The Role of Social Networks in Information Diffusion

Eytan Bakshy*
Facebook
1601 Willow Rd.
Menlo Park, CA 94025
ebakshy@fb.com

Itamar Rosenn
Facebook
1601 Willow Rd.
Menlo Park, CA 94025
itamar@fb.com

Cameron Marlow
Facebook
1601 Willow Rd.
Menlo Park, CA 94025
cameron@fb.com

Lada Adamic
University of Michigan
105 S. State St.
Ann Arbor, MI 48104
ladamic@umich.edu

ABSTRACT

Online social networking technologies enable individuals to simultaneously share information with any number of peers. Quantifying the causal effect of these mediums on the dissemination of information requires not only identification of who influences whom, but also of whether individuals would still propagate information in the absence of social signals about that information. We examine the role of social networks in online information diffusion with a large-scale field experiment that randomizes exposure to signals about friends' information sharing among 253 million subjects in situ. Those who are exposed are significantly more likely to spread information, and do so sooner than those who are not exposed. We further examine the relative role of strong and weak ties in information propagation. We show that, although stronger ties are individually more influential, it is the more abundant weak ties who are responsible for the propagation of novel information. This suggests that weak ties may play a more dominant role in the dissemination of information online than currently believed.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems; J.4 [**Social and Behavioral Sciences**]: Sociology

General Terms

Experimentation, Measurement, Human Factors

Keywords

social influence, tie strength, causality

1. INTRODUCTION

Social influence can play a crucial role in a range of behavioral phenomena, from the dissemination of information, to the adoption of political opinions and technologies [23, 42], which are increasingly mediated through online systems [17,

---

*Part of this research was performed while the author was a student at the University of Michigan.

38]. Despite the wide availability of data from online social networks, identifying influence remains a challenge. Individuals tend to engage in similar activities as their peers, so it is often impossible to determine from observational data whether a correlation between two individuals' behaviors exists because they are similar or because one person's behavior has influenced the other [5, 32, 39]. In the context of information diffusion, two people may disseminate the same information as each other because they possess the same information sources, such as web sites or television, that they consume regularly [3, 38].

Moreover, homophily – the tendency of individuals with similar characteristics to associate with one another [1, 28, 34] – creates difficulties for measuring the relative role of strong and weak ties in information diffusion, since people are more similar to those with whom they interact often [22, 34]. On one hand, pairs of individuals who interact more often have greater opportunity to influence one another and have more aligned interests, increasing the chances of contagion [11, 27]. However, this commonality amplifies the potential for confounds: those who interact more often are more likely to have increasingly similar information sources. As a result, inferences made from observational data may overstate the importance of strong ties in information spread. Conversely, individuals who interact infrequently have more diverse social networks that provide access to novel information [12, 22]. But because contact between such ties is intermittent, and the individuals tend to be dissimilar, any particular piece of information is less likely to flow across weak ties [14, 37]. Historical attempts to collect data on how often pairs of individuals communicate and where they get their information have been prone to biases [10, 33], further obscuring the empirical relationship between tie strength and diffusion.

Confounding factors related to homophily can be addressed using controlled experiments, but experimental work has thus far been confined to the spread of highly specific information within limited populations [6, 13]. In order to understand how information spreads in a real-world environment, we wish to examine a setting where a large population of individuals frequently exchange information with their peers. Facebook is the most widely used social networking service in the world, with over 800 million people using the service each month. For example, in the United States, 54% of adult Internet users are on Facebook [26].

Those American users on average maintain 48% of their real world contacts on the site [26], and many of these individuals regularly exchange news items with their contacts [38]. In addition, interaction among users is well correlated with self-reported intimacy [18]. Thus, Facebook represents a broad online population of individuals whose online personal networks reflect their real-world connections, making it an ideal environment to study information contagion.

We use an experimental approach on Facebook to measure the spread of information sharing behaviors. The experiment randomizes whether individuals are exposed via Facebook to information about their friends' sharing behavior, thereby devising two worlds under which information spreads: one in which certain information can only be acquired external to Facebook, and another in which information can be acquired within or external to Facebook. By comparing the behavior of individuals within these two conditions, we can determine the causal effect of the medium on information sharing.

The remainder of this paper is organized as follows. We further motivate our study with additional related work in Section 2. Our experimental design is described in Section 3. Then, in Section 4 we discuss the causal effect of exposure to content on the newsfeed, and how friends' sharing behavior is correlated in time, irrespective of social influence via the newsfeed. Furthermore, we show that multiple sharing friends are predictive of sharing behavior regardless of exposure on the feed, and that additional friends do indeed have an increasing causal effect on the propensity to share. In Section 5 we discuss how tie strength relates to influence and information diffusion. We show that users are more likely to have the same information sources as their close friends, and that simultaneously, these close friends are more likely to influence subjects. Using the empirical distribution of tie strength in the network, we go on to compute the overall effect of strong and weak ties on the spread of information in the network. Finally, we discuss the implications of our work in Section 6.

