Social Media and the Science of Health Behavior

Damon Centola, PhD

Social influences are a primary factor in the adoption of health behaviors.[1,2] Compliance with diet and nutrition programs, adherence to preventive screening recommendations, and maintenance of exercise routines all can depend on having contact with friends and family who also engage in these behaviors. In addition to a great deal of literature on peer effects,[3] recent studies of large network data sets have made important advances in our understanding of how social networks influence the collective dynamics of health behavior.[4,5] Research has shown that social influences can affect collective health outcomes ranging from epidemic obesity to smoking behaviors, which have important consequences both for theoretical models of social epidemiology and for the practical design of interventions and treatment strategies.[6,7] These findings have direct implications for research aimed at understanding how social influences on dieting, exercising, medication use, and getting screenings can impact behavior change affecting cardiovascular disease. The large number of health domains affected by recent research on the spread of behaviors has made social diffusion a topic of growing interest for an increasing variety of researchers and practitioners who are concerned with understanding the social dimensions of health. This article discusses the development of new methods that use social media to study these health dynamics.

Although there is widespread theoretical and practical interest in understanding how social influences affect health-related behaviors, empirical studies of the social dynamics of health face important methodological challenges. Large observational studies of population health have faced the limitation that they are unable to address problems of causal identification.[8,9] Extant studies have been able to show conclusively that health-related traits such as smoking[5] and weight gain[4] correlate with social ties in a network, yet the data do not provide a clear assessment of the degree to which social network ties directly influence behaviors versus reveal shared exposure to common influences.[9] Correlations of traits with social network ties can occur because people who are friends are exposed to the same media signals (ie, exogenous information), because connected individuals live in the same neighborhoods (ie, geographic constraints such as living near the same restaurants and gyms), because people who already have similar traits form social ties with one another (ie, choice homophily), or because connected individuals influence one another to adopt similar behaviors (ie, social influence).[10,11] Although new techniques are being developed to discriminate between each of these causal mechanisms,[4,12] determining the relative impact of these factors is very difficult and is made even more complicated by the lack of reliable data both on the timing of behavior change and on the actual social network structure of a given population.

These difficulties motivate the need for new methods that can allow health researchers to identify the role of social networks in the real-time dynamics of behavior change. The goal of this article is to demonstrate that the rapid growth of peer-to-peer social media presents an important new resource for addressing these empirical challenges. Increasing levels of public participation in a diverse range of health-related social media create a new population of subjects whose natural, everyday engagement with health behaviors can be monitored and scientifically explored with a rapidly expanding repertoire of social technologies. Building on these new capacities, recent research has begun to study how social media can be used to experimentally evaluate the effects of social influence on behavior change. This approach to using social media entails a shift in focus from the interpersonal dimensions of social interactions to the community-wide effects of social network structure on the spread of behaviors through online populations.

Social Media and Health

Social media has become an indelible part of the public health landscape.[13–15] From Web-based appointment scheduling to online coaching for smoking cessation and weight loss, the Internet provides an increasingly valuable resource for customers of health services.[16] Although the majority of these efforts concentrate on organizational tools for providing clients with improved services, an equally important use of social media has come from the emergence of peer-driven health communities.

Peer-to-peer interaction in the health sector has a long history, starting with the creation of support groups for alcohol and tobacco abstention, weight control, long-term treatment, and grief and trauma counseling.[17] Much of the value of these peer-counseling organizations derives from personal and empathetic interaction.[18] The logic of this kind of interaction has been extended into the virtual domain with the development of online tools for coaching and abstention.[19] For instance, recent Web-based social support services such as QuitNet and Free and Clear provide peer-to-peer e-mail and instant messaging systems that offer newly abstaining smokers support and counseling from members with years of abstention experience.

At first glance, it is remarkable that anonymous online communities can be effective environments for providing

From the Massachusetts Institute of Technology, Cambridge, MA.
Correspondence to Damon Centola, PhD, Massachusetts Institute of Technology, E62-462, 100 Main St, Cambridge, MA 02142. E-mail dcentola@mit.edu
(*Circulation*. 2013;127:2135-2144)

Circulation is available at http://circ.ahajournals.org
DOI: 10.1161/CIRCULATIONAHA.112.101816

productive interactions that improve participants' health behaviors. However, the idea of social support from online interactions has been around since the inception of the Internet.[19] Since the 1990s, Usenet groups and Listservs have provided tools for support groups and medical information sharing through patient networks. For instance, a long-standing Listserv for cancer patients and their families, ACOR, provides an open network for patients to share treatment experiences and to engage with an empathetic community.[20]

The increasing popularity of social media sites like Facebook and Twitter has also given rise to commercial applications that offer radical new approaches to using social media for improved health. For instance, companies such Redbrick Health, StayWell, and Healthways have begun to use online social support platforms to help promote compliance with planned health regimens. Through widespread recruitment and regular interactions, these sites create communities that encourage increased participation in exercise and diet programs among their members. In a similar spirit, a recent Internet startup called PatientsLikeMe offers an extensive social media platform with online health profiles, patient information and disease histories, and interactive tools that allow members to share comprehensive reports with one another. Members of the site can participate in multiple disease-specific communities, allowing them to find information relevant to their individual medical needs. Not unlike the Listservs and patient support chat groups from the previous generation, patients can share information about their treatments and experiences, but with the important difference that the new, more sophisticated social media technologies allow participants to interact by comparing detailed records of ongoing health status, treatment programs, and recovery plans.

