and 512GB of RAM.

difficult to conclude a statistically significant effect (vis-à-vis a higher tolerance level for Type I errors).

Table 6.5 shows the first two models, M1 and M2, both of which include the same set of mixing variables. M2 also includes a term for reciprocity (“mutual”). In M1, six of the eight positive sample regression coefficients on the mixing variables are significant. In contrast, thirteen of the mixing variables show negative coefficients, twelve of which are significant. Table 6.6 contains a summary of these findings: Results for mixing and matching variables in M1 show a positive or negative sign or a symbol for not significant. Cell values without such notation belong to the reference group of mixing relationships.⁴²

It appears that without additional usage controls, the ERGM shows a highly bifurcated network structure with many ties leading *to* intermediaries and media entities *from* intermediaries, nanotechnology firms, and scientists. On the other hand, there is a lower probability of following ties emanating (a) from nanotechnology firms to unknown actors and scientists and (b) from intermediaries to nanotechnology firms, other firms, and scientists. Actors with no recognizable affiliation, while included in C3-945 because they are followed by at least one graphene firm, are not likely to be followed by nanotechnology firms as a whole, nor are they more or less likely to follow those nanotechnology firms in turn. As suggested in the descriptive overview in Chapter 5, this finding indicates “unknown” users are not broadly embedded in the innovation network; i.e., that a real and revealed identity plays an important role in developing social connections online.

⁴² Relationships not explicitly modeled are in the reference group by default. See Section 4.2.2 and text in this section for a complete discussion.

Table 6.5: Actor mixing parameters with usage and network structural controls

Parameter	M1	M2	M3
<i>Edges</i>	-2.691 ***	-3.145 ***	-5.185 ***
<i>match.actor_type.intermediary.intermediary</i>	0.223 ***	0.145 ***	0.020
<i>mix.actor_type.media.intermediary</i>	0.006	-0.168 ***	-0.264 ***
<i>mix.actor_type.nano_firm.intermediary</i>	0.417 ***	0.681 ***	0.366 ***
<i>mix.actor_type.scientist.intermediary</i>	-0.142 ***	0.111 ***	-0.101 ***
<i>mix.actor_type.intermediary.media</i>	0.341 ***	0.376 ***	0.167 ***
<i>mix.actor_type.nano_firm.media</i>	0.814 ***	1.018 ***	0.636 ***
<i>mix.actor_type.scientist.media</i>	0.540 ***	0.697 ***	0.421 ***
<i>mix.actor_type.intermediary.nano_firm</i>	-0.783 ***	-1.104 ***	-0.606 ***
<i>mix.actor_type.media.nano_firm</i>	-0.287 ***	-0.782 ***	-0.235 ***
<i>match.actor_type.nano_firm.nano_firm</i>	0.377 ***	0.248 ***	0.572 ***
<i>mix.actor_type.unknown.nano_firm</i>	-0.965 ***	-0.842 ***	-0.520 ***
<i>mix.actor_type.other_firm.nano_firm</i>	-0.308 ***	-0.402 ***	0.035
<i>mix.actor_type.scientist.nano_firm</i>	-0.789 ***	-0.796 ***	-0.360 ***
<i>mix.actor_type.nano_firm.unknown</i>	-0.699 ***	-0.476 ***	0.010
<i>mix.actor_type.intermediary.other_firm</i>	-0.949 ***	-0.969 ***	-0.788 ***
<i>mix.actor_type.nano_firm.other_firm</i>	0.070	0.170	0.241 ***
<i>mix.actor_type.intermediary.scientist</i>	-0.977 ***	-1.044 ***	-0.639 ***
<i>mix.actor_type.media.scientist</i>	-0.259 ***	-0.581 ***	-0.140 ***
<i>mix.actor_type.nano_firm.scientist</i>	-0.265 ***	-0.065	0.163 **
<i>mix.actor_type.unknown.scientist</i>	-1.299 ***	-1.321 ***	-0.944 ***
<i>match.actor_type.scientist.scientist</i>	-0.031	-0.038	0.300 ***
<i>mutual</i>		2.760 ***	3.138 ***
<i>nodeicov.ln_followers_count</i>			0.376 ***
<i>nodeocov.ln_followers_count</i>			-0.071 ***
<i>nodeicov.ln_norm_status_cnt</i>			-0.137 ***
<i>nodeocov.ln_norm_status_cnt</i>			0.028 ***
<i>nodeicov.account_age</i>			0.000 ***
<i>nodeocov.account_age</i>			-0.000 ***
AIC	421,388	394,410	365,268

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. Notes: ** $\alpha = .01$, *** $\alpha = .001$. Alternate specifications of M3 were attempted with parameters for in- and out-bound statistics for *ln_friends_count*, but these ERGMs did not converge in fewer than 250 iterations.

Like unknown users, nanotechnology firms as a whole are less likely to be followed by the other categories of ecosystem actors. However, nanotechnology firms

are more likely to be followed by other nanotechnology firms, suggesting that firm homophily is an important social selection process in social media networks. The coefficient on within-category network ties among intermediaries is also positive and significant, indicating support for homophily as a positive social selection here as well. However, in this model specification, scientists are not more or less likely to follow each other (than the reference group).

Table 6.6: Summary of s.s. sample coefficients, as reported in M1

From	To					
	inter- mediary	media	nano firm	unknown	other firm	scientist
intermediary	+	+	-		-	-
media	n.s.		-			-
nano firm	+	+	+	-	n.s.	-
unknown			-			-
other firm			-			
scientist	-	+	-			n.s.

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Notes:* All signs significant at $\alpha = .001$. Empty cells indicate ties constituting the reference group.

In sum, M1 shows limited support for P1a because heterophily as a positive predictor of network structure only occurs when ties are directed to media and intermediary actor types. M1 also shows limited support for P1b due to the positive significant coefficients for within category ties in the intermediaries and nanotechnology firm actor classes.

M2 adds a control for reciprocity. Rather than re-interpret the model completely, I instead highlight two major differences between M1 and M2. First, in M1, the log-odds

of an actor following another actor in the reference group of mixing and matching relationships is -2.691 (6.4%). In M2, the log-odds of an actor *i* following another actor *j* in the reference group *without* a reciprocal tie (*j,i*) is -3.145 (4.1%); with a reciprocal tie it is -0.385 (40.5%). Reciprocity, therefore, is an important control: The magnitude of the “mutual” coefficient accounts for more explanatory power than any of the actor mixing and matching variables alone (*vis-à-vis* M1), and the AIC decreases from 421,388 in M1 to 394,410 in M2 just by adding this one term.

Controlling for reciprocity in M2 always results in a lower log-odds of following when the tie is not reciprocated (*vis-à-vis* the baseline statistic in M1). For example, the log-odds of intermediaries following nanotechnology firms in M2 without a reciprocal relationship is -4.249 (1.4%). If the reciprocated relationship exists, the log-odds improves to -1.489 (18.4%). This suggests that an intermediary will follow a nanotechnology firm 18.4% of the time when a nanotechnology firm follows the intermediary. In M1 where mutuality is not included as a control, the log-odds of intermediaries following nanotechnology firms is -3.474 (3.0%). In another case, the log-odds of scientists following media entities in M2 without reciprocation is -2.448 (7.9%); with reciprocation it is 0.312 (57.7%). In M1, the log-odds of scientists following media entities without controlling for mutuality is -2.151 (10.4%). This analysis shows that even when paired with certain actor types who may not be as likely to “follow back” – e.g., media entities – reciprocation positively and significantly predicts tie formation.

Second, two coefficients change directions between M1 and M2. First, while the coefficient on following relationships between media entities and intermediaries is positive but not significant in M1, it becomes negative and significant in M2. In M2, this

indicates that media entities are less likely to follow intermediaries than the reference group overall when ties are not reciprocated. Second, while the parameter on following relationships between scientists and intermediaries is negative and significant in M1, it becomes positive and significant in M2. This implies that once we control for mutuality scientists are more likely to follow intermediaries than the reference group even when the tie is not reciprocated.

In general, M1 and M2 both show the bifurcation of the network in terms of positive and significant coefficients on following relationships toward media entities and intermediaries and negative and significant statistics elsewhere (again in relation to the reference group). As described in Chapter 5, media and intermediary types exhibit the largest average follower bases vis-à-vis the other actor categories; they are also among the most active tweeters. Assuming these characteristics significantly predict network structure in C3-945, what happens to the model results once usage controls are included? M3 controls for number of followers, normalized status (tweet) activity, and account age, and it provides two sets of important results.

The first finding from M3 provides evidence substantiating actors' following choices based on actor mixing (P1a): the bifurcation in the network apparent in M1 and M2 partially recedes in M3, as we see the first signs of the type of actor mixing assumed in the innovation ecosystem construct (see Table 6.7 for a summary). One parameter on the following relationship from nanotechnology firms to scientists changes direction from negative in M2 to positive in M3. In addition, (a) the sample regression coefficient from nanotechnology firms to other firms changes from not significant to positive, and (b) the insignificant negative coefficient in M2 on the following relationship from

nanotechnology firms to scientists becomes positive and significant in M3. Also of interest is the change in sign from other firms to nanotechnology firms from negative in M2 to positive but not significant in M3. Comparing Tables 6.6 and 6.7 and reading from left to right in the third row, these changes in signs and significance are readily identifiable: In M3, which controls for usage factors and reciprocity, nanotechnology firms are more likely to follow each type of actor (besides unknown users) in the ecosystem than the reference group at large. In contrast, M2 shows the aforementioned bifurcation of following relationships where scientists, nanotechnology firms, and other firms are likely to follow across actor categories only when the tie is directed to an intermediary or media entity. Therefore, this evidence from M3 provides some support for the heterophily argument as being an important predictor of network structure, at least from the perspective of nanotechnology firms.

Table 6.7: Summary of s.s. sample coefficients, as reported in M3

From	To					
	inter- mediary	media	nano firm	unknown	other firm	scientist
intermediary	n.s.	+	-		-	-
media	-		-			-
nano firm	+	+	+	n.s.	+	+
unknown			-			-
other firm			n.s.			
scientist	-	+	-			+

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Notes:* All signs significant at $\alpha = .01$ or $\alpha = .001$. Empty cells indicate ties constituting the reference group.

The nature of homophily as a predictor of network structure also changes from M2 to M3: similarity between actor types also positively predicts network structure in M3, albeit in a slightly different way. In M3, no longer is the positive sample regression coefficient on following relationships among intermediaries significant. Like in M2, nanotechnology firms are more likely to follow each other in M3. However, in M3, we see a change in sign on the statistic in the following relationship between scientists from negative to positive vis-à-vis the reference group. This finding agrees with the reviewed literature that shows anecdotal evidence for professionalized interaction among scientists on Twitter (Bonetta, 2009; Reich, 2011).

A series of results related to the control variables – i.e. reciprocity, number of followers, normalized status (tweet) activity, and account age – offer an interesting perspective on social selection dynamics. Because the edge parameter estimate decreases in magnitude from a log-odds of -3.145 in M2 to -5.185 M3, we see that all four usage controls account for a great deal of the variance in the dependent variable and subsequently act as important predictors of network structure. For example, the log-odds of the “mutual term” increases from 2.760 in M1 to 3.138 in M3. Overall, M3 performs well (AIC: 365,268) and shows a vast improvement over M1 and M2.

M3 includes six variables for the three types of user controls, follower count, normalized status count, and account age. The ERGM includes two terms for each variable, one for the node sending the following tie and the other for the node receiving the following tie (i.e., the friend). I specify the natural log of follower count and normalized status count because of the highly skewed nature of the two distributions, though the log-log nature of the model makes the ERGM slightly more difficult to

interpret. Thus, to improve readability, I rely on language that interprets statistically significant changes in log-odds as general increases or decreases in the log-odds or probability of a tie occurring. However, the proper way to discuss the meaning of natural logged variables is to interpret the log-odds coefficients as an increase or decrease in the dependent variable as a function of scaling up the natural logged variable by a factor of e , or roughly 2.718.

The more followers a user maintains, the more likely he is to be followed by members of C3-945. This comes as no surprise due to the “Matthew Effect” of preferential attachment in social networks whereby popular actors get more popular over time (Barabási, 1999; De Bellis, 2009). Although the dataset is cross-sectional, it is possible to generalize by asserting that popular actors in the broader Twitter social network are also popular in C3-945. However, consider the significant negative coefficient on the log-odds for the follower node: The more followers this user has, the less likely he is to initiate a following relationship overall. This finding is also in line with prior work on social media, which shows that some users with high numbers of followers have very few friends (Kwak, Lee, Park, & Moon, 2010). By controlling for number of followers, it becomes possible to isolate the effects of preferential attachment from certain high follower actor types (e.g., media entities and intermediaries), a consideration that will be explored in more detail shortly.

Similarly, the older an account is, the less likely it is for that user to initiate following relationships. At the same time, newer accounts are less likely to attract following relationships. This finding also agrees with the Matthew Effect, which predicts that actors who have been around the longest are often the most popular nodes in a social

network; that is, older and established users develop large following bases due to their increased visibility in the social network vis-à-vis younger and less well connected users. At first glance, the magnitude of the coefficients on *nodeicov.account_age* and *nodeocov.account_age* appear minuscule in comparison to the other usage controls. However, recall that the parameters represent a change in tie probability given an additional *day of age* for the sender and receiver, respectively.

The more frequently an actor tweets, the less likely he is to attract followers in C3-945. Conversely, the more frequently an actor tweets, the more likely it is for him to follow another user (in the reference group). This counterintuitive finding is that excessive tweeting will not attract followers in C3-945, holding all other variables constant. Instead, users that frequently tweet are more likely to *initiate* following relationships. In S&T domains, then, social media may reflect the ecosystem community's desire for meaningful, not superfluous, content. This finding of "less tweeting attracts more followers" in C3-945 runs counter to the positive correlation reported between number of overall followers and normalized tweet activity in Table 6.1.⁴³

In terms of goodness of fit metrics, there is not a great deal of change between the in-degree and out-degree distributions in M2 or M3 and the trends shown in Figures 6.2 and 6.3. Much of the observed degree distribution is within the acceptable range of the metrics obtained through simulation, with the exception of those actors with very small in-degree and out-degree values. Controlling for in-degree and out-degree to improve

⁴³ But this finding is commensurate with the interview data showing that too much tweeting can be unproductive. "It's hard to tweet more than once or twice a day. If you share garbage, people don't like it and they'll be done with you" (F2). See Chapter 7 for further details.

goodness of fit was not successful, as parameters added to either M2 or M3 failed to produce convergent ERGMs.

Both M2 and M3 show some signs of degeneracy. For example, in M3 individual sample statistics significantly differ (at $p < .05$) from the observed values for the parameters, *mix.actor_type.other_firm.nano_firm*, *mix.actor_type.media.nano_firm*, *match.actor_type.nano_firm.nano_firm*, *nodeicov.account_age* and *nodeocov.account_age*. Geweke statistics that test whether means from two parts of the Markov chain (at 0.1 and 0.5 “windows”) are from the same distribution are significant (i.e., this null hypothesis is rejected) for the variables: *mix.actor_type.media.nano_firm*, *mix.actor_type.nano_firm.other*, *mix.actor_type.intermediary.scientist*, *mix.actor_type.unknown.scientist*, and *nodeocov.account_age*. However, autocorrelation across sample statistics does not appear to be a problem across the simulations.

6.4. The Mediating Role of Information Distance

The purpose of this section is to explore the second set of propositions addressing the relationship between information distance and two actors’ likelihood of connecting. Recall that P2a suggests a positive association between information novelty (i.e., information distance) and the presence of following relationships, while P2b assesses whether information distance explains the following decisions of uses better than actor affiliation mixing and matching alone can.

This section is organized in three parts. First, I offer descriptive results of the topic modeling method that underlies two related operationalizations of the information distance measure. Second, I present the results of including these variables into two ERGMs. Because the findings do not provide support for the first proposition, I turn to a

post-hoc analysis to shed some additional light on how disparities in information content predicts network structure in the graphene-based Twitter ecosystem.

6.4.1. Topic Modeling Descriptive Results

LDA is run on two sets of training corpora, (a) a sample of tweets issued by members of C3-945, up to 200 tweets per user, and (b) a set of 100,000 random tweets containing the word “graphene” sampled across all of Twitter and authored between 2013 and the first quarter of 2014. The rationale behind this decision stems from the divergent content (and subsequent topics) that emerge from the different training sets. As we’ll see, the topics identified from the C3-945 set of tweets encompass a variety of S&T and non-S&T related content. Conversely, the topics identified from the set of 100,000 graphene tweets appear highly related to the R&D and commercialization of nanoscience; broader areas of materials science, chemistry, and physics; and recent or emerging application areas of graphene. As discussed in Chapter 4, the number of topics K , as well as well as the smoothing parameters α and β , were determined by identifying the set of parameters that produce the maximum log-likelihood of the posterior distribution of words to topics. For the set of C3-945 tweets, $K=50$, $\alpha=.25$, and $\beta=.10$. For the set of graphene tweets, $K=150$, $\alpha=.25$, and $\beta=.025$. See the Appendix (Tables A6.1 and A6.2) for data substantiating these choices.

In this section, I provide descriptive results in terms of words comprising select topics from each of the training sets for three users. The three users represent a media entity (@PhysicsWorld), a nanotechnology firm (@DGSGraphene), and a scientist (@drskyskull). The purpose of this descriptive overview, thereafter, is to illustrate how

information distance scores change with topical similarity (or dissimilarity) between users.

Table 6.8 provides the topics trained on the sample of tweets authored by C3-945 users. Reading from left to right, Table 6.8 shows that @PhysicsWorld tweets about physics news, space, and graphene in physics research. In contrast, the scientist @drskyskull is likely to author tweets that span a range of non-S&T topics drawing on a lexicon of day-to-day activities and foreign affairs. The nanotechnology firm @DGSgraphene tweets about topics related high-technology entrepreneurship, including content on awards, the graphene market, and conferences. How do the asymmetric information distance scores dpq differ between these three users given the 50 topics trained on C3-945's tweets? The average dpq score is 6.559 with a sample standard error of 2.946, a minimum of 0.0, and a maximum of 18.280. The distance of (a) @drskyskull from @PhysicsWorld is 5.099, and 3.241 from @drskyskull to @PhysicsWorld; (b) @DGSgraphene from @PhysicsWorld is 5.578, and 5.490 from @DGSgraphene to @PhysicsWorld; and (c) @DGSGraphene from @drskyskull is 8.608, and 8.680 from @DGSGraphene to @drskyskull.

Table 6.9 contains topics trained on 100,000 random graphene tweets, and even for these same three users, the topical content changes accordingly to relate graphene centric content. Both the media user @PhysicsWorld and the scientist @drskyskull are most likely to share news commentary related to graphene, while the nanotechnology firm @DGSgraphene is likely to tweet about content pertaining to live events, research infrastructure, and liquidity events. (Indeed, this firm went public on the London Stock Exchange in 2013.) How do the asymmetric information distance scores dpq differ

between these three users given the 150 topics trained on graphene tweets? The average dpq score is 5.554 with a sample standard error of 2.431, a minimum of 0.0, and a maximum of 19.173. The distance of (a) @drskyskull from @PhysicsWorld is 3.900, and 2.687 from @drskyskull to @PhysicsWorld; (b) @DGSgraphene from @PhysicsWorld is 6.945, and 5.577 from @DGSgraphene to @PhysicsWorld; and (c) @DGSGraphene from @drskyskull is 6.138, and 5.472 from @DGSGraphene to @drskyskull.

Table 6.8: Comparison of top three topics trained on sample tweets from C3-945 for three select users

User	Most likely topic	Second most likely topic	Third most likely topic
@PhysicsWorld (media)	Topic 14 – “physics news” (15.7%) physics, cern, particle, matter, universe, dark, higgs, black, big, science	Topic 13 – “space” (12.3%) space, nasa, earth, mars, moon, telescope, water, science, planet, launch	Topic 1 – “graphene in physics” (10.5%) quantum, researchers, physics, magnetic, graphene energy, dots, computing, scientists, lab
@drskyskull (scientist)	Topic 44 – “personal” (38.1%) good, don, people, time, make, things, work, thing, love, life	Topic 12 – “intellectual property law” (15.0%) court, supreme, extrusion, law, plastics, laws, packaging, patent, media, @iplawalerts	Topic 46 – “foreign affairs” (8.5%) world, war, india, china, http, Russia, people, obama, news, state
@DGSgraphene (nanotechnology firm)	Topic 38 – “high-tech hype” (39.0%) award, prof, congratulations, congrats, research, science, awards, professor, wins, university	Topic 9 – “graphene market” (22.3%) graphene, material, graphite, nasdaq, grapheneweek, university, graphenelabs, nyse, lomiko, applications	Topic 35 – “conference” (13.9%) conference, week, booth, today, june, visit, deadline, international, open, day

Source: Twitter, up to n=200 tweets from each user’s timeline; data collected in early 2014

Table 6.9 shows more overlap in the top three topics than does Table 6.8, with the user @drskyskull sharing his top three topics with the other two users, @PhysicsWorld

and @DGSgraphene. As a consequence, the asymmetric distance scores between @PhysicsWorld and @drskyskull, and between @drskyskull and @DGSgraphene, narrow when comparing values derived from training LDA on C3-945 tweets vs. the random sample of tweets containing the word “graphene”.

Table 6.9: Comparison of top three topics trained on 100,000 random graphene tweets for three select users

User	Most likely topic	Second most likely topic	Third most likely topic
@PhysicsWorld (media)	Topic 106 – “news commentary” (14.1%) future, good, article, interesting, read, material, cool, stuff, incredible, make	Topic 29 – “more news commentary” (6.2%) video, made, material, future, making, amazing, great, watch, condoms, workable	Topic 18 – “physics research” (5.8%) arxiv, cond, hall, quantum, spin, electrons, mat, bilayer, electron, dirac
@drskyskull (scientist)	Topic 106 – “news commentary” (35.4%) [see above]	Topic 29 – “more news commentary” (7.4%) [see above]	Topic 107 – “live event” (4.3%) live, ceo, event, talk, today, polyakova, berlin, ways, conference, graphite
@DGSgraphene (nanotechnology firm)	Topic 107 – “live event” (51%) [see previous cell]	Topic 61 – “research infrastructure” (7.4%) budget, cell, centre, therapy, science, turing, data, research, centres, innovation	Topic 130 – “liquidity event” (7.1%) aim, applied, materials, float, ipo, company, million, foams, conductive, material

Source: Twitter, up to n=200 tweets from each user’s timeline; data collected in early 2014

Why is it important to compare values across different corpora using a case based approach such as the one presented above? Because of the relative novelty of using topic modeling for social science research, such an analysis provides much needed face validity to the measure as a way of examining network structure. It also reveals how

different training sets can alter the values of information distance between any two actors. I now turn to incorporating these information distance scores into a series of ERGMs.

6.4.2. Testing the Relationship

P2a suggests a positive relationship between greater levels of information distance and the presence of following ties. P2b assesses whether information distance explains the following decisions of uses better than actor affiliation mixing and matching alone, and I explore this latter proposition through a test of mediation; that is, I examine whether information distance partially mediates the relationship between actor type and network structure. In other words, to explore P2a, I test for a positive direct effect of information distance on network structure, while to study P2b, I examine the indirect effect of actor type on network structure as mediated by information distance.

Table 6.10 contains model results for both measures of information distance dpq where LDA is trained on the sample of tweets from C3-945 (M4), as well as the random sample of graphene tweets (M5). The table also includes M3 for comparison purposes. M4 is the best performing model yet in terms of AIC, while M5 also exhibits a noticeable improvement in AIC over M3. The explanatory power of M4 exceeding that of M5 is not a surprise given the topics in M4 are trained on the content produced in aggregate by C3-945.

Table 6.10: Model results adding information distance variables

Parameter	M3		M4		M5		Remarks
<i>edges</i>	-5.185	***	-4.217	***	-4.572	***	
<i>match.actor_type.intermediary.intermediary</i>	0.020		0.036		-0.032		
<i>mix.actor_type.media.intermediary</i>	-0.264	***	-0.201	***	-0.255	***	
<i>mix.actor_type.nano_firm.intermediary</i>	0.366	***	0.432	***	0.391	***	Suppressor effect
<i>mix.actor_type.scientist.intermediary</i>	-0.101	***	-0.100	**	-0.149	***	
<i>mix.actor_type.intermediary.media</i>	0.167	***	0.273	***	0.151	***	
<i>mix.actor_type.nano_firm.media</i>	0.636	***	0.669	***	0.652	***	Suppressor effect
<i>mix.actor_type.scientist.media</i>	0.421	***	0.386	***	0.383	***	
<i>mix.actor_type.intermediary.nano_firm</i>	-0.606	***	-0.564	***	-0.612	***	
<i>mix.actor_type.media.nano_firm</i>	-0.235	***	-0.270	***	-0.244	***	Suppressor effect
<i>match.actor_type.nano_firm.nano_firm</i>	0.572	***	0.431	***	0.563	***	
<i>mix.actor_type.uknown.nano_firm</i>	-0.520	***	-0.521	***	-0.493	***	
<i>mix.actor_type.other_firm.nano_firm</i>	0.035		-0.027		0.025		
<i>mix.actor_type.scientist.nano_firm</i>	-0.360	***	-0.358	***	-0.395	***	
<i>mix.actor_type.nano_firm.other</i>	0.010		0.021		0.022		
<i>mix.actor_type.intermediary.other_firm</i>	-0.788	***	-0.701	***	-0.811	***	
<i>mix.actor_type.nano_firm.other_firm</i>	0.241	***	0.247	***	0.256	***	Suppressor effect
<i>mix.actor_type.intermediary.scientist</i>	-0.639	***	-0.574	***	-0.684	***	
<i>mix.actor_type.media.scientist</i>	-0.140	***	-0.155	***	-0.151	***	Suppressor effect
<i>mix.actor_type.nano_firm.scientist</i>	0.163	**	0.223	***	0.178	***	Suppressor effect
<i>mix.actor_type.uknown.scientist</i>	-0.944	***	-1.044	***	-0.951	***	Suppressor effect
<i>match.actor_type.scientist.scientist</i>	0.300	***	0.167	***	0.224	***	
<i>mutual</i>	3.138	***	3.061	***	3.125	***	
<i>nodeicov.ln_followers_count</i>	0.376	***	0.385	***	0.374	***	
<i>nodeocov.ln_followers_count</i>	-0.071	***	-0.086	***	-0.081	***	
<i>nodeicov.ln_norm_status_cnt</i>	-0.137	***	-0.137	***	-0.142	***	
<i>nodeocov.ln_norm_status_cnt</i>	0.028	***	0.025	***	0.019	***	
<i>nodeicov.account_age</i>	0.000	***	0.000	***	0.000	***	
<i>nodeocov.account_age</i>	-0.000	***	0.000	***	0.000	***	
<i>edgocov.dpq_C3-945_tweets</i>			-0.134	***			
<i>edgocov.dpq_graphene_tweets</i>					-0.086	***	
AIC	365,268		357,920		363,620		

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. Note: ** $\alpha = .01$, *** $\alpha = .001$

In M4, the log-odds of a following tie existing between any two actors in the reference group, holding all other variables constant, is -4.217 (1.5%), whereas in M5 the log-odds of an edge existing between any two similar actors is -4.572 (1.0%). Contrary to the anticipated positive direct effect of greater levels of information distance on network structure, both M4 and M5 show that the parameters on *edgecov.dpq_C3-945_tweets* and *edgecov.dpq_graphene_tweets* are negative and statistically significant. For example, in M4, the log-odds of an edge existing between two *new* users in the reference group with no followers, no prior tweeting activity, and a one unit increase in information distance is -4.351 (1.3%). In M4, the corresponding log-odds is 4.658 (0.9%). Since the coefficient on the two information distance terms is negative, P2a cannot be substantiated, and subsequently, P2b cannot be properly evaluated for “consistent” mediation. This is due to the way in which consistent mediation occurs where the direct effects of the causal variable on the dependent variable is the same sign as the product of the signs constituting the indirect effect through the mediating variable (MacKinnon, Fairchild, & Fritz, 2007). Usually both the direct effect and indirect effect are positive in mediation analysis, an outcome which is not substantiated by either M4 or M5. See section 6.4.3 for details.

