CHAPTER 4

Special People

It may seem hard to believe in a time when "social networking" has become so commonplace an idea that it shows up in everything from feature films to Foster's beer commercials, but it wasn't that long ago—as recently as the mid-1990s—that the study of social networks was relatively obscure, pursued mostly by a small cadre of mathematically inclined sociologists interested in mapping the social interactions among individuals.[1] The field has exploded in recent years, in large part because fast computers, along with communication technologies like e-mail, cell phones, and social networking sites such as Facebook have made it possible to record and analyze these interactions with great precision, even for hundreds of millions of people at a time. Nowadays, thousands of computer scientists, physicists, mathematicians, and even biologists count themselves as "network scientists," and new discoveries about the structure and dynamics of networked systems arrive daily.[2]

SIX DEGREES OF SEPARATION

Back in 1995, however, when I was a graduate student at Cornell studying the synchronization of chirping crickets, all this was in the future. Back then, in fact, the idea that everyone in the world is connected through a giant social network

through which information, ideas, and influence might flow, was still sufficiently novel that when my father asked me during one of our regular phone conversations if I'd ever heard of the notion that "everyone in the world is only six handshakes away from the president of the United States," I assumed it was folklore. And in some sense it was. People have been fascinated with what sociologists call the small-world problem for nearly a century, since the Hungarian poet Frigyes Karinthy published a short story called "Chains" in which his protagonist boasts that he can connect himself to any other person in the world, whether a Nobel Prize winner or a worker in a Ford Motor factory, through a chain of no more than five acquaintances. Four decades later, in her polemic on urban planning *The Death and Life of Great American Cities,* the journalist Jane Jacobs described a similar game, called messages, that she used to play with her sister when they first moved to New York:

> The idea was to pick two wildly dissimilar individuals—say a headhunter in the Solomon Islands and a cobbler in Rock Island, Illinois—and assume that one had to get a message to the other by word of mouth; then we would each silently figure out a plausible, or at least possible, chain of persons through whom the message could go. The one who could make the shortest plausible chain of messages won.

But how long are these chains in reality? One way to answer the question would be to map out all the links in the social network of the whole world and then simply count by brute force how many people you can reach in one "degree of separation," how many at "two degrees," and so on, until you have reached

everyone. In Jacobs's day that was impossible, but in 2008 two computer scientists at Microsoft Research got somewhat close when they computed the length of paths connecting pairs of individuals in Microsoft's 240-million-strong instant messenger network, where being "friends" in this case meant being on each other's buddy lists.[3] On average they found that people were separated by about seven steps—remarkably close to the six handshakes that my father had mentioned. Yet this can't be the real answer to the question. The characters in Jacobs's fictional game didn't have access to this network, so they couldn't have computed the paths the way the Microsoft researchers did even if they had the computing power to do so. Clearly they must have used some other method to direct their messages. And indeed, according to Jacobs, they did:

> The headhunter would speak to the headman of his village, who would speak to the trader who came to buy copra, who would speak to the Australian patrol officer when he came through, who would tell the man who was next slated to go to Melbourne on leave, etc. Down at the other end, the cobbler would hear from his priest, who got it from the mayor, who got it from the state senator, who got it from the governor, etc. We soon had these close-to-home messengers down to a routine for almost everybody we could conjure up, but we would get tangled up in long chains at the middle until we began employing Mrs. Roosevelt. Mrs. Roosevelt made it suddenly possible to skip whole chains of intermediate connections. She knew the most unlikely people. The world shrank remarkably.[4]

Jacobs's solution, in other words, assumes that social networks are organized in a hierarchy: Messages flow up the

hierarchy from the periphery and then back down again, with high-status figures like Mrs. Roosevelt occupying the critical center. We are so used to a world of hierarchies—whether inside formal organizations, across the economy, or in society—that it is natural to assume that social networks should be hierarchical as well. Karinthy, in fact, used a similar line of reasoning to Jacobs's, where in place of Mrs. Roosevelt he invoked Mr. Ford, writing that "to find a chain of contacts linking myself with an anonymous riveter at the Ford Motor Company…The worker knows his foreman, who knows Mr. Ford himself, who in turn is on good terms with the director general of the Hearst publishing empire. It would take but one word from my friend to send a cable to the general director of Hearst asking him to contact Ford who could in turn contact the foreman, who could then contact the riveter, who could then assemble a new automobile for me, should I need one."

