Information in Digital, Economic and Social Networks

Arun Sundararajan,[1] Foster Provost,[1] Gal Oestreicher-Singer[2] and Sinan Aral[3]

(*Information Systems Research*, forthcoming)

Digital technologies have made networks ubiquitous. A growing body of research is examining these networks to gain a better understanding of how firms interact with their consumers, how people interact with each other, and how current and future digital artifacts will continue to alter business and society. The increasing availability of massive networked data have led to several streams of inquiry across fields as diverse as computer science, economics, information systems, marketing, physics and sociology. Each of these research streams asks questions which at their core involve 'information in networks' — its distribution, its diffusion, its inferential value and its influence on social and economic outcomes. We suggest a broad direction for research into social and economic networks. Our analysis describes four kinds of investigation that seem most promising. The first studies how information technologies create and reveal networks whose connections represent social and economic relationships. The second examines the content that flows through networks and its economic, social and organizational implications. A third develops theories and methods to understand and utilize the rich predictive information contained in networked data. A final area of inquiry focuses on network dynamics and how IT affects network evolution. We conclude by discussing several important cross-cutting issues with implications for all four research streams, which must be addressed if the ensuing research is to be both rigorous and relevant. We also describe how these directions of inquiry are interconnected: results and ideas will pollinate across them, leading to a new cumulative research tradition.

[1] *NYU Stern School of Business*

[2] *Tel-Aviv University, Recanati School of Business*

[3] *NYU Stern School of Business and MIT*

1. Introduction

Information technologies have made digital networks ubiquitous. Some of these digital artifacts provide a window into pre-existing social and economic networks, while others represent entirely new socioeconomic and socio-technical systems. The tremendous research interest in networks and the massive amounts of available digital trace data enable studies of population-level human interaction on scales orders of magnitude greater than what was previously possible (Lazer et al. 2009; Agarwal, Gupta and Kraut 2008).

We discuss four interrelated research streams that ask questions that at their core involve "information in networks." For this paper, we define *information in networks* as ***any quantity that reduces uncertainty or introduces novelty in the context of a relationship structure or set of relationship structures***. This definition is intentionally broad, as the interrelated research streams span many fields. Nevertheless, the basic idea is that *informatio*n is anything that reduces uncertainly or introduces novelty and *relationship structures* broadly define different conceptualizations of networks relevant to contemporary research, especially that of interest to IS. Throughout the paper we describe many concrete examples. We then develop additional guidance toward categorizing information in networks.

To discuss *relationship structure* in more detail, let's consider a common research task: examining interrelated random variables. Is that research on information in networks? More concretely, consider a researcher examining a data set to draw conclusions. Her data set comprises multiple *instances* of a *set of variables*. The way such data typically are treated falls **outside** what we will consider "information in networks" (although it also could be viewed as a degenerate case). Oversimplifying for clarity, typically the researcher imposes a "star" relationship structure on these random variables. This star structure (i) posits a dependency between a specified dependent variable (y) and each of the independent variables or *features* ($x_1 \ldots x_n$), (ii) may or may not assume conditional independence between features, (iii) assumes independence across the different instances, and (iv) asserts the same "model" or statistical interdependence relationship among variables for each instance (see figure 1 for illustration). A typical regression model would follow this star structure.

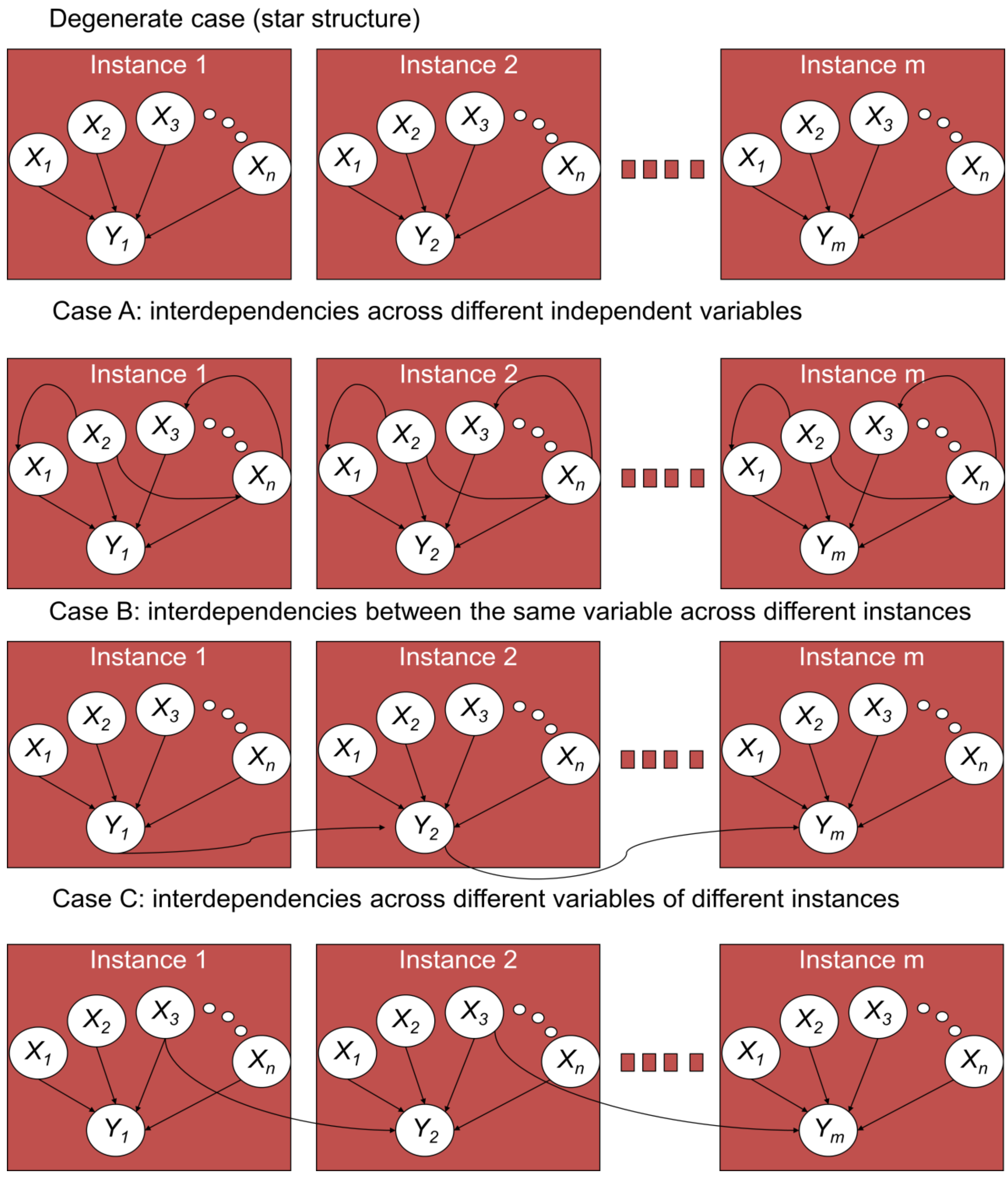

Figure 1: Illustrates the different statistical interdependence relationship structures posited by researchers

Importantly, the star relationship structure posited by the researcher does not include any statistical interdependencies:

(a) across the different independent variables of the same instance; for example, even if we were to know a consumer's (instance's) probability of purchasing (dependent variable), her income and occupation (features) may be related;

(b) between the same variable across different instances; for example, it may be that the product choices (dependent variable) of different consumers (instances) who are friends are correlated, or that the performance levels of different employees on the same team are correlated, or

(c) across different variables of different instances. For example, that the product choice (dependent variable) of one consumer might be related to the level of advertising (feature) of a friend of the consumer (different instance). Analogously, the performance of an employee might be related to the education level of a team member.

In each of these examples, the relationship structure that the researcher needs to assert is more complex than the degenerate star network illustrated above. The data required to conduct a meaningful analysis that takes this relationship structure into account may need to specify which instances are connected (for example, which consumers are friends), or more generally, which variables of which instances are connected. This specification is what yields a non-trivial "network." In this paper, we are generally interested in networks based on interdependencies of types (b) and (c) above—inter-instance networks[4].

Armed with this more complex networked relationship structure, our researcher might choose to conduct *dyadic inference*, wherein she restricts her analysis to being between pairs of variables connected in the network; or she might choose to conduct *structural inference*, wherein larger portions of the

---

[4] We do not mean to downplay the importance of research on intra-instance networks (type-a). Indeed, research on *Bayesian networks* and *Markov networks* (Pearl, 1988), changed the world's understanding of statistical inference, largely based on intra-instance applications. However, contemporary researchers have turned considerable attention to inter-instance networks, and so will we.

network are simultaneously analyzed, allowing the network's structural properties and/or the position of variables in the network to inform the analysis. The specific "inference" in question varies—it may involve prediction or explanation; it may be quantitative or qualitative (we will discuss specific examples throughout the paper). Alternatively, the researcher might choose to better understand the flows of information within this network. In each of these cases, if the analysis involves data over time, the researcher may need to account for (or may want to better understand) dynamics in the network.

As mentioned above, the study of social and economic networks spans many disciplines. These include Computer Science, Economics, Physics, Sociology, Statistics, and others. Traditionally, each discipline has tended to focus on specific pieces of a larger puzzle—of which we try to give an integrative view in this paper. Moreover, we assert that researchers who span many of the individual disciplines contributing to contemporary network research are uniquely positioned to provide and assume intellectual leadership in research on social and economic networks. For this reason we feel IS research is particularly well positioned.

In order to provide an integrative view of the study of social and economic networks, we describe *four broad directions* of inquiry. Section 2 examines how information technologies create and reveal networks whose connections represent a range of social and economic relationships. Section 3 examines the content that flows through networks and its economic, social and organizational implications. Section 4 examines theories and methods to understand and utilize the rich predictive information contained in networked data. Section 5 examines network dynamics—how networks are created, evolve and dissolve—and how IT affects network evolution. For each of these four streams, we discuss salient research challenges, summarize what is known about their solutions, motivate key questions that have arisen or are expected to arise over the coming decades, and examine the importance and feasibility of research that addresses these questions.

Our analysis also reveals *important cross-cutting issues* with implications for all four research streams, issues which are important if research is to proceed with rigor and relevance. Section 6 discusses the role of three of these cross-cutting issues: network structure, the need to inform network analysis with underlying theories from the social sciences, and the importance of causal identification in empirical network analysis.

2. How Information Technologies Create and Reveal Networks

A broad range of digital artifacts and social structures constitute what researchers label "networks." As electronic interaction spreads, it reveals new information about existing social connections and economic similarities and at the same time alters existing economic and social structures. A wealth of new research opportunity arises on account of this ongoing transition towards digital interaction.

Consider the emergence of new "digital networks," IT artifacts that are created and exist to facilitate and mediate digital interaction. Some of these, like Facebook's social graph and its associated user content, are digital approximations of (well-understood, familiar) social networks. Others, like the network of users connected by their shared social bookmarks on del.icio.us, defy traditional categorization as communities or social networks, creating new kinds of social spaces or cultural fields (Arriaga & Levina, 2008). Still others, like networks of products connected by overlapping consumer purchases, are not social networks, but instead create an entirely new kind of artifact of interconnected entities, which one might think of as approximating an underlying "economic" network (Jackson 2008, Oestreicher-Singer and Sundararajan 2012b). Some of these networks form an integral part of electronic interactions, for example, those on social networking web sites. Others, like the co-purchase network on Amazon.com, are constructed by capturing, filtering, analyzing and/or summarizing the data trails left by these electronic interactions. Nevertheless, the visible presence of the interconnections may affect interaction and outcomes and hence should be carefully researched.

As these digital networks become more visible and influential, researchers have become increasingly interested in understanding their social and economic impact. The economic and social transformations they engender could be even greater than those induced by the widespread adoption of IT in business in the past decades. Correspondingly, IS research has begun to address specific questions related to these transformations, such as, "*How does the existence of social and economic networks as digital artifacts alter individual choices and global outcomes?*" The existence of Amazon's co-purchase network has been shown to affect individual product demand (Oestreicher-Singer and Sundararajan 2012a) and to explain variations in the distribution of demand across popular products and niche products (Oestreicher-Singer and Sundararajan 2012b). Facebook's social graph alters individual choices about both consumption (Jill read an article that her friend recommended) and relationship formation (Jack formed a link to someone who became visible as his friend's friend, or someone whose picture Jack found appealing) (Aral and Walker 2011a). However, we currently have a limited understanding of changes in the individual-level process of such choices as well as of their broader social and cultural impact. Both of these represent fertile directions for future inquiry. Another research opportunity that the availability of these digital networks provides is the ability to enhance our understanding of the nature of consumer preferences and product characteristics through the co-choice networks they generate. In a sense, these networks summarize the (unmanageably) high-dimensional preference space that describes tens to hundreds of millions of diverse consumers and the characteristics space describing hundreds of thousands to millions of products (and theoretically at least an order of magnitude more in both cases). These connections are relevant precisely because they are based on shared outcomes and thus highlight the information in these "spaces" that is (or was at some point) decision-relevant, a point recognized by early literature on collaborative filtering and also implicit in the use of networks of statistically correlated entities for predictive modeling (more on this later).

