

Look Who I Found: Understanding the Effects of Sharing Curated Friend Groups

Lujun Fang
University of Michigan
2260 Hayward Avenue
Ann Arbor, MI 48109
ljfang@umich.edu

Alex Fabrikant
Google Research
1600 Amphitheatre Pkwy
Mountain View, CA 94043
fabrikant@google.com

Kristen LeFevre
Google Research
1600 Amphitheatre Pkwy
Mountain View, CA 94043
klefevre@google.com

ABSTRACT

Online social networks like Google+, Twitter, and Facebook allow users to build, organize, and manage their social connections for the purposes of information sharing and consumption. Nonetheless, most social network users still report that building and curating contact groups is a time consuming burden. To help users overcome the burdens of contact discovery and grouping, Google+ recently launched a new feature known as “circle sharing.” The feature makes it easy for users to share the benefits of their own contact curation by sharing entire “circles” (contact groups) with others. Recipients of a shared circle can adopt the circle as a whole, merge the circle into one of their own circles, or select specific members of the circle to add.

In this paper, we investigate the impact that circle-sharing has had on the growth and structure of the Google+ social network. Using a cluster analysis, we identify two natural categories of shared circles, which represent two qualitatively different use cases: circles comprised primarily of celebrities (*celebrity circles*), and circles comprised of members of a community (*community circles*). We observe that exposure to circle-sharing accelerates the rate at which a user adds others to his or her circles. More specifically, we notice that circle-sharing has accelerated the “densification” rate of community circles, and also that it has disproportionately affected users with few connections, allowing them to find new contacts at a faster rate than would be expected based on accepted models of network growth. Finally, we identify features that can be used to predict which of a user’s circles (s)he is most likely to share, thus demonstrating that it is feasible to suggest to a user which circles to share with friends.

Author Keywords

Social Network, Google+, Circle Sharing

ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebSci 2012, June 22–24, 2012, Evanston, Illinois, USA.
Copyright 2012 ACM 978-1-4503-1228-8...\$10.00.

General Terms

Experimentation.

INTRODUCTION

Every day, hundreds of millions of users enjoy sharing and consuming information using online social network sites. At the same time, it can be difficult for users to discover new contacts and to maintain contact groupings (e.g., Google+ circles, Facebook friend lists)[21, 7].

Most contact management solutions focus on only one of these two tasks. A significant amount of research focuses on link prediction, which can be used to recommend new contacts to social network users [17, 14]. These recommendations are often made based on the user’s existing connections, which means that they are less accurate for new users (the “cold-start” problem). Moreover, link prediction algorithms usually generate one recommendation at a time. On the other hand, contact grouping is notoriously difficult for users [21]. A number of data mining and machine learning approaches have been proposed and built to automatically group contacts [1, 4, 10], but none of them generates satisfactory user groups without user involvement. Further, existing tools typically cannot detect real-life communities until many of the community’s interconnections are already captured in the online system [13]. As a result, new users and users of nascent social networks are often forced to manually curate and populate lists to capture the natural groupings among their contacts.

In September 2011, Google+ launched a “circle-sharing” tool, which allows users to share their individual circles (i.e., contact groups) with other users [3]. A screenshot of the circle-sharing tool is shown in Figure 1. Recipients of a shared circle can copy the circle as-is, merge the circle into one of their existing circles, or cherry-pick people from the circle to add to their own circles.

In this paper, we provide a large-scale data-driven examination of the impact that circle-sharing has had on the Google+ social network, including a characterization of the usage patterns that have driven this impact. Our main contributions are the following:

- We observe that shared circles can be categorized into two distinct types: communities and celebrities. Based

Share circle

Share a copy of this circle for others to add to their circles.

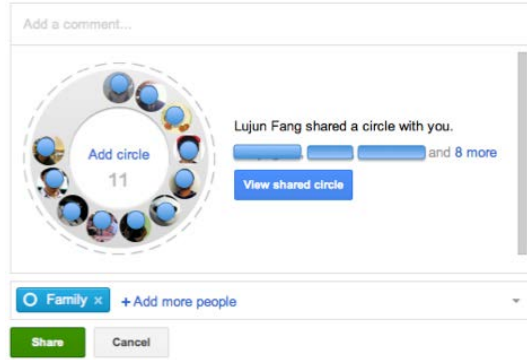


Figure 1. Screenshot of the circle-sharing tool.

on structural features of the circles themselves, we use clustering techniques to discover two predominant clusters of shared circles, which correspond to intuitive and qualitatively different use cases. Circles in the first large cluster (“communities”) are characterized by high within-circle link density, high link reciprocity with the circle owner, and relatively low popularity among circle members. Circles in the second large cluster (“celebrities”) are characterized by low within-circle link density, low link reciprocity with the circle owner, and very high popularity among circle members.

- **We provide the first large-scale study of the impact of contact-group sharing on the structure and growth of a social network.** Past research (e.g., [15]) has observed that the features and prevailing use cases of a social networking site can have a substantial effect on the growth patterns and structure of the resulting network graph. Our results demonstrate that the circle-sharing feature accelerates the “densification” of community-type circles. We also observe that circle-sharing alleviates the “cold start” problem of link prediction; if circle-sharing is prevalent in a user’s social neighborhood, this allows low-degree users to discover new contacts at a much faster rate than would be expected based on accepted models of network growth.
- **We demonstrate the feasibility of algorithmically recommending circles that a user should share.** We identify features that can differentiate shared circles from “ordinary” circles (i.e., those created by users for personal use, but never shared with others). In particular, we show that shared circles are more “commonly useful” than ordinary circles. Using this characterization, we can recommend circles that are good candidates for sharing.