As plausible as this method sounds, however, it is *not* how messages actually propagate through social networks, as we know now from a series of "small-world experiments" that began not long after Jacobs was writing. The first of these experiments was conducted by none other than Stanley Milgram, the social psychologist whose subway experiment I discussed in Chapter 1. Milgram recruited three hundred people, two hundred from Omaha, Nebraska, and the other hundred from around Boston, to play a version of the messages game with a Boston stockbroker who was a friend of Milgram's and who had volunteered to serve as the "target" of the exercise. Much as in Jacobs's imaginary version, participants in Milgram's experiment knew whom they were trying to reach, but could only send the message to someone whom they knew on a first-name basis; thus each of the three hundred "starters" would send it to a friend, who would

send it to a friend, and so on, until someone either refused to participate or else the message chain reached the target. As luck would have it, sixty-four of the initial chains made it all the way to their destination, and the average length of those that did was indeed about six; hence the famous phrase "six degrees of separation."[5]

But although Milgram's subjects were able to find paths as short as those hypothesized by Karinthy and Jacobs, it wasn't because they employed Mrs. Roosevelt or anyone like her. Instead, ordinary people passed messages to other ordinary people, tracking along the same social stratum rather than going up and down the hierarchy as both Karinthy and Jacobs imagined. Nor did the chains get tangled up in the middle as Jacobs worried they might. Instead they experienced their greatest difficulties after they had already gotten close to their targets. Social networking, it seems, is less like a pyramid than it is like a game of golf—where, the old adage goes, you "drive for show, putt for dough." When you are far away from the target, that is, it's relatively easy to jump large distances simply by sending the message to someone in the right country, and from there to someone in the right city, and then to someone in the right profession. But once you get close to the target, big jumps don't help you anymore and messages have a tendency to bounce around until they find someone who knows the target.

Nevertheless, Milgram still found that not all message handlers are created equal. In fact, of the sixty-four messages that got through, nearly half of them were delivered to the target by one of three people, and half of those—sixteen chains—were delivered by a single person, "Mr. Jacobs," a clothing merchant who was a neighbor of the target. Struck by this concentration of messages into the hands of a few individuals, Milgram speculated that what he called socio-

metric stars might be important to understanding how the small-world phenomenon worked.[6] Milgram himself didn't conclude much more than that, but three decades later, in an essay called "Six Degrees of Lois Weisberg," *New Yorker* writer Malcolm Gladwell used Milgram's finding about Mr. Jacobs to make the argument that "a very small number of people [like Mrs. Weisberg] are linked to everyone else in a few steps, and the rest of us are linked to the world through those few."[7] In other words, even though Mr. Jacobs and Ms. Weisberg are not "important" in the same way that Mrs. Roosevelt or Mr. Ford were important, from a network perspective they end up serving the same kind of function—like hubs in an airline network that we necessarily pass through in order to get from one part of the world to another.

Like Jacobs's hierarchy, the airline network metaphor is an appealing one, but it says more about how we would organize the world if given the opportunity to do so than it says about how the world is actually organized. If you think about it for minute, in fact, the metaphor is actually quite implausible. Some people clearly have more friends than others. But people are not like airports—they can't just tack on an extra wing when they need to handle more traffic. As a result, the number of friends that people have doesn't vary by nearly as much as the traffic in airports. An average person, for example, has a few hundred friends, while the most gregarious top out around a couple of thousand—roughly ten times as many. That is a big difference, but not remotely comparable to a true hub like O'Hare, which handles thousands of times as many passengers as a small airport. So how is it that connectors in social networks can nevertheless act like hubs in airline networks?[8]

In fact, they do not, as my collaborators Roby Muhamad and Peter Dodds and I found several years ago when we rep-

licated Milgram's original experiment—only this time we used e-mail in place of physical packets, allowing us to work on a much larger scale. Whereas Milgram had three hundred initial senders in two cities attempting to reach a single target in Boston, we had more than twenty thousand chains in search of one of eighteen targets in thirteen different countries around the world. By the time the experiment had ended, the chains had passed through over 60,000 people in 166 countries. Using some more up-to-date statistical analysis than Milgram had available to him, we were also able to estimate not only the length of the chains that made it to their targets, but also how long the chains that failed would have been had they continued. Our main finding was remarkably close to Milgram's—roughly half of all chains should be expected to reach their targets in seven steps or fewer. Given the differences between the two experiments, which were conducted on very different scales using different technologies and nearly forty years apart, it is actually sort of amazing that the results accorded so closely, and provides strong support for the claim that many people—although certainly not all people—can connect to one another through short chains.[9]

