Knowledge Flows and Influence in Online Social Networks: Proposing a Research Agenda

Charles M. Weber  
*Portland State University*

Nitin V. Mayande  
*Portland State University*

**Citation Details**

Knowledge Flows and Influence in Online Social Networks: Proposing a Research Agenda

Charles M. Weber, Nitin V. Mayande
Department of Engineering and Technology Management, Portland State University, Portland, Oregon, USA

Abstract — Online social networks, which have been defined as aggregated organizations that emerge from the Internet when people carry on public discussions, are increasingly becoming the vehicle of influence in social, political and economic discourse. Yet, despite its increasing importance, the nature of influence in online social networks is not really understood. Practitioners openly admit that they lack the experience to make sense of the phenomenon, and extant theory of influence in networks, which extrapolates from observations of the real world, is demonstrably inadequate when it comes to explaining influence online.

The paper introduces a novel approach to analyzing influence online, which is based on the premise that knowledge flows rather than connectivity or position determine loci and regions of influence. The authors propose an exploratory, longitudinal population study of 100 highly diverse online social networks. The study will 1) benchmark a set of metrics for influence in these networks to determine which metrics are best suited for measuring influence in a plethora of contexts; 2) characterize the nature and properties of knowledge flows within each network; and 3) determine how knowledge flows impact the (virtual) spatial and temporal distribution of influence within that network.

I. INTRODUCTION

Online social networks have been defined as aggregated organizations that emerge from the Internet when people carry on public discussions (Mayande [114], Preece [146], Rheingold [149], Schobeth & Schrott [154]). They are increasingly affecting the performance of the organizations that engage with them (Ayres [14], Chakrabarti & Berthon [54], Khammash & Griffiths [104], Longart [111]); they are also influencing social trends and political processes at an astonishing scale (Gelman et al. [87], Goggins and Petakovic [91], Silver [157]). For example, political campaigns are increasingly conducted online and in real time. In democratic countries, candidates for political office voice their reactions to current events on Twitter to mobilize their followers on short notice. Their rivals respond in kind to thwart these efforts. An inability to characterize how online social networks behave may consequently limit our understanding of how socio-political processes function within modern society (Aral, et al. [11]-[13]).

The pace at which influence in online social networks builds and dissipates can be truly breathtaking (Mayande [114]). For example, in societies that aspire democracy (e.g., the countries of the Arab Spring), revolutionary political campaigns form online communities that try to induce rapid transformational change, which the authorities cannot comprehend before decisive action has been taken. An online campaign could thus potentially disrupt social order before its intent is known, its motivations are understood, and it sources have been identified. For good or for ill, online social networks can evolve at a faster pace than the rate at which the real world can respond. And, unless we make serious efforts to understand them soon, online social networks may simply overwhelm us in the not-too-distant future.

The commercial impact of online social networks has been substantial. They have enabled firms and organizations to leverage the network value of business ecosystems (Afsarmanesh & Camarinha-Matos [3]) in activities such as marketing, customer service and product innovation (Bressler & Grantham [36]). Online social networks are at the core of many successful business models, and they are used to coordinate business and information exchanges (Feller, et al. [82]). They have also disrupted traditional marketing models, because, to an ever-increasing degree, they are composing a virtual domain in which societal trends are established (Deighton [67]). It is estimated that worldwide around 2.13 billion people will use online social network by the end of 2016, up from 1.4 billion in 2012 (Statista [159]). Millions of consumers are thus continuously involved in highly fluid conversations (Dodds, et al. [69]) in which market needs for products and services are articulated or even determined (Chakrabarti & Berthon [54]). Understanding the structure and behavior of online social networks may consequently constitute a crucial source of competitive advantage in many domains of the global economy.

Unfortunately, practicing firms that are engaging with online social networks neither have a reliable theory nor sufficient practical experience to make sense of the phenomenon (Aral, et al. [11], Li & Bernoff [109], Wiertz, et al. [193]). Extant theory in particular is based on observations of the real world, and may thus not apply to online social networks (Mayande [114]). Practicing firms may thus be misallocating a large amount of resources, simply because they do not know how the online social networks with which they interact are organized and how these networks behave (Edwards [73] [74]).

In depth theoretical understanding of the mechanisms that drive influence in online social networks will help us explain many of the economic and socio-political phenomena that we observe in the modern world. It should also enable managers in real-world organizations design routines, structures, processes and practices that help them develop radically
innovative products and provide dramatically improved services, both in the private and the public sector. The primary motivation for the research agenda described in this paper is to make a significant contribution toward gaining such an understanding.