2. RELATED WORK

Online networks are focused on sharing information, and as such, have been studied extensively in the context of information diffusion. Diffusion and influence have been modeled in blogs [2, 20, 25], email [31], and sites such as Twitter, Digg, and Flickr [8, 21, 29]. One particularly salient characteristic of diffusion behavior is the correlation between the number of friends engaging in a behavior and the probability of adopting the behavior. This relationship has been observed in many online contexts, from the joining of Live-Journal groups [7], to the bookmarking of photos [15], and the adoption of user-created content [9]. However, as Anagnostopoulos, et al. [4] point out, individuals may be more likely to exhibit the same behavior as their friends because of homophily rather than as a result of peer influence. Statistical techniques such as permutation tests and matched sampling [5] help control for confounds, but ultimately cannot resolve this fundamental problem [39].

Not all diffusion studies must infer whether one individual influenced another. For example, Leskovec et al. [30] study the explicit graph of product recommendations, Sun et al. [41] study cascading in page fanning, and Bakshy et al. [9] examine the exchange of user-created content. However, in all these studies, even if the source of a particular

contagion event is a friend, such data does not tell us about the relative importance of social networks in information diffusion. For example, consider the spread of news. In Bradley Greenberg's classsic study of media contagion [24], 50% of respondents learned about the Kennedy assassination via interpersonal ties. Despite the substantial word-of-mouth spread, it is clear that all of the respondents would have gotten the news at a slightly later point in time (perhaps from the very same media outlets as their contacts), had they not communicated with their peers. Therefore, a complete understanding of the importance of social networks in information diffusion not only requires us to identify sources of interpersonal contagion, but also requires a counterfactual understanding of what would happen if certain interactions did not take place.
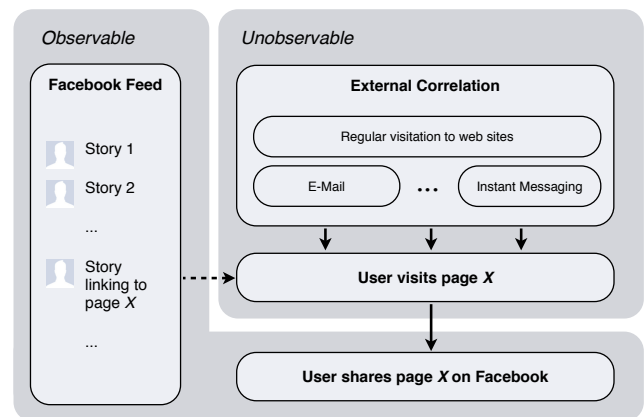

Figure 1:   Causal relationships that explain diffusion-like phenomena. Information presented in users' news feeds and other sharing behavior on `facebook.com` **are observed.   External events that cause users to be exposed to information outside of Facebook cannot be observed and may explain their sharing behavior. Our experiment blocks the causal relationship (dashed arrow) between the Facebook newsfeed and user visitation by randomly removing stories about friends' sharing behavior in subjects' feeds. Thus, our experiment allows us to compare situations where both influence via the feed and external correlations exist (the** *feed* **condition), to situations in which only external correlations exist (the** *no feed* **condition).**

3. EXPERIMENTAL DESIGN AND DATA

Facebook users primarily interact with information through an aggregated history of their friends' recent activity (stories), called the News Feed, or simply feed for short. Some of these stories contain links to content on the Web, uniquely identified by URLs. Our experiment evaluates how much exposure to a URL on the feed increases an individual's propensity to share that URL, beyond correlations that one might expect among Facebook friends. For example, friends with whom a user interacts more often may be more likely to visit sites that the user also visits. As a result, those friends may be more likely to share the same URL as the
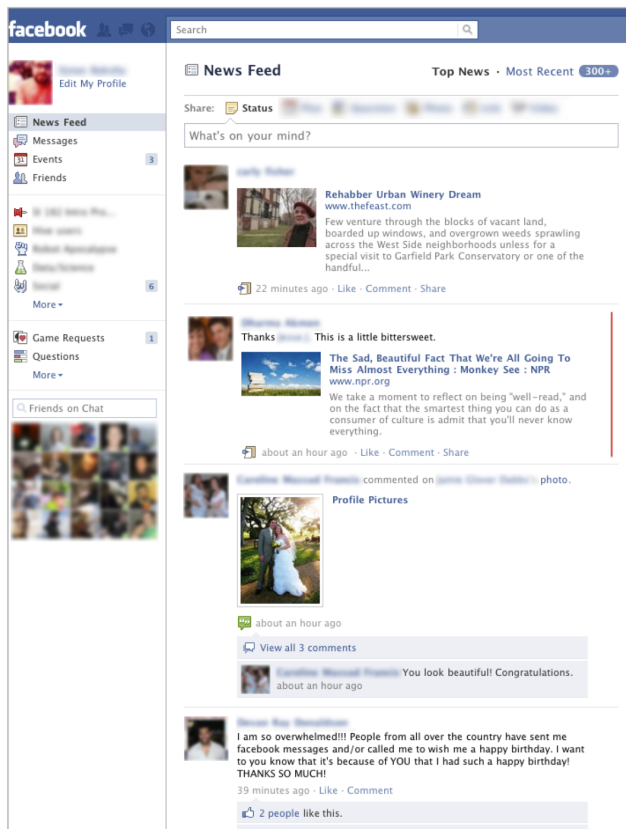
(a)   (b)

Figure 2: An example of the Facebook News Feed interface for a hypothetical subject who has a link (highlighted in red) assigned to the (a) *feed* **or (b)** *no feed* **condition.**

user before she has the opportunity to share that content herself. Additional unobserved correlations may arise due to external influence via e-mail, instant messaging, and other social networking sites. These causal relationships are illustrated in Figure 1. From the figure, one can see that all unobservable correlations can be identified by blocking the causal relationship between the Facebook feed and sharing. Our experiment therefore randomizes subjects with respect to whether they receive social signals about friends' sharing behavior of certain Web pages via the Facebook feed.