A natural "sub-question" relates to the specific impact of the *visibility* of these digital artifacts. Our social, economic and cultural connections are made more persistently visible by virtue of their being encapsulated and displayed as digital artifacts, and the visibility of these networks by itself will alter their socioeconomic impact (Oestreicher-Singer and Sundararajan 2012a). One interesting question raised by the visibility of networks is: what is the impact of remote connections or of observation of remote connections? For example, in the context of social influence, researchers have traditionally focused on the effects of one's friends or social groups on one's actions. The underlying assumption in those studies was that the influence of a friend's friends is mitigated by the interaction between the focal person and his friends. However, if the friends of my friends are made visible to me on a social networking website, I might be directly influenced by their actions. Moreover, I can now leverage not only my peers' social capital but also their second-degree or even third-degree connections' social capital. Further, the mere possibility of making economic or social actions visible might alter these actions (Rhue and Sundararajan, 2012). This can create a new form of social "influence," which rather than flowing from the actor to the observer, flows from the observer to the actor.

Looking more closely into these digital artifacts suggests that there may be an entirely new class of social spaces being created by their emergence. Many digital networks of the kind we describe are not naturally defined as communities because they do not involve direct *interaction* between the individuals that compose them (Arriaga and Levina 2008). Rather, the connection is more tacit, often revealing some kind of shared preference (for example, see the "quasi-social" content-affinity network of Provost et al. (2009)). Understanding the sociological underpinnings of the groups created and implied by these artifacts, ideally through the development of new theory, will contribute to the understanding of their short-run economic and social impact, as well as the eventual changes they will engender in social structure and human behavior as they become increasingly representative and encompassing of the totality of human interactions and social groups.

In the rest of this section we discuss three examples of such questions. Given the role of IT features and design in enabling, constructing and defining the use of different social and economic networks, IS research is well positioned to lead research is this field—on these questions and beyond.

First, the fact that these social and economic structures are encoded in and potentially altered by digital artifacts leads to the possibility of optimizing them, which in turn raises the question of their effective design and efficient structure. There are a number of dimensions related to this question. At an elementary level, there is the problem of the effective design of the "engine" that facilitates the emergence of the digital network. Restrictions on content access (such as the ability to block your page on Facebook), link formation (for example, the ability to follow any Twitter user, versus the need for mutual consent on Facebook) are usually the result of design choices. Such choices could affect the success and usage patterns of the different networks and should therefore be carefully managed. Such choices can also affect the type of information transmitted over these networks. One interesting recent example is Twitter.com. Compared with many social networking websites (including the recent Google+), Twitter provides users less control over their lists of followers, and messages posted on Twitter accounts are publicly available on the Twitter website and sent to all the user's followers. As a result, Twitter has been described as a "social broadcasting technology" rather than a "social networking technology" (Shi, Rui and Whinston, 2011). Another example is LinkedIn, where choices of privacy restrictions by the platform designers have led it to become a prominent recruiting tool. Firms or governments that seek to use IT-based systems to create similar networks will persistently face questions of optimal engine design, questions that are naturally informed by IS research. This also suggests benefits from an ongoing intellectual exchange with the IS community on design science (Hevner et al. 2004), a point underscored by the fact that the 2009 INFORMS ISS Design Science Award was given to a team working on “Social Network-based Marketing Systems.”

Similar results have been found in a somewhat different domain. In the context of knowledge management, network analysis has been used to model knowledge transfer and sharing within an organization or market (Hansen 1999), among intra-corporate and strategic alliances (Tsai 2001, Inkpen and Tsang 2005), and among open-source projects (Kuk 2006). Information technology is perceived to link sources of knowledge (people or documents), thus widening and deepening "knowledge flows" (Carlsson et al. 1996). The design of the underlying digital platforms that support such networks has been shown to affect the rate, diversity and probability of knowledge sharing. For example, Ma and Agarwal (2007) document the role of perceived identity, self-presentation and virtual co-presence in increasing knowledge sharing; Ren, Kraut and Kiesler (2007) show how highlighting interpersonal similarities can cause increased attachment and encourage direct reciprocity, and Jeppesen and Laursen (2009) find that search and integration of knowledge functions positively moderate knowledge contributions.

Second, when viewed more broadly as a socio-technical problem rather than an engineering problem, “design” also includes the appropriate choice of a set of “seed” adopters (Domingos and Richardson 2001, Kempe, Kleinberg and Tardos 2003, Sundararajan 2008, Aral, Muchnik and Sundararajan 2012). Merging the software design problem with the literature on viral marketing, word-of-mouth (Hill, Provost and Volinsky 2006, Manchanda, Xie and Youn 2008, Nam, Manchanda and Chintagunta 2010) and epidemic contagion (Pastor-Satorras and Vespignani 2001) could lead to a better understanding of the right combination of technology, incentives and sociology for seeding a network to maximize the diffusion of products or behaviors. Aral and Walker (2011a) suggest that such thinking need not be restricted to entities whose primary goal is the creation of a digital network. They describe how “viral product design” can be used to engineer products that ‘go viral’ in a social network. By adding features that increase the amount of peer-to-peer sharing and awareness of a product, technologies from Hotmail to PayPal successfully have increased their reach, and such design strategies can be incorporated into a broad range of products and services. This means that a program of research on optimal design could have wider-ranging implications for marketing and business beyond merely explaining the emergence of digital networks.

Third, a related but distinct question concerns the efficient design and use of the digital network itself. For example, what kind of visibility in the social graph facilitates the formation of valuable (rather than fleeting) friendship? What depth of transparency in LinkedIn leads to its most effective use as a recruiting platform? What kinds of co-purchase networks lead to the highest demand increases or the most desirable division of attention between products? Each of these questions relates to the impact of the structure of the digital artifact on social or economic outcomes. To illustrate this further, Oestreicher-Singer and Zalmanson (2012) provide a first set of insights into the interplay between digital networks and traditional non-networked products using data from the music streaming site Last.FM. They demonstrate that the presence of social networking features increases the adoption of paid subscriptions, which in turn improves the individual listening experience. Similarly, Sykes, Viswanath and Sanjay (2009) show that social networks affect technology adoption within organizations. As firms continue to seek evidence of the business value of social and digital networks, research establishing such empirical linkages will contribute to the presence and impact of the IS field as a source of research with business-relevant findings.

Clearly, what is "efficient" (or what the objective function for the design is) depends on the context, and thus, if IS research chooses to embrace this line of inquiry, the most productive path forward may well be one in which the initial studies are specific to fairly narrowly defined contexts. What seems equally important is that these artifacts may eventually encompass a large fraction of structures that used to exist in physical spaces, and for which the supporting artifacts themselves may be owned by private entities rather than being in the public domain. This raises important sociological and policy questions, which, while perhaps beyond the scope of IS research in the near term, are worth seriously considering. For example, in a future world where interaction is largely digital, does it seem reasonable for Facebook's design to aim to affect the average strength of human social ties, or for a private entity to be able to restrict an individual's ability to form relationships through their ownership of the digital "space of interaction"? What, if any, effect will such policy decisions have on previously accepted social constraints like Dunbar's number (Dunbar 1992)—the theoretical cognitive limit to the number of people with whom one can maintain stable social relationships, typically estimated to be approximately 150?

3. The Flow of Network Content

Many theories in sociology, marketing, economics and IS, about the strength of weak ties and brokerage (Granovetter 1973, Burt 1992), social influence and contagion (Iyengar, Van den Bulte and Valente 2011, Aral, Muchnik and Sundararajan 2009, Aral 2011, Aral and Walker 2012), the diffusion of innovations (Rogers 2003), and competition (Burt 1992), rely on assumptions about how information, knowledge, resources, transactions, influence and attention flow through networks. However, the vast majority of networks research today remains "content-agnostic" (Hansen 1999, 83). While network structure and node outcomes are privileged, direct evidence on content flowing through networked relationships is rarely used to validate theory. Scholars studying how digital technologies alter business and society have a unique leadership role to play in this particular area of network research because IT is critical both to the socio-technical process that facilitates content flow in networks and is the platform that naturally transcribes content flow into data for analysis in research. Combining the analysis of network structure with analysis of the content that flows through networks, digital or otherwise, can open new avenues for understanding how networks affect a variety of phenomena.

The assumption that network structure influences the distribution of information and knowledge in social groups (and thus characteristics of the information to which individuals have access) underpins a significant amount of theory linking outcomes to social structure. In fact, information-based mechanisms have been the centerpiece of network theories for some time, including theories of network brokerage (Burt 1992), cohesion (Coleman 1988), the strength of weak ties (Granovetter 1973), the search-transfer problem (Hansen 1999), herding behavior in markets (Bikhchandani, Hirshleifer and Welch 1991) and many others. Information flow in networks also affects a variety of outcomes of specific interest to IS scholarship, including the productivity of information workers (Cross and Cummings 2004, Cross and Sproull 2004, Aral, Brynjolfsson and Van Alstyne 2011), coordination and collaboration in software development (Grewal, Lilien and Mallapragada 2006, Hahn, Moon and Zhang 2008, Singh, Tan and Mookerjee 2008, Singh 2011), health IT and outcomes (Kane and Alavi 2008), bidding behavior in online auctions (Hinz and Spann 2008), the diffusion of electronic medical records (Miller and Tucker 2009, Angst et al. 2011), virtual team communication and coordination (Ahuja and Carley1998), and technology adoption inside firms (Burkhardt and Brass 1990, Tucker 2008, Sykes et al. 2009).

However, evidence on information flowing through networked relationships is rarely used to validate information-based network theories. As Burt (2008, 253) notes: “Empirical success in predicting performance with network models has far outstripped our understanding of the way information flow in networks is responsible for network effects… the substance of advantage, information, is almost never observed.” IS researchers are particularly well positioned to address this important challenge.

IT-focused research can combine network analysis with text-mining techniques to validate information-based network theories. For example, Aral and Van Alstyne (2011) combine analysis of the network structure of email communication with text analysis of email content to test whether diverse networks actually provide access to non-redundant information—a long-standing assumption of theories such as the strength of weak ties (Granovetter 1973) and structural holes (Burt 1992)—and Aral and David (2012) replicate and extend this work in a different organizational setting. Many other related papers combine text analysis with network analysis to observe how opinion leaders emerge in online review communities (Lu, Kinshuk and Singh 2010), how information diffuses in broadcast information networks such as Twitter (Yang and Counts 2010), and how information diffusion in organizational email networks affects information worker productivity (Aral, Brynjolfsson and Van Alstyne 2007). Simultaneous examination of network structure and information content has also been used to analyze mental models in teams (Carley 1997, Aral, Brynjolfsson and Van Alstyne 2008), to generate topical maps of relationships between concepts, and to extract the network structure that connects people, organizations and resources as they are described in text documents (Diesner and Carley 2005). As these studies demonstrate, examination of information flow in networks can help us better understand why networks seem to play such a vital role in so many social and economic phenomena.

Information technologies also play a role in orchestrating the flow of attention, awareness and influence in networks. For example, product recommendation networks suggest related content to consumers and guide user attention from one product to another, influencing consumer demand patterns (Carmi, Oestreicher-Singer and Sundararajan 2009, Oestreicher-Singer and Sundararajan 2012a). The social influence individuals have over one another is also increasingly mediated by digital networks and the diffusion of content they facilitate, a point made increasingly salient by the growing visibility of such networks (Oestreicher-Singer and Sundararajan, 2012a, Rhue and Sundararajan 2012b). For example, users of online social networking websites send reviews of products and services as well as invitations and notifications for their online and offline activities to their network peers using messages created in digital networking platforms, influencing each other's behavior (Aral and Walker 2011a). Influence and attention also flow over personal digital communication networks. Product and service adoption is known to be correlated in network space and time with telephone conversations (Hill, Provost and Volinsky

2006), instant messaging communication (Aral et al. 2009), and medical referral networks (Iyengar et al. 2011). Firms also have the ability to manipulate word-of-mouth attention and influence by proactively engineering content that appears in online social networks and opinion forums (Dellarocas 2006, Mayzlin 2006). The flow of attention and influence through digital networks, whether through personal electronic communication or through the automated recommendations that aggregate demand correlations across products, remains a central area of research into network content that can help develop our understanding of how networks shape social and economic outcomes in markets and within firms.