II. ONLINE SOCIAL NETWORKS: THE STATE OF THE ART

A. Online Marketing

The performance of organizations that develop new products or provide new services—be they public or for profit—increasingly depends upon how these organizations market these products and services in conversations that take place online. Organizations that engage in such ‘online marketing’ not only need to pay attention to these conversations (Chakrabarti and Berthon [54]); they must also try to become a part of these conversations, in order to shape them. When the conversations are positive, they can lead to free advertising and better brand recognition (Longart [111]). However, when the conversations are negative, they can do irreparable financial damage (Ayres [14], Khammash and Griffiths [104]). Online conversations can therefore make or break a product or a service.

Today’s marketers are responding to the increasing importance of online social networks by spending billions of dollars in digital marketing (Ng and Vranka [133]). They are reallocating their marketing resources to specifically target users of the highly popular networking platforms Facebook and Twitter, where the majority of online conversations about products and services take place (HBRAS [96]). However, the outcomes of these efforts have been disappointing (Edwards [73], Rusli and Eavis [151], Terlep, et al. [165]), primarily because companies deployed traditional approaches to marketing, which rely on broadcasting information that is passively consumed (Anderson [10]). Instead, advertising via social media requires users of online social networks to deliberately spread the information that they receive through word of mouth (Hodas and Lerman [100]), an approach that is demonstrably more efficient and effective than merely broadcasting information (Wolf and Scott [195]). This implies that traditional Internet marketing paradigms and processes are being upended by swiftly evolving social platforms and technology (Deighton [67]), and that billions of dollars in marketing resources have been misallocated (Edwards [74]).

With increased spending on social media, businesses are feeling the pressure to gain new insights into customer behavior. They need to know who the online influencers are and how they exert their influence (Lindsay et al. [110]). Success in marketing though online social media critically depends upon understanding the virtual community that may have a potential interest in your product or service and by identifying the key influencers that will spread your marketing message (Lindsay et al. [110]). They require analytics to transform enormous volumes of data into actionable strategies (Halavais [94]), for which companies are willing to pay large amounts of money. A report by the research firm Gartner projects that companies will spend in excess of $4 billion on analytics in 2016, and the trend is up (Colombus, [62]).

Many firms that engage in social media analytics (e.g., Klout, Kred, PeerIndex, and Trackr) have tried to analyze online social networks by finding the individuals that have the most friends and followers or generate the most output (Hurley [101]). This approach has not been particularly successful (Cha, et al. [52] [53]). Evidently, those who have the most connections or generate the most activity online are not the true influencers in social media (Cha, et al., [52] [53]), and whatever influence they have is ephemeral (Chen, et al. [55]). Instead, people appear to consume information from people they know and from people they trust (Wolf & Scott [195]), just as they do in the real world (Rogers [150]).

B. Network Flows, Network Structure, Network Phenomena

Many of the approaches that practitioners of social network analytics have deployed are grounded in theory that was developed almost entirely from observing social networks in the real world (e.g., Bailey [15], Luhmann [112], Miller [122], Parson, [138]). For example, practitioners track the deliberate propagation of information, through word of mouth, from one user to another (Granovetter [92], Rogers [150], Tichy, et al. [166]). This method of information transfer is henceforth referred to as network flows.

Social scientists have long understood the importance of network flows in spreading information (Granovetter [92]) and in the diffusion of innovations (Rogers [150]) in real-world social networks. All network flows in the real world take place between the seeker of information and the source of information, and all network flows transpire within existing social relationships (Bristor [37], Duhan et al. [72], Money et al. [124]). Interactions only happen between people who have social relationships (Burt [39], Burt & Doreian [42]). Thus an individual’s relationship network and his/her interaction network are considered to be one and the same (Burt [39]). Therefore, the structure of an individual’s relationship network or interaction network is henceforth defined as network structure.

In extant theory on social networks, network structure defines the boundaries of communities (Bailey [15], Luhmann [112], Miller [122], Parson [138]). For example, in living systems theory (Miller [122]), a system is defined as a set of interacting units and the relationships among them. The boundaries of these interacting units are determined by the processes through which these units get organized. These units are organized hierarchically. For example, two or more people and their relationships comprise a group; communities consist of two or more groups and two or more communities comprise a society. There are comparatively few barriers to information transfer within units than there are between the units. Therefore, the boundaries between units (e.g., groups, communities, societies) constrain network flows between the units (Carlile [44] [45]).
Within communities, network structure guides the network flows (Bailey [15], Luhmann [112], Parson [138]), and network flows give rise to network phenomena such as social capital (Bourdieu [34], Burt [40] [41], Coleman [61], Putnam [148]), social behavior (Allen [8], Burt [38], Granovetter [92]), economic benefit (Allen [8], Bourdieu [34], Burt [38] [40], Cartwright [47], Coleman [61], Granovetter [92]) and social influence (Cartwright [47], March [113], Simon [158]), the focus of the proposed research. Social influence in real-world networks occurs when an actor adapts his/her behavior to the behavior of other actors in the community (Cartwright [47], March [113], Simon [158]). A precondition for social influence is the availability of information, through network flows, about the other actors (Leenders [107]). The scope of the network flows within all real-world networks is constrained by factors such as connectivity (the number of actors to which an individual is connected) (Allen [8], Burt [38] [40], Granovetter [92]) and physical distance between the actors in the network (Allen [8]). Therefore, an individual’s influence in a real-world network depends upon the individual’s connectivity, his/her access to an individual with high connectivity or a combination of both.