RELATED WORK AND BACKGROUND

Fully 65% of online adults are using social networking sites [6], and Facebook alone has over 800 million active users [2]. One of the prevailing purposes of a social networking site is to allow users to add and group their contacts for the purpose of information sharing and consumption. Almost all major social networking sites provide tools to help users find and group contacts (e.g., Google+ circles, Facebook user lists,

Twitter lists, and friend suggestion tools provided by each of these sites). At the same time, finding and organizing one’s contacts on a social networking site are still difficult tasks, largely due to the complex and faceted nature of users’ on-line social spheres [20, 12].

A large body of prior work has focused on identifying and recommending potential contacts for social network users; most existing techniques involve viewing the social network as a graph (i.e., users as nodes and connections between users as edges) and recommending new edges in the graph based on existing edges in the graph [17, 14]. Such recommenders usually do not capture the underlying relationships between the recommendations. For example, although a recommender may find some of Alice’s high school friends, it could not group them together and recommend the group as a whole. There are indeed some “group recommendation” algorithms [9], however they view group memberships as features of social network users, and make recommendations about which groups to join, rather than recommendations of adding a group of users as contacts.

The other limitation of such recommenders is that they fail to provide good recommendations for users who have few existing connections. The recommenders are very dependent on the target user’s existing connections. Therefore, it is often difficult for new users to find contacts. However, the “cold start” problem of new social networking users is not solely because of the ineffectiveness of contact recommenders. Social network researchers have established theoretically [8, 19] and experimentally [18] “preferential attachment” of social network edge creation process: new edges are more likely to be connected to users of large degrees than those of small degrees. For example, under the BA model of network growth [8], a social networking user with 100 contacts is 10 times more likely to add another contact before a social networking user with 10 contacts. Google+ circle sharing tools help high-degree users to share their connections with low-degree users, potentially alleviating the cold start problem for new (low-degree) users.

There is also a body of literature about automatic group-creation algorithms, which can be used to assist users with grouping contacts [1, 4, 10]. Unfortunately, each of these techniques requires user involvement to create final groupings, and the group creation process is isolated from membership suggestion. User list creation through crowdsourcing is also a possible solution if the members in the lists are all public figures [5]. However, this technique is less applicable to personalized local communities (e.g., families). In contrast, as we will demonstrate in our paper, the Google+ circle sharing tool can be successfully used for both “celebrity circles” (circles containing popular and public figures) and “community circles” (circles containing members of a local community or group).

Finally, past research has observed that the network structure articulated by users of an online social network is often influenced by the features of the social network service and predominant use cases. For example, Kwak et al. observed

that the structure of the Twitter network is qualitatively different from other social networks, likely due to prevailing use cases (celebrity following and news consumption) [15]. Similarly, we observe that the Google+ circle-sharing feature has had a quantifiable impact on network growth and structure.

Google+ Circle Sharing Feature

In Google+, a user can create circles reflecting different facets in her social life. Each Google+ user has four default circles: friends, family, acquaintances and following. A user can also create other circles to describe other aspects of her life. If a user U_A puts another user U_B into any of her circles, then we say that U_A is *following* U_B . Connections on Google+ can be asymmetric (i.e., U_A is following U_B does not imply that U_B is following back U_A).

The circle sharing feature launched in Google+ in September 2011. This feature allows users to share their circles with other users. A user can choose to share any of her circles, and she can choose with whom she wants to share the circles. When a user notices that another user has shared a circle with her, she can decide to add some or all of the members in the shared circle as her own contacts. She can either add those members to one of her existing circles, or create a new circle for them.

OVERVIEW OF THE ANALYSES

The analyses presented in this paper are intended to answer three key questions: (1) Are there different types of shared circles, and how can we identify them? (2) What is the impact of circle-sharing on the structure and growth of the Google+ social network?, and (3) Can we recommend to users which of their circles are suitable to be shared?

Data Overview

All of our analyses are performed based on a large anonymized sample of Google+ circles and their adjacent edges. For each circle, we use identities of the person who shared it, the members of the shared circle, and the time of the circle share. We also use times when each node (member) joined Google+, and the circle membership edges in the network at the time of the study, along with the times the edges were created. All the user and circle IDs involved were then anonymized, and all other information on node and circle identities was scrubbed from the dataset before the study.

For each different analysis, we sampled a subset of these circles according to the requirements of the analysis; details of the sampling are provided for each analysis. All analyses are based on at least 5,000 circles.

Analysis Road Map

In order to understand the impact of circle-sharing, it is first important to understand how people are utilizing the circle-sharing feature (i.e., which circles they are sharing). We start by describing a clustering analysis. The analysis discovers two large categories of shared circles: “communities” and “celebrities.” Both categories of shared circles play an important role in the latter analyses.

Then, we move on to study the effect that circle sharing has had on the growth and structure of the Google+ social graph. Using aggregated statistics about edge creation times, we demonstrate that sharing both types of circles accelerates the growth of the social network. We also observe that circle-sharing accelerates the densification of community-type circles.

Finally, we develop a model to distinguish shared circles from ordinary circles (i.e., circles that are not shared). One possible use for such a model is to recommend to users which of their circles are good candidates for sharing. We identify a feature called *commonality* which is predictive of a circle being shared. Using commonality as well as some other features, we investigate the feasibility of classifying circles as “shared” or “not shared.” We observe that sharing of community circles is more easily predicted than sharing of celebrity circles.

CATEGORIZING SHARED CIRCLES

In this section, we describe a cluster analysis with the goal of identifying different types of shared circles. Based on our analysis, we identify two large clusters of shared circles: those that contain primarily *celebrities*, and those that contain *communities*, or groups of people who are socially interconnected.

Methodology

The shared-circle cluster analysis is based on a random sample of 9000 shared circles with size ≥ 10 . We use standard clustering techniques to group these circles on the basis of several key features. Some of the features (e.g., density) can be derived from understood features of communities in symmetric social networks [16, 13], while some features (e.g., reciprocity, popularity) are unique to asymmetric networks.