Beyond information and attention, several types of economic transactions and monetary and human resources also flow through networks. Unfortunately, relatively little work emphasizes how the structure of resource flow enables and constrains the types of resources that are flowing or how these contingencies affect the social and economic outcomes under consideration. Researchers have analyzed labor flows among firms (Tambe and Hitt 2010) and across labor markets (Munshi 2003); the diffusion of user-generated content in online networks (Oh, Susarla and Tan 2011, Ghose and Han 2011, Goldenberg, Oestreicher-Singer and Reichman 2012); the flow of transactions in securities markets (Baker 1984); the flow of trust in informal borrowing networks (Karlan et al. 2009); the spread of computer viruses in email networks (Newman, Forrest and Balthrop 2002), in mobile networks (Wang et al. 2009) and across the internet (Pastor-Satorras and Vespignani 2001); default and interest rates in online lending markets (Lin, Prabhala and Viswanathan 2009); and the structure and evolution of global trade networks (e.g. Wilhite 2001). Scholarship in virtual team communication and coordination could also benefit from analysis of communication content and resource flows in networks. Most of this work examines interaction structure and node attributes and outcomes (e.g. Ahuja and Carley 1998, Leenders et al. 2003). Rarely does work emphasize what actually flows, instead focusing exclusively on the structure of the network of flows and how the structure is correlated with node outcomes. Future research in this direction might draw on prior work which has examined the style of interaction between communication partners (Yates and Orlikowski 1992, Orlikowski and Yates 1994).

Given the role of IT in enabling, constraining and recording the micro-level, time-stamped flow of content through social and economic networks, IS research is well positioned to address several of the most important research questions at the core of network-based theories of human behavior. We discuss three examples of such questions next.

First, IS research can markedly improve our understanding of how content flows through networks and specifically how network structure enables and constrains the flow of different types of content. At a high level, textual information such as tweets or emails may exhibit diffusion and cascading properties in networks that differ from those of viruses (real or computer), for example. Although most current views of information access in networks define network content as the "attributes of nodes" (Rodan and Galunic 2004), information exchange is a social process, and knowledge transfer is a discretionary activity (Reagans and McEvily 2003, Wu et al. 2004; for a review see Aral et. al. 2007). A connection to an individual with a certain information endowment affords the possibility of receiving that information, but by no means guarantees it. As Wu et al. (2004, 328) point out: "There are … differences between information flows and the spread of viruses. While viruses tend to be indiscriminate, infecting any susceptible individual, information is selective and passed by its host only to individuals the host thinks would be interested in it." In fact, information sometimes is withheld even when it is known to be of interest to others, reflecting selection and discretion in social choices concerning information sharing. Network structure and its visibility in digital networks may alter what types of information people are willing to share and with whom. In addition, the design of the digital network engine may directly affect choices individuals make about information sharing and seeking. Furthermore, the diffusion of content such as email viruses can be affected by network latency, the effectiveness of prevention technologies and the bandwidth of communication channels (Dezso and Barabasi 2002, Wang et al. 2009). Understanding how information flows in networks differ from the movement of email viruses, economic resources or skills from firm to firm could add important new dimensions to our understanding of several long-standing research areas across disciplines.

Different types of information may also diffuse differently. Aral et al. (2007) show that news and discussion topics exhibit distinguishable diffusion characteristics in organizations. They find the diffusion of news, characterized by a spike in communication and rapid, pervasive diffusion through the organization, is influenced by demographic and network factors but not by functional relationships (e.g. prior co-work, authority) or the strength of ties. In contrast, the diffusion of discussion topics, which exhibit shallow diffusion characterized by 'back-and-forth' conversation, is heavily influenced by functional relationships and the strength of ties, as well as demographic and network factors. Discussion topics are more likely to diffuse vertically up and down the organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news is more likely to diffuse laterally as well as vertically, and without regard to the strength or function of relationships. One can imagine other relevant categories or characteristics of information, for example, secrets, gossip, complementary information, valuable or less valuable information, information with legal implications, etc. Understanding how different types of information exhibit different diffusion patterns through networks will help us predict and explain who in a population is informed more quickly, and, as a consequence, who is able to make better decisions or be more productive.

Second, although research has linked network structure to node outcomes across several disciplines and phenomena, the theoretical mechanisms linking structure to outcomes have remained poorly understood. As we discover how different types of information and more broadly how other types of content flow through networks, we will be better able to understand the mechanisms linking social structure (e.g. strong or weak ties and brokerage positions) to outcomes of social and economic significance (e.g. job placement, productivity, innovation, health). Several fruitful avenues of research can help lead the way. For example, as described above, testing whether weak ties actually deliver more novel information can help validate a long-held assumption of sociological theories regarding (for example) the strength of weak ties and brokerage. At the same time, examination of the interaction of structure and content can illuminate even more nuanced propositions. For example, Centola and Macy (2007) contend that the diffusion of complex contagions requires exposure from multiple network contacts, because social affirmation and reinforcement are necessary to convince people to adopt risky, complex, or socially costly behaviors. However, Aral and Van Alstyne (2011) contend that a few strong ties may be more important to the spread of complex contagions, because information from trusted sources (rather than reinforcement from multiple weak ties) is more important to spreading risky or socially costly behaviors. Analysis of the content of information exchanges between such ties and the interaction of structure and content could help untangle the social processes that enable and constrain the diffusion of behaviors in society.

Third, it is important to understand how the shift of content flows from traditional offline networks to digital online networks alters outcomes such as product demand, the productivity of IT investments, the success of new technologies, and the structure of organizations. The shift from offline channels to online digital networks is also likely to reshape content flow in networks. Digital networks make social structure visible and thus are likely to alter actors' strategic decisions and actions, from social activity in online social networking sites, to buyer-supplier relationships in electronic markets, to expertise and knowledge transfer inside firms. Although little evidence exists comparing the flow of these types of resources online and offline, some recent work suggests that the explicit links created by online networks alter the flow of content (e.g. Oh et al. 2011, Goldenberg, Oestreicher-Singer and Reichman 2012). When network connections are made explicit, the choices individuals make about whom to share content with, whom to trade with, and whom to lend to may be affected both by the visibility of their choices and by the visibility of the connections amongst those with whom they choose to share or trade (Oestreicher-Singer and Sundararajan 2012a). A long tradition in IS and organizational communications research establishes a baseline of thought on how the richness of communication media affects information and knowledge transfers (Ngwenyama and Lee 1997, Oinas-Kukkonen et al. 2010). Information richness theory, for example, posits that richer modes of communication that provide multiple social cues through both natural language and body language help reduce equivocality (Daft and Lengel 1986, Chidambaram and Jones 1993). For this reason, richer communication media, such as face-to-face communication, have two important properties that help facilitate information transfers: the ability to transmit complex and tacit information and the ability to foster trust between actors. Face-to-face communication is therefore thought to have the greatest capacity to transfer complex knowledge (Roberts 2000, Wu et al 2008). These theories provide a great starting point to guide future work on how a shift to digital communication will continue to affect the flow of content in networks. IS scholars are well positioned to examine how the instantiation of social structure and network content in digital artifacts affects the four kinds of content flow we have described.

As we undertake such work, it is important to remember that networks are not merely pipes through which content flows but are also endogenous representations of realized movements of content between network nodes. The idea that networks are both “pipes and prisms” of the market highlights the importance of how connections between networked actors enable the generation of signals of quality and reputation (Podolny 2001). As Podolny (2001, 34) argues, “a tie between two market actors is not only … a pipe conveying resources … (but also) an informational cue on which actors rely to make inferences about the underlying quality of one or both of the market actors.” Such a view enables consideration not only of how resources flow through networks, but also of how networks reflect signals of quality and reputation by conveying information about the choices other actors have made in their relationship-formation decisions. For example, in their study of an online peer-to-peer lending market, Lin et al. (2009) find that “a borrower’s social network serves as a prism through which potential lenders deduce which borrowers to fund and at what interest rate” (Lin et al. 2009, 3). Interestingly, in this case, content is not ‘flowing through’ the network, but is ‘reflected by’ the flow of resources through the network. Additional insight could again be gained by examining the effect of more complex network structures beyond summary statistics such as the number of lenders or borrowers, for example, whether the network around a borrower or lender is cohesive or characterized by many mutual ties among contacts. At the same time, as content flows through a network, the structure of the network changes, creating an endogenous co-evolution of structure and content, which we discuss in more detail below in the section on network dynamics.

4. Network-based Inference

Over the past two decades, the confluence of tremendous decreases in price/performance for data storage, data processing, and data networking, along with the broad availability of data analysis technology, has produced a striking expansion of the influence of data on large-scale decision-making. Recently, increasing numbers of data science researchers and practitioners have focused their attention on improving decision-making by taking advantage of the sorts of "networked" data we have been discussing—formally, entities interconnected by links. This is important to managers because the information embedded in digitally represented social network data has been shown to substantially improve organizational performance in various important areas, for example: (i) the identification of fraud (Fawcett and Provost 1997, Cortes, Pregibon and Volinsky 2002, Hill, Agarwal et al. 2006), (ii) targeted marketing (Hill, Provost and Volinsky 2006, Martens and Provost 2011), (iii) online advertising (Provost et al. 2009), (iv) the management of customer attrition (Dasgupta et al. 2008), (v) the identification of "bad" brokers (Neville et al. 2005), (vi) suspicion scoring for counter-terrorism (Macskassy and Provost 2005), (vii) demand prediction for networked products (Dhar et al. 2012), and undoubtedly other business and government applications.

Network inference is the subject of IS research from two different perspectives, which are both vital and are necessarily intertwined. First, IS researchers design and evaluate new techniques for doing network inference in a business/organizational context (cf., Hevner et al. 2004). These contributions are often technical or statistical, and can have substantial value both to research and to business itself. Importantly, the business-oriented perspective of IS researchers produces results with a characteristically different flavor compared with more traditional computer science-oriented research on the same topics, and the editors of the top journals in IS have come together to state explicitly their support of publishing technically oriented design-science research in IS journals (Baskerville et al. 2010). In short, IS researchers tend to approach design science research with a significantly greater focus on aligning designs with the goals of business—which can lead to substantially different designs.[5]

Second, IS research should address network inference from a more general IS perspective, asking questions about whether network inference really provides business value, what the challenges are to actually using it, and what new strategic issues it introduces, such as the new business models it engenders, and the potential value of network data assets. This research landscape is little explored by researchers in any field, and IS researchers are best equipped to contribute because our combination of focus and deep understanding of technology allows us to avoid a superficial treatment of the technology, and at the same time our broad toolkit for research on business issues allows us to avoid a superficial treatment of value. These two perspectives on IS research on network inference cannot be separated cleanly, because as with the development of actual business value from network inference, success involves a tight interplay between new technical advances and the understanding of their business value.

To discuss IS research questions about network inference, we first need to understand network inference in a little more detail. Network inference techniques use the information represented by and embedded in the network to improve the estimation/prediction of values of important variables associated with the network. Specifically, most applications of network-directed inference focus on estimating/predicting the values of attributes of the interconnected entities. At the most basic level, the difference from traditional predictive inference is that in network data we can use information on other, linked entities to affect predictions. For example, knowing whether a consumer's social-network neighbors have purchased a product can significantly improve the prediction of whether the consumer in question will purchase the product (Hill, Provost and Volinsky 2006). Moreover, even if the consumer has no purchasers as neighbors, estimating that the <u>neighbors</u> are more likely to purchase because of <u>their</u> neighbors' proclivities (and so on) can improve predictions (Hill, Provost and Volinsky 2007). In a different setting, in a social network defined by calling behavior, observing that wireless phone accounts

---

[5] As an example, compare the active data acquisition research published in the IS literature (e.g., Saar-Tsechansky and Provost 2007) with the active learning work published in the computer science literature (e.g., Cohn et al. 1994).

two steps away in the network have previously been identified to be fraudulent increases the probability substantially that the focal account has been defrauded (Fawcett and Provost 1997).

Now let's consider three broad research questions where IS can and should be contributing to our understanding of network inference for business:

First, *how can network inference actually provide business value?* Above we list some specific applications of network inference providing business value, but the existence of a dozen examples is only a first step toward our understanding of how network inference can provide business value. This question can be approached from all IS research perspectives.