Unlike Milgram's findings, however, we discovered no "hubs" in the delivery process. Rather, messages reached their targets through almost as many recipients as there were chains. We also asked people why they chose the next person in the chain, and here, too, we discovered little evidence of hubs or stars. Subjects in small-world experiments, it turns out, do not typically pass messages to their highest-status or most-connected friends. Instead, they pass them to people they think have something in common with the target, like geographic proximity or a similar occupation, or they simply pass them to people they think will be likely to continue passing it along. Ordinary individuals, in other words, are just as

capable of spanning critical divides between social and professional circles, between different nations, or between different neighborhoods, as exceptional people. When you want to get a message to a graduate student in Novosibirsk, Russia, for example, you don't think about whom you know who has a lot of friends, or goes to lots of parties, or has connections to the White House. You think about whether you know any Russians. And if you don't know any Russians, then maybe you know someone from Eastern Europe, or someone who has traveled to Eastern Europe, or has studied Russian, or who lives in a part of your city that is known for its Eastern European immigrants. Mrs. Roosevelt, or Lois Weisberg for that matter, may indeed connect many people. But those same people have many other ways of connecting as well. And it is these other, less obvious ways that they tend to actually use, if only because there are so many more of them.

The overall message here is that real social networks are connected in more complex and more egalitarian ways than Jacobs or even Milgram imagined—a result that has now been confirmed with many experiments, empirical studies, and theoretical models.[10] In spite of all this evidence, however, when we think about how social networks work, we continue to be drawn to the idea that certain "special people," whether famous wives of presidents or gregarious local businessmen, are disproportionately responsible for connecting the rest of us. Evidence, in fact, seems to have very little to do with why we think this way. After all, Jacobs was writing years before Milgram's experiments and long before anyone had the kind of data that might have supported her claim about Mrs. Roosevelt. So wherever she got the idea from, it obviously wasn't based on any actual evidence. Rather, it seems that Jacobs was drawn to the idea that a few special people connect everyone else simply because without invoking such people

it's hard to come with any explanation at all. The result is that no matter how many times the evidence rules out one kind of special person, we simply insert another. If it's not Mrs. Roosevelt, then it must be Lois Weisberg, and if it's not Lois Weisberg, then it must be Mr. Jacobs the clothing merchant. And if it's not him, then it must be our friend Ed who seems to know everyone. "It's got to be someone special." We feel compelled to conclude: "How else could it work?"

Nor is the intuitive appeal of special-people explanations restricted to problems to do with networks. The "great man" view of history explains important historical events in terms of the actions of a few critical leaders. Conspiracy theorists imbue shadowy government agents or secret cabals with near infinite capabilities to meddle with society. Media analysts credit high-profile celebrities with setting fashion trends or selling products. Corporate boards pay exorbitant amounts for a CEO whose decisions will shape the fate of the entire company. Epidemiologists worry that a few "superspreaders" can trigger an epidemic. And marketers extol the power of "influencers" to make or break a brand, change social norms, or otherwise shift public opinion.[11] In his book *The Tipping Point,* for example, Gladwell explains the origins of what he calls social epidemics, meaning everything from fads and fashions to shifts in cultural norms and sudden drops in crime rates, in terms of what he calls the law of the few. Just as superspreaders drive real epidemics and great men drive history, so too the law of the few claims that social epidemics are "driven by the efforts of a handful of exceptional people." For example, in discussing the mysterious resurgence of Hush Puppies in the mid-1990s, Gladwell explains that

> the great mystery is how those shoes went from something worn by a few fashion-forward downtown Man-

> hattan hipsters to being sold in malls across the country. What was the connection between the East Village and Middle America? The Law of the Few says the answer is that one of these exceptional people found out about the trend, and through social connections and energy and enthusiasm and personality spread the word about Hush Puppies just as people like Gaeten Dugas and Nushawn Williams were able to spread HIV.[12]

Gladwell's law of the few is catnip to marketers and businessmen and community organizers and just about anyone else in the business of shaping or manipulating people. And it's easy to see why. If you can just find these special people and influence *them*, their connections and energy and enthusiasm and personality would be put to work for you. It's a plausible-sounding story, and yet as with so many appealing ideas about human behavior, the law of the few turns out to be more a matter of perception than reality.

