

# review articles

DOI:10.1145/1400214.1400232

## Internet-based data on human interaction connects scientific inquiry like never before.

BY JON KLEINBERG

# The Convergence of Social and Technological Networks

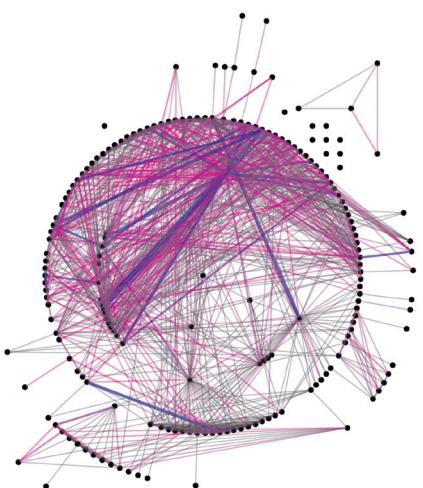
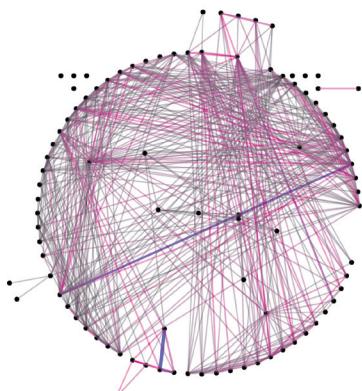
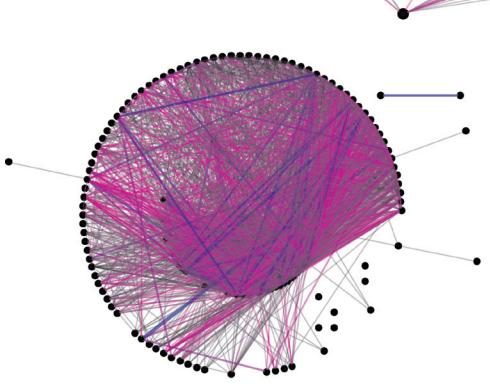
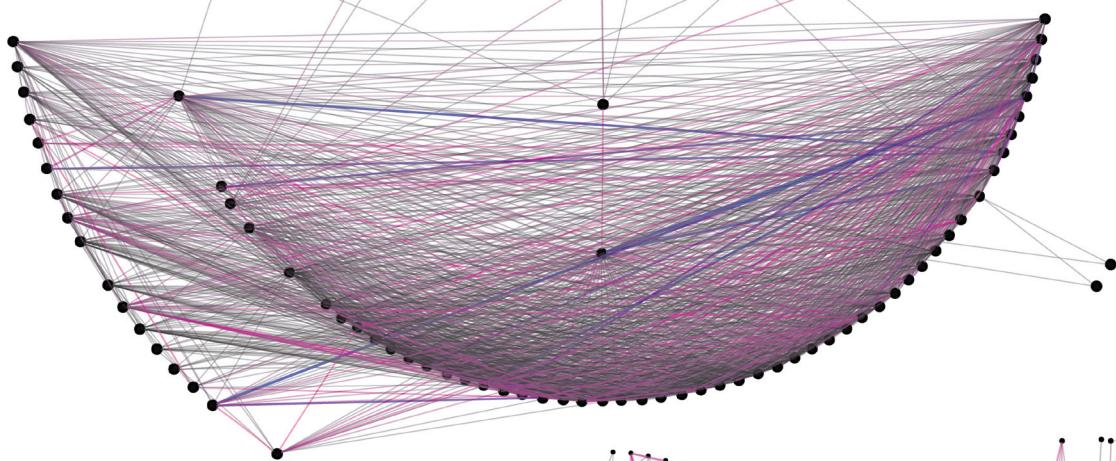
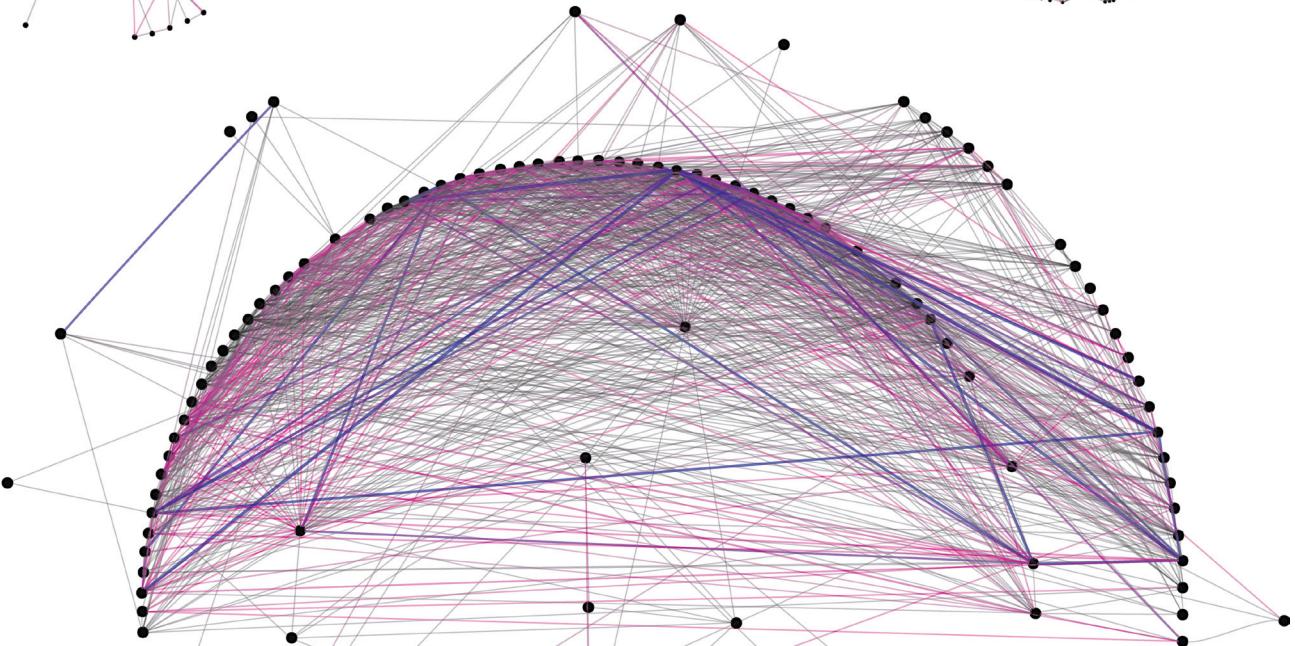
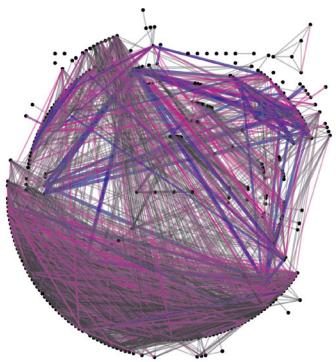
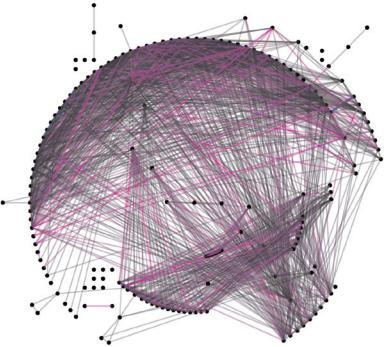
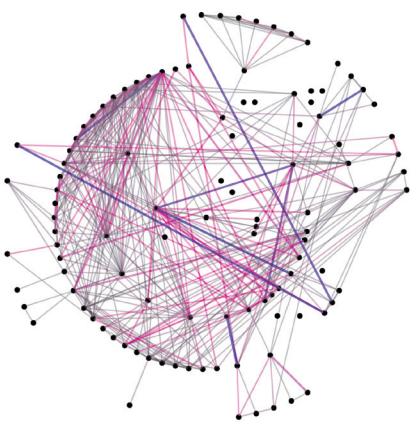
The past decade has witnessed a coming-together of the technological networks that connect computers on the Internet and the social networks that have linked humans for millennia. Beyond the artifacts that have sprung from this development—sites such as Facebook, LinkedIn, MySpace, Wikipedia, digg, del.icio.us, YouTube, and flickr—there is a broader process at work, a growing pattern of movement through online spaces to form connections with others, build virtual communities, and engage in self-expression.