There are myriad design science questions revolving around how to do such inference well in business settings. For example, in many business settings information with high predictive value, such as consumer purchase, fraudulent behavior, brand affinity, etc., is often scarce. In addition, linkage in such social networks is quite sparse and local. This means that techniques that depend on direct relational autocorrelation in the network (e.g., they focus on entities whose neighbors have purchased) may be ineffective for the vast majority of entities in the network. Researchers have begun to introduce methods for: drawing network inferences even with sparse information, such as information propagation and other forms of collective inference (Macskassy and Provost 2007, Sen et al. 2008, Hill, Provost, Volinsky 2007); changing the link structure to better take advantage of the information (Macskassy and Provost, 2007; Gallagher et al. 2008); using link structure as a predictor (Henderson et al. 2011), etc. Another network inference challenge is to draw inferences about the links rather than about the entities. For example, "friend-links" on social networking sites are not nearly as predictive as friend relationships in traditional social networks, largely because friend-links actually incorporate a variety of relationships besides what traditionally would have been called "friends." Therefore, inferring attributes of links such as type (McCallum et al. 2007) or strength (Kahanda and Neville 2009, Xiang, Neville and Rogati 2010) holds promise for improving business outcomes, such as advertising targeting. In the extreme, inferring that the strength of a link *should* be high when currently it is zero (or missing) opens up an entirely new set of possible applications, often grouped as "link prediction" (Getoor et al. 2003, Liben-Nowell and

Kleinberg 2007, Clauset, Moore and Newman 2008). For example, link prediction could be used to recommend connections on a LinkedIn-style business-connections site or new friends on a social networking site. If we apply ideas of link prediction to bipartite graphs between consumers and products, we have a network perspective on recommendation systems: recommend products where the inferred link strength is high and the link is missing (link strength is zero). There also remain considerable design science questions around how to deal with the massive networks that businesses actually have, in contrast to the substantially smaller networks that are the subject of the vast majority of existing research on network inference. Unlike other data analysis applications, network data cannot simply be scaled down; we do not know how to sample network data effectively, and it is likely that the sampling needs to be designed specifically for the target business application.[6] Unfortunately, many of the sophisticated techniques offered by computer science researchers for inference in network data do not scale up to massive networks. These are just a few of the design science questions that IS researchers should be contributing to answering. Let's now step back and look at our question from a broader perspective.

Even though these network inference techniques offer much promise, and we have various cases of successful business applications, almost all research on network inference has been from the technical perspective. What about economics-oriented or behavioral perspectives on how network inference can provide business value? Given the history of research in IS, we should know that the interplay between the different IS research perspectives can give us a much richer understanding of the true applicability of such techniques. Technical research has important strengths, but as with research generally it makes assumptions about the scenario of use. For network inference, there has been little-to-no examination of these assumptions from the broader IS perspective. Are these techniques actually broadly useable? Is it even necessary to undertake the effort to mine massive networks? Does doing so really provide value in

[6] Traditional data sampling schemes used for building and evaluating statistical predictive models in non-networked data typically assume that entities are independently distributed—an assumption patently violated by networked data, and in fact the exact characteristic that we want to take advantage of for improved inference. Research is only beginning to try to understand how to sample networks so as to provide the best predictive inference and to evaluate whether indeed we have done so (Ahmed et al. 2010, Neville et al. 2011, Wang et al. 2011).

the organizational context? Existing technical research generally assumes that the networks are more or less static, but in fact they are dynamic, and the dynamics of structure and content are possibly (likely) related endogenously to the actions taken based on the network inference. Should we look at different metrics of evaluation than those that are examined in technical research (such as predictive accuracy)? What about user acceptance? Effects on customer loyalty, etc.? Since the examination of network inference is in its infancy, even the set of important questions itself should be the subject of IS research (and this is not the sort of question normally taken up in the purely technical communities).

More specifically, network inference is performed in the context of a constellation of business constraints. Very little research addresses issues regarding the actual economic and managerial settings in which network inference takes place. For example, research that simply assumes that data are freely available is not well aligned with many actual business settings, where data are costly. One may need to apply costly human resources, give consumers costly incentives, or incur other costs in order to obtain the requisite data for high-quality inference with networked data—which includes special concerns like acquiring costly data at specific locations in the network to improve inference (Rattigan, Maier and Jensen 2007, Bilgic and Getoor 2008, Macskassy 2009), and acquiring specific links between entities (Macskassy and Provost 2005). Managers face other concerns when using data for predictive inference, concerns that deserve more attention from IS researchers. For example, inference with social network data involves using data about some people to influence predictions about others. This raises important and serious privacy concerns, both ethical issues—what *should* organizations do—and possible ways of addressing privacy concerns, in terms of both policy and technique (Zhou, Pei and Luk 2008, Narayanan and Shmatikov 2009, Provost et al. 2009).

Second, *what theory underlies network inference, and (i) how can it help in understanding the application of network inference to business problems, e.g., where it should work, where it should not work, and (ii) what new techniques should be designed, as informed by the theory?* Data-driven inference works because of statistical dependencies, and statistical dependencies exist for various reasons. Social and behavioral theories help us to understand the reasons. Network inference in particular takes advantage of networks of statistical dependencies among entities. Technically these can be modeled as Bayesian networks (Pearl 1988), Markov networks (Pearl 1988), dependency networks (Heckerman et al. 2001), and relational derivatives thereof (e.g., Friedman et al. 1999, Neville and Jensen 2007). Networks of statistical dependencies specify the probabilistic/statistical relationships between the values of the random variables that we either know or would like to infer. Importantly, these networks of statistical dependencies are not the same as the observed data networks; just because there is a data link between two entities does not necessarily mean that there is a statistical dependency between the corresponding random variables for which we would like to infer values. For example, in fraud detection in wireless communications, one can form a data network based on who calls whom. However, practitioners have observed that there is not a strong statistical dependency between fraud status of such linked entities—which is what we would like to infer. Instead, there is a strong statistical dependency between the fraud status of entities two links away from each other. On the other hand, for targeted marketing there often is a very strong statistical dependency between purchasing propensities of linked entities.

The latter example illustrates one of the existing integrations of social theory and network inference procedures: it is now well accepted that the social principle of homophily (McPherson et al. 2001) creates network autocorrelation in the values of many variables we would like to infer. This very nice illustration notwithstanding, we need much more work on providing theoretical underpinnings for other sorts of network inference. As described elsewhere in this paper, IS researchers are helping to lead the effort to understand social influence in networks, which is a different mechanism from homophily for creating networked statistical dependencies, which network inference in turn can take advantage of. (But possibly different network inference techniques—which then broadens the field for IS design science researchers.) A different underlying theoretical reason for networked statistical dependencies is that there is an underlying group structure (Girvan and Newman 2002, Clauset, Newman and Moore 2004, Neville and Jensen 2005); people are similar within the groups, and they also are linked by social relationships within the groups. This is a subtly different theoretical reason from homophily or social influence, more akin to theories of affiliation networks (Breiger 1974). Understanding network inference in markets and organizational settings requires us to understand the theoretical reasons for the networks of statistical dependencies and the corresponding structures of these networks.

Existing network influence work focuses largely on the case where the data network (the links we observe, such as communication links) is aligned with the network of statistical dependencies. In practice, we see plenty of cases where the two networks are not so well aligned. The fraud example above is one. Another is the case of bipartite networks between consumers and other entities, such as products/services, webpages, etc. Design science researchers press forward with designing methods for network inference in these scenarios, with impressive results (e.g., Provost et al. 2009, Martens and Provost 2011). This research is important and useful in its own right; however, IS research can contribute much more deeply by creating (and debating) theories to understand this sort of network inference. For example, we need to draw together the technical work on such bipartite network inference with theories of consumer behavior that explain how the consumer choices that lead to the links in the network are based on deeper (latent) tastes, interests, socioeconomic status and constraints, etc. Thus, we might conjecture that network inference produces a proxy for these unobserved factors, and that is why it is effective for estimating (for example) brand affinity (Provost et al. 2009) or product affinity (Martens and Provost 2011). Generally, with our broad set of research approaches and ties into multiple relevant reference disciplines, IS researchers are uniquely positioned to contribute to our deep understanding of network inference.

Third, *what strategic lessons arise from the careful examination of network inference techniques?* As discussed above, we know that network inference is effective in at least some business applications. Nonetheless, there is very little research providing guidance at a strategic level regarding the use (or avoidance) of network inference in different business settings. For example, what data should firms be gathering? Hill, Provost and Volinsky (2006) suggest that the striking results they present using consumer networks based on observed telecommunication links would also apply to consumer networks based on other forms of linkage, such as emails, instant messages, etc. Aral et al. (2009) indeed later show that the same sort of results indeed are observed for networks based on instant messages. This is interesting and important, but is not quite sufficient for making strategic recommendations to firms. In order for such recommendations to be robust, IS should bring all its research tools to bear: What are the implications from the economic and behavioral perspectives? How do they extend or qualify the vanilla recommendation that companies should be building consumer networks however they can, and then using them to target offers (for example)?

Network inference also introduces deeper strategic questions. For example, some firms have access to much larger networks than others do, either because of the sheer size of these firms (consider the consumer network observed by an AT&T or Verizon, or the consumer/service network observed by a CitiBank or American Express) or because of a particular large-scale data business (consider the consumer/content network observed by an online advertising exchange). We are beginning to see evidence that for network inference, large-network data can be a tremendously valuable data asset—and, importantly, the performance increases to scale with network data outstrip the increases to scale with traditional data-driven inference (Martens and Provost 2011). If such results hold generally, this implies that firms with larger data networks (often, larger firms) can get substantial competitive value from their data assets, as compared to smaller competitors. This observation can feed back to strategic decisions about data asset curation and management, as well as strategies for startups in new business areas where network inference will play a role (e.g., invest early in building very large data networks).

Finally, completely new business approaches can be built by understanding network inference techniques. Google is an example, as one of their main innovations was to use network inference to infer webpage quality, rather than using just the inherent (local) attributes of the page. As network inference enters the consciousness of more firms, we see new ways of using social networks to improve existing applications, for example, to target offers (as discussed above), and to manage fraud and customer attrition (look at whether people are connected to fraudsters/churners, (Fawcett and Provost 1997, Dasgupta et al. 2008), or whether there is an negative influencer in a customer's social circle (Richter et al. 2010). We also see brand new businesses entering and becoming successful quickly, based on the application of network inference techniques (e.g., Media6Degrees and 33Across have seen rapid growth and success very quickly by applying network inference to online ad targeting). Further, we see promise emerging in other areas, such as mobile marketing (Provost 2011) and banking (Martens and Provost 2011). IS research would do well to study more systematically where and how network inference can and should be applied, in order to give solid strategy recommendations both to existing firms and to entrepreneurs.

5. Network Dynamics

The final stream focuses on the dynamics of network evolution. After all, a fundamental underlying characteristic of digital networks—online interaction networks, digital organizational networks and product networks—is that they are not static but rather evolve over time. Much of the existing research, including that described above, includes an inherent assumption that the network under study is static, an assumption sometimes made simply because of data availability or complexity considerations. For example, when modeling the dynamic process of diffusion, researchers often assume that the network of social ties is static during the diffusion period (Muller, Peres and Mahajan 2009). Other research takes into account changes over time by using snapshots of the network in a panel setting, exploiting changes to the network structure or changes in outcome. However, the study of network dynamics – how networks are created, evolve and dissolve, and the ways in which IT affects those dynamic processes – has received scant attention from information systems researchers and warrants closer investigation (see also Trier, 2008). Two recent exceptions are by Lu et al. (2011), who use a network growth paradigm to study the emergence of opinion leaders in online review communities (such as Epinions.com), and by Kossinets and Watts (2006), who study the dynamics of an email network.

The formation and evolution of social networks have been researched extensively, in the sociology, computer science and physics literatures, focusing on describing the process by which nodes and edges are added or dropped and generative statistical models of network formation. Though a comprehensive review of this literature is beyond the scope of this paper, several observations highlight the opportunity for IS research. First, new forms of interaction and new sources of data on the evolution of social and economic networks are being made available through technology-mediated interaction and communication, creating new opportunities for researchers with access to these data and a theoretical background for interpreting and anticipating how interactions and communication are enabled and constrained by information technology. Second, most current models are typically mathematical abstractions with limited theoretical motivation from the social sciences, and they do not fully account for strategic, economic and behavioral considerations. Opportunities therefore exist to advance current thinking by more deeply considering social and economic motivations that may guide network evolution. Third, current models typically assume full information (that is, that the new node "knows" the existing degree distribution across the entire graph). These models are especially appropriate for open, growing networks, such as the World Wide Web or a social network. However, some cases include relatively closed systems, where new edges are added between existing nodes (Robins, Pattison and Woolcock 2005). Further, Faraj and Johnson (2010) have found contradicting observations in the context of online communities as well. Researchers from sociology have focused on the need to understand the underlying social processes that drive network dynamics, such as the tendency for reciprocity, transitivity or the need for group balance (see Doreian and Stokman (1997) for a review). In this context, the focus is primarily on the individual actor, making a *choice*. Further, such models have traditionally been developed to explain human interaction, and have focused mainly on interactions between individual actors. Extending this research to interaction between organizations, products, or online content ‘objects’ is another promising avenue for researchers to explore. IS researchers are perhaps uniquely qualified to provide insights into these questions and extend current theories to the context of technology-mediated communication.

