# Google+ or Google-? Dissecting the evolution of the new OSN in its first year

Roberto Gonzalez, Ruben Cuevas
Universidad Carlos III de Madrid
{rgonza1,rcuevas}@it.uc3m.es

Reza Motamedi, Reza Rejaie
University of Oregon
motamedi,reza@cs.uoregon.edu

Angel Cuevas
Telecom Sud Paris
angel.cuevas_rumin@itsudparis.eu

## ABSTRACT

In an era when Facebook and Twitter dominate the market for social media, Google has made another attempt to become a player by introducing Google+. This begs the question that "whether G+ can sustain a meaningful growth to become a relevant player in this market despite the dominance of Facebook and Twitter?". The conflicting reports on Google mostly focus on high level statistics that do not offer a meaningful answer. To tackle this question, this paper presents a detailed measurement study to characterize the evolution of the key features of Google plus over a year. Our main findings are: (i) the LCC grows at an increasing rate, however its relative size shrinks. This suggests that there is an effective (and increasing) growth in the number of interested users, however this seems to be significantly smaller than the growth reported by official statistics. (ii) The overall public activity including actions (e.g. posts) and reactions (e.g. plusones and comments) is growing. However, the public activity (both actions and reactions) is concentrated around a low fraction of popular users. (iii) The connectivity topology seems to be reaching an stable status what is an indication of maturity of the system. Furthermore, connectivity properties reveal that Google+ seems to have clear similarities (e.g. degree distribution) to a social media network like Twitter, but with a higher presence of friendship (i.e. bidirectional) relationships. (iv) Users prefer to share professional rather than personal information. This gives an indication of the professional focus that the network is acquiring. In addition, privacy sensitivity has not changed since the percentage of users that make available different profile attributes remains almost flat.

## Keywords

Online Social Networks, Google+, Measurements, Characterization, Evolution, Growth

## 1. INTRODUCTION

During the past decade, the Internet has witnessed the rise and fall (or at least moderate success) of several Online Social Networks (OSN) (*e.g.* Bebo, Friendster, MySpace, Orkut). However, two major OSNs with rather different features, namely Facebook and Twitter, have enjoyed an increasing popularity over the past few years. This raises a couple of interesting questions: First, *can a new OSN become popular despite the dominance of Facebook and Twitter* and second, *what is the evolution of the key features of new released OSNs in an already mature market with an important base of well trained users?*. Interestingly, the recent launch of a new OSN by Google, called Google+, provides an opportunity to examine these questions. Since the inception of Google+ in June 2011, periodic reports from Google claimed an impressive growth in the population of Google+ users (400M users in September 2012) [1] while some other indicators and experts called Google+ a "ghost town" [2]. Apart from their conflicting content, these reports merely focused on a couple of coarse grain characteristics of Google+ (*e.g.*, users population and average daily time that Google+ users spend on the OSN). Clearly, such a narrow view of Google+ does not offer a clear and meaningful picture of its growth and evolution and thus does not shed much light on the above questions.

Furthermore, the design of Google+ imitates properties from Facebook (*i.e.*, a friendship OSN) and Twitter (*i.e.*, a broadcast Social Media) so that users can use Google+ in any of the two manners (or a combination of both). However, to the best of the authors knowledge, *the purpose for which users are utilizing Google+?* is still an unanswered question.

This paper aims to answer the three previous questions. To this end we perform a measurement-based study to characterize the evolution of the key properties of Google+ during its 17 months of life. Leveraging the numerical IDs for Google+ users, we developed a simple, parallel technique to crawl the connectivity structure of Google+ and capture several snapshots of the largest connected component (LCC) across several months. Since sampling random Google+ users through generating random IDs is infeasible, we carefully leverage its search API to collect random users. Using these random samples, we estimate the evolution of the fraction of Google+ users that are located in its LCC, small partitions or are singletones (*i.e.*, isolated nodes).

Using Google+ API, we crawl **all** the publicly visible posts made by LCC users and their associated reactions (*e.g.* , comments, reshares and plusones) during 437 days between

Jun 28 2011 (Google+ release) and Sep 7 2012. This information enables us to characterize the aggregate growth in the visible activity in Google+ as well as the skewness in the contributed posts and attracted reactions among active users.

We also examine the evolution of the main connectivity properties of the LCC across several months. In particular we analyze the degree distribution, the links bidirectionality, the clustering coefficient and the average path length.

Finally, in order to gain further insight we compare all the analyzed connectivity and activity features with their equivalents in Facebook and Twitter.

Our main findings can be summarized as follows:

(1) The number of interested users (*i.e.*, LCC users) joining the system as well as the visible activity are growing at an steady rate after the initial phase. This dismiss the argument that Google+ is a "ghost town". However the relative size of the LCC is shrinking mostly due to a dominant increase rate of the number of Singletons joining the system. The high number of singletons must be due to the integrated registration process of Google by which a user that, for instance, creates a Gmail or a YouTube account is automatically registered in Google+. Hence, the high presence of Singletons (that are most likely considered in the official statistics) suggest that the effective growth of the system is significantly smaller than that announced by Google.

(2) Despite of the significant growth in number of nodes and activity, Google+ seem to have reached an stable status due two the following observations: First, despite the growth on number of nodes and links the distributions of the main connectivity features present marginal statistical variation in the last three months. Second, in spite of the steady growth in the overall visible activity, the contribution of posts and their associated reactions (*e.g.*, comments or plusones) are concentrated in a small fraction of users that seem to sustain the visible activity of Google+.

(3) There is a strong correlation between users popularity and activity (*i.e.*, post rate and aggregate reaction rate). In other words, popular users are the most active and those attracting more reactions from other users.

(4) The observations in the two previous points suggest that in its seemingly stable status Google+ is a broadcast social media in which a small fraction of popular users are the source of information consumed by other users and so are also the target of most reactions.

The rest of this paper is organized as follows: Section 2 introduces a short overview of Google+. In Section 3, we describe our data collection techniques and datasets. We study the relative size of different components and the growth of the LCC in Section 4. Section 5 and 6 address the evolution of activity and connectivity properties of Google+ whereas Section 7 analyzes the correlation between activity and connectivity metrics. Finally, Section 8 presents the related work and Section 9 concludes the paper.

## 2. GOOGLE+ OVERVIEW

After a few unsuccessful attempts (Buzz [7], Wave [19] and Orkut [20, 21]), Google launched Google+ on June $28^{th}$ 2011 with the intention of becoming a major player in the OSNs market. Users were initially allowed to join by invitation. The overwhelming demand of accounts made Google to stop accepting new users a couple of days after the release of the service to adapt their infrastructure to such demand. On September $20^{th}$, Google+ became open to public and the Google+ Pages service was launched on November $7^{th}$ 2011 [13, 14]. This service imitates the *FaceBook Pages* enabling businesses to connect with interested users. Furthermore, also in November 2011, the registration process was integrated with other Google services (*e.g.*, Gmail, YouTube) [17, 18]. After this integrated registration, as we explicitly verified, the creation of a new Gmail or YouTube account automatically generates a Google+ account for the user.

Google+ features have some similarity to Facebook and Twitter. Similar to Twitter (and different from Facebook) the relationships in Google+ are unidirectional. More specifically, user $A$ can follow user $B$ in Google+ and view all of $B$'s public posts without requiring the relationship to be reciprocated. We refer to $A$ as $B$'s *follower* and to $B$ as $A$'s *friend*. Moreover, a user can also control the visibility of a post to a specific subset of its followers by grouping them into *circles*. This feature imitates Facebook approach to control visibility of shared content. It is worth noting that this circle-based privacy setting is rather complex for average users to manage and thus unskilled users may not use them properly[1].

Each user has a stream (similar to Facebook wall) where any activity performed by the user appear. The main activity of a user is to make a "post". A post consists of some (or no) text that may have one or more attached files, called "attachments". Each attachment could be a video, a photo or any other file. Other users can react to a particular post in three different ways: *(i) Plusone*: this is similar to the "like" feature in Facebook with which other users can indicate their interest in a post, *(ii) Comment*: other users can make comments on a post, and *(iii) Reshare*: this feature is similar to a "retweet" in Twitter and allows other users resend a post to their followers.

Google+ assigns a numerical user ID and a profile to each user. The user ID is a 21-digit integer where the highest order digit is always 1 (e.g. 113104553286769158393). Our examination of the assigned IDs did not reveal any clear strategy for ID assignment (*e.g.*, based on time or mod of certain numbers). Note that this extremely large ID space ($10^{20}$) is sparsely populated (large distance between user IDs) which in turn makes identifying valid user IDs by generating random numbers infeasible. Moreover, the user profile includes the attributes listed in Table 6. For each attribute the user can define an independent privacy setting ranging from "public" to "no accessible to anyone".

