Assessing The Evolution of Google+ in its First Two Years

Roberto Gonzalez*, Ruben Cuevas†, Reza Motamedi‡, Reza Rejaie‡, Angel Cuevas†
*NEC Europe Ltd.
roberto.gonzalez@neclab.eu
†Universidad Carlos III de Madrid
{rcuevas, acrumin}@it.uc3m.es
‡University of Oregon
{motamedi, reza}@cs.uoregon.edu

***Abstract*—In the era when Facebook and Twitter dominate the market for social media, Google has introduced Google+ (G+) and reported a significant growth in its size while others called it a ghost town. This begs the question that "whether G+ can really attract a significant number of connected and active users despite the dominance of Facebook and Twitter?".**

This paper presents a detailed longitudinal characterization of G+ based on large scale measurements. We identify the main components of G+ structure, characterize the key feature of their users and their evolution over time. We then conduct detailed analysis on the evolution of connectivity and activity among users in the largest connected component (LCC) of G+ structure, and compare their characteristics with other major OSNs. We show that despite the dramatic growth in the size of G+, the relative size of LCC has been decreasing and its connectivity has become less clustered. While the aggregate user activity has gradually increased, only a very small fraction of users exhibit any type of activity and an even smaller fraction of these users attract any reaction. The identity of users with most followers and reactions reveal that most of them are related to high tech industry. To our knowledge, this study offers the most comprehensive characterization of G+ based on the largest collected data sets.

***Index Terms*—Online Social Networks, Google+, Measurements, Characterization, Evolution**

I. INTRODUCTION

A significant majority of today's Internet users rely on Facebook and Twitter for their online social interactions. In June of 2011, Google launched a new Online Social Network (OSN), called *Google+* (or *G+* for short) in order to claim a fraction of social media market and its associated profit. G+ offers a combination of Facebook- and Twitter-like services in order to attract users from both rivals. There has been several official reports about the rapid growth of G+ user population (540M active users in Oct 2013) [11] while some observers and users dismissed these claims and called G+ a "ghost town" [1]. This raises the following important question: *"Can a new OSN such as G+ attract a significant number of engaged users and become a relevant player in the social media market?"*. A major Internet company such as Google with many popular services, is perfectly positioned to implicitly or explicitly require (or motivate) its current users to join its OSN. Then, it is interesting to assess to what extent and how Google might have leveraged its position to make users join G+. Nevertheless, any growth in the number of users in an OSN is really meaningful only if the new users adequately connect to the rest of the network (*i.e.,* become connected) and become active by using some of the offered services by the OSN on a regular basis. We also note that today's Internet users are much more savvy about using OSN services and connecting to other users than users a decade ago when Facebook and Twitter became popular. This raises another related question: *"how does the connectivity and activity of G+ users evolve over time as users have become significantly more experienced about using OSNs?" and "whether these evolution patterns exhibit different characteristics compared to earlier major OSNs?"*. These evolution patterns could also offer an insight on whether users willingly join G+ or are added to the system by Google.

In this paper, we present a comprehensive measurement-based characterization of connectivity and activity among G+ users and their evolution during the first two years after its release in order to shed an insightful light on all the above questions. We start by providing a brief overview of G+ in Section II. One of our contributions is our measurement methodology to efficiently capture complete snapshots of G+'s largest connected component (LCC), several large sets of randomly selected users, and all the publicly-visible activities (*i.e.,* user posts) of LCC users with their associated reactions from other users. To our knowledge, this is one of the largest and more diverse collection of datasets used to characterize an OSN. We describe our datasets in Section III along with our measurement methodology and validation techniques.

In Section IV, using our LCC snapshots, we characterize the evolution of LCC size during the first two years of its operation. Furthermore, we leverage the randomly selected users to characterize the relative size of the main components (*i.e.,* LCC, small partitions, and singletons) of G+ network and the evolutions of their relative size over time along with the fraction of active users and users with publicly visible attributes in each component. Our results show that while the size of LCC has increased at an impressive rate over the first two years of system operation, its relative size has consistently decreased such that the LCC users currently make up only 27% of the network and the rest of the users are mostly singletons. The large and growing fraction of singletons appears to be caused by Google's integrated registration process that implicitly creates a G+ account for any new Google account regardless of the user's interest. Furthermore, we discover that LCC users generate most of the public posts and provide a larger number of attributes in their profile. Since LCC users form the most important component of G+ network, we focus the rest of our analysis on LCC.

We then turn our attention to the publicly visible activity of LCC users and its evolution during the entire lifetime of G+ in Section V. We discover that the aggregate number of posts by LCC users and their reactions (namely comments, plusones or reshares) from other users have been steadily growing over time. Furthermore, a very small fraction of LCC users generate posts and the post from an even smaller fraction of these users receive most of the reactions from other users, *i.e.,*user actions and reactions are concentrated around a very small fraction of LCC users. The average number of daily active users is growing around 670 users per day and only 17% of LCC users have ever become active. The comparison of user activity among G+, Twitter and Facebook reveals that G+ users are significantly less active than other two OSNs. More specifically, the number of G+ users who have ever become active during the first two years after the release of the system is 2.3 and 8.6 times smaller than that in MySpace and Twitter, respectively. In Section VI, we focus on the percentage of users making individual attributes in their profile publicly available. We also show that users are generally more willing to make their professional attributes publicly available but the fraction of such users has continuously decreased.