One particularly important avenue for IS research is to understand the extent to which the known processes of offline network evolution are influenced (or altered) by existing IT artifacts. The transition from offline social interaction to online social networking sites such as Facebook has changed social interaction and has therefore potentially altered the process by which social networks evolve. For instance, as mentioned earlier, previous research on offline networks has demonstrated a tendency toward reciprocity and transitivity in forming social ties. In online social interaction our social contacts are visible. We can therefore see our friends' friends; this visible information may influence our future linking decisions. Will this new available information increase tendencies toward reciprocity and transitivity or reduce them? Moreover, many websites currently recommend links to consumers on the basis of shared interests or consumption patterns or simply on the basis of existing ties among users. Will such recommendations result in new network structures? Will those recommendations change online network evolution and make it fundamentally and systematically different compared with offline settings? Will they enhance well-known social network characteristics, such as homophily or triadic closure? Oestreicher-Singer and Sundararajan (2012a) have shown that visible product networks significantly increase existing correlations in demand between complementary products. As the transition to electronic commerce continues and the use of such networks increases—could such networks alter fundamental patterns of demand? The inclusion of strategic and economic considerations in models of network evolution is therefore particularly important in cases where IT artifacts alter available information and as a result behavior. Models that do not take such considerations into account are likely to be less accurate.

A second important avenue of research in which the dynamics of network formation are of particular interest is strategic link formation. In some contexts, network formation is the result of strategic economic decisions. One example is the strategic placement of hyperlinks between content websites. The tradeoff between investment in producing new content and investment in linking to other high-quality content has been studied by Dellarocas, Katona and Rand (2009) in the context of online publishing and by Mayzlin and Yoganarasimhan (2011) in the context of competing blogs. Content websites are faced with the option to link to other content websites and thus reduce the cost of content production while simultaneously increasing the risk of losing consumers. Free riding and its effect on competition and public policy should also be considered in this context. Similarly, Katona and Sarvary (2008) model the evolution of the commercial World Wide Web as a function of strategic linking (advertising on a different page) and link-price decisions of web pages. A useful theoretical basis for these studies may be found in the recent literature on "network games" (Bala and Goyal 2003, Bramoulle and Kranton 2007, Galeotti et al. 2010, Sundararajan 2007), which examines how the properties of the equilibria of specific classes of IT-related games, played on a graph, depend on network structure.

Third, understanding the life cycle of online social networks is another promising direction. Following Section 2, any specific online social network (such as Facebook or LinkedIn) is an IT artifact, which can also be thought of as a product. How do such products emerge, and under what circumstances do they dissolve? Are dissolved networks replaced by other social networks? How does substitution between networks occur? Is there a generic life cycle for social networks? What is the life cycle of a network, and how is it different from the life cycle of other products? To date, researchers have mainly studied the emergence of social networks. Katona, Zubcsek and Sarvary (2009) study the diffusion of a social network, modeling the adoption decision of potential members. They find that the adoption decision is influenced by the decisions of others, as well as by the local structure (degree as well as clustering) that is being formed. Stephen and Toubia (2010) study the evolution of links between consumers in an online marketplace, focusing on the structural properties of the resulting network. One might also be interested in studying the process whereby specific roles are created within a network. For example, Lu et al. (2011) study the process of emergence of opinion leaders in online review communities, and Palla, Barabasi and Vicsek (2007) study the formation of groups and their dynamics.

The above mentioned papers study the endogenous process of network formation. A fourth question of interest to the business community is the effect of exogenous events on the resulting network structure. That is, how do sudden, exogenous economic events affect network structure and stability? To what extent do those exogenously created shocks spill over to neighboring nodes? For example, a marketing campaign featuring a certain product may generate an increase in demand for neighboring products (Carmi, Oestreicher-Singer and Sundararajan 2012a); changes in ownership of one firm may spill over to other firms in its "network," or increased attention to one blog can affect other blogs in the blogosphere. Exogenous shocks can affect the entire network rather than a specific node and can change the global network structure. This kind of research—studying the diffusion of multiple sequential exogenous shocks—is now more feasible because the underlying network is made visible by the IT artifact. For example, Trier (2008) documents structural changes to the organizational email network around the time the Enron scandal was made public; Burkhardt and Brass (1990) study the change in the organizational social network following the adoption of a new technology, and Tucker (2008) examines how an exogenous shock to demand for video-conferencing technologies during the World Cup affects the demand for network neighbors. Again, technology-mediated networks often provide us with an opportunity to investigate the effects of such events on an entire network, with the aim of fully understanding and quantifying these effects. As IT makes the interconnections between entities increasingly visible, it provides scholars with fertile ground for future research.

Finally, a nascent stream of research has observed that in certain kinds of interaction networks, edges are inherently transitory because the opportunity for interaction exists only for a specific period of time. The typical approach is to model such networks as an evolving sequence of graphs (Clauset and Eagle 2007), although there is also evidence that this kind of representation is subject to bias, and fundamentally incomplete (Scellato et al. 2010). The focus of these studies has often been on "ad-hoc networks" created by IT-based interaction (for example, networks of interaction between mobile devices using Bluetooth). However, all social and economic networks fundamentally have this kind of "evolving edge" property (relating either to the existence of edges or to the strength of the tie / bandwidth of channel implied by the edge) (see Aral and Van Alstyne 2011). Furthermore, as human interaction becomes increasingly digital, it is likely that this non-stationarity will be more pronounced. IS scholars have an especially promising opportunity to participate in the development of new models that capture the dynamics of IT-enabled networks over time in order to shed light on the social and economic implications of this growing non-stationarity (see, for example, Trier (2008)).

6. Cross-Cutting Issues

Several fundamental cross-cutting ideas are pertinent to the areas of research that are central to the interface of information systems and networks. In this section, we briefly describe three of these, namely: the structural properties of networks and their informativeness across contexts, the role of underlying social and economic theories in explaining and predicting networked behaviors, and the importance of a rigorous approach to causal inference using networked data.

6.1. Structural Properties of Networks

The exact structure of a network may (i) play an important role in moderating the impact of digital networks on social and economic outcomes, (ii) determine the nature of flows of different kinds of content, (iii) be of consequence for what predictive modeling strategies are most effective, and (iv) shape the path by which the network evolves. However, graphs are complex objects, and as a consequence, networks are typically summarized by a number of different (and simpler) structural properties. (For definitions of different network properties and measures, see Newman (2003).)

Perhaps the most widely reported structural characteristic pertaining to networks is that their degree distribution often approximately follows a *power law*. This striking empirical regularity—that many empirical networks are scale-free – persists across a wide variety of networks and has been observed in some more recently analyzed digital networks such as Yahoo!'s IM network and Amazon's co-purchase networks (Aral, Muchnik and Sundararajan 2009; Oestreicher-Singer and Sundararajan, 2012a). Documenting this regularity is important because it changes our notion of what is "normal" and how to interpret summary statistics. For researchers and managers used to Gaussian and lognormal distributions (for which the mean and the standard deviation are meaningful statistics that we have learned to interpret in a specific way), it is important to understand that averages and variances do not mean much if one's data are distributed according to a power law.

Another widely investigated structural characteristic of networks is the extent to which the network is *clustered*, which, roughly speaking, signifies the fraction of neighbors of a node that are also neighbors of each other. As social networks become digitally visible, they may become more clustered, perhaps changing the cohesiveness of groups of friends, or alternatively, the interpretation that triadic links are "stronger." There is evidence that clustering changes the nature of content and resource flows within a network (for example, the flow of attention to products on Amazon.com in Carmi et al. 2009).

Similarly, one would expect clustering to impact the choice of predictive modeling technique: many links among the same nodes should reinforce autocorrelation-based techniques that use collective inference.[7]

Not all nodes are created equal in networks; some have better "positions" than others. Position often is quantified by a variety of measures of node centrality. Node centrality is likely to affect some of the design issues discussed in Section 2 (especially those relating to socio-technical design that involve seeding), is related to how digital networks might alter economic outcomes (for example, more "centrally located" product categories have been shown by Oestreicher-Singer and Sundararajan (2012b) to have flatter demand distributions), can determine the nature of flows in networks, and can help predict "importance" in networks. Such measures of importance are best known as the basis for Google's PageRank algorithm; recently they are being used for social-media analytics, for example, to estimate which bloggers are authoritative—with the inference being that they are therefore influential and that firms should monitor them for positive or negative mentions of their brands/products/services, or even try to influence them to propagate the firms' messages (Melville et al. 2010). A related kind of "position" property that has received widespread attention is whether a link is a "bridge" between two otherwise distinct components, and in a related sense, whether a node fills a "structural hole" (Burt 1992). A number of studies have related this kind of bridging role to organizational power (Burkhardt and Brass 1990); recent evidence suggests that it is central in indicating the informativeness of content transmitted in digital email networks (Aral and Van Alstyne 2011), and that users in decentralized content networks who play this kind of bridging role alter search and consumption patterns in an economically significant way (Goldenberg, Oestreicher-Singer and Reichman 2012). Similarly, Ravindran, Susarla and Gurbaxani (2009) find that structural embeddedness in a network plays an important role in determining the duration of IT outsourcing contracts.

Structural properties of networks are widely used across different disciplines and may even be independently discovered across disciplines. There may be interesting applications of measures used in

[7] There has been work showing that whether or not the known labels are clustered in the network affects the performance of different techniques (Xiang and Neville 2008).

one kind of investigation to problems in others; for example, Kleinberg (1999) and Page et al. (1998) built on the measures of centrality of Katz (1953) in developing their respective metrics of web page importance (and also for the blogger authority metrics mentioned above). Some of these structural properties form a common basic "language" for networks researchers from different fields, who may have very different reference disciplines and research goals. As researchers in IS address questions across the different streams we have highlighted, the judicious use of this shared language could facilitate greater cross-disciplinary understanding and spillover (see Trier (2008) for such an attempt).

6.2. Social and Economic Theory about Networked Behavior

A substantial fraction of research in what is being called "network science," especially those studies that analyze large networked data sets, is only minimally informed by theories from the social sciences, adopting instead what one might call the "physics" approach to modeling and analyzing data (Newman 2003). While there is a fairly rich tradition of research from management (for example, the work of Ron Burt) and sociology (Granovetter 1973 and the research that followed) that has analyzed social networks through the lens of behavioral theories, these disciplines have been relatively slow to take advantage of the information contained in the massive digital network artifacts that are now emerging (for a review of research in social sciences using social network analysis see Borgatti et al. (2009)). Consequently, there is a tremendous opportunity for IS researchers to combine their expertise from the different "sciences of the artificial" to make fundamental contributions to explaining networked behaviors in these massive data sets that are informed by sociology, social psychology, economics and data science. These theoretical underpinnings will be central in ensuring that the relevance of our research into digital, economic and social networks has longevity, and this approach is also aligned well with the cumulative tradition of IS as a field.

For example, a widely documented empirical regularity in networks is that nodes are assortatively mixed by (shared) individual characteristics or group membership—that there is "homophily" in human social networks (McPherson, Smith-Lovin and Cook 2001), or that "birds of a feather flock together," a point first noted by Burton (1638). However, the determinants of homophily may be varied. An important distinction alluded to by Hill, Provost and Volinsky (2006) and subsequently discussed by Jackson (2008) is the distinction between homophily due to actor opportunity and homophily due to actor choice, since these suggest fairly different underlying sociological or economic mechanisms. Given the prevalence of assortative mixing in digital networks that are both social (for example, Hill et al, 2006, Aral et al.,2009) and economic (Oestreicher-Singer and Sundararajan 2012a), and the importance of homophily as an alternative explanation for network flows and a key explanation for the autocorrelation in both empirical and auxiliary networks used for prediction (Macskassy and Provost 2007), more nuanced theory at the interface of economics and sociology seems warranted.

Similarly, as discussed in Section 5, many popular models of network formation and evolution build theory using reduced-form mathematical models of network evolution (Barabasi and Albert 1999, Leskovec et al. 2005). While these are useful statistical approximations of the processes generating the networks, the questions that businesses and IS researchers might be most interested in may demand an underlying mechanism that is better informed by the social sciences. A first step in this direction may be found in models of strategic link formation (Bala and Goyal 2003), perhaps enriched by appropriate theories from sociology and social psychology. Likewise, many well-regarded models of diffusion in networks (for example, Pastor-Satorras and Vespignani 2001) model it as following a simplified stochastic process on a graph, ignoring the fact that nodes may be making strategic choices, or that diffusion may be influenced by social and economic considerations. The drawbacks of this approach become most apparent when one attempts to adapt these mathematical models to the context of, say, the spread of a product due to viral marketing, or the flow of decision information in an organizational network. The gap has been filled in part by recent work from Galeotti and Goyal (2009), López-Pintado (2008), and Jackson and Yariv (2005), as well as work that uses sociological theory to inform dynamic models where actors simultaneously maximize network and behavioral utility functions (Snijders, Steglich and Schweinberger 2006). Still, a simple model of diffusion grounded in the social sciences remains an important future contribution, one perhaps well-suited for IS research.

