

Like, Comment, Repin: Exploring the Pinterest Activity Graph

ABSTRACT

We present the results of a study of the Pinterest activity graph. Pinterest is an Online Social Network (OSN) centered around the curation and sharing of visual content. We sample its activity graph, the network formed by connecting users who interact with each other and determine that it provides more information than the follow graph, formed by users following other users. We find that the Pinterest activity graph is a distinct network that partially overlaps the follow graph but contains large numbers of links not contained in the follow graph; >70% of all incoming activity per user is done by non-followers. In those areas where the two graphs do overlap, the activity graph is much sparser: on average, only 12.3% of a user's followers interact with them. We present a model of user behavior on Pinterest based on our data, which shows that following is a second-class mechanism for content discovery on Pinterest.

Author Keywords

Pinterest; Online Social Networks; Activity Graph

General Terms

Human Factors; Measurement

INTRODUCTION

Pinterest is an Online Social Network (OSN) centered around the curation and sharing of visual content. Since its inception in 2010, it has grown extremely rapidly, reaching 10 million monthly unique visitors faster than any OSN ever¹, and boasting 70 million users by July 2013. A 2013 Pew Internet Survey [10] found that 21% of all Internet users in the US use Pinterest. The site focuses on the curation and sharing of highly visual content by its users, using the metaphor of virtual 'pinboards' on which images and other media are 'pinned'. Users can follow other users, or **pinners**, to see all of their content in their own home feed. Pinterest has attracted much attention from marketers, due to its "aspirational" nature, with users using the site to find and share products and services that they would like to buy. By late 2013, Pinterest was driving 20% of all social network referrals to purchasing sites, second only to Facebook². While Pinterest's popularity

alone may make it worthy of scholarly interest, its true value as a subject of study lies in an essential difference between Pinterest and many other popular OSNs.

Online social networks such as Facebook and Cyworld are primarily social-based: friend links demonstrate a relationship between users, with any content shared just a by-product and enhancer of their relationship. Pinterest, on the other hand, is centered around content, with all activity revolving around pins. Even microblogging services like Twitter, which also involve content, are much more social than Pinterest. Users tend to tweet about themselves, and to follow other tweeters they are interested in hearing from. The same is true of Instagram; while it is also a visual image-sharing site, its structure promotes building relationships with other users through their images - in many ways, a visual-based Twitter. On Pinterest, the focus is on content rather than the content creators; pinners are encouraged to follow *boards* (see below) whose topics they are interested in, rather than users. This difference can easily be seen in the difference between the highest-ranked users on Twitter and on Pinterest. Nearly all of the top 100 users with the highest follower counts on Twitter are celebrities or news sites.³ On Pinterest, the top 10 are all "ordinary" users whose content, not fame, is what won them followers.

The subject of influence in online social networks has received a large amount of attention over the last decade or so. Researchers have gathered and analyzed data from many social networking sites to derive insights into the spread of ideas and trends through these networks (e.g.[30, 8, 3]). Another, related area of study investigates the characteristics of user behavior on OSNs to yield insights into how people interact and build networks of relationships on social networking sites. Through both bodies of research runs a common theme: human relationships, including online ones, are complex, and many of the metrics used to model both behavior and influence fall short of their goal. For instance, PageRank [21], mentions [5], and follower count [5, 21] have all been shown to be inadequate estimators of influence. Instead, in the absence of complete information about user behavior and motivations, researchers often use the rate of reposting as a rough proxy for influence [5, 21, 5]. Content that is reposted (e.g. shared on Facebook, retweeted on Twitter, or repinned on Pinterest) has been seen, reacted to, and spread to others, who will potentially react to it as well; a user's influence in this model is thus approximated by his or her reach. Despite these findings, follower count, or popularity, is still widely perceived as extremely important. Follower (or friend, on networks with undirected edges) count is often seen as synonymous with the size of a user's audience and even the amount

¹Comscore report, March 2013

²Comscore report, October 2013

³<http://www.twitaholic.com>

of influence they have; users with large numbers of followers are viewed as very influential and are often sought after by marketers to promote their products [28].

Similarly, a large body of work shows that the network formed by linking users with their followers - the *follow graph* - contains incomplete information about users and their relationships. Studies of Facebook [31], Twitter [16], and Cyworld [9] all found that the hidden network formed by interactions between users - the *activity graph* - is very different from the follow graph. Specifically, users tend to interact with only a small subset of their friends and followers [23, 16]; thus, relationship strength, and in turn, strength of influence, are difficult to predict using the follow graph alone [5, 29, 28]. While the follow graph shows the number of users who were passively presented with some content, the activity graph allows us to see how many actually *reacted* to it - either by liking or commenting on it, or more importantly, by reposting it. Nevertheless, the follow graph is still frequently used in OSN analysis to model user relationships. In fact, several previous studies of Pinterest [7, 25, 11, 26, 13, 27] all used the follow graph to varying extents.

It is clear, then, that in many OSN's, the implicit activity graph provides additional, and more accurate, information about relationships between users than the follow graph. Is this true of content-based networks as well? Pinterest users are supposed to follow pinners, as well as interact with their pins, for the same reason - interest in the content. With following transformed from an expression of connection with a user to an expression of interest in their content, it may be reasonable to expect that interaction would closely parallel follow links. Is this indeed the case? Are the activity and follow graphs on Pinterest similar? Or do content-based networks like Pinterest also have wide disparities between their follow and activity graphs? If so, what can we learn from the activity graph that cannot be derived from the follow graph? The answers to these questions have important implications for both the academic study of influence and interaction in social networks, as well as for those looking to spread products and ideas on Pinterest.