C. Social Networks: Real-World versus Online

Online social networks differ from real world social networks in a variety of ways. First and foremost, online social networks tend to be larger than the social networks that have been studied in the real world. Known real-world social networks tend to consist of hundreds or thousands of people (e.g., Burt [39], Granovetter [92], Rogers [150], Tichy et al. [166]); online network may contain hundreds of thousands or millions (Dodds et al. [69] [70], Mislove, et al. [123]). Networks of such different scale could thus behave differently; some social processes may transpire in very large but not in comparatively small networks, and conversely. Social theories that were developed from observing real-world networks may thus not necessarily apply to online social networks.

Secondly, the ability to conduct searches in online social networks (Adamic and Adar [2], Watts et al. [173]) makes the network structure and the network flows, which result from the interaction that follows that search, highly dynamic (Dodds, et al. [69]). Real world constraints such as connectedness and distance may thus not have any significant impact on how these networks behave (Borgatti, et al. [29]-[33]). Instead, behavior may be most affected by topological organization of network structure (e.g., “scale free” (Barabasi, et al. [5] [6] [18] [19]), “assortativity” (Newman, et al. [132]) and “small world” (Watts and Strogatz [174]) or by various attributes of network flows (e.g., paths, geodesics) (Borgatti, et al. [29] [30]), which extant theory does not really consider and prior empirical studies have not explored.

As a consequence, network flows in online social networks cannot all be attributed to social relationships (Pei, et al. [139]). We know from observation of practicing firms (Wierz et al. [193]) that online social networks are an emergent phenomenon (in the sense of Drazin and Sandelands [71] [152]), and that network flows can be generated by ad hoc interactions. For example, the DARPA Network Challenge successfully tested the ability of online social networks to mobilize massive ad hoc teams to solve problems (Greeneemeier [93]), suggesting that an individual’s online social network and his/her online interaction network are not one and the same thing. We also know from observing hashtag communities that people in online social networks may interact virtually with people with whom they share a common interest; that online social networks and network flows can be ephemeral; and that they can disappear on short notice, as the common interest of the community dissipates (Weng et al. [192]). Extant theory of social networks, which assumes that strong bonds cause or enable network phenomena, may therefore not apply to online social networks.

Due to the emergent and dynamic nature of online social networks, the relationship between network structure, network flows and the resulting network phenomena in these networks is not very well understood. Recent research on network structure (Centola [51], Chomutare et al. [57], Sasidharan et al. [153]), network flow (Aral & Walker [12], Burt et al. [43], Dellarocas et al. [68], Hodas & Lerman [100]) and network phenomena (Aral & Walker [13], Khammash & Griffiths [104], Muchnik et al. [126] [127], Pei et al. [139]) focuses on these individual categories. However, studies that characterize the mechanisms through which network structure, network flows and network phenomena collectively emerge and operate are woefully lacking (Aral et al. [11]). We cannot even identify the loci of influence within an online social network reliably. Thus we are unable to explain how and why online social networks respond to a marketing message. To date, we do not know how online social networks form, how they get organized and how they evolve. As practitioners concede (Li & Bernoff [109]), firms that are considering engaging in online social networks have neither a reliable theory nor sufficient practical experience to manage these networks effectively. Even companies that are very adroit at marketing via online social networks have experienced unintended consequences when they attempted to direct and control social networks (Wierz et al. [193]). Using online social networks deliberately may consequently turn out to be challenging.

D. Pilot Study

The authors of this paper and one of their colleagues have conducted exploratory research, in the hope of enhancing the general understanding of the nature and behavior of online social networks (Mayande, et al. [114]-[119]). This research, which focused on interplay between network flows, network structure and the network phenomenon of influence in online marketing, acted as a pilot study for the proposed research agenda. It has led to the following insights, upon which the proposed research will expand.
1. The pilot study confirmed what many practitioners in social media analytics have claimed—there is no significant correlation between the number of connections that a member of an online social network has, the amount of activity he/she generates and the amount of influence he/she exerts.

2. The size and degree of activity of online communities that discuss product lines are not necessarily correlated to the popularity of the product lines that they discuss.

3. Network structure and network flow definitely impact network phenomena and each other. However, their impact cannot be taken for granted because network structure and network flow can change dramatically from minute to minute. Network phenomena, by extension, can do likewise.

4. The nature of influence within a social network cannot be understood by just analyzing an undirected or even a directed network. Influence may involve structuration (Barley and Tolbert [20], Giddens [89], Goggin & Petakovic [91], Orlikowski [137]). A person who influences how information is consumed may not necessarily influence how it is propagated, and conversely. By analyzing the consumption and the propagation of information across an online social network, it is possible to deduce the behavioral traits of individuals within the network and identify its key influencers.

