Received: 3 June 2021 | Revised: 14 May 2023 | Accepted: 15 May 2023

DOI: 10.1002/smj.3527

RESEARCH ARTICLE

STRATEGIC MANAGEMENT JOURNAL WILEY

Giant cluster formation and integrating role of bridges in social diffusion

Sungyong Chang[1] | Jeho Lee[2] | Jaeyong Song[2]

[1]London Business School, London, UK

[2]Graduate School of Business, Seoul National University, Seoul, South Korea

Correspondence
Sungyong Chang, London Business School, London, UK.
Email: schang@london.edu

Jeho Lee, Graduate School of Business, Seoul National University, Seoul, South Korea.
Email: jeho0405@gmail.com

Funding information
Institute of Management Research, Seoul National University

Abstract

Research Summary: In social networks, isolated subgroups often aggregate into a massively connected subgroup, or a giant cluster, when bridges are built across subgroups. To understand the roles of bridges in integrating subgroups, we develop models focusing on the percentage of bridges among all ties. When it is below 1%, diffusion does not affect many individuals because the system is merely a collection of fragmented subgroups. Near 1%, however, we find that a slight increase in the percentage of bridges leads to sudden widespread diffusion across many subgroups. This dramatic change stems from a threshold-like structural characteristic of the network whereby previously fragmented subgroups come together abruptly. Our findings suggest that this integrating role of bridges is an important piece missing from the literature on small-world networks.

Managerial Summary: Our findings suggest that the formation of a giant cluster could be a structural precondition for large-scale diffusion. Detection of such clusters may allow prognostication of the possibility of large-scale diffusion. With the rise of social media and the availability of large amounts of social network data, the ability to detect giant clusters seems to be more attainable than in the past; such an ability would be a source of competitive advantage. We describe methods of detecting giant clusters and analyzing their structural properties using readily available social network data. With these methods, entrepreneurs and established firms can stimulate user adoption by targeting massive clusters of aggregated subgroups and spreading viral messages about their new products or services throughout the clusters.

KEYWORDS

bridge, diffusion, giant cluster, network

1 | INTRODUCTION

The knowledge of network structures can push the limit of understanding dynamics in social phenomena, such as diffusion of an innovative idea or learning across different groups of people. An illustration of structural changes in a real-world network gives us a glimpse of how subgroups may be knit together in a social system. Figure 1 illustrates changes in the inventor collaboration network at Samsung Electronics over a period of roughly 20 years. In the early years, only small isolated subgroups or individuals made up the network (Figure 1a). As inventors began to collaborate with others and to build inter-subgroup ties, which, in this study, are called "bridges," the formerly isolated subgroups began to aggregate into a massively connected subgroup, as shown in Figure 1b,c. The emergence of such a massive subgroup in social networks often involves building of bridges between isolated subgroups.

Granovetter (1973, pp. 1370–1371) stressed the importance of bridges in understanding social dynamics as follows: "They are the channels through which ideas, influences, or information

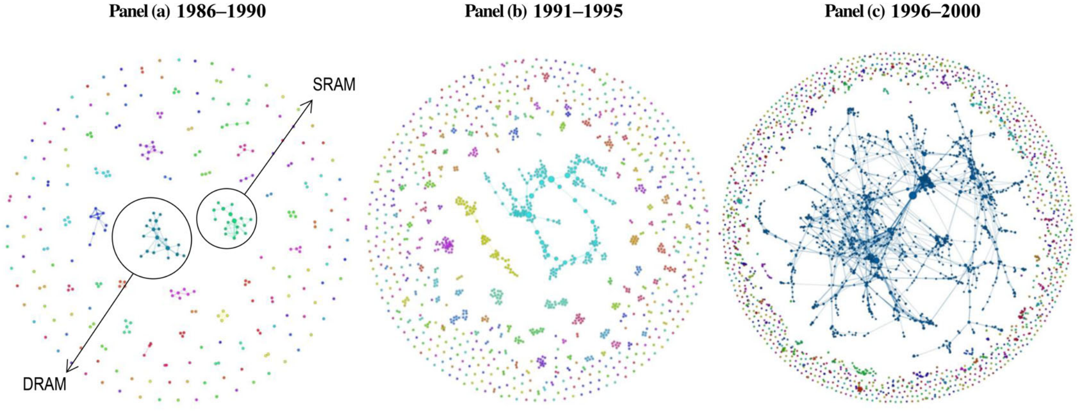


FIGURE 1 Evolution of an inventor collaboration network at Samsung. (a): 1986–1990. (b): 1991–1995. (c): 1996–2000. In the 1980s, few bridges existed between isolated subgroups in the collaboration networks at Samsung. During this period, each subgroup worked in isolation. For example, no interaction occurred between people in the DRAM and SRAM units. A giant cluster developed in the 1990s, growing larger over time. (Color online) Each node in this network represents an inventor. A link between two nodes indicates that the two inventors filed one or more patents together during the given period. Node sizes reflect betweenness centralities, which represent the level of bridging. Different node colors indicate different subgroups detected using a modularity optimization algorithm (Blondel et al., 2008). This figure was constructed with Gephi using the Fruchterman-Reingold and force-atlas algorithms.

socially distant from ego may reach him." Understanding various structural properties associated with bridges provide an idea of how a diffusion process unfolds in the social system. Building on Granovetter's structural view of social dynamics, prior work has advanced understanding in the contexts of learning (e.g., Cattani & Kim, 2021; Fang et al., 2010; Miller et al., 2006; Schilling & Fang, 2014), cultural polarization (Adams & Roscigno, 2005; Axelrod, 1997; Centola & Macy, 2007; Flache & Macy, 2011; Shibanai et al., 2001), imitation (Posen et al., 2013; Posen et al., 2020), winner-take-all competition (Lee et al., 2006; Lee et al., 2016), and diffusion of innovative ideas (Balachandran & Hernandez, 2018; Cattani & Ferriani, 2008).

On the theoretical front, research on small-world networks (Watts, 1999; Watts & Strogatz, 1998) has offered sophisticated tools for studying the dynamic role of bridges more systematically. One of the key findings is that an increase in the proportion of bridges to all ties in a system accelerates a diffusion process throughout the system. For example, by applying these tools, Fang et al. (2010) find that as the number of bridges in an organization increases, learning across subgroups accelerates. However, excessively fast learning with a large percentage of bridges is shown to have deleterious effects on cross-subgroup learning.

Although progress in this stream of research has been impressive, we argue that use of small-world network tools may result in invalid implications of social dynamics due to the simplifying assumption that all individuals in a given network are connected to one another with no disconnected parts. Over the last two decades, however, numerous empirical studies suggest that many large, social networks have only a tiny fraction of bridges (e.g., Cattani et al., 2008; Fleming et al., 2007; Gulati et al., 2012; Kogut et al., 2007; Onnela et al., 2007; Phillips, 2011; Uzzi & Spiro, 2005). As a consequence, the presence of isolated individuals or subgroups is not uncommon. For example, although the largest subgroup at Samsung in 2006 included over 75% of all inventors working at the firm, bridges in the network that year made up only 1.7% of all ties, and 25% of inventors was unconnected to the largest subgroup. Despite this reality, it has been challenging to relax the connected network assumption above due to the difficulties of formalizing isolated parts in a network. In particular, Watts (1999: p. 506) articulates those challenges as follows: "Disconnected graphs pose a problem because they necessarily have $L = \infty$ (i.e., average social distance becomes infinity), and this makes them hard to compare with connected graphs or even each other."

To improve our understanding of the dynamic implications of social networks that include fragmented subgroups, we develop computational models by employing tools from percolation theory (Christensen & Moloney, 2005; Stauffer & Aharony, 2018), which have been developed at the intersections of mathematics, polymer science, and statistical physics. With these tools, we elucidate a less-well-understood, dynamic role of bridges, or what we call "integrating role," by relaxing the connected network assumption above. The Samsung case serves as a guide to our modeling of social networks (e.g., parameter calibration) and the interpretation of our results.

Unlike the accelerating role in research on small-world networks, the integrating role is less related to the speed of an idea's diffusion and more related to the breadth of diffusion. In the context of cross-subgroup learning, we find that when the percentage of bridges among all ties is in the vicinity of 1%, a tiny percentage increase in bridges leads to a quantum jump in the efficacy of learning.

How can such a tiny change bring about the dramatic improvement in learning? We find that this dramatic effect stems from the threshold-like structural characteristic associated with the number of bridges in the system. As shown in Figure 2, overall connectivity across subgroups is a function of the percentage of bridges. When the percentage of bridges is below a certain threshold, the whole system is merely a collection of fragmented parts. Our numerical

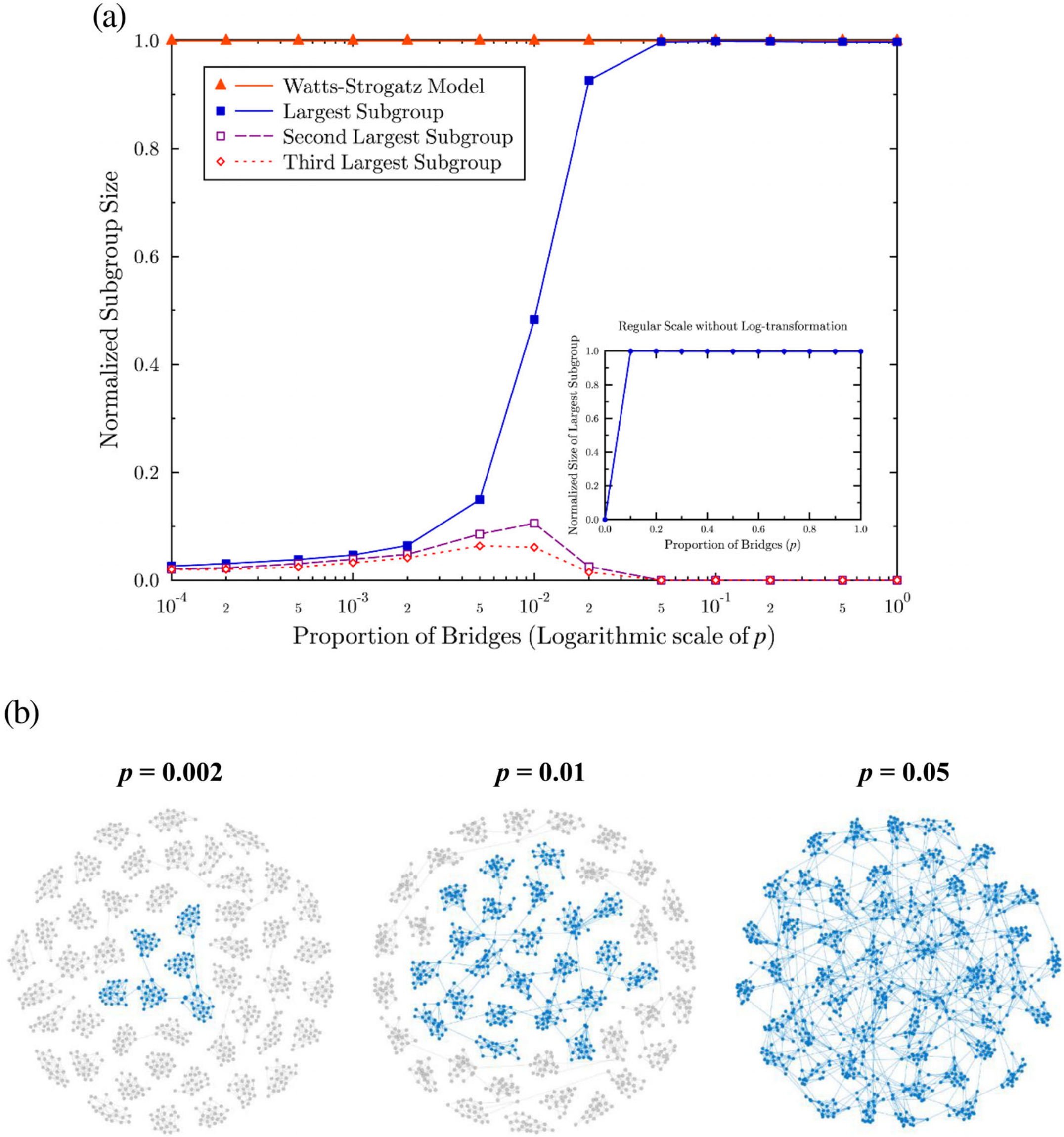


FIGURE 2 Effect of bridges on growth patterns of different subgroups. (a) Subgroup size by increasing bridges. (b) Illustration of largest subgroups with selected values of p. When the proportion p of bridges is below threshold p_c (i.e., $p < p_c$), the system is merely a collection of disconnected parts, and the largest subgroup will be small like other subgroups. When p is near the threshold, however, a slight increase in the proportion of bridges brings members of otherwise unconnected subgroups come together suddenly, leading to a quantum jump in the size of the largest subgroup. This abrupt change occurs even when bridges account for only about 1% of all ties in the system. For $p > p_c$, we define this largest connected group as a "giant cluster," which exhibits a marked deviation in size from the rest of the population. (Color online) For all networks, the number of nodes is 1000, and the number of links is 6000. The ratio of links to nodes is 6, which is observed in Samsung's network in 2006. Each data point here is averaged over 200 simulations.

result indicates that this threshold is approximately 1%. In the vicinity of this threshold, adding a tiny fraction of bridges causes a sizeable percentage of subgroups to be connected together all of a sudden. This sudden structural change, in turn, prompts the diffusion process to affect a considerably larger fraction of the population by connecting previously isolated subgroups, thereby boosting exchanges of diverse ideas across subgroups.