Even as these new media have led to changes in our styles of communication, they have also remained governed by longstanding principles of human social

interaction—principles that can now be observed and quantified at unprecedented levels of scale and resolution through the data being generated by these online worlds. Like time-lapse video or photographs through a microscope, these images of social networks offer glimpses of everyday life from an unconventional vantage point—images depicting phenomena such as the flow of information through an organization or the disintegration of a social group into rival factions. Science advances whenever we can take something that was once invisible and make it visible; and this is now taking place with regard to social networks and social processes.

Collecting social-network data has traditionally been hard work, requiring extensive contact with the group of people being studied; and, given the practical considerations, research efforts have generally been limited to groups of tens to hundreds of individuals. Social interaction in online settings, on the other hand, leaves extensive digital traces by its very nature. At the scales of tens of millions of individuals and minute-by-minute time granularity, we can replay and watch the ways in which people seek out connections and form friendships on a site like Facebook or how they coordinate with each other and engage in creative expression on sites like Wikipedia and flickr. We can observe a news story suddenly catching the attention of millions of readers or witness how looming clouds of controversy gather around a community of bloggers. These are part of the ephemeral dynamics of ordinary life, now made visible through their online manifestations. As such, we are witnessing a revolution in the measurement of collective human

*The Nexus friend grapher application, created by Ivan Kozik, allows Facebook account holders to generate graphs illustrating their social network of friends. The resulting spheres not only demonstrate how friends are connected, but also indicate the interests shared by different groups of friends. For more information, or to create a graph, visit <http://nexus.ludios.net>.*



behavior and the beginnings of a new research area—one that analyzes and builds theories of large social systems by using their reflections in massive datasets.

This line of investigation represents a flow of ideas between computing and the social sciences that goes in both directions. Using datasets on collective human behavior, together with an algorithmic language for modeling social processes, we can begin to make progress on fundamental social-science questions, informed by a computational perspective. Meanwhile, social scientists' insights into these problems, which predate the Internet, are essential to understanding the current generation of computing systems. Indeed, most of the high-profile Internet applications to emerge over the past half-decade are governed not just by technological considerations but also by recurring and quantifiable principles of human social interaction; both technological and social forces, working together, shape the inherent operating constraints in such systems.

The resulting research questions arise from a coming-together of different styles of research, and it is important to recognize that analyses of truly massive social networks provide us with both more and less than we get from detailed studies at smaller scales. Massive datasets can allow us to see patterns that are genuine, yet literally invisible at smaller scales. But working at a large scale introduces its own difficulties. One doesn't necessarily know what any one particular individual or social connection signifies; and the friendships, opinions, and personal information that are revealed online come in varying degrees of reliability. One is observing social activity in aggregate, but at a fine-grained level the data is more difficult to interpret. The true challenge is to bridge this gap between the massive and the detailed, to find the points where these lines of research converge.

With that goal in mind, we discuss two settings where this research strategy is being pursued. We begin with the "small-world phenomenon" in social networks—the principle that we are all connected by short chains of acquaintances—and then look at the re-

lated problem of how ideas spread contagiously through groups of people.

### **The Small-World Phenomenon and Decentralized Search**

When the playwright John Guare coined the term "six degrees of separation,"<sup>15</sup> describing the notion that we are all just a few steps apart in the global social network, he was referring to a series of experiments performed by the social psychologist Stanley Milgram in the 1960s.<sup>38</sup> Milgram's work provided the first empirical evidence for this idea, and it is useful to consider the structure of his experiments and their significance.

Inspired in part by the work of the political scientist Ithiel de Sola Pool with the applied mathematician Manfred Kochen,<sup>7</sup> Milgram asked each of a few hundred people in Boston and the Midwest to try directing a letter toward a designated "target" in the network—a stockbroker who lived in Sharon, Massachusetts. The participants in the experiment were given basic personal information about the target, including his address and occupation; but each participant could only mail the letter to someone he or she knew on a first-name basis, with the instructions to forward it on in this way toward the target as quickly as possible. The mail thus closed in on the town of Sharon, moving from friend to friend, with the successful letters reaching the target in a median of six steps.<sup>38</sup> This kind of experiment, constructing paths through social networks to distant target individuals, has been repeated by a number of other groups in subsequent decades.<sup>9, 12, 23</sup>

Milgram's experiment and its follow-ups come with many caveats. In particular, they have tended to be most successful when the target is affluent and socially accessible; and even then, many chains fail to complete. Nevertheless, the striking fact at the heart of the results—that such short paths can be discovered in social networks—has been borne out by many subsequent analyses of large-scale network data. Quite recently, Leskovec and Horvitz built a social network from nearly a quarter-billion instant-messaging accounts on MSN Messenger, connecting two individuals if they engaged in a conversation over a one-month obser-

vation period.<sup>26</sup> The researchers found the average length of the shortest path between any two people on this system to be around 6.6—a number remarkably close to Milgram's, and obtained by utterly different means.