6.3. Causality and Identification in Networks

Attention to the estimation of cause and effect in network studies is essential to creating knowledge that is robust to alternative scientific explanations and relevant to policy. Networks are important because they explain patterns of social and economic outcomes across interconnected populations. Rigorous treatment of identification ensures that we avoid mistaking confounding factors and spurious correlations for causal relationships that can form the basis of managerial action.

The difficulty of identifying endogenous social effects, popularly referred to as the reflection problem (Manski 1993), is a critical impediment to the formulation of effective social and managerial policy in networked settings.[8] Several sources of bias can confound causal statistical estimation in networks, including simultaneity (Godes and Mayzlin 2004), unobserved heterogeneity (Van den Bulte and Lilien 2001), homophily (Aral et al. 2009), time-varying factors (Van den Bulte and Lilien 2001), and other contextual and correlated effects (Manski 1993). It is well known that these factors can complicate the identification of endogenous social effects and cause significant upward bias in estimates of social contagion and other peer effects (Aral et al. 2009). Approaches to identification in networked studies will therefore be essential to the development of rigorous scientific results that can effectively guide policy.

A new line of research is emerging that uses randomized trials to identify peer influence and contagion in social networks (Aral and Walker 2011a, b, Aral and Walker 2012, La Fond and Neville 2010, Oktay, Taylor and Jensen 2010, Goldenberg, Oestreicher-Singer and Reichman 2012, Aral and Taylor 2011). Randomization is an effective method for identifying social effects because unobserved heterogeneity and a lack of truly exogenous variation limit the ability of more traditional identification techniques to cleanly estimate causal peer influence in networks (Falk and Heckman 2009). The logic of randomization is simple. Since in reality individuals who are exposed to a treatment typically differ from those who are not, comparing the treated to the untreated without random assignment of the treatment creates a selection bias that reflects differences in the potential (untreated) outcomes of treatment and

[8] Technically, the “Reflection Problem” is but one source of difficulty in identifying peer effects, as we discuss below.

comparison groups. Randomization solves this problem because individuals assigned to the treatment and control groups differ in expectation only through their exposure to the treatment (Duflo, Glennerster and Kremer 2006). In networked settings, randomization is potentially more complicated, as controlling the entire social environment of a networked experimental subject is difficult. Aral and Walker (2011a, b), for instance, describe an "inside-out" strategy for estimating contagion effects in networked randomized trials. The conventional approach to estimating peer influence and social contagion involves estimating the influence of an individual's social environment 'inward' on the individual's own behavior. However, experimental analysis is difficult in this setting because comprehensively controlling the network environment of each user in the study is typically infeasible at scale. The inside-out approach described by Aral and Walker treats a user and observes the effect of treatment "outward" on the outcomes of the user's peers; the peers' social environment thereby is controlled (see Aral and Walker 2011 a,b for more details). At the same time, networked environments increase the risk of contamination, leakage or interference in network experiments (Aral and Walker 2011b, Aronow and Sammi 2012), which in turn create violations of the Stable Unit Treatment Value Assumption (SUTVA) that can bias inference. Future work modeling the degree to which contamination or leakage in networked experiments affects inference and estimation strategies that avoid bias from interference will therefore be essential.

While randomization can provide a clean causal estimation strategy, the vast majority of data available to firms and governmental organizations remains observational, making the improved understanding of causal peer influence estimation in such data critical to our knowledge of what drives behavioral contagions in social networks, and of how we might attempt to promote or contain such contagions. Several approaches to the identification of peer effects in observational data have been proposed in various literatures including peer effects models (e.g. Bramoulle, Djebbari and Fortin 2009, Oestreicher-Singer and Sundararajan 2012a), actor-oriented models (e.g. Snijders et al. 2006), instrumental-variable methods based on natural experiments (e.g. Sacerdote 2001, Tucker 2008), dynamic matched sample estimation (Aral et al. 2009), structural models (e.g. Ghose and Han 2011), and ad hoc approaches (Christakis and Fowler 2007).

We encourage econometric rigor before advancing causal claims based on networked data. However, it is also important to recognize that although correlation does not imply causation, there is typically no causation without correlation. The myriad sources of endogeneity in networked data which might preclude making strong causal statements should not unduly impede the reporting of interesting patterns of network correlation, such as patterns of assortative mixing in new digital networks. These patterns, if carefully assessed, may provide useful business insights by themselves even if the mechanisms generating them are not fully explained, and further, their timely publication could encourage additional cumulative scientific investigation; follow-up studies could lead to important insight into the causal mechanisms (or lack thereof). Furthermore, even without a causal mechanism, such results could have impact on the research and practice of predictive modeling, as in many cases predictive modeling can be strikingly effective based on correlation alone.[9] An excessive focus on causation could be detrimental to progress in this exciting new area. Editorial boards should be judicious in their balance between demanding empirical rigor and facilitating the communication of interesting new discoveries from rapidly growing digital networked data, especially at this nascent stage in the area's evolution.

Finally, it is important to consider exactly what it means for someone to influence or be influenced by his or her peers, and what theoretical conceptualizations of peer influence mean for our ability to identify it in observational data. For example, if we conceptualize peer influence as how peer behaviors change one's expected utility and thus the likelihood that or extent to which one will engage in a certain behavior, then we necessarily define influence as causal and exclude correlated and confounding effects, making causal estimation essential to peer influence identification. For example, highly central individuals or individuals of high degree are not necessarily influential by this definition. In order to be influential, individuals must cause behavior change in the network, whether by changing peers' information and awareness or changing their preferences, rather than simply being connected to or passing information on to a significant number of people without changing their behavior. This point is

[9] Presuming the requisite care and diligence to ensure generalizability.

discussed in detail in Aral (2011). Similarly, Sundararajan (2012) discusses how measurement issues might be leading to an excessive focus on a form of peer influence that results simply from the refocusing of attention of boundedly rational agents on choices made by their peers, rather than on contexts where the behavioral changes are due to genuine "influence" or persuasion, or actual consumption complementarity. Of course, it is also possible that there are endogenous effects or changes in behaviors on account of peer characteristics as well as their behaviors (Manski 1993), in which case, the ways in which characteristics and behaviors of connected entities aggregate into the theorized influence (for example, is it the sum of one's neighbors' behaviors or their average which matters) can be a critical determinant of whether the peer effect is identified, as illustrated by the econometric model of Oestreicher-Singer and Sundararajan (2012a). Defining peer influence clearly is therefore an essential precursor to identifying it in networked data.

7. Concluding Remarks

This is an exciting time for the social and data sciences. The availability of massive, networked electronic data that contain information about individual-level connections among people, products, web content, and other entities provides researchers with an unprecedented microscopic view into the nature of commercial and social interaction. Digital networked data are revolutionizing empirical research in the social sciences in the same way that the microscope revolutionized empirical research in the biological sciences. We hope that this commentary provides a much-needed roadmap for future research focused on digital, economic and social networks in information systems. As we have discussed, a wide range of economic, business and organizational outcomes are determined by the creation, evolution and informativeness of these networks and the content that flows through them. This makes a compelling argument for systematically integrating the presence and role of these networks into the development of new theory.

Beyond these broader scientific objectives, research into information in networks has a number of more pragmatic and business-oriented research goals. Certain aspects of formulating strategy can benefit from a better understanding of the associated "networks." For example, many marketing efforts today exploit either a digital network or an underlying social network to attract and retain customers. The design of teams and, more broadly, of organizations is informed by the analysis of email networks representing employee interaction and collaboration patterns. A growing fraction of organizational knowledge is encapsulated in such networks, making them central to effective knowledge management strategy. Further, an increasing number of corporate information systems in modern organizations are characterized by some kind of underlying digital network, and answering some of the research questions we have raised will contribute towards superior design of these systems. In each of these examples, a better understanding of each of the four topics we have highlighted – how IT creates and reveals these networks and the associated design issues, how content flows through these networks, how networked data can be used for prediction and inference, and how these networks evolve over time – might further the sophistication of a firm's strategic thinking.

Given that these networks are constructed because of growing interaction (commercial, social or organizational) that is mediated by information technologies, network research is especially suited for our field. It is not surprising that many of the initial answers to these questions in all four areas are being provided by IS researchers, and that this line of IS research is gaining visibility in a number of other disciplines. A final issue to keep in mind in discussing the benefits of the emergence of digital artifacts that contain and reveal increasingly detailed information about our interactions relates to *information privacy* in networks. Among the many research questions in this domain, one of especially critical policy importance is, who should own the personal data contributed by users and contained in these massive digital network artifacts? Current law transfers ownership to the platform provider once any content is posted by a user. It may be more reasonable, however, for a user to maintain ownership of some of these data as their own "intellectual property," and for a legal statute to determine a standard, limited transfer of rights to the platform and perhaps to the other users of the platform or the platform's partners. Pragmatic firms might meanwhile follow the prescription of "aligning intent with use" (Dhar, Hseih and Sundararajan, 2011). While this issue arises for any form of digital data shared as a by-product of electronic interaction, it is exacerbated in networks because of the explicit sharing of data with network neighbors, and the implicit rights thereby conferred to the members of local networks.

References

Agarwal, R., A. K. Gupta, R. Kraut. 2008. Editorial overview: The interplay between digital and social networks. *Inform. Syst. Res.* **19**(3) 243–252.

Ahmed, N., J. Neville, R. Kompella. 2010. Reconsidering the foundations of network sampling. *Proceedings of the 2nd Workshop on Information in Networks*.

Ahuja, M. K., K. M. Carley. 1998. Network structure and virtual organizations. *Journal of Computer Mediated Communication* **3**(4).

Angst, C. M., R. Agarwal, V. Sambamurthy, K. Kelley. 2011. Propensity, susceptibility, infectiousness, and proximity: Predicting organizational diffusion of electronic medical records. *Manage. Sci.* **57**(1) 1077-1093.

Aral, S. 2011. Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 217-223.

Aral, S., E. Brynjolfsson, M. Van Alstyne. 2007. Productivity effects of information diffusion in networks. *Proceedings of the 28th Annual International Conference on Information Systems*. Montreal, CA.

Aral, S., E. Brynjolfsson, M. Van Alstyne. 2008. "Sharing mental models: Antecedents and consequences of mutual knowledge in teams." Working paper, NYU Stern School of Business. Available at SSRN: http://ssrn.com/abstract=1299260.

Aral, S., Brynjolfsson, E. & Van Alstyne, M. 2012. "Information, Technology and Information Worker Productivity." *Information Systems Research*, Articles in Advance, March 12: 1-19.

Aral, S. & David, V. 2012. "The Anatomy and Dynamics of Vision Advantages." *Proceedings of the 33rd Annual International Conference on Information Systems*, Orlando, FL.

Aral, S., L. Muchnik, A. Sundararajan. 2009. Distinguishing influence based contagion from homophily driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci.* U S A, 106(51) 21544–21549.

Aral, S., L. Muchnik, A. Sundararajan. 2012. Engineering social contagion: optimal network seeding and incentive strategies. Working Paper, NYU Stern.

Aral, S., S. Taylor. 2011. Viral incentive systems: A randomized field experiment. *Workshop on Information Systems Economics (WISE)*; Shanghai, China.

Aral, S., M. Van Alstyne. 2011. The diversity-bandwidth tradeoff. *American Journal of Sociology* **117**(1) 90–171.

Aral, S., D. Walker. 2011a. Creating social contagion through viral product design: A randomized trial of peer influence in networks. . *Manage. Sci.* 57(9) 1623-1639.

Aral, S. & Walker, D. 2011b. "Identifying Social Influence in Networks Using Randomized Experiments." *IEEE Intelligent Systems*, September-October, 26(5): 91-96.

Aral, S., & Walker, D. 2012. "Identifying Influential and Susceptible Members of Social Networks." *Science*, July 20: 337-341.

Aronow, P. M. & Samii, C. 2012. "Estimating Average Causal Effects Under General Interference," *New York University Working Paper*.

Arriaga, M., N. Levina. 2008. Social dynamics in online cultural fields. ICIS 2008 Proceedings, Paper 120. http://aisel.aisnet.org/icis2008/120.

Baker, W. E. 1984. The social structure of a national securities market. *Am. J. Sociol.* **89** 775–811.

Bala, V., S. Goyal. 2003. A noncooperative model of network formation. *Econometrica* **68**(50) 1181-1229.

Barabasi, A. L., R. Albert. 1999. Emergence of scaling in random networks. *Science* **286**(5439) 509–512.

Baskerville, R., K. Lyytinen, V. Sambamurthy, D. Straub. 2011. A response to the design oriented information systems research memorandum. *Euro. J. of . Inform. Syst.* **20** 11–15.

Bikhchandani, S. D., D. Hirshleifer, I. Welch. 1991. A theory of fads, fashions, custom and cultural change as information cascades. *J. Politi. Econ.* **100** 215–227.

Bilgic, M., L. Getoor. 2008. Effective label acquisition for collective classification. *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 43–51.