3.1 Assignment Procedure

Subject-URL pairs are randomly assigned at the time of display to either the *no feed* or the *feed* condition. Stories that contain links to a URL assigned to the *no feed* condition for the subject are never displayed in the subject's feed. Those assigned to the *feed* condition are not removed from the feed, and appear in the subject's feed as normal (Figure 2). Pairs are deterministically assigned to a condition at the time of display, so any subsequent share of the same URL by any of a subject's friends is also assigned to the same condition. To improve the statistical power of our results, twice as many pairs were assigned to the *no feed* condition. Because removal from the feed occurs on a subject-URL basis, and we include only a small fraction of subject-URL pairs in the *no feed* condition, a shared URL is on average delivered to over 99% of its potential targets.

All activity relating to subject-URL pairs assigned to either experimental condition is logged, including feed exposures, censored exposures, and clicks to the URL (from the feed or other sources, like messaging). Directed shares, such as a link that is included in a private Facebook message or explicitly posted on a friend's wall, are not affected by the assignment procedure. If a subject-URL pair is assigned to an experimental condition, and the subject clicks on content containing that URL in any interface other than the feed, that subject-URL pair is removed from the experiment. Our experiment, which took place over the span of seven weeks, includes 253,238,367 subjects, 75,888,466 URLs, and 1,168,633,941 unique subject-URL pairs.

3.2 Ensuring Data Quality

Threats to data quality include using content that was or may have been previously seen by subjects on Facebook prior to the experiment, content that subjects may have seen through interfaces on Facebook other than feed, spam, and malicious content. We address these issues in a number of ways. First, we only consider content that was shared by the subjects' friends only after the start of the experiment. This enables our experiment to accurately capture the first time a subject is exposed to a link in the feed, and ensures that URLs in our experiment more accurately reflect content that is primarily being shared contemporaneously with the timing of the experiment. We also exclude potential subject-

Demographic Feature (% of subjects)	feed	no feed
Gender		
FEMALE	51.6%	51.4%
MALE	46.7%	47.0%
UNSPECIFIED	1.5%	1.5%
Age		
17 OR YOUNGER	12.8%	13.1%
18-25	36.4%	36.1%
26-35	27.2%	26.9%
36-45	13.0%	12.9%
46 OR OLDER	10.6%	10.9%
Country (top 10 & other)		
UNITED STATES	28.9%	29.1%
TURKEY	6.1%	5.8%
GREAT BRITAIN	5.1%	5.2%
ITALY	4.2%	4.1%
FRANCE	3.8%	3.9%
CANADA	3.7%	3.8%
INDONESIA	3.7%	3.5%
PHILIPPINES	2.1%	2.3%
GERMANY	2.3%	2.3%
MEXICO	2.0%	2.1%
226 OTHERS	37.5%	37.7%

Table 1: Summary of demographic features of subjects assigned to the *feed* ($N = 160,688,092$) and *no feed* ($N = 218,743,932$) condition. Some subjects may appear in both columns.

URL pairs where the subject had previously clicked on the URL via any interface on the site at any time up to two months prior to exposure, or any interface other than the feed for content assigned to the *no feed* condition. Finally, we use the Facebook's site integrity system [40] to classify and remove URLs that may not reflect ordinary users' purposeful intentions of distributing content to their friends.

3.3 Population

The experimental population consists of a random sample of all Facebook users who visited the site between August 14th to October 4th 2010, and had at least one friend sharing a link. At the time of the experiment, there were approximately 500 million Facebook users logging in at least once a month. Our sample consists of approximately 253 million of these users. All Facebook users report their age and gender, and a user's country of residence can be inferred from the IP address with which she accesses the site. In our sample, the median and average age of subjects is 26 and 29.3, respectively. Subjects originate from 236 countries and territories, 44 of which have one million or more subjects. Additional summary statistics are given in Table 1, and show that subjects are assigned to the conditions in a balanced fashion.

3.4 Evaluating Outcomes

The assignment procedure allows us to directly compare the overall probability that subjects share links they were or were not exposed to on the feed. The causal effect of exposure via the Facebook feed on sharing is simply the expected probability of sharing in the *feed* condition minus the expected probability in the *no feed* condition. This quantity, known as the average treatment effect on the treated (or alternatively, the absolute risk increase), can vary when conditioning on other variables, including the number of friends

and tie strength, which are analyzed in Sections 4 and 5. Alternatively, the difference in probabilities can be viewed as a ratio (the relative risk ratio), which quantifies how many times more likely an individual is to share as a result of being exposed to content on the feed.

Although the assignment is completely random, subjects and URLs may differ in ways that impact our measurements. For example, certain users may be highly active on Facebook, so that they are assigned to experimental conditions more often than other users. If these users were to vary significantly in terms of their information sharing propensities, such as sharing or re-sharing greater or fewer links than others, the disproportionate inclusion of these users may bias our measurements and threaten the population validity of our findings. Similarly, very popular URLs may also introduce biases; they may be more or less likely to be re-shared because of their inherent appeal or more likely to be discovered independently of Facebook because of their relative popularity amongst friends.