Recall that Google+ connections can be asymmetric. Intuitively, the members of some of a user’s circles (e.g., the user’s *Cousins* or *Book Club* circles) are more likely to follow the user back than the members of other circles (e.g., the user’s *Music Stars* circle). To capture the extent to which the users in a circle follow back the circle’s owner, we define the *reciprocity* feature of a circle.

DEFINITION 1. *Reciprocity* *The reciprocity of a circle is defined as the proportion of the circle members who follow back the circle owner.*

EXAMPLE 1. *Figure 2 describes a social network of 4 users: Alice, Bob, Claire, and Dan. Suppose that users have the circles shown, with an edge from A to B indicating that B is in some of A’s circles. Alice’s circle K contains Bob, Claire and Dan. The reciprocity of K is $1/3 = 0.33$, since among the members of K, only Bob follows Alice.*

To better understand the reciprocity feature, for all of the shared circles in our sample, we compute their reciprocities and plot the probability density function for their reciprocities (Figure 3(a)).¹ It is interesting to observe that this dis-

¹Note that probability density at a given point can be larger than 1.

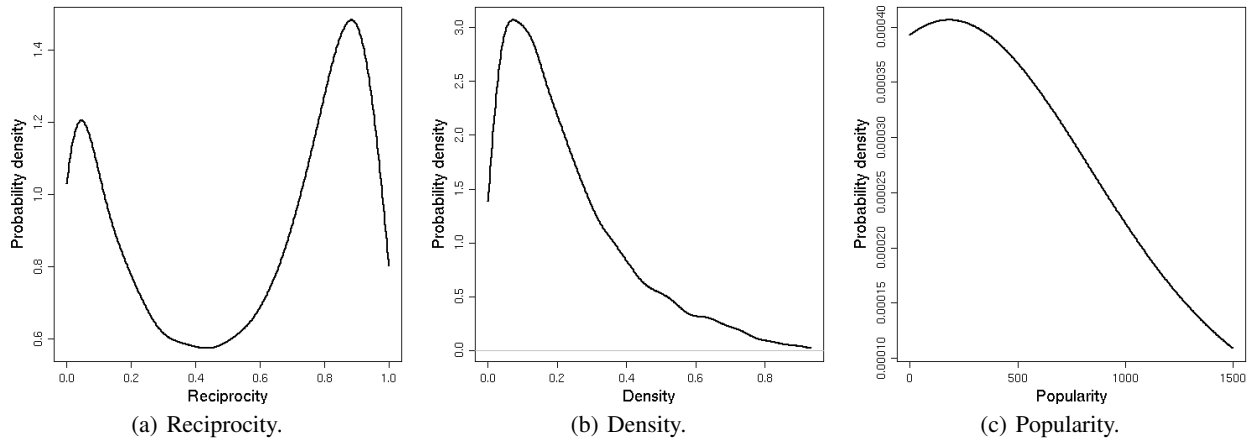


Figure 3. Probability density distributions of different circle features.

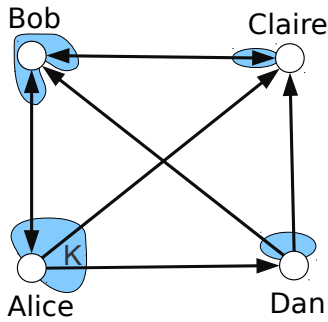


Figure 2. An example social network of 4 users. Each user has exactly one circle, and circle memberships are represented by outgoing edges.

tribution is heavily bimodal; in other words, shared circles tend to have either high or low reciprocity.

In addition to the owner, the individual members of a circle can be connected to one another. For example, we would expect members of a family circle to be well connected to one another. To capture the degree to which the members of a circle are interconnected, we define the *density* feature of a circle.

DEFINITION 2. *Density* The density of a circle is defined as the actual number of bi-directional edges between circle members divided by the maximum possible number of bi-directional edges (i.e., $\frac{n(n-1)}{2}$ if the size of the circle is n).

EXAMPLE 2. In Figure 2, among members of Alice’s circle K , there is only one bi-directional edge $Bob \leftrightarrow Claire$. The maximum possible number of such edges is $3 \cdot \frac{3-1}{2} = 3$, so the density of K is $\frac{1}{3}$.

Figure 3(b) shows the probability density function of circle density, as measured from our sample of shared circles. The function reaches its peak at 0.1, although there are indeed circles with density of 1, indicating existences of fully connected circles.

Finally, we define the *popularity* of a circle based on the number of people who are following the circle’s members.

DEFINITION 3. *Popularity* The popularity of a user is defined as the in-degree (i.e., the number of followers) of the user in the social network. The popularity of a circle is defined as median popularity of its members.

EXAMPLE 3. In Figure 2, K ’s members have popularities 3 (Bob), 3 (Claire), and 1 (Dan). The popularity of K is thus 3 (the median of $\{1, 3, 3\}$).

Note that we use *median* instead of *mean* of member popularities as the circle popularity because the distribution of individual popularity is very heavy-tailed: a few users have upward of millions of followers, but most have a modest number, which would make the mean popularity dominated by a circle’s most popular members. Figure 3(c) shows the probability density distribution of circle popularity, measured using our sample of shared circles. We observe that the function reaches its peak around 200, although there are still a significant number of circles with very high popularity (e.g., >1000).

There are undoubtedly other features (besides reciprocity, density, and popularity) that are useful for characterizing circles. Circle name is another logical feature to consider. For example, we would expect a circle named *Family* to represent a community (with high density and high reciprocity); we would expect a circle named *Following* to include a set of celebrities (with low reciprocity and high popularity). Unfortunately, in many cases, the circle name alone is insufficient. For example, a circle named *Photographer* could represent a community or a group of celebrity photographers; in order to distinguish the two cases, we would end up looking at the structure of the network graph. For these reasons, the remainder of our analysis focuses on structural features, but future work could, with appropriate privacy safeguards, incorporate semantic signals from circle names, circle-share post annotations, and more sophisticated signals of user engagement with the circles.