THE INFLUENCERS

The culprit again is common sense. As marketing consultants Ed Keller and Jon Berry argue, "Some people are better connected, better read, and better informed. You probably know this from your own experience. You don't turn to just anyone when you're deciding what neighborhood to live in, how to invest for retirement, or what kind of car or computer to buy."[13] As a description of our perceptions, this statement is probably accurate—when we think about what we're doing when we seek out information, access, or advice, it does indeed seem that we focus on some people over others. But as I've already discussed, our perceptions of how we behave are far from a perfect reflection of reality. A number of studies,

for example, have suggested that social influence is mostly subconscious, arising out of subtle cues that we receive from our friends and neighbors, not necessarily by "turning to them" at all.[14] Nor is it clear that when we are influenced in these other, less conscious ways, we recognize that we have been influenced. Employees, for example, may well influence their bosses as much as their bosses influence them, but they are not equally likely to name each other as sources of influence—simply because bosses are supposed to be influential, whereas employees are not. In other words, our perceptions of who influences us may say more about social and hierarchical relations than influence per se.

One of the most confusing aspects of the influencer debate, in fact, is that no one can really agree on who the influencers are in the first place. Originally the term referred to "ordinary" people who just happened to exert extraordinary influence over their friends and neighbors. But in practice all sorts of people are referred to as influencers: media giants like Oprah Winfrey; gatekeepers like Anna Wintour, the editor of *Vogue;* celebrity actors and socialites; popular bloggers; and so on. All of these people may or may not be influential in their own way, but the kind of influence they exert varies tremendously. Oprah Winfrey's advocacy of an unknown book, for example, may dramatically improve its chances of appearing on the bestseller lists, but that is mostly because her individual influence is magnified enormously by the media empire that she runs. Likewise, a fashion designer might be well advised to have a famous actress arrive at the Oscars wearing his dress, but that is again because her arrival is being recorded, broadcast, and commented upon by the mass media. And when a popular blogger expresses his enthusiasm for a particular product, potentially thousands of people read his opinion. But is his or her influence analogous

to that of an Oprah endorsement, a personal recommendation from a friend, or something else?

Even if we narrow down the problem to direct, interpersonal influence of the kind that excludes the media, celebrities, and bloggers of the world, measuring influence is a lot more difficult than simply measuring the length of message chains. For example, to demonstrate just one incident of influence between two friends, Anna and Bill, you need to demonstrate that whenever Anna adopts a certain idea or product, Bill is more likely to adopt the same idea or product as well.[15] Even keeping track of just one such relationship would not be easy. And as researchers quickly discovered, doing it for many people simultaneously is prohibitively difficult. In place of observing influence directly, therefore, researchers have proposed numerous proxies for influence, such as how many friends an individual has, or how many opinions they voice, or how expert or passionate they are about a topic, or how highly they score on some personality test—things that are easier to measure than influence itself. Unfortunately, while all these measures are plausible substitutes for influence, they all derive from assumptions about how people are influenced, and no one has ever tested these assumptions. In practice, therefore, nobody really knows who is an influencer and who isn't.[16]

This ambiguity is confusing, but it's still not the real source of the problem. If we could invent a perfect instrument for measuring influence, presumably we would find that some people are indeed more influential than others. Yet some people are also taller than others and that is not necessarily something about which marketers should care. So why are they so excited about influencers? Consider, for example, that many studies count someone as an influencer if at least three acquaintances named them as someone to whom they would

turn for advice. Now, in a world where the average person influences just one other person, influencing three others makes you 300 percent as influential as average—a big difference. But on its own it doesn't solve the kinds of problems that marketers care about, like generating a hit product, driving public health awareness, or influencing a political candidate's election chances. All these problems require influencing millions of individuals. So even if each one of your influencers can influence three other ordinary people, you will still need to find and influence a million of them, which is rather different from what the law of the few promises. As it turns out, there's a solution to this problem as well, but it requires that we incorporate another related but distinct idea from network theory—that of social contagion.