In addition to the social support that participants receive, a significant motivation for participating in these environments is the new informational channels that they create across traditional health communities.[21] A great deal of literature in social epidemiology studies the ways in which spatial, geographic, social, and economic constraints can significantly affect patients' health outcomes.[22] This research has found that a major variable affecting population health can be knowledge about and access to medical treatments and technologies.[23,24] For instance, studies have shown that physician practices can be highly localized, varying dramatically from one geographic region to another.[22,25] Consequently, information about treatments, medicines, and screenings may be disproportionately available to some patient populations and not to others.[26] The remarkable growth of Internet-based health and wellness communities allows patients from a variety of social and geographic backgrounds to share information about novel health resources, ranging from information about diet and nutrition to opportunities to learn about patient advocacy, preventive health screenings, and new treatment technologies.[27,28]

Health Research

This growing variety of social technologies provides an array of new opportunities for health researchers. As shown in Table 1, extant technologies fall into 2 broad categories of social interactions and health: open forms of social media such as Facebook and Twitter and intentionally designed online health communities.

Open technologies are large-scale virtual communication infrastructures that are designed for social interactions across many substantive domains. They are not specifically designed for health-related interactions, nor do they explicitly target any particular health community. Despite this, technologies such as Facebook, Twitter, Google+, and a variety of other social tools have created novel opportunities to trace the interactions between social connectivity and health. Recent work using open social technologies has provided important new insights into the dynamics of opinion propagation on health behaviors. For instance, studies on Twitter networks have found that sentiment about vaccines can be propagated through chains of Twitter feeds.[15] Similarly, attitudes toward smoking, weight loss, and cholesterol and blood pressure medications can also have a viral quality. The reinforcement and diffusion of attitudes has important consequences for the kinds of preventative behaviors that people are willing to engage in, particularly when the behaviors are difficult (like smoking cessation and exercise), or costly (like new medications and better nutrition).[29,30] Open social resources provide access to attitudinal records of hundreds of millions of people, which offers an unprecedented opportunity for large-scale inferences about the sentiments of the population at large, and the kinds of messaging strategies that may be most effective for reaching them. However, these large-scale data do not provide context-specific interactional observations or the capacity for clear causal identification of interaction patterns on population health. For this, we need to turn to intentionally designed online health networks.

Intentionally designed health communities are composed of members with an explicit interest in health and health behaviors. They can serve a variety of ends, from promoting regular exercise routines to providing counseling for smoking cessation. For researchers interested in the social dynamics of health, these sites offer the novel opportunity to collect data on participants' recorded health behaviors (such as exercise minutes per week, or daily dietary intake), while also tracking those behaviors among the members' social networks. On some sites, these data may suffer from self-reporting bias, for instance the accuracy of daily exercise reports or diet entries may depend on a participant's memory or be colored by a desire to appear fit. However, many of these sites provide tools that solve these problems by allowing participants to upload digitally recorded exercise data, or real-time medical

Table 1. Examples of Open and Intentionally Designed Online Social Networks

Types of Online Networks	Examples
Open social networks	Facebook, Twitter, Google+
Intentionally designed social networks	The Healthy Lifestyle Network, QuitNet, PatientsLikeMe

Open networks such as Twitter and Facebook provide social interactions on any topic. Intentionally designed networks such as The Healthy Lifestyle Network and QuitNet provide participants with targeted interactions around health-related goals.

records, which eliminates self-report bias and provides a means for timely social interactions on relevant health behaviors. Further, subjects' levels of participation in these sites – eg, the detail of their entries, their engagement with other members, and the overall frequency of log-ins – provide a direct behavioral measure of participants' involvement with the health community. The data extracted from these sites can be useful for establishing correlations between features of participants' social networks and subjects' commitments to smoking cessation, exercise routines, or good nutrition practices. Since the social networks in these sites are completely known to researchers, the impact of social factors such as homophily (ie, the tendency for social contacts to be similar to one another), clustering (ie, the tendency for people's contacts to be connected to one another), and degree (ie, the number of contacts that each subject regularly engages with) can be measured and evaluated with precision.

As the social value of these online technologies increases with scale, an important benefit for the medical community is the opportunity that they present for improving on traditional methods of health research. As shown in Table 2, traditional methods for studying the social dynamics of health have faced difficulties getting regular, reliable measurements of when and how behavior change takes place. These difficulties are compounded by the inability of traditional observational methods to get reliable network measurements, and to identify the causal impact of social influences on changes in behavior. Open social media technologies have begun to address these problems by introducing the possibility of collecting regular, reliable data on health activities in a way that seamlessly integrates with existing online behaviors. Combined with the ability to record accurate social network data, and trace behavior change over time, these open technologies create a capacity to improve not just the quantity, but also the quality of data available for health research. Yet, these technologies are primarily observational. Going one step further, intentionally designed communities create the opportunity to address the difficulties that have plagued traditional efforts to identify the causal impact of social factors on health behaviors.

The gold standard for medical research is the randomized, controlled trial. For scientific evaluation of medical treatments, this method provides an invaluable means of determining the relative efficacy of new pharmaceuticals. Small group social psychology has also made use of this method for testing the effects of interpersonal interactions on behavior change, identifying how key factors, such as status and gender, can affect social influence.[31] However, large-scale social dynamics and network effects have been impossible to study in controlled settings due to significant practical barriers, such as limitations on behavioral realism and population size. The creation of intentionally designed health communities using peer-to-peer social media has begun to eliminate these barriers, allowing the development of new experimental methods for designing controlled studies of the real-time collective dynamics of health behaviors.

To provide an overview of the methodological advances created by open and intentionally designed social media, I discuss these technologies in light of the limitations faced by previous methods of social and behavioral research on health (shown in Table 2). For each of the difficulties faced by traditional methods – *scale, measurement, behavioral fidelity, structural control, and reproducibility* – I identify how emerging social technologies offer new solutions.