When the percentage of bridges is above 5%, all individuals in our models are connected to one another with no disconnected parts. Beyond this point, additional bridges tend to reduce learning performance because faster diffusion of some innovative ideas quickly drives out other valuable ideas that are scattered across subgroups. This finding is similar to that of Fang et al. (2010). However, the numerous empirical studies cited above suggest that such high percentages of bridges with no disconnected parts would be unrealistic. Furthermore, adding more bridges beyond the 5% level not only leads to an excess of bridges but also causes deformation of within-subgroup ties in our model. This deformation of subgroup structures also deviates from the reality of social networks. Therefore, the deleterious effects of additional bridges on learning seem to be a theoretical artifact in the parameter range that diverges from social reality.

Our key findings speak to the literature on the role of bridges in diffusion. In his seminal work, Granovetter (1973, p. 1360) identified a fundamental issue: "[H]ow interaction in small groups aggregates to form large-scale patterns (e.g., widespread diffusion of innovative ideas) eludes us in most cases." The substantial research on small-world networks has addressed this issue by paying attention to the role of bridges in accelerating diffusion. However, Granovetter's fundamental issue, the essence of which lies in the integration of small isolated subgroups, has not been fully addressed, as this research stream has assumed away isolated parts in social networks. Our work offers a more nuanced view by casting theoretical light not only on the accelerating role of bridges, but also on their integrating role. When building bridges across subgroups is difficult (e.g., inventor collaboration networks), our findings suggest that the integrating role of bridges will be far more pronounced than the accelerating role.

This article is organized as follows. First, we review the literature and outline the theoretical underpinnings of the two building blocks of social networks: (1) subgroups and (2) bridges. Second, we model aggregation of isolated subgroups, including these two building blocks. Next, we show under what conditions increasing the number of bridges will be conducive to the diffusion of innovative ideas and the promotion of learning across subgroups. Finally, we conclude by highlighting key theoretical insights and our contributions.

2 | LITERATURE REVIEW

Diffusion involves the spread of innovative ideas or influences from one individual to another until many are contacted. In the case of diffusion of an innovation, for example, Abrahamson and Rosenkopf (1997) noted that information about the innovation spreads from one potential adopter to another. A case in point is Hotmail, which was the fastest-growing user-based media company in the late 1990s (Subramani & Rajagopalan, 2003). Hotmail stimulated user adoption by promoting their slogan, "Get your free e-mail at Hotmail," which was inserted in the form of a tagline at the bottom of every e-mail sent out by Hotmail users. The company launched its service in 1996, and about 1 year later, 10 million subscribers had been attracted, ushering in the age of "viral marketing."

In a broad sense, a diffusion process includes not only simple diffusion of information, but also more complex forms of diffusion, such as learning or imitation, in which multiple ideas compete for an adopter's attention. Adoption of each idea depends on the adopter's payoff of that idea under uncertainty and bounded rationality (e.g., Cattani & Kim, 2021; Fang et al., 2010; March, 1991; Posen et al., 2013). However, all diffusion processes have one essential characteristic in common—social actors are susceptible to other actors' influences.

2.1 | Two essential properties of social networks: Subgroups and bridges

Social networks play an important role in how information spreads from one actor to another (Abrahamson & Rosenkopf, 1997). Research on social networks identified two essential building blocks for the architecture of social networks: (1) subgroups and (2) bridges. First, let us consider subgroups. Each subgroup in a social network is more or less insular from others, and certain barriers constrain membership overlap across subgroups (Milgram, 1967; Newman et al., 2002). Centola (2015: p. 1301) described such barriers as follows: "Our embeddedness in social contexts reflects our interests, appetites, and ambitions, which constrain the social contacts we have…" For example, Girvan and Newman (2002) showed that the coauthor network among scientists at the Santa Fe Institute came together through similarities of either research topic or methodology, and that collaboration across subgroups was limited (e.g., scientists working on the structure of RNA did not collaborate with mathematical ecologists).

Subgroup structures in social networks often reflect redundant ties or overlapping acquaintances. Overlap is likely to occur as people become acquainted by meeting friends of their friends (Granovetter, 1973). In particular, Gans (1982) observed how such overlap occurred when he was a participant-observer in the Italian community of Boston's West End. Gans (1982, pp. 340–341) noted:

> [He and his wife] were welcomed by one of our neighbors and became friends with them. As a result, they invited us to many of their evening gatherings and introduced us to other neighbors, relatives and friends… As time went on … other West Enders … introduced me to relatives and friends, although *most* of the social gatherings at which I participated were those of our *first* contact and their circle.

Cliquishness among friends and overlap are measured by clustering, which is defined as the average probability that two friends of a focal individual are also friends of each other (Newman, 2010; Watts, 1999). Research showed that social networks are characterized by far higher levels of clustering than nonsocial networks (Newman & Park, 2003).

The second essential building block of social networks is the bridge, which represents between-subgroup ties. The theoretical foundation of this concept can be traced back to the 1970s, when research on social networks focused mostly on small subgroups at the micro level, while ignoring how they are tied together at the macro level. Granovetter (1973) saw a paucity of micro–macro linkages as a fundamental weakness in the early literature. To fill this gap, he introduced the concept of bridge, which is conceived as a cross-level linchpin connecting different subgroups. This consideration of bridge as a cross-level linchpin was an important first step for building theories of when and how diffusion dynamics within a subgroup may travel to other subgroups. In particular, Granovetter (1973, p. 1360) envisioned the importance of bridges as follows:

> But how interaction in small groups aggregates to form large-scale patterns eludes us in most cases… the analysis of processes in interpersonal networks provides the most fruitful micro-macro bridge. In one way or another, it is through these networks that small-scale interaction becomes translated into large-scale patterns, and that these, in turn, feed back into small groups.

If a subgroup is isolated such that few bridges connect it to outside communities, information originating from that subgroup will not go far beyond itself. The lack of bridges could also be detrimental to learning or innovation. As individuals learn from one another in a small isolated subgroup, everyone quickly begins looking like everyone else, and knowledge diversity suffers. In such circumstances, it is impossible for any one individual to learn from others. Burt (2004, p. 349) described the flip side of this phenomenon as follows: "… people connected across groups are more familiar with alternative ways of thinking and behaving, which gives them more options to select from and synthesize." Balachandran and Hernandez (2018) empirically showed that the level of bridges between firms with different institutional boundaries is strongly associated with innovation. Similarly, Cattani and Ferriani (2008, p. 825) found that creative performance in the Hollywood film industry depends on the presence of bridging ties between two different parts of the world, where one part maintains "exposure to alternative sources of inspirations and novel ideas," whereas the other part provides "the base of legitimacy and support." Over the last three decades, there has been substantial progress in understanding the micro–macro linkages due to both methodological and theoretical advances (see, e.g., Adams & Roscigno, 2005; Axelrod, 1997; Baum et al., 2003; Centola & Macy, 2007; Doreian & Stokman, 1997; Flache & Macy, 2011; Shibanai et al., 2001; Snijders, 2001; Stokman & Doreian, 2001).[1]

2.2 | Limitations of applying the small-world network tools for understanding social dynamics

Major theoretical progress on the role of bridges in diffusion dynamics was made when Watts and Strogatz (1998) developed a formal model of "small-world networks," in which an increase in the percentage of bridges in the system reduces its average path length or the number of degrees of separation. This reduction in average path length was shown to accelerate the diffusion of ideas or influences. In the management area, the small-world network model has garnered considerable attention because this reduction substantially accelerates learning (e.g., Fang et al., 2010), imitation (Posen et al., 2020), winner-take-all competition (Lee et al., 2006; Lee et al., 2016), and information diffusion (Balachandran & Hernandez, 2018; Cattani & Ferriani, 2008; Watts, 1999; Watts & Strogatz, 1998).

[1]Some studies cited above suggested that even with a sufficient level of bridges, increased diversity rather than assimilation among individuals could be possible. Beside structural factor, other factors could affect diffusion dynamics. For example, individual factors like hatred or religious belief can affect cultural polarization. Complex factors like symmetry breaking and pattern formation in physics can reinforce division among actors in unexpected ways (Sayama et al., 2000). In light of these complex factors, the spatial population structure is shown to trigger local ethnic violence (Lim et al., 2007). Since our focus is primarily on structural factors, consideration of such additional factors is beyond the scope of the present work. Our position here is simply that the number of bridges can be an important precondition for large-scale diffusion.

In this literature, the assumption of full connectedness with no disconnected parts in the system has been taken for granted primarily because researchers believed that they had no choice but to ignore disconnected parts; otherwise, researchers would face a daunting challenge of formalizing isolated parts in a network (Watts, 1999). However, in reality, many large social networks have isolated parts (e.g., Cattani et al., 2008; Fleming et al., 2007; Gulati et al., 2012; Onnela et al., 2007; Phillips, 2011; Uzzi & Spiro, 2005). In the mobile call network of a European mobile operator, for example, 15.9% of subscribers were not connected to the largest subgroup (Onnela et al., 2007). In the inventor collaboration network at Samsung, about 25% of inventors were not connected to the largest subgroup.

The small-world network approach limits our understanding of "how interaction in small groups aggregates to form large-scale patterns" (Granovetter, 1973, p. 1360). In addressing this issue, researchers using the small-world network tools face two problems. First, these tools do not allow researchers to model isolated parts. Second, these modeling tools embrace a theoretical ideal, where the average path length is the shortest, thereby creating too many bridges and unrealistically low numbers of within-subgroup ties. This ideal is, again, inconsistent with the empirical findings of actual social networks, in which the number of within-subgroup ties far exceeds that of bridges, or between-subgroup ties (Girvan & Newman, 2002; Newman & Park, 2003; Onnela et al., 2007). For example, 98.3% of all ties at Samsung's inventor collaboration network are within-subgroup ties. Furthermore, the preponderance of bridges implies that it is rather easy and costless to build a bridge between two individuals who live in different worlds. This assumption violates the realities of embeddedness of social actors in subgroups, which constrain social contacts across those subgroups (Centola, 2015; Girvan & Newman, 2002; Newman et al., 2002). Thus, we argue that blind use of small-world network tools may result in invalid implications of social dynamics. In particular, excessively fast and widespread diffusion of ideas or influences is unlikely if the percentage of bridges in the system is not as large as assumed.

2.3 | Large-scale diffusion can happen even when an increase in the percentage of bridges is small

We relax the aforementioned connected network assumption to highlight a less-well understood role of bridges in social networks for large-scale diffusion, that is, the integrating role of bridges in tying members of subgroups together into a large conglomerate, or a giant cluster. Peitgen et al. (1992, p. 16) defined percolation or giant cluster formation as follows: "When a structure changes from a collection of many disconnected parts into basically one big conglomerate, we say that percolation occurs." The integrating role of bridges in networks is associated with how widely ideas and influences will diffuse throughout the system. In this study, we focus more on the breadth of diffusion than the speed of diffusion.

Empirical research suggests that giant clusters in social networks are extreme outliers deviating markedly in size from the rest of the subgroup population. For example, Onnela et al. (2007) reported that in the mobile call network of a European mobile operator, the giant cluster consists of 84.1% of its subscribers. In Samsung's inventor collaboration network in 2006, the giant cluster included over 75% of all inventors. One may conjecture that a large percentage of bridges must be necessary for so many disconnected parts to aggregate into a giant cluster. Percolation theory (e.g., Christensen & Moloney, 2005; Stauffer & Aharony, 2018), which was developed to study giant cluster formation in nonsocial networks in the early years, suggests that this conjecture may not be invalid in many nonsocial networks.[2] However, the percentage of bridges in either of the social networks above is rather small—far smaller than that of within-subgroup ties. For example, bridges made up only 1.7% of all ties in the Samsung case.

A small percentage of bridges for interweaving many subgroups into a giant cluster seems to have important implications for social dynamics. Kirkpatrick (2011, p. 7) illustrated how a large-scale diffusion (e.g., an upsurge of nation-wide protests) can happen via social media like Facebook: "Ideas on Facebook have the ability to rush through groups and make many people aware of something almost simultaneously, spreading from one person to another and on to many with unique ease—like a virus, or meme." However, such virus-like diffusion may not be always possible. Katz (1961) noted: "It is as unthinkable to study diffusion without some knowledge of the social structures in which potential adopters are located as it is to study blood circulation without adequate knowledge of the veins and arteries." Percolation theory suggests that large-scale diffusion in a nonsocial system (e.g., a massive forest fire) requires a structural precursor—that is, the formation of a giant cluster. In a social network, we argue that even without a large increase in the percentage of bridges, a giant cluster could be formed. This formation may be a precondition for large-scale social change.

In the next section, we will validate the two claims: (1) large-scale diffusion can happen in large social networks if giant clusters emerge and previously isolated subgroups become integrated; this scenario is likely even when an increase in the percentage of bridges is small and (2) the excessively fast and widespread diffusion that is possible in the presence of abundant bridges cannot occur in large social networks in reality. We confirm these claims by relaxing the aforementioned idealized assumptions in small-world network research, while embracing the essential structural properties of social networks.

3 | MODEL OF BRIDGES AS INTEGRATORS

In this section, we examine the role of bridges as integrators tying together fragmented subgroups. We do this by gradually increasing the percentage of bridges, which is the key control parameter. The theoretical issues surrounding this parameter are addressed through formal modeling.