In this paper, we sample the implicit activity graph on Pinterest and compare it to the follow graph. To guide our exploration of the activity graph, we formulate the following research questions.

RQ1: How is activity distributed?

How is activity distributed across Pinterest? The distributions of likes, comments, and repins across users, boards, and pins are an important source of information about user behavior. Are some pins much more popular than others? Are some users more successful than others at inspiring activity on their boards? Do pins on the same board all get about the same number of likes?

RQ2: What is the relationship between followers and activity?

If the activity graph in Pinterest is similar to the follow graph, we would expect to find a strong correlation between degree in the activity graph (that is, activity links) and degree in the

follow graph (number of followers). Does such a correlation indeed exist? Do followers and activity follow similar patterns in their relationships with other metrics, such as a user's number of pins?

RQ3: What are the roles of followers and non-followers?

Pinterest is an 'open' network; all users can see and interact with all other users' content without having to follow them. This characteristic inspires a number of questions. What is the role of followers on Pinterest when it comes to activity? Are most followers active on their followees' boards? Do non-followers of a user ever interact with the user's pins? If so, to what extent?

RQ4: How do users find content?

How do users find content that they like? Do they stick primarily to their home feeds, and therefore only see (and potentially interact with) content from the users they follow? Or do they explore other boards as well, perhaps through the "Popular" or "Everything" feeds, the category pages, or the search box? (For definitions, see the description of Pinterest in the next section.)

The answers to these questions together create a fascinating portrait of user behavior on Pinterest.

The contributions of this work are as follows:

- We determine that the Pinterest activity graph does not closely parallel its follow graph. The activity graph is also not a proper (and small) subset of the follow graph, as is the case in Facebook [23] and Cyworld [9]. Rather, it overlaps the follow graph to some extent, but a large percentage of its edges are not present in the follow graph at all. A large amount of valuable information about user activity can therefore be derived from the activity graph that is invisible when looking only at follow relationships.
- We find that where the activity graph does overlap with the follow graph, the activity graph is significantly sparser. In particular, only a small percentage of a user's followers interact with the user's content. On average, only 12.3% of a user's followers like, comment, or repin any of their pins.
- A large majority (>70%) of the activity on a user's posts comes from users who are not their followers. This vast amount of viewing, reacting to, and most importantly, spreading (via repins) of content between users is entirely invisible when looking at only the follow graph.
- We present a model of user behavior on Pinterest derived from our data that describes the way that users find content they are interested in, and how they interact with that content and with each other. We conclude that following is a second-class mechanism for content discovery on Pinterest, a significant difference from many other OSNs.

The remainder of this paper is structured as follows. In the Background section, we provide an overview of Pinterest and define the terms we use in the rest of the paper. We then discuss related work. We devote the next section to describing our sampling and data collection methods, along with the resulting dataset. In the Data Analysis section, we report the

results obtained from our analysis of the dataset; we then discuss their implications in a section entitled Discussion. Finally, we conclude and propose some ideas for future work.

BACKGROUND

Overview of Pinterest

Pinterest describes itself as “A tool for collecting and organizing things you love.” It’s billed as a “virtual pinboard” service, where users can easily ‘pin’ digital content they find interesting or useful and share it with others. The central entity on Pinterest is the **pin**. A pin is an image or video, often accompanied by a caption. Pins can be uploaded by the user or reposted from somewhere else on the Web; these link back to the original source when clicked on. Users, or **pinners**, pin content onto their **boards** - pages, usually organized around a specific theme, where pins are laid out in an informal style reminiscent of a physical pinboard. Pinterest’s trademark layout is designed for maximum visual appeal: pins are displayed in neat rectangles of varying heights in a grid pattern that continuously loads new content as the viewer scrolls down. Clicking on a pin opens it on a separate page with more detail.

User profiles on Pinterest are pretty basic: users can add a profile picture, a brief description, and their location, as well as links to Facebook, Twitter, and/or a personal website. Also displayed are the user’s number of boards, pins, and likes, and their number of followers and following (other users the user follows). The rest of the profile page is devoted to the user’s boards. Like pins, these are laid out as small rectangles in a grid, and display several images from the board. Every board on Pinterest belongs to a category. A continuously changing sample of pins from boards in each category are reposted on the category pages, accessible from a drop-down menu on the site header. There is also a “popular” page, where popular pins from around the site are displayed, and a search box for finding pins, boards, or users matching a search term. Pinterest users can connect with other users by **following** them. New pins from followed boards or users show up in a follower’s **home feed**, which they see when they open Pinterest. Follow edges in Pinterest are directed; pinners can follow other users who do not follow them back. In addition, following someone does not require their permission. Since pinners often have boards on many different topics, users often follow only those boards which interest them [33]; board followers are counted in the user’s general follower count on his/her profile page. Users can also create and join group boards, some of which have thousands of pinners posting content to them. Each user is also allowed up to 3 secret boards, which are only visible to the owner(s). There are three types of active social interaction on Pinterest: likes, comments, and repins. According to the Pinterest help page, “Like a pin when you want to say Hey you! Neat idea!” When a user likes a pin, it is also saved to the “Likes” page in her profile, so pinners will often like pins that they want to be easily able to find later. Users can also comment on a pin; comments are displayed beneath the pin on the board. Finally, users can repin a pin to one of their own boards. Repinning is the equivalent of retweeting on Twitter or reblogging on Tumblr and is the

mechanism by which pins can “go viral”. Pinterest encourages business to use its site by providing specialized business accounts with options that help businesses better market their products on Pinterest.

Definitions

Below we define some of the terms we use.