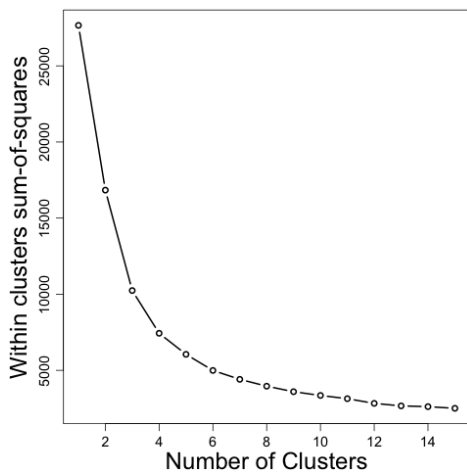


Figure 4. Within clusters sum-of-squares for different k when performing k-means circle clustering.

Circle Clustering

Using reciprocity, popularity, and density as features, we applied a standard clustering technique (k-means) to the shared circles in our sample. Of course, circles (as well as their feature values) change over time, and we use the feature values at the time when each circle was shared.

Before clustering, we pre-processed the data in two ways: (1) Because the popularity value is heavily skewed, we transform this feature by taking its log. (2) We applied the *scale()* function in R to normalize each of the features.

As a second preliminary step, we computed the within-clusters sum-of-squares for different possible values of k ($k = 2..15$), and selected $k = 4$ by visually observing the natural “knee” in the trend plot [11] of the within-cluster sum-of-squares, in Figure 4.

The result of clustering based on the processed features is shown in Figure 5. Each triple of bars represent the mean processed feature values of a circle cluster. Since the feature values are normalized, the numbers in the figure indicate a feature’s relative, rather than absolute, value. The aggregate results of real feature values (after reversing the normalization) are shown in Table 1.

The first two clusters of circles are of high reciprocity and relatively low popularity, indicating that members of those circles are most likely to be ordinary users who are friends with the circle owners, and the circles are very likely to describe real life communities like families or groups of friends. Therefore we call them “community circles”. We also notice that the circles in Cluster 1 are more dense than those in Cluster 2, which suggests that some community circles have been well-developed, while others are still nascent. These two clusters of circles combined comprise 52% of all shared circles.

In contrast, the circles in Clusters 3 and 4 are of high popularity and low reciprocity. This is in particular true for circles

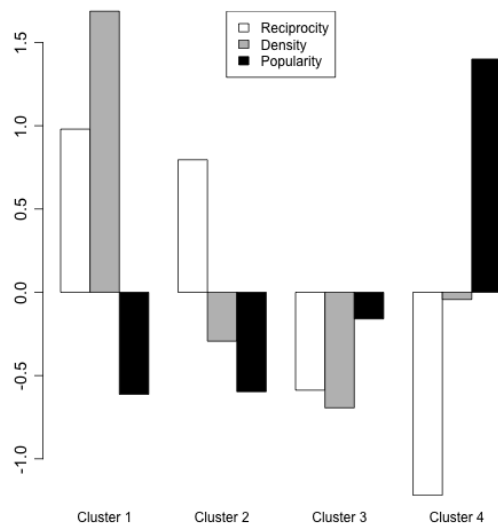


Figure 5. Shared circle clustering using k-means ($k = 4$) algorithm.

in Cluster 4; their median popularity is more than 20000, and the mean reciprocity is only 0.11. We call circles in these two clusters “celebrity” circles, since they mostly contain famous (i.e., high in-degree) people, and the connections to them are mostly single-directional. It is interesting to observe that celebrity circles, especially those in Cluster 4, have moderate densities. This suggests that some of those celebrities are connected to each other. Circles in Clusters 3 and 4 comprise the remaining 48% of all the shared circles.

Cluster ID	Reciprocity (mean)	Density (mean)	Popularity (median)
1	0.86	0.52	233
2	0.80	0.17	212
3	0.32	0.10	605
4	0.11	0.21	22561

Table 1. Aggregated statistics of circle clusters.

IMPACT OF SHARED CIRCLES

In this section, we now turn our attention to understanding the impact that the Google+ circle-sharing feature has had on the growth and structure of the network. We describe a large-scale quantitative study, the results of which are the following important observations:

- Circle-sharing events accelerate the creation of edges in the network. In particular, we find that circle-sharing events accelerate the densification of community circles. We hypothesize that circle-sharing accelerates the popularity of celebrities, but we are not able to confirm this hypothesis for reasons described in detail below.
- We find that circle-sharing disproportionately accelerates the growth of edges involving low-degree users. After being exposed to a shared circle, the degrees of low-degree users increase at a rate higher than predicted by accepted models of network growth.

- Among users who are exposed to shared circles, circle-sharing accelerates the rate at which circles are created, and the rate at which new people are added to circles.

Methodology

To understand how circle-sharing events have affected circles and users, we identify important circle- and user-related metrics (e.g., the density of a circle), and measure their values before and after the circle or user is affected by the circle-sharing feature (we will define what we mean by “affected” for each analysis). Since each circle (user) is affected by circle-sharing at a different time, to summarize the changes of multiple circles (users), we group circles (users) in our dataset into cohorts according to the week in which they are affected by circle sharing. For each cohort of circles (users), we can then measure the changes in these metrics over time to understand the effect of circle sharing.

Accelerating Edge Growth

We start by investigating whether and how circle-sharing events affect the speed at which new edges are added to the social graph. Intuitively, we expect that when a circle is shared, it will draw the attention of other users (the recipients of the shared circle) to its members. As a result, we expect that the number of people following the circle members (in-edges) will increase very quickly soon after the circle is shared.