Scale

Scale is an essential feature of studying the dynamics of behavior in social networks. This is because the effects of social interactions on collective outcomes are qualitatively different in small groups than they are in large networks. For instance, a population of twenty people with diverse opinions may be unable to reach a consensus on the risks of smoking; however mathematical models show that increasing the size of the population by two orders of magnitude can allow consensus on smoking risks to emerge.[32] This is because while very different people will not influence each other, people with similar beliefs will. The more people there are in the population, the more likely it is that someone will find another person with a similar belief, which allows a regression to the mean of their beliefs. The more people who can interact, the easier coordination will be. These local interactions can generate a process of global consensus across the population.[33,34] Changes in norms about obesity, beliefs about the need for diabetes mellitus screenings, and attitudes toward taking prescription medications to lower cholesterol all take place within large-scale social networks. Empirically studying these dynamics of belief formation, social influence, and behavior change in large social networks is impossible in a laboratory setting since only a small number of subjects can interact at a time. By contrast, thousands, even tens of thousands of people can interact on social media websites, where participants learn about new ideas, information, and behaviors from one another. These online environments can thus allow researchers to observe the large-scale dynamics of social influence in real, connected populations.

Table 2. Comparison of Methods for Studying Social Influences on Health Behaviors

	Traditional Observational Data	Laboratory Experiment	Digital Observational Data	Internet Experiment
Scale	✓	X	✓	✓
Measurement	X	✓	✓	✓
Structural control	X	✓	X	✓
Replication	X	✓	X	✓
Behavioral fidelity	✓	X	✓	✓

Traditional observational data, laboratory experiments, and digital observational data each have complementary advantages and disadvantages. Internet experiments combine the advantages of each approach.

Measurement

Another essential factor for studying the collective dynamics of health behavior is the ability to measure behavior change. What behaviors do people adopt, in what order, and because of which social influences? Detailed measurements of population connectivity, the sequences of adoption, and the real-time social influences on both adopters and nonadopters alike are necessary in order to understand the effects of social factors on behavior change. For instance, one of the most well-known empirical challenges in measuring behavior change is identifying individual thresholds for adopting new behaviors.[35–38] In mathematical models of behavior change, thresholds are typically represented as the number of adopters an individual needs to be exposed to before she will be convinced to also adopt. However, in order to measure individual thresholds researchers must count the number of social exposures that each person has. This requires knowing both the number of connections that each subject has, and who among those connections adopted before the subject. Getting accurate data for these simple facts has turned out to be a very hard problem to solve. By contrast, in social media networks, every connection and every action is time-stamped and recorded. So, the state of each individual, the connectedness of the population, and the path of behavior change through the population are all known quantities.

Behavioral Fidelity

Social media can allow us to study social interactions at scale and with precise measurement. Yet, how do we know that the networks measured in the online space are accurate representations of the actual influences on the behaviors of interest? A third feature of social media is the remarkable, global importation of daily social experiences into the online domain. While traditional laboratory studies of social behavior have faced the difficulty that measured behaviors and social interactions are explicitly artificial, in online settings detailed measurements of interactions over social media record people's natural behaviors of people in online settings. Open social technologies such as Facebook, Twitter, Google+ and other popular online venues have provided a minimal infrastructure for everyday social life, which allows individuals to interact seamlessly offline and online, blurring the distinction between the two worlds. This flexibility allows online environments to take on a remarkable familiarity, which offers researchers a high-fidelity record of subjects' everyday online interactions.

In addition to recording real social interactions, online social media increasingly provides an accurate record of health behaviors. As traditional health-related activities such as shopping for health and beauty products, signing up for gym memberships, contacting doctors, and making appointments for routine checkups become increasingly routine online activities, these digital traces of everyday life provide direct observations of the health-related behaviors that people engage in. Moreover, as these activities give rise to new intentionally designed online health communities, larger numbers of people maintain their own personal records of exercise and diet behavior in online form, providing as accurate a record of daily health behavior as there has ever been. Combined with the increasing trend toward online integration of physician records of medication maintenance, along with regular weight and cardiovascular reports, connections between individualized online health records and health-tailored social media may well become a new gold standard for accurate measures of real-time changes in population health.

Structural Control

Observational studies of any ilk, even well-designed natural experiments and field experiments, typically face the difficulty that population structure is not controlled. Consequently, specific features of social connectivity or confounding interactions between variables such as social familiarity, the structure of the social network, and the similarity of social contacts cannot be explicitly controlled or varied independently. This makes it essentially impossible to causally identify which social factors directly affect collective changes in health behaviors. However, as new, intentionally designed health communities and social media applications are developed, it is increasingly common for certain health-relevant relationships to take place entirely within these online environments. This means that the full sequence of messages, notifications, and contacts related to a specific behavior can be recorded and studied. For behavioral and medical scientists, this invites a new kind of research opportunity. Individual interactions not only can be precisely monitored but also can be seamlessly and strategically structured to study how changes in the pattern of social interactions influence changes in behavior. This also suggests that new kinds of behavioral interventions may be able to be deployed through social media tools, which can turn the online record of interactions and behaviors into a scientifically comprehensive record of how social structures affect the diffusion of health outcomes. These innovations suggest the possibility of laboratory-quality measurements that can, for instance, offer direct causal evidence on how alterations to the i) kinds of social contacts that subjects are exposed to ii) how many people they interact with, or iii) the connectedness of their network clusters can systematically affect collective changes in their health behaviors.