THE ACCIDENTAL INFLUENTIALS

Contagion—the idea that information, and potentially influence, can spread along network ties like an infectious disease—is one of the most intriguing ideas in network science. As we saw in the last chapter, when everyone is being influenced by what other people are doing, surprising things can happen. But contagion also has important implications for influencers—because once you include the effects of contagion, the ultimate importance of an influencer is not just the individuals he or she influences directly but also all those influenced indirectly, via his neighbors, his neighbors' neighbors, and so on. It is through contagion, in fact, that the law of the few gets its real power. Because if just the right influencers can trigger a social epidemic, then influencing four million people may in fact require only a few of them. That's not a good deal—that's a *great* deal. And because finding and influencing just a few people

is quite different from finding and influencing a million, it qualitatively changes the nature of influence.[17]

What it means, though, is that the law of the few is not one, but two hypotheses that have been mashed together: first that some people are more influential than others; and second, that the influence of these people is greatly magnified by some contagion process that generates social epidemics.[18] It was therefore this combination of claims that Peter Dodds and I set out to test a few years ago in a series of computer simulations. Because these simulations required us to write down explicit mathematical models of how influence spreads, we had to specify all the assumptions that are typically left unstated in anecdotal descriptions of influencers. How should an influencer be defined? Who influences whom? What kinds of choices are individuals making? And how are these choices influenced by others? As I've discussed, no one really knows the answers to any of these questions; thus it's necessary, as in any modeling exercise, to make a number of assumptions, which could of course be wrong. Nevertheless, to cover our bases as much as possible, we considered two very different classes of models, each of which has been studied for decades by social and marketing scientists.[19]

The first was a version of Granovetter's riot model from the previous chapter. Unlike Granovetter's model, however, where everyone in the crowd observed everyone else, the interactions among individuals were specified by a network in which each individual got to observe only some relatively small circle of friends or acquaintances. The second model was a version of the "Bass model," named for Frank Bass, the marketing scientist who first proposed it as a model of product adoption, but closely related to an even older model used by mathematical epidemiologists to study the spread

of biological diseases. In other words, whereas Granovetter's model assumes that individuals adopt something when a certain fraction of their neighbors do, the Bass model assumes that adoption works like an infection process, with "susceptible" and "infected" individuals interacting along network ties.[20] The two models sound similar, but they're actually very different—which was important, because we didn't want our conclusions about the effect of influencers to depend too much on the assumptions of any one model.

What we found was that under most conditions, highly influential individuals were indeed more effective than the average person in triggering social epidemics. But their relative importance was much less than what the law of the few would suggest. To illustrate, consider an "influencer" who directly influences three times as many of his peers as the average person. Intuitively, one would expect that, all other things being equal, the influencer would also influence three times as many people indirectly as well. In other words, the influencer would exhibit a "multiplier effect" of three. The law of the few, it bears noting, claims that the effect would be much greater—that the disproportionality should be "extreme"—but what we found was the opposite.[21] Typically the multiplier effect for an influencer like this was less than three, and in many cases, they were not any more effective at all. Influencers may exist, in other words, but not the kind of influencers posited by the law of the few.

The reason is simply that when influence is spread via some contagious process, the outcome depends far more on the overall structure of the network than on the properties of the individuals who trigger it. Just as forest fires require a conspiracy of wind, temperature, low humidity, and combustible fuel to rage out of control over large tracts of land, social epidemics require just the right conditions to be satisfied by the

network of influence. And as it turned out, the most important condition had nothing to do with a few highly influential individuals at all. Rather, it depended on the existence of a critical mass of *easily influenced* people who influence other easy-to-influence people. When this critical mass existed, even an average individual was capable of triggering a large cascade—just as any spark will suffice to trigger a large forest fire when the conditions are primed for it. Conversely, when the critical mass did not exist, not even the most influential individual could trigger any more than a small cascade. The result is that unless one can see where particular individuals fit into the entire network, one cannot say much about how influential they will be—no matter what you can measure about them.

When we hear about a large forest fire, of course, we don't think that there must have been anything special about the spark that started it. Indeed, such an idea would be laughable. Yet when we see something special happen in the social world, we are instantly drawn to the idea that whoever started it must have been special also. And of course, whenever a large cascade did take place in our simulations, it was necessarily the case that someone had to have started it. However unexceptional that person might have seemed in advance, in retrospect they would seem to fit exactly the description of the law of the few: the "tiny percentage of people who do the majority of the work." What we knew from our simulations, however, was that there really *was* nothing special about these individuals—because we had created them that way. The majority of the work was being done not by a tiny percentage of people who acted as the triggers, but rather by the much larger critical mass of easily influenced people. What we concluded, therefore, is that the kind of influential person whose energy and connections can turn your book

into a bestseller or your product into a hit is most likely an accident of timing and circumstances. An "accidental influential" as it were.[22]

"ORDINARY INFLUENCERS" ON TWITTER

As many people immediately pointed out, this conclusion was based entirely on computer simulations. And as I've already mentioned, these simulations were highly simplified versions of reality, and made a large number of assumptions, any of which could have been wrong. Computer simulations are useful tools that can generate great insight. But in the end they are more like thought experiments than real experiments, and as such are better suited to provoking new questions than to answering them. So if we really want to know whether particular individuals are capable of stimulating the diffusion of ideas, information, and influence—and if these influencers exist, which attributes distinguish them from ordinary people—then we need to run experiments in the real world. But studying the relationship between individual influence and large-scale impact in the real world is easier said than done.