5. Scale matters. It looks as if very large social networks are driven by processes and social phenomena to which extant social network theory does not apply.

6. Current measures of influence are not particularly effective. Centrality measures derived from graph theory (e.g., Freeman [84] [85]) do not really measure influence. Eigenvector centrality (Bonacich [27] [28]) measures influence as a function of the spread of information but not as a function of the speed of information spread.

7. The evolution of online social networks is clearly path dependent in the sense of Suárez & Utterback [170]. Every online social network has a unique history, context and knowledge base.

8. Evolutionary processes can unfold much more rapidly online than they do in the real world. They may even transpire at a rate that is faster than the rate that we can learn about them.

Despite yielding valuable insights, the research described above exhibits a significant limitation—it measures the impact of network flows, which are information flows rather than knowledge flows. However, knowledge is more than information. It is information that is sufficiently certain (Shannon & Weaver [156]) and sufficiently contextualized to enable human action (Stehr [160]). Contextualized knowledge flows should thus impact network structure and the network phenomenon of influence much more than non-contextualized network flows do (Goggin & Petakovic [91], Nonaka, et al. [135] [136]). Repeating the pilot study in multiple contexts and including variables that measure knowledge-related phenomena would consequently yield insights into the organizational behavior of online social networks that are much more significant and generalizable than those that have emerged from the pilot study.

III. KNOWLEDGE ABOUT TECHNOLOGICAL KNOWLEDGE

Recent studies of organizational learning, knowledge and intellectual capital in high technology organizations have yielded the following potentially transformative findings.

1. Continuous improvement at the subsystem level can induce radical improvement in performance at the organization level without generating an upheaval of an organization’s routines, processes, structure and practices [177] [178]. The date for a surge in performance can be planned years in advance and executed according to plan in a timely manner (Weber [179], Weber & Yang [190]). In other words, radical improvement in performance can occur without a disruption of the ‘deep structure’ (as defined by Gersick [88], Prigogine & Stengers [147] and Tushman & Anderson [168]).

2. Interplay between innovative activities that transpire at the subsystem-level of organizations and those take place in their extended value network contribute significantly to radical improvement in the performance of the organization and its value network. The keystone firms of some business ecosystems (Iansiti & Levien [102]) entrain (e.g., Ancona & Chong [9]) their subsystems and their value networks to deliver a cascade of revolutions in organizational performance that transpire in a timely manner (Gabella & Weber [86], Yang, et al. [197]). In other words, a ‘broad structure’ governs timely revolutions in organizational performance.

3. The ‘broad structure’ of the semiconductor manufacturing ecosystem consists of hundreds of high tech firms, their suppliers, the suppliers of these suppliers and industry trade organizations, who synchronize their innovative activities, even if they are not aware of each other’s existence (Yang, et al. [197]). A significant portion the global economy may thus march to the drumbeat of the semiconductor manufacturing ecosystem’s keystone firms.

4. Timely revolutions in organizational performance critically depend upon how a business ecosystem manages the flow of its technological knowledge, which Bohn [26, p. 62] defines as “understanding the effect of input variables on the output.” Bohn notes that technological knowledge goes from being tacit, which is hard to encode or express verbally, to being explicit, which is easy to encode or express verbally, as an industrial process matures. Technological knowledge...
flows more readily with increasing process maturity—it becomes easier to transfer (Szulanski, et al. [162] [163]), co-create (Nonaka, et al. [135] [136]), or co-transform (Carlile [44] [45]). The cost of knowledge transfer, knowledge co-creation, and knowledge co-transformation, also known as the stickiness of knowledge (von Hippel [171]), drops accordingly, and the industrial process becomes easier to control (Bohn [26]). A recent study also suggests that the keystone of a synchronized ecosystem consist of multiple firms, which sustain competitive advantage by controlling the rate at which technological knowledge is converted from tacit to explicit (Yang, et al. [197]).

Among the unforeseen positive consequences of recent research of high technology organizations are novel insights into the nature of technological knowledge. These include identifying and observed the following properties of technological knowledge.

**Knowledge Impedance.** One of the authors has defined knowledge impedance as “the degree of difficulty with which a particular type of knowledge is transferred between two or more entities, co-created by two or more entities, or transformed by two or more entities” (Weber & Yang [189]). These entities can be individuals, groups, organizations or firms, which can be located in different regions or countries (Allen, et al. [7], Espinosa, et al., [76]-[79], Evaristo, et al. [80], Farshcian [80], Hinds & Mortensen [99] [125], Sengupta, et al. [155], Zolin, et al. [203]). According to this definition, impedance to the flow of knowledge between two entities decreases as an industrial process matures. Knowledge impedance may even drop to the point where a “knowledge short circuit”—an inadvertent transfer of knowledge with low impedance from one entity to another—can occur. Such mishaps can result in highly adverse consequences for the competitive position of the transmitting entity, and they can threaten the stability of global business ecosystems (Yang, et al. [197]).