In addition to accelerating edge growth overall, we also hypothesize that circle-sharing events will affect the network differently, depending on whether the shared circle is a community or celebrity circle. Specifically, anecdotal evidence suggests that community circles (e.g., the *Knitting Club* circle) are often shared with users who are also members of the community. Thus, we suspect that circle-sharing will contribute to the densification of the community, as members adopt the shared circle. In contrast, we expect that shared celebrity circles (e.g., the *Rock Stars* circle) will serve primarily to accelerate the popularity of circle members.

To verify these hypotheses, we use the same sample of shared circles as in the previous section, first categorizing them into community and celebrity circles, and then dividing them into cohorts based on the week during which they were shared.

Density increase of community circles In the previous section we defined circle density. However, the density of a circle at any point in time is dependent not only on the number of edges in the circle, but also on the number of members in the circle. In order to reason about changes in density due to edge growth, in the following analyses, in this section we use “density” to specifically refer to the density of edges among a circle’s members at a globally-fixed date shortly before the beginning of the study period.

For each weekly cohort of community circles, we compute their mean density over time and plot the trend. Figure 6(a) shows the density trend over time of the circle cohort $C_{\text{Nov}2}$ (i.e., circles shared during the week of November 2-8). We notice that, aside from week of November 2, the growth of

circle density is mostly linear. However, during the week when the circle sharing events happen, we notice an obvious jump in circle density. The same observation also holds for other weeks.

To better understand the density increase trend and the acceleration of density increase during the circle-sharing week, we compute the density increase for each week, and compare the weekly density increase of the circle sharing-week to that of other weeks. The weekly density increase value $\Delta D_w(c)$ of a circle c for timestamp w , expressed in weeks, is defined by:

$$\Delta D_w(c) = D_{w+1.0}(c) - D_w(c). \quad (1)$$

Based on weekly density increases, we compute the sharing-week acceleration rate R_D , which captures the amount of density increase during the week when the circle got shared, w_c (rounded to the beginning of the week), as compared to the previous week:

$$R_D(c) = \frac{\Delta D_{w_c}(c)}{\Delta D_{w_c-1.0}(c)}, \quad (2)$$

The mean R_D for all the shared circles in our sample is 2.5. In other words, the mean density acceleration is 150% during the week when the circle is shared.

Finally, we perform a one-sample t-test to see if the density increase during the circle sharing week is significantly better than other weeks. We computed the p-value for each circle cohort separately, and the density increase acceleration brought by circle-sharing was statistically significant with $p < 0.05$ for all weeks.

Impact on popularity in celebrity circles We also performed a similar analysis to test the hypothesis that circle-sharing accelerates the popularity in celebrity circles. We see anecdotal evidence that in some cases, circle-sharing events are helping celebrity circles attract a significant number of new followers. However, our analysis did not show such a growth with statistical significance.

One possible explanation is that celebrity circles usually attract hundreds or thousands of followers, while circles are often shared with smaller groups of people. Even if the circle owner shares it publicly, the impact of the action is likely mostly limited to those who follow the sharer. Thus, while the circle-sharing event may bring in new edges, the total number of new edges is likely to be small in comparison to the number of users already following the celebrities. Nonetheless, multiple shares of the same circle of celebrities can attract larger audiences, and a closer look at the impact of being included in many shared circles is an interesting topic for future research.

Structural Impact on Edge Growth

So far we have demonstrated how circle-sharing events are accelerating the network growth in term of edge additions, but we also want to see if circle-sharing events are affecting the structural properties of the network. Most social network growth exhibits a phenomenon called *preferential attach-*

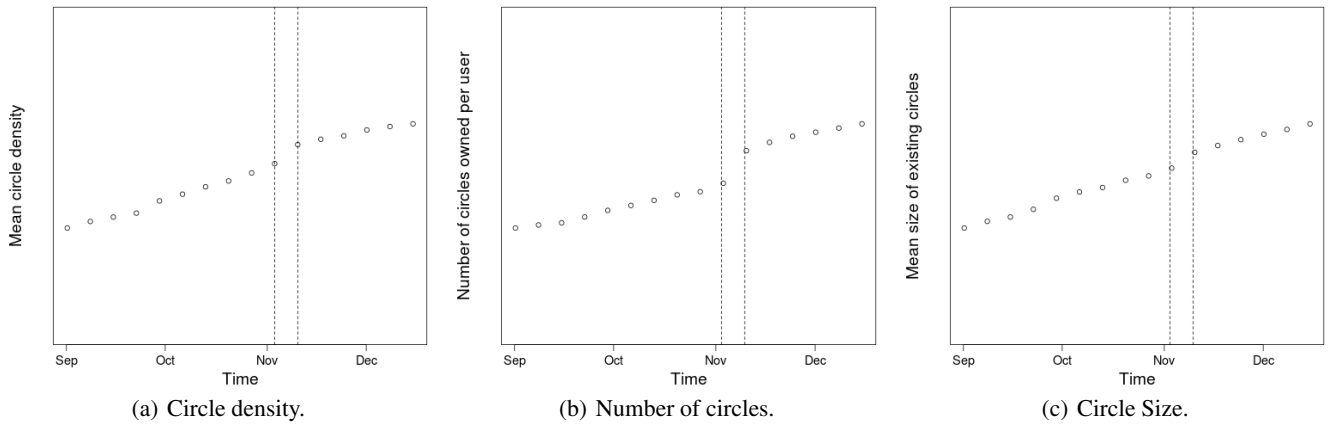


Figure 6. Mean values of various circle metrics, for users who became circle-sharing-touched (Figure 6(b) and 6(c)) or for circles got shared (Figure 6(a)) during the week of November 2–8. The beginning and end of the circle sharing week are indicated by the dashed lines. (The y-axis has been descaled to protect proprietary information.)

ment [8, 19, 18]; new edges are more likely to be connected to large-degree nodes than smaller-degree nodes. The circle-sharing feature makes it easier for both low-degree and high-degree users to discover groups of contacts, and low degree users might even benefit more since they may find more new contacts from a shared circle. Therefore we expect the difference between edge growth rates for low- and high-degree users becomes smaller as a result.