The main problem is that you need an enormous amount of data, and most of it is very hard to collect. Just demonstrating that one person has influenced another is difficult enough. And if you wanted to make the connection to how they influence larger populations, you need to gather similar information for whole chains of influence, in which one person influences another who in turn influences another, and so on. Pretty soon, you're talking about thousands or even millions of relationships, just to track how a single piece of information was spread. And ideally you would want to study many such cases. It's an over-

whelming amount of data to test what seems to be a relatively straightforward claim—that some people matter more than others—but there's no getting around it. It also helps explain why diffusion research, as it is known, has remained such a myth-laden business for so long: when it's impossible to prove anything, everyone is free to propose whatever plausible story they like. There's no way to decide who is right.

As with experiments like Music Lab, however, the Internet is starting to change this picture in important ways. A handful of recent studies have begun to explore diffusion in social networks on a scale that would have been unimaginable just a decade ago. Blog postings diffuse among networks of bloggers. Fan pages diffuse among networks of friends on Facebook. Special capabilities called "gestures" diffuse among players on the online game Second Life. And premium voice services have been shown to diffuse among networks of IM buddies.[23] Inspired by these studies, my Yahoo! colleagues Jake Hofman and Winter Mason and I, along with Eytan Bakshy, a talented graduate student at the University of Michigan, decided to look for the diffusion of information in the largest communication network we could get our hands on: Twitter. In the process, we would look for influencers.[24]

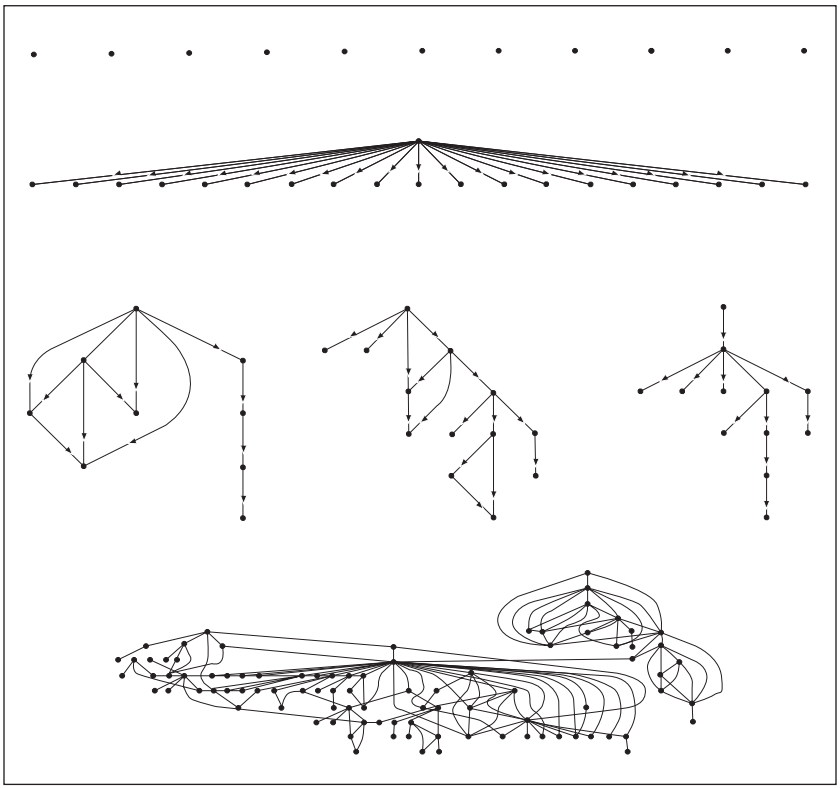
In many respects, Twitter is ideally suited to this objective. Unlike Facebook, say, where people connect to one another for a multitude of reasons, the whole point of Twitter is to broadcast information to other people—your "followers"—who have explicitly indicated that they want to hear from you. Getting people to pay attention to you—influencing them, in other words—is what Twitter is all about. Second, Twitter is remarkably diverse. Many users are regular people whose followers are mostly friends interested in hearing from them. But many of the most followed users on Twitter

are public figures, including bloggers, journalists, celebrities (Ashton Kutcher, Shaquille O'Neal, Oprah), media organizations such as CNN, and even government agencies and nonprofits (the Obama administration, No. 10 Downing Street, the World Economic Forum). This diversity is helpful because it allowed us to compare the influence of all manner of would-be influencers—ordinary people all the way up to Oprah and Ashton—in a consistent way.