Reproducibility

The logic of experimental replication over independent observations provides the foundation for causal inference. But, how can we reproduce a behavioral epidemic? Or how can we independently replicate the diffusion of a health technology? Despite its intuitive appeal, the logic of replication has traditionally eluded empirical studies of social dynamics. The reason is that most large-scale observational studies cannot be reproduced under identical structural circumstances, with the same measurement capabilities, and with equivalent distributions of subject populations. However, the ability to design and control studies of behavioral dynamics with intentionally designed social media provides a way of overcoming these obstacles. With a controlled experimental design, subjects participating in intentionally designed sites can be randomized to independent trials in which entire populations can be independently "treated" with socially targeted interventions. These population-sized trials can thus be replicated by randomizing pools of subjects to completely independent social worlds with demonstrably divergent health outcomes. Extending the randomized, controlled designs from medicine and psychology, health scientists can gather repeated observations not just of individuals but of entire populations.

The first-order question that has traditionally occupied researchers of health behavior is whether it is possible to identify social influences on behavior change.[4,8,9] However, the union of these 5 features of social media—scale, measurement, behavioral fidelity, control, and reproducibility— creates the novel capacity not only to identify, but to investigate, the dynamics of behavioral influence on health in large populations. Once methodological barriers to identification are overcome, the number and variety of important new questions that can be studied increase dramatically. For instance, theoretical research in social epidemiology suggests not only that social influence takes place but also that changes to the pattern of network ties (or network topology) in a population may dramatically alter whether and how influence occurs. These theories of social influence suggest compelling new ideas about how network theory might be used to promote changes in population health. Could introducing changes into people's social networks really have large-scale impacts on their weight, cholesterol levels, and smoking behaviors? It is an intriguing proposition; however, when it comes to empirically evaluating these theories, traditional observational techniques do not offer many solutions. Here, new social media-based methods for behavioral research may allow researchers to empirically investigate for the first time how theoretically proposed changes to people's social networks might promote significantly different outcomes for population health.

A Case Study: Social Networks and Diffusion

Social epidemiologists have long been familiar with a truism called "the strength of weak ties."[36] The basic idea is that if you want to find information about new forms of medical treatment, innovations in health technology, or recent trends in diets that lower cholesterol and blood pressure, the best people to talk to are not the people whom you know well but rather your casual acquaintances. People you know well (ie, your close friends and family) are referred to as your affectively "strong ties" because of the strong emotional bond in the relationship. Correspondingly, casual acquaintances are called "weak ties" because there is very little affect in these relationships.

By definition, you do not know your weak ties well. So, you are not likely to know who their close friends or family are. You probably would not ask them to watch your children or offer to lend them a large amount of money. However, they provide an important social service. They connect you to parts of the social network that are far away. In other words, weak ties are also long-distance ties in the social sense that they connect people in the social network who would otherwise be socially remote. In contrast, close ties tend to know each other's friends, and their friends know each other. This creates triangles in the social network (see Figure 1). Each triangle is connected to other triangles, which are embedded in still more strong-tie triangles. An important consequence of this is that if someone tells a few of her close friends that she is looking to find out about alternative treatments for high blood pressure, and they pass the word on to their friends, chances are that many of them will wind up repeating this information to each other.

From the point of view of social diffusion,[a] eg, getting the word out about a search for medical treatment options, these redundant messages shared among close friends are wasted signals because the search will get circulated among many of the same people without reaching new sources of information. In a diffusion process, each time a piece of information gets repeated to someone who has already heard it, the social link fails to create a new exposure, and the search travels comparatively more slowly than it would have if the link had gone to an unknown person. Consequently, an informational search winds up bouncing around networks of people who have many of the same ideas about how to find health resources, similar exposure to medical treatments, and similar beliefs about the available treatments.

In contrast, if none of your social contacts have any contact with one another, that is, if every tie is a weak tie, then each person who is told about a search for high blood pressure treatments will repeat the message to entirely new people. And if all of those people's contacts are also nonredundant ties, then no links are wasted, and after a few steps of people repeating the message to their contacts, the social contagion will spread to exponentially more people than if it had been spreading through clustered triangles of close friends.

The inefficiency of clustered social networks and, in contrast, the remarkable effectiveness of nonredundant ties for accelerating the dynamics of social diffusion are the key insights behind the strength of weak ties. Building on this principle, the small world model of social networks decisively demonstrated the power of low-redundancy networks for improving the spread of social contagions (Figure 2).[39] This model showed that by randomly rewiring some of the ties in an otherwise clustered social network, the rate of diffusion of a social contagion could be radically improved.[40] From the perspective of public health, this finding has mixed implications. For the spread of infectious diseases, the small world model showed that casual contact networks with low redundancy could be a vehicle for rapid diffusion, accelerating the spread of infectious diseases like H1N1 across large

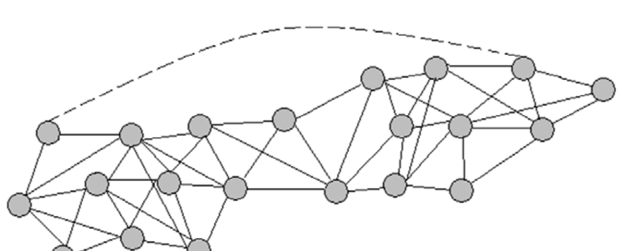

Figure 1. Clustered network with a single weak tie. Each individual's neighbors in the clustered network share neighbors with each other, creating triangles in the network. The addition of a weak tie into the network (dashed line) creates a link that connects otherwise distant individuals, allowing new ideas and information to spread more quickly.

[a]In the present discussion, "search" and "diffusion" are conflated for ease of exposition. In some contexts, diffusion and search processes behave qualitatively differently from one another.

populations. However, the small world model also suggested that for information and behavioral diffusion these networks would be just as helpful. Although casual contact networks might accelerate the growth of an epidemic, they might also be exploited to spread information about preventive behavior that would be effective at stemming the epidemic. This offers the hope that low-redundancy, rapid-diffusion networks could be used to spread desirable behavior change just as rapidly as diseases.