Borgatti, S. P., A. Mehra, D. J. Brass, G. Labianca. 2009. Network analysis in the social sciences. *Science* **323**(5916) 892–895.

Bramoulle, Y., H. Djebbari, B. Fortin. 2009. Identification of peer effects through social networks. *J. Econometrics* **150**(1) 41–55.

Bramoulle, Y., R. Kranton. 2007. Public goods in networks. *J. Econ. Theory* **127**(1) 478–494.

Breiger, R. L. 1974. The duality of persons and groups. *Social Forces* **53** 181-190.

Burkhardt, M. E., D. J. Brass. 1990. Changing patterns or patterns of change: The effects of a change in technology on social network structure and power. *Administrative Science Quarterly* **35**(1) 104-127.

Burt, R. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Burt, R. 2008. Information and structural holes: Comment on Reagans and Zuckerman. *Ind. Corp. Change* **17**(5) 953–969.

Burton, R. 1638. *The Anatomy of Melancholy*. Farrar and Rinehart, New York, NY (1927 ed.).

Butts, C. 2008. A relational event framework for social action. *Sociological Methodology* **38**(1) 155–200.

Carley, K. M. 1995. Communication technologies and their effect on cultural homogeneity, consensus, and the diffusion of new ideas. *Sociological Perspectives* **38**(4) 547–571.

Carley, K. 1997. Extracting team mental models through textual analysis. *J. Organ. Behav.* **18** 533–558.

Carlsson, S. A., O. A. El Sawy, I. Eriksson, R. A. 1996. Gaining competitive advantage through shared knowledge creation: In search of a new design theory for strategic information systems. *ECIS 1996 Proceedings.*

Carmi, E., G. Oestreicher-Singer, A. Sundararajan. 2009. Spreading the Oprah effect: The diffusion of demand shocks in a recommendation network. *Proceedings of the INFORMS Conference on Information Systems and Technology (CIST 2009)*.

Centola, D., M. Macy. 2007. Complex contagions and the weakness of long ties. *American Journal of Sociology* **113**(November) 702–34.

Chidambaram, L., B. Jones. 1993. Impact of communication medium and computer support on group perceptions and performance: A comparison of face-to-face and dispersed meetings. *MIS Quarterly* **17**(4) 465-491.

Christakis, N. A., J. H. Fowler. 2007. The spread of obesity in a large social network over 32 years. *New Engl. J. Med.* **357**(4) 370–379.

Clauset, A., N. Eagle. 2007. Persistence and periodicity in a dynamic proximity network. DIMACS Workshop on Computational Methods for Dynamic Interaction Networks. New Brunswick, NJ.

Clauset, A., C. Moore, M. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191) 98–101.

Clauset, A., M. Newman, C. Moore. 2004. Finding community structure in very large networks. *Phys. Rev. E* **70**(6) 66111.

Cohn, D., L. Atlas, R. Ladner. 1994. Improving generalization with active learning. *Machine Learning* **15**(2) 201-221.

Coleman, J. S. 1988. Social capital in the creation of human capital. *Am. J. Sociol.* **94** S95–S120.

Cortes, C., D. Pregibon, C. Volinsky. 2002. Communities of interest. *Intell. Data Anal.* **6**(3) 211–219.

Cross, R., J. N. Cummings, 2004. Tie and network correlates of individual performance in knowledge-intensive work. *Acad. Manage. J.* **47**(6) 928–937.

Cross, R., L. Sproull. 2004. More than an answer: Information relationships for actionable knowledge. *Organ. Sci.* **15**(4) 446-462.

Daft, R. L., R. H. Lengel. 1986. Organizational information requirements, media Richness and structural design. *Management Science* **32**(5) 554-571.

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. Nanavati, A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. *Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology*. ACM, New York, NY, 668–677.

Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Manage*. *Sci*. **52**(10) 1577–1593.
Dellarocas, C., Z. Katona, W. Rand. 2009. Media, aggregators and the link economy. Working paper.
Dezso, Z., A-L. Barabasi. 2002. Halting viruses in scale-free networks. *Phys*. *Rev*. *E* **65** 1–4.
Dhar, V., Geva, T., G. Oestreicher-Singer, A. Sundararajan. 2012. The gestalt in graphs: Prediction using economic networks. Working paper, NYU Stern School of Business.
Dhar, V., Hseih, J., and A. Sundararajan. Comments on "Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers". Available at *http://www.ftc.gov/os/comments/privacyreportframework/00458-57993.pdf*
Diesner, J., K. Carley. 2005. Revealing social structure from texts: Meta-matrix analysis as a novel method for network text analysis. V. K. Narayanan, D. J. Armstrong, eds. *Causal Mapping for Research in Information Technology*. IGI Publishing, Hershey, PA, 81–108.
Domingos, P., M. Richardson. 2001. Mining the network value of customers. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 57–66.
Doreian, P., F. N. Stokman. 1997. *Evolution of Social Networks*. Gordon and Breach, Amsterdam, The Netherlands.
Dorogovtsev, S., J. Mendes. 2003. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, New York, NY.
Duflo, E., R. Glennerster, M. Kremer. 2006. *Using randomization in development economics research: A toolkit*. Handbook of Development Economics **4** 3895–3962.
Dunbar, R.I.M. 1992. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* **22** 469-493.
Falk, A., J. J. Heckman. 2009. Lab experiments are a major source of knowledge in the social sciences. *Science* **326**(5952) 535–538.
Faraj, S., S. L. Johnson. 2010. Network exchange patterns in online communities. *Org. Sci.* (forthcoming).
Fawcett, T., F. Provost. 1997. Adaptive fraud detection. *Dat*. *Min*. *Knowl*. *Disc*. **3** 291–316.
Friedman, N. and Getoor, L. and Koller, D. and Pfeffer, A. 1999. Learning probabilistic relational models. In *Proceedings of the 1999 International Joint Conference on Artificial Intelligence*, 1300-1309.
Galeotti, A., S. Goyal. 2009. Influencing the influencers: a theory of strategic diffusion. *RAND J. Econ.* **40**(3) 509–532.
Galeotti, A., S. Goyal, M. O. Jackson, F. Vega-Redondo. 2010. Network games. *Rev. Econ. Stud.* **77**(1) 218–244.
Gallagher, B., H. Tong, T. Eliassi-Rad, C. Faloutsos. 2008. Using ghost edges for classification in sparsely labeled networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 256-264.
Getoor, L., N. Friedman, D. Koller, B. Taskar. 2003. Learning probabilistic models of link structure. *J*. *Mach*. *Learn*. *Res*. **3** 679–707.
Ghose, A., S-P. Han. 2011. An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet. *Management Science* (forthcoming).
Girvan, M., M. Newman. 2002. Community structure in social and biological networks. *Proc*. *Natl*. *Acad*. *Sci*. *U S A* **99**(12), 7821–7826.
Godes, D., D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Market*. *Sci*. **23**(4) 545–560.
Goldenberg, J., G. Oestreicher-Singer, S. Reichman. 2012. The quest for content: How User-Generated Links Can Facilitate Online Exploration. *J. Marketing Res*. **49(**4) 52-468.
Granovetter, M. 1973. The strength of weak ties. *Am*. *J*. *Sociol*. **78** 1360–1380.
Grewal, R., G. L. Lilien, G. Mallapragada. 2006. Location, location, location: How network embeddedness affects project success in open source systems. *Manage*. *Sci*. **52**(7) 1043-1056.

Hahn, J., J. Y. Moon, C. Zhang. 2008. Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Inform*. *Syst*. *Res*. **19**(3) 369–391.
Hansen, M. 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Admin*. *Sci*. *Quart*. **44**(1) 82–111.
Heckerman, D., D. M. Chickering, C. Meek, R. Rounthwaite, C. Kadie. 2001. Dependency networks for inference, collaborative filtering, and data visualization. *JMLR* **1**, 49-75.
Henderson, K., B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, C. Faloutsos. 2011. It's who you know: graph mining using recursive structural features. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery in data mining*, 663-671.
Hevner, A. R., S. T. March, J. Park, S. Ram. 2004. Design science in information systems research. *MIS Quarterly* **28**(1) 75–106.
Hill, S., D. K. Agarwal, R. Bell, C. Volinsky. 2006. Building an effective representation for dynamic networks. *J*. *Comput*. *Graph*. *Stat*. **15**(3) 584–608.
Hill, S., F. Provost, C. Volinsky. 2006. Network-based marketing: Identifying likely adopters via consumer networks. *Stat*. *Sci*. **21**(2) 256–276.
Hill, S., F. Provost, C. Volinsky 2007. Learning and inference in massive social networks. The 5th International Workshop on Mining and Learning with Graphs, Florence, Italy.
Hinz, O., M. Spann. 2008. The impact of information diffusion on bidding behavior in secret reserve price auctions. *Inform*. *Syst*. *Res*. **19**(3) 351–368.
Holland, P. W. and S. Leinhardt. 1977. A dynamic model for social networks. *Journal of Mathematical Sociology* **5**(1) 5-20.
Inkpen, A. C., E. W. K. Tsang. 2005. Social capital, networks, and knowledge transfer. *The Acad*. *Manage*. *Rev*. **30**(1) 146-165.
Iyengar, R., C. Van den Bulte, T. W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Market*. *Sci*. **30**(2) 195-212.
Jackson, M. O. 2008. *Social and Economic Networks*. Princeton University Press, Princeton, NJ.
Jackson, M. O., L. Yariv. 2005. Diffusion on social networks. *Économie Publique* **16** 3–16.
Jeppesen, L. B., K. Laursen. 2009. The role of lead users in knowledge sharing. *Res. Policy*. **38**(10) 1582-1589.
Kahanda, I., J. Neville. 2009. Using transactional information to predict link strength in online social networks. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence*, Menlo Park, CA.
Kane, G. C., M. Alavi. 2008. Casting the net: A multimodal network perspective on user-system interactions. *Inform*. *Syst*. *Res*. **19**(3) 253–272.
Karlan, D., M. Mobius, T. Rosenblat, A. Szeidl. 2009. Trust and social collateral. *Q*. *J*. *Econ*. **124**(3) 1307–1361.
Katona, Z., M. Sarvary. 2008. Network formation and the structure of the commercial World Wide Web. *Market*. *Sci*. **27**(5) 764–778.
Katona, Z., P. P. Zubcsek, M. Sarvary. 2009. Network effects and personal influences: The diffusion of an online social network. Working paper, University of California, Berkeley.
Katz, L. 1953. A New Status Index Derived from Sociometric Analysis. *Psychometrika*. **18** 39-43.
Kempe, D., J. Kleinberg, E. Tardos. 2003. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY.
Kleinberg, J. 1999. Authoritative sources in a hyperlinked environment. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms.* ACM, New York, 668–677.
Kossinets, G., D. Watts. 2006. Empirical analysis of an evolving social network. *Science* **311**(5757) 88–90.
Kuk, G. 2006. Strategic Interaction and knowledge sharing in the KDE developer mailing list. *Manage. Sci*. **52**(7) 1031-1042.

La Fond, T., J. Neville. 2010. Randomization tests for distinguishing social influence and homophily effects. *Proceedings of the 19th International Conference on World Wide Web*. Raleigh, NC, 601–610.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, A. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne. 2009. Computational social science. *Science* **323**(5915) 721–723.

Leenders, R. 1995. Models for network dynamics: a Markovian framework. *Journal of Mathematical Sociology* **20** 1-21.

Leenders, R., J. Engelen, J. Kratzer. 2003. Virtuality, communication, and new product team creativity: a social network perspective. *Journal of Engineering and Technology Management* **20** 69-92.

Leskovec, J., J. Kleinberg, C. Faloutsos. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. ACM, New York, NY.

Liben-Nowell, D., J. Kleinberg. 2007. The link-prediction problem for social networks. *J*. *Am*. *Soc*. *Inform*. *Sci*. **58**(7) 1019–1031.

Lin, M., N. R. Prabhala, S. Viswanathan. 2009. Judging borrowers by the company they keep: Social networks and adverse selection in online peer-to-peer lending. Working paper, University of Maryland.

López-Pintado, D. 2008. Diffusion in complex social networks. *Games and Economic Behavior* **62**(2) 573-590.

Lu, Y., J. Kinshuk, P. V. Singh. 2010. The emergence of opinion leaders in online review communities. Working paper, Carnegie Mellon University.

Ma, M., R. Agarwal. 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Inform*. *Syst*. *Res*. **18**(1) 42–67.

Macskassy, S. 2009. Using graph-based metrics with empirical risk minimization to speed up active learning on networked data. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 597–606.

Macskassy, S. A., F. Provost. 2005. Suspicion scoring based on guilt-by-association, collective inference, and focused data access. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.4577&rep=rep1&type=pdf.

Macskassy, S., F. Provost. 2007. Classification in networked data: A toolkit and a univariate case study. *J*. *Mach*. *Learn*. *Res*. **8** 983.