*Modeling the Phenomenon.* Mathematical models of this phenomenon start by asking *why* social networks should be so rich in short paths. In an influential 1998 paper, Watts and Strogatz sought to reconcile this abundance with the seemingly contrasting observation that the world is highly clustered, consisting of acquaintances who tend to be geographically and socially similar to one another.<sup>40</sup> They showed that adding even a small number of random social connections to a highly clustered network causes a rapid transition to a small world, with short paths appearing between most pairs of people. In other words, the world may look orderly and structured to each of us—with our friends and colleagues tending to know each other and have similar attributes—but a few unexpected links shortcircuiting through the network are sufficient to bring us all close together.

There is a further aspect to the Milgram experiment that is striking and inherently algorithmic: the experiment showed not just that the short paths were there but that people were able to find them.<sup>20</sup> When you ask someone in Omaha, Nebraska, as Milgram did, to use his or her social network to direct a letter halfway across the country to Sharon, Massachusetts, that person can't possibly know the precise course it will follow or whether it will even get there. The fact that so many of the letters zeroed in on the target suggests something powerful about the social network's ability to "funnel" information toward far-off destinations. The U.S. Postal Service does this when it delivers a letter, but it is centrally designed and maintained at considerable cost to do precisely this job; why should a social network, which has grown organically without any central control, be able to accomplish the same task with any reliability at all?

To begin modeling this phenomenon, suppose we all lived on a two-dimensional plane, spread out with a roughly uniform population density,

and that we each knew our next-door neighbors for some distance in each direction. Now, following Watts and Strogatz, we add a small number of random connections—say, each of us has a single additional friend chosen uniformly at random from the full population. Short paths appear, as expected, but one can prove that there is no procedure the people living in this world can perform—using only local information and without a global “bird’s-eye view” of the social network—to forward letters to faraway targets quickly.<sup>20</sup> In other words, in a structured world supplemented with purely random connections, the Milgram experiment would have failed: the short paths would have been there, but they would have been unfindable for people living in the network.

By extending things a little bit, however, we can get the model to capture the effect Milgram saw in real life. To do this, we keep everyone living on a two-dimensional plane but revisit the random connections, which are supposed to account for the unexpected far-flung friendships that make the world small. In reality, of course, these links are not completely random; they too are biased toward closer and more similar people. Suppose, then, that each person still has a random far-away friend, but that this friend is chosen with a probability that decays with the individual’s distance in the plane—say, by a “gravitational law” in which the probability of being friends with a person at a distance  $d$  decays as  $d^{-r}$  for some power  $r$ . Thus, as the exponent  $r$  increases, the world gets less purely random—the long-range friendships are still potentially far away, but overall they are more geographically clustered. What effect does this have on searching for targets in the network?

Analyzing this model, one finds that the effectiveness of Milgram-style search with local information initially gets better as  $r$  increases—because the world is becoming more orderly and easy to navigate—and then gets worse again as  $r$  continues increasing—because short paths actually start becoming too rare in the network. The best choice for the exponent  $r$ , when search is in fact very rapid, is to set it equal to 2. In other words, when the

## A rumor, a political message, or a link to an online video—these are all examples of information that can spread from person to person, contagiously, in the style of an epidemic.

probability of friendship falls off like the square of the distance, we have a small world in which the paths are not only there but also can be found quickly by people operating without a global view.<sup>20</sup> The exponent of 2 is thus balanced at a point where short paths are abundant, but not so abundant as to be too disorganized to use.

Further analysis indicates that this best exponent in fact has a simple qualitative property that helps us understand its special role: when friendships fall off according to an inverse-square law in two dimensions, then on average people have about the same proportion of friends at each “scale of resolution”—at distances 1–10, 10–100, 100–1000, and so on. This property lets messages descend gradually through these distance scales, finding ways to get significantly closer to the target at each step and in this way completing short chains, just as Milgram observed.