Note that the signs of all coefficients between M3 and M4 and between M3 and M5 remain consistent, suggesting that the addition of the topic modeling based information distance scores might still enhance the baseline interpretation of M3. For example, it may be possible that information distance acts as a type of “negative” suppressor variable that exhibits positive correlations with the actor type mixing variables as well as the network structure dependent variable. In theory, when included in the full

model, the negative suppressor variable takes on a negative parameter estimate because it explains more in terms of the irrelevant information between itself and the other predictor variable (actor mixing) than it does between itself and the outcome variable (network structure) (Maassen & Bakker, 2001). However, after examining the results of a bivariate ERGM on network structure as a function of information distance using either topic modeling measure, a negative and significant relationship persists. In addition, a suppressor variable will often inflate the values of the primary predictor variable (actor mixing), which it does not systematically do in either M4 or M5. Therefore, based on this evidence, information distance does not appear to be a suppressor variable.

What do these results tell us about how information distance qualifies patterns of interaction in the online graphene ecosystem? The simple answer is that individuals, at least in C3-945, group together in structures where decreasing levels of differences in tweet content result in higher probabilities of following ties. So, while there is some evidence that some types of actor mixing generate network ties, information diversity has the opposite effect, regardless of the underlying corpus used to train LDA.

A goodness-of-fit analysis on M4 in terms of in-degree and out-degree distributions of the observed network vis-à-vis the simulated networks shows very similar results to Figures 6.2 and 6.3. Again, the observed degree distribution is within the acceptable range of the metrics obtained through simulation, with the exception of those actors with very small in-degree and out-degree values. Similar to M3, M4 shows limited signs of degeneracy. In M4 individual sample statistics significantly differ (at $p < .05$) from the observed values for the parameters, *match.actor_type.intermediary.intermediary*, *mix.actor_type.nano_firm.media*,

mix.actor_type.scientist.media, *mix.actor_type.intermediary.nano_firm*, *nodeicov.account_age* and *nodecov.account_age*. Geweke statistics are significant for the variables: *mix.actor_type.media.nano_firm*, *mix.actor_type.nano_firm.other*, *mix.actor_type.intermediary.scientist*, *mix.actor_type.unknown.scientist*, and *nodecov.account_age*. Like in M4, autocorrelation across sample statistics does not appear to be a problem across the simulations produced in M3.

6.4.3. Post-Hoc Analysis

The negative coefficient on information distance using the topic modeling measure stands in contrast to the expected direct effect between information distance and network structure: It was proposed that actors choose whom to follow based on the perceived novelty of information. Yet, given the focus of C3-945 as a nanotechnology R&D community, and given the reach of the topics, some of which do not ostensibly focus on graphene per se, the *dpq* edge covariate may capture information distance across too many lexical dimensions – even if those dimensions are broadly related to graphene R&D. To address such a limitation, I turn to the difference in tweeting activity between actors in C3-945 in terms of the number of tweets with the word containing “graphene”. Recall that to be included in C3-945, users must have authored at least one graphene tweet; however there are differences even within this measure that may inform the development of network structure.

In this post-hoc analysis, I test the proposition that differences in graphene tweeting positively predict network structure. Since C3-985 exists as a community in the combined friend network, wherein users follow one another to seek information, it is reasonable to expect that greater distances in terms of the number of graphene tweets

issued by two users predicts tie existence. The rationale here is that the term “graphene” acts as a signal to readers about a specific topic that may be relevant to many different applications and industries, given the GPT nature of the nanomaterial: While some users share a lot of information about graphene, others seek information about it through the following (i.e., friend) connection.

I operationalize differences in graphene tweeting via two steps. First, the natural log of the count values is taken due to the highly skewed nature of the distribution. Second, I specify an interaction term for nodal attribute mixing, similar to nodemix, called “absolute difference” or absdiff (Morris, Handcock, & Hunter, 2008). However, unlike with nodemix, the attribute to model is quantitative, not categorical. “absdiff” adds a statistic to the model equal to the sum of $|\ln_{\text{graphene_tweets}}(i) - \ln_{\text{graphene_tweets}}(j)|$ for each possible directed tie (i,j) in the network.

To reiterate, larger distances of this measure are expected to predict network structure, and the results show this is indeed the case (Table 6.11). Table 6.11 includes M3 as a comparison case; M6 adds the absolute difference term. The direct effect on absdiff is positive and significant. In M6, when holding usage factors constant at zero, the log-odds of a non-reciprocated edge existing in the network between two new actors in the reference group with no followers and no tweet history is -5.357, or 0.5%. With reciprocation, the log odds improve to -2.226 (9.7%). When the linear distance between the natural log number of graphene tweets of two actors increases by one unit, the log-odds of a tie appearing improves to -2.149 (10.4%) with reciprocation.

Having established the positive direct effect of different levels of graphene tweeting on network structure, what about the question of how information distance mediates the relationship between actor mixing and network structure? Baron and Kenny (1986) identify two main criteria that establish a mediating relationship. First, the three variables in question (e.g., actor mixing, information distance, and network structure) should correlate with one another. I provide the following evidence to prove correlation exists.

1. M1 shows that actor mixing predicts network structure (i.e., the existence of network ties).
2. A separate ERGM (not reported in detail) containing an edges parameter and another explanatory variable, the absolute difference of graphene tweeting between two actors (i,j), shows that differences in graphene tweeting positively predicts tie formation.
3. Of the eight positive and statistically significant mixing type relationships reported in M3, all but one are positively and significantly correlated with greater levels of information distance, as measured by differences in graphene tweeting (bivariate output results not reported in detail).

Second, Baron and Kenny argue that once the mediating variable is included in the regression model, the effects of the exogenous variable (in this case, the set of actor mixing types) should approach zero. Many times, however, the effect of the independent variable(s) (in this case, actor mixing) still remains significant but the magnitude of the coefficient declines. Before the results are discussed, I briefly turn to a technical

discussion of testing mediation through indirect effects; this provides a more nuanced interpretation of the findings ascertained from Baron and Kenney's approach.

Advanced tests that evaluate the statistical significance of an indirect effect given a mediating variable focus on either (a) the difference between the total effect of the primary explanatory variable on the outcome and the corresponding direct effect, or (b) evaluating the product of coefficients constituting the indirect effect (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) (Figure 6.4). The total effect c can be isolated by regressing network structure on the presence of a specific actor type mixing relationship, whereas the direct effect c' can be identified by adding the mediating variable to the specification. The indirect effect lies on the path from the actor mixing categorical variable to the outcome variable of network structure *via* the mediator, information distance (a and b).

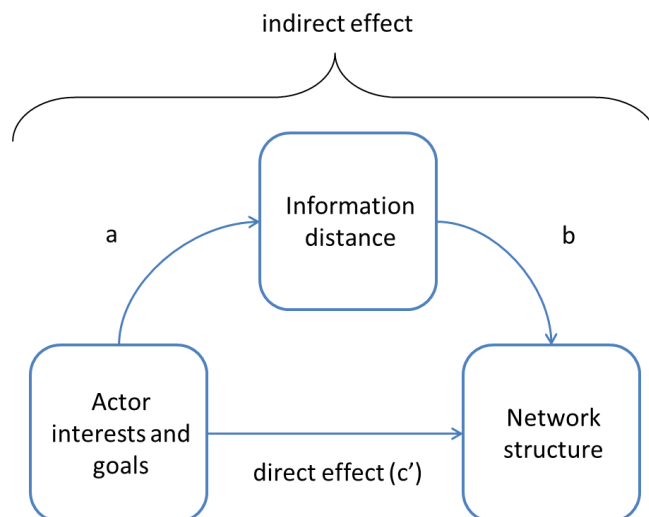


Figure 6.4: Path diagram showing direct and indirect effects

This first approach to testing mediation effects – i.e., calculating the difference between the total and direct effects – is computationally incorrect in cases where either the outcome or mediating variable is dichotomous (MacKinnon et al., 2007). This is because the residuals in logit and probit models (to which ERGMs are related) are fixed, and therefore any given parameter estimates derived from this family of models depend on other parameter estimates. However, Iacobucci (2012) conveniently introduces a well-articulated approach to computing whether mediation occurs regardless of whether the variables in question are continuous or categorical. This process requires standardizing the sample regression coefficients comprising the indirect effect (z_a and z_b), taking their product (z_{ab}), and formulating a “collected” standard error. A “z-mediation” score can then be calculated as follows and tested against a standard normal the usual significance levels; e.g., mediation occurs if $|z_{mediation}| > 2.576$ at the $\alpha=.01$ significance level for a two-sided test.

$$z_{mediation} = \frac{z_a z_b}{\sqrt{z_a^2 + z_b^2 + 1}}$$

Table 6.11: Adding an information distance term for differences in graphene tweeting

Parameter	M3	M6	Remarks	z.med
edges	-5.185 ***	-5.357 ***		
match.actor_type.intermediary.intermediary	0.020	0.025		
mix.actor_type.media.intermediary	-0.264 ***	-0.273 ***		
mix.actor_type.nano_firm.intermediary	0.366 ***	0.343 ***	+ magnitude ↓	15.355 ***
mix.actor_type.scientist.intermediary	-0.101 ***	-0.077 **		
mix.actor_type.intermediary.media	0.167 ***	0.164 ***	+ magnitude ↓	14.880 ***
mix.actor_type.nano_firm.media	0.636 ***	0.613 ***	+ magnitude ↓	18.803 ***
mix.actor_type.scientist.media	0.421 ***	0.430 ***	+ magnitude ↑	2.805 ***
mix.actor_type.intermediary.nano_firm	-0.606 ***	-0.649 ***		
mix.actor_type.media.nano_firm	-0.235 ***	-0.282 ***		
match.actor_type.nano_firm.nano_firm	0.572 ***	0.513 ***	+ magnitude ↓	16.994 ***
mix.actor_type.unknown.nano_firm	-0.520 ***	-0.558 ***		
mix.actor_type.other_firm.nano_firm	0.035	-0.009		
mix.actor_type.scientist.nano_firm	-0.360 ***	-0.393 ***		
mix.actor_type.nano_firm.other	0.010	-0.003		
mix.actor_type.intermediary.other_firm	-0.788 ***	-0.777 ***		
mix.actor_type.nano_firm.other_firm	0.241 ***	0.225 **	+ magnitude ↓	7.898 ***
mix.actor_type.intermediary.scientist	-0.639 ***	-0.620 ***		
mix.actor_type.media.scientist	-0.140 ***	-0.135 ***		
mix.actor_type.nano_firm.scientist	0.163 **	0.156 **	+ magnitude ↓	9.703 ***
mix.actor_type.unknown.scientist	-0.944 ***	-0.898 ***		
match.actor_type.scientist.scientist	0.300 ***	0.335 ***	+ magnitude ↑	-21.327 ***
mutual	3.138 ***	3.131 ***		
nodeicov.ln_followers_count	0.376 ***	0.380 ***		
nodeocov.ln_followers_count	-0.071 ***	-0.068 ***		
nodeicov.ln_norm_status_cnt	-0.137 ***	-0.143 ***		
nodeocov.ln_norm_status_cnt	0.028 ***	0.021 ***		
nodeicov.account_age	0.000 ***	0.000 ***		
nodeocov.account_age	-0.000 ***	-0.000 ***		
absdiff.ln_graphene_tweets		0.077 ***		
AIC	365,268			

Source: Twitter; network consists of 945 actors and 57,935 edges; data collected in early 2014. *Note:* ** $\alpha = .01$, *** $\alpha = .001$

Based on this discussion, we are most interested in what happens to the eight positive and statistically significant actor mixing combinations in M3 vis-à-vis M6 (Table

6.11). Referring to Figure 6.4, we know b is positive, that is greater levels of disparities in graphene tweeting positively predict network structure. In addition, of the 21 categorical mixing variables, eight are of particular interest because their sample regression coefficients positively predict network structure. These mixing variables live on path c' . Comparing M3 to M6, six of the eight positive coefficients decrease in magnitude, though all are still significant. So, for example, nanotechnology firms are less likely to follow intermediaries, media entities, scientists, and other nanotechnology firms once information distance in terms of graphene tweeting is accounted for. Because these six positive coefficients decrease in magnitude (in path c' vs. c), this indicates that these mixing relationships on average are associated with greater levels of disparities in tweets containing graphene (i.e., through the indirect effect of actor mixing on network structure beginning with path a in Figure 6.4).

That is, nanotechnology firms find value not only in following a diverse set of users but also in the amount of graphene-related content that those relationships provide. Note that users tweeting about graphene represent a variety of actor categories, and thus these results do *not* suggest that disparities in graphene tweeting predict network structure simply because nanotechnology firms are more likely than other users to tweet about graphene. (Recall for instance that to be included in C3-945, all actors must have tweeted about graphene at least once.)

Surprisingly, the two mixing types that experience increases in coefficient sizes are following ties from scientists to media entities and from scientists to other scientists. This finding suggests that scientists do not follow media entities and other scientists to access additional quantities of graphene content. Indeed, as indicated in the sign of

$z_{\text{mediation}}$, the bivariate correlation between intra-scientist following relationships and information distance (path a in Figure 6.4) is negative.

The $z_{\text{mediation}}$ scores are very unlikely to occur if the null hypothesis were true; i.e., if mediation did not occur. It is worth noting, however, that $z_{\text{mediation}}$ scores are the results of a basic path analysis excluding important endogenous network and exogenous usage controls. While it makes the calculation $z_{\text{mediation}}$ easier to compute, this simplification does not consider omitted variables that could confound the results.

6.5. Summary

This chapter examines the first two propositions of this research study, namely that actors choose whom to follow by mixing and matching across affiliation types, and that these following decisions are also informed by subjective perceptions of information novelty. The findings provide preliminary support for P1b by showing that graphene firms are more likely to follow one another than all other actors. The results also support P1b by showing larger homophily trends within the nanotechnology firm and scientist actor categories, once a series of other exogenous and endogenous variables are included as controls. In terms of mixing across actor categories (P1a), the results show that media entities and intermediaries are likely to attract following relationships from all actor categories. In later model specifications, we see that nanotechnology firms are also more likely to follow across the other firm and scientist actor categories, suggesting that Twitter provides a suitable platform for high-technology firms to build awareness of and potentially engage in ecosystem activities. The role of social media in improving awareness will be explored further in the qualitative results (Chapter 7).

P2a, which supposes ecosystem actors choose whom to follow based on perceived access to information novelty, is not substantiated using the measures derived from topic modeling. Because of this result, P2b, which examines the mediating influence of information distance on the relationship between mixing and matching of actor affiliation and following decisions, was not pursued further in these model specifications. However, when incorporating the difference in graphene tweeting as a measure of information distance, evidence for the positive direct effect and mediation relationship was found.

In summary, these findings support the notion that the online innovation ecosystem is an amalgam of different types of actors who follow one another – and thereby gain access to content – albeit in some traditional ways. For example, users are likely to follow media entities for information. At the same time, however, the Twitter social graph shows that diversity is also important, particularly for nanotechnology firms, which are likely to maintain links to scientists and other firms. People without professional affiliations – those “unknown” users – exist in the network but do not attract a widespread following. In terms of information content, network linkages appear to develop around topical areas of shared interest, even within the diverse actor composition of C3-945. However, perhaps because graphene is a technology with significant interest across industries and sectors, it appears that low volume tweeters in the graphene space are likely to follow high-volume tweeters. (Alternatively, it could be the case that high volume tweeters also follow low volume graphene tweeters.)

The findings in this chapter are subject to three main limitations. First, while C3-945 exhibits ecosystem qualities and thus lends itself to closer examination in terms of

the study's first two propositions, this community is ultimately a network with boundaries defined by the constraints of the selected ERGM implementation. As such, even though the total number of observations is large, the results may not generalize to other communities such as C3, the combined follower network, or the entire population of Twitter users involved in ST&I. Assuming better tool availability, additional regression modeling using the same set of covariates should be employed on larger networks within the graphene space and even in non-related emerging technology domains (e.g., synthetic biology).

Second, with different networks comes the opportunity to test alternative specifications of the ERGMs presented in this chapter. The empirical literature on ERGMs advises the inclusion of complex network self-organizing statistics (e.g., geometrically weighted terms) as important controls (Hunter, Handcock, Butts, Goodreau, & Morris, 2008). With these controls, however, such approaches always resulted in non-converging models. Yet including additional network controls could produce important findings should models converge on different datasets.

Finally, the results are also qualified by the enabling and constraining features of Twitter as a communication platform. For example, it is relatively easy to connect with strangers on Twitter, and yet the character limits on tweets, as well as their public nature, may restrict the type of communication that occurs on the platform in the public timeline. As a result, these findings, particularly with respect to information distance, may not be applicable to other online platforms or offline modes of interacting where community formation around similar topics is less important than pursuing weak ties that can transmit diverse information.

CHAPTER 7: QUALITATIVE FINDINGS

The prior two chapters addressed the first two propositions based on quantitative evidence. The first proposition examined whether actors choose following relationships based on social selection processes of similarity or dissimilarity in professional affiliation. The results show that innovation networks on social media are not random and that there is a certain type of interaction – among traditional innovation actors, as well as the media – that predicts friend networks in graphene firm-centric networks.

The second proposition builds on theory that stresses information novelty, as subjectively viewed by network members, as a scarce resource that offers particular value and relevance in the social media domain, where information traverses freely across traditional boundaries. This proposition specifically assesses whether dissimilarity in information across ties predicts network connections. Contrary to expectation, the results show a negative effect; that is, as information distance increases, the probability of a network tie existing goes down. However, in a post-hoc analysis that evaluates the difference in the number of ‘graphene’ hashtags tweeted by one party vis-à-vis another potential user, as the distance between graphene tweeting increases, so too does the probability of a linkage.

This chapter considers innovation outcomes that social media usage may help advance. Qualitative evidence is supplied through nine interview transcripts showing noticeable differences in each respondent’s orientation towards Twitter and other social media usage (Section 7.1; refer to Table 7.1 for an overview of the sample). The outcomes of interest include increased awareness (Section 7.2), better problem solving ability (Section 7.3), and enhanced linkages via private (and often non-Twitter based)

modes of communication (Section 7.4). Other potential outcomes include sales and marketing gains for firms, impacts to the job market for scientific talent, and overall changes to the way science is adjudicated and disseminated (Section 7.5). Each section presents relevant usage evidence that helps set the context as to why certain outcomes are more or less perceptible to the respondents. An analysis (Section 7.6) is then undertaken to interpret the qualitative findings and then briefly integrate the findings from both the quantitative and qualitative results chapters.

The results from this chapter suggest mixed evidence with regard to social media's effects on innovation additionalities: For most interviewees, Twitter in particular helps improve awareness of ecosystem topics but rarely contributes to enhanced problem solving outcomes. Similarly, some users transition public Twitter interaction to private modes of interaction (e.g., via other online platforms including LinkedIn), while others do not. Finally, Twitter and social media may facilitate sales and marketing efforts for nanotechnology and graphene firms, and it may also offer some benefits to job seekers in terms of job market related information and contacts. Many respondents were enthusiastic about social media's impacts on the scientific enterprise in terms of public discourse and accessing information in related fields, but they also conveyed reservations about how well tweeting could capture truly complex ideas. A word table (Table 7.2) summarizes the results on these various outcome dimensions.

7.1. Sample Overview and Usage Patterns

The sample consists of nine interviewees with three intermediaries, three firms, two scientists, and one media entity. The interviews were conducted between March

2014 and January 2015 and lasted between 20 and 90 minutes, with most taking between 30 to 45 minutes. Each of the interviewee's usage characteristics is outlined in Table 7.1. To ensure anonymity, respondent identifying information is obscured from interval values to ordinal categories where appropriate. For example, instead of disclosing number of followers or friends, which can be easily viewed online, labels such as "few hundred" "nearly two thousand" are given instead. See the table's caption for a description of the classification scheme.

Most of the users interviewed in this study have been on Twitter for several years with older accounts experiencing more followers and friends, on average, than newer accounts. However, in terms of the ratio of followers to friends, it is not possible to infer anything about certain types of accounts having more followers than friends, or vice-versa. For example, a scientist and two intermediaries show more followers than friends, but all other accounts maintain more friends than followers.

Table 7.1: Interviewee profile and usage characteristics

Id	Account Type	AccountAge (years)	Followers	Friends	Ratio (followers to friends)	Normalized Status Count	Graphene Tweets	Interview Date and Length (min)
F1	Firm	< 4	Couple hundred	Many hundreds	0.50	Moderate	4	19/3/2014, 55
F2	Firm	< 5	Many hundreds	Near two thousand	0.48	Moderate	27	27/3/2014, 45
F3	Firm	< 6	Over a thousand	Near two thousand	0.59	Prolific	16	31/3/2014, 90
I1	Intermediary	< 6	Few hundred	Many hundreds	0.76	Moderate	0	1/4/2014, 65
I2	Intermediary	< 3	Many hundreds	Many hundreds	1.01	Infrequent	0	3/9/2014, 30
I3	Intermediary	< 7	Millions	Couple hundred	20002.73	Prolific	0	19/9/2014, 30
M1	Media	< 2	Tens	Tens	0.67	Moderate	261	14/3/2014, 20
S1	Scientist	< 6	Over a thousand	Over a thousand	1.27	Prolific	2	21/11/2014, 30
S2	Scientist	< 7	Tens	Tens	0.76	Moderate	8	23/1/2015, 35

Sources: Interview transcripts and Twitter, n=9 users; Twitter data collected in early 2014, and interview data collected in 2014 and early 2015. *Notes:* *Account age* is measured in years with “< N years” indicating that the account has been active between N-1 and N years. *Follower* and *friend* ordinal categories include tens, couple hundreds, many hundreds, over a thousand, nearly two thousand, and millions. Recall that *normalized status count* refers to the number of tweets issued by the account since inception normalized by number of days since the account has been online. Infrequent denotes fewer than 0.10 tweets per day; moderate represents between 0.10 and ~1 tweets per day; prolific indicates greater than ~1 tweet per day.

Table 7.2: Word table summarizing qualitative findings

Outcome	Comment 1	Comment 2	Comment 3
Awareness	I would say I try to learn trends from Twitter... Maybe one example is graphene. (F3)	Because I work at the interface of two research areas... I find out a lot more about what's going on in some of these areas. (S1)	It's less about getting information from the friend network, because usually the stories that people tweet on graphene, I get that morning on Google alerts. (F1)
Problem solving capacity	I don't think social media has helped us achieve our objectives. There are very few discussions that take place. (I2)	No, [Twitter does not help solve problems]. (S2)	You of course want to create some spreadsheets about how effective social media and websites are.... We learned that it depends a lot on the product. (F1)
Transitioning from public to private interaction	We made that connection on social media and followed-up on phone calls and over email. (I1)	We're very cautious about working with companies [directly] because of Federal laws. If you look at who we're following, we don't even follow our contractors. (I3)	I am looking for people who tweet and then show up at other places. (F2)
Science jobs	I don't think people are finding jobs, though they may be finding out about jobs. (I1)	If you found an opportunity on Twitter, like as a career change, but that's probably pretty dramatic. I never looked to Twitter for an opportunity. (F3)	
Sales and marketing	In marketing, you need to make a lot of contacts in order for one or two to work out... It might be Twitter, LinkedIn, email, phone calls on Skype, then a period of silence, and then they get back to you about a project. This works more organically. (F2)	The biggest benefit of Twitter right now is that we can share information in the field as thought leaders (e.g., conferences). It's not just about getting more sales. (F1)	It's a little harder to connect the dots using social media as a business tool. I never remember making a sale on Twitter. (F3)
Impacts on science	It's [Twitter is] a good thing as long as the debate is civil. (S1)	If I follow [someone tweeting about] 2D materials – I might learn about transition metal dichalcogenides – e.g., WS ₂ or WSe ₂ . These are the next set of 2D materials that are up and coming after graphene. (S2)	Speaking simply is important for describing your idea in a short amount of time. (I1)

Sources: Interview transcripts, n=9 users, collected in 2014 and early 2015. *Note:* Pseudonym legend available in Table 7.2.

Most respondents clearly identified their accounts as professional and not for personal use; even when they consume non-ST&I related tweets, the respondents are careful to author and retweet content that their followers would find interesting. In general, individual users not readily representing the interests of their organizations tend to take a more laissez-faire attitude toward their following and tweeting choices than individuals intending to further some set of strategic goals. That is, while active on the platform, some individuals do not alter their behavior to attract resources or new followers, at least not in the way social capital theory might predict. Also important is the observation that these non-strategic users are less particular than strategic users about how they view and interact on different platforms (e.g. LinkedIn vs. Twitter). As shown elsewhere in this chapter, this difference between usage styles is due in part because less strategic users *expect* few tangible outcomes resulting from their participation. For example, one user has not changed much about his tweeting behavior since he first started using Twitter:

I've [actively] used Twitter for 2.5-3 years. I don't know that it [my usage] has changed. If you look at what I tweet, it's been mostly technology and leadership. My daughter is on Twitter, and I don't use it as a communication tool with her – perhaps once a year I might mention her handle. Other than that, I keep it tech focused.

[Do you focus content strategically to be of value to these different communities?] *No, I do what is interesting to me. That's life, the way Twitter works... [People can just unfollow you, right?] Yes, and they have. People do unfollow me, but I don't track them down to figure out who they are... I don't follow all the rules. I don't retweet when I'm supposed to.*
(F3)

Another user echoed this informal approach to growing his network based on tweeting interesting content:

I haven't really engaged with people to build my network. There are some Twitter lists – I was on the Huffington Post's top list of Twitter physics users. If you say something that people are interested in, then you will automatically build a network. (S1)

The journey to develop a presence on social media took a more serpentine route for one small business owner looking to capitalize on building his online exposure. By developing and providing a service to the nanotechnology community, he is able to funnel contacts from Twitter and LinkedIn to his newsletter.

I went on Twitter in 2009. It wasn't particularly active at that point. I thought of tweeting something about myself, but it was difficult to find something meaningful to send out to the world. Even if you're a large company, it's hard to tweet more than once or twice a day. If you share garbage, people don't like it and they'll be done with you...

Then I had a chat with some people that live near me. We all read a lot to what is relevant to what we do, and what our clients do. So I decided to tweet what I've read. As a result, people started following me and retweeting me – from all walks of life, people who are interested in nano. These are active followers. These articles I get through Google Alerts.

I spoke with a few people in [my locality], and they agreed that using Twitter is beneficial for the community. You start building your own community, rather than tweeting about a taxi ride or lunch. From that I decided to consolidate my contacts through Twitter and LinkedIn through my newsletter. I have [many] people who receive this newsletter, [people spanning] from government officials, companies, etc. The result is non-intrusive marketing – there is valuable information but I'm not selling them anything. I'm just reminding them of myself. (F2)

Two of the three intermediaries (I1 and I3) are large organizations consisting of various divisions; a social media policy implemented by a dedicated staff governs how official accounts disseminate information on the platform. To one of these intermediaries, each social media platform offers a distinct purpose for community engagement:

On Twitter, we have so many accounts, they vary in the size of the following: 15,000 to a couple hundred. On LinkedIn, we have just a group

for (only active) members, 20,000. We have a big following on Facebook too: 50,000 likes. It [Facebook] is less used for professional purposes; it's more helpful for outreach and the "lighter side of science". There is some interest in popular research on Twitter (e.g., dark chocolate is helpful for your health or a new polymer material for stitches). The community on Twitter is much more about people actively involved with the organization or heavily involved in the [redacted] sciences. (I1)

A third smaller intermediary (I3) maintains no such formal policy, but participants do discuss during meetings what to post and how to coordinate across a network of inter-institutional accounts.