For a more detailed description of Google+ functionality we refer the reader to [11, 12].

## 3. MEASUREMENT METHODOLOGY AND DATASETS

### 3.1 Capturing LCC Structure

To capture the connectivity structure of the Largest Connected Component (LCC), we use a few high-degree users as starting seeds and crawl the structure using a breadth-first search (BFS) strategy. Our initial examination revealed that the allocated users IDs are very evenly distributed across the ID space. We leverage this feature to speed up our crawler

[1] A clear example of this complexity if the provided diagram to guide users to determine their privacy setting [8].

| Name | #nodes | #edges | Start Date | Duration (days) |
|---|---|---|---|---|
| LCC-Dec* | 35.1M | 575M | 11/11/12 | 46 |
| LCC-Apr | 51.8M | 1,049M | 15/03/12 | 29 |
| LCC-Aug | 79.2M | 1,643M | 20/08/12 | 4 |
| LCC-Sep | 85.3M | 1,742M | 17/09/12 | 5 |
| LCC-Oct | 89.76M | 1,799M | 15/10/12 | 5 |
| LCC-Nov | 93.08M | 1,850M | 28/10/12 | 6 |

**Table 1: Main characteristics of LCC snapshots (LCC-Dec was collected in [39] whereas for LCC-Apr we accounted with less resources than for the other snapshots)**

as follows: We divide the ID space into 21 equal-size zones and assign a crawler to only crawl users whose ID falls in a particular zone. Given user $u$ in zone $i$, the assigned crawler to zone $i$ collects the profile along with the list of friends and followers for user $u$. Any newly discovered users whose ID is in zone $i$ are placed in a queue to be crawled whereas discovered users from other zones are periodically reported to a central coordinator. The coordinator maps all the reported users by all 21 crawlers to their zone and periodically (once per hour) sends a list of discovered users in each zone to the corresponding crawler. This strategy requires infrequent and efficient coordination with crawlers and enables them to crawl their zones in parallel. Crawl of each zone is completed when there is no more users in that zone to crawl. After some tuning, the average rate of discovery for each crawler reached 800K users per day or 17M users per day on average for the whole system. With this rate, it takes 4-6 days to capture a full snapshot of the LCC. Table 1 summarizes the main characteristics of our LCC datasets. We examined the connectivity of all the captured LCC snapshots and verified that all of them form a single connected components.

## 3.2 Sampling Random Users

Our goal is to collect random samples of Google+ users for our analysis. To our knowledge, none of the prior studies on Google+ achieved this goal. The sparse utilization of the extremely large ID space makes it infeasible to identify random users by generating random IDs. To cope with this challenging problem, we leverage the Google+ search API to efficiently identify a large number of seemingly random users. The search API provides a list of up to 1000 users whose name or surname matches a given input keyword. Careful inspection of search results for a few surnames revealed that Google+ appears to order the reported users based on their level of connectivity and activity, *i.e.* users with a higher connectivity or activity (that are likely to be more interesting) are listed at the top of the result. Since searching for popular surnames most likely results in more than 1000 users, the reported users are biased samples. To avoid this bias, we selected a collection of 1.5K random American surnames from the US[2] 2000 census [9] with low to moderate popularity and used the search API to obtain matched Google+ users. We consider the list of reported users only if it contains less than 1000 users. These users are assumed to be random samples because Google+ must report all matached

[2]US is the most represented country in G+ [39, 44]. Furthermore, the high immigration level of US allows to find surnames from different geographical regions

| Name | #nodes | #edges | Start Date | Duration (days) |
|---|---|---|---|---|
| Rand-Apr | 2.24M | 144.76M | 18/04/12 | 23 |
| Rand-Oct | 5.73M | 262.95M | 15/10/12 | 10 |
| Rand-Nov | 3.46M | 157.22M | 28/10/12 | 13 |

**Table 2: Main characteristics of Random datsets**

| Active Users | Postings | Attachments | +1 | Replies | Reshares |
|---|---|---|---|---|---|
| 13.6M | 218.48M | 299.41M | 352.01M | 202.21M | 63.78M |

**Table 3: Main characteristics of Activities dataset (collected in Sep 2012)**

users, and there is no correlation between surname popularity and the connectivity (or activity) of the corresponding users. Table 2 summarizes the main characteristics of our random datasets. Note that the timing of each one of the random datasets is aligned with a LCC dataset.

To validate the above strategy, we collect two groups of more than 140K samples from the search API, users whose name match popular and unpopular ($<$ 1000 users) surnames. We focus on samples from each group that are located in the LCC since we have a complete snapshot of the LCC that can be used as ground truth. In particular, we compare the connectivity of samples from each group that are located in LCC with all users in a LCC snapshot. Figure 1 plots the summary distribution of the number of followers and friends for these two groups of samples and all users in the LCC, respectively. These figures clearly demonstrate that only the collected LCC samples from unpopular surnames exhibit very similar distributions of followers and friends with the entire LCC. A Kolgomorov-Smirnov test confirms that they are indeed the same distribution. The collected samples from popular surnames have a stronger connectivity and thus are biased.

## 3.3 Capturing User Activity

We consider user activity as a collection of all posting by individual users and the reaction (*i.e.* PlusOnes, Comments and Reshares) from other users to these posting. User activity is an important indicator of user interest and thus the aggregate activity (and reactions) across users is a good measure of an OSN popularity. Despite its importance, we are not aware of any prior study that examined this issue among Google+ users. Toward this end, we focus on user activity in the most important element of the network (*i.e.* the LCC). We leverage the Google+ API to collect all the public posts and their associated reactions for all LCC-Sep users between Google+ release date (Jun $28^{th}$ 2011) and the date our measurement campaign started (Sep $7^{th}$ 2012), *i.e.* 437 days. Given the cumulative nature of recorded activity for each user, a single snapshot of activity contains all the activities until our data collection time. Furthermore, since each post has a timestamp, we are able to determine the temporal pattern of all the posts from all users. Note, that Google+ API limits the number of daily queries to 10K per registered application. Then, we use 303 accounts to collect the referred data in 68 days. Table 3 summarizes the main features of the activity dataset. In particular, note that only 13.6M (out of 85M) LCC-Sep users made at least one public post in the analyze period.

## 3.4 Other datasets

Along the paper we use the following datasets for comparison purposes: (*i*) Authors from [37], [26] have kindly provided data from complete snapshots of Twitter graph collected in 2009; (*ii*) Authors from [39] have also shared data from a Google+ snapshot collected between November and December 2011. Note that we use this snapshot to analyze the long term evolution of the LCC characteristics for Google+. Along the paper we refer to this snapshot as LCC-Dec and its main features are summarized in Table 1; (*iii*) We use information related to Twitter users activity from our previous dataset [43]; (*iv*) Due to the unavailability of information for Facebook users activity we have implemented our own crawler tool and collected the public posts of 12K (random) Facebook users between Jun 1st 2011 and September 7th 2012. Furthermore, we have also crawled the degree and public profile attributes for 480K and 75K (random) Facebook users, respectively.

# 4. MACRO-LEVEL STRUCTURE & ITS EVOLUTION

Similar to any other OSN, the macro-level connectivity structure among Google+ users should consist of three components. (*i*) A largest connected component (LCC), (*ii*) A number of partitions (with at least 2 users) that are smaller than LCC, and (*iii*) Singletons or isolated users.

Our goal is to estimate the relative size of each component and its evolution over time. Toward this end, we determine the mapping of users in a random dataset across the three main components to estimate their relative sizes. The LCC users can be easily detected using the corresponding LCC snapshot (*e.g.*, LCC-Oct for Rand-Oct). For all the users outside LCC, we perform a BFS crawl from each user to verify whether a user is a singleton or part of a partition, and in the latter case determine the size of the partition.

The first part of Table 4 presents the relative size of all three components using our random datasets in April, October and November of 2012[3]. The other two parts of this table present the fraction of users that have any public posting or public attributes in their profiles and map them to each component. This table shows that the LCC, partitions and Singletons made up 43%, 1.4% and 55%[4] of Google+ network in April 2012, respectively. Six months later in October of 2012, the size of LCC drops to 32% while partitions and singletons grow to 1.7% and 66% of the network, respectively. Finally, in the last snapshot (in November), the increasing (decreasing) trend for Singletons (LCC) continues while we observe a reverse trend for Partitions. Interestingly, the LCC is significantly less representative in Google+ than in other OSNs. For instance, Facebook's LCC included 99.91% of the registered users as of May 2011 [46] whereas in Twitter the LCC represented 94.8% of the nodes with just 0.2% Singletons in August 2009 [26]. Furthermore, Leskovec et al. [38] showed that the relative size of the LCC typically grows up to represent more than 90% in several studied networks (e.g., AS-topology or ArXiv citation graph). The unnatural growth pattern of the relative size for different Google+ components and, in particular, the large fraction of Singletons must be due to the integrated registration procedure of Google: when a user creates a Google account to utilize a specific Google service, such as Gmail or YouTube, she is automatically registered in Google+[5]. It is likely that many of these users may be unaware (not interested) to be (become) members of Google+ and then, they become Singletons.