**Selective Absorption.** An organizational entity’s capacity to absorb new knowledge (Cohen & Levinthal [60]) does not just depend on the presence of prior related knowledge within the entity (Todorova & Durisin [167], Zahra & George [198]). Research by the authors and colleagues suggests that absorptive capacity could also be a function of the source of external knowledge, the knowledge pathway into the entity, the source of complementary or substitutive knowledge that resides within the entity, and the mission to which the knowledge contributes (Bresman [35], Nemanich, et al., [131], Ploykitikoon [142], Ploykitikoon & Weber [143] [145]). This insight leads to the speculation that practicing managers can enhance the competitiveness of their organizations through knowledge filtration (Yang, et al. [196]). They can modulate their organizations’ capacity to absorb external knowledge selectively by pursuing practices that let specialized knowledge flow into entities where it is particularly useful. This focused approach can enhance the performance of specific organizational entities very effectively (Ploykitikoon & Weber [143]). What is not yet known is whether organizations can engage in selective transmission of knowledge to resolve Kogut and Zander’s paradox [105]. Can managers modulate the impedance of knowledge selectively and in a timely manner (Szulanski [163])? Will this constrain imitability prior to product release (e.g., prevent a knowledge short circuit)? Will modulation of knowledge impedance maximize the efficiency and effectiveness of knowledge dispersion thereafter, in order to enhance marketing communication and accelerate the transfer of production knowledge to the organizational entities that will manufacture the product?

**Knowledge Vacuum.** Aristotle’s dictum scio nescio (I know that I don’t know) may serve as a source of enhanced performance for the innovating organization. A preliminary study (Weber, Hasenauer & Mayande [183]) indicates that awareness of nescience within innovative startup organizations tends to spawn research in domains that the organizations had hitherto considered out of bounds. Synergy between prior internal knowledge and the result of said research frequently results in breakthrough innovations. Knowledge may thus abhor a vacuum, especially if its impedance is low, and competitive advantage may go to the organizational entity that is prepared to enhance the capacity to absorb the right kind of knowledge at the right time.

**Knowledge centrality** denotes the opposite of a knowledge vacuum—it refers to the loci within an organization or network in which knowledge is concentrated. Identifying these loci of knowledge is critical for problem solving (Weber, et al. [186]), but only if these loci are willing and able to transmit the knowledge that they have aggregated (Mayande [114]). This suggests knowledge impedance is asymmetric and that knowledge flows are directional. In an analogy to electrical engineering, knowledge impedance resembles a diode rather than a resistor.

**Knowledge Modularity.** The authors define knowledge modularity in a manner that is analogous to how physiologists (McClelland & Rumelhart [120], Plaut [141]) and designers (Baldwin & Clark [16] [17]) have defined modularity. Knowledge, like the brain or a design, is modular when it exhibits a structure in which the parameters and tasks are interdependent within units (modules) and independent across them. A system whose connections exhibit low knowledge impedance within modules and high knowledge impedance between modules is considered a modular knowledge system. By contrast, a system whose connections exhibit relatively high knowledge impedance within modules and relatively low knowledge impedance between modules is considered an integral knowledge system. Thus, knowledge modularity is more than knowledge centrality. It identifies organizational, geographic or virtual regions in which knowledge is concentrated.

Research of the kind described in this section has made contributions to management theory and practice. It has increased the effectiveness of industrial engineering and
technology management processes in the semiconductor industry (Cohen, et al. [59], Weber, et al. [175]-[176]), and it provided economic models that characterized a variety of high tech industries (e.g., Berglund, et al., [22], Bieglo, et al. [23], Hasenauer, et al. [97]-[98], Weber, et al. [177]-[178], Weber & Yang [187]-[191]). It has also contributed to novel theory in the fields of consumer behavior (Albar & Jetter [4], Zenobia, et al., [199]-[202]), knowledge management (Ploykitikoon & Weber [143], Weber [190], Weber, et al. [195a], Weber & Yang [189]), R&D management (Ploykitikoon & Weber [144]-[145]), open innovation (Mayande, et al. [116]-[119], Yang, et al. [197], Ploykitikoon & Weber [145]) and organizational change (Yang, et al. [197]). Most importantly, additional knowledge about technological knowledge has provided the authors of this paper insight into the diversity of contexts in which the proposed research must be conducted, in order to generate potentially transformative results.

IV. RESEARCH OBJECTIVES

The objective of the proposed research agenda is to gain a significantly enhanced understanding of how knowledge flows impact the network phenomenon of influence in online social networks. The following research questions are of particular interest to the authors and their colleagues.