To test this hypothesis, we chose a random sample of users who were *members* of a circle that got shared, and divided them into cohorts based on the number of bidirectional edges they had before the relevant circle sharing event, and measured, for 3 example cohorts, the change in the number of new bidirectional edges created the week before the relevant circle share, and the week after. The results, with the degree change figures descaled, are shown in Table 2.

As we have seen in the previous analysis, all the users can benefit from circle sharing in terms of making new connections. However, this is particularly true for low-degree users. During the week immediately after circle-sharing events, users of degree 10 make 1.63 times more connections than they did the week before. In contrast, users of degree of 100 make 1.07 times as many as the week before. Before circles are shared, users of degree 100 add 4 times as many connections as users of degree 10. After circle-sharing events, users of degree 100 only add 2.6 times as many connections as users of degree 10. Therefore, circle sharing is indeed changing the network growth process by giving low degree users better chances to make new connections.

Degree when shared	10	50	100
Weekly link creations before share	87	195	348
Weekly link creations after share	142	252	372
Link creation acceleration ratio	1.63	1.29	1.07

Table 2. Degree of user vs. new bidirectional link creations per week before and after a circle-sharing event. (The six weekly link creation rate averages are rescaled to protect proprietary information.)

Circle Creation and Expansion of Recipients

Next we examine whether and how shared circles are adopted or used by their recipients. Upon seeing a shared circle, if the recipient decides to add some or all of the contacts in the shared circle, she has two choices: add the contacts to one of her existing circles, or create a new circle for the contacts. To verify the adoption of these two types of shared circle-adoption behaviors, we select groups of users that are recipients of shared circles and see if they are expanding their existing circles and creating new circles as a result of seeing shared circles. (Note that data about shared-circle uptake events was not available, so we had to observe these behaviors indirectly by observing changes in circle sizes and changes in the number of circles owned by a user.)

To perform the analyses, we randomly sampled 10000 users that became *circle-sharing-touched* between September and December, 2011. We say a user becomes circle-sharing-touched if the user shares a circle or is a member of a shared circle. There are other ways to define circle-sharing-touched, but our main goal is to isolate a set of users that are likely to have been recipients of a shared circle. Let $w(u)$ denote the timestamp, in weeks, of when user u was first touched by circle-sharing, rounded down to the beginning of the calendar week to define weekly cohorts.

Number of circles owned per user. We first compute the mean number of circles owned by different cohorts of users over time. If users are adopting shared circles they see and creating new circles for them, then we would expect the mean number of circles owned by users to increase faster when the users become circle-sharing-touched. For each user cohort, we compute the mean number of circles owned by the users over time; we show the trend of one example weekly cohort (those who became circle-sharing-touched during the week of November 2–8) in Figure 6(b). We see that users create more circles during the week they become circle-sharing-touched. (Similar observations can be made for other groups, but are omitted for space.)

Following the same process we used in the previous analysis for circle density, we compute the weekly increase in circle count C :

$$\Delta C_w(u) = C_{w+1.0}(c) - C_w(c),$$

and then compute sharing week acceleration rate as:

$$R_C(u) = \frac{\Delta C_w(u)}{\Delta C_{w(u)-1.0}(u)}.$$

The mean $R_C(u)$ for all selected users is 2.2. A one-sample t -test showed that users create statistically significantly more circles after getting touched by circle sharing, with $p < 0.05$ for each weekly cohort separately. This is a strong indication that these users are creating new circles based on the shared circles they see.

Mean circle size. Finally, we measure the mean sizes of circles owned by each cohort of users, before and after the owners become circle-sharing-touched. If users are adopting the shared circles they see by adding all or some members of the shared circle into their existing circles, then we would expect the mean size of existing circles owned by users to increase more quickly when the users become touched by circle sharing. For each user group, we compute the mean size of the associated circles over time and show the trend of the example cohort (first touched by circle sharing during the week of Nov 2) in Figure 6(c). We see that circles expand faster during the week when their owners first became circle-sharing-touched. (Again, the same observations are true for other user groups.)

Similar to the circle count case, we also compute the sharing week acceleration rate for circle size increase and compute the p -values for statistical significance test. The mean acceleration rate for circle size is 1.9, and all of the p -values for different cohorts are below 0.05. These results demonstrate that users expand their existing circles faster when they become circle-sharing-touched, which is a strong indication that users are adding contacts to existing circles from the shared circles they see.

RECOMMENDING CIRCLES TO SHARE

With the impact of circle sharing events in mind, in this section, we focus our efforts on distinguishing shared circles from ordinary circles (i.e., those that do not get shared). Ultimately, this has interesting applications, including recommending to the user which of his circles are good candidates for sharing. We identify a quantitative feature of circles, which we call *commonality*, and we demonstrate how commonality can be used to recommend circles for sharing.

One of the main goals of circle sharing is to let other users reuse all or part of a shared circle to create similar circles. Thus, intuitively, we expect that the circles that are the best candidates for sharing are those that are of common interest, or useful to many people. Following this reasoning, we suspect that if many users have already constructed the same circle (or a circle containing a very similar set of people), then that is a good indication that the circle is a prime candidate for sharing.

Following this intuition, we define a property of a circle c called *commonality*, which summarizes the extent to which other users have constructed a circle that is similar to c . Before describing the details of the commonality definition, we first define the *co-existence probability* of two users to capture the frequency with which two users co-occur in the same circles. In particular, the co-existence probability of two users is defined as the average conditional probability that having one user in a circle would result the other user also being in the same circle².