Furthermore our representative sampling technique is able to identify tens of thousands of Partitions, then we are able to study the distribution of size across partitions in each snapshot. The results show that 99% partitions had a size $\leq$ 4, 3 and 3 for our April, October and November snapshots, respectively. Moreover, the size of the largest partition also decreases from 55 (April) to 42 (October) and to 36 (November). Therefore, Google+ Partitions are rather small, in addition, bigger partitions seems to be absorbed by the LCC what leads to the observed reduction in the 99 percentile of the size distribution.

In terms of active users (with at least one posting), Table 4 demonstrates that only 10% of users were active in April and this fraction dropped to 8.6% after seven months in November of 2012. The absolute majority of these active users (roughly 80% of them) are located in LCC, and most of the remaining ones are Singletons. Finally, the fraction of users that publicly share at least one attribute (in addition to sex that is mandatory) in their profile, dropped from 29% in April to 24% in November. Again, a decreasing majority of these users are in LCC and the increasing minority of them are Singletons.

## 4.1 Evolution of LCC Size

Most LCC users, contrary to Singletons, show an interest to participate in the network by creating (accepting) connections to (from) other users. Capturing complete snapshots of the LCC enables us to examine the evolution of the total number of users in the LCC along with the number of users who join (show interest) and leave (lose interest) the LCC between two snapshots as shown in Figure 1 using log-log scale. This figure illustrates that the overall size of LCC has roughly doubled (from 26M to 51M) in four months between December 2011 to April 2012 at an average growth rate of 155K users per day. This average rate has increased between April and November 2012 to 207K users per day. This shows that Google+ is attracting an increasing number of interested users that is a clear symptom of "health". However, the pace at which interested users join the system is significantly smaller to that depicted by official statistics [1] that most likely also include Singletons in their reports. Furthermore, it is interesting to notice that the growth rate presents some short term variability. This is consistent with recent results by Gaito et al. [28] that show the bursty nature of the evolution of OSNs.

Moreover, it is worth to examine whether users who lose their interest in Google+ (*i.e.*, were part of the LCC and

[3]It is possible that our approach incorrectly categorizes user $u$ as a Singleton if $u$ has a private list of friends and followers and, in addition, all of $u$'s friends and followers have a private list of followers and friends, respectively. However, we believe this is rather unlikely. Indeed our BFS on the LCC identified about 7.5% users with private friend and follower lists who were detected through their neighbors.

[4]Note that, based on probabilistic theory, the large random sample of nodes collected (in the order of millions) guarantees the accuracy of the obtained results with an error smaller than 0.1%.

[5]We indeed confirmed this observation for new Gmail and YouTube accounts.

| Element | % users | | | % users public posts | | | % users public attr. | | |
|---|---|---|---|---|---|---|---|---|---|
| | Apr | Oct | Nov | Apr | Oct | Nov | Apr | Oct | Nov |
| LCC | 43.49 | 32.33 | 32.23 | 8.93 | 7.03 | 6.92 | 27.43 | 17.92 | 17.63 |
| Partitions | 1.41 | 1.73 | 1.45 | 0.13 | 0.18 | 0.15 | 0.5 | 0.60 | 0.51 |
| Singletons | 55.10 | 65.94 | 66.32 | 1.27 | 1.57 | 1.57 | 1.77 | 5.69 | 6.17 |
| All | 100 | 100 | 100 | 10.33 | 8.78 | 8.64 | 29.69 | 24.21 | 24.31 |

**Table 4: Fraction of Google+ users, active users and users with public attributes across Google+ components along with the evolution of these characteristics from April to November of 2012 (based on the corresponding Random datasets)**

then departed) present any distinguishable characteristic[6]. Toward this end, we compute the distribution of the number of followers, friends, public attributes and public post among departed LCC users between our first and last LCC snapshots compared with the same characteristics for all the remaining LCC users. The obtained results show that departed nodes present similar connectivity and information sharing properties to LCC nodes but are less activity. This suggests that the lack of incentives to actively participate in the system motivates these users to leave.

*In summary, our results reveal that 2/3 (and increasing fraction) of Google+ users are Singletons while the other 1/3 are mostly in the LCC except for a small fraction (1%) that form small Partitions. The large and increasing fraction of Singletons is due to Google's integrated registration service. Apart from Singletons, Google+ is attracting an increasing rate of interested users and those who leave the system are among the least active users. Finally, only 10% of users exhibit any public activity and 25% of users post any public attributes, and a significant majority of these users are in LCC.*
Since the LCC is the well connected component that contains a majority of active users, we focus our remaining analysis only on LCC.

# 5. PUBLIC ACTIVITY & ITS EVOLUTION

In this section, we characterize publicly visible (or in short "public") posts by users in LCC as well as other users' reactions (including users outside LCC) to these posts. We focus only in public posts because collecting private information have serious ethical issues, and cannot be done (at the needed scale to derive meaningful conclusions) without the unlikely consent and collaboration of hundreds of thousands (or millions of) users. In addition, public posts and their reactions are an important indicator of activity in Google+ for the following reasons: First, earlier studies have reported that more than 30% of post during the initial phase were public [35]. Second, note that only the public posts are indexable by search engines (including Google), and thus visible to others (apart from Google) for marketing and mining purposes [15]. Thus, characterizing public posts provides an important insight about the publicly visible part of Google+.

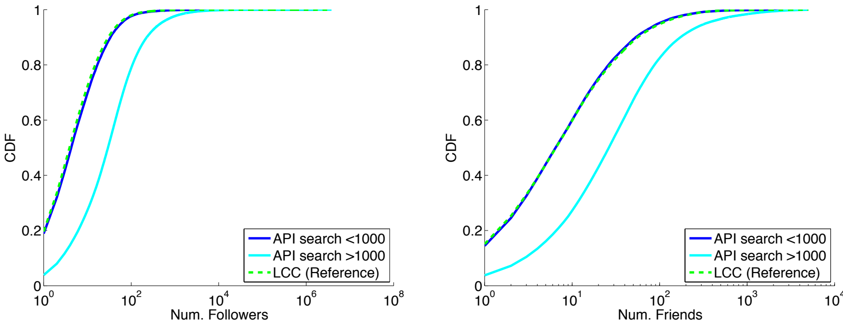
We focus on public activity only for users in LCC since they are responsible for a significant majority of activities among users as we reported in Section 4. We recall that the main *action* by individual users is to generate a "post" that may have one or more "attachments". Each post by a user may trigger other users to react by making a "comment", indicate their interest by a "plusone" or "reshare" the post with their own followers. Note that for scalability reasons we only collect the timestamp for individual posts (but not for reactions to each post). We use the timestamp of each post as a good estimate for all of its reactions because most reactions often occur within a short time after the initial post. To validate this assumption we have collected the timestamp of 4M comments associated to 700K posts and confirm that more than 80% of the comments are done within the 24 hours following the posting time.

**Temporal characteristics of public activity:** Using the timestamp for all the posting and associated reactions enables us to examine the temporal characteristics of public activity among Google+ user in LCC during the entire 15 months of Google+ operation.
Figure 3(a) depicts the total number of daily posts by LCC users along with the number of daily posts that have attachments, have at least one plusone, have been reshared or have received comments. Note that a post may have any combinations of attachment, plusone, reshare and comment (*i.e.*, these events are not mutually exclusive). The pronounced repeating pattern in this figure (and other similar results) is due to the weekly change in the level of activity among Gooogle+ users that is a significantly lower during the weekend and much higher during weekdays. Most of the observed spikes in this (and other related) figure(s) appear to be perfectly aligned with specific events as follows[7]: *(i)* on Jun 30 caused by the initial release of the system (by invitation) [3]; *(ii)* on Jul 11, as users reaction to a major failure on Jul 9 when the system ran out of disk [4]; *(iii)* on Sep 20, caused by the public release of the system [3]; *(iv)* on Nov 7, due to the release of Google+ Pages service [14]; *(v)* on Jan 17, caused by new functionalities for auto-complete and adding text in the photos [5, 6]; *(vi)* on Apr 12, caused by a major redesign of Google+[16].

Figure 3(a) also demonstrates that the aggregate number of daily posts has steadily increased after the first five months (*i.e.*, the initial phase of operation). A significant majority of the posts have attachments but the fraction of posts that trigger any reaction by other users is much smaller, moreover plusones seem to be the most common reaction. Note that this figure presents the number of daily posts with at-

[6]Note that it is possible that the seemingly departure of these nodes be due to a change in their privacy setting. However, this is unlikely since in most cases we would be able to identify these nodes through their contacts.