A. Questions regarding knowledge: How is knowledge in online social network structured? Are there regions high and low knowledge impedance? Where are loci of knowledge? How do I find the person that knows what I need to know? How does knowledge spread? How rapidly does it spread? What are the pathways by which it spreads? What enhances the absorptive capacity and transmission capacity of the various nodes (people) or structures (subsystems) within an online social network? How are knowledge vacuum filled? How are knowledge short circuits avoided?

B. Questions regarding network structure: How do knowledge flows affect network structure and how do changes in network structure affect knowledge flows? Can the rate of structural change be moderated by modulating (changing the context of) knowledge flows?

C. Questions regarding influence: Which variables act as the best measures of influence in particular social networks? How do knowledge flows and network structure affect the nature of influence? How will modulating knowledge flows impact influence?

Three tasks, which are denoted below, have to be completed to address these areas of interest.

1. Benchmark metrics for influence in a variety of online social networks that are very different from each other. This endeavor will determine which metrics are best suited for measuring influence in a plethora of particular contexts.

2. Characterize the nature of knowledge flows in a highly diverse set of online social networks.

3. Characterize the impact of knowledge flows on the nature of influence in this diverse set of online social networks.

V. RESEARCH METHODS

Research methods for the proposed studies incorporate the lessons learned by the authors and their colleagues while performing the pilot study. The proposed research on online social networks will utilize methods that were successfully tested in the pilot study (Mayande, et al. [114]-[119]). It will consist entirely of exploratory, longitudinal population studies of online communities, which follow the research framework depicted in figure 1 in the appendix.

A. Research Design

Theoretical Sampling. Mayande, Jetter & Weber [115] have identified six purposes for online social networks—relationship building; sharing and trading; stakeholder engagement; fostering common interest; advancing a common cause; and improving government services—and many more may exist. Different metrics may measure influence in each of these categories, and knowledge flows may affect influence in different ways. The authors consequently propose to study multiple online social networks in each of these categories and perhaps more in categories that have not yet been identified. These studies will be inherently exploratory in nature because prior research of this kind is lacking. In addition, extant theory for online social networks is not very descriptive, and its normative value is not particularly high.

Data Sources. Data come from records of online conversations that take place on social media platforms (e.g., Twitter, Facebook). Collaborators from industry, who have purchased these data from the platform firms, will provide these data to the authors at no cost (see Facilities and Equipment). The authors have applied for funds for purchasing online data from additional sources.

Network Constraints. The pilot study has shown that scale and directionality affect network phenomena. Thus scale and directionality will act as control variables in the proposed research. Network size will become an important sampling consideration, in order to observe scale effects. Every online social network under study will be analyzed four times—without directionality, with directionality, with a focus on information consumption and with a focus on information propagation.

Population Studies. Modern data extraction capabilities on the Internet enable the study of whole populations. This practice eliminates sample selection bias; it also ensures that the observed results are valid and generalizable to the entire population under study. This is especially important in studies that involve networks, as selecting only a sample instead of
the population can break a network into multiple small networks (Goggins and Petakovic [91]), leading to faulty results (Mayande [114]).

Research Setting. A diverse set of online communities, which serves the abovementioned purposes of online social networks identified in the pilot study [115], has been chosen as the setting for this research. The proposed studies will track and analyze context-specific conversations within these communities, which serve as the unit of analysis for the proposed research. The knowledge flows, network structure and patterns of influence that these conversations reveal will be compared to each other.

B. Research Variables

Independent Variables reflect basic network characteristics (see figure 1) such as the number of nodes (individuals) and ties (relationships) in the network, the cluster coefficient (Newman, et al. [132], Wasserman & Faust, [172], Watts and Strogatz [174]), network density (# of ties / # of nodes) and reciprocity (the fraction of ties in a directed network that are bi-directional).

Moderating Variables. Two sets of variables moderate the independent variables in figure 1. One set measures network organization; the other measures network flow. Network organization variables include the “small world metric” (Watts & Strogatz [174]), the “scale free metric” (Barabási & Albert [5] [6] [18]) and assortativity (Newman, et al. [132]). Many of the variables that measure network flow come from graph theory. They include the number of paths, the number of geodesics (shortest paths), the graph diameter (of the network), the average (virtual) path length and the average (virtual) geodesic length (Borgatti [29]). Variables that involve paths act as proxies for the extent to which information spreads; variables that involve geodesics act as proxies for the speed at which it spreads. Variables that measure impedance to network flow come from applications of information theory (e.g., Abramson [1], Beckmann [21], Cover & Thomas [64], Hartley [95], Kullback [106], Weber, et al. [185] [186]). The power law distributions of paths and geodesics per node have been included as measures of network organization, because social networks are frequently characterized by a power law distribution of their connections (Barabási et al. [18] [19], Castellano et al. [48], Clauset et al. [58], Muchnik et al. [127]). Formulas for the abovementioned variables are given in the cited references and in chapter 4 of Mayande [114].