3.1 | Emergence of a giant cluster in Samsung's inventor collaboration network

Before introducing our model, we describe how increases in the number of bridges over time tied isolated individuals or small groups together into a giant cluster in the inventor collaboration networks of the semiconductor division at Samsung Electronics. This case serves as a guide to our modeling of social networks (e.g., parameter calibration) and interpretation of our

[2]A forest fire cannot spread across gaps in the trees. A necessary condition for the outbreak of a massive forest fire is that a sufficiently large percentage of trees should act as bridges in the gaps across isolated parts of the forest. For example, 800,000 acres in Yellowstone National Park burned in 1988. Malamud et al. (1998, p. 1841) noted: "Until 1972, Yellowstone National Park had a policy of suppressing many of its fires, resulting in a large accumulation of dead trees, undergrowth, and very old trees." This left the forest susceptible to a massive fire. Similarly, large-scale electrical conductivity requires a large percentage of conductive materials to increase their density per unit volume, thereby bridging gaps across isolated parts (Last & Thouless, 1971).

results. We construct the network using patent data from the United States Patent and Trademark Office database. Here, we assume that if two inventors patented together, a direct tie existed between them. We analyze a total of 15,777 US patents (semiconductor-related) filed by Samsung Electronics from 1982 to 2006.[3] In the early years, only small isolated subgroups or individual isolates existed (Figure 1a). Over time, however, a large subgroup emerged, as shown in Figure 1b,c. In 2006, the largest subgroup included over 75% of all inventors in Samsung's semiconductor division with a clustering coefficient of 0.721. We call this subgroup a giant cluster, as it exhibits a marked deviation in size from the rest of the subgroup population.

To understand how the giant cluster formed, we interviewed key informants.[4] We learned that the initial isolation of inventors at Samsung was largely caused by the natural boundaries that form between subunits made up of inventors with distinct expertise. People from different subunits tended to work in information silos. For example, the two isolated clusters circled in Figure 1a represent collaborators in the DRAM and SRAM subunits. One of the key informants noted that inventors in these subunits did not talk to each other in the 1980s. Top managers at Samsung believed that these information silos were not conducive to information diffusion and learning across different subunits. They introduced policies and practices to build bridges across different subunits, primarily through job rotation and cross-functional meetings. In addition, Samsung's patent office regularly examined all filed patents, identified researchers with similar interests, and encouraged collaboration between them. As a consequence, bridges between subunits at Samsung Electronics increased over time, but the proportion of bridges to all ties in the network remained tiny (see Appendix A). For example, the proportion of bridges in 2006 was only 0.017,[5] which seems to reflect the substantial costs of crossing the boundaries of subunits with distinct specialties. Indeed, job rotation was enacted for only a limited number of experienced people. However, a giant cluster emerged over time. How was this possible? In the next section, we address this question by developing a model of social networks.

3.2 | Comparison of our network model with related network models

We develop our network model by building on and extending prior work, which includes the ER model (Erdős & Rényi, 1960), the WS model (Watts & Strogatz, 1998), and the FLS model (Fang et al., 2010). Appendix B summarizes similarities and differences between our model and these models.

First, our model departs from the WS and FLS models in that we relax the connected network assumption, which is that all individual agents in a system are connected to one another as parts of a single large network. Given the empirical evidence that many real-world social networks consist of isolated subgroups of different sizes (e.g., Cattani et al., 2008; Fleming et al., 2007; Gulati et al., 2012; Onnela et al., 2007; Phillips, 2011; Uzzi & Spiro, 2005), we consider the relaxation of this assumption as a necessary step to go deeper into the dynamic role of bridges in integrating fragmented subgroups. In this regard, our model is more closely akin to

[3]Yayavaram and Ahuja (2008) identified 30 semiconductor-related patent classes. Following their selection criteria, we collect a total of 15,777 US patents filed by Samsung Electronics from the USPTO database. Then, following Fleming et al.'s (2007) methodology, we construct inventor collaboration networks with moving 5-year windows. For presentation bravity, we spell out only the last year for each of 5-year periods.

[4]The key informants included the former president of the semiconductor division, chief technology officer, chief intellectual property officer, HR manager, and inventors.

[5]We identified bridging ties among all ties based on the link classification method proposed by Lee et al. (2010).

the ER model. Like the ER model, ours start with individual isolates or isolated subgroups in a system. With incremental percentage increases in the number of bridges, both models show a percolation transition from a collection of fragmented parts to the formation of a giant cluster. In contrast, such a transition is not of theoretical interest to the WS and the FLS models, given their connected network assumption.

Second, as discussed previously, the existence of distinct subgroups is central to research on social networks (e.g., Centola, 2015; Newman et al., 2002; Onnela et al., 2007). The ER model ignores this social reality completely. As such, random connectivity in a large sparse network tends to result in poor clustering and a preponderance of bridges in a system; these conditions are inconsistent with the empirical regularity of social networks—that is, within-subgroup connectivity is far larger than between-subgroup connectivity in social networks (e.g., Girvan & Newman, 2002; Newman & Park, 2003). Unlike the ER model, our model allows for some parameter range within which this social reality is embraced.

3.3 | Modeling of giant cluster formation with increasing bridges

The essence of our model of giant cluster formation lies in an increasing proportion p of bridges to all ties in the network, which contributes to the growth of the largest subgroup up to a theoretical extreme, where all individuals in the system are interconnected. We start with the following basic conditions: n individuals exist within a system, which consists of subgroups of an equal size g at $p = 0$. On average, each individual has six ties to other individuals within the same subgroup. This was the situation at Samsung Electronics in 2006, in which the condition for large, sparse networks was satisfied (Watts & Strogatz, 1998). A high clustering coefficient of 0.721 precludes consideration of the ER model (Erdős & Rényi, 1960), which is characterized by poor clustering for the given sparse network condition. High clustering in each subgroup represents a social circle characterized by a shared context (Blau & Schwartz, 1984; Centola, 2015). In the Samsung case, a subgroup represents a functional unit for carrying out a certain task (e.g., the DRAM subunit). All parameters used in this simulation are specified in Appendix C.

To understand the role of bridges in integrating fragmented subgroups, we specify the following procedure for all individuals: Each individual's tie to other individuals within his or her subgroup is removed with probability p, and a new tie for the focal individual is formed to a randomly selected individual outside the subgroup—thus, p represents an average proportion of between-subgroup ties, or bridges, to all ties in the network. When $p = 0$, the unconnected subgroup structure above is preserved. That is, individuals maintain ties with others only within their subgroup boundaries. Therefore, no bridges exist between subgroups, and the whole network is a collection of fragmented parts.

Figure 2a shows that the formation of a giant cluster is characterized by threshold-like behavior with increasing p. The results demonstrate the relationship between the proportion p of bridges to all ties in the network and the size of the largest subgroup S, which is measured as a fraction of the overall system size n. Let p_c denote the threshold. When $p < p_c$, even the largest subgroup will be small, and it is not much larger than other subgroups. For example, when $p = .002$, the largest subgroup in Figure 2b is small. Here, bridges, on average, account for 0.2% of all ties, and 99.8% are within-subgroup ties. When p is just above p_c, however, there is a quantum jump in S with respect to a slight increase in p. To enable the observation of the fine details of the transition near the threshold, we plot p (the horizontal axis) on a log scale of base 10, which understates the sharp changes at the transition (see the inset without the log).

According to our estimation, $p_c \cong .01$ for the chosen parameter conditions here. This means that bridges, on average, account for 1% of all ties, and 99% are within-subgroup ties. In the vicinity of $p = p_c$, the addition of a tiny percentage of bridges causes sudden connections among a sizeable percentage of subgroups and their members. For $p > p_c$, we define this largest subgroup as a "giant cluster," which is an extreme outlier deviating markedly in size from the rest of the subgroup population. The low value of the threshold indicates that a giant cluster can emerge even if bridges constitute only a tiny fraction of all ties in the system.

As shown in Figure 2b, at $p = .05$, the growth of S tapers off and all subgroups are connected to one another, and the whole system becomes a single, connected network. This is what we call the "saturation point." As mentioned earlier, empirical research on large, sparse social networks indicates that many real-world social networks include disconnected parts, suggesting that the costs of building bridges may be high, or some individuals may not want to build bridges with distant others. That is, the range of $.05 \leq p \leq 1$ may be a theoretical artifact for such social networks.

Figure 3 shows another noticeable divergence between the giant cluster and other subgroups in terms of their growth trajectories. Let us define the growth rate as the magnitude of change in subgroup size per incremental increase in p. In the vicinity of $p = p_c$, the growth rate for the largest subgroup peaks. Why does the growth rate peak near the threshold? As p increases, the subgroups grow at first by linking to other, isolated subgroups. In particular, when $p << p_c$, new bridges are more likely to connect two isolated subgroups. In this range, the merged subgroups look still small relative to the system size n. In the vicinity of $p = p_c$, however, new bridges are more likely to connect subgroups that have already gone through mergers several times—that is, merged subgroups tend to crosslink to form a giant cluster. This is an underlying reason for the peak growth in the largest subgroup. Far above the threshold (i.e., $p >> p_c$), however, its growth trajectory is positive, but at a steadily decreasing rate as the value approaches zero. As p approaches the saturation point ($p = .05$), where all nodes are interconnected either directly or indirectly, the growth rate tapers off to zero since there is no more room for further growth.

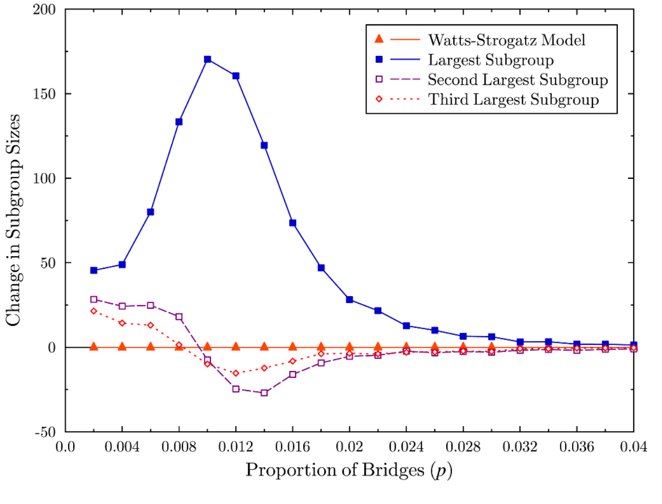


FIGURE 3 Changes in subgroup size with incremental increases in the proportion of bridges. For all networks, the number of nodes is 1000, and the number of links is 6000. The ratio of links to nodes is 6, which is observed in Samsung's collaboration network in 2006. Each data point here is averaged over 1000 simulations.

Unlike in the giant cluster, declining growth rates are observed in the second- and third-largest subgroups, values for which become negative before all subgroups become parts of a single connected network. Such growth trajectories below zero naturally arise because for $p > p_c$, the giant cluster gobbles up the other subgroups as more bridges are added to the system. To illustrate why this happens, suppose that the size of the second-largest subgroup is sufficiently larger than that of the third-largest subgroup. Then, the second-largest subgroup is more likely to be merged with the giant cluster in a few steps. This is because whenever a bridge is created with a random rewiring step, the probability that two given subgroups will be merged is proportional to the number of possible links between them which, in turn, is equal to the product of their respective sizes (numbers of nodes). If this possibility materializes, the second-largest subgroup will be a part of the giant cluster, and the third-largest subgroup will become the second-largest subgroup. As assumed earlier, the size of the third-largest subgroup is smaller than that of the second-largest subgroup. Therefore, the new, second-largest subgroup experiences a negative change in size. Again, unlike our model, the Watts–Strogatz model reveals no change in the size of the largest group since all individuals across subgroups stay connected to one another regardless of changes in rewiring parameters.

4 | SIMPLE DIFFUSION FROM AN INTEGRATOR PERSPECTIVE

To understand when and to what extent new information diffuses, we first examine a simple diffusion process, namely the diffusion of a single piece of information from the perspective of bridges as integrators. We show that the detection of a giant cluster (i.e., if $S > S_c$, or if $p > p_c$), which lies at the heart of determining whether a whole network is integrated or fragmented, allows one to prognosticate the possibility of large-scale diffusion. Suppose that there are n individuals in the system and that only one of them serendipitously comes across new information and diffuses it to her direct contacts with some positive transmission probability. Here, each individual will be in one of two states with regard to the new information: adoption or no adoption. For simplicity, the initial adopters of this information will also diffuse it to their direct contacts with the same transmission probability,[6] and this process repeats until no further direct contacts are left.

To what extent will the information diffuse throughout the system? The answer depends on whether the focal individual is connected to a small, isolated subgroup or a large subgroup. If one is interested in the average behavior of the system, the answer depends on the size of the largest subgroup, S. When $S < S_c$, or when $p < p_c$, the extent of information diffusion will be negligible because the size of the largest subgroup will be negligible relative to system size n. As shown previously, the entire network is simply a collection of many small, isolated parts, and there will be no giant cluster. On the other hand, when $S > S_c$, or when $p > p_c$, a giant cluster will emerge whose size will be proportional to n, whereas the sizes of the second-largest and other smaller subgroups will remain small, as shown earlier. Then, the extent of information diffusion will be disproportionately more influenced by the largest subgroup than by any other subgroup. The more detailed analysis in Figure 4a shows that the size of the largest subgroup S positively affects the extent of information diffusion. The (blue) closed circle indicates the

[6] Without loss of generality, we assume that the transmission probability is one. The long-term result will not be affected by changing it to smaller probability values, which only slow down the diffusion process.