Follow graph: a graph representation of the users in a social network where vertices represent users and the edges between users are formed by users following (directed) or friending (undirected) each other

Activity graph: a graph representation of the users in a social network where the directed edges between users represent interactions between them. Two users A and B are connected in the activity graph if there was an interaction of some sort between them. **Activity link:** an edge in the activity graph, formed by an interaction between the users at its ends.

Follow link: an edge in the follow graph, formed by a friend/follower relationship.

Pinterest activity link: a like, comment, or repin *received* by a user. Though the interaction is technically with the pin itself, it creates a link to the pin’s owner. Due to the nature of interaction on Pinterest, we focus on incoming activity in this work.

Activity rate: the total amount of activity we collect for a user, divided by the number of the user’s sampled pins. This normalization step helps discount the simple effect of larger numbers of pins yielding more activity. (See the Analysis section for a fuller discussion of the activity rate.)

RELATED WORK

Pinterest is a fairly new site, and its lack of an API has created an additional barrier to its study. A few analyses of Pinterest, however, have been published very recently. [13] attempt to determine what drives user behavior on Pinterest by calculating the contribution that various factors (such as the gender and nationality of the original pinner) have to the likelihood of a pin’s being repinned and the number of followers a user attracts. They also compare the language used by the same users on Pinterest and on Twitter and determine that there are significant differences. [12] study user behavior on Pinterest, but they confine their analysis to the static follow graph; they also study the content and categories of pins, particularly popular ones. Mittal, et. al. analyze various aspects of Pinterest, including some user characteristics, the distribution of user locations, and pin sources. They also address privacy and copyright issues and find many instances of personal data leakage and copyright violations on Pinterest. Finally, they find that they can predict gender of Pinterest users with high accuracy [26]. Gender differences in Pinterest are a popular topic of study; [27] quantify the differences in Pinterest behavior between male and female users. [7] also study gender, specifically, the types of content favored by, and degree of specialization of, the two genders. They also report that homophily - here defined as similarity in interests - has a large influence on repinning, but a smaller one on following. [18] build a model to automatically recommend boards that users

might like. [33] studied Pinterest from a quantitative perspective through user interviews. Their findings confirm that users see Pinterest as a content provider rather than a social network.

The concept of the implicit activity network in an online social network and the fact that it differs from the explicit follow network was first proposed by [9], who analyzed the topographical characteristics of both the follow and activity networks of Cyworld, a Korean OSN. They found that the one-way interaction network had a similar topology to the follow network, but the reciprocal “friends” network was quite different, more similar to known topologies of offline social networks than to the usual characteristics of OSNs. [2] had previously made a similar observation about the testimonial network on Cyworld, but did not extend their results to the concept of the activity graph in general. [31] performed a very similar analysis on Facebook, referring to the implicit network as the interaction graph. They found significant differences between the follow graph and the interaction graph, once again finding that the interaction graph displays the small-world properties typical of OSN graphs to a lesser extent than the follow graph does. The Twitter interaction graph was studied by [16], who compared “friends” (other users the user directed at least two tweets at) with declared followers and found that most users have many more followers than friends; that is, they interact closely with only a small subset of their followers. This disparity was confirmed in the case of Facebook by the Facebook Data Science team, who, with access to all of Facebook’s user data, showed that the number of active reciprocal relationships per user was much smaller than the user’s friend count [23]. In a similar vein, [5] challenged common assumptions about influence by showing that follower count was not strongly correlated with influence. [28] performed a similar analysis, creating a method for calculating influence in Twitter that accounts for the high amount of passivity among users, which makes the identification of active content forwarders essential for the propagation of information; they state specifically that follower count is not a good measure of influence.

SAMPLING

As discussed in the introduction, we concentrated on the activity graph rather than the follow graph. Instead of crawling follower edges, we followed activity links between users (though we did collect follow links as well for crawled users). As is common in OSN analysis, analyzing the entire Pinterest graph was impractical; we therefore did our analyses on a sample of the network. Since unique ids on Pinterest are text-based, random sampling was difficult; it was also undesirable in our case, since we wanted to collect as well-connected a graph as possible. Instead, we used modified Breadth-First Search (BFS) (that is, Snowball Sampling [15]) on the full graph $G(V, E)$ to collect a sample $S(V', E')$, $V' \subset V$, $E' \subset E$ beginning from several randomly chosen seeds and moving outward by selecting a random subset of edge clusters and crawling all edges in each cluster. This is accomplished by randomly choosing k boards of each crawled user, and then crawling all activity links on each board. While BFS, and by

extension snowball sampling, [17] have been shown to be biased towards high-degree nodes [19, 14, 22], we chose BFS because we wished to capture the interactions between tightly connected groups of users. This goal is well served by BFS, which excels at fully covering small regions of a graph [20]. BFS has been used for many analyses of OSNs, among them [2, 24], and [31]. Due to the extreme time complexity of gathering data from Pinterest (see the next section for details) we chose to collect a large subgraph of users by crawling a sample of k boards from each user. Here, we set $k = 5$. Likewise, we limited the number of pins collected per board to the first 300, since 90% of boards have 300 or fewer pins. On a similar OSN graph, [4] were able to estimate many graph properties using, as we do, only a random sample of k edges from each node, even with very small values of k . Due to the extreme difficulty and expense of crawling large follower lists, we limited our analysis of activity from followers (see the end of Data Analysis) to users with 10,000 followers or fewer.