Finally, although many tweets are mundane updates ("Having coffee at Starbucks on Broadway! It's a beautiful day!!"), many of them refer either to other online content, like breaking news stories and funny videos, or to other things in the world, like books, movies, and so on, about which Twitter users wish to express their opinions. And because the format of Twitter forces users to keep every message to no more than 140 characters, users often make use of "URL shorteners," such as bit.ly, to replace the long, messy URL of the original website with something like http://bit.ly/beRKJo. The nice thing about these shortened URLs is that they effectively assign a unique code to every piece of content broadcast on Twitter. Thus when a user wishes to "retweet" something, it's possible to see whom it came from originally, and thereby trace chains of diffusion across the follower graph.

In total, we tracked more than 74 million of these diffusion chains initiated by more than 1.6 million users, over a two-month interval in late 2009. For each event, we counted how many times the URL in question was retweeted—first by the original "seed" user's immediate followers, then by their followers, and their followers' followers, and so on—thereby tracing out the full "cascade" of retweets triggered by each original tweet. As the figure on page 102 shows, some of these cascades were broad and shallow, while others were narrow

and deep. Others still were very large, with complex structure, starting out small and trickling along before gaining momentum somewhere else in the network. Most of all, however, we found that the vast majority of attempted cascades—roughly 98 percent of the total—didn't actually spread at all.



Cascades on Twitter

This result is important because, as I'll discuss in more detail in the next chapter, if you want to understand why some things "go viral"—those occasional YouTube videos that attract millions of downloads, or funny messages that circu-

late wildly through e-mail or on Facebook—it's a mistake to consider only the rare few that actually succeed. In most settings, unfortunately, it is only possible to study the "successes" for the simple reason that nobody bothers to keep track of all the failures, which have a tendency to get swept under the rug. On Twitter, however, we can keep track of every single event, no matter how small, thereby enabling us to learn who is influential, how much more influential than average they really are, and whether or not it is possible to tell the differences between individuals in a way that could potentially be exploited.

The way we went about this exercise was to imitate what a hypothetical marketer might try to do—that is, using everything known about the attributes and past performance of a million or so individuals, to predict how influential each of them will be in the future. Based on these predictions, the marketer could then "sponsor" some group of individuals to tweet whatever information it is trying to disseminate, thereby generating a series of cascades. The better the marketer can predict how large a cascade any particular individual can trigger, the more efficiently it can allocate its budget for sponsored tweets. Actually running such an experiment is still extremely difficult in practice, so we instead did our best to approximate it using the data we had already collected. Specifically, we divided our data in two, artificially setting the first month of our time period as our "history" and the second half as the "future." We then fed all our "historical" data into a statistical model, including how many followers each user had, how many others they were following, how frequently they tweeted, when they had joined, and how successful they had been at triggering cascades during this period. Finally, we used the model to "predict" how influen-

tial each user would be in our "future" data and checked the model's performance against what actually transpired.

In a nutshell, what we found was that individual-level predictions are extremely noisy. Even though it was the case that on average, individuals with many followers who had been successful at triggering cascades of retweets in the past were more likely to be successful in the future, individual cases fluctuated wildly at random. Just as with the *Mona Lisa,* for every individual who exhibited the attributes of a successful influencer, there were many other users with indistinguishable attributes who were not successful. Nor did this uncertainty arise simply because we weren't able to measure the right attributes—in reality we had more data than any marketer would normally have—or to measure them accurately. Rather, the problem was that, like the simulations above, much of what drives successful diffusion depends on factors outside the control of the individual seeds. What this result suggests, in other words, is that marketing strategies that focus on targeting a few "special" individuals are bound to be unreliable. Like responsible financial managers, therefore, marketers should adopt a "portfolio" approach, targeting a large number of potential influencers and harnessing their average effect, thereby effectively reducing the individual-level randomness.[24]