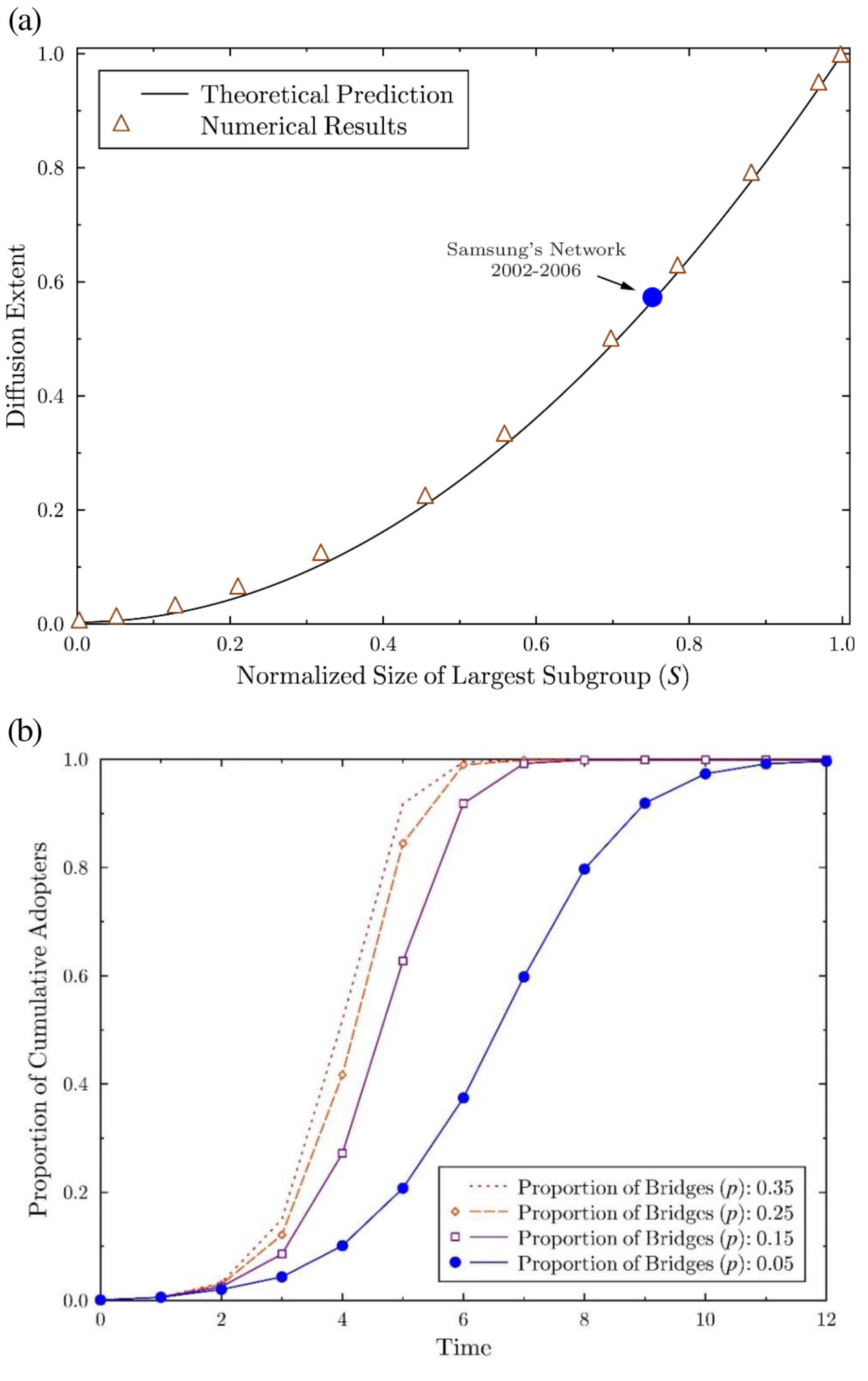


FIGURE 4 Results of information diffusion. (a) Effect of largest subgroup size on extent of information diffusion. (b) Effect of bridges on the speed of information diffusion above saturation point. Panel (a) shows that as the size of the largest subgroup increases, new information diffuses more widely. Panel (b) shows that as the number of bridges increases, the diffusion speed tends to accelerate when the system is above the saturation point. For all networks, the number of nodes is 1000, and the number of links is 6000. Each data point here is an average over 1000 simulations.

average result of 1000 simulation runs in the real-world network at Samsung for the period between 2002 and 2006. The solid line represents the analytical result based on our idealized network models, whereas the triangles represent numerical results, which are averaged over 1000 simulation runs in our idealized network models. The results show that the extent of information diffusion measured by the number of information adopters is positively associated with the size of the largest subgroup S—the larger the size of the largest subgroup, the higher the extent of information diffusion. More precisely, the extent of information diffusion is proportional to S^2 (see Appendix D for detailed analytical results).

We also examine the effects of additional bridges on diffusion beyond the saturation point, where additional bridges will have no effect on the size of the largest subgroup (recall that S reaches its theoretical limit at this point). As shown in Figure 4b, an increase in the proportion of bridges to all ties in the network beyond the saturation point tends to accelerate information diffusion up to a certain point because given that the giant cluster gobbled up all previously isolated subgroups, additional bridges will only reduce the average path length in the network. This result is equivalent to the acceleration effect in much of prior work (e.g., Fang et al., 2010; Lee et al., 2016; Watts & Strogatz, 1998).

In sum, our analysis of information diffusion suggests that the presence of a giant cluster can make the system more susceptible to large-scale diffusion. In other words, if we have some idea of threshold p_c, we can get a fairly good idea about the extent of information diffusion—information diffusion will be far more widespread if $p > p_c$. To validate this claim more systematically, we run 100 repeated simulations by using the estimate of the threshold ($p_c \cong .01$) for determining the presence of a giant cluster. For each simulation of information diffusion in a network with no giant cluster, a value of p is randomly drawn from a uniform distribution between 0 and 0.01. For information diffusion in a network with a giant cluster, we split our analysis into two parameter conditions: (1) the small percentage of bridges and (2) the large percentage of bridges. For the small percentage case, a value of p is randomly drawn from a uniform distribution between .01 and .05, whereas for the large percentage case, a value of p is randomly drawn from a uniform distribution between .05 and 1. The results in Table 1 show

TABLE 1 Effect of giant cluster formation on information diffusion and learning.

	Average penetration rate of information	**Average learning performance**
(1) Network without a giant cluster ($0 \le p < .01$)	0.054	0.154
(2) Network with a giant cluster if the percentage of bridges is small ($.01 \le p < .05$)	0.917	0.757
(3) Network with a giant cluster if the percentage of bridges is large ($.05 \le p \le 1$)	0.998	0.506
Difference between (2) and (1)	0.863 (39.72)	0.603 (36.79)
Difference between (2) and (3)	−0.081 (−4.17)	0.251 (13.34)

Note: T statistics in parentheses. All p-values for the differences are less than .0001. This table reports the average penetration rate of new information and average learning performance for three different network architectures: (1) those without a giant cluster and (2) those with a giant cluster (when the percentage of bridges is low; $.01 \le p \le .05$), and (3) those with a giant cluster (when the percentage of bridges is high; $.05 \le p \le 1$). For each simulation in a network with no giant cluster, a value of p is randomly drawn from a uniform distribution between 0 and .01. For each simulation in a network with a giant cluster for architecture (2), a value of p is randomly drawn from a uniform distribution between .01 and .05. For each simulation in a network with a giant cluster for architecture (3), a value of p is randomly drawn from a uniform distribution between .05 and 1. The results here are averaged over 100 repeated simulations.

that on average, 5.4% of individuals adopt the new information in the absence of a giant cluster, whereas 91.7% adopt the information when a giant cluster is present for $.01 \leq p < .05$. This difference is statistically significant. In addition, when the percentage of bridges is sufficiently large ($.05 \leq p \leq 1$), 99.8% of individuals adopt the new information. The results here support the claim that information diffusion will be far more widespread in the presence of a giant cluster than without it. In fact, even when the percentage of bridges is rather small (i.e., ranges from 1 to 5%), the claim is supported.

5 | COMPLEX DIFFUSION FROM AN INTEGRATOR PERSPECTIVE

In this section, we analyze the dynamics of a more complex diffusion, cross-subgroup learning, from the perspective of bridges as integrators. By focusing on the size of the largest subgroup S, we show that the presence of a giant cluster (i.e., if $S > S_c$, or if $p > p_c$) enhances cross-subgroup learning. As described previously, however, these processes are more complex in that more than one innovative idea is diffused via interpersonal learning. As such, adding more bridges to the system is not always conducive to learning.

Given this complexity, we confirm two key claims. First, we test the claim that large-scale diffusion can happen in a large social network if a giant cluster emerges and brings together previously isolated subgroups. In the learning context, this means that a quantum jump in the efficacy of cross-subgroup learning is likely, as an increase in bridges integrates previously isolated subgroups, thereby boosting exchanges of diverse ideas across subgroups. We expect that this integration effect is likely even with a small increase in the percentage of bridges. Second, we will show the acceleration effect, or the deleterious effect of excessively fast learning, when the percentage of bridges is large. In light of findings from numerous empirical studies, we then argue that this effect cannot occur in large, social networks in reality.

5.1 | Model of interpersonal learning

We first build a family of networks with varying p values as described in the network model section. For example, in Figure 2, we create 13 networks by selecting 13 different values for p (see the details in Appendix C). Then, we run learning simulations for each value of p separately. We do this to isolate learning dynamics from potentially confounding effects due to changes in p.

Our learning model consists of two other components: (1) reality and (2) n individuals with m ideas. Here, we primarily employ the modeling convention of organizational learning established by March (1991) and modified by Fang et al. (2010). Reality $\boldsymbol{R}$ reflects the environment, which determines what is good and what is bad. It is represented by m elements, each of which takes on a value of either 1 or 0. Activity vector $\boldsymbol{a_i}$ for individual i is represented by m elements, each of which takes on a value of either 1 or 0; these values represent alternative ideas. In sum, each individual tries to select an idea on each dimension of m to match its corresponding value in $\boldsymbol{R}$ under conditions of uncertainty and bounded rationality. The idea is considered "good" if it matches. Otherwise, the idea is considered "bad."

To evaluate the performance of each individual, we build on the payoff function $\Phi(x)$ proposed by Fang et al. (2010), where $\boldsymbol{a_i}$ represents an m-bit string with $\boldsymbol{a_i} = (x_1, x_2, ..., x_m)$. Rosenberg (1979, p. 30) noted: "Improvements in performance in one part are of limited significance without simultaneous improvements in other parts..." To capture this feature in our payoff function, an m-bit string is partitioned into equally-sized independent subsets, within which there are ψ bits whose performance is coupled as follows:

$$\Phi(x) = \frac{\Psi}{m}\left(\prod_{j=1}^{\Psi}\delta_j + \prod_{j=\Psi+1}^{2\Psi}\delta_j + \cdots + \prod_{j=m-\Psi+1}^{m}\delta_j\right)$$

where $\delta_j = 1$ if x_j matches the value of the jth element in $\boldsymbol{R}$, and $\delta_j = 0$ otherwise. Note that $1 \leq \psi \leq m$.

In this payoff function, ψ serves as a tunable parameter that controls the difficulty of searching for the right combination to match reality. When $\psi = 1$, a search is the easiest because improvements in performance in one part are independent of improvements in other parts. As the value of ψ increases, the search problem becomes more difficult as parts become more interdependent.

Now, let us consider interpersonal learning processes. Initially, each individual brings her own idea on each dimension of m, which is randomly drawn from two possible states: 1 or 0. The probability that each idea is good is 0.3—this probability in prior studies is set at 0.33 (e.g., Fang et al., 2010; March, 1991).[7] At time step t, each individual learns from other individuals with direct ties whose connection topology was described in our network model earlier in this study. In so doing, each individual is likely to assimilate a subset of ideas from "superior performers," those individuals whose performance is better than that of the focal individual at $t-1$. If there is only one superior performer, a focal individual adopts the superior performer's idea on the jth dimension with probability θ, which is the learning rate. If there are multiple superior performers, and if there is no disagreement among them, the focal individual adopts the agreed idea on the jth dimension with probability θ. On the other hand, if there is disagreement, the individual adopts the idea held by the majority of superior performers with probability θ. A system's performance is then measured as the average performance across all individuals in the system.

5.2 | Simulation results of basic model

Figure 5a shows the results of the cross-subgroup learning model with $n = 1000$ when learning dynamics reach a steady state, where no individuals change their ideas. The details of the other parameter values are specified in Appendix C. When the fraction of bridges is insufficient (i.e., $p < p_c$ and $p_c \cong .01$), the system is fragmented into many small, isolated subgroups, and learning performance is poor. As p increases, learning performance improves bit by bit at first. However, learning performance improves dramatically with a slight increase in the fraction of

[7]We examine sensitivity to the probability that each idea is good at an initial period. In the baseline model, we set the probability at 0.3. Here, we vary this probability from 0.1 to 0.5. Appendix G shows that as this probability increases, the performance curve shifts upward. Nonetheless, the shapes of the curves look similar unless the probability is too high (e.g., 0.5). This is not inconsistent with the literature on learning.