Dependent Variables. The dependent variables of the proposed research measure aspects of influence in a variety of ways. (Their formulae can be found in the references cited in this paragraph.) Centrality metrics from graph theory (e.g., Freeman [84] [85]) act as proxies for measuring the communication activity of a particular node (degree centrality); the control a node can exert on the communication process in a network (betweenness centrality); and the efficiency of a node’s communication process (closeness centrality). Eigenvector centrality (Bonacich [27]) has been included as a dependent variable because it has been shown to measure the extent of a node’s influence very effectively (Bonacich [28], Mayande [114]). Entropy centrality (Mayande & Weber [116], Nikolaev, et al. [134], Tutzauer [169]) has been added to the list of dependent variables because it measures the amount of information that can concentrate in a particular node. The power law distributions (Clauset, et al. [58], Muchnik, et al. [126]) of centrality metrics have been included as dependent variables to assess the impact of the power law distributions of network structure variables on network phenomena.

Associated with every centrality metric is a centralization metric, which measures the differences in centrality between the nodes that are the most central and all others in the network (e.g., Freeman [84] [85]). Centralization consequently is a property of the network as a whole, rather than the property of any individual node. Centralization metrics also act as proxies for modularity (including knowledge modularity) when they are applied to subsections or (virtual) “regions” of the network.

C. Virtual Field Work

Virtual field work consists of collecting data about a particular social network from the Internet, analyzing the data using statistical methods and repeating the analysis in a variety of specific contexts in which the networks under study operate. These contexts will be explored in keyword searches, an approach that has been deployed successfully in the pilot study (Mayande [114]) and by other researchers (Jansen, et al. [103], Teevan, et al. [164], Williams, et al. [194]). Different sets of keywords will act as proxies for different kinds of knowledge.

Data Collection. Data for the proposed studies will be collected retrospectively in continuous time so that the number and sequence of events and the duration between them can all be calculated. The main advantage of this approach lies in the greater detail and precision of information (Blossfeld & Rohwer [25]). It also reduces the time required to collect data, and it enhances the chances of recognizing the overall patterns (Leonard-Barton [108]). The pilot study has demonstrated that networks with more than 20,000 active nodes can be analyzed on a daily basis, suggesting that network phenomena that cannot be explicated by extant theory will be observed readily. Data extraction occurs through application program interfaces provided by the platform firms.

Data Analysis follows the procedures that were deployed successfully in the pilot study (see [114] for details). To assess whether a network’s structural features have been identified, the output data of every network under study will be compared to a simulated random network with the same number of nodes and ties (Erdős & Rényi, [75]). A correlation analysis will determine the degree of interdependence between variables and measure criterion validity (Cooper & Emory [63], Murphy & Davidshofer [130], Pennington [140]). An
exploratory factor analysis (Cattell [49]) with Varimax rotation will be conducted to find the smallest number of interpretable factors that can adequately explain the correlations among the set of variables (Field [83], p. 619). A Scree test (Cattell [50]) will then be performed to produce a more interpretable solution. Cronbach’s alpha (Cronbach [65]) will be used as a measure of internal consistency and by implication as a measure of reliability. Multiple linear regression analysis will help determine the relative impact of the independent variables on the dependent variables, as well as the impact of moderating effects. The values of R square and the adjusted R square will be used to provide a statistical test of the model’s ability to predict the dependent variables (Field [83], pp. 179).

**Keyword Modulation.** Initially, networks will be analyzed without any context-specific keywords, in order to characterize the nature of non-contextualized network flows and their impact on network phenomena. The analysis will be repeated, ceteris paribus, by introducing sets of context-specific keywords. This keyword modulation will show how particular knowledge flows, which are after all context-specific network flows, differ in nature from each other, as well as from network flows that have no particular context. Changes in centrality and centralization that result from the introduction of keywords will determine the impact of contextualization (i.e. knowledge) on the network phenomenon of influence.

**VI. WORK PLAN**

The exploratory nature of the proposed research mandates a work plan that allows for iterative learning such as the one presented in figure 2 in the appendix. In this plan, virtual field work and theory building run in parallel. Insights gained from one activity are likely to influence the execution of another. The diagonal arrows indicate the flow of information between concurrent activities. Insights gained from virtual field work contributes to novel theory pertaining to online social networks, because new models and novel theory will be built from fresh empirical evidence the will be generated in the virtual field studies. Conversely, novel theoretical insights by the authors and other researchers may impact the design of subsequent virtual field work and the choice of which online social networks should be studied next.

Figure 2 shows that the proposed research begins with setup activities, which commence at t=0 and should finish within less than 6 months. Setup activities consist of capturing the requirements for the project, fine tuning the research and conducting a pre-test, which should shed light on the extent to which research methods deployed in the pilot study are applicable to the proposed studies. For example, algorithms that execute statistical analyses automatically are under development to reduce the time required for statistical analysis by an order of magnitude. If these are deployed, then the authors believe that they can comprehensively analyze more than 100 online social networks within three calendar years.