EXAMPLE 4. Consider the social network in Figure 2. Claire is in 3 circles and Dan is in 1 circle. They co-occur in 1 circle. The conditional probability that a circle including Claire would also include Dan is $1/3 = 0.33$, the conditional probability that a circle including Dan would also include Claire is $1/1 = 1$. Therefore, the co-existence probability of Claire and Dan is $(0.33 + 1)/2 = 0.67$.

Based on the co-existence probability of two users, we can then define *commonality* as follows:

DEFINITION 4. (Global) Commonality The commonality of a circle is defined as the average co-existence probability, taken over all pairs of users in the circle.

If there exist many other circles (created by other users) that are similar to circle c , then we expect c to have high commonality; otherwise, it should have low commonality. Since we consider circles owned by all social network users when computing the co-existence conditional probability, we also call it *global commonality* (in analogy to *local commonality*, which we will define later).

EXAMPLE 5. Consider the social network in Figure 2. Alice's circle contains three pairs of users: Bob and Claire with co-existence probability of 1, Bob and Dan with co-existence probability of 0.67, Claire and Dan with co-existence probability of 0.67. Therefore, the commonality of Alice's circle is $(1 + 0.67 + 0.67)/3 = 0.78$.

To compare shared circles and ordinary circles, we randomly selected 9000 shared circles and 9000 ordinary circles. To make sure the owners of ordinary circles are aware of the option of circle sharing, when sampling the ordinary circles, we only consider circles owned by a user who has shared at least one circle. Using this data set, we compute the probability density function of global commonality, for both shared circles and ordinary circles (Figure 7(a)). As expected, shared circles tend to have higher global commonality than ordinary circles.

Note that global commonality considers all social network users' circles when computing co-existence probabilities. We suspect that this will be less meaningful for community circles, since the members of a community circle are likely to be of interest to only a small subset of the social network's users. (On the other hand, members of celebrity circles tend

²For consistency, we assume the circle owners are also in their own circles when computing co-existence probability.

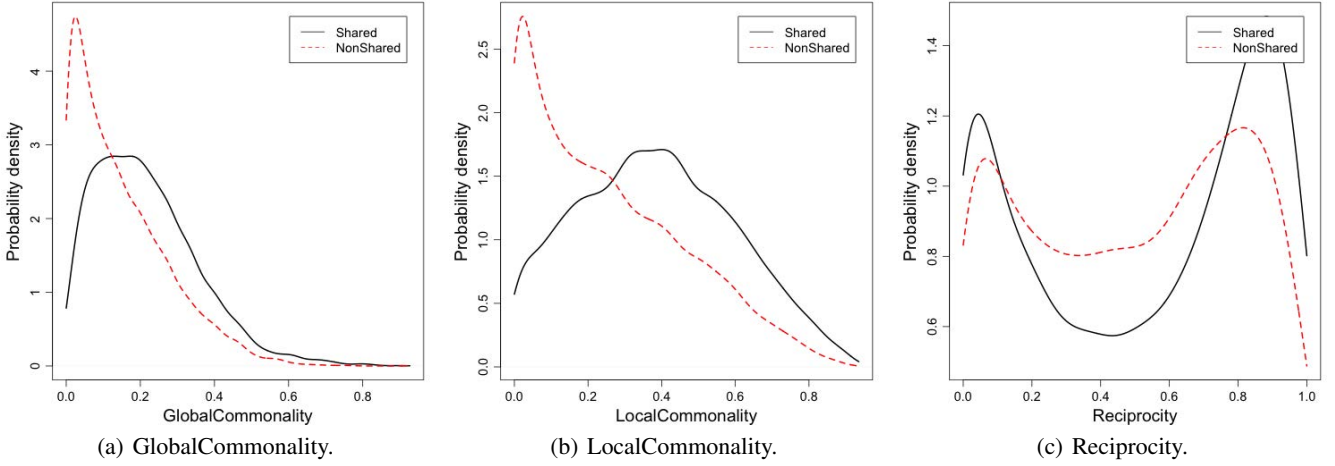


Figure 7. A comparison of shared and ordinary circles based on the probability density function of different features.

to be of more global interest.) To capture this intuition, we define *local commonality* as follows:

DEFINITION 5. Local Commonality The local commonality of a circle is defined as the average co-existence probability (considering only those circles owned by members of the given circle), computed over all pairs of users in the circle.

EXAMPLE 6. Consider the social network in Figure 2, and imagine there is an additional user Eva who has Bob, Claire and Dan in her circle. When computing the local commonality for Alice, Eva’s circle would be ignored since Eva is not in Alice’s circle; however in the case of global commonality, Eva’s circle would be considered.

We show the probability density functions of local commonality for both shared and ordinary circles in Figure 7(b). Similar to the global commonality case, shared circles are of higher local commonality comparing to ordinary circles, although the difference is even larger comparing to the global commonality case. This indicates that local commonality could be a better feature to distinguish shared and ordinary circles than global commonality.

Aside from local and global commonalities, the features mentioned in previous sections (e.g., reciprocity, density, popularity) can also be used to distinguish shared and ordinary circles. For example, we show the probability density functions of reciprocity for shared and ordinary circles in Figure 7(c). Compared to ordinary circles, shared circles are more likely to have very high or low reciprocity. We also notice that, even for non-shared circles, there is a tendency for circles to have either very high or very low reciprocity, which indicates that the categorization of circles into two types – celebrity and community – is applicable to circles in general, but that the phenomenon is more pronounced for shared circles.

In the following, we categorize all the circles (i.e., the union of all sampled shared and ordinary circles) in our dataset

into celebrity and community circles and compute the correlation between each feature (reciprocity, popularity, density, local and global commonality) and the circle sharing decision. Of course, some outlier circles do not fit into either of the two categories (celebrity or community). However, this is actually a good indication that they are less likely to be shared (e.g., see Figure 7(c)). Therefore, it is less sensitive to which category we put them into. The Pearson correlation coefficients between circle features and sharing decisions for both celebrity and community categories are shown in Figure 3. We notice that in both the celebrity and community cases, global commonality, local commonality, popularity and density have positive correlation with circle sharing. As expected, reciprocity is positively correlated with circle sharing for community circles, but negatively correlated with circle sharing for celebrity circles. We also notice that, for all of these features, they are more correlated with sharing behavior for community circles, indicating that recommendation for community circles can be made with better accuracy than celebrity circles.