[7]We could not identify any event associated with the peaks on May 3rd, Jun 4th and Aug 7th.




(a) Num Followers

(b) Num Friends

**Figure 1: Random dataset validation: Distribution of #followers and #friends for (*i*) users collected from the Search API for popular surnames, (*ii*) users collected from the Search API and unpopular surnames (< 1000 users) and (*iii*) all users in LCC**

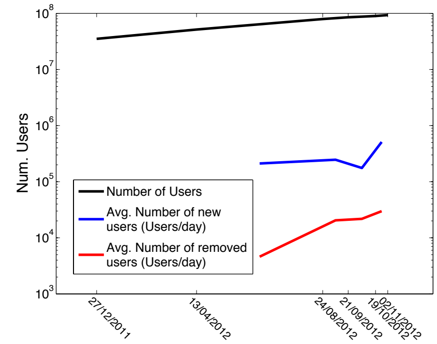
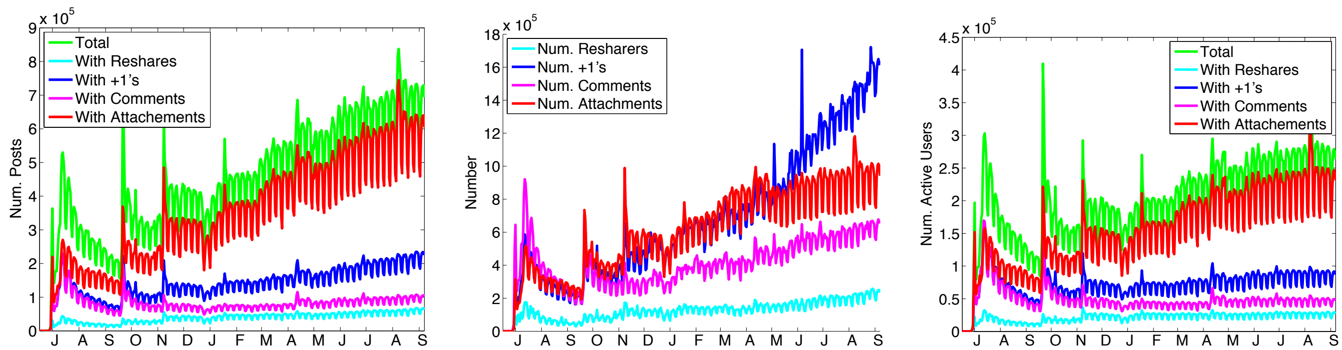



**Figure 2: Evolution of LCC size and #arriving and #departing users over time**




(a) Aggregate num. of daily posts and num. of daily posts with at least one attachement, plusone, reshare and comment

(b) Aggregate num. of daily attachements, plusones, reshares and comments.

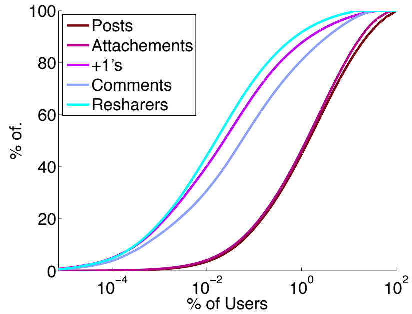
(c) Aggregate num. of daily active users and num. of daily active users that receive at least one attachement, plusone, reshare and comment

**Figure 3: Aggregate and Per user daily activity from 28 Jun 2011 to 7 Sep 2012 (the first and last months indicated in the x axis are July 2011 and Sep 2012, respectively)**

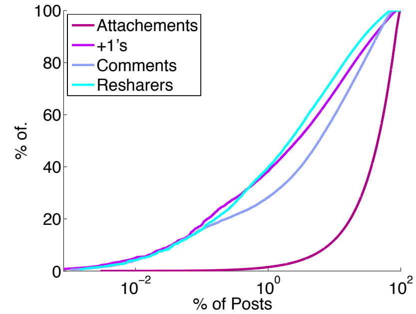
tachment or reactions but does not reveal the total daily number of attachment or reactions. To this end, Figure 3(b) depicts the temporal pattern of aggregate daily attachments, plusones, comments and reshares for all the daily posts by LCC users, *i.e.*, multiple attachments or reactions to the same post are counted separately. This figure paints a rather different picture. More specifically, the total number of comments and in particular plusone reactions has been rapidly growing after the initial phase.

Figure 3(c) plots the temporal pattern of user-level activity by showing the daily number of active LCC users along with the number of users for whom their posts have attachments or triggered at least one type of reaction. This figure reveals that the total number of users with posting has been steadily growing (after the initial phase). While a large proportion of these users create posts with attachments, the number of daily users whose posts trigger at least one plusone, comment or reshare has consistently remained roughly below 1M, .5M and .25M, respectively, despite the major growth in the number of LCC users.

**Skewness in activity contribution:** Previous results suggest that a relatively small and stable number of users with interesting posts attract most reactions. This raises the question of how skewed is the contribution of posts and associated reactions among users in Google+. Figure 4(a) presents the fraction of all posts in our activity dataset that are generated by the top $x\%$ LCC users during the 15 months of life of Google+ (the x axis uses a log-scale). Other lines in this figure show the fraction of all attachments, plusones, comments and reshares that are associated with the top $x\%$ users that receive most reactions in each category. This figure clearly demonstrates that the contribution of the number of posts and the total number of associated attachments across users is similarly very skewed. For example, the top 10% of users contribute 80%. Furthermore, the distribution of contribution of reactions to a user's posts is significantly more skewed than the contribution of total posts per user. In particular 1% of users receive roughly 80% of comments and 90% of plusones and reshares. These findings offer a strong evidence that only a very small fraction of users (around 1M) create most posts and even a smaller fraction of them create posts that trigger any reaction by others. In short, both the action and reaction aspects of public activity are centered around a very small fraction of users. Furthermore, we repeat a similar analysis at the post level. In particular,




(a) % of posts, attachements, plusones, reshares, comments associated to top x% users




(b) % of attachements, plusones, reshares, comments associated to top x% posts

**Figure 4: Skewness of actions and reactions contribution per user and post**

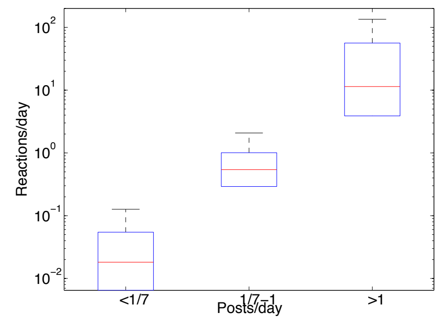



**Figure 5: Post-rate (x axis) vs aggregate reaction rate (y axis) correlation**

Figure 4(b) shows the fraction of attachements, plusones, comments and reshares associated to the top $x\%$ posts. The distribution for attachements presents an non-skewed shape that indicates that most posts include one or few attachements. For reactions we observe roughly one order of magnitude less skewed distributions for posts than for users, that is a rather expected result, since reactions spread across users' posts.

**Correlation between actions and reactions:** Our analysis so far has revealed that actions and reactions are concentrated in an small group of users, but we cannot conclude whether the members of these two groups are overlapped[8]. To address this issue, first we analyze the correlation between the post and the aggregate reaction rate for different groups of users grouped by their average level of activity as follows: users that post every day, users that post less than once a day but more than once a week and users that has less than one post per week. Figure 5 shows the distribution of daily reaction rate for each one of the described groups using boxplots. The result suggests that the reaction rate grows exponentially with the level of activity of the users. For instance, users that post every day receive more than 10 reactions per day (in median). Hence, the small group of users that contribute most posts is also receiving the major portion of reactions. To gain further insight in the correlation between users' actions and reactions, Table 5 shows the result for the Rank Correlation (RC) [31] between different types of actions and reactions for our activity dataset. Note that RC shows the correlation between the rank of a group of users by two different parameters. It offers values between -1 (ranks are reversed) and 1 (ranks are the same), where 0 indicates that ranks are independent. We observe that there is a high correlation between actions and reactions except for the case of reshares that present a lower correlation with the number of posts (0.3) and even smaller with the number of attachements (0.05).