One media entity, a moderate tweeter with the greatest number of graphene tweets across the sample, surprisingly does not log-in to Twitter on a regular basis.

Twitter is not a major focus [for me]. I use several social network platforms, usually Facebook and LinkedIn. It's all automatic on Twitter via RSS [Really Simple Syndication]. I only have a little over [hundreds of] followers, it doesn't generate a lot of traffic. There are a lot of external links [that drive traffic to my site] – it's mainly external links, Google or direct. The social networks are not very important for me. (M1)

In sum, Table 7.1 shows that the sample is diverse not only in account types but also with respect to tweeting rates and the extent to which those tweets contain graphene related content. Several users, while found in the graphene firm networks, do not tweet much about graphene at all. So, users need not tweet about graphene to be followed by graphene firms, and this loosely supports the notion that graphene firm networks on social media reveal interaction outside of a specific scientific domain to other sectors of the economy. This observation can also be discerned through the combined friend network visualization (Figure 5.3) in Chapter 5.

7.2. Awareness

Most participants in the sample acknowledged that social media, and Twitter in particular, facilitate increased awareness of tangential knowledge domains outside of the user's core area of knowledge. Twitter is not a platform for "drilling deep" into niche content domains but rather adds context to discourse. Additionally, in most cases, Twitter does not offer proprietary information; it is essentially a platform for distributing public knowledge.⁴⁴ One user (F3) uses Twitter to stay abreast of industry developments when it is not possible to travel:

The benefits are that you stay on top of trends in the industry. I probably did that before through talking with people. But because of the economy, we can't travel around so much anymore – there's no money for that – so LinkedIn and Twitter help with that. For example, I learned that Cannon acquired MMI, a nanoimprint company.

I would say I try to learn trends from Twitter, but I can't say I've ever discerned one. Maybe one example... is graphene. In graphene, much of what I know is more through Twitter than through industry contacts. (F3)

A theoretical physicist specializing in 2-D materials (including graphene) described his experience finding new articles to cite on Twitter:

I learned about some of my friend's publications on Twitter... The publications are not necessarily related to graphene. I was reading a paper [that was tweeted] from a friend, and then in the introduction I found and read another paper that was related to my area but that I didn't know about. I cited that. It's like a bar hopping thing. You find something interesting that you weren't ever aware of before (S2).

Another scientist related that his use of Twitter helped him gain knowledge of related scientific domains. Though this user set up his Twitter account to initially engage with students, he soon found his students were not on the platform and instead turned to

⁴⁴ Social media and Twitter data can offer propriety insights through analytics and analysis, as seen in this current research.

accessing different “spheres of interest”. Many of these interests are embedded in distinct communities.

Because I work at the interface of two research areas, in biophysics, I find out a lot more about what’s going on in some of these areas. I can’t read in-depth, niche publications. I see it [Twitter] as a news feed... I’m exposed to things I wouldn’t see otherwise. For example, I’ve become aware of certain conferences that I’ve later attended. (S1)

At one intermediary whose job it is to fund and increase the visibility of nanotechnology EHS research, a user finds relevant research on Twitter – but it’s not as if this were somehow unavailable through other channels. Rather, it’s about the speed with which he can access new journal literature from his mobile device.

I become aware of papers that have been published by other groups. It’s about immediate access, but these papers would come across my desk at some point anyway. I don’t check my account everyday but my cell phone gives me notifications. (I2)

Timely access to scientific content is certainly not a new concept. Consumers of scholarly material, including scientists, have increasingly relied on online sources and indices to satisfy their information needs (Hemminger, Lu, Vaughan, & Adams, 2007), and the results indicate that social media is an additional (albeit more contemporary) channel continuing this longer-term trend.

One professional society supports many sub-organizational accounts. Followers receive tweets about a specific area of interest and this may build relationships which on aggregate result in community development. Indeed, this interviewee implicitly views followers as community members:

There are specialties in [a specific field], doing a lot of outreach to 32 areas, e.g., in health and safety there is a lot of news that is generated to their members. The audience can see the value of belonging to that group. I can’t say their attendance is increasing at meetings, but I can say that it

is more about reaching out and building a direct relationship. It helps when you have that type of community existing. (I1)

Awareness is also related to engagement. Since information on Twitter is nearly costless to access and share (but not necessarily to assimilate), users benefit from learning about new developments and easily disseminating that information to their networks. In addition, by observing the tweets of others, it becomes possible to measure the degree of conflict or agreement within a community regarding contentious issues in nanotechnology R&D.

Yes, it has improved my awareness. I find articles on Twitter, and can retweet it. Sometimes I offer my view on the article.

Twitter can be used as a measure of the positive or negative reaction to an action. I always wanted to get into it. There is a lot of discussion about the safety of nano, and so far, surveying public opinion on this is extremely difficult. The survey design has always affected the results. If you're a specialist, you can design the question [to elicit a certain response]. For Twitter, you have a statistically independent community of lots of people thinking about various things, e.g., articles and news. The sentiment that people use in tweets allows you to monitor [their reactions] in real-time. (F2)

An intermediary with millions of followers uses Twitter almost exclusively to generate intimacy with its user base (I3). Before social media gained traction there, this respondent noted the organization produced a quality website and award winning multimedia content, but it was having a hard time building broad awareness of its mission and buy-in from the public. For this account, Twitter is a mechanism for broadcast communication but with a unique element that encourages active listening: If another user says something interesting, they respond. Moreover, in prior eras at this organization, scientists and other technical personnel were reluctant to engage with the public. With Twitter, these employees were also reticent initially, but as news and

visibility of their work increased with tweeting and online interaction, so too did their interest in participating regularly. Scientists and technicians were encouraged to set up official accounts so that the primary organization account could retweet their messages. In this way, Twitter becomes a mechanism to extract interesting and valuable information from down within the organization to a “master communications” account. This continues to serve not only the followers of the main account with improved awareness of the organization’s activities but also offers research groups and individual employees greater visibility and outlets for their work.

Not all users experience improved awareness of the broader R&D ecosystem, however. As noted above, one media user (M1) does not “use” Twitter directly and rarely logs-in to read tweets; the account automatically tweets via an RSS feed that pulls content from a website. Another graphene SME, while active on Twitter for building a community around graphene and for enhancing the organization’s reputation as a thought leader in the domain, builds awareness of new developments through Google Alerts.⁴⁵ In addition, this firm does not use Twitter to advance its R&D capacity: “People who do that work [R&D] stay on the cutting edge of things. They look at scientific papers [as their primary source for information]” (F1).

7.3. Problem solving

The literature on innovation networks stresses the importance of combining knowledge across disparate intellectual domains, geographic locations, and industries (Desrochers, 2001; Nahapiet & Ghoshal, 1998; Neal et al., 2008; Phelps et al., 2012). In

⁴⁵ Google Alerts accepts a user’s query and scours the web for new matching results, which are then delivered via email.

the literature review I consider the possibility (Chapter 2) that social media may encourage the development of expansive knowledge networks and facilitate discourse that results in beneficial problem solving outcomes.

The interview data provide insufficient evidence for such an assertion. In short, tweeting is more amenable to high-level summary statements that quickly diffuse across networks, not the type of exposition and conversation that expound upon an innovation's technical, economic, and social dimensions. While dialogue can be found on Twitter, one interviewee (I2) commented that for his public organization this is not the case.

I don't think social media has helped us achieve our objectives. There are very few discussions that take place. This usually happens through email and meetings. Twitter helps us make our work more public. (I2)

One respondent noted that Twitter doesn't help solve problems, but it can encourage reflection on particular problems that need solving:

No, [Twitter does not help solve problems]. But I do get excited. For example, on New Year's Day, I read about a breakthrough, and it germinated an idea in my mind. Someone reported an interesting result, but I said hey, I could solve this problem and give a model. But then I took a step back because there are many other things that need prioritization. But that's the beauty of science. There are many things that I can do, but [only] when I'm an independent researcher and have graduate students to tell [i.e., delegate to]! (S2)

Yet tweeting does provide some value to participants in terms of applied problem solving outcomes, just not in ways that are directly relevant to material science innovation. For example, one organization's reporters are able to scour social media to find better sources for their reporting. Two additional interviewees had been involved with initiatives to directly leverage social media data for analytics projects. The first described his firm's social media campaign in terms of gauging its effectiveness and modifying usage accordingly based on reporting insights.

You of course want to create some spreadsheets about how effective social media and websites are.... There were times where we were not really sure what kind of tweets were really engaging the most (e.g., with products), so I took a two month period of time, and captured what the tweet was and the type of tweet (e.g., a news story, humorous or not). Then we tracked the amount of engagement in terms of number of tweets, number of [new] followers, and number of favorites.

[So what did you learn?] *We learned that it depends a lot on the product.*
(F1)

“Depending a lot on the product” in this context means that certain product types were more likely to elicit interest from the community than other types of products, for example given the current state of the overall technology’s development and its relation to other popular news stories (see Box 7.1 below).

The second interviewee (S1) recalled an instance where a large US public agency convened an ideation forum to brainstorm and prototype solutions to pressing problems aligned with the agency’s mission. Although this “hack jam” was largely data-oriented, the outputs of the forum were presented at a leading US conference that “ended up in funded research [in the UK]” (S1).

7.4. Transitioning from Public to Private Interaction

Five of nine interviewees related positive experiences with transitioning public Twitter interactions into *sustained* private modes of communication; for these users, social media offers a way to expand one’s directory of contacts, and some of these contacts may emerge as valuable relationships requiring other modes of communication to develop and sustain. Four interviewees mentioned some type of technology or Internet-mediated transition from their Twitter usage to LinkedIn, email, Skype, or phone

calls. The responses below identify the utility of Twitter vs. other social media platforms where possible and as identified by the interviewee.

One user (F3) received a tweet from a leading science agency, and by following a link and filling out a web form, was invited to attend an exclusive service commemorating the legacy of a retired program. Still, this event did not result in an ongoing relationship as, to this user, social media does not facilitate “deep” relationship-building. In contrast, for one professional society, working across sectors to promote its agenda on social media is commonplace, and communication naturally shifts to other modes of higher fidelity interaction, including in-person meetings:

There are members and non-members who I've developed relationships with, e.g. a woman with our small [redacted] business division. She has a small analytical [redacted] business. She does a lot of work in business and marketing, so she's also somebody who will advertise meetings. She has coordinated tweets at our informal meet-up events. She tweets under both her personal and [redacted professional] handles. (I1)

For profit-seeking firms, the calculus of pursuing linkages on Twitter likely depends on the business model of the organization and the degree to which social media affects normal business activities. For example, the founder of a service oriented firm linked the utility of online communities with professional circles that appear in other contexts.

Social media is another way of interacting. It doesn't create a special community, and it doesn't create a subcommunity. If it does create a special subcommunity, then it's because of people who are not real. I am looking for people who tweet and then show up at other places... For example, on Twitter, I got into contact with NNI [the US National Nanotechnology Initiative]. They have difficult jobs in terms of strategy and administration, but they do read non-refereed articles. People do find that web-sourcing of technology news is helpful. You may have to read 1-2-3 articles before you know [what is accurate].

The NNI was more of a casual introduction. “This guy was reading

interesting things, and so am I.” These aren’t permanent contacts; they’re not my clients. (F2)

Some firms are able to mobilize social capital developed on Twitter in tangible ways. This same interviewee (F2), for example, noted his ability to build relationships with clients from Twitter and other social media, as discussed in Section 7.5.1. A representative from a second graphene firm also described the benefits that Twitter linkages confer with respect to sales and marketing efforts, as covered in greater detail below (Box 7.1).

While small nanotechnology firms may reach out to other types of organizations and users for information and to network, the reverse is not always true. That is, the asymmetrical nature of social media (directed) relationships, as discerned through the quantitative results, suggests a lack of online engagement between certain user types and others. For instance, one account owner from a large intermediary conveyed his organization’s conservative policy towards following companies in the larger innovation ecosystem.

We’re very cautious about working with companies [directly] because of Federal laws. If you look at who we’re following, we don’t even follow our contractors. (I3)

A second interviewee echoed limited interaction with nanotechnology firms online:

In terms of companies, [I have] not [pursued] specific relationships. I do follow companies, e.g., for lasers. It gives me information about products and marketing [in microscopy], which link into research that uses their products. Sometimes they just share pretty pictures. (S1)

Three interviewees experienced very few additional offline linkages as a result of their Twitter usage. Recall that one user, a media entity, rarely logs-in to Twitter and benefits from RSS website feeds that automatically author tweets. This user relies on

LinkedIn, not Twitter, to initiate industry contacts, and Google and direct URL access to drive traffic to his website, which enables his business model in terms of original content, advertising, and professional services.

I have developed relationships, but not through Twitter. Many of my connections have developed through LinkedIn, my personal account. Sometimes I don't remember where the relationships originate, but just a couple weeks ago, I was talking to some graphene industry people... (M1)

An intermediary that infrequently tweets and follows the researchers it funds, also reported limited interaction with users outside of its funding umbrella.

In terms of followers, the interviewee noted:

I think most of the people are related to the research activity we fund. Now it is true that there are some faces [followers] that I've never met. But these are people mostly in the involved research cluster or broader research domain. (I2)

Finally, a scientist noted that the nature of his field of study precluded certain types of interaction on social media from becoming productive offline relationships. Whereas applied physicists may collaborate with hundreds or even thousands of their colleagues, the same is not true for theoretical physicists: “Unfortunately, I am a theoretician, so industry or lab collaboration is limited. I don't have much access to people. I meet people at conferences [but not through Twitter] (S2).”

7.5. Other Innovation Outcomes

This section explores three innovation outcomes not directly tied to the exploratory propositions and literature review but still salient given the research context and the types of responses elicited during their interview sessions. Firstly, sales and marketing impacts are one significant set of outcomes for nanotechnology and graphene SMEs. Secondly, shifts in the labor market is a concern for companies, intermediaries,

and individuals. Finally, all interviewees discussed their view of how social media is influencing the broader scientific enterprise: To these respondents, social media and Twitter in particular continues to act as a progressive change agent in scholarship across many scientific disciplines.

7.5.1. Sales and Marketing

Two of the three firms interviewed as part of this work use social media to indirectly increase sales. For one respondent, cross-platform communications were first enabled through social media introductions, which then grew to email, direct private messaging, Skype, and in-person interviews as part of an elongated sales cycle for nanotechnology services.

In marketing, you need to make a lot of contacts in order for one or two to work out. For me it takes 4-5 interactions, to get them to sign-up. It might be Twitter, LinkedIn, email, phone calls on Skype, then a period of silence, and then they get back to you about a project. This works more organically. With small businesses like ours, we can't barge in with a contract – they don't like that. We do have a niche to work with small and medium sized companies... (F2)

This respondent would not provide additional details, but he indicated that both informal and formal discussions resulted from his interaction on Twitter and lead to new business opportunities.

A second interviewee representing a graphene SME recounted his company's early entry into social media, the diverse types of users he engages with, and some of the strategies he employs to increase interest in the firm's product lines (see Box 7.1). For this user, experimenting with content and engaging on social media leads to a combination of greater visibility in the space and increased sales through non-traditional customers, e.g. hobbyists, parents, and scientists as opposed to downstream firms. Yet,

marketing and sales are not the sole objective of social media participation because in professional circles, the firm is (now) already well known. Instead, social media expands the reach of the company beyond a system of competitors and distributors into the broader innovation ecosystem.

Box 7.1: A graphene SME's (F1) social media usage strategy and outcomes

[Tell me about your strategy with respect to Twitter and social media?] *When I first started at the company, we didn't have a social media presence. I started the Twitter account. My boss was really interested in it to sell products and generate buzz. It was difficult at first because... we're selling a single layer of atomic carbon. We first learned about being personal about it [our tweeting], not just necessarily posting scientific content... We instead focus on applications and uses. For example, we got a lot of traction when we posted about graphene for [a particular application area]; people were very interested in that. We want to get people more interested in our content, get more followers. We don't want to overrun what we're doing with sales pitches. Say that a paper came out [example given], that was a big thing. Then we'll take a look at a product that was used in the paper and relate it back to a similar product on our website.*

[Have people followed up on this?] *Yes, we've had parents contact us. For example, parents with children in science fairs will ask us for help. Most inquiries we get are more business related [but those are not necessarily done through Twitter]. Most of the people that get interested in our company at this point are just hobbyists. The biggest benefit of Twitter right now is that we can share information in the field as thought leaders (e.g., conferences). It's not just about getting more sales.*

[How do you decide who to follow?] *[Since] the end goal is conversion to sales, [we] follow PhD students or Chemistry students and follow people who will potentially use our materials. If they're not a private account, then I'll follow them. Usually if people are talking about graphene, it's about a Gizmodo article or stock or finance related, but it's not often that you'll see someone studying the material, talking about it on Twitter. We're already following most of the journalists and analysts in this space.*

[What kinds of relationships have you developed with companies and research institutions?] *We have done custom projects [with people] that have first contacted us on Twitter... Sometimes people want to partner with us, because they see that we have a social media campaign, but we'll do our due diligence anyway. A serious business person contacting us on Twitter – that would give us pause.*

[Can you describe a situation where you obtained any new resources or business opportunities as a result of your Twitter usage?] *Yeah definitely. The biggest benefit has been through journalists or market people who follow us and learned about us on Twitter... When we first started, the content was drier. I didn't have a grasp on the materials. It took a couple months, but within 6 months, we were contacted by [major media outlets].. I'm not*

sure how they discovered us, but I'd like to think that that [i.e., social media] helped... To that degree, we don't get any more meaningful relationships. If any company is looking for a relationship (distributor), they already know their space and competitors, so there's not much to learn from Twitter.

The user from the third SME had not experienced any sales leads as a result of his social media usage; however, he created his Twitter account for personal (albeit professional) use more so than to represent his firm. “I don't think I can [recall a sales event tied to social media interaction]. It's a little harder to connect the dots using social media as a business tool. I never remember making a sale on Twitter” (F3).

An analysis of usage patterns (see Section 7.1 and Box 7.1) reveal that both F1 and F2 *engage* their followers by synthesizing inbound information flows (e.g., from Google alerts) and then simply tweeting that content. But these users go one step further by contextualizing or intermingling newsworthy content with editorial matter (e.g., opinions), press releases, and their products and services. According to the interviewees, this type of online presence results in an authentic way of engaging in ecosystem communications through thought-leadership rather than pushing sales.⁴⁶ However, in contrast to Marwick and boyd's (2010) finding that some users achieve authenticity through a combination of personal and professional tweeting styles, this research shows that innovative nanotechnology SMEs communicate on Twitter for professional purposes only. This orientation is due to the SME's desire to increase its credibility in the space.

In sum, the first two firms' responses show a strategic use of Twitter to achieve low-cost, unobtrusive marketing and to tap into an additional sales channel. At the same

⁴⁶ In practitioner circles, this is often referred to as “unobtrusive marketing”

time, social media is part of a larger sales cycle that spans many communication technologies, as well as social networks that exist both on and offline. Alignment between usage patterns and account strategy appears as a primary determinant of business outcomes: if an account is never intended for marketing purposes, then it is unlikely that sales opportunities will emerge as a result of usage. That said, the two firms that successfully use Twitter and other social media for sales and marketing purposes do so by monitoring nanotechnology and/or graphene news items and carefully curating their tweet content to provide salient and timely information. They also offer a glimpse into how their products and services fit into a larger narrative on graphene and nanotechnology R&D.

7.5.2. Jobs

The same user expressing skepticism about social media resulting in sales opportunities also acknowledged limitations of the platform in facilitating career moves. “If you found an opportunity on Twitter, like as a career change, but that’s probably pretty dramatic. I never looked to Twitter for an opportunity” (F3). In contrast, a user from an intermediary that tracks changes in the scientific job market described a more nuanced understanding of social media’s place in the workforce.

I don't think people are finding jobs, though they may be finding out about jobs. I think it's more about education stuff, resources for interviewing, how you'd network, writing a resume. The nuts and bolts. But it's specifically for this audience of scientists because that's going to be different than for an MBA, e.g., publications are important...

The job market has changed for scientists. One of the things we're hoping to do is encourage entrepreneurship and innovation for scientists. There have been huge changes, comparing the 1960s to today's job market. Sometimes we need to think about different careers. Some of it is about creating new opportunities, or becoming more flexible, e.g., leaving the country. (I1)

7.5.3. Impacts on the Broader Scientific Community

Anecdotal and survey evidence from popular science magazines and journals suggests that most scientists are not avid social media users (Piwowar, 2013; Van Noorden, 2014). However, social media usage increases with younger cohorts of junior scientists that have “grown up” around the web. There are additional field level differences: Emerging areas of study enabled by computer networks and software are more active on social media than older disciplines without such computing involvement (I2). In any case, many of these stories relate to the extent to which social media continues to have a transformative effect on the peer-review cycle. As noted by one respondent, when papers present breakthrough results that cannot be easily replicated, or that reveal a methodological flaw, scientists turn to blogs and other forms of online communication to raise their concerns.

I follow people that are engaged with nano research, for example the striping nano particle controversy. These Italian researchers suggested that the structural layers and molecular groupings behave in a certain way: they used AFM [Atomic Force Microscopy] to show that there is a structure these molecules are deposited in. They published in high impact journals, but there has been a lot of dissent. For example, there are folks from the UK and Scotland that have gone through great lengths to challenge this work.

[Is this a good thing – debate on social media?] It's a good thing at long as the debate is civil. If you can get people to engage and criticize in a public way with attribution, particularly in a public forum, that's good for science. (S1)

High-levels of interaction coupled with real-time responses allow for lively debates which may eventually resurface within the confines of peer reviewed literature (Mandavilli, 2011). In terms of challenging the narratives that emerge around some hyped technologies, Cossins (2014) reports that scientists – especially the younger ones –

are likely to interface with the popular press to temper “unchecked hype” as breakthroughs appear.

Given the increasing complexity of scientific inquiry, how well are scientific ideas communicated on Twitter across sectors, not just within scientific communities? For one scientist, bridging complex nanotechnology topics is possible on Twitter, though the capacity to do so is grounded in his ability to translate research findings into a context transcending any particular technology.

I am interested in graphene solar cells, for example. People are working hard on the theory side, and the 2D side of things. [Did you come across any related information on Twitter?] Twitter is helpful. I did find publications on graphene and solar cells in particular. But it's not just about graphene. If I follow [someone tweeting about] 2D materials – I might learn about transition metal dichalcogenides – e.g., WS₂ or WSe₂. These are the next set of 2D materials that are up and coming after graphene. (S2)

For another interviewee, links to additional web-based content offer enhanced information value within the constraints of 140 characters⁴⁷; rarely in his network do these links direct to academic journal content, however.

I don't think this [Twitter] works well unless you follow the links to where they're going. 90% of my tweets have links to articles [but] very seldom do they link back to a journal with even more depth. Twitter helps you maintain a high level understanding of something, but it's probably not the best thing for drilling down. But [at the same time] when you drill down, you're getting away from Twitter. (F3)

The inability to “drill-down” may be a limitation that deters some scientists interested in acquiring further specialized knowledge from using the platform. A second interviewee spoke in similar terms: The constraints of the character limit encourage a

⁴⁷ Indeed Table 3.1 shows that 92.60% of graphene tweets, which may be representative of the broader nanotechnology discourse on Twitter, contain urls. This is a much higher percentage of tweets compared to a random sample.

new set of skills for concisely sharing information. While possibly anathema to rigorous scientific exposition, tweeting may appeal to a wheel-and-deal mentality common in business settings or in ordinary conversation.

Speaking simply is important for describing your idea in a short amount of time, for example when seeking venture capital funding. Things like Twitter are important because you have to explain in an almost insanely brief way, maybe with a link to the article. It helps with that. (I1)

7.6. Summary and Synthesis

The results from this chapter reveal that most interviewees who actively use Twitter (i.e., those who log-in and read their tweets) experience increased awareness of ecosystem topics both within and outside of their substantive knowledge domains (Table 7.1). Social media does not supplant existing information channels but rather enables 1) more timely access to information, 2) a better way to “listen-in” on public sentiment related to nanotechnology R&D and commercialization, and 3) an effective mechanism to latch-on to subject area specific communities.

Awareness improves inbound knowledge flows, as is the case with technology scouting, which seeks to identify trends rather than details in the technological landscape (Parida, Westerberg, & Frishammar, 2012). In particular, higher levels of awareness of external sources of knowledge have been shown to improve behavioral additionality, as measured by a firm’s ability to improve internal processes related to innovation management, at low levels of experiential (hands-on) learning (Clarysse et al., 2009). However, because Twitter is a media platform, improved awareness is the most likely (and possibly least remarkable) type of outcome to assess because it moderates other

types of more impactful outcomes, such as hands-on learning, and does not directly result in improvements to behavioral additionality per se.

This study's results support the notion that improved awareness does not inevitably lead to other types of short-term innovation outcomes. In terms of problem solving (e.g., hands-on learning), for example, the interviewees could not name instances where Twitter helped address problems within nanotechnology R&D per se, though some were able to identify problem solving benefits in other areas such as analytics and data-oriented ideation. Other potential short-term outcomes depicted in the logic model in Figure 4.2 include enhanced participation, community/brand development, and customer and revenue growth. Five actors experienced these non-awareness outcomes as a result of their Twitter usage (Table 7.3).

Aside from the lack of problem solving outcomes, a pattern appears to emerge, where certain types of usage behaviors result in distinct benefits accruing to individual users (Table 7.3): The analysis shows that to achieve returns beyond better awareness, some users embark on a two-step process. First, they build and mobilize their reputation and credibility via their following network by sharing interesting information and *engaging* with other users. (This is the first step in the value creation process, and relies on awareness of ecosystem discourse.) Engagement appears to increase the strength of relationships, which initially develop as weak ties that become stronger through repeated interaction across multiple communication channels. Second, depending on the nature of any given interaction, some users marshal their contacts' resources to achieve some purposive end. (This is the second step in the value creation process.) In sum, assembling resources on social media is not decreed by order; it is a discursive,

sometimes spontaneous process that unfolds over time and over multiple communication channels.

As part of the second step in the value creation process, there is a need to transition communication from the public timeline to private channels, often completely outside of Twitter. Rarely did Twitter usage lead directly to a face-to-face encounter; other communication technologies, such as LinkedIn, email, Skype, and telephone calls mediated this transition. In total, five of the nine interviewees meaningfully transitioned a public Twitter interaction or relationship to a private channel, and the same five interviewees were able to achieve some type of longer-term outcome that enhanced individual returns or indirectly did so by increasing ecosystem participation. (Participation is a key outcome for two of the three intermediaries in Table 7.3.)

Taken together, the analysis reveals that information consumption on its own may lead to improved awareness. However, strategic sharing of information and engagement on Twitter, combined with an ability to transition conversations to private channels to marshal network resources, may result in innovation outcomes. These two key findings are discussed in further detail in the following chapter.

There are three notable limitations to this analysis, the first of which includes the obvious small sample-size of the interview panel, and an apparent correlation between usage behaviors and outcomes of interest. Specifically, respondents could not (or in one case would not) describe specific usage approaches that resulted in specific outcomes, and consequently, a causal argument cannot be made between usage and outcomes. A larger, more systematic study could isolate and quantify the potential determinants outlined here and measure the degree to which users are likely to experience specific

beneficial outcomes as a result of their social media usage. A research design that records user behaviors in a controlled setting could be one way to capture the necessary data.