Manchanda, P., Y. Xie, N. Youn. 2008. The role of targeted communication and contagion in product adoption. *Market*. *Sci*. **27**(6) 961–976.

Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *Rev*. *Econ*. *Stud*. **60**(3) 531–542.

Martens, D., F. Provost. 2011. Pseudo-social network targeting from consumer transaction data. NYU Working Paper No. CEDER-11-05, 2011. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1934670

Mayzlin, D. 2006. Promotional chat on the Internet. *Market*. *Sci*. **25**(2) 155–976.

Mayzlin, D., H. Yoganarasimhan. 2011. Link to success: How blogs build an audience by promoting rivals.*Market*. *Sci*. **30**(2) 224-229.

McCallum, A., X. Wang, A. Corrada-Emmanuel. 2007. Topic and role discovery in social networks with experiments on Enron and academic email. *Journal of Artificial Intelligence Research* **30**(1) 249-272.

McLure-Wasko, M., S. Faraj. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quart*. **29**(1) 35-37.

McPherson, M., L. Smith-Lovin, J. M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annu*. *Rev*. *Sociol*. **27** 415–444.

Melville, P., K. Subbian, C. Perlich, R. Lawrence, E. Meliksetian. 2010. A predictive perspective on measures of influence in networks. The 2010 Workshop on Information in Networks (WIN 2010).

Miller, A. R., C. Tucker. 2009. Privacy protection and technology diffusion: The case of electronic medical records. *Manage*. *Sci*. **55**(7) 1077–1093.
Monge, P. R., N. S. Contractor. 2003. *Theories of Communication Networks*. Oxford University Press.
Morris, M., M. Handcock, D. R. Hunter. 2008. Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software* **24**(4) 1548-7660.
Muller, E., R. Peres, V. Mahajan. 2009. Innovation diffusion and new product growth. *MSI Working Series*. Marketing Science Institute, Cambridge, MA.
Munshi, K. 2003. Networks in the modern economy: Mexican migrants in the US labor market. *Q. J. Econ*. May 549–599.
Nam, S., P. Manchanda, P. K. Chintagunta. 2010. The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Market*. *Sci.* **29**(4) 1–11.
Narayanan, A., V. Shmatikov. 2009. De-anonymizing social networks. *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*. IEEE Computer Society, Washington, DC, 173–187.
Neville, J., B. Gallagher, T. Eliassi-Rad, T. Wang. 2011. Correcting evaluation bias of relational classifiers with network cross validation. *Knowledge and Information Systems*.
Neville, J., D. Jensen. 2005. Leveraging relational autocorrelation with latent group models. *Fifth IEEE International Conference on Data Mining*. IEEE Computer Society, Washington, DC, 322–329.
Neville, J., D. Jensen. 2007. Relational dependency networks. *Journal of Machine Learning Research* **8** 653-692.
Neville, J., O. Şimşek, D. Jensen, J. Komoroske, K. Palmer, H. Goldberg. 2005. Using relational knowledge discovery to prevent securities fraud. *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. ACM, New York, NY, 449–458.
Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Rev*. **45** 167–256.
Newman, M., A. Barabasi, D. J. Watts. 2006. *The Structure and Dynamics of Networks (Princeton Studies in Complexity)*. Princeton University Press, Princeton, NJ.
Newman, M. E. J., S. Forrest, J. Balthrop. 2002. Email networks and the spread of computer viruses. *Phys*. *Rev*. *E*, **66** 1–4.
Ngwenyama, O. K., A. S. Lee. 1997. Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS* Quarterly **21**(2) 145-167.
Oestreicher-Singer, G., A. Sundararajan. 2012a. The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets. . *Manage*. *Sci*. (forthcoming).
Oestreicher-Singer, G., A. Sundararajan. 2012b. Recommendation networks and the long tail of electronic commerce. *MIS Quart*. **36**(1) 65-83.
Oestreicher-Singer, G., L. Zalmanson. 2012. Paying for content or paying for community? The effect of consumer involvement on willingness to pay on media web sites. *MIS Quart*. (forthcoming).
Oh, J., A. Susarla, Y. Tan. 2011. Social networks and the diffusion of user-generated content: Evidence from YouTube. *Inform*. *Syst*. *Res*. (forthcoming).
Oinas-Kukkonen, H., K. Lyytinen, Y. Yoo. 2010. Social networks and information systems: Ongoing and future research streams. *Journal of the Association for Information Systems* **11** 61-68.
Oktay, H., B. J. Taylor, D. Jensen. 2010. Causal discovery in social media using quasi-experimental designs. 1st Workshop on Social Media Analytics (SOMA ’10), Washington, DC.
Orlikowski, J. A., J. Yates. 1993. Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly* **39**(4) 541-574.
Page, L., S. Brin, R. Motwani, T. Winograd. 1998. The PageRank citation ranking: Bringing order to the web. Stanford Digital Libraries Working Paper.
Palla, G., A.-L. Barabási, T. Vicsek. 2007. Quantifying social group evolution. *Nature* **446** 664–667.
Pastor-Satorras, R., A. Vespignani. 2001. Epidemic spreading in scale-free networks. *Phys*. *Rev*. *Lett*. **86**(14) 3200–3203.
Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Francisco, CA.

Perry, P. O., P. J. Wolfe. 2011. Point process modeling for directed interaction networks. Working Paper, NYU Stern School of Business.

Podolny, J. 2001. Networks as the pipes and prisms of the market. *Am*. *J*. *Sociol*. **107**(1) 33–60.

Price, D. S. 1965. Networks of scientific papers. *Science* **149**(3683) 510–515.

Provost, F. 2011. Geo-social targeting for privacy-friendly mobile advertising: Position paper. NYU Working Paper No. CEDER-11-06a, 2011 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1946695.

Provost, F., B. Dalessandro, R. Hook, X. Zhang, A. Murray. 2009. Audience selection for on-line brand advertising: privacy-friendly social network targeting. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 707–716.

Rattigan, M., M. Maier, D. Jensen. 2007. Exploiting network structure for active inference in collective classification. *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*. IEEE Computer Society, Washington, DC, 429–434.

Ravindran, K., A. Susarla, V. Gurbaxani. 2009. Social networks and contract enforcement in IT outsourcing. *Proceedings of the 30th Annual International Conference on Information Systems*. Phoenix, AZ.

Reagans, R., B. McEvily. 2003. Network structure and knowledge transfer: The effects of cohesion and range. *Admin*. *Sci*. *Quart*. **48**(2) 240–267.

Ren, Y., R. E. Kraut, S. Kiesler. 2007. Applying common identity and bond theory to design of online communities. *Org. Studies*. **28**(3) 377-408.

Rhue, L. and A. Sundararajan, 2012a. An empirical analysis of digital visibility. *Eighth Statistical Challenges on Electronic Commerce Symposium,* Montreal

Rhue, L. and A. Sundararajan, 2012b. Digital visibility and anticipatory social influence. Mimeo, New York University

Richter, Y., E. Yom-Tov, N. Slonim. 2010. Predicting customer churn in mobile networks through analysis of social groups. In *Proceedings of the 2010 SIAM International Conference on Data Mining* (SDM 2010).

Roberts, J. 2000. From know-how to show-how? Questioning the role of information and communication technologies in knowledge transfer. *Technology Analysis and Strategic Management* **12**(4) 429-443.

Robins, G., P. Pattison, J. Woolcock. 2005. Small and other worlds: Global network structures from local processes. *American Journal of Sociology* **110** 894-93.

Robins, G., T. Snijders, P. Wang, M. Handcock, P. Pattison. 2007. Recent developments in exponential random graph (p*) models for social networks. *Social Networks* **29**(2) 192-215.

Rodan, S., D. Galunic. 2004. More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Manage*. *J*. **25** 541–562.

Rogers, E. M. 2003. *Diffusion of Innovations*. Free Press, New York, NY.

Saar-Tsechansky, M. and F. Provost. 2007. Decision-centric active learning of binary-outcome models. *Information Systems Research* **18**(1) 4-22.

Sacerdote, B. 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Q*. *J*. *Econ*. **116**(2) 681–704.

Scellato, S., M. Musolesi, C. Mascolo, V. Latora. 2010. On nonstationarity of human contact networks. *Proceedings of 2nd Annual Workshop on Simplifying Complex Networks for Practitioners (SIMPLEX 2010).* Genoa, Italy.

Sen, P., G. Namata, M. Bilgic, L. Getoor, B. Gallagher, T. Eliassi-Rad. 2008. Collective classification in network data. *AI Magazine* **29**(3) 93-106.

Shi, Z., H. Rui, A. Whinston. 2011. Information sharing in social broadcast: Evidences from Twitter. Workshop on Information Systems and Economics.

Singh, P. V. 2011. The small world effect: The influence of macro level properties of developer collaboration networks on open source project success. *ACM T*. *Soft*. *Eng*. *Meth*. (forthcoming)

Singh, P. V., Y. Tan, V. Mookerjee. 2008. Network effects: The influence of structural social capital on open source success. Working paper, Carnegie Mellon University.

Snijders T., C. Steglich, M. Schweinberger. 2006. Modeling the co-evolution of networks and behavior. K. V. Montfort, J. Oud, A. Satorra, eds. *Longitudinal Models in the Behavioral and Related Sciences*. Lawrence Erlbaum, Mahwah, NJ.

Stephen, A. T., O. Toubia. 2010. Deriving value from social commerce networks. *J. Marketing Res*. **47**(2) 215–228.

Sundararajan, A. 2006. Network seeding. 18th Workshop on Information Systems and Economics. Available at http://digital.mit.edu/wise2006/papers/3A-2_Network_Seeding-Arun_S.pdf.

Sundararajan, A. 2007. Local network effects and complex network structure. *The B.E. Journal of Theoretical Economics* **7**(1).

Sundararajan, A., 2012. Deconstructing social influence in networks: on the need for models over measurement. Mimeo, New York University.

Sykes, T. A., V. Viswanath, G. Sanjay. 2009. Model of acceptance with peer support: A social network perspective to understand employees' system use. *MIS Quart*. **33**(2) 371–393.

Tambe, P., L. Hitt. 2010. Job hopping, information technology spillovers, and productivity growth. *http://ssrn.com/abstract=1302637*

Trier, M. 2008. Towards dynamic visualization for understanding evolution of digital communication networks. *Information Systems Research* **19**(3) 335-350**.**

Tsai, W. 2001. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Acad. Manage. J.* **44**(5) 996-1004.

Tucker, C. 2008. Identifying formal and informal influence in technology adoption with network externalities. *Manage*. *Sci*. **54**(12) 2024–2038.

Van den Bulte, C., G. L. Lilien. 2001. Medical innovation revisited: Social contagion versus marketing effort. *Am*. *J*. *Sociol*. **106**(5) 1409–1435.

Wang, P., M. Gonzales, C. Hidalgo, A-L. Barabasi. 2009. Understanding the spreading patterns of mobile phone viruses. *Science* **324**(5930) 1071–1076.

Wang, T., J. Neville, B. Gallagher, T. Eliassi-Rad. 2011. Correcting bias in statistical tests for network classifier evaluation. In *Proceedings of the 21st European Conference on Machine Learning*.

Wasserman, S. 1979. A stochastic model for directed graphs with transition rates determined by reciprocity. In *Sociological Methodology*, K. F. Schuessler (Ed.) Jossey-Bass, San Francisco, 392-412.

Wasserman, S., G. Robins. 2005. An introduction of random graphs, dependence graphs and p*. In *Models and Methods in Social Network Analysis*, P. J. Carrington, J. Scott, S. Wasserman (Eds.), Cambridge University Press, New York.

Wilhite, A. 2001. Bilateral trade and small world networks. *Comput*. *Econ*. **18** 49–64.

Wu, L., Waber, B., Aral, S., Brynjolfsson, E., & Pentland, S. 2008. "Mining Face to Face Interaction Networks Using Sociometric Badges: Evidence Predicting Productivity in IT Configuration" *Proceedings of the 29th Annual International Conference on Information Systems*, Paris, France.

Wu, F., B. Huberman, L. Adamic, J. Tyler. 2004. Information flow in social groups. *Physica A* **337** 327–335.

Xiang, R., J. Neville. 2008. Pseudolikelihood EM for within-network relational learning. *Proceedings of the 8th IEEE International Conference on Data Mining*.

Xiang, R., J. Neville, M. Rogati. 2010. Modeling relationship strength in online social networks. *Proceedings of the NIPS Workshop on Analyzing Networks and Learning with Graphs*. ACM, New York, NY, 981–990.

Yang, J., S. Counts. 2010. Predicting the speed, scale and range of information diffusion in Twitter. *Proceeding of the International AAAI Conference on Weblogs and Social Media*. Washington, DC.

Zhou, B., J. Pei, W.-S. Luk. 2008. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *ACM SIGKDD Explorations* **10**(2) 12–22.