However, the implications of network structure for diffusion turned out to be more complex. Although research on small world networks and weak ties offers the promise of using fast-diffusion, low-redundancy networks for disseminating desirable health behaviors, the "complex contagions" model of diffusion showed that the social network structures that accelerate the diffusion of information and disease can, in fact, slow down and inhibit the spread of behavioral contagions.[38] The reason is that information and disease are typically simple contagions, which means that they can be transmitted by a single contact. This allows a long tie across a social network to function as a highly effective conduit for diffusion. More long ties create more bridges and faster spreading dynamics. However, for complex contagions such as behavior change, people often require multiple sources of social reinforcement before they are willing to change their diet, increase their exercise levels, or seek out treatment for a chronic heart condition. The complex contagions model showed that the need for social reinforcement decisively alters the dynamics of diffusion. The redundant signals that are wasted for simple contagions, slowing down the processes of information and disease diffusion, can actually be necessary for complex contagions. Using a series of mathematical models, these studies demonstrated that as redundancy is reduced (eg, by rewiring the ties in a clustered social network), the channels of social reinforcement that are necessary for behavioral diffusion are eliminated.[38] Without sufficient social reinforcement, the spread of behaviors slows down and ultimately stops before reaching most of the network.

The complex contagions model of diffusion suggests a reversal of the traditional wisdom in networks research. Instead of the strength of weak ties, it showed the weakness of long ties. Behaviors spread faster and produce greater overall adoption when traveling through clustered social networks with redundant ties than when traveling through the putatively more efficient small world networks. The implications of these results are far-reaching. The spread of new products, social activism, and healthcare innovations is often a complex contagion.[38] The more difficult, costly, or unfamiliar the behavior is, the more important social reinforcement becomes for promoting adoption. Thus, the more complex the social contagion is, the more that successful diffusion depends on clustered triangles in the social network.[29]

These findings have significant implications for how social interactions affect cardiovascular health. Word-of-mouth campaigns that spread information about the different types of cholesterol, the risks of high blood pressure, or the advantages of routine doctor's visits for patients with a history of acute myocardial infarction may diffuse widely through casual contact social networks. However, the complex contagions model suggests that campaigns designed to promote the adoption of unfamiliar treatments (eg, new kinds of cholesterol or blood pressure medication) or technologies (eg, new tests to keep track of blood pressure or arrhythmia) may be unlikely to spread through these social networks. Instead, campaigns designed to promote these behaviors may have greater success targeting clustered, strong tie networks than trying to exploit widespread acquaintance networks.

Furthermore, regular screenings, the adoption of good nutritional habits, and the maintenance of regular exercise routines can be time-consuming and sometimes painful. Being convinced to change these health behaviors can often depend on receiving reinforcement from social contacts who provide social proof for the value and effectiveness of these behaviors. Although a great deal of literature has been dedicated to showing the benefits of social support for promoting healthy behaviors,[1,2,41] social support can also improve the widespread diffusion of desirable behavior change throughout a population. Most important, these models show that changes in the structure of a social network can significantly impact the effectiveness of social ties for improving the success of efforts by physicians and public health campaigns dedicated to changing people's health behaviors.

The essential question that emerges from these theoretical results is whether their implications can be empirically tested. Although traditional methods of studying health behavior have been unable to evaluate network theories of social diffusion, key features of social media—scale, measurement, fidelity, control, and reproducibility—introduce the possibility of

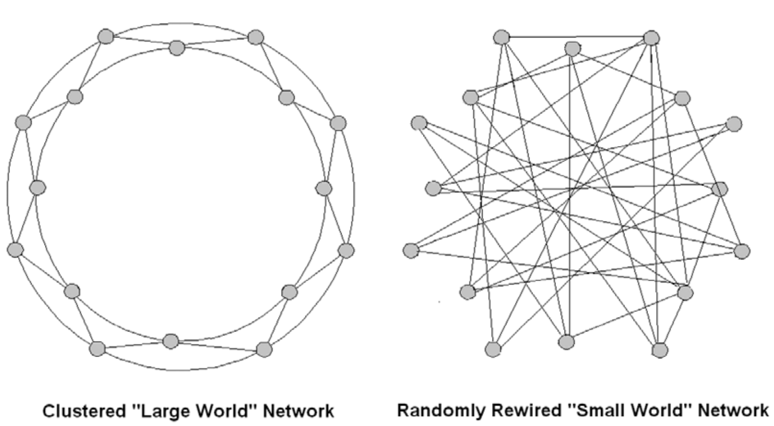

Figure 2. Clustered large world and randomized, small world networks. In a clustered lattice, the number of steps, or links, to travel from any node to any other randomly chose node is on the order of the size of the network (N). Randomly rewiring links dramatically reduces the travel time between any 2 nodes in the network.

experimentally evaluating these models of diffusion. To do this, an experimental design would have to control the confounding factors that are often correlated with network structure, such as affect, homophily, social capital, and access to resources. Furthermore, the design would need to identify the participants' social networks and study the spread of a measurable behavior, eliminating all exogenous factors that might introduce unobserved heterogeneity. And finally, the design would need to be replicated so that the effects of structure on diffusion could be evaluated across independent trials.

Studying Health Behavior With Social Media

Harnessing the new opportunity offered by social media, recent research has begun to pioneer the use of online technologies to investigate the effects of social structure on the spread of health behaviors. Using a novel social media platform, researchers studied the spread of behavior through an experimentally designed online health community.[30] The study recruited 1500 participants to join a health Web site called The Healthy Lifestyle Network in which participants were embedded in anonymous, online communities. The flow of subjects from registration through participation is shown in Figure 3. Once participants joined the study, they were randomized into one of two network conditions; one community was designed with tightly clustered, strong tie networks, and one community was designed with randomly structured weak tie or small world networks (Figure 4). A subject's neighbors within her social network constituted his or her health buddies, ie, other anonymous members of the online community from whom participants could receive information about new health behaviors and to whom they could send notifications about behaviors that they adopted.