DATA COLLECTION

Collecting data from Pinterest was an extremely complex challenge. Unlike most large social network services, Pinterest does not have an API for downloading user data. We therefore created a crawler to download publicly available data visible on the site itself. As of this writing, Pinterest does not have any privacy controls, so all users and their boards (with the exception of secret boards) are publicly visible. Pinterest’s trademark design, which utilizes infinite scrolling (a design technique where new content is continuously loaded as the user scrolls down), and its method of storing a single user’s data on many separate pages, makes automatic data collection difficult. Crawling a single user with just an average number of boards, pins, followers, and social interaction can require over 1700 server calls! To add to the difficulty, we found that the actual number of an entity’s followers, pins, or likes/comments/repins frequently differed, often significantly, from the number displayed on the corresponding page. This required us to use other validation methods to insure that the data collected was accurate and complete. The lack of an API also meant that some data was not available to us; for instance, temporal data on Pinterest is extremely coarse-grained (‘1 month ago’, ‘1 year ago’), so we were unable to utilize it for any comparison purposes. Shortly before we began this project, Pinterest introduced a complete redesign of the site, one part of which makes standard http-request-based crawling impossible. Instead, we used Selenium Webdriver, a browser automator that can simulate the action of scrolling a page. Browser automation is significantly slower than issuing and parsing http requests, so the change was effectively a strict rate-limiting mechanism. The redesign included many changes aimed at altering the way that users interact with the site. Our dataset, therefore, contains important information about social interaction in the ‘new Pinterest.’

Crawler Architecture

We utilize a multithreaded crawler architecture for data collection. The controller maintains a shared FIFO queue of usernames to crawl as well as a set containing all usernames

already crawled. Each thread contains a separate browser instance, which handles the task of visiting the profile page and 5 board pages for each user and downloading all available data. The next step is accessing the separate pages (described in the previous section) storing any likes and repins for each pin. Like pages display partial profile information for all users who liked the pin, and repin pages contain information about the board the pin was repinned to. User, board, pin, and like/repin/comment data is all added to a graph database (discussed below) as entities connected by relationships, as shown in the data model in Figure 1. Finally, we download all followers for each crawled user in a separate process and add them to the database as well. We ran the crawler for 5 weeks in December 2013 and January 2014 and collected the data described in Table 1.

Data

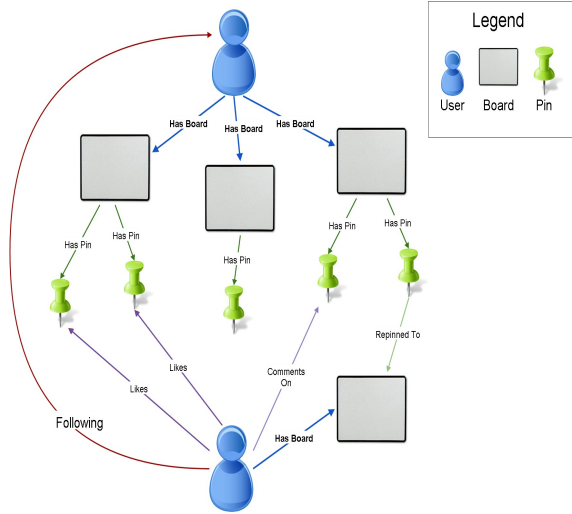


Figure 1: The Pinterest data model. Note that it is not necessary to follow a user in order to repin, like, or comment on their pins.

We collect all available data about each of these entities and their relationships. Since our data forms a network structure, it is well suited for storage in a graph database, which represents entities and the relationships between them as nodes and edges in a graph. We use the Neo4j database system for our storage, and its Cypher query language for data extraction from the graph. [1] The details of our dataset are presented in Table 1. “Partial data” includes real name, number of pins and number of followers; “users touched” includes the two categories above it, as well as users who repinned one of the pins in our dataset, for whom we have only their usernames. Some basic statistics for the data can be found in Table 2; corresponding distributions are shown in Figure 2.

DATA ANALYSIS

Activity Links (RQ1)

Our first research question concerns the distribution of the three types of activity across users, boards, and pins. Table 3 summarizes the activity links contained in our dataset.

Dataset Details	
Crawled Users (with boards and pins)	31,359
Users (partial data)	4.5 million
Total Users touched	5.4 million
Crawled Boards	150,000
Boards (partial data)	200,000
Total Boards Touched	5.1 million
Total Pins Crawled	14 million
Total Repins	7 million
Total Likes	1.56 million
Total Comments	47,557

Table 1: Data Description

	Mean	Med.	Mode	Stdev	Max
Boards per User	30.1	19	12	41	2310*
Pins per Board	138	39	1	563	100,228**
Followers per User	604	64	0	26,639	4,283,442

*All are group boards.

**This is the largest number of pins on a board with a single pinner. Group boards can have millions of pins.

Table 2: Statistics for the dataset. All mins are 0.

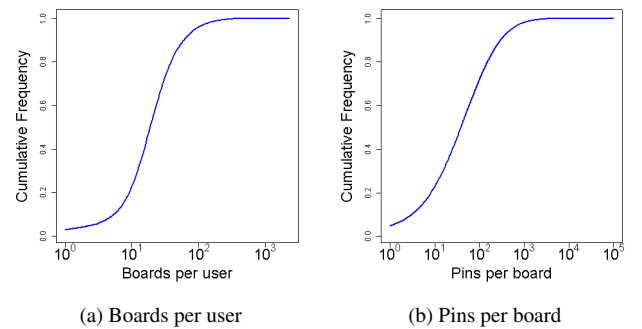


Figure 2: Cumulative frequency distributions of boards per user and pins per board, in log scale.