**Identity of top users:** We have manually inspected the identity of the top 20 users with a largest number of public posts as well as those that receive a largest number of

| | posts | attach. | plusones | comments | reshares |
|---|---|---|---|---|---|
| posts | - | 0.83 | 0.78 | 0.71 | 0.3 |
| attach. | 0.83 | - | 0.6 | 0.51 | 0.05 |
| plusones | 0.78 | 0.6 | - | 0.93 | 0.71 |
| comments | 0.71 | 0.51 | 0.93 | - | 0.8 |
| reshares | 0.3 | 0.05 | 0.71 | 0.8 | - |

**Table 5: Rank Correlation between actions and reactions associated to active users.**

reactions. While the analysis of the first group does not reveal any interesting finding, we observe that among the Top 20 users attracting more reactions 18 are related to music groups of young girls from Japan and Indonesia (*e.g.,* nmb48, ske48, akb48, hkt48 from Japan or jkt48 from Indonesia) all of them produced by the same person (Yasushi Akimoto) and that seem to be very popular in those countries.

## 5.1 Comparison with Other OSNs

We compare a few aspects of user activity among G+, Twitter and Facebook to investigate the level of engagement among users in these three OSNs. For this purpose we leverage a Twitter dataset from our previous study [43] and our Facebook dataset including activity from random samples described in Section 3. In our analysis we only consider the active users within each dataset that represent 17%, 34%, and 73% of the samples for Google+, Facebook and Twitter[9], respectively.

**Posting Rate:** Figure 6(a) shows the distribution of average activity rate per user across users of each one of these three OSNs where activity rate is measured as the total number of actions (*i.e.,* posts in Google+ and Facebook or tweets in Twitter) divided by the time between the timestamp of a user's first collected action and our measurement time. This figure offers a broad sense of the level of user activity in these three OSNs and reveals that the level of activity among Facebook and Google+ users are more homogeneous than across Twitter users. Furthermore, Facebook users are most active (with typical rate of 0.19 posts/day) while Google+ users

[8] It might be that celebrities that post infrequently accumulate most of the reactions in their few posts.

[9] The common public setting of Twitter in contrast to more privacy restrictive settings in Google+ and Facebook justify the observed difference.

exhibit the least level of activity (with the typical rate of 0.08 posts/day).

**Recency of Last Post:** An important aspect of user engagement is (in addition to how much) how often individual users post. We can obtain the timestamp of the last post for all active users and then compute the *recency* of their last post as the the time between the last post timestamp and our measurement time. The distribution of this metric across a large number of users provides an insight on how frequently individual users post in an specific system. Figure 6(b) depicts the distribution of recency of the last post across different groups of Google+, Twitter and Facebook users that are grouped based on their average activity rate as described for Figure 5. We observe that Facebook users post in the system more frequently. Furthermore, they present a much more homogeneous behaviour since the difference between the 25 and 75 percentile is significantly smaller than in Twitter or Google+. If we compare now Twitter and Google+, we observe that low and average active users post more frequently in Twitter, however this trend reverses for very active users. Finally, we can observe that posting volume and frequency are correlated across systems and for users within an specific system.

**Public User Attribute:** Similar to other OSNs, Google+ users have a profile that has 17 fields where users can provide a range of information and pointers (*e.g.* ,to their other pages) about themselves. However, providing this information (except for sex) is not mandatory for creating an account and thus users may leave some (or all) attributes in their profile empty. Furthermore, users can limit the visibility of specific attributes by defining them as "non-public"[10]. Roughly 48% of all Google+ users in the LCC were initially providing at least one attribute (in addition to their sex) and this ratio decreased and then stabilized around 44% in the past few months. We further examine the distribution of the number of visible attributes across LCC users and compare it with the 480K profiles collected from Facebook (note that we identify 21 different attributes in Facebook profile) in Figure 7. The result shows that users are willing to share more information in Facebook than in Google+. In particular, half of the users publicly share at least 6 attributes in Facebook and less than 10% share that number of attributes in Google+.
Table 6 presents a more detailed view by showing the fraction of LCC users that provide public information in each specific field of their profile. Overall, users seem more inclined to share attributes related to the professional aspects such as "Studies", "location", "Profiles" and "Profession". In contrast, they are less willing to share attributes that reveal rather more private aspects of their life such as their relationships (*e.g.*, single, married) or what they are looking for? (*e.g.*, friendship, love). This may be an indication that Google+ is being used for professional purposes (or by professional rather then average users). To double-check this hypothesis we have retrieved the identity of the Top 20 users from Twitter, Facebook and Google+ (*i.e.*, users with more followers in Twitter and Google+, and Facebook pages with more fans) and manually inspected their professions. We observe that in Twitter and Facebook all the Top 20 publishers are *celebrities* (politicians, musicians, actors, soccer players, etc) and some companies (*e.g.*, YouTube, Twitter, FaceBook). However, in Google+ we found, along with some celebrities, professionals from the hi-tech sector (*e.g.*, Google CEO, Virgin CEO, Myspace founder), photographers or bloggers in the Top 20. This confirms that Google+ seems to be acquiring a more professional focus despite of have been launched as a general purpose OSN. Some of these observations are aligned with results presented in [39].

[10]Note that it is not possible to distinguish whether a non visible attribute is empty or non-public.

| attribute | DEC* | APR | AUG | SEP | OCT | NOV |
|---|---|---|---|---|---|---|
| Alias | - | 0 | 0,01 | 0,01 | 0,02 | 0,01 |
| Indexable | - | 0 | 0 | 0 | 0 | 0 |
| Other names | 4.39 | 4,08 | 4,2 | 5,05 | 5,08 | 3,34 |
| Introduction | 7.80 | 8,42 | 9,74 | 9,93 | 9,98 | 6,52 |
| Employment | 13.27 | 11,47 | 13,32 | 13,59 | 13,65 | 8,93 |
| Bragging rights | 3.90 | 3,77 | 3,93 | 4,75 | 4,78 | 3,14 |
| Relationship | 4.31 | 3,94 | 3,99 | 4,74 | 4,74 | 3,11 |
| Web | - | 1,22 | 1,1 | 1,64 | 1,72 | 1,15 |
| Work | 0.22 | 0,34 | 0,16 | 0,16 | 0 | 0,32 |
| Looking for | 2.74 | 2,64 | 2,61 | 2,19 | 2,13 | 2,03 |
| Contributor to | 13.15 | 11,95 | 11,59 | 9 | 8,68 | 8,1 |
| Places lived | 26.75 | 26,98 | 28,36 | 25,48 | 25,47 | 24,15 |
| Other profiles | 13.48 | 10,7 | 10,54 | 15,01 | 17,52 | 17,44 |
| Education | 27.11 | 24,32 | 24,72 | 21,47 | 21,05 | 20,13 |
| Home | 0.21 | 0,01 | 0,26 | 0,24 | 0,24 | 0,23 |
| Links | 3.63 | 3,26 | 3,3 | 2,76 | 2,67 | 2,51 |
| Gender | 97,67 | 95,82 | 95,76 | 94,6 | 94,44 | 94,15 |

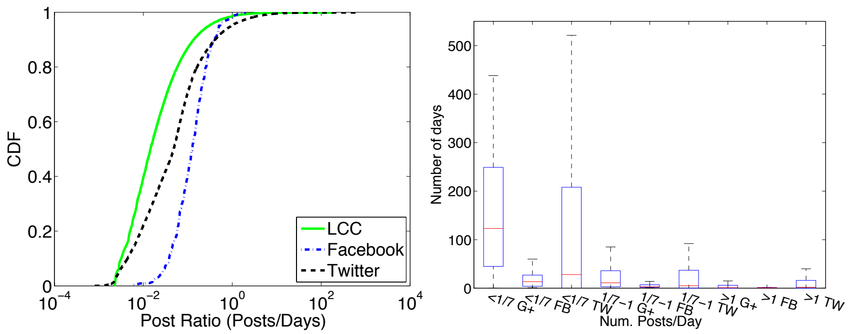
**Table 6: Percentage of LCC users that make public each attribute for each dataset**

*In summary, our results reveal that Google+ public activity has been growing at a linear rate in the last year. However, despite of this growth, posting activity and reactions are concentrated in an small fraction of users that sustain the overall public activity of Google+. This suggests that (at least) the visible part of Google+ has reached a rather stable status that present a broadcast nature in which a small fraction of users generate the information consumed by others and thus are the target of their reactions. Finally, our comparison with the public activity in other OSNs shows that Facebook and Twitter users post more (and more frequently) than Google+ users so that Google+ has still a long run before generating an activity volume comparable to its two main competitors.*

# 6. LCC CONNECTIVITY & ITS EVOLUTION

In this section, we focus on characterizing different aspects of connectivity among LCC users, the evolution of these characteristics over time as the system becomes more populated, and the comparison of these features with other OSNs. This analysis has a twofold purpose: on the one had we can evaluate the level of "maturity" achieved by the LCC of Google+ in its (roughly) one and a half year of existence. On the other hand, comparing the main properties of LCC of Google+ with those of Twitter and Facebook we can deepen our insight on the purpose for which Google+ is being used, as a friendship OSN (like FaceBook) or as a broadcast social media (like Twitter)?.