The study evaluated the spread of membership in a new online environment for rating and evaluating health resources. The diffusion dynamics were initiated in each network using a randomly chosen "seed" node. Signals were sent from this node to its neighbors, inviting the neighbors to join the Community Health Forum. If one of the seed's health buddies joined the forum, messages were then sent to his or her health buddies inviting them to adopt; and, if any of these neighbors adopted, messages were then sent to their neighbors, and so forth, producing an observable cascade of adoption behavior through the social network. Using this design, the online environment maintained a perfect record of how many signals the subjects received and whether these signals were effective for convincing them to join the Community Health Forum.

Across 6 independent trials of this experimental design, the results were surprisingly clear. In every trial, the behavior spread to a greater fraction of the population in the clustered social networks than in the random networks (on average, 54% of participants adopted in the clustered network, whereas only 38% of the population adopted in the randomized, small world networks). Moreover, in all trials, the behavior spread faster in the clustered networks than in the random networks. Finally, the results also showed that commitment to the behavior, ie, longer-term engagement with the forum, was greater among members who had received reinforcing signals from multiple neighbors.

The results from The Healthy Lifestyle Network study suggest not only that social networks can be structured to promote the spread of behavior but also that the methodology of using intentionally designed online communities may be a general means for investigating strategies for promoting healthy behaviors. In a second study, researchers used participants in an online fitness program to study the impact of rearranging social relationships—based on similarities in age, sex, and body mass index—on participants' adoption of an online dieting tool.[42] The results showed that these similarities, or homophily, in health characteristics significantly improved adoption both in the overall population and specifically among the obese members of the fitness program. These findings are highly suggestive and reveal an open territory of research questions about how specific groups can be targeted to improve behavior change. How do age and gender differences among people's social contacts affect their influence on behavior change? Who are the most influential members of the network? Are social influences on exercise and diet behavior the same as those on smoking cessation and medication compliance? How can we seed new behaviors into these populations most effectively? With the new capacity to address traditional problems of identification, the scope and scale of new questions that can be asked and behaviors that can be studied through the use of social media technologies is remarkably broad.

Limitations and Methodological Provisos

As with any methodology, social media also has its limitations. One of the primary considerations is the privacy of data collected from social media sources. The digital traces of social life present new kinds of political and institutional concerns about participant privacy.[43] Based on the National Research

Figure 3. Registration and participation in an intentionally designed online network. Potential participants initially arrive at a landing page, where they find out about the study. They are then provided with an informed consent page and the ability to opt in or to decline participation in the study. Participants are then provided with information about their buddies in the social network. Finally, participants have opportunities to share recommendations with other members of the online community.

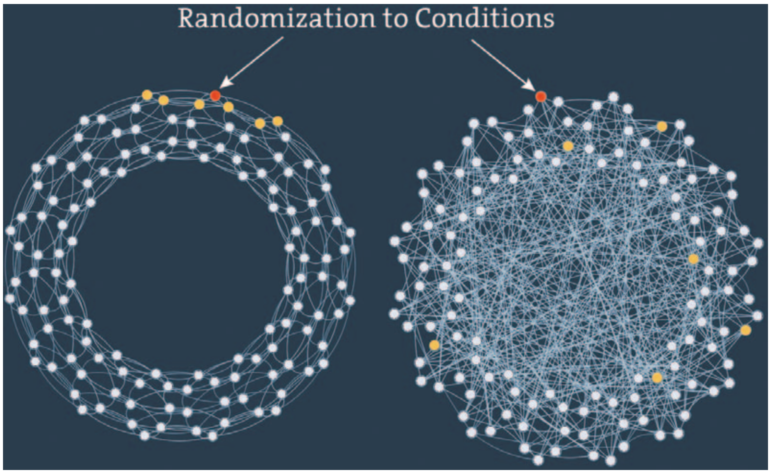

Figure 4. Randomization to conditions (n=128, *z*=6). Subjects were randomly assigned to clustered lattice and random network conditions in each trial of the study. In each condition, the red node shows the focal node of a neighborhood to which an individual has been assigned, and the orange nodes correspond to that individual's neighbors in the network. In the clustered lattice network, the orange nodes share neighbors with each other, whereas in the random network they do not.

Act of 1974, federal and institutional requirements on the use of human subjects are designed to ensure that individual privacy is maintained in the use of digital data for studying social behavior. Although early attempts to collect and use data from open sources of social media raised concerns,[44] new techniques for anonymizing data from online sources can be used, enhancing the strategies that have been used for offline sources of sensitive data. For instance, along with more sophisticated techniques, the elimination of data fields, randomization of identifiers, and resequencing methods can be used to code and protect data that contain sensitive information.

A related concern is that sources of data from open online technologies may also present challenges for acquiring informed consent, for instance, when tens of millions of subjects are in a Facebook or Twitter data set. For intentionally designed social media sites, the situation is simplified by the inclusion of explicitly created informed consent pages in the registration process. These sites allow participants to opt out of experimental studies, providing a transparent means for protecting subjects' rights to privacy and offering full disclosure of how their data will be used.