Feature	Correlation to sharing	
	(celebrity)	(community)
GlobalCommonality	0.10	0.30
LocalCommonality	0.15	0.36
Reciprocity	-0.09	0.26
Popularity	0.09	0.22
Density	0.16	0.32

Table 3. Correlation of sharing with various features. For both community and celebrity circles.

In summary, these results suggest that we can recommend to a user to share a circle if either (1) it is a community circle, and it has high reciprocity, popularity, density, local and global commonality, or (2) it is a celebrity circle, and it has low reciprocity, high popularity, high density, and high local and global commonality.

We built such a recommender using an SVM classifier and the proposed features to test the feasibility of such recom-

mendation. This is a difficult problem since a user might make the sharing decision for various unpredictable reasons (e.g., some users might just want to try out the circle sharing feature and randomly pick some circles to share). To evaluate the precision and recall of the recommendation, for both the celebrity and community circles, we use 2/3 of them as training data to train a classifier using the circles features mentioned above, then we compute the precision and recall for the recommendations on the remaining 1/3 testing data. The results are shown in Table 4. Compared to celebrity circles, sharing of community circles can be predicted more accurately, although recalls and predictions in both cases are not very high. Better predictions might be achieved by considering additional features like time of sharing, the sharer's online activity history, etc., and the details are left as future work.

<i>Circle group</i>	<i>Precision</i>	<i>Recall</i>
Community	0.66	0.78
Celebrity	0.63	0.60

Table 4. Circle sharing prediction.

Finally, recalling that circles can be shared publicly or to selected smaller audiences, we examine the ACL'd recipients of shared circles. For simplicity, we consider just two categories (*public* to everyone, and *selective*, meaning that the circles was shared with a smaller group of people). For both celebrity and community cases, most circles are shared privately, although, unsurprisingly, celebrity circles are more likely to be shared publicly than community circles.

<i>Circle group</i>	<i>Public</i>	<i>Selective</i>
Community	25%	75%
Celebrity	37%	63%

Table 5. Targets of shared circles.

CONCLUSION

In this paper, we provided the first large-scale study of the usage and impact of a contact-group sharing tool, the Google+ circle-sharing feature. We identified two different types of shared circles, “communities” and “celebrities,” which are characterized by different structural properties (density, reciprocity, and popularity), and which also represent qualitatively different use cases for the feature.

We also observed that the circle-sharing feature has had measurable effects on the growth and structure of the social network graph. Edges among circle members grow 150% faster during the week the circle gets shared. Recipients of shared circles create significantly more new circles and add significantly more people to their existing circles based on the shared circles.

Finally, we demonstrate the feasibility of recommending to users which circles they should share with friends. We propose a feature called *commonality* that captures the potential benefits to share a circle. Using commonality and other circle features, we build a recommender and show that circle sharing events, especially those associated with community

circles, can be predicted with reasonable precision and recall.

In the future, we plan to study the interaction among different circle-sharing events. It would be interesting to know if one circle-sharing event often triggers others, and if yes, how such events propagate through the social network. We also plan to explore how to combine the power of contact-group sharing tools with the intelligence of friend recommenders based on link prediction.

ACKNOWLEDGEMENTS

We're grateful to Ed Chi for his help with brainstorming at the inception of this work, and to Owen Prater and Andy Hertzfeld for building the Google+ circle sharing feature and helping us understand the implementation details. Lujun Fang was partially supported by NSF Grant CNS-1017149.

REFERENCES

1. Facebook smart list. <http://blog.facebook.com/blog.php?post=10150278932602131>.
2. Facebook statistics. <http://newsroom.fb.com/content/default.aspx?NewsAreaId=22>.
3. Google circle-sharing feature release news. <http://plusheadlines.com/google-circle-sharing-is-here/2378/>.
4. Katango. <http://www.katango.com/>.
5. Listorious. <http://listorious.com/>.
6. Social networking site coverage. <http://www.pewinternet.org/reports/2011/social-networking-sites.aspx>.
7. P. Adams. *Grouped: How small groups of friends are the key to influence on the social web*. New Riders Press, 2011.
8. R. Z. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002.
9. L. Backstrom, D. Huttenlocher, and J. Kleinberg. Group formation in large social networks: Membership, growth, and evolution. In *SIGKDD*, 2006.
10. K. Bacon and P. Dewan. Towards automatic recommendation of friend lists. In *CollaborateCom*, 2009.
11. B. Everitt and T. Hothorn. *A Handbook of Statistical Analyses Using R*. Chapman and Hall/CRC, 2006.
12. S. Farnham and E. Churchill. Faceted identity, faceted lives: social and technical issues with being yourself online. In *CSCW*, 2011.
13. S. Fortunato. Community detection in graphs. *Physics Reports*, 2010.
14. L. Getoor and C. P. Diehl. Link mining: A survey. *SigKDD Explorations Special Issue on Link Mining*, 2005.
15. H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a soial network or a news media? In *WWW*, 2010.
16. J. Leskovec, K. Lang, and M. Mahoney. Empirical comparison of algorithms for network community detection. In *WWW*, 2010.
17. D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *CIKM*, 2003.
18. A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *WOSN*, 2008.
19. M. Newman. Clustering and preferential attachment in growing networks. *Physics Review E*, 2001.
20. F. Ozenc and S. Farnham. Life modes in social media. In *CHI*, 2011.
21. M. Skeels and J. Grudin. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In *Group*, 2009.