**Degree Distribution:** The distribution of node degree is one of the basic features of connectivity. Since Google+ structure is a directed graph, we separately examine the distribution of the number of followers in Figure 8(a) and friends in Figure 8(b). Each figure shows the corresponding distribution across users in each one of our LCC snapshots, the Google+ snapshot collected in Dec 2011 [39] and the snapshot of Twitter captured in Aug 2009 [26]. Furthermore




(a) Avg. Post/Tweet rate

(b) Recency of Activity

**Figure 6: Comparison of activity metrics for Google+, Twitter and Facebook**

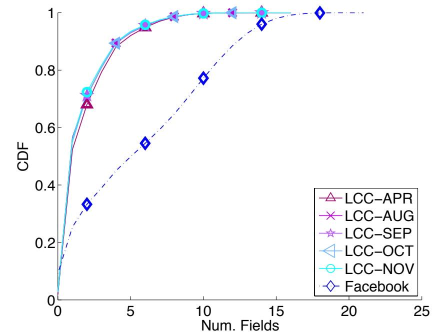
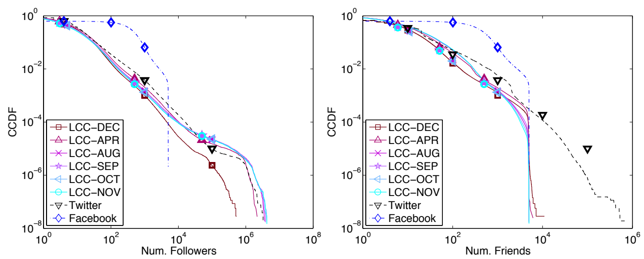


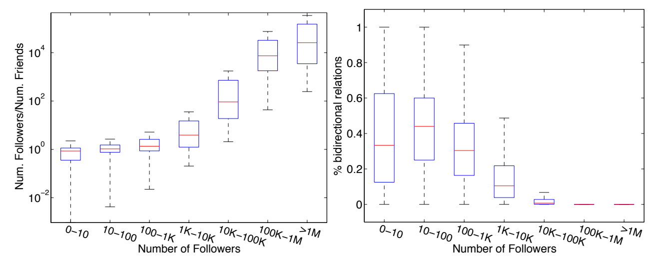

**Figure 7: Distribution of number of public attributes for Google+ and Facebook**




(a) Num Followers

(b) Num Friends

**Figure 8: Degree Distribution for different snapshots of Google+, Twitter and Facebook**




(a) #followers/#friends

(b) % bidirectional relationships

**Figure 9: The level of imbalance and reciprocation for different group of users based on their number of followers.**

we present in both figures the distribution of neighbors for our random Facebook samples[11].

The distributions of number of followers and friends can be approximated by a power law distribution with $\alpha = 1.26$ and 1.39 for our most recent LCC snapshot, respectively. This is a common property across different studies OSNs like Twitter [37], RenRen [34], Flickr or Orkut [42]. Interestingly, we observe that despite the overall growth of the LCC has been (roughly) similar in the periods Dec 2011-Apr 2012 and Apr 2012-Nov 2012 (see Figure 2), the change in the degree distributions is more notable in the first period whereas the evolution since August 2012 has been marginal. This pattern suggests that in spite of the still significant growth rate of the system the degree distribution across users have become stable. Moreover, it is worth to mention that most popular users that had > 10K followers in December 2011 have increased their popularity one order of magnitude in one year.

This figure also demonstrates that the overall shape of the distributions for the Twitter graph (including around 60M nodes and collected 3 years after Twitter release) are surprisingly similar to our most recent LCC snapshots (including slightly more than 80M nodes and collected shortly 1 year after Google+ release). The most significant difference appears in the distribution of number of friends and is due to the limit of 5K friends imposed by Google+ [10].

Finally, although not directly comparable due to its bidirectional nature, we observe that Facebook users present a significantly higher degree than in Twitter and Google+[12]. Specifically, 56% of Facebook users have more than 100 neighbors, whereas in Google+ and Twitter less than 3.6% and 0.8% of the users present that number of friends and followers, respectively. Then, Facebook shows a denser network. It is also worth noting that Facebook also establish a limit of 5K neighbors, that seems to be used as reference by Google+ to set up its own limit on the number of friends.

**Balanced Connectivity & Reciprocation:** We calculate the percentage of bidirectional relationships across the LCC snapshots that evolves from 32% in Dec 2011 [39] to 22.8% in April 2010, and to 22.6% (21.9%, 21.5%) and 21.3% in August (September, October) and November. We observe again that Google+ LCC seems to have reached a quasi-stable status (with an slight reduction of this metric in the last months) despite of still being growing. Kwak et al. [37] reported 22% bidirectional relationships in a complete Twitter snapshot from July 2009 and in Facebook all the links are bidirectional. These results suggest that in its seemingly mature status Google+ presents a significantly smaller and similar percentage of bidirectional relationships to Facebook and Twitter, respectively.

[11]Note that Facebook forces bidirectional relationships, therefore the distribution for Facebook in both subfigures is the same.

[12]Note that has demonstrated in [46] and shown in Figure 8, Facebook degree distribution does not follow a power law.

In order to gain further insight we now focus on studying individual nodes. In particular, the balance between the in-degree and outdegree of each node coupled with the fraction of bidirectional connections for each node provides a valuable insight about the user level connectivity in OSNs and reveals whether users exchange or simply relay information. To examine the level of balance in the connectivity of individual nodes, Figure 9(a) plots the summary distribution of the ratio of friends to followers (using box plot) for different group of users based on their number of followers for our most recent LCC snapshot (LCC-Nov).

Furthermore, we calculate the percentage of bidirectional relationships for a node u, $BR(u)$, as expressed in Equation 1 where Friend(u) and Follower(u) represent the set of friends and followers for u, respectively. Note that the numerator captures the total number of bidirectional relationships whereas the denominator computes the total number of unique (either uni- or bidirectional) relationships for u.

$$BR(u) = \frac{Friend(u) \cap Follower(u)}{Friend(u) \cup Follower(u)} \quad (1)$$

Figure 9(b) presents the summary distribution of $BR$ for different groups of Google+ users in LCC based on their number of followers for our LCC-Nov snapshot[13].

As expected, popular users (> 10k followers) present a high ratio #friends/#followers between 10 and $10^5$ (depending on the user's popularity) and a percentage of bidirectional relationships close to 0. More interestingly, we observe that low-degree users (< 1K followers), despite of presenting a balanced number of friends and followers, do not often reciprocate since they show between 30 and 45% bidirectional relationships (in median).

Finally, we identified around 5% of LCC users who reciprocate more than 90% of their edges and examined their characteristics to gain a deeper insight. 90% of them present less than 3 friends/followers and less than 5% of them have public posts. Then, we can conclude that these users are typically inactive and poorly connected.

**Clustering Coefficient:** Figure 10 depicts the summary distribution of the undirected version of the clustering coefficient (CC) among LCC users in different snapshots[14] along with the CC for Twitter snapshot from Cha et al. [26]. This figure clearly illustrates that during the roughly one year period (from Dec 2011 to Nov 2012), the CC among the bottom 90% of the users remains below 0.5 and it continuously decreases. On the other hand, the CC for the top 10% of users is very stable. In essence, the Google+ structure has become less clustered as new users joined over the one year period. This same trend has been recently reported for a popular Chinese OSN [49] and then we conjecture that this must be a common trend among popular OSNs. In addition, contrary to previous studied connectivity properties, the difference in the CC between the August and November snapshots cannot be considered marginal. Then it seems that the still significant growth of the network is expected to reduce even further the CC.[DOUBLECHECK WITH THE FINAL VERSION]

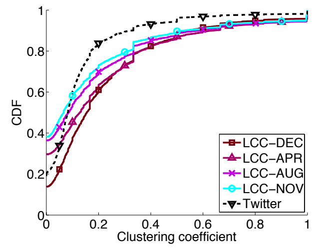



**Figure 10: Clustering Coefficient**

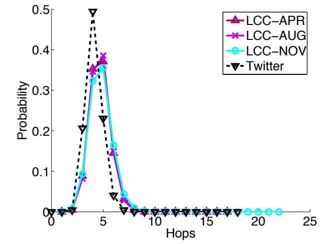



**Figure 11: Average Path Length**

Compared to Twitter where CC is less than 0.3 for 90% of users, Google+ is still more clustered although the existing gap is shrinking as new nodes join Google+. Furthermore, using the approximation presented in [39] we conclude that just 1% of the nodes in a complete Facebook snapshot collected in May 2011 [46] have a CC > 0.2 in comparison with the 10% and 28% in the Twitter and Google+ (November) snapshots. Hence, Google+ is the most clustered network among the three.