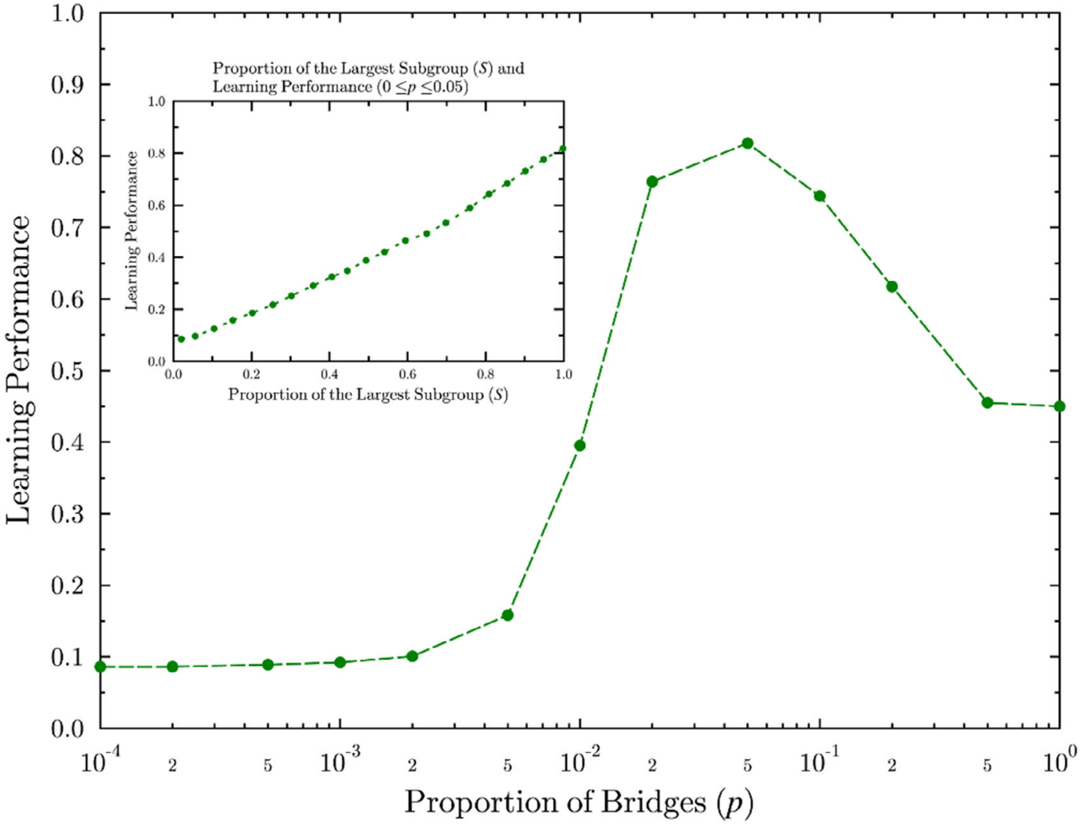


FIGURE 5 Bridges, largest subgroup size, and learning performance. The result shows that there is a dramatic jump in learning performance in the range of p between .01 and .05. The inset shows that learning performance monotonically increases with the size of the largest subgroup in that range. However, when p is above .05, additional bridges tend to reduce learning performance. The problem complexity is set at $\psi = 4$. Each member starts recombination with 30% correct bits at period 0. The learning rate is set at $\theta = 0.5$. Learning performance is normalized by dividing each outcome by the highest one. Each data point here is averaged over 200 simulations.

bridges near $p = p_c$. Especially in the narrow range for $.01 \leq p \leq .05$, learning performance continues to increase substantially, and the highest performance observed in this analysis is 0.82 at $p = .05$. Beyond this point, however, additional bridges tend to reduce the efficacy of learning.

5.3 | Integration effect: Efficacy of cross-subgroup learning due to giant cluster formation

In this section, we validate the claim that a giant cluster can emerge in a social system with a small percentage increase in bridges and that it can act as a knowledge integrator by bringing together diverse good ideas and knowledge from different subgroups. Our model setup allows us to test this claim by controlling p. As shown before, there are substantial increases in performance for $.01 \leq p < .05$. We know that in this narrow range, the largest subgroup S turns into the giant cluster by gobbling up other subgroups. Recall that S increases suddenly with a slight increase in p when p is just above p_c ($p_c \cong .01$) and that S has room for further growth by merging with other isolated subgroups when more bridges are added. Therefore, the jumps in learning efficacy may stem from the increases in S. This result seems to capture the integration effect, which facilitates the diffusion of good ideas between previously unconnected subgroups, promoting learning widely across subgroups.

To validate this integration effect more systematically, we run 100 repeated simulations by using the range for $.01 \leq p < .05$. For each simulation of cross-subgroup learning in a network with a giant cluster, a value of p is randomly drawn from a uniform distribution between .01 and .05. For each simulation of cross-subgroup learning in a network with no giant cluster, a value of p is randomly drawn from a uniform distribution between 0 and .01. As shown in Table 1, the repeated simulations show that average learning performance is significantly higher with a giant cluster than without it. Specific values for learning performance are 0.757 and 0.154, respectively. This finding suggests that the giant cluster plays a role of knowledge integrator in learning across subgroups even when the increase in the percentage of bridges is rather small.

Argote and Ingram (2000) introduced the notion of knowledge reservoirs, in which diverse ideas and knowledge are preserved for future use. The literature on learning suggests that learning processes enhance system performance as lower performers adopt good ideas from higher performers, while discarding bad ideas over time. Improvement in organizational learning mainly comes from the availability of diverse knowledge in the system (Cattani & Kim, 2021; Fang et al., 2010; March, 1991; Posen et al., 2013). To check whether the integration effect stems from such diversity, we investigate the level of diversity in ideas across a different range of S for $0 \leq p < .05$. We use the measure of diversity developed by Posen et al. (2013), which reflects how well individuals in the largest subgroup collectively preserve good ideas, or those ideas that match the values of corresponding elements in reality. Even when S is small for $0 \leq p < .01$, the system starts with as many diverse good ideas as the case for $.01 \leq p < .05$. However, the problem in the former is that individuals cannot exploit the available diversity to improve their performance because they primarily work in information silos—that is, few bridges allow individuals in one subgroup to capture good ideas from other subgroups. Consequently, isolated subgroups as a whole tend to lose a substantial number of good ideas that were available at the beginning (see Appendix E.1). In contrast, when a giant cluster is present for $.01 \leq p < .05$, it acts as a knowledge integrator whose role is to bring together diverse good ideas and knowledge by connecting previously isolated subgroups. What is surprising here is that the percentage increase in bridges is rather small, ranging between 1 and 5% compared to the lower bound (i.e., 0%) of the parameter space. In sum, good ideas are better retained in the system when there is a giant cluster.

5.4 | Acceleration effect: Negative effect of fast cross-subgroup learning due to abundant bridges

Now, we show the acceleration effect, which occurs due to excessively fast learning when bridges become abundant. Our model setup permits us to test this claim by controlling p. Note that beyond the saturation point (i.e., $p \geq .05$), the integration effect vanishes completely since no more isolated subgroup remain for the giant cluster to integrate with additional bridges. On the other hand, we expect that additional bridges will reduce average path length and accelerate learning processes.

To show this acceleration effect more systematically, we run 100 repeated simulations by using the range for $.05 \leq p \leq 1$. For each simulation of cross-subgroup learning in this range of larger percentages of bridges, a value of p is randomly drawn from a uniform distribution between .05 and 1. Then, we compare this simulation result with the previous one when a percentage of bridges is smaller (i.e., $.01 \leq p < .05$).

As shown in Table 1, the repeated simulations show that average learning performance is significantly lower (with T-value of 13.34) in the larger percentage case than in the smaller percentage case (values for learning performance are 0.506 and 0.757, respectively). This finding implies that cross-subgroup learning becomes less efficacious when the percentage of bridges is larger, given that a giant cluster is formed.

Now, one may wonder why learning performance declines after reaching a peak at the saturation point of $p \cong .05$. Recall that beyond the saturation point, all individuals in the system are connected to one another with no disconnected parts—that is, the value of S is at its theoretical maximum of $S = 1$. Performance degradation, then, may stem from the decreasing average path length with the increasing percentage of bridges. Our analysis in fact shows that additional bridges after the saturation point tend to reduce the average path length L (e.g., when $p = .05$, $L = 6.93$, when $p = .15$, $L = 5.04$, and when $p = .35$, $L = 4.33$). The literature on small-world networks has established that smaller average path length leads to faster diffusion of an idea (e.g., Fang et al., 2010; Watts & Strogatz, 1998). In the context of learning, ideas held by high-performing individuals are more likely to be appreciated and assimilated by others. Our results in Appendix E.1 indeed show that performance degradation is associated with the addition of bridges beyond the saturation point, which tends to reduce the number of good ideas over time. Our experiment in Appendix E.2 further reveals that the acceleration effect is associated with a weakened ability to preserve diverse ideas and knowledge for future use. In particular, faster diffusion of some good ideas hinders the preservation of other good ideas.

5.5 | The negative effects of abundant bridges may not represent social reality

In summary, our model shows that adding more bridges to the system is not always conducive to learning. Beyond the saturation point (i.e., $p \geq .05$), additional bridges tend to have the deleterious effects on learning outcomes. In this parameter range, additional bridges no longer act as integrators because no more isolated subgroups are left to be integrated. Numerous empirical studies, however, have shown that when a giant cluster is observed, there exist isolated parts as well (e.g., Cattani et al., 2008; Fleming et al., 2007; Gulati et al., 2012; Onnela et al., 2007; Phillips, 2011; Uzzi & Spiro, 2005).

Furthermore, when p approaches 1, our model looks more like a random network model with too many bridges and attenuation of within-subgroup ties. Although most small-world network models embrace these idealized properties, they deviate from real-world social networks, where bridges tend to account for a tiny fraction of all ties. For example, bridges in Samsung's inventor collaboration network in 2006 made up only 1.7% of all ties. Although the Santa Fe Institute was renowned for interdisciplinary research, bridging ties across disciplinary boundaries accounted for only a tiny fraction of all ties in the collaboration network at the institute (Girvan & Newman, 2002). In general, within-subgroup ties are far more numerous than between-subgroup ties in social networks (e.g., Girvan & Newman, 2002; Newman & Park, 2003).

In sum, all these findings suggest that the deleterious effects observed above may be a theoretical artifact. In particular, when it is costly for individuals to build bridges across subgroups, as in inventor collaboration networks, a giant cluster can emerge, but it is unlikely to gobble up other disconnected parts completely. Then, the aforementioned deleterious effect may not be realized under this condition, where additional bridges are more likely to be conducive to cross-subgroup learning.

6 | SIMULATIONS ON DATA FROM A REAL-WORLD NETWORK

We now run simulations on data from a real-world network to examine the robustness of our key findings. Given the data constraints, we show two things. First, an increase in the number of bridges speeds up simple information diffusion. Second, learning improvement is better with a giant cluster than without it.

The results presented in Figure 6a are generated by running simulations of our information diffusion model on a real-world network. In this simulation, we consider the largest cluster at Samsung in 2006 as a giant cluster, utilizing it as a starting point. To add more bridges to this cluster, we choose a node at random from the cluster. With probability r, we remove the existing link for the chosen node and reconnect this node to another node chosen uniformly at random in the giant cluster, with duplicate links forbidden. Note that the numbers of nodes and links must remain constant across different networks in order to minimize confounding effects due to arbitrary increases in these numbers. Given this rewiring procedure, the real-world network has the lowest number of bridges.

With an increase in rewiring probability r, the number of bridges within the network increases, while the numbers of nodes and links remain constant. For the given range $0 \leq r < .3$, additional bridges tend to reduce average path length L. For example, when $r = 0$, $L = 5.84$, and when $r = .1$, $L = 4.92$. As was the case in the previous simulation results using our network model, a positive association is observed between the number of bridges and the speed of information diffusion—the larger the number of bridges, the faster the information diffusion.

Now, we show the robustness of our key findings with respect to cross-subgroup learning. The results in Figure 6b were generated by running simulations of learning dynamics on the real-world network. Like the previous simulation results using our network model, the results here show that learning performance is significantly higher in the presence of a giant cluster than otherwise, suggesting that the giant cluster plays the role of a knowledge integrator in cross-subgroup learning. The results on the right were obtained by applying the same rewiring procedure that was applied in the simple information diffusion simulation above. We consider the largest subgroup at Samsung in 2006 as a giant cluster and then add new bridges to the largest subgroup by tuning r, while simultaneously removing the existing link within a given subgroup. The results suggest that the acceleration effect (i.e., the negative effect of excessively fast learning) would exist even when we run simulations with the data from the real-world network.[8]

[8]Here, we rewire ties between inventors only from the largest cluster of Samsung 2002–2006 (i.e., adding bridges only between inventors in the largest cluster) and run learning simulations within this cluster. In other words, we ignore unconnected parts in this experiment because our focus here is to demonstrate that the acceleration effect will exist even when we run simulations with the real-world network. Therefore, in this experiment, the integration effect from connecting isolated clusters must not exist *a fortiori*.