*Validating and Applying the Model.* When such models were first proposed, it was unclear not only how accurate they were in real life but also how to go about collecting data to measure the accuracy. To do so, you would have to convince thousands of people to report where they lived and who their friends were—a daunting task.

But of course, the public profiles on social-networking sites readily do just that, and as these sites began to grow explosively in 2003 and 2004, Liben-Nowell et al. developed a framework for using this type of data to test the predictions of the small-world models.<sup>30</sup> In particular, they collected data from the friendship network of the public blogging site LiveJournal, focusing on half a million people who reported U.S. hometown locations and lists of friends on the site. They then had to extend the mathematical models to deal with the fact that real human population densities are highly nonuniform. To do so, they defined the distance between two people in an ordinal rather than absolute sense: they based the probability that a person  $v$  forms a link to a person  $w$  on the number of people who are closer to  $v$  than  $w$  is, rather than on the physical distance between  $v$  and  $w$ . Using this more flexible definition, the distribution of friendships in the data could

then in fact be closely approximated by the natural generalization of the inverse-square law.

It was difficult not to be a bit surprised by the alignment of theory and measurement. The abstract models were making very specific predictions about how friendships should depend on physical distance, and these predictions were being approximately borne out on data arising from real-world social networks. And there remains a mystery at the heart of these findings. While the fact that the distributions are so close does not necessarily imply the existence of an organizing mechanism (for example, see Bookstein<sup>5</sup> for a discussion of this general issue in the context of social-science data), it is still natural to ask why real social networks have arranged themselves in a pattern of friendships across distance that is close to optimal for forwarding messages to faraway targets. Further, whatever the users of LiveJournal are doing, they are not explicitly trying to run versions of the Milgram experiment—if there are selective pressures driving the network toward this shape, they must be more implicit, and it remains a fascinating open question whether such forces exist and how they might operate.

Other research using online data has considered how friendship and communication depend on nongeographic notions of “distance.” For example, the probability that you know someone is affected by whether you and they have similar occupations, cultural backgrounds, or roles within a large organization. Adamic and Adar studied how communication depends on one such kind of distance: they measured how the rate of email messaging between employees of a corporate research lab fell off as they looked at people who were farther and farther apart in the organizational hierarchy.<sup>1</sup> Here too, this rate approximated an analogue of the inverse-square law—in a form adapted to hierarchies<sup>21, 39</sup>—although the messages in the researchers’ data were skewed a bit more toward long-range contacts in the organization than short-range ones.

Finally, these models can rapidly turn into design principles for distributed computing systems as well. Modern peer-to-peer file-sharing systems

are built on the principle that there should not be a central index of the content being shared (in contrast, for example, to the way in which search engines like Google provide a central index for Web pages). As a result, looking up content in a peer-to-peer system follows a Milgram-style approach in which the hosts participating in the system must forward requests with only a local view.<sup>31</sup> Mathematical models of small worlds—originally built with human networks in mind—can provide insights into the design of efficient solutions for this distributed search problem as well.

We’ve thus seen how viewing such models in the online domain can help us understand the global layout of social-networking sites, the flow of communications within organizations, and the design of peer-to-peer systems. We now look at how the insights we’ve gained here can provide perspective on an important related problem—the spread of information through large populations.

### **Social Contagion and the Spread of Ideas**

Milgram’s experiment was about focusing a message on a particular target, but much of the information that flows through a social network radiates outward in many directions at once. A rumor, a political message, or a link to an online video—these are all examples of information that can spread from person to person, contagiously, in the style of an epidemic. This is an important process to understand because it is part of a broader pattern by which people influence one another over longer periods of time, whether in online or offline settings, to form new political and social beliefs, adopt new technologies, and change personal behavior—a process that sociologists refer to as the “diffusion of innovations.”<sup>35</sup> But while the outcomes of many of these processes are easily visible, their inner workings have remained elusive.