To provide control for these biases, we use bootstrapped averages clustered by the subject or URL. We find that in all of our analyses, clustering by the URL rather than the subject yields nearly identical probability estimates that have marginally wider confidence intervals, so we have chosen to present our results using means and 95% confidence intervals clustered by URL. Risk ratios are obtained using the 95% bootstrapped confidence intervals of likelihood of sharing in the *feed* and *no feed* conditions. To compute the lower bound of the ratio, we divide the lower bound of the probability of sharing in the *feed* condition by the upper bound for the *no feed* condition. The upper bound of the ratio is computed by dividing the upper bound in the *feed* condition by the lower bound of the *no feed* condition. The additive analog of the same procedure is used to obtain confidence intervals for probability differences.

4. HOW EXPOSURE TO SOCIAL SIGNALS AFFECTS DIFFUSION

We find that subjects who are exposed to signals about friends' sharing behavior are several times more likely to share that same information, and share sooner than those who are not exposed. To measure the relative increase in sharing due to exposure, we compute the risk ratio: the likelihood of sharing in the *feed* condition (0.191%) divided by the likelihood of sharing in the *no feed condition* (0.025%), and find that individuals in the *feed* condition are 7.37 times more likely share (95% $CI = [7.23, 7.72]$). Although the probability of sharing upon exposure may appear small, it is important to note that individuals have hundreds of contacts online who may see their link, and that on average one out of every 12.5 URLs that are clicked on in the *feed* condition are subsequently re-shared.

4.1 Temporal Clustering

Contemporaneous behavior among connected individuals is commonly used as evidence for social influence processes (e.g. [4, 9, 8, 15, 16, 19, 20, 25, 29, 36, 43]). We find that subjects who share the same link as their friends typically do so within a time that is proximate to their friends' sharing time, even when no exposure occurs on Facebook. Figure 3 illustrates the cumulative distribution of information lags between the subject and their first sharing friend, among
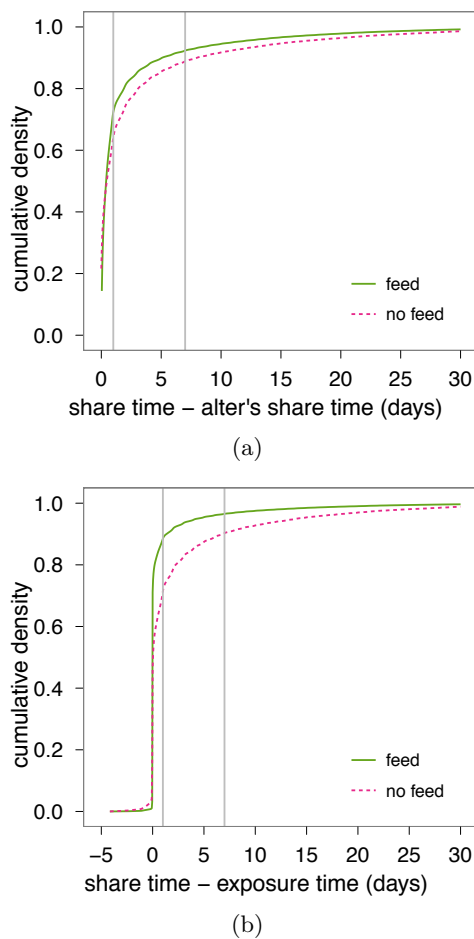
(a)

(b)

Figure 3: Temporal clustering in sharing the same link as a friend in the *feed* and *no feed* conditions. (a) The difference in sharing time between a subject and their first sharing friend. (b) The difference between the time at which a subject was first to exposed (or was to be exposed) to the link and the time at which they shared. Vertical lines indicate one day and one week.

subjects who had shared a URL after their friends. The top panel shows the latency in sharing times between the subject and their friend for users in the *feed* and *no feed* condition. While a larger proportion of users in the *feed* condition share a link within the first hour of their friends, the distribution of sharing times is strikingly similar. The bottom panel shows the differences in time between when subjects shared and when they were (or would have been) first exposed to their friends' sharing behavior on the Facebook feed. The horizontal axis is negative when a subject had shared a link after a friend but had not yet seen that link on the feed. From this comparison, it is easy to see that users in the *feed* condition are most likely to share a link immediately upon exposure, while those who share it without seeing it in their feed will do so over a slightly longer period of time.

To evaluate how exposure on the Facebook feed relates to the speed at which URLs appear to diffuse, we consider URLs that were assigned to both the *feed* and *no feed* condi-

tion. We first match the share time of each URL in the *feed* condition with a share time of the URL in the *no feed* condition, sampling URLs in proportion to their relative abundances in the data. From this set of contrasts, we find that the median sharing latency after a friend has already shared the content is 6 hours in the *feed* condition, compared to 20 hours when assigned to the *no feed* condition (Wilcoxon rank-sum test, $p < 10^{-16}$). The presence of strong temporal clustering in both experimental conditions illustrates the problem with inferring influence processes from observations of temporally proximate behavior among connected individuals: regardless of access to social signals within a particular online medium, individuals can still acquire and share the same information as their friends, albeit at a slightly later point in time.