Similarly, a second limitation of the analysis is the potential conflation of outcomes produced as a result of Twitter usage vs. social media usage more broadly. Some respondents in the panel could not remember the source of a new relationship (e.g. as originating on Twitter or LinkedIn), and thus the results may either under- or over-report the outcomes of interest.

Finally, because this study is exploratory in nature, no particular outcome of interest was assumed as being more or less likely *a priori* to the empirical work commencing. Since this research finds problem solving and deep contextual learning are unlikely outcomes of Twitter (and perhaps other social media) usage, future work could drop this variable for analysis. In this way, a more parsimonious theoretical framework could be developed with fewer determinants and outcome variables of interest.

Table 7.3: Usage and outcome evidence by respondent

Id	Information sharing	Improved awareness	Engagement behavior	Transitioning from public to private interaction	Sustained outcome including behavioral additionalities (but not including improved awareness)
M1	Syndication of content through RSS; does not log-in	No	No	No	None
I2	Tweeting specific information originally authored in a newsletter	Yes, for intra-organizational visibility	No	No	None
S2	Retweet science related news and stories	Yes, information within and across expertise areas	No	No	None
F3	Share whatever is interesting in two communities	Yes, discerned graphene as a trend	No	Attended one-time event at US agency	None long-term
F1	Share information in the field as a thought leader	No, uses Google Alerts	Relate products back to research breakthroughs, applications, and uses	Media exposure, sales from scientists and researchers and “custom projects”	Brand and community development, sales lift
F2	Tweet reading material as a service to the community and thereby attract followers and retweets	Yes, new articles and for public sentiment	4-5 interactions are routine, across channels as needed	Newsletter now has over 1,000 subscribers, partly due to Twitter and other social media; slowly build relationships leading to eventual sales	Brand development through unobtrusive marketing, community development, sales lift
I1	Anything being promoted or shared on website	Yes, indirectly citing the audience benefiting from members’ social media participation	Promotions, coordinating hashtags for events, being aware of what people are talking about.	Yes, through local members’ meetings (e.g., via Twitter, long-term) and through collaboration with a multi-national (e.g., via social media in general, short-term project)	Improved participation and community development (via Twitter in particular) and collaboration (social media in general)
I3	Retweet content from official accounts	Yes, for inter-agency news and intra-organizational visibility	Listening and conversing, encouraging a community of volunteers	In-person volunteer meet-ups	Enhanced ecosystem participation where volunteers take an active role; also engage scientific community within agency
S1	Mostly professional, sometimes personal	Yes, Twitter as a newsfeed tapping into research communities	Ad hoc, follows users across spheres of interest	Lectures and conferences	Community development across research areas

Sources: Interview transcripts, n=9 users, collected in 2014 and early 2015.

Recall that the regression analysis found non-random following relationships among actor types in the graphene firm-centric networks (P1): Actors choose whom to follow based on actor type similarity (i.e., homophily, P1a) and dissimilarly (i.e., heterophily, P1b). Yet, the quantitative results also show that at the micro-level actors choose whom to follow based on similarities in information content (thus providing no support for P2a). Clearly there remains much to be explained from the quantitative analysis. For example, what explains the differences in interaction across actor types accounting for the homophily and heterophily effects? Why does information distance – as measured by a granular topic modeling information distance measure – show a negative effect when Twitter is often portrayed in both the popular press and academic studies as an effective means for networking across interdisciplinary domains and across sectors?

The qualitative data suggest that different actor types have different concerns and constraints, and these factors drive behavior in meaningful ways on Twitter. However, this actor diversity stands in contrast to overarching similarities in information needs and tweeting patterns which self-organizing communities ultimately reveal. Subjective perceptions of belonging to communities can be traced to online and offline benefits, giving users a reason for continuing to invest in their social media presence. However, because individuals are embedded in communities, the quantitative method, which models micro-level dyadic linkages, does not adequately capture the boundary spanning following behaviors that some users exhibit. The next and final chapter delves into this conclusion with greater precision and extracts the management and policy implications of identifying, analyzing, and valuing these communities on social media.

CHAPTER 8: DISCUSSION AND CONCLUSION

The prior three results chapters explore this study's three main propositions using visualization, quantitative, and qualitative evidence. This final chapter summarizes these findings and discusses the results in the context of the literature reviewed in Chapter 2. It also introduces innovation policy and management implications for using social media to further the goals of individuals, firms, organizations, and the broader innovation ecosystem. The chapter concludes with a series of limitations, opportunities future work, and some final thoughts.

In terms of propositions, the findings generally support the assertions underlying P1a, P1b, and P3; however, the findings do not support P2a (Table 8.1). P1a and P1b suggested that following relationships are not random, and that users in the graphene innovation ecosystem follow others based on revealed identity and potential resources (e.g., certain types of information) that actor affiliation intimates. The results indeed show that Twitter users in the graphene innovation ecosystem follow one another in non-random ways that subscribe to patterns of both homophily (P1a) and heterophily (P1b). Note that P1a and P1b are not diametrically opposed to one another, but rather they are complimentary such that both types of following behavior can and do exist at the same time but under different circumstances.

Take for example the nanotechnology firm: The findings reveal that nanotechnology firms (which include the graphene firms in C3-445) are likely to follow all types of actors who tweet about graphene (e.g., media entities, intermediaries, scientists, and other firms), but these other types of actors are not likely to follow nanotechnology firms in turn. In addition, both the descriptive and regression results

show that media entities are likely to attract followers from across all other actor types; this is not surprising given the broadcast communication design functionality built into Twitter. Yet, the propensity to follow across actor affiliation types is not universal. For instance, intermediaries are not likely to follow nanotechnology firms, other firms, or scientists who tweet about graphene; and scientists are not likely to follow intermediaries and nanotechnology firms that tweet about graphene

The results also show that following within an actor category is limited to select cases. Nanotechnology firms are likely to follow other nanotechnology firms, just as scientists are likely to follow other scientists, once controlling for confounding factors including reciprocity, tweeting intensity, account age, and number of followers. However, intermediaries are not more or less likely to follow one another, once accounting for the same set of controls.

P2a proposes that actors choose whom to follow based on the perceived novelty of information accessible through network linkages.. The quantitative results do not support this proposition through the topic modeling approach used to operationalize information distance in the nanotechnology R&D-based community (“C3-945”). In fact, the ERGM specification shows that as information distance increases, the likelihood of tie existence goes down.⁴⁸ In a post-hoc test, however, greater differences in graphene tweeting positively predict tie formation. The mediating effect of information distance on the relationship between actor type and network structure (P2a) was not explored as planned because of the initial result undermining P2a.

⁴⁸ This finding stands in contrast to the interview evidence, which highlights that decisions on whom to follow are based on the type and quality of information that users tweet.

P3 suggests that innovation outcomes are more likely to occur in strategically-developed and information-rich social media networks. The qualitative findings reveal that information awareness is the most wide-spread advantage of using Twitter: Some users experience more timely access to information, while others – especially those who are attuned to their placement in diverse communities – learn about content related to their specific interests or professions. Although other outcomes are difficult to tie directly to Twitter participation, respondents – depending on their role and objectives for using the platform – were able to access and/or mobilize non-information related resources leading to sales lift, community and brand development, and enhanced participation. A key differentiator between improving information awareness and experiencing more significant tangible returns is the ability to engage the network and transfer communication from the public Twitter timeline to private channels of interaction.

Table 8.1: Summary of propositions, findings, and remarks

Proposition	Supported	Findings	Brief Interpretation
<i>P1a: Actors choose whom to follow by mixing across affiliation types (i.e. via heterophilous relationships).</i>	Partially	<ul style="list-style-type: none"> • Nanotechnology firms are likely to follow all types of actors who tweet about graphene, but these actors are not likely to follow nanotechnology firms in turn. • All actor groups are most likely to follow media entities. • Intermediaries are not likely to follow nanotechnology firms, other firms, or scientists who tweet about graphene • Scientists are not likely to follow intermediaries and nanotechnology firms that tweet about graphene 	Nanotechnology and graphene firms plug-in to existing communities to facilitate their growth and survival. However, following relationships are generally not reciprocated at the level of actor types (though at the micro-level, reciprocity does predict the existence of a following tie). This asymmetric characterization of the following network suggests nanotechnology firms use Twitter as a coping mechanism to traverse the “Darwinian Sea”. That is, entrepreneurs attach to the surrounding ecosystem, but the ecosystem has fewer connections to these firms in turn.
<i>P1b: Actors choose whom to follow by matching on affiliation type (i.e. via homophilous relationships).</i>	Partially	<ul style="list-style-type: none"> • Nanotechnology firms are likely to follow other nanotechnology firms • Scientists are likely to follow other scientists 	There is some evidence that homophily is an important driver among actors who tweet about graphene.
<i>P2a: Actors choose whom to follow based on the perceived novelty of information accessible through network linkages.</i>	Not supported	Greater levels of information distance do not positively predict network structure	Following relationships appear most often within communities sharing similar topical interests. However, ego network visualizations show that graphene firms may bridge distinct communities through structural holes.
<i>P2b: Information distance explains the following decisions of users better than actor affiliation mixing and matching alone can.</i>	Not explored as proposed	Recall that differences in graphene tweeting between two actors positively predicts tie formation, and that for some actor groups, this has the effect of mediating the relationship between actor mixing and network structure	Graphene is a hype technology, and within the nanotechnology R&D community, there is a demand for graphene-related content.
<i>P3: Innovation outcomes are likely to occur in strategically-developed and information-rich social media networks.</i>	Supported	Engagement in social media networks builds on participation and bidirectional communication to transfer relationships offline, and to enable important resource transfers.	Being embedded in a community helps focus attention and encourage dialog that can build relationships across channels and help actors achieve their goals. Community interfaces access social contexts in which strategic actors seek to construct narratives

This chapter synthesizes these findings in greater depth, particularly in terms of linking the results derived from exploratory propositions with other analytic insights (Sections 8.1 and 8.2). I also discuss this work in the context of theory development (Section 8.3), policy and management implications (Section 8.4), limitations (Section 8.4) and avenues for future work (Section 8.6). The chapter concludes with some final thoughts (Section 8.7).

8.1. Summary and Integration of Findings

The Twitter-based graphene innovation ecosystem can be characterized by a diverse blend of user types and communities, as observed by the combined friend network of the 33 graphene firms in the sample. Some firms, for example Lokimo Metals in the mining industry and Graphene3D in the 3d printing space, build friend networks predominantly in their specific domain of activity. Firms embedded within visually discernible communities suggest a high percentage of non-overlapping connections among their friends and other firms' friends. While these outlier firms may be interested in graphene, their positioning of the technology within the value chain (e.g., upstream sourcing of graphite, or downstream in the 3d printing arena) is unique across the sample firms. Out of the 33 firms in the sample, 17 place within the nanotechnology R&D community (C3) with another 10 found in the energy cluster (C0).

As the two ego networks in Chapter 5 show, however, firms maintain linkages across different communities depending on their revealed information and resource needs. (Some of these smaller communities simply do not appear in the larger combined friend network.) Recall that Zyvex Technologies is a Columbus, Ohio, based spin-out of Zyvex founded in 2007, and Graphene Frontiers is a University of Pennsylvania spin-out

established in 2011. Both firms follow regional actors within their hometowns of Columbus and Philadelphia, respectively. The presence of the regional innovation system (RIS) in each firm's network is not by accident: A growing consensus of scholars views regional linkages as a source of competitive advantage (Asheim & Coenen, 2006), in part because (a) regions along with their knowledge institutions act as hubs from which firms can absorb tacit knowledge and build intellectual property (Shapira & Youtie, 2008) and (b) regions promote "flexible specialization" wherein any liabilities of vertical disintegration are offset through agglomerated, agile supply chains (Simmie, 2005). The interesting point here is that these very local linkages can be clearly identified as dense communities even on a globally far-reaching, virtual platform such as Twitter.

Zyvex Technologies and Graphene Frontiers also exhibit linkages beyond their respective RISs. Both firms follow actors in a larger nanotechnology R&D community (C3 in the combined friend network, Figure 5.3). Zyvex Technologies follows two additional distinct clusters (composites and defense and aviation, C0 and C1, respectively), while Graphene Frontiers follows a sizeable venture capital community (C6). Given where both of these firms are in their respective growth cycles, the analysis in Chapter 5 cites external sources of evidence to weave a narrative around how these ego networks reflect the real-world, multi-faceted dispositions of these SMEs. Here we see the value of social media data underlying findings typically associated with multiple theoretical lenses – e.g., regional embeddedness, value chain placement, financing, and science communication – within a single empirical setting. The ego network visualizations succinctly convey how high-technology SMEs bridge multiple communities in the innovation ecosystem; consequently these diagrams lend some

credibility to the way in which the network sample was constructed, i.e., using graphene firms as the sample to access a relevant cross-section of ecosystem actors.

The results of the quantitative findings show that users in the graphene ecosystem – C3-945 to be exact – follow each other in ways that are non-random and that conform to *some* expected modes of interaction. For example, scientists are likely to follow scientists, and nanotechnology firms are likely to follow nanotechnology firms. At the same time, this homophily is counterbalanced by significant interaction tendencies across groups. As noted, all user types are less likely to follow identity-less “unknown” accounts, while intermediaries are less likely to follow nanotechnology firms or other firms in general. The findings highlight that in contrast to traditional linear models of innovation essentially ignoring the role of media intermediaries and individuals, these two actor classes appear to be critically important in defining the structure of the social media based ecosystem.

Yet, there exists a stark asymmetry between nanotechnology firms and other ecosystem actors: nanotechnology firms are likely to follow intermediaries, media entities, other nanotechnology firms, other firms, and scientists, but essentially none of these other actor classes are likely to follow nanotechnology firms (when controlling for general reciprocity and other confounding factors). Scholars often cite information asymmetries between high-tech entrepreneurs and customers, suppliers, funding agents, employees, and other ecosystem actors as a primary hurdle in developing a business (Audretsch et al., 2007), though high quality information reduces uncertainty and contributes to decision making through better informed conclusions (Newman, 2010). Indeed, the results show that entrepreneurs may face the same hurdles on social media as

they do in offline realms, e.g., in their inability to attract influential followers.

Consequently, while nanotechnology SMEs may use Twitter to traverse the perilous “Darwinian Sea” (Auersald, 2007; Auerswald & Branscomb, 2003) in the search for information and other resources, there is limited evidence to support that influential online ecosystem users systematically appreciate or are even aware of nanotechnology firms’ participation.

Surprisingly, in C3-945, users are less likely to follow one another as information distance increases, holding all else equal. That is, as a whole, Twitter users do *not* choose whom to follow based on the perceived novelty of information accessible through network linkages. Recall that this sampled and filtered community is highly focused around graphene and nanotechnology R&D. So, in a stylized sense, one can imagine that as a user’s discourse forays into politics, popular news, and other non-S&T domains, actors in this specialist cluster become less likely to follow that user. Given the insights from the qualitative interviews, this logic presents a conundrum: Users join Twitter to be exposed to multiple information sources, but yet the quantitative results show that increasing levels of information distance negatively predict tie formation. What could be the cause of this apparent discrepancy?

The answer may lie in two possible explanations. The first has to do with the underlying method. On Twitter, the dynamics of micro-level behavior produce one-off friend requests. ERGMs model the probability of a given network tie appearing given on a one unit increase (or decrease) in an explanatory variable, holding all else equal. As seen with the sample network visualizations, users build their presence and attach to multiple communities. These users may benefit from different types of information

across these communities, but in general, within the same community, content similarity likely prevails. Therefore, the statistical approach models individual interactions but misses cross-community meso-level phenomena. Future work is needed to test whether inter-community interaction can be appropriately modeled in the ERGM framework, and if so, whether the results validate this proposition of micro- versus vs. meso-level information dynamics.

The finding that increasing levels of information distance decrease the probability of network ties can also be an artifact of the specialized focus of C3 as a nanotechnology cluster. Here, users follow one another based on merit (i.e., value), representing the general ideals of the larger scientific enterprise. In particular, users may tacitly join this community to learn and share about graphene research and product breakthroughs, observe or participate in the emerging value chain, assess EHS risks and public sentiment, etc. Indeed, the result showing that the likelihood of tie formation increases as the difference in pairwise graphene tweeting rises offers a strong indicator that this is indeed the case. Consequently, if a user tweets about a broad set of topics outside of this information space, then they may be less likely to attract followers, at least within the same community. Because differential levels of graphene suggest a producer-consumer market for graphene related information,⁴⁹ tweeting across following relationships in C3-945 also implies that the graphene hype effect is still ongoing, and that Twitter is an especially suitable platform for distributing newsworthy content to pique further interest in the material.

⁴⁹ As opposed to a more even distribution of graphene information sharing across community members.

Taken together, the quantitative and qualitative results provide three data points suggesting that Twitter usage improves awareness of graphene topics. First, the comparison presented in the Data and Methods chapter shows that news items are more often tweeted than mundane musings in the graphene S&T ecosystem vis-a-vis a random sample of all public timeline tweets. Thus, in the professional S&T Twitter universe, information sharing is based on news-worthiness. Second, the combined and ego network visualizations show that users build following and friend relationships around topics; they seek out those who offer some kind of distinct information value. Third, the qualitative interviews reveal that some users are mindful about presenting themselves on Twitter with an eye toward providing information value as a strategy for gaining followers. Interviewees also identified learning about a variety of new topics through their Twitter usage, a finding also confirmed recently by Ooms et al. (2015) within a large firm context.

Given that users seek and share valuable information, it follows that the social networking component of Twitter – i.e., the relationships directing information flows – are not random. Users with identities are more likely than “unknown” accounts to attract a broad set of relationships; that is, users may follow an “unknown” account, but if that account does not represent a media entity, scientist, firm, or intermediary, then it is unlikely to attract additional followers. This speaks to the highly professionalized circles that operate on Twitter vis-à-vis more traditional network contexts where friends of friends are often friends.⁵⁰ In the conventional sense, people who know and often socialize with each other form dense cliques of relationships based on close-knit

⁵⁰ In fact, ERMG specifications controlling for the phenomenon where “friends of friends are friends with one another” did not converge with either simple triangle or more sophisticated k-star parameters.

friendship circles (Coleman, 1988). In the geographically dispersed, graphene-based innovation ecosystem, Twitter does contribute to inbound knowledge flows as conceived in the open innovation literature (e.g., Chesbrough, 2006), but users may not know one another and therefore rely on other proxies (e.g., information value, identity) when determining who to follow.

As suggested above, users construct their networks and strategically orient their behavior around some notion of “value”. Before discussing the types of beneficial innovation outcomes that ensue as a result of social media participation, I review how participation varies, mediates, and helps predict certain types of outcomes. In short, I argue below that all active social media users can benefit from increased information awareness but only some users can affect resource mobilization.

8.2. A Communication-Engagement Model to Predict Innovation Outcomes

The literature on innovation networks acknowledges the difference between exploration in ideation and exploitation of opportunity (March, 1991; Powell, 1990; Rothaermel & Alexandre, 2009). Exploration assumes scanning behaviors, whereby participants view the topography of a domain space and generate new ideas by recombining knowledge, finding applications for new technologies, or applying existing technologies to new problems. In contrast, exploitation through networks often relies on the importance of persistent relationships. As Uzzi (1996, 1997) notes, these “embedded” relationships facilitate trust building, using heuristics to simplify decision making, transferring tacit and complex information, and iterating on interparty problem solving.

The results from this study reveal that social media is squarely within the *exploration* domain of innovation activity; that is, social media participation and

communication exposes the latent qualities of a network to enhance relationship building and information diffusion. To understand how social media encourages innovative activity, I present an analogy of “working the room” – albeit in a virtual environment. In particular, this example uses a conference setting as a way to show how social media presents novel opportunities and challenges for communicating about and furthering outcomes related to innovation. The plausible series of outcomes are supported by the cases study evidence provided in the qualitative results chapter, while the findings from the quantitative and visualization results chapters offer context to the type of interaction occurring within and across groups of actors.

First let us begin with a hypothetical (but familiar) control scenario with face-to-face interaction at an academic conference: Imagine walking into a large convention center; the vastness of the space results in some isolated conversations because of incomplete knowledge of the participant and conversation space. How does one begin to interact? Upon entry, the participant may choose to introduce himself casually. This introduction may be based on prior acquaintance or recognition, subject matter interest, demographic homophily, random selection, etc. After a brief period of time, the gregarious onlooker may transition from one corner of the floor to another using his or her selection criteria (which may change) to engage different groups or individuals.

Now compare this to entering into an ecosystem of virtual actors on social media, and particularly on Twitter. First, there is no need to introduce oneself because identity, while perhaps helpful to attract followers, is not a prerequisite to access information and the structure of the latent network. The entry strategy is therefore asymmetric because following one user does not require reciprocation to participate, and this is akin to

surveying the entire landscape of the convention center and eavesdropping on many conversations at once. It is here that the first outcome of social media usage is most apparent with an increase in awareness of ecosystem activities vis-à-vis traditional communication channels (or in comparison to non-participation on Twitter).

Is this function of social media – that is, of surveying the landscape of the ecosystem in a short period of time – that much different from other communication technologies? I argue that social media participation indeed increases awareness in ways that differ from mailing lists, blogs, and private online networks. For example, a listserv can offer more timely access to information vis-a-vis weekly conference calls or in-person meetings. The interview data suggest social media offers a tangible mechanism to join communities and listen-in on public sentiment related to nanotechnology R&D and commercialization. A mailing list's (or blog's) content is not curated through individually tailored communities; in other words, one graphene firm will never receive the same information inputs as another firm unless their networks are identical. Just as a mailing list offers the same content to all subscribers, so too does a blog with its viewers. Therefore, notwithstanding saliency of information content, social media users experience information awareness as a function of their network participation *as well as* their ability to assimilate content, frequency of use, etc. In some cases (e.g., F2, F3, S1), respondents were keenly aware of the role that networks – and specifically communities – play in effectively accessing and sharing valuable information.⁵¹

“Participation” in the social media network is an ambiguous way of describing interaction. As the qualitative results show, some interviewees are more or less *engaged*

⁵¹ In other cases, community development was an outcome for mobilizing resources other than information.

with their audiences. For example, recall that the operator of a specialized graphene media outlet barely logs-in to read his tweets and instead defers to automatic tweeting enabled by RSS feeds from his website. This user built a network over time by essentially following users in an ad hoc way or following the set of users that Twitter recommended. Compare this passive type of participation to one of the graphene firms working proactively to respond to inquiries, relate popular news stories to its product line, and follow students and researchers to extend the boundaries of its potential customer-base. The second set of outcomes going beyond information access via a “news feed” appears to be mediated by the level of engagement across communities and/or the broader ecosystem.

Engagement in this study was exhibited by responding to tweets, holding Twitter conversations, and posting relevant and useful information to followers, e.g., as part of a community orientation. As one respondent noted, “The result is non-intrusive marketing – there is valuable information but I’m not selling them anything. I’m just reminding them of myself” (F2). I surmise that the response of this type of engagement is *attention*, a latent variable not expressly measured as part of this study’s research design. That is, on a platform where information may overwhelm users, engagement may lead to an attentive audience. For some users, this results in community building where firms become thought leaders (e.g., F2), where intermediaries attract the interest and participation of members or volunteers in offline meetings (e.g., I1 and I3), and where scientists may attract followers from different research domains (e.g., S1)

Beyond community development and enhanced participation, other sustained outcomes include increased sales through expanded marketing reach and improved brand

awareness. Less tangible outcomes with potential longer-term implications for ecosystem capacity building involve disseminating job market related best practices and opportunities for non-traditional positions and furthering discourse of nanotechnology related R&D both within the scientific community and across sectors. However, Twitter as a social media platform rarely resulted in a *direct* sales opportunity, job lead, or problem solving gain. Instead the platform mediates communication that moves beyond the public timeline into the realm of private channels, e.g., individual inboxes, phones or Skype accounts, or LinkedIn private messaging. Transitioning to private modes of communication is arguably a tangible outcome in professional circles, R&D or otherwise; it signals the germination of ideas, acquaintances, and potential partnerships or business arrangements. In this way, the offline relationship building process is no different than exchanging personal contact information at a conference. However, the funnel is arguably an infinite times wider. Figure 8.1 shows an individual user's communication pipeline beginning with participation to engagement, and the various types of outcomes that ensue in each mode of interaction. I term this the social media communication-engagement model.

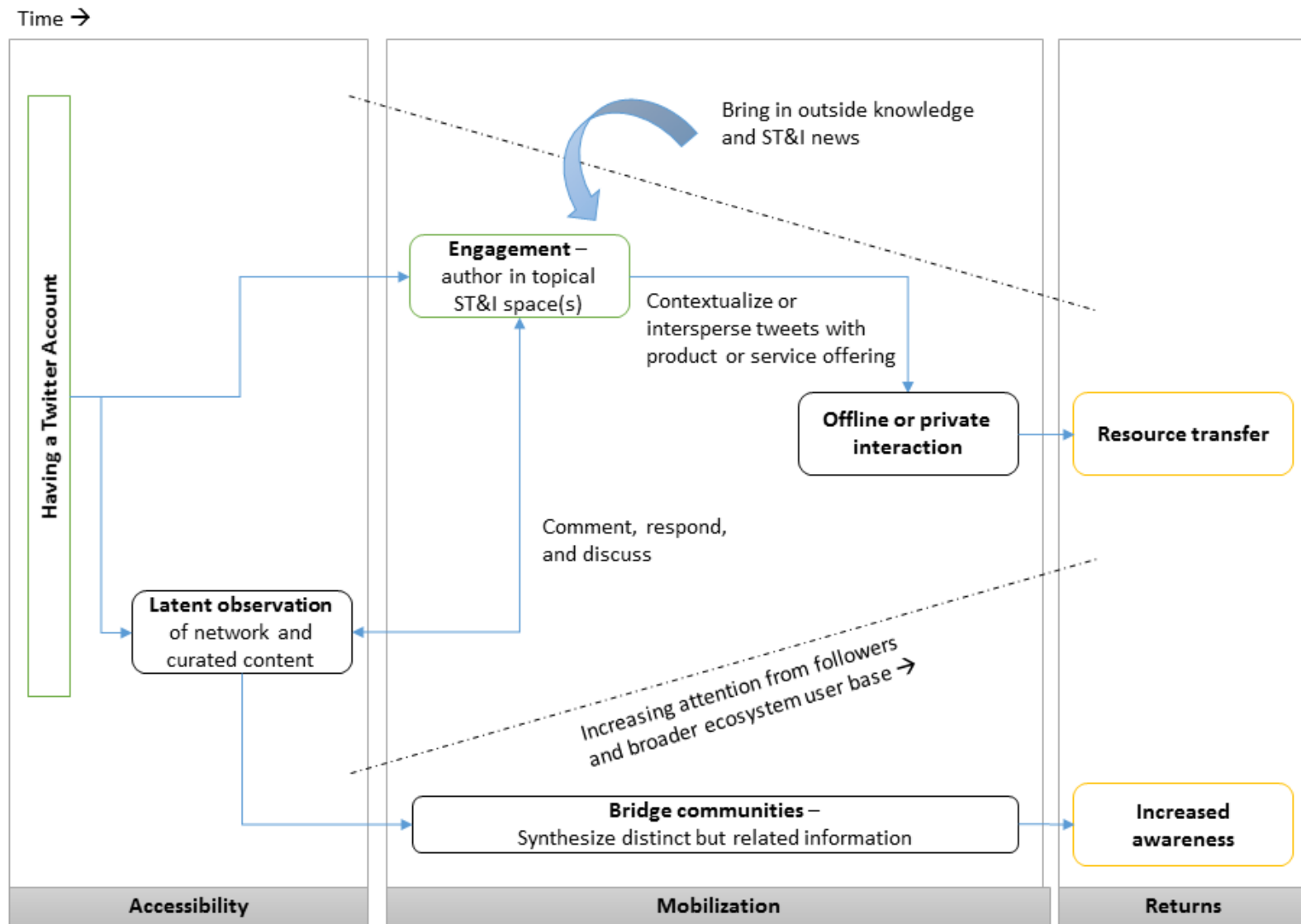


Figure 8.1: A social media communication-engagement model building on Lin’s (2001) model of social capital development

With little to no engagement, actors' latent observation of the network and curated content result in increased awareness of ecosystem topics. These are passive users. In this *modus operandi*, there is limited opportunity for other forms of resource transfer. However, with some degree of engagement, e.g., retweeting, directing conversation, and co-creating narratives, users are able to capture attention and direct communication to non-public channels. These are active users. In more intimate settings, resource transfer may (eventually) ensue. The bottom of Figure 8.1 depicts the “capitalization” and “effects” components of Lin's (2001) social capital path diagram (Figure A.2). In essence, there is very little mobilization of resources in the bottom path, whereas the top part of the diagram shows users accessing network contacts and resources to achieve “instrumental returns”.