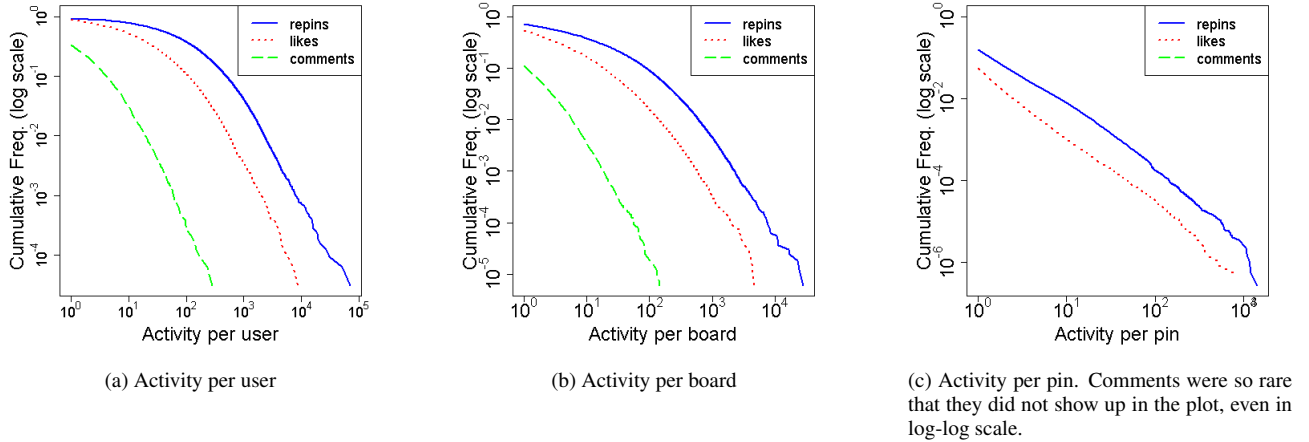


Figure 3: Complementary cumulative frequency distributions of activity per: user (3a), board (3b), and pin (3c). The distributions are plotted on a log-log scale so that they can be shown on the same plot; despite this, the number of comments per pin is so small that the distribution line disappears off the side of even a log-log plot.

While overall, there are 58 interactions (likes, comments, or

	Mean	Med.	Stdev	Max
Activity per user	265.4	67	951.7	78,336
Activity per board	51.5	5	262.6	32,640
Repins per pin	0.47	0	4.4	1424
Likes per pin	0.11	0	1.5	755
Comments per pin	N/A	0	N/A	N/A

Table 3: Statistics for activity distribution. All mins and modes are 0.

repins) for every 100 pins, only 17.7% of pins have even a single interaction, thanks to the skew in the distribution of activity per pin. Repins are by far the most common, about 4 times more common than likes and nearly 150 times more than comments. We attribute this to the nature of Pinterest itself, where the primary goal is discovering and curating content (i.e. pins), with social interaction coming in a distant second. Comments by nature involve far more social interaction than repins; repinning just means that the repinner wants the content for herself, while commenting is usually a form of communication with others.

Figure 3 shows the distributions of likes, comments, and repins across users, boards, and pins. Two things are immediately evident: the first is that repins are by far the most common type of activity, making up 81% of total activity recorded and, on average, 79% of activity on each board. Likes are the next most frequent, and comments the least. Comments, in fact, are extremely rare; just .3% of pins have even one comment. By contrast, 15.2% of the pins in our dataset were repinned at least once. The second is the shapes of the distributions. All of the distributions are very heavy-tailed, as is clear from the plots. Clearly, some users are very successful at inspiring activity on their pins, while others have next to no activity at all. This may simply be a result of some users

having many more followers than others, and therefore more activity; we investigate this possibility later in the paper. The distribution of activity per pin is even more skewed; there are a few pins that tend to be interacted with repeatedly while others appear to be ignored. These distributions, in fact, were so skewed that they had to be plotted on a log-log scale in order to be clearly visible. This skew in distribution is consistent with that observed on other social networks, such as Orkut [2], Flickr [6, 24], and Twitter [21, 5].

Activity Rate

In this work, we use activity links as edges in the activity graph. When comparing activity between users, however, or even when comparing to the follow graph, the raw activity count is not a useful metric. Different users have different numbers of pins, so a large amount of activity may just be a function of having many pins. Ideally, pinners would like to maximize their *activity rate* - that is, the amount of activity *per pin*; activity rate can be thought of as the rate of return on the investment of pinning. Activity (specifically reposting) per message is used as a measure of influence on Twitter by [29], and on Pinterest by [13]. We therefore use activity rate throughout this paper instead of raw activity counts. As would be expected from the activity distributions shown above, activity rates vary widely. Some users have large amounts of activity per pin, while others have next to none: the minimum activity rate is 0, while the maximum activity rate belongs to the Pinterest account of the popular children’s retailer Carter’s - the brand averages 105 likes, repins, or comments per pin. The distribution of activity rate is shown in Figure 4.

Followers and Activity (RQ2)

We next address RQ2, the relationship between number of followers and activity rate. One might expect there to be a strong correlation; the more people see a pin, the higher the

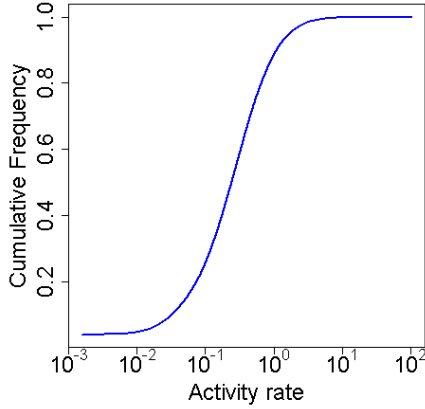


Figure 4: The cumulative distribution of activity rate per user.