Finally, it is interesting to point out the presence of an X% and Y% of users with CC = 0 and 1, respectively. A careful investigation reveal that the former group is formed by users with a very high degree and very low reciprocation whereas the latter group include very low degree nodes that form triangles with its neighbors [DOUBLECHECK IN THE LAST VERSION].

**Path Length:** Figure 11 plots the probability distribution function for the pairwise path length between nodes in different snapshots of Google+ and a snapshot of Twitter [37]. Note that we use the same methodology described in [37, 39] to estimate the path-length distribution using few thousands random samples. We observe that roughly 97-99% of the pairwise paths are between 2 to 7 hops long and roughly 68-74% of them are 4 or 5 hops. The diameter of the Google+ graph is increase from 17 hops (in April) to 22 hops (in November). The two visibly detectable changes in this feature of G+ graph as a result of its growth are: a negligible decrease in typical path length (from April to November) and the increase of its diameter in the same period. Moreover, Table 7 summarizes the Average and Mode Path Length, the diameter and the efficient diameter [38] (*i.e.*, 90 percentile of pairwise path distances) for Google+ (November), Twitter [37] and Facebook [23] snapshots. We observe that Google+ and Facebook exhibit similar Average (and Mode) Path Length but Facebook shows a longer diameter (probably) because its graph is roughly 1 order of magnitude larger than Google+ LCC. Twitter is the network showing the shortest Average (Mode) path length and Diameter. We conjecture that this difference is due to the lack of restriction in the maximum number of friends that allows Twitter users to connect to a larger number of core (high degree) nodes.

*In summary, the careful analysis of the evolution of the connectivity features of Google+ across one year and its comparison with Facebook and Twitter reveal the following insights:*

[13] Note that metrics presented in Figure 9 evolve similarly to previously discussed connectivity properties, that is, the differences across snapshots become marginal since Aug 2012. We do not include results for other snapshots in the paper due to space limitations.

[14] Due to the expensive computational cost of this metric we calculate it for a random selection of 50K nodes.

| | LCC-Nov | FB | Twitter |
|---|---|---|---|
| Path Length (Avg) | 4.7 | 4.7 | 4.1 |
| Path Length (Mode) | 5 | 5 | 4 |
| Eff. Diameter | 6 | - | 4.8 |
| Diameter | 22 | 41 | 18 |

**Table 7: Summary of Path Length and Diameter characteristics for Google+, Facebook and Twitter**

| | num friends | num followers |
|---|---|---|
| num posts | 0.45 | 0.53 |
| num attach. | 0.36 | 0.43 |
| num plusones | 0.4 | 0.56 |
| num comments | 0.35 | 0.5 |
| num reshares | -0.17 | -0.02 |

**Table 8: Ranking Correlation**

*(1) After 15 months from its release, and despite of still growing at an increasing rate, the LCC of Google+ seems to have reached a stable status since some of the most important connectivity characteristics (e.g., degree distribution, reciprocity or average path length) show a marginal statistical evolution in the last 3 months. Google+ has achieved this seemingly stable status shortly after 1 year, presumably much faster than other previous OSNs such as Twitter or Facebook that experienced a much lower growth in their first years. This suggests that OSNs released nowadays are expected to evolve fast to reach a mature status, probably due to the high penetration level of this technology among the users that have become "experts" in the utilization of OSNs. (2) The comparison with Twitter and Facebook shows that Twitter presents stronger similarities with Google+ (degree distribution, percentage of bidirectional relationships or clustering coefficient) than Facebook. This observation reinforces the results obtained from our analysis of Google+ visible activity and confirms the broadcast social media nature of this OSN.*

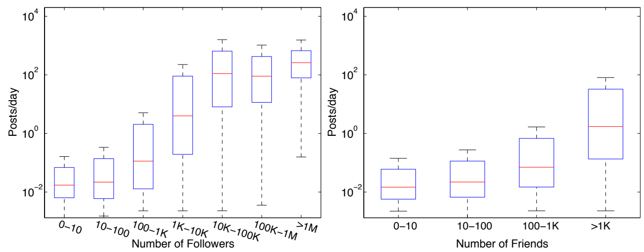
# 7. RELATING USER ACTIVITY & CONNECTIVITY

In earlier sections, we separately characterize different aspects of user activity and connectivity. One interesting question is whether and how different aspects of connectivity and activity of individual users are related. We tackle this question at both broad and more detailed levels.
To determine how correlated the connectivity of a user (#followers, #friends) are with different aspects of its activity (#Posts, #Plusones, #Comments, #Reshares), we compute the Rank Correlation (RC) between all 8 pairs of these properties across active users and show it in Table 8. The results suggest a relatively high correlation between the popularity (#followers) and the activity metrics with values of RC around 0.5. The only exception is the number of reshares that is independent from users' popularity. Furthermore, we observe that the correlation between #friends and activity metrics is smaller.
To take a closer look at the relationship between user connectivity and activity, we examine how the distribution of actions and reactions among a group of users change if we divide users into groups based on their #followers or #friends. The two plots in Figure 12 show the summary distribution of posts/day for different groups of users based on #followers, and #friends using log scale for both axis. Figure 12(a) illustrates that the rate of generated posts by users rapidly increases with their number of followers and the rate of increase is especially large as we move from users with 100-1K followers to those with 10K-100K followers. Figure 12(b) shows that there is also a positive correlation between #friends and rate of posts. However, the rate of increase is much smaller than what we observed for grouping based on #followers in Figure 12(a). observation.
Figure 13 presents the summary distribution of average reaction rate (*i.e.*, for 3 types of reactions) for different group of users based on #followers and #friends. Again, we observe a very strong correlation between the reaction rate to a user and its number of followers especially for users with more than 100 followers. The reaction of users does increase with the number of friends but at a much lower rate. The stronger correlation between #followers and the rate of reaction by others is reasonable since only the followers of a user see her posts (without taking any action) and thus have the opportunity to react.

*In summary, given that users with many followers, that in turn are the most active, have a small fraction of bidirectional edges (as we shown in Section 6), we can confirm that G+ users use the system primarily for broadcasting information as suggested our separated study of activity and connectivity properties in previous sections.*

# 8. RELATED WORK

**OSN characterization:** The importance of OSNs has motivated researchers to characterize different aspects of the most popular OSNs. The graph properties of Facebook [46, 23], Twitter [37, 26] and other popular OSNs [42] have been carefully analyzed. Note that all these studies use a single snapshot of the system to conduct their analysis, instead we analyze the evolution of the Google+ graph over a period of one year. In addition, some other works leverage passive (e.g., click streams) [24, 45] or active [48, 32] measurements to analyze the user activity in different popular OSNs. These papers are of different nature than ours since they use smaller datasets to analyze the behaviour of individual users. Instead, we use a much larger dataset to analyze evolution of the aggregate public activity along time as well as the skewness of the contribution to overall activity across users in Google+. Finally, few works have also analyzed the users' information sharing through their public attributes in OSNs such as Facebook [41].
**Evolution of OSN properties:** Previous works have separately studied the evolution of the relative size of the network elements for specific OSNs (Flickr and Yahoo 360) [36], the growth of an OSN and the evolution of its graph properties [40, 22, 49, 28, 29, 43] or the evolution of the interactions between users [34] and the user availability [25]. In this paper, instead of looking at an specific aspect, we perform a comprehensive analysis to study the evolution of different key aspects of Google+ namely, the system growth, the representative of the different network elements, the LCC connectivity and activity properties and the level of information sharing.
**Google+ Characterization**: Google+ has recently attracted the attention of the research community. Mango et al. [39] use a BFS-based crawler to retrieve a snapshot of




(a) #followers vs Avg. Post Rate

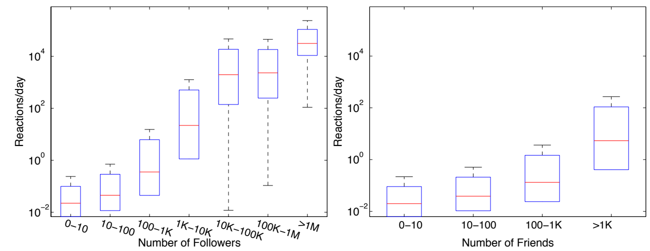
(b) #friends vs Avg. Post Rate

**Figure 12: Correlation between Post Rate and Connectivity (#followers and #friends) properties in Google+**

the Google+ LCC between Nov and Dec 2011. They analyze the graph properties, the public information shared by users and the geographical characteristics and geolocation patterns of Google+. Schiöberg et al. [44] leverage Google's site-maps to gather Google+ user IDs and then crawl these users' information. In particular, they study the growth of the system and users connectivity over a period of one and a half months between Sep and Oct 2011. Unfortunately, as acknowledged by the authors the described technique was anymore available after Oct 2011. Furthermore, the authors also analyze the level of public information sharing and the geographical properties of users and links in the system. Finally, Gong et al. [30] use a BFS-based crawler to obtain several snapshots of the Google+ LCC in its first 100 days of existence. Using this dataset the authors study the evolution of the main graph properites of Google+ LCC in its early stage. Our work presents a broader focus than these previous works since in addition to the graph topology and the information sharing we also analyze (for first time) the evolution of both the public activity and the representativeness of the different network elements. Furthermore, our study of the graph topology evolution considers a 1 year window between Dec 2011 and Nov 2012 when the network is significantly larger and present important differences to its early status that is the focus of the previous works. In another interesting, but less related work, Kairam et al. [35] use the complete information for more than 60K Google+ users (provided by Google+ administrators) and a survey including answers from 300 users to understand the selective sharing in Google+. Their results show that public activity represents 1/3 of the Google+ activity and that an important fraction of users make public posts frequently. Finally, other papers have study the video telephony system of Google+ [47], the public circles feature [27] and collaborative privacy management approaches [33].