Figure 2 shows that over 24 months of virtual field work and theory building follow the setup activities. This prolonged period is required to complete the three tasks that have been identified in section IV. Some of these tasks have to be executed in sequence because the outcome of one or more tasks influences the research design for the next task. For example, task 3 has to start once task 1 has been completed because the outcome of the metrics benchmarking study will reveal the best metrics for the network phenomena to be investigated in the impact study. Similarly, task 3 should not begin until significant insights into the nature of knowledge flows in online social networks have been gained.

Milestones 1 and 2, which transpire at t=1 year and t=2 years respectively, are defined as follows:

**Milestone 1:**
- The authors expect to have completed the pre-test and a significant amount of virtual field work.
- Basic insights into knowledge flows, their impact on network structure and network phenomena, as well as the rudiments of a behavioral theory of online social networks should have emerged.

**Milestone 2:**
- For a wide range of online social networks, the nature of knowledge flows and its impact on influence has been characterized and useful performance metrics have been identified.

Changes in the course of the proposed research will be considered at milestones 1 and 2. If findings contain unexpected insights of significant magnitude, then subsequent research may go in a direction that looks more promising than the one that was originally expected. No course correction will occur, if the findings contain no unexpected insights or if unexpected insights are of insufficient magnitude to warrant a change in direction. Some course corrections are anticipated in response to feedback from colleagues at other universities who are conducting complementary research on online social networks. This feedback is likely to occur at conferences at which the authors hope to present their results.

The last six months of the project will be spent summarizing all the findings and preparing for final dissemination of the results. A final report will be created at the end of the third year of the project. It will contain, in addition to the findings of the proposed research, recommendations for further research. This research may cover topics that pertain to technologies that have not yet been invented and to phenomena that have not been observed as of yet.

**VII. DISCUSSION**

This paper has reviewed a substantial portion of the academic literature that pertains to knowledge flows, networks structure and influence in online social networks. It has discussed key concepts pertaining to technological knowledge, and introduced a few new ones like knowledge impedance and
knowledge modularity into the discussion. The basic premise of this paper is that knowledge flows can be described in terms of these concepts, and that knowledge flows strongly affect the network phenomenon of influence in online social networks. Following through with the proposed research agenda will determine whether this premise is correct.

The authors believe that by performing the proposed research agenda they are likely gain theoretical insights into how online social networks get organized. To what degree are they emergent? To what extent can they be designed and controlled by the real world? To what degree can they influence the real world? To what extent does knowledge impact organizational structure and the network phenomenon of influence?

The authors are particularly hopeful that they will identify metrics for knowledge and influence online. For example, can knowledge impedance be measured by metrics from information theory? By extension, can the Kullback-Leibler formula help determine the direction of future research? To what extent can keyword searches contextualize knowledge flows? To what degree does knowledge modularity, as measured by centralization metrics, determine or predict specific organizational competences? More fundamentally, to what degree are conclusions about particular online social networks generalizable? To what degree are they context specific?

The proposed studies of many online social networks that operate in a variety of contexts is likely to generate empirical evidence for the existence of phenomena, which cannot be explained by extant theory of innovation and organizational change. Each network will be examined with and without taking directionality, information consumption and information propagation into consideration. Furthermore, the research setting—online communities—differs significantly from that of most studies of social networks in the real world. Due to all these opportunities for contrast, the proposed research is well-positioned to have a broad theoretical impact. At a minimum, it should be relatively easy to assess the generalizability of any theoretically significant insight that emerges from the data.

The limitations of the proposed research agenda should be stated at its outset. For example, the proposed research does not investigate specific attributes of particular social media platforms, because that has already been performed by other researchers (e.g., Goggins & colleagues [24] [90] [91]). Likewise, the proposed research does not cover leadership in the competitive advantage for innovative firms, which has already been mentioned, can translate into a competitive advantage for regions and nations, if firms, organizations and individuals with a high level of understanding of online social networks concentrate geographically. Security and defense implications of the proposed research agenda need to be mentioned in this context because highly fluid social media campaigns can disrupt social order, for good or for ill. An asymmetric understanding of how online social networks behave will enhance the ability of existing, real-world institutions to cope with them.

Every researcher probably wishes that his/her research will yield results that are transformative, but that cannot be guaranteed a priori, especially if said research is by nature exploratory. The authors of this paper have high hopes that the research described in this proposal has the potential of gaining some transformative insights, simply because so little is definitively known about the subject matter at hand. Yet, comprehending the subject matter at hand is critically important to understanding the socio-technical world in which we live. The authors consequently believe that the proposed research is definitely worth pursuing.

REFERENCES


APPENDIX

Figure 1: Research design including variables.

Figure 2: Work plan for the proposed project