Some of the basic mathematical models for the diffusion of innovations posit that people’s adoption of new behaviors depends in a probabilistic way on the behaviors of their neighbors in the social network: as more and more of your friends buy a new product or join a new activity, you

are more likely to do so as well.<sup>13</sup> Recent studies of online data have provided some of the first pictures of what this dependence looks like over large populations. In particular, Leskovec, Adamic, and Huberman studied how the probability of purchasing books, DVDs, and music from a large online retailer increased with the number of email recommendations a potential customer received.<sup>25</sup> Backstrom et al. determined the probability of joining groups in a large online community as a function of the number of friends who already belonged to the group.<sup>4</sup> And Hill, Provost, and Volinsky<sup>16</sup> analyzed how an individual’s adoption of a consumer telecommunications service plan depended on his or her connections to prior adopters of the service.

While the probability of adopting a behavior increases with the number of friends who have already adopted it, there is a “diminishing returns” pattern in which the marginal effect of each successive friend decreases.<sup>4,25</sup> In many cases, however, an interesting deviation from this pattern is observed—a “0–1–2 effect,” in which the probability of joining an activity when two friends have done so is significantly more than twice the probability of joining when only one has done so.<sup>4</sup>

*The structure of cascading behavior.* Beyond these local mechanisms of social influence, it is instructive to trace out the overall patterns by which influence propagates through a large social network. In recent work, David Liben-Nowell and I investigated such global-scale processes by gathering data on chain-letter petitions that had spread widely over the Internet.<sup>29</sup> A particularly pervasive chain letter, which spread in 2002 and 2003, purported to organize opposition to the impending invasion of Iraq. Each copy of the petition contained the list of people who had received that particular copy, in the order in which they added their names and then passed it on to others in their email address books. In the process, several hundred of these copies had been sent to Internet mailing lists; by retrieving them from the mailing lists’ archives, we could reconstruct a large fragment of the branching tree-like trajectory by which the chain letter had spread.

The structure of the tree was surprising, as it challenged our small-world intuitions. Rather than fanning out widely, reaching many people with only a few degrees of separation, the chain letter spread in a deep and narrow pattern, with many paths consisting of several hundred steps. The short chains in the social network were still there, but the chain letter was getting to people by much more roundabout means. Moreover, we found a very similar structure for the one other large-scale chain letter on which we could find enough mailing-list data, this one claiming to be organizing support for National Public Radio.

Why this deep and narrow spreading pattern arises in multiple settings remains something of a mystery, but there are several hypotheses for reconciling it with the structure of a small world. In our work on chain letters, we analyzed a model based on the natural idea that people take widely varying amounts of time to act on messages as they arrive: when recipients forward the chain letter at different times to highly overlapping circles of friends, it can in effect “echo” through dense clusters in the social network, following a snaking path rather than a direct one. Simulations of this process on real social networks such as the one from LiveJournal produce tree structures very similar to the true one we observed.<sup>29</sup>

It is also plausible that the nature of social influence—properties such as the 0–1–2 effect in particular—play an important role. Suppose that most people in the social network need to receive a copy of the letter at least twice before actually signing their name and sending it on. As Centola and Macy have recently argued, our long-range friendships may be much less useful for spreading information in situations such as these: you can learn of something the first time from a far-flung friend, but to get a second confirmatory hearing you may need to wait for the information also to arrive through your more local contacts.<sup>6</sup> Such a pattern could slow down the progress of a chain letter, forcing it to slog through the dense structure of our local connections rather than exploit the long-range shortcuts that make the world small.



**The availability of such rich and plentiful data on human interaction has closed an important feedback loop, allowing us to develop and evaluate models of social phenomena at large scales and to use these models in the design of new computing applications.**



*Contagion as a design principle.* As with the decentralized search problem at the heart of the small-world phenomenon, the idea of contagion in networks has served as a design principle for a range of information systems. Early work in distributed computing proposed the notion of “epidemic algorithms,” in which information updates would be spread between hosts according to a probabilistic contagion rule.<sup>8</sup> This has led to an active line of research, based on the fact that such algorithms can be highly robust and relatively simple to configure at each individual node.

More recently, contagion and cascading behavior have been employed in proposals for social computing applications such as word-of-mouth recommendation systems,<sup>25</sup> incentive mechanisms for routing queries to individuals possessing relevant information,<sup>22</sup> and methods to track the spread of information among Weblogs.<sup>2, 14</sup> Large-scale social contagion data also provides the opportunity to identify highly influential sets of people in a social network—the set of people who would trigger the largest cascade if they were to adopt an innovation.<sup>11</sup> The search for such influential sets is a computationally difficult problem, although recent work has shown that when social influence follows the kind of “diminishing returns” pattern discussed here, it is possible to find approximate methods with provable guarantees.<sup>19, 32</sup>

### Further Directions

Research on large-scale social-network data is proceeding in many further directions as well. While much of what we have been discussing involves the dynamic behavior of individuals in social networks, an important and complementary area of inquiry is how the structure of the network itself evolves over time.