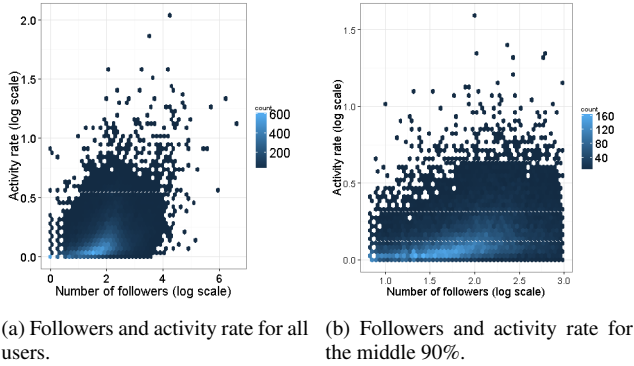


Figure 5: Followers and Activity for: 5a) all users; 5b) the middle 85% of users by follower count.

likelihood that some will interact with it. A strong relationship between the two would also be predicted if the follow and activity graphs were similar; in that case, edges in the two graphs would parallel each other closely. We find, however, that this is not the case. Figure 5a shows the number of followers per user plotted against the user’s activity rate, both on a log scale so differences are clearly visible; each hexagon represents multiple points, as shown on the key to the right. The correlation between the number of followers and the activity rate is moderate (Spearman’s $\rho = .55$). On closer investigation, however, we find that this correlation is heavily influenced by outliers; when we calculate the follower-activity correlation for only those users between the 10th and 95th percentiles by follower count (5 - 949 followers), the correlation for these “average users” is much weaker - just $\rho = .44$. The lack of visible relationship is evident on the scatterplot. (Figure 5b) Clearly, then, edges in the follow graph (followers) and activity graph (activity rate) are not similar, even in number.

There is also an interesting ‘bump’ at the y axis, consisting of a surprisingly large number of users who have no followers at all but nevertheless have activity on their pins. These

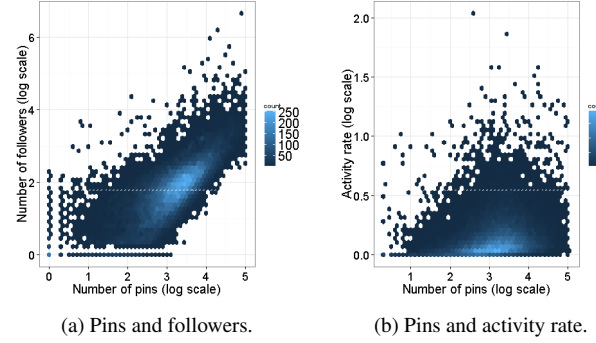


Figure 6: Number of pins (x axis) plotted against edges in the follow graph (6a) and the activity graph (6b). (All axes are log scale.)

users make up 1.4% of our dataset and strengthen the case for the importance of the activity graph. These are users who would have been dead ends in the follow graph, but they have a reasonable (and in some cases, large) amount of influence on other users, as is clear from the amount of interaction their pins received. Conversely, there are significant numbers of users all along the x axis who have little to no activity but do have followers - in many cases, large numbers of them. Approximately 3.6% of the users in our sample have followers but no activity. These users may have been considered connected when only the follow graph is taken into account, but the activity graph shows that they are not very successful at inspiring interaction and reposting of their content.

Pins, Followers, and Activity Links

To illustrate the additional information yielded by a careful study of the activity graph, we turn to an examination of a user’s pin count. Figure 6a plots number of pins against follower count for each user in our sample. There is a strong relationship between the two, with a Spearman’s correlation coefficient of 0.78. From the follow graph alone, it would seem that simply posting more pins increases a user’s influence. The activity graph, however, tells a different story. We measure influence in the activity graph by the **activity rate**, Figure 6b shows the activity rate plotted against the number of pins, using the same methodology as Figure 6a. There is not much of a relationship; Spearman’s ρ is just 0.32.

It would seem, then, that when measuring influence as number of followers, increasing the number of pins does increase follower count. It may be that people with more followers are more motivated to post, or conversely, that users with more pins garner more followers. The lack of correlation with amount of activity, however, is very interesting. This would seem to contradict [32]’s finding of a positive feedback loop between feedback on posts and posting frequency. Do pinners see only follower count as a positive feedback measure and ignore activity? These questions require much more study to answer, but they are extremely intriguing. Similar results have actually been reported for Twitter: [29] found a linear relationship between retweets and followers, but little correlation between numbers of tweets and retweets. For a direct

comparison, we correlated just the repin rate with the number of pins. Our results are similar to what Suh, et. al. report for retweets: pins and repins were barely correlated, ($\rho = 0.34$), while the number of pins and followers have a strong relationship, as reported above.

Active Followers - RQ3

In the next two subsections, we discuss RQ3 and RQ4: the roles of followers and non-followers, and how users find interesting content. One of the main arguments for the superiority of activity graphs over follow graphs as a measure of actual relationships is that only a small percentage of users' followers actively interact with them. This has been shown for Facebook, Twitter, and Cyworld (see the related work section). We find that this is true of Pinterest as well: on average, only 12.3% of a user's followers have ever engaged in a single interaction (like, comment, or repin), with any of the user's pins. The distributions for all users with any activity on the boards we crawled is shown in Figure 7. Likes and comments once again have a more uneven distribution than repins; so not only do those users who interact comment and like much less than they repin, there are also fewer users who engage in either of the two activities than who repin. Even repins, however, are only done by a few followers: very few users have repins from more than 20% of their followers.