8.3. Theoretical Implications

This work makes two significant contributions to the literature. First, it bridges two different but complimentary research streams, the innovation systems and entrepreneurship literature. By blending institutional theories of innovation activity with micro-level phenomena, this research combines exogenous and endogenous parameters that account for both perspectives. Second, through a novel methodological approach, it evaluates how information distance impacts the formation of network formation. This empirical test continues an emerging trend in social network research that observes and quantifies the role of information in network construction, particularly by using online data sources such as email and social media. Implications for theory are assessed at individual and agglomerated units of analysis.

8.3.1. Integrating Two Research Streams through Narrative Construction

Most theoretical contributions are incremental, and the following text is no different in its ambition; it intends to synthesize and reconcile institutional and agency-based explanations of innovation through the co-creation of narratives and social contexts. In particular, the treatment below examines “who”, “when”, “where” and “how” aspects (c.f. Whetten, 1989) of narrative construction in communities active in innovation ecosystems. The “who” speaks to a diversity of ecosystem actors; the “when” conveys a specific type of interaction of engagement, not just passive participation, while “where” delineates the boundaries of the innovation ecosystem on social media. Finally, the “how” element proposes a three-step process by which narrative construction unfolds. To appropriately frame this exposition, I begin with a high-level description of innovation systems and entrepreneurship theoretical underpinnings.

The innovation systems literature explains regional, national, and supranational differences in innovative activity as a function of a geography’s institutional infrastructure (Cooke, 2001; Doloreux & Parto, 2005; Metcalfe, 1995; Neal et al., 2008). Actors such as universities, firms, public research institutions, and incubators enable and steer knowledge flows in ways that improve or limit innovation outcomes such as economic growth and competitiveness (Nelson & Rosenberg, 1993; OECD, 1997). In this framework, public policies intervene into the innovation system topography to improve institutional interaction. In many cases, especially at the national level, innovation systems exhibit path dependent trajectories as a result of the slow moving nature of existing market, regulatory, intellectual property, taxation, and financing laws and rules. This approach, like many other institutional frameworks, does not specifically

acknowledge agency at the individual level; explanatory mechanisms reflect top-down drivers of individual activity (Autio, Kenney, Mustar, Siegel, & Wright, 2014).

In contrast to the macro-focus of the innovation system literature, entrepreneurship research often concentrates on the individual (or small firm or team) as a primary unit of analysis (Moroz & Hindle, 2012). Depending on the theoretical lens, the entrepreneur experiences beneficial outcomes by harnessing resources at the individual, network, or societal (i.e., institutional) levels (Garud, Gehman, & Giuliani, 2014). New value creation enhances the environment through either incremental or radical change. While some frameworks show new venture creation as a series of optimizing steps for the resource constrained yet goal setting entrepreneur, other approaches assume the entrepreneur first surveys the environment and available resources before setting goals and optimizing (Sarasvathy, 2001).

Autio et al. (2014) contend that research on “entrepreneurial innovation” can and should integrate the two streams on innovation system and entrepreneurship by recognizing a series of contextual factors. Among these various factors is a social component acknowledging that “knowledge is widely dispersed among many heterogenous agents and that the interactions and exchanges between them are crucial for new knowledge production, and hence entrepreneurial innovation” (p. 1101). This focus on heterogenous agents and their relationships is the primary theoretical motivation of this research, which argues that a descriptive understanding of the micro-foundations of organic network formation in high-technology environments necessarily involves a diversity of actors not usually observed using traditional research methods.

A significant theoretical implication of this work, then, stems from the social contexts exposed through the online innovation ecosystem. As the empirical results indicate, there is no one social context to identify; instead, graphene firms access multiple social contexts through their online social media participation. Furthermore, actors either passively or actively participate in communities via *online interfaces* to those communities:⁵² I posit that each user's friend network, and the communities that may exist therein, acts as a postern into a given social context, which may represent a technology or science domain, regional interests, a value chain, venture capital, etc. While communities may consist of a homogenous set of actors – e.g., the venture capital community reveals a financing social context – online communities can incorporate a wide variety of user types not typically conceived of by traditional theoretic lenses.⁵³ Figure 8.2 summarizes this framework constituting the high-tech entrepreneur's orientation towards the online ecosystem via several community interfaces and the social contexts they reveal.

Recent theoretical work on social contexts transcends the individual vs. institutional dichotomy. As a theoretical construct, social context is useful on several fronts (Garud et al., 2014): It provides the context by which entrepreneurial opportunities are discovered or created; it also affords a frame of reference by which agency and environmental factors co-emerge through recursive processes that seek to stabilize a capricious technology and market landscape. Finally, social contexts help explain how narratives unfold in entrepreneurial innovation, a topic which was examined in some

⁵² Several interviewees recognized the difference between an offline community of individuals and their online personas constituting a representation of that community online.

⁵³ For example, the network visualizations present a venture capital community with embedded media entities, as well as a nanotechnology R&D community with a number of unidentifiable users.

detail in Chapter 3. The primary assertion submitted then was that entrepreneurs continually seek to position a technology and their role in the co-creation of that technology into a socially acceptable narrative in order to gain stakeholder buy-in and build a reputation.

The narrative concept is not new to science and technology studies, but insights gained from this research may better explain the processes by which narratives unfold in the online (or offline) innovation ecosystem. First, generating narratives is not a foregone conclusion as a result of ecosystem participation, as discussed above with the communication-engagement model (Figure 8.1). Entrepreneurs seeking to co-create or exploit narratives think carefully about audiences, the social contexts those audiences represent, and the best way to enter into and contribute to mediating communities. As shown in Figure 8.2., it is possible for more than one community to access a social context, e.g., in the case of a politically divisive issue such as EHS risk of nanotechnology, so entering one community vis-à-vis another can be an important consideration.

Second, the high-tech entrepreneur interprets content from one or more community. Based on his or her discretion and anticipation of how the community will perceive any number of responses, the entrepreneur then chooses a reply. The strategic intent behind a reply can be reproduced in future communications (e.g., tweets) to solidify the message. Thus, to the extent possible, the social context is now co-created through a narrative that intertwines the entrepreneur's endogenous choice in messaging with the narrative of the community.

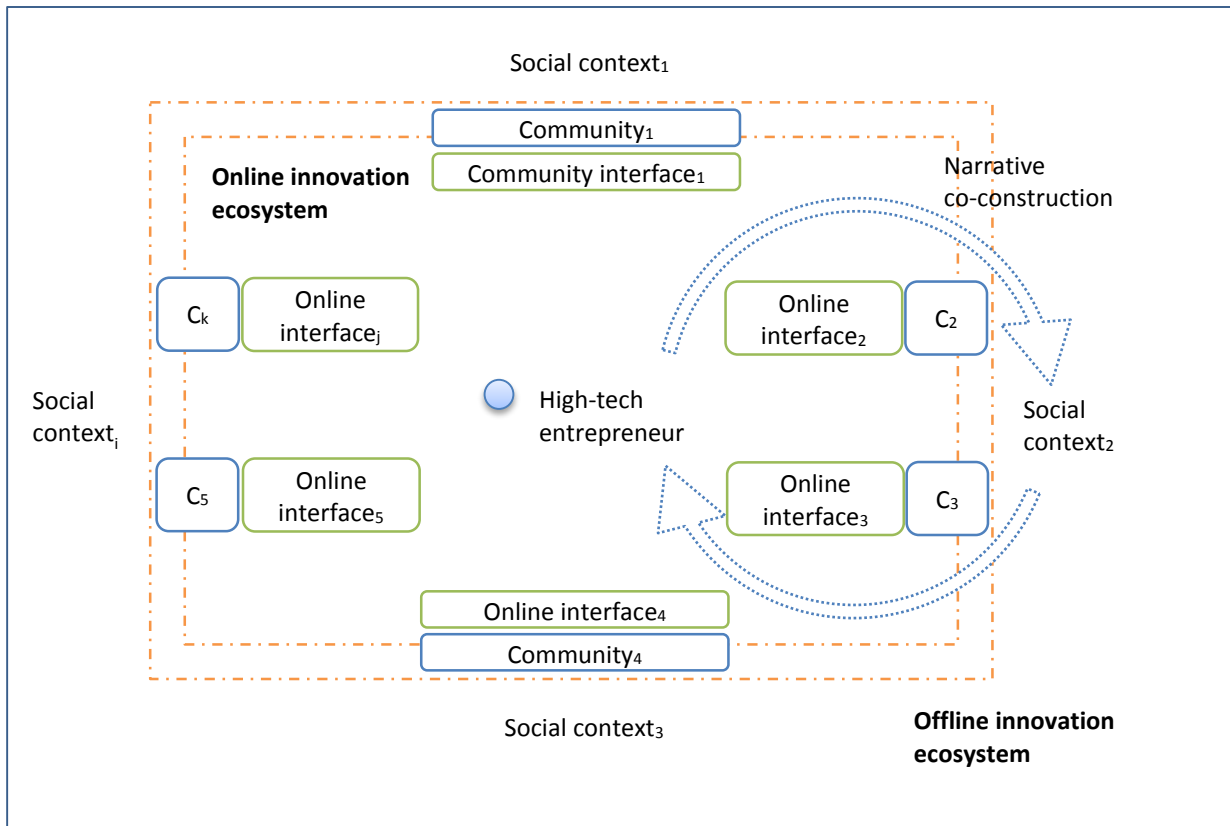


Figure 8.2: Narrative co-creation process

Third, the narrative (or more accurately, a piece of the narrative) intended for one community may diffuse across one or more communities. This is a peculiar aspect of social media, and Twitter in particular, because of its broadcast communication design: Because followers receive all tweets authored by a user, individual users may be exposed to and begin participating in multiple communities over some time. This may result in a slow but perceptible blending of social contexts over time, the implications of which could potentially impact individual entrepreneurs and other ecosystem actors. This potential feedback between ecosystem participants and the social context, as mediated through community participation and engagement and as revealed through co-created narratives, is shown in the cyclical arrows in Figure 8.2.

As discussed throughout this work, the high-tech entrepreneur does not act alone in setting the social context; the ecosystem consists of a diverse set of actors. Consequently, rather than viewing the narrative as one that is encountered and manipulated by just the entrepreneur, it is more accurate (though methodologically more difficult) to capture contributions from the plethora of other actors involved in the narrative co-construction process. The theoretical framework depicted in Figure 8.2 can therefore place the scientist, intermediary, venture capitalist, media entity, or large firm at the center of the diagram. As a result, the theoretical contributions of this section are not limited to just innovation and entrepreneurship but may facilitate further insights and empirical testing in disciplines ranging from science communication and economic geography to strategic management.

8.3.2. Social Capital and Information Value

A stylized synopsis of social capital theory positions network density and cohesiveness vis-à-vis the importance of bridging relationships, structural holes, and weak ties (Burt, 2001; Granovetter, 1973; Kadushin, 2011).⁵⁴ Denser networks consist of strong ties characterized by frequent communication and the propagation of redundant information. In contrast, structural holes and the bridging relationships that link them tend to be associated with less frequent communication and access to non-redundant information.

Aral and Walker (2011) subtly examine this continuum between structural holes and areas of high network density by isolating the effects of network diversity and channel bandwidth on information diversity and novelty. Using a dataset consisting of

⁵⁴ Refer to Chapter 2+ (Literature Review) for additional details.

corporate emails from an executive headhunting agency, they determine that increases in network diversity result in higher levels of information diversity and novelty. This finding mirrors the established theoretical relationship between weak ties and their ability to transfer non-redundant information into a cohesive group of actors.

The authors proceed to theorize and test the mediating effect of channel bandwidth on the relationship between network diversity and information novelty/diversity, arguing that as networks become more *diverse*, channel bandwidth reduces and therefore less novel and diverse information can be communicated. Their line of reasoning assumes that channel bandwidth and network diversity are inversely related; i.e., as network diversity increases, channel bandwidth decreases. While this finding may be true for most communication channels where conversation flows privately from one party to another, broadcast mediums such as those found on social media platforms suggest a different way of theorizing this relationship.

Counter to initial expectations, the findings in this research show increasing levels of information diversity result in fewer network linkages overall.⁵⁵ Therefore, most actors with following relationships share the same topical content. Yet it is likely that some of the same linkages that represent boundary spanning connections actually produce a significant amount of information for a focal user to process (e.g., consider media entities, who may be found in any community and who tweet more than any other user type). Therefore, it is possible that network diversity and channel bandwidth are not inversely related on social media, and that perceived information novelty and diversity

⁵⁵ Note the causal direction of this generalized relationship is reversed from what is posed by Aral and Walker (2011), who contend that network diversity results in greater levels of information diversity.

increase with network diversity until channel bandwidth saturates a user's cognitive processing capacity.

Under these assumptions, weak ties remain “weak” because of infrequent bidirectional communication, yet they simultaneously produce a *high volume* of diverse and novel content that users may assimilate to their benefit, depending on their absorptive capacity and ability to contextualize that information for their immediate problem domain and organizational setting. Recent commentary in the practitioner literature recommends that Twitter users wanting to increase the diversity of their networks should be less interested in following more people than in pruning their networks periodically to consistently infuse heterogeneity in opinion (Parise, Whelan, & Todd, 2015). The commentary also highlights a need to transition communication to higher-fidelity modes (e.g., in person meetings) to convert weak ties into strong (offline) ties and to further explore complex ideas not amenable to Twitter's character limits. Such advice is commensurate with the communication-engagement model presented in Figure 8.1, which implies a set of optimal usage patterns for online ecosystem users.

8.4. Innovation Policy and Management Implications

This research offers several innovation management and policy implications. I consider policy implications primarily for intermediaries and scientists, while management implications are viewed from the perspective of the high-tech entrepreneur or startup firm.

The first public policy implication speaks to the diversity of online ecosystems and the types of interactions one would expect to find in robust innovation systems. The innovation system literature posits that institutions – be it at the national, regional, or

supranational levels – work together to shape innovation policy, the production of knowledge, and the diffusion of technologies (Cooke, 2001; Doloreux & Parto, 2005; Metcalfe, 1995; Nelson & Rosenberg, 1993). For example, universities conduct basic and applied research which then may be adopted by firms, whose profit seeking motives require them to exploit one or more markets. Public policy informs market and regulatory conditions and the funding of certain types of research. In regional settings, clusters of firms, their suppliers, and universities may create a critical mass of dynamic knowledge spillovers that also further innovation outcomes.

The findings in this research show that the innovation ecosystem construct, when applied to an online setting, reveals a trend among follower relationships that does not always conform to expected (i.e., ideal) modes of interaction. For instance, many of the intermediaries tweeting about graphene maintain few follower linkages to firms and scientists; however, many of these firms and scientists are likely to follow intermediaries. This suggests an information asymmetry in terms of what intermediaries provide to the rest of the ecosystem: There may be ample opportunity for intermediaries to broadcast news items, events, etc. to their audience but fewer chances to listen-in and therefore learn from and bridge between other actors in the ecosystem.

While any institution would have a hard time grappling with the massive amount of information provided in their Twitter timeline if they followed everyone who followed them, the results here show that communities are an important component of how the ecosystem takes shape online. Being able to discern where an intermediary sits in terms of communities and following an ample number of users in other related communities may provide the university, funding agency, regulatory body, regional authority, non-

profit think tank, etc. with a better understanding of how they are perceived and how they may go about furthering their missions online. Additionally, the qualitative results show that some intermediaries are unable to or lack the knowledge or resources to follow other users in such a way. Public policies could intervene where appropriate to encourage intermediaries to follow users – or at least passively monitor ecosystem development – whilst avoiding the impression of advocating for one type of user base vs. another. For example, the technique of science maps described as future work below could be one useful tool in facilitating an organization’s understanding of where they fit within the broader ecosystem.

Intermediaries lacking social media expertise could learn from their peers. The qualitative results reveal a degree of information sharing with respect to lessons learned and best practices among social media experts at US federal agencies. Another science and technology intermediary, however, struggles to fashion a social media niche that fulfills initial expectations within the constraints of available time and resources. Formal and informal knowledge sharing sessions could facilitate training opportunities to help agencies, universities, etc. in the same region or country better utilize social media to achieve desirable outcomes.

Scientists too are in a unique position to follow users in a broad arena of topics to further science communication. Prior work on “open science” stresses two important roles of online interaction: (1) sharing data and information, and (2) providing opportunities for subject matter experts to identify themselves and spontaneously contribute in substantive ways (Woelfle, Olliaro, & Todd, 2011). The results here show that scientists are likely to follow other scientists but are relatively less likely to follow

firms and other users. As a result, scientists may be missing an important opportunity to ascertain how their research indirectly impacts and aligns with other actor interests and communities. By following a broad array of user types, scientists increase the likelihood of users following them. More following relationship could result in greater visibility of individual and collective research efforts, not limited to but including the potential for higher citation counts. From the perspective of other users, scientists who engage with non-scientists may help close gaps in knowledge by providing valuable information, e.g., technical knowledge, professional guidance, etc. Of course, Twitter and social media is not the only channel in the long-term communication-engagement model, but rather the first step.

In sum, public policies seeking to facilitate information exchange and learning may encourage social media participation and engagement to help bridge knowledge and networking gaps between all actor types. Implicit in this argument is the importance of maintaining a diversity and tolerance of ideas in innovative cultures (Wallner & Menrad, 2011). That is, enhancing the vitality of the online ecosystem with the intent of improving a culture of innovativeness should be the goal for innovation policies aimed at governing social media. As discussed, when left to their own devices, users may not reach out to diverse actors in terms of role identity and information content. Thus, it may be helpful for public sector *and* non-profit institutions to leverage existing in-person conferences, networking events, etc. to facilitate the movement of offline relationships to online platforms. Not only will relationships persist even after periods of inactive communication but on some platforms such as Twitter these linkages can be discovered and reproduced by other actors. Returns to such an intervention may include enhanced

knowledge sharing, better system-wide learning, and increased inter-sector collaboration opportunities. For impact assessment, quantifying actual short and long-term outcomes would entail carefully identifying the value-add of the intervention in the presence of the counterfactual (i.e., if some social media relationships did not materialize as a result of the intervention) (Georghiou, 2002; also see K. B. Smith & Larimer, 2009).

In terms of management implications, social media offers challenges and opportunities, and I consider both angles from the perspective of the high-tech startup firm or entrepreneur grappling to use social media (and more particularly, Twitter) in a cost-effective way. First, the qualitative results show that most ecosystem users begin on Twitter in an ad hoc manner. For example, they experiment by following certain kinds of users and tweeting certain types of content. The results and subsequent synthesis presented here suggest that new users should begin by learning how competitors or other firms in the network are using the platform. It is not necessary for users to visualize a network; instead they can examine the following and friend relationships of key user accounts and attempt to reproduce a network in the same or similar manner. I call this the “Twitter as a template” approach, which after some amount of replication, allows for later fine-tuning. For example, the high-tech entrepreneur may first plug into the regional innovation system and then follow notable firms in the global nanotechnology supply chain. Then, depending on need, specific scientists or venture capitalists can be followed to provide additional information and resource value.

It is assumed that media entities will be intermingled in all communities given their pervasive presence on Twitter; however, it may make sense to distinguish between strict information providers and those actors in the network that may not tweet as often.

For instance, by not following media entities, or by placing media entities in separate user-generated lists, tweets from scientists, other firms, and intermediaries may rise to the surface in the entrepreneur's news feed: Recall that part of the challenge with Twitter, as documented in this research and elsewhere, is to overcome information overload. That said, media entities are still a critical element of the ecosystem fabric, as they have been shown in other settings to be the primary diffusers of information across a network (Cha, Benevenuto, Haddadi, & Gummadi, 2012).

For entrepreneurs and high-tech firms already with active Twitter accounts, one way to get more from the social media platform is take an insular approach to examining extant networks. Here, it is possible to visualize a focal ego network and identify various communities in a manner similar to the one presented in this research. The goal of this exercise is to determine topical focus and information value deriving from each community in the network. The entrepreneur may then examine the findings in the context of the strategic needs of his or her business coupled with an understanding of the limitations and opportunities afforded by Twitter. This method of gap analysis can be used to promote "Twitter as a gateway" to access additional communities.

On one hand, the quantitative results show that higher levels of information distance, when modeled on a per edge basis, are less likely to result in a following relationship; on the other hand, theoretical insights and the network visualizations show community clustering occurs via topical interest areas. Using "Twitter as a gateway" recognizes that following recommendations from the platform largely encourage users to stay within their current communities. In the S&T domain, however, there is reason to believe that exploration beyond one's immediate environment could add incremental

benefits to network participation. For example, there is an opportunity for firms to look further upstream or downstream in the value chain to gain a better perspective of their current position and anticipated placement in the ecosystem. Commensurate with exploration and search, graphene firms manufacturing the material as an intermediate input could follow end-user product firms to learn more about that downstream industry, and vice-versa for downstream firms aiming to learn more about supply-side dynamics.

For the entrepreneur or high-tech startup with an existing Twitter account, this means choosing carefully between communities to gain the most benefit from participation. For instance, following scientists within the exact sphere of research conducted by the entrepreneur or high-tech firm may represent reputational or personal comradery benefits rather than incremental informational value. However, the entrepreneur or high-tech firm may find it particularly helpful to follow scientists in related fields to track tangential developments that could potentially offer insights into the firm's primary R&D efforts. In other words, while Twitter does not act as a substitute for journal literature, it can complement other sources of learning about new scientific content (e.g., search engines and indexes, publisher notices, informal conversation, etc.)

In this way, the high-tech entrepreneur or startup is uniquely positioned to bridge several different communities in the innovation ecosystem. Indeed, the network visualizations show that the full graphene innovation ecosystem incorporates a diverse suite of actor types that do not necessarily follow one another. As discussed with the communication-engagement model, it is not enough for high-tech entrepreneurs and firms to participate passively. To extract resources, to transfer public interaction into private channels of communication, and to facilitate sales, these actors must engage with

other users in ways that solicit conversation and exchange;⁵⁶ engagement is all the more important to attest to these communities that the firms exist in the first place, and that the firms can contribute to a meaningful narrative. To do this well, firms may wish to invest in dedicated (yet part-time) social media staff, who can shape the network and who can effectuate the ecosystem *through* their engagement. As one interviewee remarked, consistent engagement meant sharing salient information to its followers; eventually the firm was contacted by well-known media entities who would increase its visibility throughout the ecosystem and popular press.

I conclude this section with a platform suggestion for Twitter. Given the emphasis on community development in professional contexts (such as the graphene innovation ecosystem), Twitter may wish to explore its ability to increase its users' sense of belonging to communities. Consequently, instead of recommending users, the platform could recommend joining communities. This would allow users to examine the vast social network as a (subjective) set of a handful of communities in which they could participate and to which they would like to belong. To some extent, a first step in this direction has already been accomplished with user defined lists, but word clouds or network visualizations could be an alternative option for enhancing the user interface. After all, knowing that communities exist and exposing those communities for users to see and interact with are two very different things.

8.5. Limitations

This research is limited in three ways. These limitations span a set of research design and conceptual aspects, including sampling and statistical methods, dynamism of

⁵⁶ In truth, this same argument could be made for all types of users, not just firms per se.

the underlying social media platform (data source), and theoretical foundations for explaining social media participation in online innovation ecosystems.

8.5.1. Alternative Explanations

This research argues that users in the graphene innovation ecosystem participate on Twitter and build network linkages to access diverse information and resources. Diversity is approximated by heterophilous following and friend relationships across different actor types; diversity is also ascertained by evaluating information distance between users. The findings show some evidence that Twitter indeed promotes mixing of actor types and information, and moreover, that some users are keenly aware of this benefit. For example, participants may enhance their reputations, firms may increase their sales, scientists may access adjacent research fields, and intermediaries may capture more attention for their work. This theoretical framework is supported by the literature on innovation networks, research collaboration, and entrepreneurship.

In contrast to explaining and predicting network linkages via social capital theory, which posits that users access, mobilize, and transfer resources online for individual benefit, alternative explanations exist: Some users may participate on Twitter not for economic gain but to further other types of utility. Toubia and Stephen (2013) argue that non-commercial users derive *intrinsic* and *image-related* utility from participation on Twitter. Intrinsic utility reflects an individual's innate satisfaction in tweeting to his or her audience; it increases as a function of a user's number of followers. Image-related utility, on the other hand, measures the importance of enhancing one's online footprint, largely construed as a sense of "self-worth" and social acceptance; this type of utility is measured as a nondecreasing concave function of a user's followers. Through empirical

exploration and a dynamic discrete choice model, Toubia and Stephen observe that both types of utility may operate at the same time. At first, when users have few followers, they tweet to increase their intrinsic utility, and this behavior increases the number of followers. As time progresses, and as their follower network grows, users tweet less often because image-related utility becomes more important than intrinsic utility. The qualitative findings show that most users improve their awareness of ecosystem concepts and news through their network curated timelines. This awareness utility is distinct from both the intrinsic and image-related varieties described above. There are also resource transfers that take place in the graphene innovation ecosystem; this type of utility is akin to traditional notions of economic utility (albeit mediated by networks and not “informal” markets).

The interview evidence suggests that social capital facilitates network transfer of information and resources for many users. However, some participants (e.g., F2 and F3) engage on Twitter to increase intrinsic and image-related utility: Interviewee F3, for example, tweets about lithography and leadership topics of interest to him, more or less irrespective of the appeal of such content to his followers. From this perspective, F3 subscribes to the intrinsic utility model of tweeting. Interviewee F2, in contrast, tweets to attract followers and to enhance his image-related utility. This user then converts interest gained from follower (and friend) relationships according to the social capital model presented in Figure A.2.

A second alternative explanation accounting for social media participation can be traced to the “collective assets” component (trust, norms, etc.) of Lin’s (2001) social capital model (Figure A.2). Suppose for a moment that individuals within the graphene

innovation ecosystem operate in diverse offline networks to further their professional goals. Then, the online component of this innovation ecosystem may be a reflection of this offline reality. Here the collective asset could be considered as a baseline community commitment to diversity and obligation to participate in similarly diverse ways online. For instance, both F2 and I2 acknowledged that significant portions of their Twitter networks exist in the “offline world.” F2 in particular noted that the Twitter community he cultivates should exist in the “real-world” because only then does it have meaning.

Given the other empirical evidence, however, this alternative explanation does not seem entirely satisfactory because some users clearly follow others they do not know (e.g. celebrities, institutions, and media entities). Additionally, many interviewees conceded would not be able identify all of their followers. Consequently, I conclude that network development on Twitter is in part a discovery-driven process, one that does not rely exclusively on existing offline relationships.