A second challenge that accompanies online data, particularly for intentionally designed communities, is the need for security. The recruitment of participants from the World Wide Web opens the door for unscrupulous individuals who may attempt to "hack" into these sites. Taking these concerns very seriously, researchers have acknowledged that high levels of security need to be used in the design of research environments, and organizations like Amazon.com have put significant effort into developing tools to ensure that these data sources remain secure. Additionally, as more data are collected and stored online, security protocols for protecting these data are continually under development, creating standardized tools for ensuring that access to all data is restricted and the data are encrypted.[45,46]

A third important concern about social media is the representativeness of the subject population. The significant digital divide between the technologically literate and those who do not have access to Internet tools presents an important concern for researchers who want to use social media to study underserved populations.[47] In the United States, Internet literacy is increasing as schools provide technological training as part of the basic educational curriculum. However, access is more restricted among older, lower-income populations in the United States. Furthermore, in underserved parts of the world, technological access can be a significant barrier to inclusion in online studies and thus may present an important limitation on the representativeness of subject populations studied online.

Consequently, new opportunities for research using social media are complementary, and do not eclipse the need for traditional approaches. On the one hand, considering the limitations on traditional studies of social influences on population health, the ability to recruit at large from the Internet, gaining access to >1 billion potential subjects with diverse health backgrounds, is an important advance. On the other hand, improvements still need to be made to increase our abilities both to study health behaviors across more diverse populations and to use these technologies to inform and improve these behaviors. Thus, it is essential that research using social media complement other valuable modes of research on the social determinants of health that do not rely on these technologies. Field studies, observational data from office visits, and large-scale data collection efforts (such as the Framingham Heart Study) are still necessary and valuable means of connecting patients' everyday social lives with their health outcomes. What social media methods offer is an opportunity to significantly expand our scientific capacity for identifying the impact of social influences on patients' regular exercise, diet, and behavioral routines.

Finally, perhaps one of the most often discussed limitations of social media for health research is the connection between online data and real health outcomes. The natural concern about studies of self-reported behavior such as the use of online diet tools or exercise programs is that these data are not an accurate representation of people's real activities. This is a general concern for any self-reported health data, and online data are no exception. However, the increasing trend toward digital integration of offline activities provides some promising new directions that mitigate this concern and make social media an increasingly attractive resource for collecting accurate health data. For instance, the use of health sensors in handheld communication devices provides new tools for real-time data collection of health behaviors using digital technologies.[48] Furthermore, as medical records and regular physician reports become available online, social media tools provide a valuable opportunity to connect people's online identities with diverse sources of information about their offline health behaviors and outcomes, creating a more complete and timely picture of how their health evolves over time and in a social context.

Conclusions

Growing interest in social media from consumers and producers alike has accelerated the development of new forms of social connectivity, creating an increasingly valuable opportunity for medical researchers interested in the social dimensions of health. As more behaviors are recorded and inevitably connected with social technologies, the variety of designs available for exploring the effects of social interactions on health outcomes continues to expand, creating a field of new possibilities for scientists interested in using the power and ubiquity of social media to study the dynamics of health.

This article has largely focused on the possibilities that these emerging technologies present for conducting randomized, controlled trials using social media. However, it is important to emphasize that the growing constellation of technologies available for studying social networks and health offers important opportunities for multimethod approaches to medical research, which can, in turn, inform how these technologies may also be used to promote public health. For instance, large-scale data from Twitter feeds or Facebook postings on attitudes toward smoking cessation, opinions about the post--acute myocardial infarction use of aspirin, or evaluations of the side effects of medications used to control high blood pressure can provide important new insights into the correlations between patients' sentiments about treatments and factors such as demographic characteristics or location. These data can also reveal correlations between the frequency of retweets of health messages and the institutional or interpersonal features of those messages. For instance, are messages from the Centers for Disease Control and Prevention, the American Heart Association, or the American Stroke Association disregarded because of perceived institutional bias? Would these messages be more effective if they were disseminated through peer networks with stronger affective content? Observational and ethnographic methods applied to open social media networks may suggest specific experimental interventions that can be used to design large-scale health campaigns deployed over social media.

These new scientific resources may also offer important new directions for applying theories of social influence to the design of social institutions aimed at improving population health.[49] The combination of network theory and social media has shown that strategically structured online communities can create social environments that promote behavior change. What remains to be discovered is how far-reaching these implications will be for the development of new online health communities aimed at promoting long-term healthcare and behavioral improvement.[15,16,41] For instance, social media sites focusing on heart health might combine medical records with personal reports about medication side effects or testimonials about the social value of exercise partners for increased fitness. For patients, such sites would be valuable resources for learning about new treatments, and receiving encouragement to continue their existing routines. For clinicians and health researchers, these sites would provide valuable social and behavioral data on patient interactions, which could be correlated with patients' real health trajectories. The capacity and design of such environments will necessarily mature as new technologies become available. Over the last decade, we have learned that these venues for social interaction can be highly valuable sources of data that can be used both to gain new scientific insight into how social processes affect health behaviors, and to foster greater understanding in the physician community about patients' needs. The positive results from recent studies, along with the new opportunities for connecting offline health behaviors with social media tools, invite many new possibilities for improving our understanding of the social dimensions of heart health.[50]

Acknowledgments

I thank the James S. McDonnell Foundation and the Robert Wood Johnson Foundation for research support.

Disclosures

None.