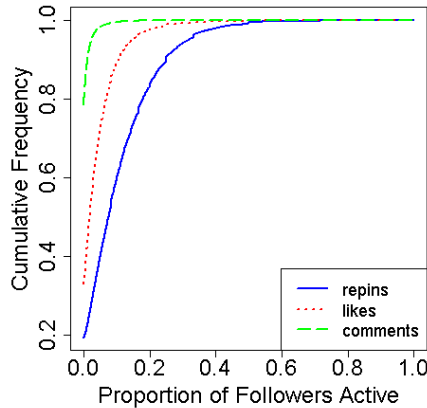


Figure 7: Proportion of a user's followers who engaged in activity on the user's boards, by type of activity.

Activity by Followers and Non-followers (RQ3 & RQ4)

Unlike Facebook and similar OSNs, Pinterest is an open network - that is, users are not limited to interacting with their friends⁴. Any user can like, comment on, or repin any other user's pins, without having to follow that user. In this sense, Pinterest is similar to Twitter, where anyone can see, respond to, and retweet anyone else's tweets. Pinterest is even more open than Twitter, since Twitter allows users to make their tweets visible to only their friends; this allows for the creation of small, closed circles of friends who use Twitter to foster their relationship rather than solely to share content.

⁴On Facebook, any user can comment on public posts, but the main interaction takes place between friends.

On Pinterest, all boards are publicly visible; secret boards are limited to just three and cannot even be viewed by followers⁵. In this manner, Pinterest emphasizes its goals of content curation and sharing over social interaction and creating relationships. We are interested in finding out whether this capability is utilized by Pinterest users. Do they interact with users whom they don't follow? If so, how often? What percentage of activity does non-follower activity represent? The answers to these questions should shed some light on RQ4, how users find content they like. Activity can be viewed as traces that users leave behind when they view content and can therefore be used to trace user viewing patterns that would otherwise be invisible.

Our data shows that users seem to be getting a large amount of the content they view from sources other than the users they follow. For the majority of users, only a small percentage of the interaction on their boards is from their followers - the median is just 24%. The remaining 76% are "drive-bys" - likes, comments, and repins done by users who do not want to be shown all of the user's content (or even one of their boards), but are interested in individual pins enough to interact with them. To ascertain whether the percentages are skewed by a few non-followers engaged in a large amount of activity, we extracted the number of *unique* non-followers who interacted with each user's content. The median is 34, showing that most users have a significant number of non-followers interacting with their content. These non-following interactors make up a full 78%, on average, of all unique users interacting with a user's pins. While these non-followers together make up the majority of activity on an average user's boards, each one does tend to interact less than those followers who do interact. The average number of interactions done by each active non-follower is just 1.4; the average active follower generates 3.4 interactions, almost two-and-a-half times as much. Figure 8 shows the distribution of percentage of repins, likes, and comments done by followers, for each user with at least one instance of the corresponding activity on their crawled boards.

DISCUSSION

Our findings paint an intriguing picture of user behavior on Pinterest. In particular, they call into question the importance of the following relationship. Most users follow a large number of other users (the median is 106), thereby bringing those users' (or boards, in the case of board following) content into their own home feeds. The lack of correlation between number of pins and activity rate, however, coupled with the fact that only a small percentage of a user's followers react to the user's pins, suggests that most users only passively consume a large majority of the content in their feeds. They may see the pins and not react to them, or they may not see them at all. Instead, they interact heavily with content from sources other than their home feeds, as shown by the high percentage of activity by non-followers. Lacking site usage data, we cannot say conclusively what these sources are, but qualitative data from Zarro, et al [33] shows that users use the site's built-in search to find pins they are interested in. They also use the

⁵Unless they are invited to contribute.

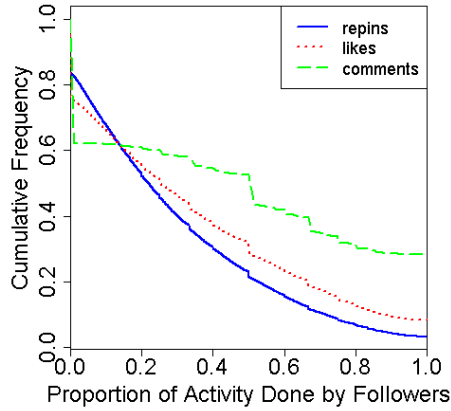


Figure 8: Proportion of the activity on each user’s boards done by the user’s followers.

category pages, each of which display a selection of pins in that category from all over Pinterest. Interestingly, users do not seem interested in following the pinners of a lot of the content they view; the fact that non-followers each perform only half as many actions as followers suggests that when users repin, like, or comment on pins from places other than their home feeds, they don’t go on to follow the pinners of those pins.

The above findings suggest that Pinterest is fundamentally different from other, social-based social networks. Not only are social interactions like comments minimized, even following - the glue that binds other social networks together - is underutilized as a method of content acquisition. Users seem to be using Pinterest as a visual mini-Internet: a place to surf, search, and find useful information; the following relationship and the home feed that it creates is de-emphasized. On Pinterest, content is king - no matter where it is found.