4.2 Effect of Multiple Sharing Friends

Classic models of social and biological contagion (e.g. [23, 35]) predict that the likelihood of "infection" increases with the number of infected contacts. Observational studies of online contagion [4, 9, 15, 30] not only find evidence of temporal clustering, but also observe a similar relationship between the likelihood of contagion and the number of infected contacts. However, it is important to note that this correlation can have multiple causes that are unrelated to social influence processes. For examle, if a website is popular among friends, then a particularly interesting page is more likely to be shared by a users' friends independent of one another. The positive relationship between the number of sharing friends and likelihood of sharing may therefore simply reflect heterogeneity in the "interestingness" of the content, which is clustered along the network: the more popular a page is for a group of friends, the more likely it is that one would observe multiple friends sharing it.

We first show that, consistent with prior observational studies, the probability of sharing a link in the *feed* condition increases with the number of contacts who have already shared the link (solid line, Figure 4a). But the presence of a similar relationship in the *no feed* condition (grey line, Figure 4a) shows that an individual is more likely to exhibit the sharing behavior when multiple friends share, even if she does not necessarily observe her friends' behavior. Therefore, when using observational data, the naïve conditional probability (which is equivalent to the probability of sharing in the *feed* condition) does not directly give the probability increase due to influence via multiple sharing friends. Rather, such an estimate reflects a mixture of internal influence effects and external correlation.

Our experiment allows us to directly measure the effect of the feed relative to external factors, computed as either the difference or ratio between the probability of sharing in the *feed* and *no feed* conditions (Figure 4bc). While the difference in sharing likelihood grows with the number of sharing friends, the relative risk ratio falls. This contrast suggests that social information in the feed is most likely to influence a user to share a link that many of her friends have shared, but the relative impact of that influence is highest for content that few friends are sharing. The decreasing relative effect is consistent with the hypothesis that having multiple sharing friends is associated with greater redundancy in information exposure, which may either be caused by homophily in visitation and sharing tendencies, or external influence.
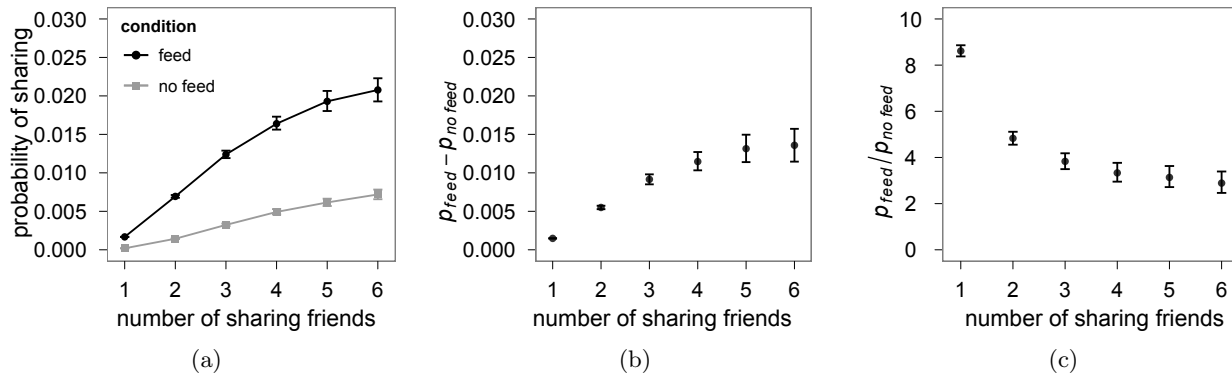
(a)   (b)   (c)

Figure 4: Users with more friends sharing a Web link are themselves more likely to share. (a) The probability of sharing for subjects that were (*feed*) and were not (*no feed*) exposed to content increases as a function of the number sharing friends. (b) The causal effect of the feed is greater when subjects have more sharing friends (c) The multiplicative impact of the feed is greatest when few friends are sharing. Error bars represent the 95% bootstrapped confidence intervals clustered on the URL.

5.   TIE STRENGTH AND INFLUENCE

Next, we examine the relationship between tie strength, influence, and information diversity by combining the experimental data with users' online and offline interactions. Following arguments originally proposed by Mark Granovetter's seminal 1973 paper, *The Strength of Weak Ties* [22], empirical work linking tie strength and diffusion often utilize the number of mutual contacts as proxies of interaction frequency. Rather than using the number of mutual contacts, which can be large for pairs of individuals who no longer communicate (e.g. former classmates), we directly measure the strength of tie between a subject and her friend in terms of four types of interactions: (i) the frequency of private online communication between the two users in the form of Facebook messages[1]; (ii) the frequency of public online interaction in the form of comments left by one user on another user's posts; (iii) the number of real-world coincidences captured on Facebook in terms of both users being labeled by users as appearing in the same photograph; and (iv) the number of online coincidences in terms of both users responding to the same Facebook post with a comment. Frequencies are computed using data from the three months directly prior to the experiment. The distribution of tie strengths among subjects and their sharing friends can be seen in Figure 5.