8.5.2. Methodological Reflections

The methodological limitations in this study derive from two main components: the initial sampling approach and the nature of the ERGM itself. The sample size for the interviews was only nine; limitations associated with the qualitative findings and interpretations are given at the end of Chapter 7. While the number of interviews could be increased to improve validity, the findings presented here are generally commensurate with recent work on social media in related contexts. For example, work by Ooms et al. (2015) cited in this research suggests that firms use social media to increase inbound knowledge flows and facilitate absorptive capacity. (This is essentially the awareness

improving finding across all users recounted above.) Lovejoy et al. (2012) find that non-profits on Twitter use the platform not as a tool for engagement but rather as a medium for one-way mass communication. (This result corresponds to this research’s finding that intermediaries often appear to use Twitter as a way to share news, with little interaction across actor types.)

In addition, the apparent correlation between usage patterns and outcomes – i.e., passive users experiencing only improved awareness, while more active users mobilize attention and resources for other non-awareness outcomes – emerged relatively early on in the interview process and repeated thematically in subsequent sessions. Differences in perspective quickly “saturated”, such that additional interviews conveyed less and less new information, thus alleviating the need for a larger sample size.⁵⁷

The sampling method for the combined friend and follower networks began with a sample of 37 graphene firms. I captured the friend and follower networks of each of these firms in one dataset and then sought to identify linkages among each of the non-graphene firm users. In this way, the sampling method is quasi-snowball. (Further sampling to increase the span of hops to all friends of friends would have exhausted available time and resources given this study’s use of the free Twitter API.)

The set of users captured by the sample does not reflect the broader spectrum of graphene related discourse and activity on Twitter. One way to measure the magnitude of this disparity is to compute the percentage of users in the combined friend network *not* tweeting about graphene: Out of 8,621 users, only 1,877 (21.8%) tweeted about graphene. This percentage appears relatively low at first, but recall that the networks of

⁵⁷ Refer to Mason (2010) for a discussion on saturation sampling in qualitative research designs.

these firms span a broad spectrum of users who tweet about topics and represent professions, regions, or industries, outside of the graphene R&D arena. In short, it is difficult to measure exactly where a technology-based innovation ecosystem on social media begins and where it ends. In this work, it begins with a set of SME graphene firms, but in another study, it could begin with graphene researchers or intermediaries or bloggers. The focus on graphene SMEs as the seeds for sampling conveys this research's fundamental interest in firm-centric innovation networks, which comprise part of the larger online innovation ecosystem on social media.

A related limitation is the mode of filtering applied to the combined friend network to run the necessary ERGMs to examine P1 and P2. Recall that given the broad spectrum of users across different topical communities, only those in the nanotechnology R&D community (C3) were used in the regression analysis. Moreover, the selected ERGM implementation in R could not run complex simulations with actor and edge continuous variables on large networks⁵⁸, so an additional filter was applied to remove users who had not tweeted about graphene (C3-945). In sum, the sampling and filtering approaches can be construed as a threat to external validity because the sample on which the propositions were explored may not reflect the entire population. If the set of active Twitter users interested in and tweeting about graphene spans relationships outside of this sample, and if the actor type composition and behavior of these users differs greatly from

⁵⁸ Technical issues were brought to the attention of the statnet development team, and although acknowledged, no timely corrections were released to the code base. In addition, the time required to complete a single ERGM estimation on these large networks – even without the continuous variables – exceeded hundreds of iterations and 24+ hours of continuous processing, only to produce model results that did not converge within say a maximum of 250 or even 500 iterations. This work was carried out on a high-performance 4 x 8-core Intel E5-4650 2.7GHz CPU (32 cores total) with 512GB memory. PNet, an alternative software implementation of ERGMs, did not fare much better in terms of model results and timeliness of estimation.

then the sample, the results presented here cannot be generalized any further than the scope of this research. In this way, the novelty of the research design may also be its greatest limitation.

As noted, the filtering approach was pursued as a direct consequence of problems related to the ERGM implementation in R. The ERGM is a statistical model that addresses network complexity through the mechanics of modeling both exogenous and endogenous network phenomenon. While a theoretically powerful tool, it has experienced slow adoption in applied empirical research due to two main reasons (Goodreau, 2007). First, conventional regression models reveal noteworthy findings in terms of testing causal relationships, regardless of whether the hypothesized effects are confirmed. With ERGMs, Markov chain Monte Carlo estimation approaches may lead to severe degeneracy such that finite point estimates of model parameters cannot be identified. In other words, parameter estimates cannot “settle” on specific values because the observed network cannot be easily reproduced given the ERGM specification. As a result, a model that does not converge cannot be used for interpretation as the resulting parameter estimates are essentially unfit for any type of inference. Second, if a model fails to converge, it provides no clear heuristics that allow the analyst to easily try related specifications that are more likely to reproduce the observed network. In this research, it took many months of trial-and-error to land on a model specification that converged. This experience portends that researchers not heavily vested into a specific research topic and method may find this time commitment intractable.

8.5.3. Usage Dynamism

People use Twitter in different ways, and this usage evolves over time. As the results indicate, users change their strategies for content sharing and networking as they learn how to use the platform, and the platform itself changes over time (e.g., through hashtags, a non-standard feature of the platform initially).

Twitter itself is a for-profit company in the highly dynamic Internet industry. Internet firms offer “platform technologies”, which spur complimentary innovations. On Twitter, this may materialize in novel, targeted advertising campaigns; in other cases, Twitter itself may introduce specific experiments or interventions to incent users to behave in some way. The evolving nature of the platform suggests that users will not use Twitter in the same way as in years past. This fact makes it difficult to interpret a certain behavior or measure as representing the same logical construct over time. For example, if Twitter were to introduce a feature where it automatically followed some users based on ascertained profile information, then this research’s conceptualization of the friend network as being a result of individual agency could not be defended. In a similar vein, Twitter’s method of recommending friends has evolved to provide more granular information; e.g., it is now possible to determine whether the recommendation to follow someone is based on other specific users’ actions. This increasing level of context likely sways users’ behaviors in ways that cannot be easily isolated or controlled in research design. In brief, the social media platform evolves in such a way as to conflate and intermingle endogenous factors that govern how people communicate and build network linkages.

Given the number of Twitter users in general and the ephemeral boundaries of the graphene innovation ecosystem as operationalized in this study, it is plausible that firm-centric networks will continue to grow over time. The dynamic nature of Twitter – that is, users joining a communication platform and discovering one another through following and friend networks – suggests that any cross-sectional snapshot of the network cannot capture the evolving nature of the graphene innovation ecosystem. I consider opportunities for longitudinal study in the next section on future work.

8.6. Future Work

The online innovation ecosystem consists of traditional institutional actors and other non-traditional participants; the ecosystem as revealed through social media shows permeable boundaries where experts, firms, media entities, universities, “unknown users”, etc. follow each other in ways that show the complexity of innovation activities in an evolving emerging technology. As suggested above, there are a number of opportunities for future work to extend this research’s scope. In particular, theoretical implications with respect to social contexts and narratives and social capital and information value would certainly benefit from confirmatory empirical testing. In this section, I focus on two additional research questions:

1. How does the innovation ecosystem on social media grow and/or contract over time to reflect changes in actor means and goals?
2. Can networks of topically adjacent communities portend future developments in an emerging technology?

8.6.1 Temporality

Introducing temporality into network studies reflects the importance of process-based analytic frameworks in both innovation and entrepreneurship research. For example, in innovation research, an invention is often considered as only the first step in the innovation process (Hounshell, 1980; Stinchcombe, 1990). Further work is needed to socialize that invention within an organization to build support for its successful deployment (Stinchcombe, 1990). In innovation networks, firms may engage externally across the value chain to increase supply chain efficiency, improve learning, and access complimentary assets (Ahuja, 2000; Autio, 1997; Rothwell, 1989; Teece, 1986). In brief, numerous studies have shown that a mix of deep and dense ties, as well as weak ties, offer distinct and synergistic advantages, and these innovation networks are dynamic and change in response to market conditions, regulatory regimes, and firms' strategic and tactical decisions.

In entrepreneurship research, scholars have defined process-based frameworks for explaining how new venture creation unfolds (Levie & Lichtenstein, 2010; Moroz & Hindle, 2012). Many of these frameworks implicitly or explicitly leverage social network theory and methods to motivate the research context and explain causal mechanisms. For example, in Sarasvathy's (2001) conceptualization of *effectuation*, entrepreneurial activity unfolds as a boundedly rational process wherein entrepreneurs assess available means and personal aspirations to determine a set of desirable (and plausible) events.⁵⁹ The entrepreneur is less likely to make decisions based on expected

⁵⁹ This view stands in contrast to causation processes which assume a reversal in cognition processes such that the entrepreneur first identifies a desirable event and then works backwards to procure or develop the necessary means to achieve that event.

return and is instead more likely to prioritize decisions based on affordable losses and acceptable risks. As a result, competition is less important in positioning the new venture than networks, which confer a broader set of means than the entrepreneur could otherwise devise alone.

In sum, these process-based views of entrepreneurship and innovation suggest nonlinear and dynamic network evolution to meet the strategic and tactical needs of ecosystem participants. The present research acknowledges the importance of this time-based perspective but does not directly evaluate how the ecosystem network evolves, at least not quantitatively; this cross-sectional orientation stems in part from the methodical limitations encountered with the ERGM method. However, future work should construct a time-series research design to account for changes in actor means and goals, possibly through the Temporal ERGM (TERGM) or Separable Temporal ERGM (STERM) approaches (Hanneke, Fu, & Xing, 2009; Krivitsky & Handcock, 2014).⁶⁰ It remains to be seen whether a set of deductive hypotheses can be developed and confirmed using these methods, but it should be possible to use egocentric data (see Krivitsky, 2012), such as those collected for the present study, to investigate the merit of informal propositions and to perform other exploratory analysis.

Time series data sets and research designs allow the investigator to more clearly identify causal mechanisms at play, instead of relying on potentially spurious correlations (Wooldridge, 2003). For the innovation ecosystem, this means being able to be better

⁶⁰ The TERGM extends the ERGM by accounting for temporal changes in a network. The STERGM separates the notion of tie incidence – i.e., the rate at which new ties are created over time – from that of duration – i.e., how long the tie lasts.

tease out how social media encourages ongoing exploration of resources, information sharing, and resource transfer to further outcomes of interest.

Table 8.2: A set of temporality-driven propositions for future work

Proposition	Explanation
At the micro level, ecosystem actors unfollow users who tweet about topics that diverge from their existing network's topical purview	As noted in this cross-sectional study, topical distance predicts a lower probability of a following tie. Ergo, it stands to reason that ecosystem users who “test trial” following relationships may unfollow whenever the new relationship fails to provide relevant information. Fischer and Reuber (2011) provide qualitative evidence substantiating this proposition. Relevance here may be modeled linearly to predict a simple negative relationship between topical distance vis-à-vis an actors' existing timeline of tweets (as curated through his network of contacts), or alternatively as an inverse U-shaped relationship. That is, a user is more likely to unfollow when topical distance is either very low or very high. The “sweet spot”, or the apex of the inverse-U reflects an area where cognitive burden may be lowest (see Gilsing et al., 2008).
At the macro and meso levels, ecosystem topical diversity grows with time	The qualitative findings show that many ecosystem users participate on social media to access diverse content. While not directly corroborated by the quantitative results here (which use a given tie as the outcome variable of interest), this proposition could examine whether diversity in content is more readily observed at the community level, and whether this diversity grows over time to reflect a broader array of technology, business, sociopolitical, and/or regional issues.
New entrants to the ecosystem attach to the network via (a) media entities and institutional actors and (b) other users with high numbers of followers	In science and technology studies, the Mathew Effect is a well-known phenomenon where highly cited papers (authors) attract additional citations faster than papers (authors) with fewer citations. The same philosophy holds true on social media. In addition, since the innovation ecosystem appears to be driven by information access (recall that most of the tweets in the graphene sample have URLs to various website content), it follows that entrants to the ecosystem will follow those actors offering the most relevant (and perhaps unbiased) news content, namely media entities and possibly institutional actors.
Social media based innovation ecosystems behave differently from non-innovation related social media networks	This proposition is unlike the three above in that it introduces a control from which to test the unique behavior of social media representations of innovation ecosystems vis-à-vis unrelated samples. Being able to identify the unique characteristics of the social media based innovation ecosystem will help further identify the value of social media to innovation studies, and to better extrapolate policy and management implications. While a cross-sectional research design could reveal some areas of divergence, differences over time could be more fruitful in terms of major findings and theory construction.

8.6.2. Science Maps

To understand any science map, one should first consider its standard mathematical definition in graph theoretic terms. A map, simply put, is a graph consisting of some number of vertices connected by some number of edges. A map of science, then, represents the relationships between fields, journals, subject areas/disciplines, authors, papers, etc. The relationships or edges may be defined by co-occurrences of citations, references of words, or co-authorships. Science maps are almost always visually represented in spatial terms as a network whereby nodes connect to one another by the edges, and they are particularly revealing because they offer a relational perspective of one field vis-à-vis another. This is important because “science seeks relatedness as much as it seeks truth, and the value of a research field is measured in large part by the contributions it makes and the clarifications it affords to adjacent fields” (De Bellis, 2009, pg. 167).

Science maps are sensitive to database biases and threshold cutoffs, and consequently they are simplified representations of reality (De Bellis, 2009; Klavans & Boyack, 2008). In addition, it is difficult to relate research outputs produced by a single organization or institution within a discipline to the rest of that discipline as a whole, at least from a content perspective (Rafols, Porter, & Leydesdorff, 2010). Two recent trends, computerization (i.e., automation) and the development of hybrid maps (i.e., combining content and citations), provide two attractive ways to produce more accurate and revealing science maps (De Bellis, 2009).

Here, I argue that social media can offer an additional source of data to produce science maps focusing on the interpretation and commercialization of inventions and

technologies. Future work extending this research could examine topical communities on social media to compare and contrast maps derived from social media with traditional ones produced with bibliometric data. Social media data do not capture the production of and knowledge flows across scientific fields, but to the extent that different communities exist around specialized science-based emerging technologies, it may be possible to show linkages across these communities. For example, one might expect that a nanomaterial-based ecosystem may be closely aligned to the optics community, which in turn may link to the medical community. In fact, this research shows the first relationship between nanotechnology and optics (refer to the follower network in Figure 5.8), but because the sample is limited to graphene firms' relations, a broader understanding cannot be gained. Herein lies the opportunity for future work.

The implications of producing science maps on Twitter or related social media could inform public policies concerning the funding of science and its uptake in the larger (non-publishing) ecosystem. These maps may hint at areas of interdisciplinary commercialization efforts not currently documented and understood by the academic literature. For example, social media content may expose key areas of lead user innovation.⁶¹ In addition, funders of science could assess social media maps to examine novel research grant opportunities: Scholars have argued that for science maps to be useful to policymakers, they need to provide interactive interfaces which respond to user queries and provide dynamic information (Klavans & Boyack, 2008). The scientist should internally validate the use of keywords used to obtain a (local) science dataset,

⁶¹ Lead user innovation takes place when users recombine know-how and apply it for product innovation. An isolated lead user innovation may not appear in social media data, but a sizeable community of lead users would.

and the user community should externally validate the map to gauge its utility. Social media data vary in real-time and are amenable to the dynamic querying and exploration that would make it a suitable source for interaction with various users, given the right human computer interface.

Social media science maps also could improve organizational intelligence. Rafols et al. (2010) present a science overlay method as a technique to evaluate the research contributions of an institution and/or organization. They do this by first constructing a complete science map based on relationships between categories, e.g., through co-citation relationships. Then, they overlay numbers of authors, documents, or citations etc. in each category on top of the nodes. The overlay, as a result, depicts those areas of inquiry that the organization or institution has contributed to in terms of publications, citations, etc. Thus, this method offers a mechanism for organizations or policymakers to make more informed decisions about entering into unknown, emerging, or foreign “cognitive spaces”, especially as those spaces relate to the research target’s other competencies (Rafols et al., 2010). With social media data, the overlay method may compare and contrast commercialization efforts and ecosystem orientations across different bodies to draw more wide ranging policy conclusions (for a similar policy argument, see Rafols, Leydesdorff, O’Hare, Nightingale, & Stirling, 2012).

Implementing social media based science maps would require an initial sample covering many types of seeds and then snowballing from there. The sampling objective would be to cover as many scientific areas as possible without capturing vast amounts of data unrelated to scientific domains. To accomplish this difficult task, an intelligent automated sampling mechanism could be used to cull the dataset in near-real time to

avoid non-relevant actors, and therefore to mitigate the amount of unnecessary snowballing. Regardless of whether this approach would classify actors as being in-scope through a rules-based engine or via machine learning classification algorithms, any sampling implementation would likely suffer from the same limitation as maps derived from bibliometric data: That is, all maps suffer from database biases and threshold cutoffs, and consequently they are simplified representations of reality. Still this alternative view could be a telling indicator of the broader innovation ecosystem's receptivity to a given emerging technology.

8.7. Concluding Thoughts

This work is among a growing set of research studies that examines online behavior on social media as applied to innovation networks and entrepreneurial contexts. According to the President's Council of Advisors on Science and Technology (PCAST) (2008), the innovation ecosystem is a “dynamic system of *interconnected* institutions, persons, and policies that are necessary to propel technological and economic development” [emphasis added] (pg. 1).

The findings reveal the *interconnected* nature of individual users with enterprises, institutions, and other stakeholder organizations, thus offering a micro-level view of traditionally unobservable phenomena: Communities organize around a limited number of topical themes, and information diversity exists across communities (i.e., by bridging structural holes). The online innovation ecosystem, then, can be compared to a bazaar (or conference) where there are few tangible costs to entry, and where everyone can share (or promote) ideas and opinions. Narratives unfold as users observe and potentially engage with various underlying social contexts, and to some extent, users that strategically

engage with the ecosystem attract resources that would not otherwise be available without their broader social media usage. For profit-seeking firms, this may mean modest sales lifts through a product or service sales pipeline transcending social media into other communication channels; for other users, reputational benefits, increased public support, enhanced awareness, and broader professional networks and opportunities constitute other positive outcomes of social media usage.

The communication-engagement model shows in simple terms that passive observation is not enough to derive benefits from social media; participation is a necessary but not sufficient condition for achieving tangible outcomes. That said, many users continue to maintain active accounts in the hope of achieving some set of benefits that currently are not realized. This indicates that social media in professionalized, S&T contexts represents varying levels of sophistication in its use and that networking and communication dynamics will continue to evolve as the platform and user base matures. Consequently, future work examining innovation ecosystem *dynamics* on social media will provide an especially fruitful (albeit methodologically complex) environment to capture endogenous and exogenous changes in behavior.

In aggregate, social media may have limited impact on a technology's development. In the graphene setting, many interviewees active in this R&D space found little value in using Twitter to shape the material's scientific trajectory. After all, users with an acumen for digesting scientific minutiae consult the journal literature to track the latest theoretical or laboratory developments. Instead, users participate in the graphene ecosystem to gain valuable social context and to follow users knowledgeable (or who appear to be knowledgeable) about ecosystem topics, including graphene R&D. At the

very least, the graphene online ecosystem embodies some degree of hype wherein users follow the technology because it is one of science's trendiest materials. Graphene's evolution through the hype cycle and through future stages of development will continue to offer a telling case study of the robustness and staying-power of online social media networks in the innovation ecosystem.

In conclusion, this research cannot point to the social media and innovation ecosystem construct as "doing" any one thing: I believe it does many things at once; i.e., with graphene it represents a hype technology with a diverse following; it consists of a set of passive observers who may or may not improve their awareness as a result of being on Twitter; it comprises another set of users who seek to co-create narratives and effectuate the transfer of resources; it is a forum for sharing news; and it most certainly dovetails into other related ecosystems and contexts which ostensibly have little to do with science and technology issues. What this research does show, however, is that the innovation ecosystem has an online footprint, and that there is a bright and growing horizon for future scholarship to trace and interrogate its constitutive nature and socio-technical and economic impacts.

APPENDIX

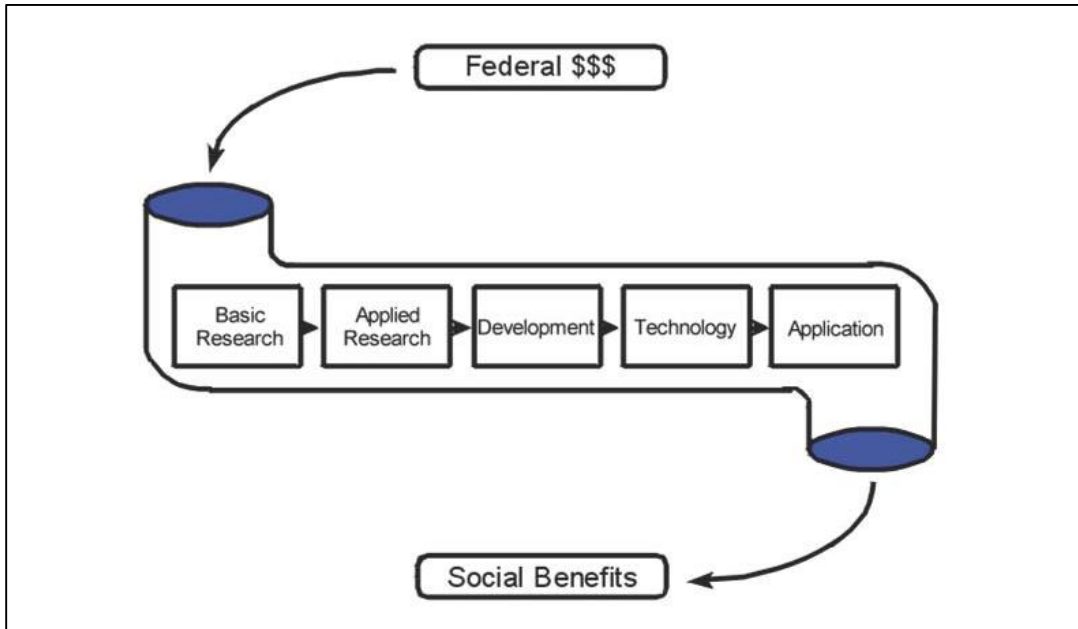


Figure A.1: The linear model of innovation. *Source:* <http://curryja.files.wordpress.com/2013/05/research-linear.jpg> (retrieved on September 12, 2013).

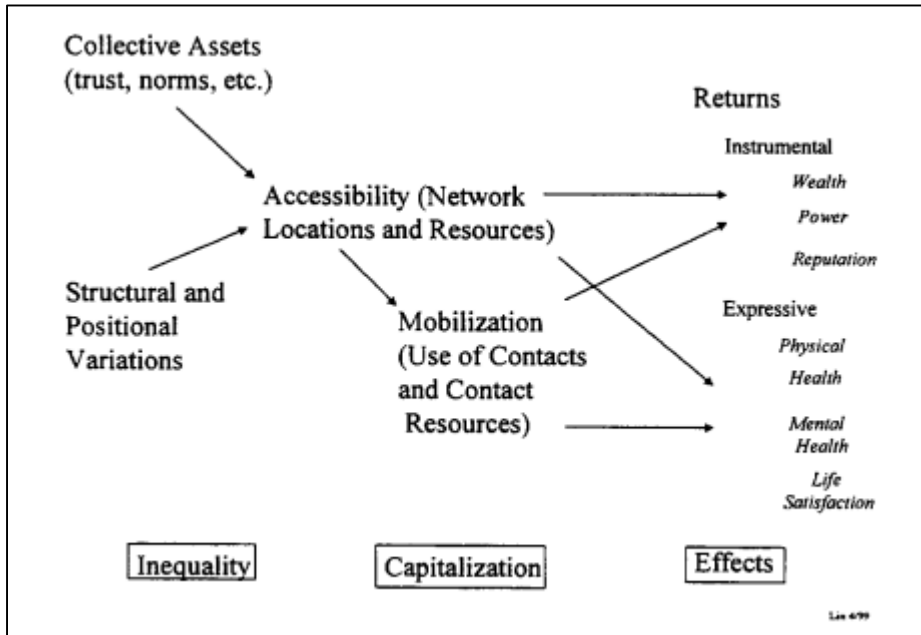


Figure A.2: Modeling a theory of social capital. Source: Lin (2001)

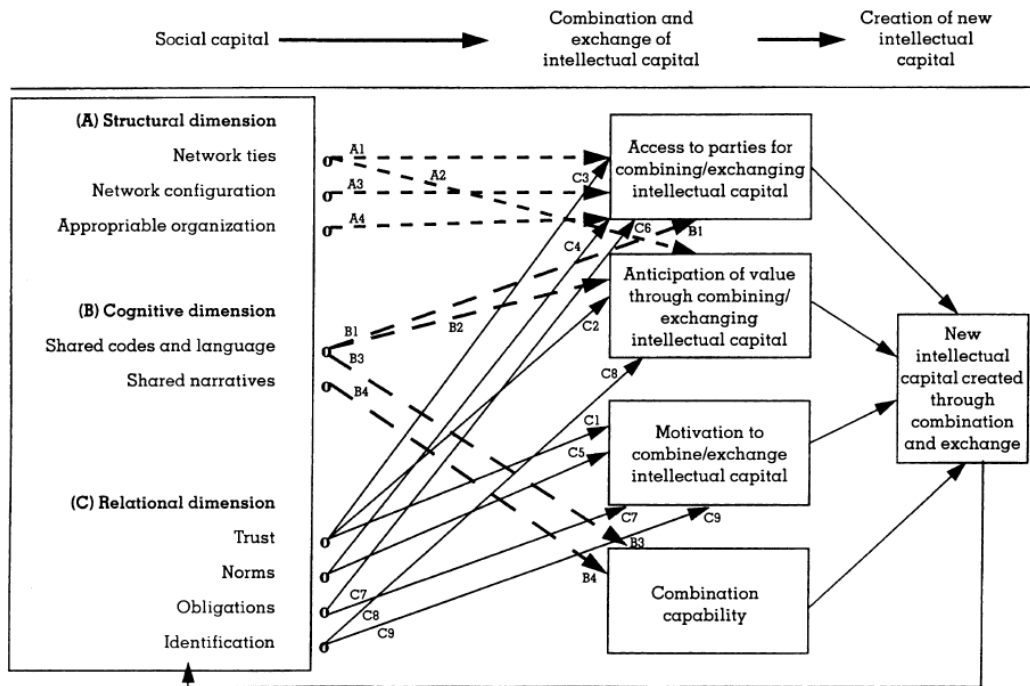


Figure A.3: Social capital and the development of intellectual capital.
Source: Nahapiet and Ghoshal (1998)