CONCLUSION AND FUTURE WORK

In this paper, we study social interaction on Pinterest using its activity graph, the network formed by activity links (repins, likes, and comments) between users, and compare it to the static follow graph. We find that the distributions of nearly everything on Pinterest - boards per user, pins per board, activity per pin, and many others as well, follow power laws, similar to those found repeatedly for other OSNs. We find that a user’s incoming interaction rate is not well-correlated with their number of followers, nor with their number of pins. We also discover that the majority of repins, likes, and comments on a user’s pins are not done by their followers, and that only a small percentage of a user’s followers interact with the user’s content. These results show that there is a large amount of information about user interactions available in the activity graph that is not visible in the follow graph. Along with our other findings, they provide strong support for carefully examining assumptions about number of followers being an accurate measure of influence on a social curation site. Given some of our findings, these assumptions may not

be correct. In this paper, we have mostly discussed characteristics of user interaction that can be learned from the activity graph. In the future, we would like to collect a larger, more fully connected dataset and examine various graph properties of the activity graph itself. We hope that this analysis will shed more light on Pinterest and the unique dynamics of user activity on a content-centered social network site. In particular, we are interested in the study of influence on Pinterest. Influence in social networks is an important topic of study because of the opportunities it presents for those who want to maximize the spread of ideas or advertising. Given the findings we present in this paper, we have reason to believe that the mechanisms of influence on Pinterest are both similar and different to those in other OSNs, and we would like to explore them in greater detail.

REFERENCES

1. Neo4j the world’s leading graph database.
2. Ahn, Y.-Y., Han, S., Kwak, H., Moon, S., and Jeong, H. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*, ACM (2007), 835–844.
3. Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. Everyone’s an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, ACM (2011), 65–74.
4. Bonneau, J., Anderson, J., Anderson, R., and Stajano, F. Eight friends are enough: social graph approximation via public listings. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, ACM (2009), 13–18.
5. Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, P. K. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of ICWSM*, vol. 10 (2010), 10–17.
6. Cha, M., Mislove, A., and Gummadi, K. P. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*, ACM (2009), 721–730.
7. Chang, S., Kumar, V., Gilbert, E., and Terveen, L. Specialization, homophily, and gender in a social curation site: Findings from pinterest.
8. Chen, W., Wang, Y., and Yang, S. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2009), 199–208.
9. Chun, H., Kwak, H., Eom, Y.-H., Ahn, Y.-Y., Moon, S., and Jeong, H. Comparison of online social relations in volume vs interaction: a case study of cyworld. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, ACM (2008), 57–70.

10. Duggan, M., and Smith, A. Social media update 2013.
11. Feng, Z., Cong, F., Chen, K., and Yu, Y. An empirical study of user behaviors on pinterest social network. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on*, vol. 1, IEEE (2013), 402–409.
12. Feng, Z., Cong, F., Chen, K., and Yu, Y. An empirical study of user behaviors on pinterest social network. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on*, vol. 1, IEEE (2013), 402–409.
13. Gilbert, E., Bakhshi, S., Chang, S., and Terveen, L. I need to try this?: a statistical overview of pinterest. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 2427–2436.
14. Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. Walking in facebook: A case study of unbiased sampling of osns. In *Proceedings of INFOCOM*, IEEE (2010), 1–9.
15. Goodman, L. A. Snowball sampling. *The Annals of Mathematical Statistics* 32, 1 (1961), 148–170.
16. Huberman, B., Romero, D. M., and Wu, F. Social networks that matter: Twitter under the microscope. *First Monday* 14, 1 (2008).
17. Illenberger, J., Flötteröd, G., and Nagel, K. An approach to correct biases induced by snowball sampling. 08–16.
18. Kamath, K. Y., Popescu, A.-M., and Caverlee, J. Board recommendation in pinterest.
19. Kurant, M., Markopoulou, A., and Thiran, P. On the bias of bfs (breadth first search). In *Proceedings of ITC*, IEEE (2010), 1–8.
20. Kurant, M., Markopoulou, A., and Thiran, P. Towards unbiased bfs sampling. *Selected Areas in Communications, IEEE Journal on* 29, 9 (2011), 1799–1809.
21. Kwak, H., Lee, C., Park, H., and Moon, S. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, ACM (2010), 591–600.
22. Lee, S. H., Kim, P.-J., and Jeong, H. Statistical properties of sampled networks. *Physical Review E* 73, 1 (2006), 016102.
23. Marlow, C., Byron, L., Lento, T., and Rosenn, I. Maintained relationships on facebook.
24. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, ACM (2007), 29–42.
25. Mittal, S., Gupta, N., Dewan, P., and Kumaraguru, P. Pinned it! a large scale study of the pinterest network.
26. Mittal, S., Gupta, N., Dewan, P., and Kumaraguru, P. The pin-bang theory: Discovering the pinterest world. *arXiv preprint arXiv:1307.4952* (2013).
27. Ottoni, R., Pesce, J. P., Las Casas, D., Franciscani, G., Kumaraguru, P., and Almeida, V. Ladies first: Analyzing gender roles and behaviors in pinterest. *Proceedings of ICWSM* (2013).
28. Romero, D. M., Galuba, W., Asur, S., and Huberman, B. A. Influence and passivity in social media. In *Machine learning and knowledge discovery in databases*. Springer, 2011, 18–33.
29. Suh, B., Hong, L., Pirolli, P., and Chi, E. H. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proceedings of SocialCom*, IEEE (2010), 177–184.
30. Ver Steeg, G., and Galstyan, A. Information transfer in social media. In *Proceedings of the 21st international conference on World Wide Web*, ACM (2012), 509–518.
31. Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., and Zhao, B. Y. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, Acm (2009), 205–218.
32. Wu, F., Wilkinson, D. M., and Huberman, B. A. Feedback loops of attention in peer production. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4, IEEE (2009), 409–415.
33. Zarro, M., Hall, C., and Forte, A. Wedding dresses and wanted criminals: Pinterest. com as an infrastructure for repository building. In *Proceedings of ICWSM* (2013).