Recent studies of large datasets have shed light on several important principles of network evolution. A central one, rooted in early work in the social sciences, is the principle of “preferential attachment”—the idea that nodes that already have many links will tend to acquire them at a greater rate.<sup>33</sup> An active line of research has shown how preferential attachment can lead to the highly

skewed distributions of links that one sees in real networks, with certain nodes acting as highly connected "hubs."<sup>33</sup>

Another principle, also a key issue in sociology, is the notion of "triadic closure:" links are much more likely to form between two people when they have a friend in common.<sup>34</sup> Recent work using email logs has provided some of the first concrete measurements of the effect of triadic closure in a social-communication network.<sup>24</sup>

Further principles have begun to emerge from recent studies of social and information networks over time, including "densification effects," in which the number of links per node increases as the network grows, and "shrinking diameters," in which the number of steps in the shortest paths between nodes can actually decrease even as the total number of nodes is increasing.<sup>27</sup>

It is also intriguing to ask whether machine-learning techniques can be effective at predicting the outcomes of social processes from observations of their early stages. Problems here include the prediction of new links, the participation of people in new activities, the effectiveness of groups at collective problem-solving, and the growth of communities over time.<sup>4, 16, 17, 18, 28, 37</sup> Recent work by Salganik, Dodds, and Watts raises the interesting possibility that the outcomes of certain types of social-feedback effects may in fact be *inherently unpredictable*.<sup>36</sup> Through an online experiment in which participants were assigned to multiple, independently evolving versions of a music-download site—essentially, a set of artificially constructed "parallel universes" in which copies of the site could develop independently—Salganik et al. found that when feedback was provided to users about the popularity of the items being downloaded, early fluctuations in the popularities of different items could get locked in to produce very different long-term trajectories of popularity. Developing an expressive computational model for this phenomenon is an interesting open question.

Ultimately, across all these domains, the availability of such rich and plentiful data on human interaction has closed an important feedback loop, allowing us to develop and evaluate models of social phenomena at

large scales and to use these models in the design of new computing applications. Such questions challenge us to bridge styles of scientific inquiry—ranging from subtle small-group studies to computation on massive datasets—that traditionally have had little contact with each other. And they are compelling questions in need of answers—because at their heart, they are about the human and technological connections that link us all, and the still-mysterious rhythms of the networks we inhabit.

### Acknowledgments

I thank the National Science Foundation, the MacArthur Foundation, the Cornell Institute for the Social Sciences, Google, and Yahoo for their support and the anonymous reviewers of this manuscript for their comments and feedback. **C**