References

1. Smith K, Christakis N. Social networks and health. *Annu Rev Sociol*. 2008;34:405–429.
2. Pampel FC, Krueger PM, Denney JT. Socionomic disparities in health behaviors. *Annu Rev Sociol*. 2010;36:349–370.
3. Umberson D, Crosnoe R, Reczek C. Social relationships and health behavior across life course. *Annu Rev Sociol*. 2010;36:139–157.
4. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357:370–379.
5. Christakis N, Fowler J. Dynamics of Smoking behavior in a large social network. *N Engl J Med*. 2008;358:2249–2258.
6. Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. *Annu Rev Public Health*. 2007;28:69–93.
7. Hershey J, Asch D, Thumasathit T, Meszaros J, Waters V. The roles of altruism, free riding, and bandwagoning in vaccination decisions. *Organ Behav Hum Decis Process*. 1994;59:177–187.
8. Cohen-Cole E, Fletcher JM. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J Health Econ*. 2008;27:1382–1387.
9. Lyons R. The spread of evidence-poor medicine via flawed social-network analysis. *Stat Politics Policy*. 2011;1–26.
10. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational data. *Sociol Methods Res*. 2011;40: 211–239.
11. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci USA*. 2009;106:21544–21549.
12. Steglich C, Snijders T, Pearson M. Dynamic networks and behavior: separating selection from influence. *Soc Method* 2010; 40:329–393.
13. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. *J Med Internet Res*. 2009;11:e48.
14. McNab C. What social media offers to health professionals and citizens. *Bull World Health Organ*. 2009;87:566.
15. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol*. 2011;7:e1002199.
16. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health Aff (Millwood)*. 2009;28:361–368.
17. Kiesler CA. Policy implications of research on social support and health. In: Cohen S, Syme SL, eds. *Social Support and Health*. San Diego, CA: Academic Press; 1985:347–364.
18. Kaplan BH, Cassel JC, Gore S. Social support and health. *Med Care*. 1977;15(suppl):47–58.
19. White M, Dorman SM. Receiving social support online: implications for health education. *Health Educ Res*. 2001;16:693–707.

20. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Internet use and social support in women with breast cancer. *Health Psychol*. 2002;21:398–404.
21. Thackeray R, Neiger BL, Hanson CL, McKenzie JF. Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. *Health Promot Pract*. 2008;9:338–343.
22. Kawachi I, Berkman LA. *Neighborhoods and Health*. New York: Oxford University Press; 2003.
23. Berkman LA, Kawachi I. A historical framework for social epidemiology. In: Berkman L, Kawachi I, eds. *Social Epidemiology*. New York: Oxford University Press; 2000:3–12.
24. Krieger N. A glossary for social epidemiology. *J Epidemiol Community Health*. 2001;55:693–700.
25. Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. *N Engl J Med*. 1992;326:1097–1101.
26. Shipman SA, Lan J, Chang CH, Goodman DC. Geographic maldistribution of primary care for children. *Pediatrics*. 2011;127:19–27.
27. Erickson T. Social interaction on the net: virtual community as participatory genre. In: *Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences*; Maui, Hawaii: Shidler College of Business, University of Hawaii at Manoa; 1997;6:13–21.
28. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. *Health Info Libr J*. 2007;24:2–23.
29. Centola D, Eguiluz V, Macy M. Cascade dynamics of complex propagation. *Physica A*. 2007;374:449–456.
30. Centola D. The spread of behavior in an online social network experiment. *Science*. 2010;329:1194–1197.
31. Correll SJ, Ridgeway CL. *Handbook of Social Psychology*. Delamater J, ed. New York: Kluwer Academic Publishers; 2003;29–51.
32. Axelrod R. The dissemination of culture: a model with local convergence and global polarization. *J Conflict Resolut*. 1997;41:203–226.
33. Klemm K, Eguíluz VM, Toral R, Miguel MS. Global culture: a noise induced transition in finite systems. *Phys Rev E*. 2003;67:045101 R.
34. Centola D, Gonzalez-Avella JC, Eguiluz V, San Miguel M. Homophily cultural drift, and the co-evolution of cultural groups. *J Conflict Resolut*. 2007;51:905–929.
35. Valente T, Vlahov D. Selective risk taking among needle exchange participants. *Am J Pub Heal*. 2001;91:406–411.
36. Granovetter M. The strength of weak ties. *Am J Sociol*. 1973:1360–1380.
37. Schelling T. *Micromotives and Macrobehavior*. New York: WW Norton & Co; 1978.
38. Centola D, Macy M. Complex contagions and the weakness of long ties. *Am J Sociol*. 2007;113:702–704.
39. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. *Nature*. 1998;393:440–442.
40. Watts DJ. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press; 1999.
41. Christakis N, Fowler J. *Connected*. Boston: Little, Brown and Co; 2009.
42. Centola D. An experimental study of homophily in the adoption of health behavior. *Science*. 2011;334:1269–1272.
43. Narayanan A., Shmatikov V. Robust de-anonymization of large sparse datasets (how to break anonymity of the Netflix prize dataset). In: *Proceedings of the 29th IEEE Symposium on Security and Privacy*. 2008:111–125, Oakland, CA.
44. Bilton N. Facebook security flaw exposes private chats. *The New York Times*. May 10, 2010.
45. Malin B, Sweeney L. How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *J Biomed Inform*. 2004;37:179–192.
46. Sweeney L. Weaving technology and policy together to maintain confidentiality. *J Law Med Ethics*. 1997;25:98–110, 82.
47. DiMaggio P, Hargittai E, Celeste C, Shafer S. Digital inequality: from unequal access to differentiated use. In: Neckerman K, ed. *Social Inequality*. New York: Russell Sage Foundation; 2004;355–400.
48. Berke EM, Choudhury T, Ali S, Rabbi M. Objective sensing of activity and sociability: mobile sensing in the community. *Ann Fam Med*. 2011;9:344–350.
49. Valente TW. Network interventions. *Science*. 2012;337:49–53.
50. Van der Leij M. Experimenting with buddies. *Science*. 2011;334:1220–1221.

Key Words: behavior therapy ■ exercise ■ mass screening ■ public policy ■ social support

Circulation

American Heart Association®

Social Media and the Science of Health Behavior

Damon Centola

Circulation. 2013;127:2135-2144
doi: 10.1161/CIRCULATIONAHA.112.101816
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7322. Online ISSN: 1524-4539