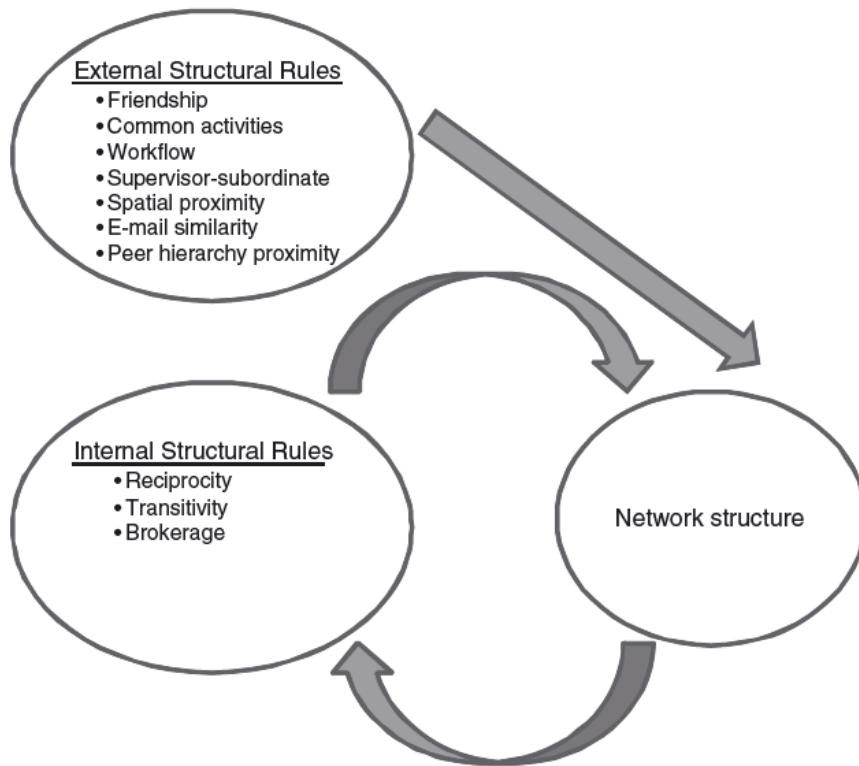


Figure A.4: A structural model of network emergence.
Source: Whitbred et al. (2011)

Table A.0.1: Coding Outsourcing Job Description

MEMO FOR OUTSOURCING OF ACTOR CODING WORK

Prerequisites for working on this project

1. Have a Twitter account and be interested in social media more broadly
2. Basic proficiency with Excel
3. Be meticulous as well as a quick decision maker
4. Be a fast learner and good communicator

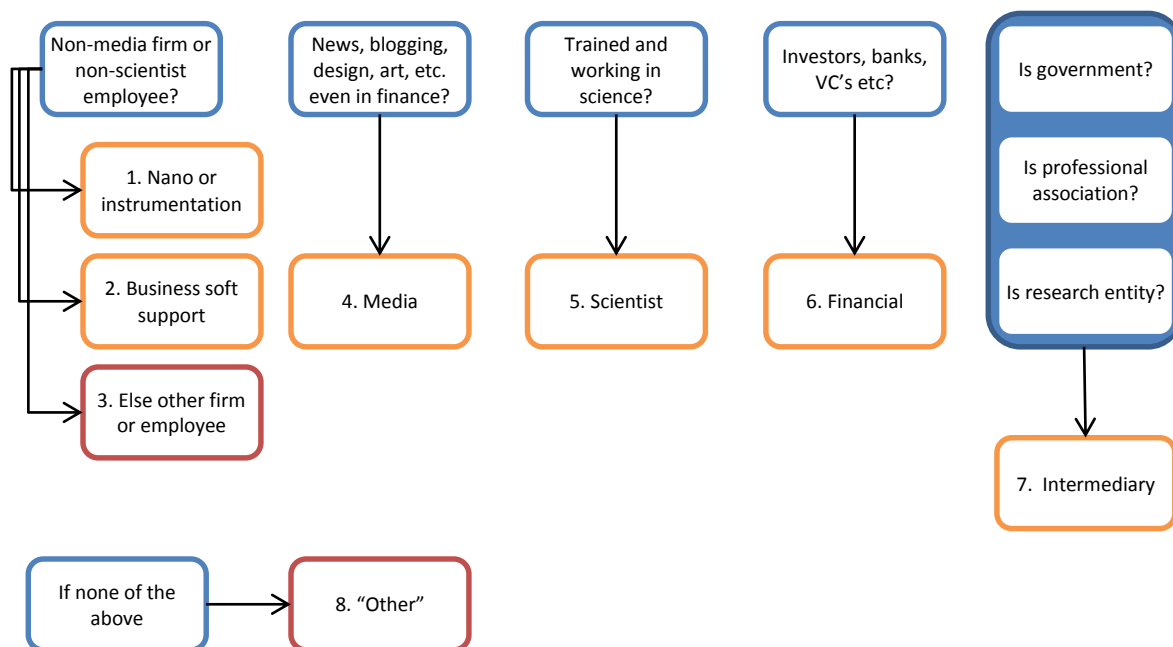
Project description and work task overview

The purpose of this research project is to understand how people use social media to commercialize emerging technologies. In particular, I'm studying a specific nanotechnology material science innovation called "graphene" on Twitter. Thus far, I've identified about 10-15 thousand users and need to code each user in one of eight different categories:

1. Nanotechnology firm
2. Firm offering strategy, consulting, online marketing and advertising, etc. soft support
3. Other firms (for example in health care, computers, or automotives)
4. Finance related firm
5. Media entity (newspaper, magazine, television station, media personality, author, writer, blogger, etc.)
6. Scientist
7. Intermediary, including universities, professional associations, trading platforms that bring together buyers and sellers, government agencies, and many non-profit organizations
8. Other; these are accounts which do not readily fit into the prior seven categories.

Some users are very easy to classify into one of the main categories. However, some users will not easily fit in one category. For example you may find a scientist who owns his own nanotechnology firm, or a professional association that also publishes a number of academic journals. In these cases, I ask that you make a best guess, mark the user as needing follow-up review, and move on.

For more information regarding the coding scheme, please see the slide in the PowerPoint file, "Actor coding typology and decision chart.pptx." The PowerPoint shows that if a person is a scientist, even if he works for a firm, then that is the most important attribute. In this case, mark the user as a scientist. However, if a firm is marketing person working for a chemicals laboratory not dealing in nanotechnology, then mark that user as "other firm". To determine whether a company operates in nanotechnology or graphene, you can search for either term in google with the website name; for example, in Google, "nanotechnology site:www.polymersolutions.com" shows that this company has enough relevant to hits to be considered as a nanotechnology company.



Data specifics

The next few paragraphs describe the data in Excel and the coding scheme. Each bullet point below corresponds to a column in Excel.

- A. screen_name. This is the user's Twitter screen_name. For example, if you come across the screen name, "floatbottle", you can add a "@" sign to search for that user on Twitter
- B. user description. This is the user's short description of the account. Sometimes its autobiographical, or about the company. Sometimes it's helpful, and sometimes it's not. This can change, so if you check Twitter, the description may not match what is in the spreadsheet. If this or the url is missing, you'll probably want to check the user's Twitter profile by searching for his or her screen_name.
- C. url. Twitter allows users to specify an url, which is visible just below the user description. If you copy the url into your browser, it can give you very helpful contextual information. For example, some urls link to LinkedIn, while others link to a blog or company website. By reviewing urls, you can make a more informed judgment regarding the appropriate user category. However, reviewing every url is time consuming.
- D. Please disregard columns D-J.
- K. Columns K-L contain the eight user categories mentioned above. Note that there can only be one '1' in these cells for each user; that is, the categories are mutually exclusive such that a user can only be a firm, or a scientist, or a media entity, etc. Even though there will likely be a value assigned to one of these categories, you may have to change it.
- S. check. If you are uncertain about a user, please insert an 'x' into this cell. I will review the coding later and make adjustments if needed.

- T. notes. Please include a short but helpful explanation of the user account or your coding choice, if needed. For example, if a scientist has a blog, I might insert 1 into the 'scientist' column in P but then indicate "also a blogger" in column T.
- U. seen. This field tracks when you've coded the user. Please insert your initials and the date in the field. (You can easily copy and paste the value for sets of records all at once). For example, "ska_3.28.14".

When you send back a spreadsheet, please add your initials and date to the versioning. This is just to make sure that I can keep track of everything.

Table A.0.2: Log-likelihood estimates

		beta				
Number of topics	alpha	0.01	0.025	0.05	0.1	
50	0.25	-9.861545342	-9.835262731	-9.819377457	-9.818944842	
	0.5	-9.928670203	-9.901281155	-9.891529561	-9.895492217	
	1	-10.09970927	-10.08809497	-10.06183314	-10.05372296	
	1.5	-10.21954406	-10.16275245	-10.17525451	-10.18110744	
	2	-10.29169392	-10.27632438	-10.23965015	-10.22828628	
100	0.25	-9.938833546	-9.919620816	-9.892036716	-9.88709755	
	0.5	-10.07369043	-10.03531296	-10.03541984	-10.02927335	
	1	-10.29044206	-10.255329	-10.22502416	-10.22497246	
	1.5	-10.38424514	-10.34667848	-10.33551547	-10.35457357	
	2	-10.48224887	-10.4449179	-10.44885808	-10.48278913	
150	0.25	-10.0101591	-9.98784447	-9.975040388	-9.96562734	
	0.5	-10.184892	-10.1509374	-10.15094785	-10.13425783	
	1	-10.36124976	-10.33173416	-10.31404152	-10.35863989	
	1.5	-10.49278668	-10.4624485	-10.47904695	-10.54247141	
	2	-10.63874301	-10.60222282	-10.63282826	-10.6832915	
200	0.25	-10.07085302	-10.03718725	-10.037084	-10.02242603	
	0.5	-10.26955006	-10.22951283	-10.21463362	-10.24432211	
	1	-10.44773131	-10.42204414	-10.43667642	-10.48638494	
	1.5	-10.58084957	-10.5866346	-10.63219095	-10.68268819	
	2	-10.7327061	-10.74083219	-10.77367241	-10.81547933	
250	0.25	-10.12657619	-10.09715736	-10.08203394	-10.09050107	
	0.5	-10.33372976	-10.28736618	-10.28211839	-10.33897257	
	1	-10.49298287	-10.48590935	-10.54157351	-10.60437982	
	1.5	-10.68265224	-10.67615415	-10.75408237	-10.78303576	
	2	-10.82773496	-10.86131993	-10.88756579	-10.91425124	

Source: Twitter, model trained on n= 161,141 tweets authored by members of C3-945; data collected in early 2014. *Notes:* Estimates produced after 1,000 iterations for varying number of topics and alpha and beta smoothing parameters. Green shading indicates higher values whereas orange and red shading indicates lower values.

Table A.0.3: Log-likelihood estimates

		beta				
Number of topics	alpha	0.01	0.025	0.05	0.10	
25	0.25	-7.130	-7.114	-7.104	-7.089	
	0.5	-7.104	-7.086	-7.142	-7.106	
	1.0	-7.145	-7.136	-7.124	-7.113	
	1.5	-7.161	-7.171	-7.147	-7.153	
50	0.25	-6.869	-6.865	-6.861	-6.836	
	0.5	-6.894	-6.871	-6.876	-6.836	
	1.0	-6.899	-6.897	-6.912	-6.881	
	1.5	-6.947	-6.952	-6.917	-6.891	
100	0.25	-6.655	-6.630	-6.635	-6.614	
	0.5	-6.669	-6.638	-6.649	-6.623	
	1.0	-6.739	-6.686	-6.661	-6.657	
	1.5	-6.793	-6.746	-6.722	-6.727	
150	0.25	-6.529	-6.502	-6.514	-6.491	
	0.5	-6.556	-6.538	-6.528	-6.494	
	1.0	-6.641	-6.609	-6.596	-6.583	
	1.5	-6.778	-6.719	-6.708	-6.722	

Source: Twitter, model trained on n=100,000 random graphene tweets collected in 2013 and early 2014. *Notes:* Estimates produced after 1,000 iterations for varying number of topics and alpha and beta smoothing parameters. Green shading indicates higher values whereas orange and red shading indicates lower values.

REFERENCES

- Aamodt, A., & Nygård, M. (1995). Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration. *Data & Knowledge Engineering*, 16(3), 191–222. doi:10.1016/0169-023X(95)00017-M
- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: a longitudinal study. *Administrative Science Quarterly*, 45(3), 425–455.
- Ahuja, G., Lampert, C. M., & Tandon, V. (2008). Moving Beyond Schumpeter: Management Research on the Determinants of Technological Innovation. *The Academy of Management Annals*, 2(1), 1–98. doi:10.1080/19416520802211446
- Amatucci, F. M., & Sohl, J. E. (2007). Business angels: Investment processes, outcomes, and current trends. In A. Zacharakis & S. Spinelli (Eds.), *Entrepreneurship: the engine of growth, Vol. 2* (pp. 87–108). Westport, CT: Praeger.
- Aral, S., & Van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, 117(1), 90–171. doi:10.1086/661238
- Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science (New York, N.Y.)*, 337(6092), 337–41. doi:10.1126/science.1215842
- Arora, A., & Gambardella, A. (1994). The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research Policy*, 23(5), 523–532. doi:10.1016/0048-7333(94)01003-X
- Arora, S. K., Youtie, J., Shapira, P., Gao, L., & Ma, T. (2013). Entry strategies in an emerging technology: a pilot web-based study of graphene firms. *Scientometrics*. doi:10.1007/s11192-013-0950-7
- Asheim, B. T., & Coenen, L. (2006). Contextualising Regional Innovation Systems in a Globalising Learning Economy: On Knowledge Bases and Institutional Frameworks. *The Journal of Technology Transfer*, 31(1), 163–173. doi:10.1007/s10961-005-5028-0
- Ashurst, C., Freer, A., Ekdahl, J., & Gibbons, C. (2012). Exploring IT-enabled innovation: A new paradigm? *International Journal of Information Management*, 32(4), 326–336. doi:10.1016/j.ijinfomgt.2012.05.006
- Aslesen, H. W., & Freel, M. (2012). Industrial Knowledge Bases as Drivers of Open Innovation? *Industry & Innovation*, 19(7), 563–584. doi:10.1080/13662716.2012.726807

- Audretsch, D. B., & Beckman, I. A. M. (2007). From Small Business to Entrepreneurship Policy. In D. B. Audretsch, I. Grilo, & A. R. Thurik (Eds.), *Handbook of Research on Entrepreneurship Policy* (pp. 36–53). Cheltenham, UK: Edward Elgar.
- Audretsch, D. B., Grilo, I., & Thurik, A. R. (2007). Explaining entrepreneurship and the role of policy: a framework. In D. B. Audretsch, I. Grilo, & A. R. Thurik (Eds.), *Handbook of Research on Entrepreneurship Policy* (pp. 1–17). Cheltenham, UK: Edward Elgar.
- Audretsch, D. B., Weigand, J., & Weigand, C. (2002). The Impact of the SBIR on Creating Entrepreneurial Behavior. *Economic Development Quarterly*, 16(1), 32–38. doi:10.1177/089124240201600104
- Auersald, P. E. (2007). The simple economics of technology entrepreneurship: market failure reconsidered. In D. B. Audretsch, I. Grilo, & A. R. Thurik (Eds.), *Handbook of Research on Entrepreneurship Policy* (pp. 18–35). Cheltenham, UK: Edward Elgar.
- Auerswald, P. E., & Branscomb, L. M. (2003). Valleys of Death and Darwinian Seas: Financing the Invention to Innovation Transition in the United States. *The Journal of Technology Transfer*, 28(3/4), 227–239. doi:10.1023/A:1024980525678
- Autio, E. (1997). New, technology-based firms in innovation networks symplectic and generative impacts. *Research Policy*, 26(3), 263–281. doi:10.1016/S0048-7333(96)00906-7
- Autio, E., Kanninen, S., & Gustafsson, R. (2008). First- and second-order additionality and learning outcomes in collaborative R&D programs. *Research Policy*, 37(1), 59–76. doi:10.1016/j.respol.2007.07.012
- Autio, E., Kenney, M., Mustar, P., Siegel, D., & Wright, M. (2014). Entrepreneurial innovation: The importance of context. *Research Policy*, 43, 1097–1108. doi:10.1016/j.respol.2014.01.015
- Babbie, E. R. (2004). *The practice of social research* (3rd ed.). Belmont, CA: Thomson/Wadsworth.
- Baird, C. H., & Parasnis, G. (2011). From social media to social customer relationship management. *Strategy & Leadership*, 39(5), 30–37. doi:10.1108/10878571111161507
- Baker, W. E. (1984). The Social Structure of a National Securities Market. *American Journal of Sociology*, 89(4), 775–811. Retrieved from <http://www.jstor.org/stable/info/2779252>

- Barabási, A. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509–512. doi:10.1126/science.286.5439.509
- Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99–120. doi:10.1177/014920639101700108
- Baron, R. A., & Markman, G. D. (2003). Beyond social capital: the role of entrepreneurs' social competence in their financial success. *Journal of Business Venturing*, 18(1), 41–60. doi:10.1016/S0883-9026(00)00069-0
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Bernstein, M. S., Marcus, A., Karger, D. R., & Miller, R. C. (2010). Enhancing directed content sharing on the web. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 971). New York, New York, USA: ACM Press. doi:10.1145/1753326.1753470
- Bichteler, J., & Ward, D. (1989). Information-Seeking Behavior of Geoscientists. *Special Libraries*, 80(3), 169–178.
- Blanchard, A. L., & Markus, M. L. (2007). Technology and Community Behavior in Online Environments. In C. Steinfield, B. T. Pentland, M. Ackerman, & N. Contractor (Eds.), *Communities and Technologies 2007* (pp. 323–350). London: Springer-Verlag.
- Blau, P. M. (1955). Consultation among colleagues. In *The dynamics of bureaucracy: A study of interpersonal relations in two government agencies*. Chicago: University of Chicago Press.
- Bleda, M., & del Río, P. (2013). The market failure and the systemic failure rationales in technological innovation systems. *Research Policy*, 42(5), 1039–1052. doi:10.1016/j.respol.2013.02.008
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77. doi:10.1145/2133806.2133826
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. doi:10.1088/1742-5468/2008/10/P10008
- Bonetta, L. (2009). Should you be tweeting? *Cell*, 139(3), 452–3. doi:10.1016/j.cell.2009.10.017
- Bonte, W. (2011). What do scientists think about commercialization activities? In D. B. Audretsch, O. Falck, S. Heblich, & A. Lederer (Eds.), *Handbook of research on innovation and entrepreneurship* (pp. 337–356). Cheltenham, UK: Edward Elgar.
- Bower, J. L., & Christensen, C. M. (1995). Disruptive Technologies: Catching the Wave. *Harvard Business Review*, 73(1), 43–53.
- Brass, D. J., Galaskiewicz, J., Greve, H. R., & Tsai, W. (2004). Taking stock of networks and organizations: a multilevel perspective. *Academy of Management Journal*, 47(6), 795–817.
- Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies “Engines of growth”? *Journal of Econometrics*, 65(1), 83–108. doi:10.1016/0304-4076(94)01598-T
- Brossard, D. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl, 14096–101. doi:10.1073/pnas.1212744110
- Brotman, S. N. (2014). Why the innovation conversation needs to be broadened beyond technology’s influence. *Tech Tank: Improving Technology Policy*. Retrieved July 4, 2015, from <http://www.brookings.edu/blogs/techtank/posts/2014/08/7-broader-definition-of-innovation>
- Brouwer, E. (1999). Innovative output, and a firm’s propensity to patent. An exploration of CIS micro data. *Research Policy*, 28(6), 615–624. doi:10.1016/S0048-7333(99)00003-7
- Brynjolfsson, E. (2010). The 4 ways IT is revolutionizing innovation: An interview with Erik Brynjolfsson by M.S. Hopkins. *Sloan Management Review*, 51(3), 51–56. Retrieved from <http://sloanreview.mit.edu/article/it-innovation-brynjolfsson-article/>
- Burt, R. S. (1992). *The social structure of competition. Networks and organizations: Structure, form, and action*. Boston: Harvard Business School Press.
- Burt, R. S. (2001). Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. S. Burt (Eds.), *Social Capital: Theory and Research* (pp. 31–56). New Brunswick, NJ: Aldine Transaction.

- Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2), 349–399. doi:10.1086/421787
- Burton, S., & Soboleva, A. (2011). Interactive or reactive? Marketing with Twitter. *Journal of Consumer Marketing*, 28(7), 491–499. doi:10.1108/07363761111181473
- Bush, V. (1945). *Science: The Endless Frontier*. Washington, DC: Government Printing Office. Retrieved from <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>
- Carpenter, S. (2010). A study of content diversity in online citizen journalism and online newspaper articles. *New Media & Society*, 12(7), 1064–1084. doi:10.1177/1461444809348772
- Case, D. O. (2012). *Looking for information: A survey of research on information seeking, needs and behavior* (3rd ed.). Bingley, UK: Emerald Group Publishing.
- Cha, M., Benevenuto, F., Haddadi, H., & Gummadi, K. (2012). The World of Connections and Information Flow in Twitter. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 42(4), 991–998. doi:10.1109/TSMCA.2012.2183359
- Chen, Y.-B., Liu, J. S., & Lin, P. (2013). Recent trend in graphene for optoelectronics. *Journal of Nanoparticle Research*, 15(2), 1454. doi:10.1007/s11051-013-1454-3
- Cheng, C. C. J., & Huizingh, E. K. R. E. (2014). When Is Open Innovation Beneficial? The Role of Strategic Orientation. *Journal of Product Innovation Management*, n/a–n/a. doi:10.1111/jpim.12148
- Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School Press.
- Chesbrough, H. W. (2006). Open innovation: A new paradigm for understanding industrial innovation. In H. W. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *Open innovation: Researching a new paradigm* (pp. 1–12). Oxford, UK: Oxford University Press.
- Chesbrough, H. W. (2010). Business Model Innovation: Opportunities and Barriers. *Long Range Planning*, 43(2-3), 354–363. doi:10.1016/j.lrp.2009.07.010
- Choi, S., Park, J., & Park, H. W. (2011). Using social media data to explore communication processes within South Korean online innovation communities. *Scientometrics*, 90(1), 43–56. doi:10.1007/s11192-011-0514-7
- Choo, C. W. (2006). *The knowing organization*. New York: Oxford University Press.
- Christopherson, S., & Clark, J. (2007a). Power in Firm Networks: What it Means for

- Regional Innovation Systems. *Regional Studies*, 41(9), 1223–1236.
doi:10.1080/00343400701543330
- Christopherson, S., & Clark, J. (2007b). *Remaking Regional Economies: Power, Labor, and Firm Strategies in the Knowledge Economy (Routledge Studies in Economic Geography)*. Routledge. Retrieved from <http://www.amazon.com/Remaking-Regional-Economies-Strategies-Knowledge/dp/0415357438>
- Clarysse, B., Wright, M., & Mustar, P. (2009). Behavioural additionality of R&D subsidies: A learning perspective. *Research Policy*, 38(10), 1517–1533.
doi:10.1016/j.respol.2009.09.003
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128.
doi:10.2307/2393553
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94 Supplem, S95–S120.
- Coleman, J. S. (1993). The Rational Reconstruction of Society: 1992 Presidential Address. *American Sociological Review*, 58(1), 1–15.
- Colliander, J., & Dahlen, M. (2011). Following the Fashionable Friend: The Power of Social Media - Weighing the Publicity Effectiveness Of Blogs versus Online Magazines. *Journal of Advertising Research*, 51(1), 313. doi:10.2501/JAR-51-1-313-320
- Consult, I., & Falk, R. (2009). *Does Europe change R&D-behaviour? Assessing the behavioral additionality of the Sixth Framework Programme*. Brussels. Retrieved from http://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/evaluation_studies_and_reports/evaluation_studies_and_reports_2009/assessing_the_behavioural_additionality_of_the_sixth_framework_programme.pdf
- Cooke, P. (2001). Regional Innovation Systems, Clusters, and the Knowledge Economy. *Industrial and Corporate Change*, 10(4), 945–974. doi:10.1093/icc/10.4.945
- Corsaro, D., Cantù, C., & Tunisini, A. (2012). Actors' Heterogeneity in Innovation Networks. *Industrial Marketing Management*, 41, 780–789.
doi:10.1016/j.indmarman.2012.06.005
- Cossins, D. (2014). Setting the Record Straight. *The Scientist*. Retrieved from <http://www.the-scientist.com/?articles.view/articleNo/41056/title/Setting-the-Record-Straight/>
- Council, N. R. (1997). *Industrial Research and Innovation Indicators: Report of a*

- Workshop*. (R. S. Cooper & S. A. Merrill, Eds.). The National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=5976
- Dahl, A., Lawrence, J., & Pierce, J. (2011). Building an innovation community. *Research-Technology Management*, 54(5), 19–27. doi:10.5437/08956308X5405006
- Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, 39(6), 699–709. doi:10.1016/j.respol.2010.01.013
- Dann, S. (2010, November 24). Twitter content classification. *First Monday*. Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/2745/2681>
- David, P. A., Hall, B. H., & Toole, A. A. (2000). Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy*, 29(4-5), 497–529. doi:10.1016/S0048-7333(99)00087-6
- De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: Scarecrow Press.
- De Carolis, D. M., & Saporito, P. (2006). Social Capital, Cognition, and Entrepreneurial Opportunities: A Theoretical Framework. *Entrepreneurship Theory and Practice*, 30(1), 41–56. doi:10.1111/j.1540-6520.2006.00109.x
- Desrochers, P. (2001). Local Diversity, Human Creativity, and Technological Innovation. *Growth and Change*, 32(3), 369–394. doi:10.1111/0017-4815.00164
- Doloreux, D., & Parto, S. (2005). Regional innovation systems: Current discourse and unresolved issues. *Technology in Society*, 27(2), 133–153. doi:10.1016/j.techsoc.2005.01.002
- EDA. (n.d.). Investment Priorities and Mission. *U.S. Economic Development Administration*. Retrieved September 12, 2013, from <http://www.eda.gov/investmentPriorities.htm>
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109–123. doi:10.1016/S0048-7333(99)00055-4
- Fagerberg, J. (2004). Innovation: A guide to the literature. In J. Fagerberg, D. C. Mowery, & R. R. Nelson (Eds.), *The Oxford Handbook of Innovation* (pp. 1–26). Oxford University Press.
- Felin, T., & Hesterly, W. S. (2007). The knowledge-based view, nested heterogeneity, and new value creation: Philosophical considerations on the locus of knowledge. *Academy of Management Review*, 32(1), 195–218. doi:10.5465/AMR.2007.23464020

- Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? *Journal of Business Venturing*, 26(1), 1–18. doi:10.1016/j.jbusvent.2010.09.002
- Freeman, C. (1987). *Technology and economic performance: Lessons from Japan*. London: Pinter.
- Freeman, C. (1991). Networks of innovators: A synthesis of research issues. *Research Policy*. doi:10.1016/0048-7333(91)90072-X
- Freeman, R. B. (2011). The economics of science and technology policy. In K. H. Fealing, J. I. Lane, J. H. Marburger III, & S. S. Shipp (Eds.), *Science of Science Policy* (pp. 85–103). Stanford University Press.
- Gallaughar, J. M., Auger, P., & BarNir, A. (2001). Revenue streams and digital content providers: an empirical investigation. *Information & Management*, 38(7), 473–485. doi:10.1016/S0378-7206(00)00083-5
- Ganley, D., & Lampe, C. (2009). The ties that bind: Social network principles in online communities. *Decision Support Systems*, 47(3), 266–274. doi:10.1016/j.dss.2009.02.013
- Garud, R., Gehman, J., & Giuliani, A. P. (2014). Contextualizing entrepreneurial innovation: A narrative perspective. *Research Policy*. doi:10.1016/j.respol.2014.04.015
- Geels, F. W., & Smit, W. A. (2000). Failed technology futures: pitfalls and lessons from a historical survey. *Futures*, 32(9-10), 867–885. doi:10.1016/S0016-3287(00)00036-7
- Geim, A. K., & Novoselov, K. S. (2007). The rise of graphene. *Nature Materials*, 6(3), 183–91. doi:10.1038/nmat1849
- Georghiou, L. (2002). Impact and additionality of innovation policy. In P. Boekholt (Ed.), *Innovation policy and sustainable development: can innovation incentives make a difference* (pp. 57–65). Brussels: IWT-Observatory. Retrieved from http://www.6cp.net/downloads/02brussels_review.pdf#page=57
- Giamanco, B., & Gregoire, K. (2012). NoTweet me, friend me, make me buy. *Harvard Business Review*, (July-August 2012). Retrieved from <https://hbr.org/2012/07/tweet-me-friend-me-make-me-buy>
- Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (p. 211). New York, New York, USA: ACM Press.

doi:10.1145/1518701.1518736

- Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & van den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy*, 37(10), 1717–1731. doi:10.1016/j.respol.2008.08.010
- Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology & Human Values*, 31(6), 639–667. doi:10.1177/0162243906291865
- Godin, B., & Gingras, Y. (2000). The place of universities in the system of knowledge production. *Research Policy*. doi:10.1016/S0048-7333(99)00065-7
- Gok, A., & Edler, J. (2012). The use of behavioural additionality evaluation in innovation policy making. *Research Evaluation*, 21(4), 306–318. doi:10.1093/reseval/rvs015
- Goodreau, S. M. (2007). Advances in Exponential Random Graph (p^*) Models Applied to a Large Social Network. *Social Networks*, 29(2), 231–248. doi:10.1016/j.socnet.2006.08.001
- Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a Feather, Or Friend of a Friend?: Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*, 46(1), 103–125. doi:10.1353/dem.0.0045
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3), 481–510. Retrieved from <http://www.jstor.org/stable/info/2780199>
- Hanneke, S., Fu, W., & Xing, E. (2009). Discrete Temporal Models of Social Networks. doi:10.1214/09-EJS548
- Hansen, M. T. (1999). The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly*, 44(1), 82. doi:10.2307/2667032
- Hautz, J., Hutter, K., Fuller, J., Matzler, K., & Rieger, M. (2010). How to Establish an Online Innovation Community? the Role of Users and Their Innovative Content. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1–11). IEEE. doi:10.1109/HICSS.2010.221
- Helpman, E., & Trajtenberg, M. (1994). *A time to sow and a time to reap: Growth based on general purpose technologies*. NBER Working Paper No. 4854.