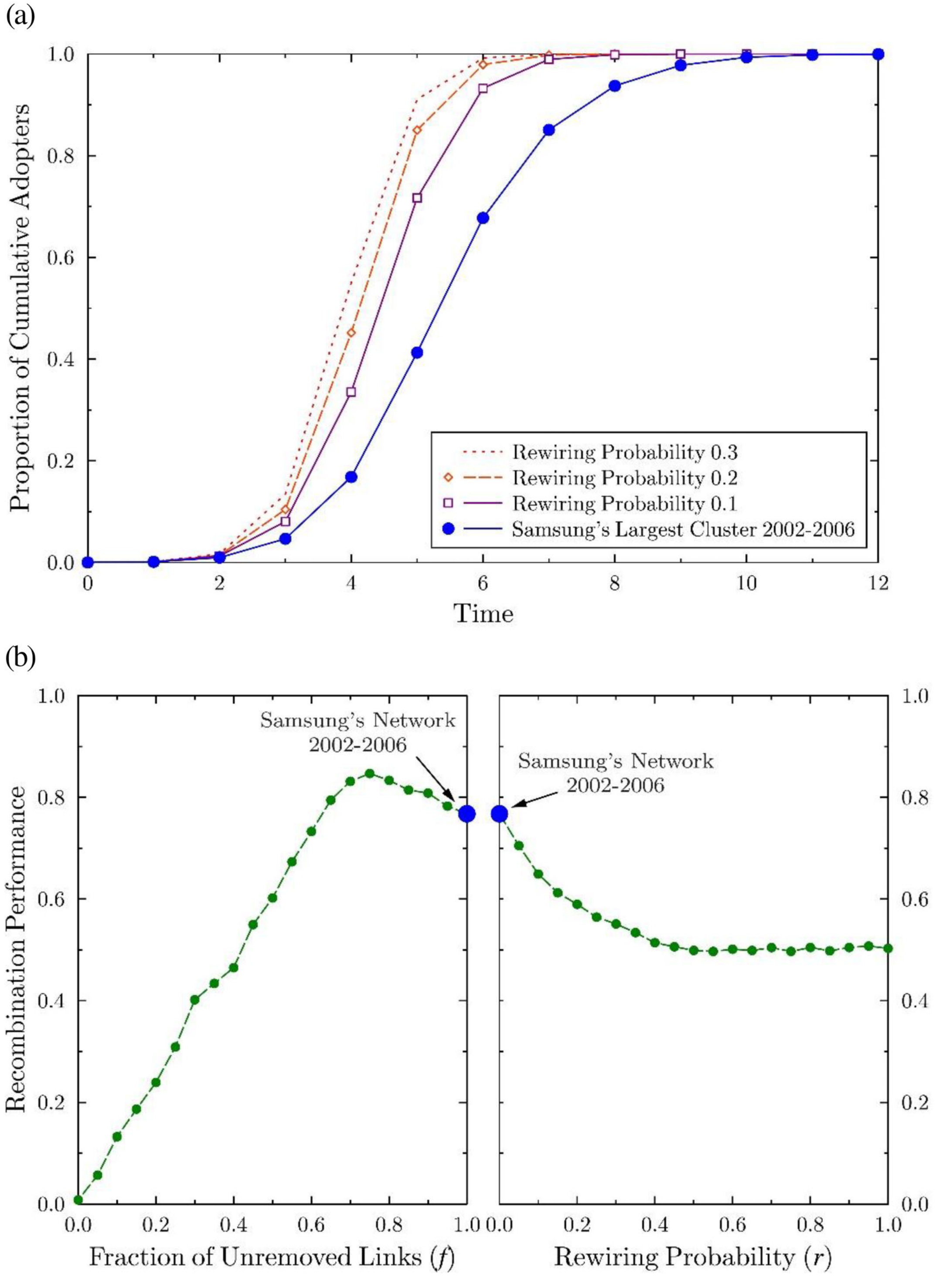


FIGURE 6 Legend on next page.

In the results on the left, we investigate how the removal of bridges from the giant cluster affects learning performance. This type of analysis is known as "breakdown analysis" or "reverse percolation" in statistical physics (e.g., Albert et al., 2000; Newman, 2010). Again, we consider the same largest subgroup at Samsung (2002–2006) as a starting point. First, we calculate the betweenness centralities of all the links in the network. Then, we remove the links from the real-world network with the highest betweenness centralities. Let f denote the fraction of unremoved links. For example, if we remove the top 70% of links, $f = 0.3$; then, the links in the bottom 30% in terms of betweenness centrality remain in the network. The results show that when $f < 0.75$, interpersonal learning performance increases. When $f \geq 0.75$, on the other hand, learning performance declines. The overall pattern in learning performance in Figure 6b is more or less consistent with that from our model in Figure 5.

In sum, the results in Figure 6a,b altogether show the robustness of our key findings. First, given that the numbers of nodes and links remain constant, the addition of bridges to the system results in accelerating information diffusion. Second, learning improvement across subgroups is higher with a giant cluster than without it. When there are too many bridges, learning is less efficacious.

7 | DISCUSSION

In his seminal work, Granovetter (1973, p. 1360) raised a fundamental question of why large-scale social dynamics (e.g., an upsurge of nation-wide protests or widespread diffusion of innovative ideas) occur sometimes and why they do not at other times. He emphasized the roles of bridges in addressing this question. Watts and Strogatz's (1998) theoretical work on small-world networks has had a major impact on our understanding of the dynamic implications of social networks, one of which is that an increase in the proportion of bridges to all ties in a system accelerates diffusion throughout the system. Research has made significant progress in elucidating how bridges affect diffusion phenomena (e.g., Balachandran & Hernandez, 2018; Cattani & Ferriani, 2008; Fang et al., 2010; Lee et al., 2016; Posen et al., 2020; Vasudeva et al., 2013). However, the small-world network framework has limitations, one of which results from the simplifying assumption that all individuals are connected to one another with no disconnected parts. Empirical studies indicate that this assumption tends to be violated in many large social networks (e.g., Cattani et al., 2008; Fleming et al., 2007; Gulati et al., 2012; Kogut et al., 2007; Onnela et al., 2007; Phillips, 2011; Uzzi & Spiro, 2005).

FIGURE 6 Simulations on real-world networks (Samsung's Network 2002-2006). (a) Effect of bridges on diffusion speed. (b) Effect of bridges on learning performance. Results from simulations on the real-world networks are consistent with those from the idealized model. In panel (a), the speed of information diffusion is positively associated with the proportion of bridges. On the right side of panel (b), learning performance is negatively associated with the number of additional bridges, since the addition of more bridges to the system reduces its average path length for the given range. On the left side of panel (b), when we remove up to 25% of the links in the network, learning performance increases. When the removal exceeds 25%, however, learning performance declines. The pattern of learning performance in panel (b) is, therefore, consistent with the results from our idealized model in Figure 5. The number of nodes in the largest cluster of Samsung 2002–2006 is 5273. On the left side, the links with the lowest centralities remain in the network. For example, if the fraction of unremoved links is 30% ($f = 0.3$), we remove the top 70% of links, and the links in the bottom 30% in terms of betweenness centralities remain in the network.

To improve our understanding of the dynamic implications of social networks in the presence of fragmented subgroups, we develop computational models by employing tools from percolation theory (Christensen & Moloney, 2005; Stauffer & Aharony, 2018), which allows us to relax the connected network assumption above. We find a less-well understood role of bridges, which is related to how widely ideas, influences, and information will be diffused. In the context of cross-subgroup learning, we find that when the percentage of bridges in the system is below 1%, the efficacy of learning is poor. In the vicinity of 1%, however, a slight increase in the number of bridges leads to a quantum jump in the efficacy of learning. This dramatic impact on learning stems from the threshold-like structural characteristic in our network model, which reflects the essence of some large, social networks in reality. When the percentage of bridges is below threshold p_c ($p_c \cong .01$), the whole system is merely a collection of fragmented parts. In the vicinity of this threshold, however, adding a tiny fraction of bridges causes a sizeable percentage of subgroups to be connected together all of a sudden. This sudden structural change with the emergence of a giant cluster, in turn, prompts the diffusion process to affect a considerably larger fraction of the population by connecting previously isolated subgroups, thereby boosting exchanges of diverse ideas across subgroups.

7.1 | Danger of blindly applying the small-world network tools

Our work suggests that blind application of small-world network tools in social dynamics research may potentially result in invalid implications. In this study, we revisit the received view that adding too many bridges to a system results in excessively fast diffusion of some ideas while driving out others, thereby impairing learning (Fang et al., 2010). Our results show that cross-subgroup learning benefits from the addition of bridges insofar as subgroups are not completely connected to one another. Once all subgroups are completely connected and no isolated parts remain, additional bridges tend to have deleterious effects on learning performance, as in the received view. In this parameter range, however, adding bridges via rewiring procedures eventually leads to both an excess of bridges and attenuation of within-subgroup ties, deforming the subgroup structure inherent in social networks. That is, the resultant network structures tend to diverge from social reality. The upshot is that danger of applying the small-world network tools lies in misinterpreting such results as if they could occur in reality.

7.2 | Formation of a giant cluster could be a structural precondition for sudden large-scale change

Understanding the threshold-like structural condition identified in this study provides a glimpse of how dynamics can unfold throughout a social system. Kirkpatrick (2011) illustrated how anti-government influences can spread from one person to another and to many via Facebook, resulting in an upsurge of nation-wide protests. Abrahamson and Rosenkopf (1997) articulated that in the case of diffusion of innovation, information about the product is spreading from one potential adopter to another. For example, Hotmail stimulated user adoption via one of the first viral marketing tools, attracting 10 million adopters within a year.

Our findings suggest that giant cluster formation could be a structural precondition for large-scale social change. The ability to detect a giant cluster may facilitate prognostication of the possibility of large-scale diffusion. In Appendix F, we outline our methods for detecting giant clusters and analyzing their structural properties using readily available social network data from social media. Applying these methods, entrepreneurs and established firms can stimulate user adoption by targeting clusters of massively aggregated subgroups and spreading viral messages about their new products or services.

In addition, companies can enhance learning and collaboration across different subunits by shaping the evolution of informational networks to form a giant cluster. Such cross-fertilization is known to be crucial for learning and innovation (Balachandran & Hernandez, 2018; Henderson & Cockburn, 1996). However, we often observe that people in different subunits tend to work in information silos, largely due to the natural boundaries that form between subunits with distinct expertise. As discussed earlier, Samsung was no exception in the 1980s. To mitigate this silo effect, top managers at Samsung introduced policies and practices to build bridges across different subunits through job rotation and cross-functional meetings. In addition, Samsung's patent office regularly examined all filed patents, identified researchers with similar interests, and encouraged collaboration between them. Over time, the company witnessed a sudden emergence of a giant cluster, where individuals could easily gain access to information and facilitate exchanges of ideas and knowledge with one another. This recipe for enhancing cross-functional learning can be benchmarked and further developed by other companies.

ACKNOWLEDGMENTS

We are grateful for comments received from Sarath Balachandran, Mitali Banerjee, Felipe Csaszar, Jerker Denrell, J.P. Eggers, Christina Fang, Scott Ganz, Deepak Hegde, Jerry Kim, Bruce Kogut, Damon Phillips, Joe Porac, Phanish Puranam, Hart Posen, Jasjit Singh, Jose Uribe, Dan Wang, Phebo Wibbens, and Maggie Zhou. We also wish to thank the seminar participants at the Wharton School of the University of Pennsylvania, Columbia University, INSEAD, New York University, London Business School, the University of Michigan, and the University of Wisconsin-Madison.

DATA AVAILABILITY STATEMENT

Data availability statement is irrelevant to our paper.

ORCID

Sungyong Chang https://orcid.org/0000-0002-7374-3644
Jeho Lee https://orcid.org/0000-0002-8471-3627
Jaeyong Song https://orcid.org/0000-0002-7436-7265

REFERENCES

Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, *8*(3), 289–309.

Adams, J., & Roscigno, V. J. (2005). White supremacists, oppositional culture and the World Wide Web. *Social Forces*, *84*(2), 759–778.

Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attach tolerance of complex networks. *Nature*, *406*(6794), 378–382.

Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes*, *82*(1), 150–169.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*(2), 203–226.

Balachandran, S., & Hernandez, E. (2018). Networks and innovation: Accounting for structural and institutional sources of recombination in brokerage triads. *Organization Science*, *29*(1), 80–99.

Baum, J. A., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from? *Industrial and Corporate Change*, *12*(4), 697–725.

Blau, P. M., & Schwartz, J. E. (1984). *Crossing social circles*. Academic Press.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399.

Cattani, G., & Ferriani, S. (2008). A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. *Organization Science*, *19*(6), 824–844.

Cattani, G., Ferriani, S., Negro, G., & Perretti, F. (2008). The structure of consensus: Network ties, legitimation, and exit rates of US feature film producer organizations. *Administrative Science Quarterly*, *53*(1), 145–182.

Cattani, G., & Kim, J. H. (2021). Variety-seeking, learning and performance. *PLoS One*, *16*(3), e0247034.

Centola, D. (2015). The social origins of networks and diffusion. *American Journal of Sociology*, *120*(5), 1295–1338.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734.

Christensen, K., & Moloney, N. R. (2005). *Complexity and criticality*. Imperial College Press.

Doreian, P., & Stokman, F. N. (1997). *Evolution of social networks* (Vol. 1). Psychology Press.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*(1), 17–60.

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organization Science*, *21*(3), 625–642.

Flache, A., & Macy, M. W. (2011). Small worlds and cultural polarization. *Journal of Mathematical Sociology*, *35*(1–3), 146–176.

Fleming, L., King, C., III, & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, *18*(6), 938–954.

Gans, H. J. (1982). *Urban villagers*. Free Press.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Gulati, R., Sytch, M., & Tatarynowicz, A. (2012). The rise and fall of sm all worlds: Exploring the dynamics of social structure. *Organization Science*, *23*(2), 449–471.

Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *The Rand Journal of Economics*, *27*(1), 32–59.

Katz, E. (1961). The social itinerary of technical change: Two studies on the diffusion of innovation. *Human Organization*, *20*(2), 70–82.

Kirkpatrick, D. (2011). *The Facebook effect: The inside story of the company that is connecting the world*. Simon and Schuster.