5.1   Effect of Tie Strength

We measure how the difference in the likelihood of sharing a URL in the *feed* versus *no feed* conditions varies according to tie strength. To simplify our estimate of the effect of tie strength, we restrict our analysis to subjects with exactly one friend who had previously shared the link. In both conditions, a subject is more likely to share a link when her
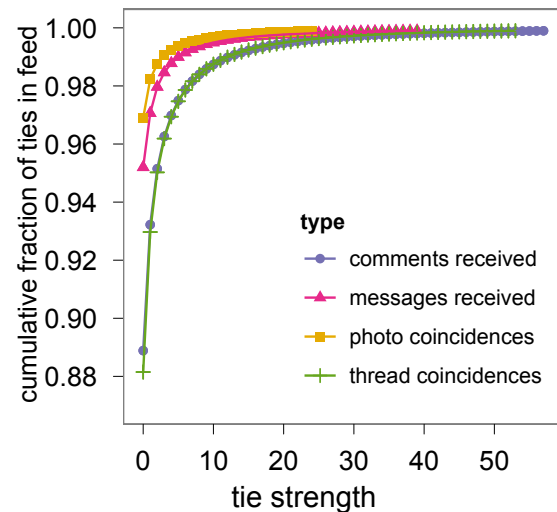

Figure 5: Tie strength distribution among friends displayed in subjects' feeds using the four measurements. Points are plotted up to the 99.9^{th} **percentile. Note that the vertical axis is collapsed.**

sharing friend is a strong tie (Figure 6a). For example, subjects who were exposed to a link shared by a friend from whom the subject received three comments are 2.83 times more likely to share than subjects exposed to a link shared by a friend from whom they received no comments. For those who were not exposed, the same comparison shows that subjects are 3.84 times more likely to share a link that was previously shared by the stronger tie. The larger effect in the *no feed* condition suggests that tie strength is a stronger predictor of externally correlated activity than it is for influence on feed. From Figure 6a, it is also clear that individuals are more likely to be influenced by their stronger

---

[1]We quantify message and comment interactions as the number of communication events the subject received from their friend. The number of messages and comments sent, and the geometric mean of communications sent and received, yielded qualitatively similar results, so we plot only the single directed measurement for the sake of clarity.
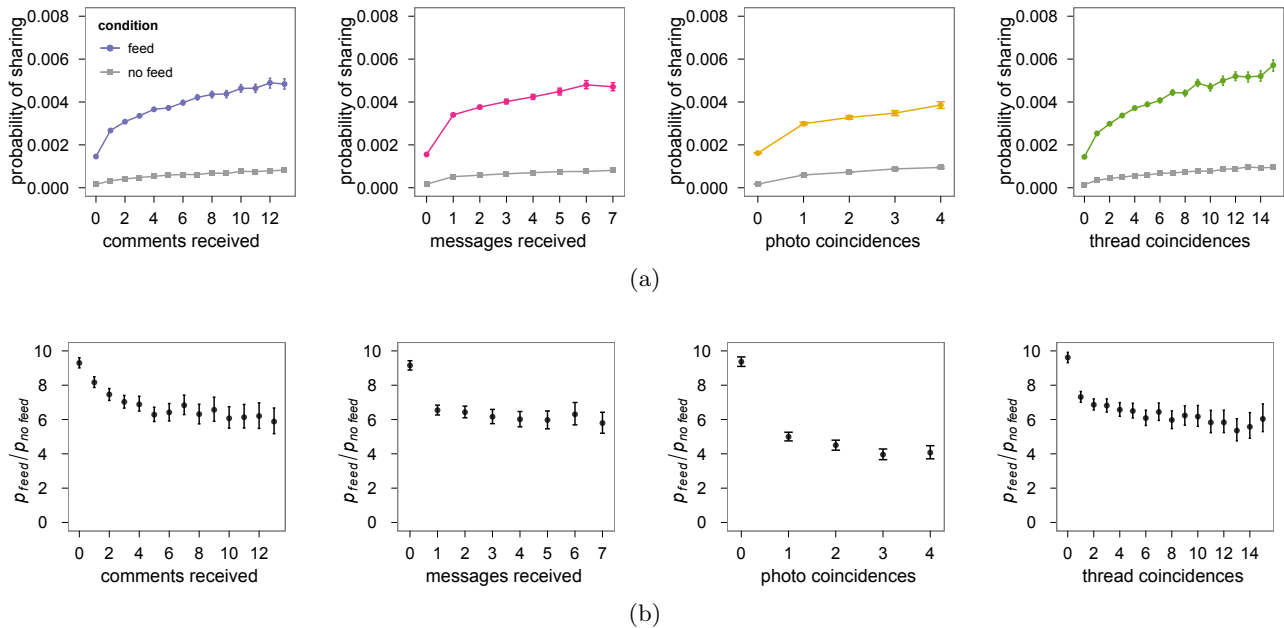
(a)

(b)

Figure 6: Strong ties are more influential, and weak ties expose friends to information they would not have otherwise shared. (a) The increasing relationship between tie strength and the probability of sharing a link that a friend shared in the *feed* and *no feed* conditions. (b) The multiplicative effect of feed diminishes with tie strength, suggesting that exposure through strong ties may be redundant with external exposure, while weak ties carry information one might otherwise not have been exposed to.

ties via the feed to share content that they would not have otherwise spread.

Furthermore, our results extend Granovetter's hypothesis that weak ties disseminate novel information into the context of media contagion. Figure 6b shows that the risk ratio of sharing between the *feed* and *no feed* conditions is highest for content shared by weak ties. This suggests that weak ties consume and transmit information that one is unlikely to be exposed to otherwise, thereby increasing the diversity of information propagated within the network.