Although promising in theory, a portfolio approach also raises a new issue, of cost effectiveness. To illustrate the point, consider a recent story in the *New York Times* that claimed that Kim Kardashian, the reality TV actress, was getting paid $10,000 per tweet by various sponsors who wanted her to mention their products. Kardashian at the time had well over a million followers, so it seems plausible that paying someone like her would generate more attention than paying some

ordinary person with only a few hundred followers. But how did they come up with that particular figure? Ordinary people, that is, might be prepared to tweet about their products for much less than $10,000. Assuming, therefore, that more visible individuals "cost" more than less visible ones, should marketers be targeting a relatively small number of more influential, more expensive, individuals or a larger number of less influential, less expensive individuals? Better yet, how should one strike the optimal balance?[25]

Ultimately, the answer to this question will depend on the specifics of how much different Twitter users would charge prospective marketers to sponsor their tweets—if indeed, they would agree to such an arrangement at all. Nevertheless, as a speculative exercise, we tested a range of plausible assumptions, each corresponding to a different hypothetical "influencer-based" marketing campaign, and measured their return on investment using the same statistical model as before. What we found was surprising even to us: Even though the Kim Kardashians of the world were indeed more influential than average, they were so much more expensive that they did not provide the best value for the money. Rather, it was what we called ordinary influencers, meaning individuals who exhibit average or even less-than-average influence, who often proved to be the most cost-effective means to disseminate information.

CIRCULAR REASONING AGAIN

Before you rush out to short stock in Kim Kardashian, I should emphasize that we didn't actually run the experiment that we imagined. Even though we were studying data from the real world, not a computer simulation, our statistical models still

made a lot of assumptions. Assuming, for example, that our hypothetical marketer could persuade a few thousand ordinary influencers to tweet about their product, it is not at all obvious that their followers would respond as favorably as they do to normal tweets. As anyone whose friend has tried to sell them on Amway products would know, there is something a little icky about a sales message embedded in a personal communication. People who follow Kim Kardashian, however, might have no such concerns; thus she may be far more effective in real life than our study could determine. Or perhaps our measure of influence—the number of retweets—was the wrong measure. We measured retweets because that's what we could measure, and that was definitely better than nothing. But presumably what you really care about is how many people click through to a story, or donate money to a charitable cause, or buy your product. Possibly Kardashian followers act on her tweets even when they don't retweet them to their friends—in which case, once again, we would have underestimated her influence.

Then again, we may not have. In the end, we simply don't know who is influential or what influencers, however defined, can accomplish. Until it is possible to measure influence with respect to some outcome that we actually care about, and until someone runs the real-world experiments that can measure the influence of different individuals, every result—including ours—ought to be taken with a grain of salt. Nevertheless, the findings I have discussed—from the small-world experiment, from the simulation studies of influence spreading on networks, and from the Twitter study—ought to raise some serious doubts about claims like the law of the few that explain social epidemics as the work of a tiny minority of special people.

It's not even clear, in fact, that social epidemics are the right way to think about social change to begin with. Although our Twitter study found that epidemic-like events do occur, we also found that they are incredibly rare. Of 74 million events in our data, only a few dozen generated even a thousand retweets, and only one or two got to ten thousand. In a network of tens of millions of users, ten thousand retweets doesn't seem like that big a number, but what our data showed is that even that is almost impossible to achieve. For practical purposes, therefore, it may be better to forget about the large cascades altogether and instead try to generate lots of small ones. And for that purpose, ordinary influencers may work just fine. They don't accomplish anything dramatic, so you may need a lot of them, but in harnessing many such individuals, you can also average out much of the randomness, generating a consistently positive effect.

Finally, and quite apart from any specific findings, these studies help us to see a major shortcoming of commonsense thinking. It is ironic in a way that the law of the few is portrayed as a counterintuitive idea because in fact we're so used to thinking in terms of special people that the claim that a few special people do the bulk of the work is actually extremely natural. We think that by acknowledging the importance of interpersonal influence and social networks, we have somehow moved beyond the circular claim from the previous chapter that "X happened because that's what people wanted." But when we try to imagine how a complex network of millions of people is connected—or worse still, how influence propagates through it—our intuition is immediately defeated. By effectively concentrating *all* the agency into the hands of a few individuals, "special people" arguments like the law of the few reduce the problem of understanding how network structure affects outcomes to the much simpler problem of

understanding what it is that motivates the special people. As with all commonsense explanations, it sounds reasonable and it might be right. But in claiming that "X happened because a few special people made it happen," we have effectively replaced one piece of circular reasoning with another.