### References

1. Adamic, L. and Adar, E. How to search a social network. *Social Networks* 27, 3 (2005), 187–203.
2. Adar, E., Zhang, L., Adamic, L.A., and Lukose, R.M. Implicit structure and the dynamics of blogspace. In *Proceedings of the Workshop on the Weblogging Ecosystem*, 2004.
3. Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (2002), 47–97.
4. Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.
5. Bookstein, A. Inferometric distributions, Part II: Resilience to ambiguity. *J. American Society for Information Science* 41, 5 (1990), 376–386.
6. Centola, D. and Macy, M. Complex contagions and the weakness of long ties. *American J. of Sociology* 113 (2007), 702–734.
7. de Sola Pool, I. and Kochen, M. Contacts and influence. *Social Networks* 1, 1 (1978), 5–51.
8. Demers, A.J., Greene, D.H., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H.E., Swinehart, D.C., and Terry, D.B. Epidemic algorithms for replicated database maintenance. In *Proceedings of the 6th ACM Symposium on Principles of Distributed Computing* (1987), 1–12.
9. Dodds, P., Muhamad, R., and Watts, D. An experimental study of search in global social networks. *Science* 301 (2003), 827–829.
10. Dodds, P. and Watts, D. Universal behavior in a generalized model of contagion. *Physical Review Letters* 92 (2004).
11. Domingos, P. and Richardson, M. Mining the network value of customers. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2001), 57–66.
12. Garfield, E. It's a small world after all. *Current Contents* 43 (1979), 5–10.
13. Granovetter, M. Threshold models of collective behavior. *American J. Sociology* 83 (1978), 1420–1443.
14. Gruhl, D., Liben-Nowell, D., Guha, R.V., and Tomkins, A. Information diffusion through blogspace. In *Proceedings of the 13th International World Wide Web Conference*, 2004.
15. Guare, J. *Six Degrees of Separation: A Play*. Vintage Books, 1990.
16. Hill, S., Provost, F., and Volinsky, C. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science* 21, 2 (2006), 265–278.
17. Hoff, P.D., Raftery, A.E., and Handcock, M.S. Latent space approaches to social network analysis. *J. American Statistical Association* 97 (2002).
18. Kearns, M., Suri, S., and Montfort, N. An experimental study of the coloring problem on human subject networks. *Science* 313 (2006), 824–827.
19. Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence in a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2003), 137–146.
20. Kleinberg, J. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing* (2000), 163–170.
21. Kleinberg, J. Small-world phenomena and the dynamics of information. In *Proceedings of the 14th Advances in Neural Information Processing Systems* (2001), 431–438.
22. Kleinberg, J. and Raghavan, P. Query incentive networks. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science* (2005), 132–141.
23. Korte, C. and Milgram, S. Acquaintance networks between racial groups: Application of the small world method. *J. Personality and Social Psychology* 15 (1978).
24. Kossinets, G., and Watts, D. Empirical analysis of an evolving social network. *Science* 311 (2006), 88–90.
25. Leskovec, J., Adamic, L., and Huberman, B. The dynamics of viral marketing. *ACM Transactions on the Web* 1, 1 (May 2007).
26. Leskovec, J. and Horvitz, E. Worldwide buzz: Planetary-scale views on an instant-messaging network. In *Proceedings of the 17th International World Wide Web Conference*, 2008.
27. Leskovec, J., Kleinberg, J.M., and Faloutsos, C. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2005), 177–187.
28. Liben-Nowell, D. and Kleinberg, J. The link-prediction problem for social networks. *J. American Society for Information Science and Technology* 58, 7 (2007), 1019–1031.
29. Liben-Nowell, D. and Kleinberg, J. Tracing information flow on a global scale using Internet chain-letter data. In *Proceedings of Natl. Acad. Sci.* 105, 12 (Mar. 2008), 4633–4638.
30. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., and Tomkins, A. Geographic routing in social networks. In *Proceedings of Natl. Acad. Sci.* 102, 33 (Aug. 2005), 11623–11628.
31. Liu, E.K., Crowcroft, J., Plas, M., Sharma, R., and Lim, S. A survey and comparison of peer-to-peer overlay network schemes. *IEEE Commun. Surveys and Tutorials* 7, 2 (2005), 72–93.
32. Mossel, E. and Roch, S. On the submodularity of influence in social networks. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, 2007.
33. Newman, M.E.J. The structure and function of complex networks. *SIAM Review*, 45 (2003), 167–256.
34. Rapoport, A. Spread of information through a population with socio-structural bias I: Assumption of transitivity. *Bulletin of Mathematical Biophysics* 15, 4 (Dec. 1953), 523–533.
35. Rogers, E. *Diffusion of Innovations*, 4th Edition. Free Press, 1995.
36. Salganik, M., Dodds, P., and Watts, D. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311 (2006), 854–856.
37. Sarkar, P. and Moore, A.W. Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, 2005.
38. Travers, J. and Milgram, S. An experimental study of the small world problem. *Sociometry* 32, 4 (1969), 425–443.
39. Watts, D.J., Dodds, P.S., and Newman, M.E.J. Identity and search in social networks. *Science* 296 (May 2002), 1302–1305.
40. Watts, D.J. and Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* 393 (1998), 440–442.

**Jon Kleinberg** (kleinber@cs.cornell.edu) is a professor of computer science at Cornell University, Ithaca, NY. His work focuses on issues at the interface of networks and information, with an emphasis on the social and information networks that underpin the Web and other online media. He is a recipient of MacArthur, Packard, and Sloan Foundation Fellowships, and the Nevanlinna Prize from the International Mathematical Union.