5.2 Collective Impact of Ties

Strong ties may be individually more influential, but how much diffusion occurs in aggregate through these ties depends on the underlying distribution of tie strength (i.e. Figure 5). Using the experimental data, we can estimate the amount of contagion on the feed generated by strong and weak ties. The causal effect of exposure to information shared by friends with tie strength k is given by the average treatment effect on the treated:

$$ATET(k) = p(k, feed) - p(k, no\ feed)$$

To determine the collective impact of ties of strength k, we multiply this quantity by the fraction of links displayed in all users' feeds posted by friends of tie strength k, denoted by $f(k)$. In order to compare the impact of weak and strong ties, we must set a cutoff value for the minimum amount of interaction required between two individuals in order to consider that tie strong. Setting the cutoff at $k = 1$ (a single interaction) provides the most generous classification of strong ties while preserving some meaningful distinction between strong and weak ties, thereby giving the most influence credit to strong ties.

Under this categorization of strong and weak ties, the estimated total fraction of sharing events that can be attributed to weak and strong ties is the average treatment effect on the treated weighted by the proportion of URL exposures from each tie type:

$$T_{weak} = ATET(0) * f(0)$$

$$T_{strong} = \sum_{i=1}^{N} ATET(i) * f(i)$$

We illustrate this comparison in Figure 7, and show that by a wide margin, the majority of influence is generated by weak ties[2]. Although we have shown that strong ties are individually more influential, the effect of strong ties is not large enough to match the sheer abundance of weak ties.

6. DISCUSSION

Social networks may influence an individual's behavior, but they also reflect the individual's own activities, interests, and opinions. These commonalities make it nearly impossible to determine from observational data whether any particular interaction, mode of communication, or social environment is responsible for the apparent spread of a behavior through a network. In the context of our study, there are three possible mechanisms that may explain diffusion-like phenomena: (1) An individual shares a link on Facebook,

---

[2]Note that for the purposes of this study, it is not necessary to model the effect of tie strength for users with multiple sharing friends, since stories of this kind only constitute 4.2% of links in the newsfeed, and their inclusion would not dramatically alter the balance of aggregate influence by tie strength.
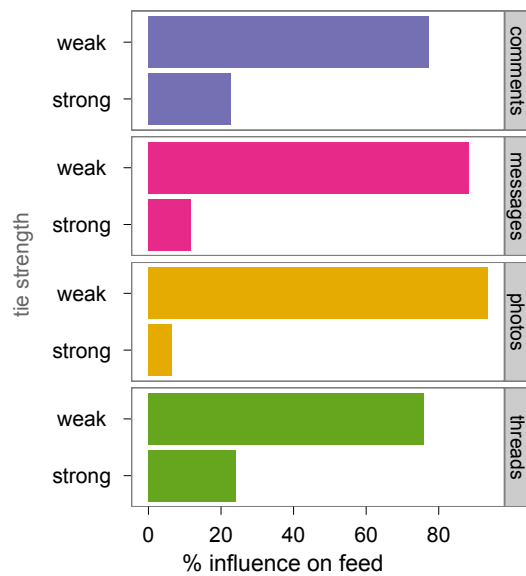
Figure 7: Weak ties are collectively more influential than strong ties. Panels show the percentage of information spread by strong and weak ties for all four measurements of tie strength. Although the probability of influence is significantly higher for those that interact frequently, most contagion occurs along weak ties, which are more abundant.

and exposure to this information on the feed causes a friend to re-share that same link. (2) Friends visit the same web page and share a link to that web page on Facebook, independently of one another. (3) An individual shares a link within and external to Facebook, and exposure to the externally shared information causes a friend to share the link on Facebook. Our experiment determines the causal effect of the feed on the spread of sharing behaviors by comparing the likelihood of sharing under the *feed* condition (possible causes 1-3) with the likelihood under the *no feed* condition (possible causes 2-3).

Our experiment generalizes Mark Granovetter's predictions about the strength of weak ties [22] to the spread of everyday information. Weak ties are argued to have access to more diverse information because they are expected to have fewer mutual contacts; each individual has access to information that the other does not. For information that is almost exclusively embedded within few individuals, like job openings or future strategic plans, weak ties play a necessarily role in facilitating information flow. This reasoning, however, does not necessarily apply to the spread of widely available information, and the relationship between tie strength and information access is not immediately obvious. Our experiment sheds light on how tie strength relates to information access within a broader context, and suggests that weak ties, defined directly in terms of interaction propensities, diffuse novel information that would not have otherwise spread.

Although weak ties can serve a critical bridging function [22, 37], the influence that weak ties exert has never before been measured empirically at a systemic level. We

find that the majority of influence results from exposure to individual weak ties, which indicates that most information diffusion on Facebook is driven by simple contagion. This stands in contrast to prior studies of influence on the adoption of products, behaviors or opinions, which center around the effect of having multiple or densely connected contacts who have adopted [6, 7, 14, 13]. Our results suggest that in large online environments, the low cost of disseminating information fosters diffusion dynamics that are different from situations where adoption is subject to positive externalities or carries a high cost.

Because we are unable to observe interactions that occur outside of Facebook, a limitation of our study is that we can only fully identify causal effects within the site. Correlated sharing in the *no feed* condition may occur because friends independently visit and share the same page as one another, or because one user is influenced to share via an external communication channel. Although we are not able to directly evaluate the relative contribution of these two potential causes, our results allow us to obtain a bound on the effect on sharing behavior within the site. The probability of sharing in the *no feed* condition, which is a combination of similarity and external influence, is an upper bound on how much sharing occurs because of homophily-related effects. Likewise, the difference in the probability of sharing within the *feed* and *no feed* condition gives a lower bound on how much on-site sharing is due to interpersonal influence along any communication medium.