Kogut, B., Urso, P., & Walker, G. (2007). Emergent properties of a new financial market: American venture capital syndication, 1960–2005. *Management Science*, *53*(7), 1181–1198.

Last, B. J., & Thouless, D. J. (1971). Percolation theory and electrical conductivity. *Physical Review Letters*, *27*(25), 1719–1721.

Lee, D., Goh, K. I., Kahng, B., & Kim, D. (2010). Complete trails of coauthorship network evolution. *Physical Review E*, *82*(2), 026112.

Lee, E., Lee, J., & Lee, J. (2006). Reconsideration of the winner-take-all hypothesis: Complex networks and local bias. *Management Science*, *52*(12), 1838–1848.

Lee, J., Song, J., & Yang, J. S. (2016). Network structure effects on incumbency advantage. *Strategic Management Journal*, *37*(8), 1632–1648.

Lim, M., Metzler, R., & Bar-Yam, Y. (2007). Global pattern formation and ethnic/cultural violence. *Science*, *317*, 1504–1544.

Malamud, B. D., Morein, G., & Turcotte, D. L. (1998). Forest fires: An example of self-organized critical behavior. *Science*, *281*(5384), 1840–1842.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 60–67.

Miller, K. D., Zhao, M., & Calantone, R. J. (2006). Adding interpersonal learning and tacit knowledge to March's exploration-exploitation model. *Academy of Management Journal*, *49*(4), 709–722.

Newman, M. E. (2010). *Networks*. Oxford University Press.

Newman, M. E., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, *68*(3), 036122.

Newman, M. E., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(suppl 1), 2566–2572.

Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., & Barabási, A. L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7332–7336.

Peitgen, H.-O., Jurgens, H., & Saupe, D. (1992). *Chaos and fractals: New frontiers of science*. Springer-Verlag.

Phillips, D. J. (2011). Jazz and the disconnected: City structural disconnectedness and the emergence of a jazz canon, 1897–1933. *American Journal of Sociology*, *117*(2), 420–483.

Posen, H. E., Lee, J., & Yi, S. (2013). The power of imperfect imitation. *Strategic Management Journal*, *34*(2), 149–164.

Posen, H. E., Yi, S., & Lee, J. (2020). A contingency perspective on imitation strategies: When is "benchmarking" ineffective? *Strategic Management Journal*, *41*(2), 198–221.

Rosenberg, N. (1979). Technological interdependence in the American economy. *Technology and Culture*, *20*(1), 25–50.

Sayama, H., Kaufman, L., & Bar-Yam, Y. (2000). Symmetry breaking and coarsening in spatially distributed evolutionary processes including sexual reproduction and disruptive selection. *Physical Review E*, *62*, 7065–7069.

Schilling, M. A., & Fang, C. (2014). When hubs forget, lie, and play favorites: Interpersonal network structure, information distortion, and organizational learning. *Strategic Management Journal*, *35*(7), 974–994.

Shibanai, Y., Yasuno, S., & Ishiguro, I. (2001). Effects of global information feedback on diversity: Extensions to Axelrod's adaptive culture model. *Journal of Conflict Resolution*, *45*(1), 80–96.

Snijders, T. A. (2001) The statisticalevaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361–395.

Stauffer, D., & Aharony, A. (2018). *Introduction to percolation theory*. CRC Press.

Stokman, F. N., & Doreian, P. (2001). Evolution of social networks. Part II. *Journal of Mathematical Sociology*, *25*, 1–4. Special issue.

Subramani, M. R., & Rajagopalan, B. (2003). Knowledge-sharing and influence in online social networks via viral marketing. *Communications of the ACM*, *46*(12), 300–307.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, *111*(2), 447–504.

Vasudeva, G., Zaheer, A., & Hernandez, E. (2013). The embeddedness of networks: Institutions, structural holes, and innovativeness in the fuel cell industry. *Organization Science*, *24*(3), 645–663.

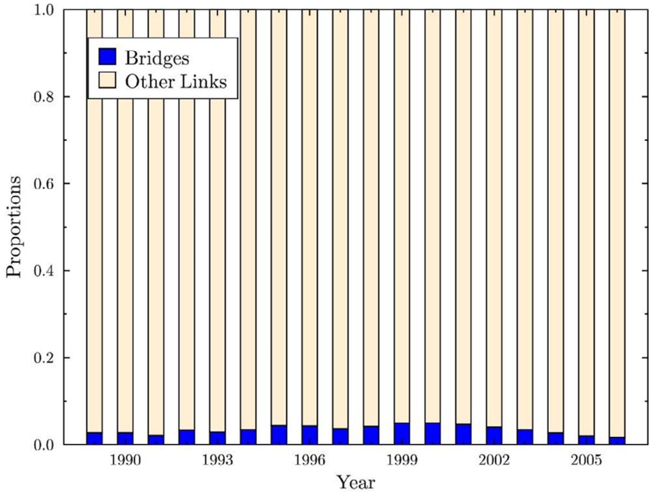
Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, *105*(2), 493–527.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.

Yayavaram, S., & Ahuja, G. (2008). Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Administrative Science Quarterly*, *53*(2), 333–362.

How to cite this article: Chang, S., Lee, J., & Song, J. (2023). Giant cluster formation and integrating role of bridges in social diffusion. *Strategic Management Journal*, 1–36. https://doi.org/10.1002/smj.3527

APPENDIX A: RELATIVE GROWTH IN THE NUMBER OF BRIDGES AT SAMSUNG



The proportion of bridges to the total number of ties in Samsung's network remained quite limited for many years. The average proportion of bridges was 0.035. The largest proportion was 0.05 in 2000. We identified bridges from other links using the link classification methodology of Lee et al. (2010).

APPENDIX B: THEORETICAL STUDIES OF NETWORKS

	Connectedness assumption	Emergence of a giant cluster	Existence of subgroups	Within-subgroup connectivity	Between-subgroup connectivity
Watts and Strogatz (1998)	Yes	Irrelevant	Yes	Tunable	Tunable
Fang et al. (2010)	Yes	Irrelevant	Yes	Tunable	Tunable
Erdős and Rényi (1960)	No	Relevant	No	Unrealistically low	Unrealistically high
Our model	No	Relevant	Yes	Tunable	Tunable

APPENDIX C: PARAMETERS

Model	Parameters	Remarks	Range of parameter values analyzed
Network model	n	Number of individuals in a network	1000
	g	Number of individuals in each subgroup when $p = 0$	20
	p	Proportion of between-subgroup ties (bridges)	0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1
Interpersonal learning model	m	Dimension for ideas	120
	ψ	Degree of complexity	4
	θ	Learning rate	0.5

APPENDIX D: ANALYTICAL RESULTS OF DIFFUSION EXTENT

Our model includes n individuals who are equally divided into n/g subgroups ($n/g < n$). As described in the main text, the diffusion process begins with individual i who is exposed to new information. Let $s_i(n/g)$ denote the normalized size of the subgroup to which a randomly selected individual i belongs. Let $p_i(n/g)$ denote the probability that i will be chosen from this subgroup. In the limit of large time steps, the extent of information diffusion, $D_i(n/g)$, will be equal to the size of that subgroup. Thus, the average diffusion extent $E(D_i(n/g))$ can be described as follows:

$$E(D_i(n/g)) = \sum\nolimits_{i=0}^{n} D_i(n/g) = \sum\nolimits_{i=0}^{n} p_i(n/g) \cdot s_i(n/g) \quad (1)$$

To determine the relationship between Equation (1) and the normalized size of the largest subgroup, we need to investigate its size distribution. Consider first the network of each subgroup. Random graph theory predicts that when the average link per node is larger than 1, all individuals will be interconnected if their ties are randomly made. Since the average link per node is 6 in our model, all individuals in each subgroup tend to be interconnected, implying that the average size of the largest subgroup in each subgroup approaches the size of a subgroup, g.

Let $L(n/g)$ denote the normalized size of the largest subgroup in the entire network. As $L(n/g)$ increases, the probability of having an initial spreader of new information from the largest subgroup increases proportionally. In this case, the probability of having the information spreader from the largest subgroup is $L(n/g)$, and the diffusion extent is $L(n/g)$. On the other hand, the probability of having the information spreader from the set of all other subgroups will be 1-$L(n/g)$, and the diffusion extent is approximately equal to g. Therefore, the average diffusion extent can be approximated as follows:

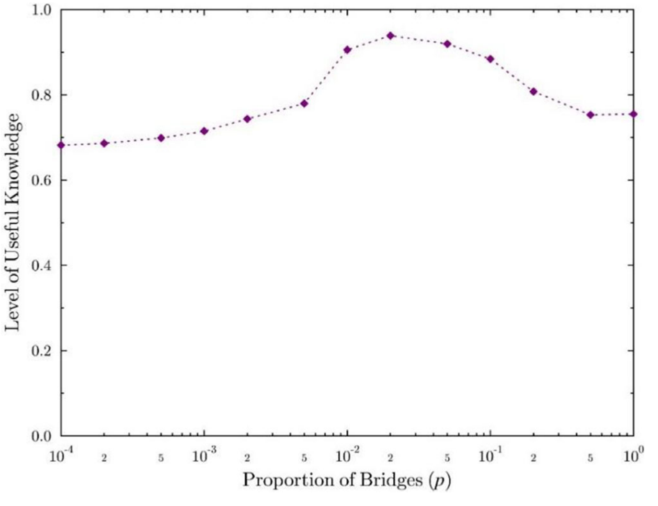
$$E(D_i(n/g)) \approx L(n/g)L(n/g) + (1 - L(n/g))g = L(n/g)^2 + (1 - L(n/g))g \quad (2)$$

When the number of subgroups *n/g* is sufficiently large, *g* will be close to 0. For example, in our toy model, $n/g = 50$ and the size (proportion of individuals in a subgroup to the total number of individuals in the organization) of each subgroup = 0.02. Thus, when the number of subgroups in the organization is sufficient, the average extent of knowledge diffusion can be approximated as follows:

$$E(D_i(n/g)) \approx L(n/g)^2 \quad (3)$$

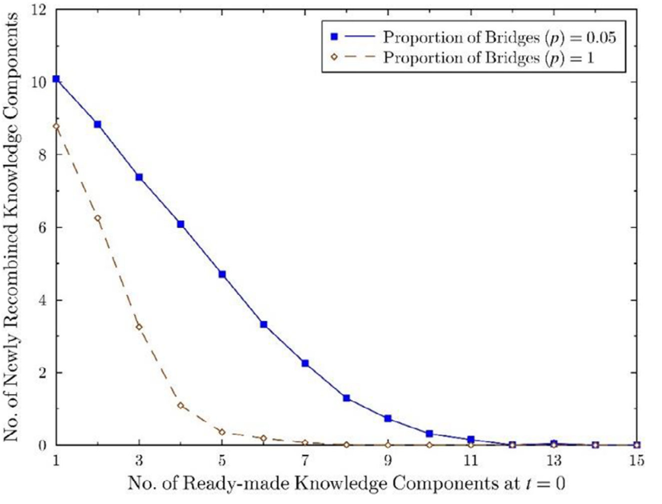
APPENDIX E: BRIDGES, LARGEST SUBGROUP, AND KNOWLEDGE DIVERSITY

E.1 | Effect of bridges on number of good ideas



Panel E.1 shows that in the range of *p* between 0 and 0.01, the system tends to lose a substantial number of good ideas over time. When a giant cluster is present in the range of *p* between 0.01 and 0.05, the addition of bridges to the system tends to help retain more good ideas over time. Beyond $p \cong .05$, however, additional bridges tend to reduce the number of good ideas.

E.2 | Cost of being "too small"

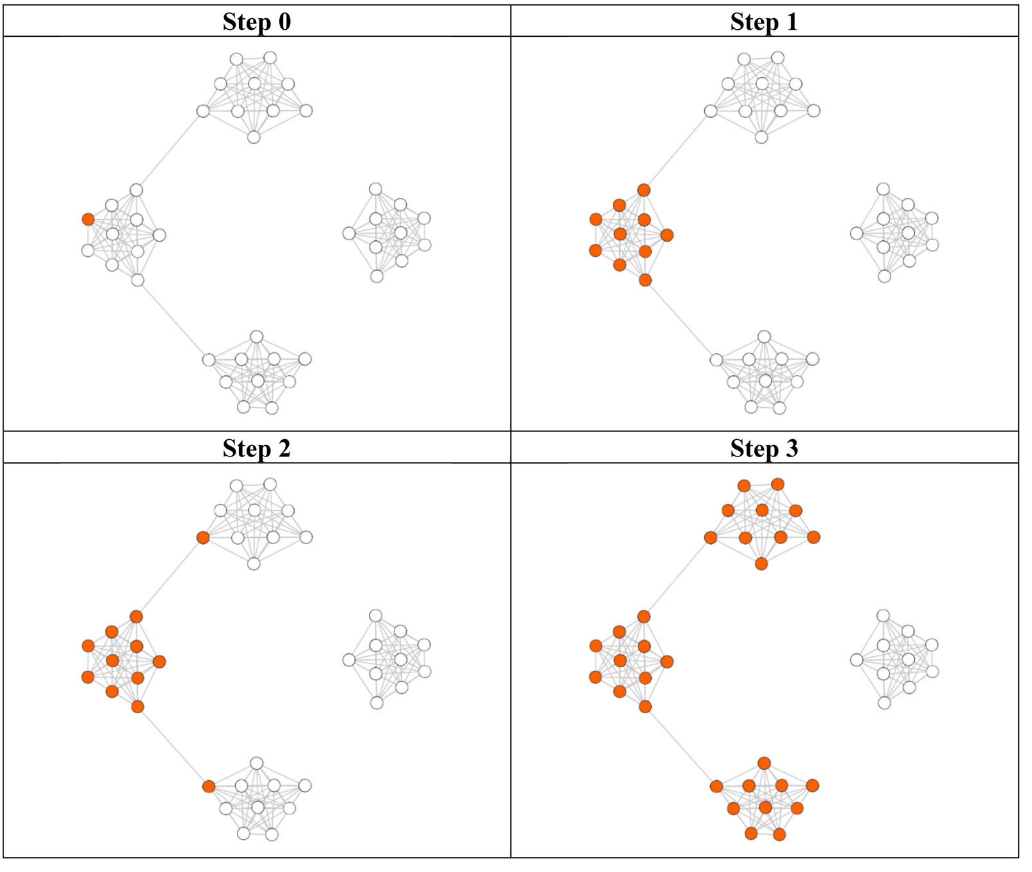


The results show that the number of newly recombined knowledge components via cross-subgroup learning is smaller in the "too-small-world" case (i.e., $p = 1$) than in the modest setting (i.e., $p = 0.05$). We define knowledge component as a unit of knowledge that carries a positive value only if all ideas in the component are good. In the-too-small-world case, the faster diffusion of the ready-made knowledge components, which are given initially by the experimental setup, deters building of new knowledge components. The number of nodes in the networks for both cases is 1000, and the number of links is 6000. Each data point here is averaged over 200 simulations.