## 9. CONCLUSION

In this paper we study the evolution of the key features of the last major player released in the OSN market,*i.e.*, Google+. For this purpose we leverage (to the best of our knowledge) one of the largest collection of datasets used to characterize an specific OSN so far. These datasets include information related to the connectivity, activity and information sharing properties of Google+ over a period of more than one year.




(a) #followers vs Avg. Reaction rate

(b) #Friends vs Avg. Reaction rate

**Figure 13: Correlation between Aggregate Reaction Rate and Connectivity (#followers and #friends) properties in Google+**

The careful analysis of this data reveals the following main insights:
(1) Contrary to some widespread opinion we cannot claim that G+ is a "ghost town". First, the number of interested users (*i.e.*, those joining the LCC) is growing at an increasing rate. However, this rate is lower than the one depicted by official statistics that most likely consider users that are forced to join Google+ as consequence of the integrated registration of Google in their ciphers.Second, the overall visible activity (actions and reactions) is steadily growing, what is a positive indicator about the level of engagement of users.
(2) Despite of this (still) important growth, Google+ seem to have reach an statistical stability in both the connectivity and activity properties that suggest that the network has reached a mature status in slightly more than 1 year. This is a clear indicator that "new" generation OSNs are expected to deploy and define its "signature" (*i.e.* , main features) in a short period of time presumably much faster than first generation OSNs.
(3) In this seemingly mature status our detailed analyses of connectivity and activity properties as well as the correlation between both reveal that Google+ is a broadcast social media system in which a relative small group of popular and very active users sustain (at least) the visible activity of the system and attract most users reactions.

## 10. REFERENCES

[1] `http://google-plus.com/category/statistics/`.
[2] `http://online.wsj.com/article/SB10001424052970204653604577249341403742390.html?mod=WSJ_hp_LEFTTopStories`.
[3] `http://en.wikipedia.org/wiki/Google+`.
[4] `https://plus.google.com/107117483540235115863/posts/YUniwagZuKZ`.
[5] *Announcement of New Feature for Google+.* `http://www.workinghomeguide.com/9918/google-new-features-hashtag-auto-complete-text-to-photos-and-video-status`.
[6] *Announcement of New Feature to add text to photos in Google+.* `https://plus.google.com/107814003053721091970/posts/D7gfxe4bU7o`.
[7] *Buzz shut down announcement.* `https://support.google.com/mail/bin/answer.py?hl=en&answer=1698228`.
[8] *Flow Chart for Google+ Sharing.*

http://googleplushowto.com/2011/07/will-user-a-see-my-post-in-google-plus/.

[9] *Genealogy Data: Frequently Occurring Surnames from Census 2000. US Census Bureau.* http://www.census.gov/genealogy/www/data/2000surnames/index.html.

[10] *Google+ (maximum number of contacts in your circles).* http://support.google.com/plus/bin/answer.py?hl=en&answer=1733011.

[11] *Google+ official learn more site.* http://www.google.com/intl/en/+/learnmore/.

[12] *Google+ official support site.* http://support.google.com/plus/.

[13] *Google+ Pages.* http://www.google.com/+/business/.

[14] *Google+ Pages announcement.* http://googleblog.blogspot.com/2011/11/google-pages-connect-with-all-things.html.

[15] *Google Privacy Policy.* http://www.google.com/policies/privacy/.

[16] *Google+ Redesign.* http://mashable.com/2012/04/11/google-plus-redesign/.

[17] *Google Registration. Arstechnica.* http://arstechnica.com/gadgets/2012/01/google-doubles-plus-membership-with-brute-force-signup-process/.

[18] *Google Registration. Blogspot.* http://googlesystem.blogspot.com/2012/01/new-google-accounts-require-gmail-and.html.

[19] *Google wave shut down announcement.* http://support.google.com/bin/answer.py?hl=en&answer=1083134.

[20] *Orkut Official Site.* http://www.orkut.com.

[21] *Orkut Statistics, Wikipedia.* http://en.wikipedia.org/wiki/Orkut.

[22] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *WWW*, 2007.

[23] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four Degrees of Separation. *CoRR*, abs/1111.4570, 2011.

[24] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *ACM IMC*, 2009.

[25] A. Boutet, A.M. Kermarrec, E. Le Merrer, and A. Van Kempen. On the impact of users availability in osns. In *ACM SNS*, 2012.

[26] M. Cha, H. Haddadi, F. Benevenuto, and K.P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *AAAI ICWSM*, 2010.

[27] L. Fang, A. Fabrikant, and K. LeFevre. Look who i found: Understanding the effects of sharing curated friend groups. In *ACM WebSci*, 2012.

[28] S. Gaito, M. Zignani, G.P. Rossi, A. Sala, X. Wang, H. Zheng, and B.Y. Zhao. On the bursty evolution of online social networks. In *ACM KDD HotSocial Workshop*, 2012.

[29] S. Garg, T. Gupta, N. Carlsson, and A. Mahanti. Evolution of an online social aggregation network: an empirical study. In *ACM IMC*, 2009.

[30] N.Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song. Evolution of attribute-augmented social networks: Measurements, modeling, and implications using google+. In *ACM IMC*, 2012.

[31] Z. Govindarajulu. Rank correlation methods. *Technometrics*, 34(1):108–108, 1992.

[32] L. Gyarmati and T.A. Trinh. Measuring user behavior in online social networks. *Network, IEEE*, 24(5):26–31, 2010.

[33] H. Hu, G.J. Ahn, and J. Jorgensen. Enabling collaborative data sharing in google+. In *IEEE Globecom*, 2012.

[34] Jing Jiang, Christo Wilson, Xiao Wang, Peng Huang, Wenpeng Sha, Yafei Dai, and Ben Y. Zhao. Understanding Latent Interactions in Online Social Networks. In *ACM IMC*, 2010.

[35] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi. Talking in circles: Selective sharing in google+. In *ACM CHI*, 2012.

[36] R. Kumar, J. Novak, and A Andtomkins. Structure and evolution of online social networks. In *ACM KDD*, 2006.

[37] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a Social Network or a News Media? In *WWW*, 2010.

[38] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *ACM SIGKDD*, 2005.

[39] Gabriel Magno, Giovanni Comarela, Diego Saez-Trumper, Meeyoung Cha, and Virgilio Almeida. New kid on the block: Exploring the google+ social graph. In *ACM IMC*, 2012.

[40] A. Mislove, H.S. Koppula, K.P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *WOSN*, 2008.

[41] A. Mislove, B. Viswanath, K.P. Gummadi, and P. Druschel. You are who you know: inferring user profiles in online social networks. In *ACM WSDM*, 2010.

[42] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and Analysis of Online Social Networks. In *ACM IMC*, 2007.

[43] R. Rejaie, M. Torkjazi, M. Valafar, and W. Willinger. Sizing Up Online Social Networks. *IEEE Network*, 24(5):32 –37, Sept-Oct 2010.

[44] D. Schiöberg, F. Schneider, H. Schiöberg, S. Schmid, S. Uhlig, and A. Feldmann. Tracing the birth of an osn: Social graph and profile analysis in google. In *ACM WebSci*, 2012.

[45] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In *ACM IMC*, 2009.

[46] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The Anatomy of the Facebook Social Graph. *CoRR*, abs/1111.4503, 2011.

[47] Y. Xu, C. Yu, J. Li, and Y. Liu. Video telephony for end-consumers: Measurement study of google+, ichat, and skype. In *ACM IMC*, 2012.

[48] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user

posting behavior on social media. In *ACM SIGIR*, 2012.

[49] X. Zhao, A. Sala, C. Wilson, X. Wang, S. Gaito, H. Zheng, and B.Y. Zhao. Multi-scale dynamics in a massive online social network. In *ACM IMC*, 2012.