- Hemminger, B. M., Lu, D., Vaughan, K. T. L., & Adams, S. J. (2007). Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, *58*(14), 2205–2225. doi:10.1002/asi.20686
- Hoang, H., & Antoncic, B. (2003). Network-based research in entrepreneurship. *Journal of Business Venturing*, *18*(2), 165–187. doi:10.1016/S0883-9026(02)00081-2
- Hobday, M. (2005). Firm-level Innovation Models: Perspectives on Research in Developed and Developing Countries. *Technology Analysis & Strategic Management*, *17*(2), 121–146. doi:10.1080/09537320500088666
- Hough, J. R., & White, M. A. (2003). Environmental dynamism and strategic decision-making rationality: an examination at the decision level. *Strategic Management Journal*, *24*(5), 481–489. doi:10.1002/smj.303
- Hounshell, D. A. (1980). Edison and the pure science ideal in 19th-century america. *Science (New York, N.Y.)*, *207*(4431), 612–7. doi:10.1126/science.207.4431.612
- Howells, J. (2006). Intermediation and the role of intermediaries in innovation. *Research Policy*, *35*(5), 715–728. doi:10.1016/j.respol.2006.03.005
- Huizingh, E. K. R. E. (2011). Open innovation: State of the art and future perspectives. *Technovation*, *31*(1), 2–9. doi:10.1016/j.technovation.2010.10.002
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.
- Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, *22*(4), 582–594. doi:10.1016/j.jcps.2012.03.006
- Jaffe, A. B. (2011). Analysis of public research, industrial R&D, and commercial innovation: Measurement issues underlying the science of science policy. In K. H. Fealing, J. I. Lane, J. H. Marburger III, & S. S. Shipp (Eds.), *Science of Science Policy* (pp. 183–192). Stanford University Press.
- Jones, B. D. (2003). Bounded Rationality and Political Science: Lessons from Public Administration and Public Policy. *Journal of Public Administration Research and Theory*, *13*(4), 395–412. doi:10.1093/jpart/mug028
- Jones, B. F. (2009). The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *Review of Economic Studies*, *76*(1), 283–317. doi:10.1111/j.1467-937X.2008.00531.x
- Kadushin, C. (2011). *Understanding Social Networks: Theories, Concepts, and Findings* (Kindle.). New York: Oxford University Press.

- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Katz, B., Vey, J. S., & Wanger, J. (2015). *One year after: Observations on the rise of innovation districts*. *Innovation Districts Series*. Washington, DC. Retrieved from <http://www.brookings.edu/research/papers/2015/06/24-one-year-innovation-districts-katz-vey-wagner>
- Katzy, B., Turgut, E., Holzmann, T., & Sailer, K. (2013). Innovation intermediaries: a process view on open innovation coordination. *Technology Analysis & Strategic Management*, 25(3), 295–309. doi:10.1080/09537325.2013.764982
- Kiesler, S., Kraut, R. E., Resnick, P., & Kittur, A. (2012). Regulating behavior in online communities: Evidence-based social design. In R. E. Kraut & P. Resnick (Eds.), *Building Successful Online Communities* (pp. 125–178). Cambridge, MA: The MIT Press.
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. doi:10.1016/j.bushor.2011.01.005
- Klavans, R., & Boyack, K. W. (2008). Thought leadership: A new indicator for national and institutional comparison. *Scientometrics*, 75(2), 239–250. doi:10.1007/s11192-007-1854-1
- Kobourov, S. G. (2012). Spring Embedders and Force Directed Graph Drawing Algorithms. *CoRR*, abs/1201.3.
- Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383–397.
- Kolodzy, J., Grant, A. E., DeMars, T. R., & Wilkinson, J. S. (2014). The Convergence Years. *Journalism & Mass Communication Educator*, 69(2), 197–205. doi:10.1177/1077695814531718
- Koskinen, J., & Daraganova, G. (2013). Exponential Random Graph Model Fundamentals. In *Exponential Random Graph Models for Social Networks*. New York: Cambridge University Press.
- Krivitsky, P. N. (2012). *Modeling of Dynamic Networks based on Egocentric Data with Durational Data* (No. 12-01). University Park, PA.
- Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 76, 29–

46. doi:10.1111/rssb.12014

- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web - WWW '10* (p. 591). New York, New York, USA: ACM Press.
doi:10.1145/1772690.1772751
- Leavey, J. (2013). *Social media and public policy: What is the evidence?* Retrieved from <http://www.alliance4usefulevidence.org/assets/Social-Media-and-Public-Policy.pdf>
- Lee, S., Park, G., Yoon, B., & Park, J. (2010). Open innovation in SMEs—An intermediated network model. *Research Policy*, 39(2), 290–300.
doi:10.1016/j.respol.2009.12.009
- Levie, J., & Lichtenstein, B. B. (2010). A Terminal Assessment of Stages Theory: Introducing a Dynamic States Approach to Entrepreneurship. *Entrepreneurship Theory and Practice*, 34(2), 317–350. doi:10.1111/j.1540-6520.2010.00377.x
- Leydesdorff, L. (2012). The Triple Helix, Quadruple Helix, ..., and an N-Tuple of Helices: Explanatory Models for Analyzing the Knowledge-Based Economy? *Journal of the Knowledge Economy*, 3, 25–35. doi:10.1007/s13132-011-0049-4
- Leydesdorff, L., & Meyer, M. (2006). Triple Helix indicators of knowledge-based innovation systems. *Research Policy*, 35(10), 1441–1449.
doi:10.1016/j.respol.2006.09.016
- Lichtenthaler, U. (2011). Open innovation: Past research, current debates, and future directions. *The Academy of Management Perspectives*, 25(1), 75–93.
- Lin, N. (2001). Building an network theory of social capital. In N. Lin, K. Cook, & R. S. Burt (Eds.), *Social Capital: Theory and Research* (p. 352). New Brunswick, NJ: Aldine Transaction.
- Loch, C. H., & Huberman, B. A. (1999). A Punctuated-Equilibrium Model of Technology Diffusion. *Management Science*, 45(2), 160–177.
doi:10.1287/mnsc.45.2.160
- Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, 38(2), 313–318. doi:10.1016/j.pubrev.2012.01.005
- Luker, K. (2008). *Salsa dancing into the social sciences: Research in an age of info-glut*. Cambridge, MA: Harvard University Press.
- Lundvall, B.-A. (2007). National Innovation Systems—Analytical Concept and Development Tool. *Industry & Innovation*. doi:10.1080/13662710601130863

- Lusher, D., & Robbins, G. (2013). Formation of social network structure. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential Random Graph Models for Social Networks* 2. New York: Cambridge University Press.
- Lynn, G., Morone, J. G., & Paulson, A. S. (1996). Marketing and Discontinuous Innovation: The Probe and Learn Process. *California Management Review*, 38(3), 8–37.
- Maassen, G. H., & Bakker, A. B. (2001). Suppressor variables in path models definitions and interpretations. *Sociological Methods & Research*, 30(2), 241–270.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614. doi:10.1146/annurev.psych.58.110405.085542
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83–104. doi:10.1037//1082-989X.7.1.83
- Mahoney, J., & Gertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis*, 14(3), 227–249. doi:10.1093/pan/mpj017
- Mandavilli, A. (2011). Peer review: Trial by Twitter. *Nature*, 469(7330), 286–7. doi:10.1038/469286a
- Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). *Disruptive technologies: Advances that will transform life, business, and the global economy*. San Francisco: McKinsey Global Institute.
- March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1), 71–87. doi:10.1287/orsc.2.1.71
- Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, 16, 435–463. Retrieved from <http://www.jstor.org/stable/info/2083277>
- Marwick, a. E., & Boyd, D. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi:10.1177/1461444810365313
- Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 11).
- Metcalf, S. (1995). The economic foundations of technology policy: equilibrium and evolutionary perspectives. In P. Stoneman (Ed.), *Handbook of the economics of innovation and technological change*. Oxford, UK: Wiley Blackwell.

- Michaelidou, N., Siamagka, N. T., & Christodoulides, G. (2011). Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands. *Industrial Marketing Management*, *40*(7), 1153–1159. doi:10.1016/j.indmarman.2011.09.009
- Michelino, F., Caputo, M., Cammarano, A., & Lamberti, E. (2014). Inbound and outbound open innovation: Organization and Performances. *Journal of Technology Management & Innovation*, *9*(3).
- Miller, G. (2011). Social scientists wade into the tweet stream. *Science (New York, N.Y.)*, *333*(6051), 1814–5. doi:10.1126/science.333.6051.1814
- Mitchelstein, E., & Boczkowski, P. J. (2009). Between tradition and change: A review of recent research on online news production. *Journalism*, *10*(5), 562–586. doi:10.1177/1464884909106533
- Monge, P. R., & Contractor, N. (2003). *Theories of Communication Networks*. New York: Oxford University Press.
- Moroz, P. W., & Hindle, K. (2012). Entrepreneurship as a Process: Toward Harmonizing Multiple Perspectives. *Entrepreneurship: Theory and Practice*, *36*, 781–818. doi:10.1111/j.1540-6520.2011.00452.x
- Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software*, *24*(4), 1–24. Retrieved from <http://www.jstatsoft.org/v24/i04>
- Muller, E., & Zenker, A. (2001). Business services as actors of knowledge transformation: the role of KIBS in regional and national innovation systems. *Research Policy*, *30*(9), 1501–1516. doi:10.1016/S0048-7333(01)00164-0
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266. doi:10.5465/AMR.1998.533225
- Narula, R. (2004). R&D collaboration by SMEs: new opportunities and limitations in the face of globalisation. *Technovation*, *24*(2), 153–161. doi:10.1016/S0166-4972(02)00045-7
- Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on Twitter. In *ACM WebSci '11*. Koblenz, Germany. Retrieved from http://www.websci11.org/fileadmin/websci/Papers/50_paper.pdf
- Neal, H. A., Smith, T., & McCormick, J. (2008). *Beyond Sputnik: U.S. Science Policy in*

- the 21st Century*. Ann Arbor, MI: University of Michigan Press.
- Nelson, R. R., & Rosenberg, N. (1993). Technical innovation and national systems. In R. Nelson (Ed.), *National innovation systems: a comparative analysis* (pp. 1–18). New York: Oxford University Press.
- Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.
- Newman, A. L. (2010). What You Want Depends on What You Know: Firm Preferences in an Information Age. *Comparative Political Studies*, 43(10), 1286–1312. doi:10.1177/0010414010369068
- Nordhaug, O. (1993). *Human capital in organizations: Competence, training, and learning*. Oslo, Norway: Scandinavian University Press.
- North, D., & Smallbone, D. (2000). The Innovativeness and Growth of Rural SMEs During the 1990s. *Regional Studies*, (2), 145–157. doi:10.1080/00343400050006069
- OECD. (1997). *National Innovation Systems*. Paris.
- Ooms, W., Bell, J., & Kok, R. A. W. (2015). Use of Social Media in Inbound Open Innovation: Building Capabilities for Absorptive Capacity. *Creativity and Innovation Management*, 24(1), 136–150. doi:10.1111/caim.12105
- Ott, I., Papilloud, C., Zülsdorf, T., & others. (2009). What drives innovation? Causes of and consequences for nanotechnologies. *Managing Global Transitions*, 7(1), 5–26.
- Owen-Smith, J., & Powell, W. W. (2004). Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organization Science*, 15(1), 5–21. doi:10.1287/orsc.1030.0054
- Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication*, 6(2), 181–201. doi:10.1177/1750481312437441
- Parida, V., Westerberg, M., & Frishammar, J. (2012). Inbound Open Innovation Activities in High-Tech SMEs: The Impact on Innovation Performance. *Journal of Small Business Management*, 50(2), 283–309. doi:10.1111/j.1540-627X.2012.00354.x
- Parise, S., Whelan, E., & Todd, S. (2015). How Twitter Users can Generate Better Ideas. *MIT Sloan Management Review*, (Summer 2015).
- Pear Analytics. (2009). *Twitter Study - August 2009*. Retrieved from <http://www.pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August->

2009.pdf

- Petty, J. S., & Gruber, M. (2011). "In pursuit of the real deal." *Journal of Business Venturing*, 26(2), 172–188. doi:10.1016/j.jbusvent.2009.07.002
- Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda. *Journal of Management*, 38(4), 1115–1166. doi:10.1177/0149206311432640
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, 493(7431), 159. doi:10.1038/493159a
- Polt, W., & Streicher, G. (2005). Trying to capture additionality in Framework Programme 5 - main findings. *Science and Public Policy*, 32(5), 367–373. doi:10.3152/147154305781779344
- Powell, W. W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12, 295–336.
- Powell, W. W., Owen-Smith, J., & Smith-Doerr, L. (2011). Sociology and the science of science policy. In K. H. Fealing, J. I. Lane, J. H. Marburger III, & S. S. Shipp (Eds.), *Science of Science Policy* (pp. 56–84). Stanford University Press.
- Powell, W. W., & Snellman, K. (2004). The Knowledge Economy. *Annual Review of Sociology*, 30(1), 199–220. doi:10.1146/annurev.soc.29.010202.100037
- Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4), 1132–1205. doi:10.1086/421508
- Prahalad, C. K., & Ramaswamy, V. (2003). The New Frontier of Experience Innovation. *MIT Sloan Management Review*, 44(4), 12–18.
- President's Council of Advisors on Science and Technology. (2008). *University-Private Sector Research Partnerships in the Innovation Ecosystem*. Retrieved from http://www.whitehouse.gov/files/documents/ostp/PCAST/past_research_partnership_report_BOOK.pdf
- Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–78.
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, 41(7), 1262–1282. doi:10.1016/j.respol.2012.03.015

- Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, *61*(9), 1871–1887. doi:10.1002/asi.21368
- Reich, E. S. (2011). Researchers tweet technical talk. *Nature*, *474*(7352), 431. doi:10.1038/474431a
- Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Random House LLC.
- Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*, *36*(1), 91–115. doi:10.1146/annurev.soc.34.040507.134743
- Robins, G., & Daraganova, G. (2013). Social selection, dyadic covariates, and geospatial effects. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential Random Graph Models for Social Networks*. Cambridge University Press.
- Robins, G., Lewis, J. M., & Wang, P. (2012). Statistical Network Analysis for Analyzing Policy Networks. *Policy Studies Journal*, *40*(3), 375–401. doi:10.1111/j.1541-0072.2012.00458.x
- Robins, G., & Lusher, D. (2013). Simplified account of an exponential random graph model as a statistical model. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential Random Graph Models for Social Networks*. New York: Cambridge University Press.
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, *29*(2), 173–191. doi:10.1016/j.socnet.2006.08.002
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.
- Rothaermel, F. T., Agung, S. D., & Jiang, L. (2007). University entrepreneurship: a taxonomy of the literature. *Industrial and Corporate Change*, *16*(4), 691–791. doi:10.1093/icc/dtm023
- Rothaermel, F. T., & Alexandre, M. T. (2009). Ambidexterity in Technology Sourcing: The Moderating Role of Absorptive Capacity. *Organization Science*, *20*(4), 759–780. doi:10.1287/orsc.1080.0404
- Rothwell, R. (1989). Small firms, innovation and industrial change. *Small Business Economics*, *1*(1), 51–64. doi:10.1007/BF00389916
- Runge, K. K., Yeo, S. K., Cacciatore, M., Scheufele, D. A., Brossard, D., Xenos, M., ... Su, L. Y.-F. (2013). Tweeting nano: how public discourses about nanotechnology

- develop in social media environments. *Journal of Nanoparticle Research*, 15(1), 1381. doi:10.1007/s11051-012-1381-8
- Saldanha, T. J. V., & Krishnan, M. S. (2012). Organizational Adoption of Web 2.0 Technologies: An Empirical Analysis. *Journal of Organizational Computing and Electronic Commerce*, 22(4), 301–333. doi:10.1080/10919392.2012.723585
- Samuel, A. (2009). Will social media consultants practice what they preach? *Havard Business Review*. Retrieved November 25, 2015, from <https://hbr.org/2009/10/social-media-hypocrisy>
- Sarasvathy, S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, 26(2), 243–263. doi:10.5465/AMR.2001.4378020
- Saxenian, A. (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press. Retrieved from <http://www.amazon.com/Regional-Advantage-Culture-Competition-Silicon/dp/0674753402>
- Schilke, O. (2013). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. *Strategic Management Journal*, n/a–n/a. doi:10.1002/smj.2099
- Scott, W. R., & Davis, G. F. (2006). *Organizations and Organizing: Rational, Natural and Open Systems Perspectives*. Upper Saddle River: Pearson Prentice Hall.
- Segal, M. (2009). Selling graphene by the ton. *Nature Nanotechnology*, 4(10), 612–4. doi:10.1038/nnano.2009.279
- Shandrow, K. L. (2013, January 6). 10 questions you must ask when hiring a social media consultant. *Entrepreneur*. Retrieved from <http://www.entrepreneur.com/article/225335>
- Shapira, P., & Youtie, J. (2008). Learning to Innovate: Building Regional Technology Development Learning Networks in Midsized Cities. *European Planning Studies*, 16(9), 1207–1228. doi:10.1080/09654310802401631
- Shapira, P., Youtie, J., & Arora, S. (2012). Early patterns of commercial activity in graphene. *Journal of Nanoparticle Research*, 14(4). doi:10.1007/s11051-012-0811-y
- Sharif, N. (2006). Emergence and development of the National Innovation Systems concept. *Research Policy*, 35, 745–766. doi:10.1016/j.respol.2006.04.001
- Sharma, N. K., Ghosh, S., Benevenuto, F., Ganguly, N., & Gummadi, K. (2012). Inferring who-is-who in the Twitter social network. In *Proceedings of the 2012*

- ACM workshop on Workshop on online social networks - WOSN '12* (p. 55). New York, New York, USA: ACM Press. doi:10.1145/2342549.2342563
- Shibayama, S., Walsh, J. P., & Baba, Y. (2012). Academic Entrepreneurship and Exchange of Scientific Resources: Material Transfer in Life and Materials Sciences in Japanese Universities. *American Sociological Review*, 77(5), 804–830. doi:10.1177/0003122412452874
- Shinn, T. (2002). The Triple Helix and New Production of Knowledge: Prepackaged Thinking on Science and Technology. *Social Studies of Science*. doi:10.1177/0306312702032004004
- Simmie, J. (2005). Critical surveys edited by Stephen Roper innovation and space: A critical review of the literature. *Regional Studies*, 39(6), 789–804. doi:10.1080/00343400500213671
- Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: MIT Press. Retrieved from <http://www.amazon.com/The-Sciences-Artificial-3rd-Edition/dp/0262691914>
- Simon, H. A. (1997). *Administrative Behavior* (4th ed.). New York, New York, USA: Free Press.
- Simoni, C., & Labory, S. (2007). The influence of social capital on entrepreneurial behavior. In M. Minniti (Ed.), *Entrepreneurship: The Engine of Growth, Vol. 1* (pp. 101–118). Westport, CT: Praeger.
- Smith, K. (2004). Measuring Innovation. In J. Fagerberg, D. C. Mowery, & R. R. Nelson (Eds.), *The Oxford Handbook of Innovation* (pp. 148–177). Oxford University Press.
- Smith, K. B., & Larimer, C. (2009). *The Public Policy Theory Primer*. Boulder, CO: Westview Press.
- Sproull, L., & Kiesler, S. (1986). Reducing Social Context Cues: Electronic Mail in Organizational Communication. *Management Science*, 32(11), 1492–1512. doi:10.1287/mnsc.32.11.1492
- Sternberg, R., & Tamásy, C. (1999). Munich as Germany's No. 1 High Technology Region: Empirical Evidence, Theoretical Explanations and the Role of Small Firm/Large Firm Relationships. *Regional Studies*, 33(4), 367–377. doi:10.1080/713693560
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum.

- Stinchcombe, A. L. (1990). *Information and Organizations*. Berkeley, CA: University of California Press. Retrieved from <http://www.amazon.com/Information-Organizations-California-Political-Economy/dp/0520067819>
- Stone, K. V. W. (2004). *From Widgets to Digits: Employment Regulation for the Changing Workplace*. Cambridge University Press. Retrieved from <http://www.amazon.com/From-Widgets-Digits-Employment-Regulation/dp/0521829100>
- Stoye, E. (2015, June 19). Graphene beyond the hype. *Chemistry World*. Retrieved from <http://www.rsc.org/chemistryworld/2015/06/graphene-beyond-hype>
- Street, C. T., & Cameron, A. (2007). External Relationships and the Small Business: A Review of Small Business Alliance and Network Research*. *Journal of Small Business Management*, 45(2), 239–266. doi:10.1111/j.1540-627X.2007.00211.x
- Sweet, L., & Strohm, B. (2006). Nanotechnology—Life-Cycle Risk Management. *Human and Ecological Risk Assessment: An International Journal*, 12(3), 528–551. doi:10.1080/10807030600561691
- Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6), 285–305. doi:10.1016/0048-7333(86)90027-2
- Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2-3), 172–194. doi:10.1016/j.lrp.2009.07.003
- Thomas, L. D. W., & Autio, E. (2012). Modeling the ecosystem: A meta-synthesis of ecosystem and related literatures. In *DRUID*. Copenhagen.
- Toubia, O., & Stephen, A. T. (2013). Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter? *Marketing Science*, 32, 368–392. doi:10.1287/mksc.2013.0773
- Uzzi, B. (1996). The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. *American Sociological Review*, 61(4), 674–698. doi:10.2307/2096399
- Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks : The Paradox of Embeddedness. *Administrative Science Quarterly*, 42(1), 35–67.
- Vaast, E., Davidson, E. J., & Mattson, T. (2013). Talking about Technology: The Emergence of a New Actor Category Through New Media. *MIS Quarterly*, 37(4), 1069 – 1092.
- van de Vrande, V., de Jong, J. P. J., Vanhaverbeke, W., & de Rochemont, M. (2009).

- Open innovation in SMEs: Trends, motives and management challenges. *Technovation*, 29(6-7), 423–437. doi:10.1016/j.technovation.2008.10.001
- van den Hoof, B., Huysman, M., & Agterberg, M. (2007). Embeddedness and Media Use in Networks of Practice. In C. Steinfield, B. T. Pentland, M. Ackerman, & N. Contractor (Eds.), *Communities and Technologies 2007* (pp. 371–394). London: Springer-Verlag.
- Van Noorden, R. (2011). Chemistry: The trials of new carbon. *Nature*, 469(7328), 14–6. doi:10.1038/469014a
- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512(7513), 126–129. doi:10.1038/512126a
- Vanhaverbeke, W. (2006). The internorganizational context of open innovation. In H. W. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *Open innovation: Researching a new paradigm* (pp. 205–219). Oxford, UK: Oxford University Press.
- Vohora, A., Wright, M., & Lockett, A. (2004). Critical junctures in the development of university high-tech spinout companies. *Research Policy*, 33(1), 147–175. doi:10.1016/S0048-7333(03)00107-0
- von Hippel, E. (1988). *Sources of innovation*. Oxford: Oxford University Press.
- von Hippel, E. (1994). “Sticky Information” and the Locus of Problem Solving: Implications for Innovation. *Management Science*, 40(4), 429–439. doi:10.1287/mnsc.40.4.429
- Wallner, T., & Menrad, M. (2011). Extending the innovation ecosystem framework. In *XXII ISPIM Conference*. Hamburg, Germany. Retrieved from http://www.researchgate.net/profile/Thomas_Wallner3/publication/268006642_Extending_the_Innovation_Ecosystem_Framework/links/546f5d4f0cf2d67fc0310f6f.pdf
- Wellman, B. (2001). Computer networks as social networks. *Science (New York, N.Y.)*, 293(5537), 2031–4. doi:10.1126/science.1065547
- Wessner, C. W. (2007). Government programs to encourage innovation by start-ups and SMEs: the role of US innovation awards. In D. B. Audretsch, I. Grilo, & A. R. Thuik (Eds.), *Handbook of Research on Entrepreneurship Policy* (pp. 172–185). Cheltenham, UK: Edward Elgar.
- West, J., & Lakhani, K. R. (2008). Getting Clear About Communities in Open Innovation. *Industry & Innovation*, 15(2), 223–231. doi:10.1080/13662710802033734
- Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of*

- Management Review*, 14(4), 490–495.
- Whitbred, R., Fonti, F., Steglich, C., & Contractor, N. (2011). From Microactions to Macrostructure and Back: A Structural Approach to the Evolution of Organizational Networks. *Human Communication Research*, 37(3), 404–433. doi:10.1111/j.1468-2958.2011.01404.x
- White, H. C. (1981). Where Do Markets Come From? *American Journal of Sociology*, 87(3), 517–547.
- Woelfle, M., Olliaro, P., & Todd, M. H. (2011). Open science is a research accelerator. *Nature Chemistry*. doi:10.1038/nchem.1149
- Wooldridge, J. M. (2003). *Introductory Econometrics* (2nd ed.). Mason, OH: South-Western.
- Xia, M., Huang, Y., Duan, W., & Whinston, A. B. (2007). Implicit Many-to-One Communication in Online Communities. In C. Steinfeld, B. T. Pentland, M. Ackerman, & N. Contractor (Eds.), *Communities and Technologies 2007* (pp. 265–274). London: Springer-Verlag.
- Yin, R. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Youtie, J., Iacopetta, M., & Graham, S. (2007). Assessing the nature of nanotechnology: can we uncover an emerging general purpose technology? *The Journal of Technology Transfer*, 33(3), 315–329. doi:10.1007/s10961-007-9030-6
- Youtie, J., Porter, A., Shapira, P., Tang, L., & Benn, T. (2011). The Use of Environmental, Health and Safety Research in Nanotechnology Research. *Journal of Nanoscience and Nanotechnology*, 11(1), 158–166. doi:10.1166/jnn.2011.3840
- Youtie, J., & Shapira, P. (2008). Building an innovation hub: A case study of the transformation of university roles in regional technological and economic development. *Research Policy*, 37(8), 1188–1204. doi:10.1016/j.respol.2008.04.012
- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493. doi:10.1002/asi.20508