APPENDIX F: METHODS FOR DETECTING THE PRESENCE OF A GIANT CLUSTER

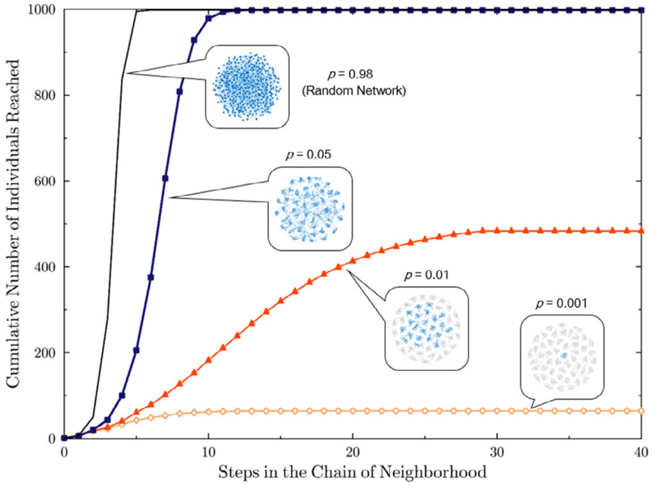
Here, we introduce methods to detect a giant cluster and determine its size across a variety of social media. These methods may be utilized in viral marketing campaigns when companies launch their new products or services. At the heart of them is distribution sequence, which represents expansion of the chain of a neighborhood of a subgroup (Lee et al., 2016; Watts, 1999). For example, let us consider the individual marked with the closed circle at step 0 in Appendix F.1. This individual has nine nearest neighbors, or direct contacts, all of whom are one step away from the focal individual. If we take one step further to the second-nearest neighborhood, there are two new neighbors. If we take one more step to the third-nearest neighborhood, there are 18 new neighbors. After this, we cannot take a further step because the subgroup on the right is isolated from the larger one on the left. To sum up, the number of new neighbors at every step in the chain of neighborhood relations expands: 9, 2, 18. A distribution sequence is a cumulative sequence of this new neighbor sequence, which expands: 9, 11, 29.

F.1. | Illustration of expansion of distribution sequence through chain of neighborhood



The visualization of a distribution sequence makes it easier for empiricists and practitioners to detect the presence of a giant cluster and determine its size, thereby prognosticating the possibility of large-scale diffusion. The examples in Appendix F.2 illustrate typical expansion patterns of the largest subgroup for varying values of p in our model of social networks. Each graph shows an increase in the cumulative number of people reached as a randomly chosen individual's influence in a given subgroup moves through the chain of its neighborhood one step at a time. As the proportion of bridges increases, the largest subgroup becomes larger and larger up to $p = .05$, beyond which an S-shaped pattern becomes more pronounced and steeper with increasing p .

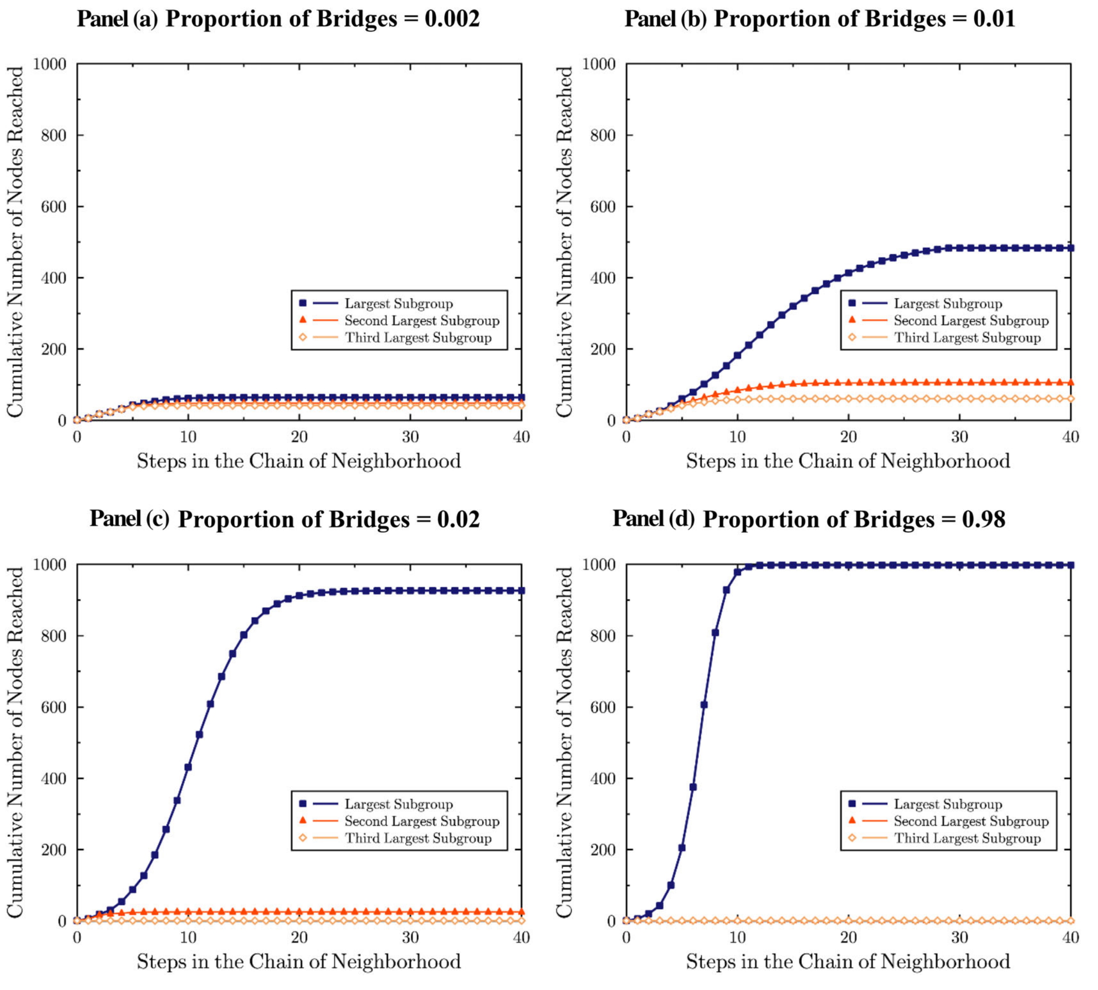
F.2. | Expansion patterns of distribution sequence for varying values of *p*



Note. (Color online) To calculate the cumulative number of nodes reached, we first choose an individual in the largest subgroup of each network. Then, from the selected individual, we calculate a series of distribution sequences by tracing additional steps in the chain of neighborhood. In our idealized networks with varying values of *p*, the blue-colored nodes are individuals in the largest subgroup in those networks. For all networks, the number of nodes is 1000, and the number of links is 6000. The ratio of links to nodes is 6, which is observed in Samsung's collaboration network in 2006.

Appendix F.3 shows a comparison of expansion patterns of distribution sequences for different subgroups. Obviously, the expansion pattern of the giant cluster with larger *p* is markedly larger and steeper (with a more pronounced S-shaped pattern) than those of others. Although these patterns are structural properties reflecting the different characteristics of chains of neighborhoods, these patterns have dynamic implications—the dynamics of simple diffusion of new information will roughly match these patterns, as shown in the Simple Information Diffusion section. In other words, the knowledge of such an expansion pattern offers a glimpse of how a diffusion process may propagate throughout the system if it actually occurs.

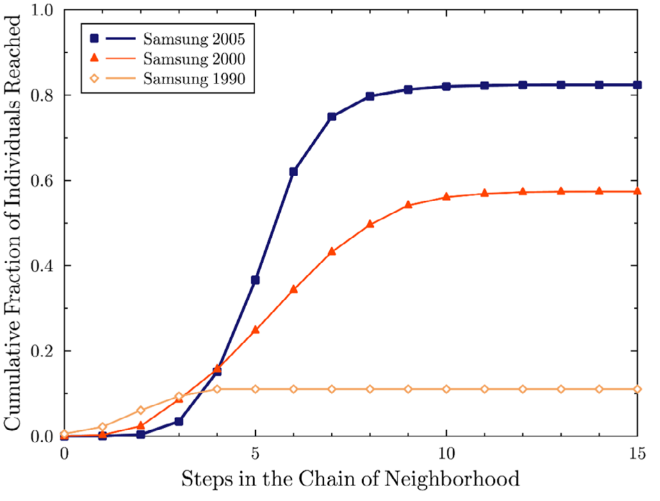
F.3. | Comparison of expansion patterns of distribution sequence by subgroup type



For all networks, the number of nodes is 1000, and the number of links is 6000. The ratio of links to nodes is 6, which is observed in Samsung's collaboration network in 2006. Each data point here is averaged over 200 simulations.

In addition, the method above can be utilized to detect the presence of a giant cluster and determine its size in a real-world network. For example, we apply this method to detect the structural properties of Samsung's inventor collaboration networks over time. Appendix F.4 exhibits the expansion patterns of distribution sequences for three selected largest subgroups in Samsung's inventor collaboration networks for 1990, 2000, and 2005, respectively. The largest subgroup for 2005, which includes 82% of all inventors at Samsung's semiconductor division at that time, is larger than the other two. Furthermore, the S-shaped pattern of this subgroup is steeper than those of the others.

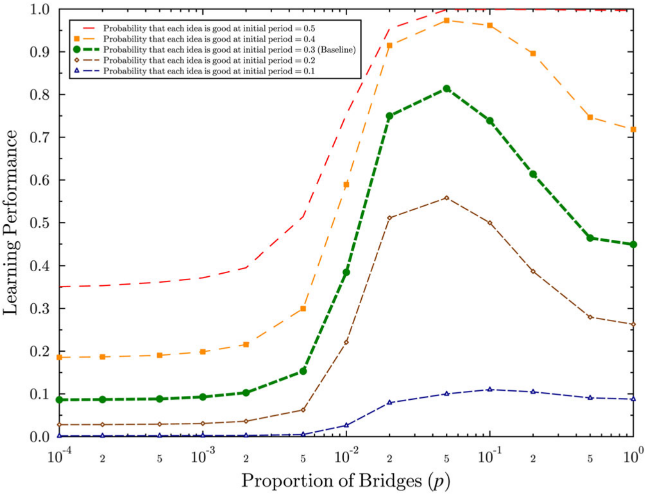
F.4. | Expansion patterns of distribution sequence of largest subgroups at Samsung



Note. (Color online) For each network, we choose a random person in the largest subgroup and calculate the distribution sequence. In Samsung's 2005 network, the number of individuals in the largest subgroup is 4925 (82.4% of the total of 5977 individuals). In Samsung's 2000 network, the number of individuals in the largest subgroup is 1592 (57.4% of the total of 2774 individuals). In Samsung's 1990 network, the number of individuals in the largest subgroup is 20 (11.05% of the total of 181 individuals).

Recall that the patterns here do not represent intertemporal diffusion dynamics. They only reflect the structural properties of the largest subgroups as well as their chains of the neighborhood relations. Based on this structural information, however, we can predict that if new information spreads through these subgroups, it will be far more widespread in the largest subgroup for 2005 than in the others. The method here can also be utilized for viral marketing campaigns when companies launch their new products or services. For instance, one can apply this method to detect the presence and determine the size of a giant cluster across a variety of social media. We believe that this type of analysis can help companies in determining suitable marketing channels for spreading their viral messages.

APPENDIX G: SENSITIVITY TO THE PROBABILITY THAT EACH IDEA IS GOOD AT INITIAL PERIOD



We examine sensitivity to the probability that each idea is good at period 0. Here, we vary this probability ranging from 0.1 to 0.5. In the baseline model, we set the probability at 0.3. As this probability increases, the performance curve shifts upward. Nonetheless, the shapes of curves look similar unless the probability is too high (e.g., 0.5). The problem complexity is set at $\psi = 4$. The learning rate is set at $\theta = 0.5$. Learning performance is normalized by dividing each outcome by the highest one. Each data point here is averaged over 200 simulations.