The mass adoption of online social networking systems has the potential to dramatically alter an individual's exposure to new information. By applying an experimental approach to measuring diffusion outcomes within one of the largest human communication networks, we are able to rigorously quantify the effect of social networks on information spread. The present work sheds light on aggregate trends over a large population; future studies may investigate how properties of the individual, such as age, gender, and nationality, or features of content, such as popularity and breadth of appeal, relate to the influence and its confounds.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25:211–230, 2001.

[2] E. Adar and A. Adamic, Lada. Tracking information epidemics in blogspace. In *2005 IEEE/WIC/ACM International Conference on Web Intelligence*, Compiegne University of Technology, France, 2005.

[3] E. Adar, J. Teevan, and S. T. Dumais. Resonance on the web: web dynamics and revisitation patterns. In *Proceedings of the 27th International Conference on Human factors in Computing Systems*, CHI '09, pages 1381–1390, New York, NY, USA, 2009. ACM Press.

[4] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th Internal Conference on*

Knowledge Discover & Data Mining, pages 7–15, New York, NY, USA, 2008. ACM Press.

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

[6] S. Aral and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, Aug. 2011.

[7] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, New York, NY, USA, 2006. ACM.

[8] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *3rd ACM Conference on Web Search and Data Mining*, Hong Kong, 2011. ACM Press.

[9] E. Bakshy, B. Karrer, and L. Adamic. Social influence and the diffusion of user-created content. In *Proceedings of the tenth ACM conference on Electronic commerce*, pages 325–334. ACM, 2009.

[10] H. R. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer. The problem of informant accuracy: The validity of retrospective data. *Annu. Rev. Anthropol.*, 13:495–517, 1984.

[11] J. J. Brown and P. H. Reingen. Social ties and word-of-mouth referral behavior. *J. Consumer Research*, 14(3):pp. 350–362, 1987.

[12] R. S. Burt. *Structural holes: The social structure of competition.* Harvard University Press, Cambridge, MA, 1992.

[13] D. Centola. The Spread of Behavior in an Online Social Network Experiment. *Science*, 329(5996):1194–1197, September 2010.

[14] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *Am. J. Sociol.*, 113(3):702–734, Nov. 2007.

[15] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 721–730, New York, NY, USA, 2009. ACM.

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

[17] S. Fox. The social life of health information. Technical report, Pew Internet & American Life Project, 2011.

[18] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, CHI '09, pages 211–220, New York, NY, USA, 2009. ACM.

[19] M. Gladwell. *The Tipping Point: How Little Things Can Make a Big Difference.* Little Brown, New York, 2000.

[20] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 1019–1028, New York, NY, USA, 2010. ACM.

[21] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 241–250, New York, NY, USA, 2010. ACM.

[22] M. S. Granovetter. The strength of weak ties. *Am. J. Sociol.*, 78(6):1360–1380, May 1973.

[23] M. S. Granovetter. Threshold models of collective behavior. *Am. J. Sociol.*, 83(6):1420–1443, 1978.

[24] B. S. Greenberg. Person to person communication in the diffusion of news events. *Journalism Quarterly*, 41:489–494, 1964.

[25] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501. ACM, 2004.

[26] K. Hampton, L. S. Goulet, L. Rainie, and K. Purcell. Social networking sites and our lives. Technical report, Pew Internet & American Life Project, 2011.

[27] S. Hill, F. Provost, and C. Volinsky. Network-Based marketing: Identifying likely adopters via consumer networks. *Stat. Sci.*, 21(2):256–276, May 2006.

[28] G. Kossinets and D. J. Watts. Origins of homophily in an evolving social network. *Am. J. Sociol.*, 115(2):405–450, September 2009.

[29] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[30] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 228–237, New York, NY, USA, 2006. ACM.

[31] D. Liben-Nowell and J. Kleinberg. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633, 2008.

[32] C. F. Manski. Identification of endogenous social effects: The reflection problem. *Rev. Econ. Stud.*, 60(3):531–42, July 1993.

[33] A. Marin. Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks*, 26(4):289–307, Oct. 2004.

[34] M. McPherson, L. S. Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.*, 27(1):415–444, 2001.

[35] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1):016128, Jul 2002.

[36] J.-P. Onnela and F. Reed-Tsochas. Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences*, 107(43):18375–18380, 2010.

[37] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási.

Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, May 2007.

[38] K. Purcell, L. Rainie, A. Mitchell, T. Rosenstiel, and K. Olmstead. Understanding the participatory news consumer. Technical report, Pew Internet & American Life Project, 2010.

[39] C. R. Shalizi and A. C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods and Research*, 27:211–239, 2011.

[40] T. Stein, E. Chen, and K. Mangla. Facebook Immune System. In *EuroSys Social Network Systems*, 2011.

[41] E. S. Sun, I. Rosenn, C. A. Marlow, and T. M. Lento. Gesundheit! modeling contagion through facebook news feed. In *Proceedings of the 3rd Int'l AAAI Conference on Weblogs and Social Media*, San Jose, CA, 2009. AAAI.

[42] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[43] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *ACM Conference on the World Wide Web*, Hyderbad, India, 2011. ACM Press.