Finally, we explore the evolution of connectivity features of LCC in Section VII and show that many of its features have initially evolved but have stabilized in recent months despite the continued significant growth in its population. Interestingly, many connectivity features of the G+ network have a striking similarity with the same features in Twitter but are very different from Facebook. More specifically, the fraction of reciprocated edges among LCC users are small (and mostly associated with low degree and non-active users) and the LCC network has become increasingly less clustered. Furthermore, we observe a strong positive correlation between the user popularity (*i.e.,* number of followers) and the user rate of posts or the rate of reactions by other G+ users to those posts in Section VIII. In summary, the similarity of connectivity features for G+ and Twitter network coupled with the concentration of posting activity (and their reaction) on a small fraction of popular users and the small fraction of bidirectional relationships in LCC indicate that G+ network is primarily used for broadcasting information. Section IX summarizes the prior related work and we conclude the paper in Section X.

II. Google+ Overview

After a few unsuccessful attempts (Buzz [7], Wave [20] and Orkut [21], [22]), Google launched G+ 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. On September 20^{th}, G+ became open to public and the G+ Pages service was launched on November 7^{th} 2011 [14], [15]. 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) [18], [19].

G+ features have some similarity to Facebook and Twitter. Similar to Twitter (and different from Facebook) the relationships in G+ are unidirectional. More specifically, user A can follow user B in G+ 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 it properly[1].

Each user has a stream (similar to Facebook wall) where any activity performed by the user appears. 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 to resend a post to their followers.

G+ assigns a numerical user ID and a profile to each user. The inferred strategy for assigning user ID by G+ is as follows, each 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 impractical. Similar to other OSNs, G+ users have a profile that has 21 fields where they can provide a range of information and pointers (*e.g.*, to their other pages) about themselves. However, providing this information is not mandatory (except for the sex) 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 (even for the sex) by defining them as "private" and thus visible to a specific group[2]. For a more detailed description of G+ functionality we refer the reader to [12], [13].

III. Measurement Methodology and Datasets

This section presents our techniques for data collection (and validation) and then a summary of our datasets that we use for our analysis[3].

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

[2] Note that it is not possible to distinguish whether a non visible attribute is private or not specified by the user.

[3] The collected datasets are publicly available upon request.

Name	#nodes	#edges	Start Date	Duration (days)
LCC-Dec11*	35.1M	575M	11/11/12	46
LCC-Apr12	51.8M	1.1B	15/03/12	29
LCC-Aug12	79.2M	1.6B	20/08/12	4
LCC-Sep12	85.3M	1.7B	17/09/12	5
LCC-Oct12	89.8M	1.8B	15/10/12	5
LCC-Nov12	93.1M	1.9B	28/10/12	6
LCC-Dec12	105.1M	2.2B	12/11/12	8
LCC-Jan13	119.8M	2.5B	28/12/12	9
LCC-Feb13	134.8M	2.8B	11/02/13	10
LCC-Mar13	149.0M	3.0B	21/03/13	11
LCC-Apr13	155.1M	3.1B	12/04/13	11
LCC-May13	173.1M	3.5B	22/05/13	12
LCC-Jul13	190.0M	3.8B	12/07/13	13

TABLE I
MAIN CHARACTERISTICS OF LCC SNAPSHOTS

Name	#nodes	#edges	Start Date	Duration (days)
Rand-Apr12	2.2M	145M	08/04/12	23
Rand-Oct12	5.7M	263M	15/10/12	10
Rand-Nov12	3.5M	157M	28/10/12	13
Rand-Jan13	5.0M	321M	08/01/13	8
Rand-Mar13	1.1M	77M	15/03/13	4
Rand-Apr13	3.6M	249M	23/04/13	7
Rand-Jul13	3.0M	234M	12/07/13	8

TABLE II
MAIN CHARACTERISTICS OF RANDOM DATASETS

Users	Postings	Attachments	Plusones	Replies	Reshares
32.5M	541M	444M	1B	408M	140M

TABLE III
MAIN CHARACTERISTICS OF ACTIVITIES AMONG ACTIVE USERS IN LCC (COLLECTED IN JUL-OCT 2013)

Label	OSN	Date	Info
TW-Pro	Twitter	Jul 2011	Profile (80K rand. users)
TW-Con [29]	Twitter	Aug 2009	Connectivity (55M users)
TW-Act [48]	Twitter	Jun 2010	Activity (895K rand. users)
MS-Act [48]	MySpace	Jan 2010	Activity (239k rand. users)
FB-Pro	Facebook	Jun 2012	Profile (480K rand. users)
FB-Con	Facebook	Jun 2012	Connectivity (75K rand. users)
FB-Act	Facebook	Sep 2012	Activity (16k rand. users)

TABLE IV
FEATURES OF OTHER DATASETS IN OUR ANALYSIS

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 as follows: We divide the ID space into 21 equal-size zones since we had 21 servers available to use as crawlers and assign each server 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. The 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 16.8M users per day for the whole system[4]. With this rate, it takes 4-13 days to capture a full snapshot of the LCC connectivity and users' profiles. Table I summarizes the main characteristics of our LCC datasets. We obtained the LCC-Dec11 snapshot from an earlier study on G+[44]. We examined the connectivity of all the captured LCC snapshots and verified that all of them form a single connected components.

Sampling Random Users: Our goal is to collect random samples of G+ users for our analysis. To our knowledge, none of the prior studies on G+ 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 search function of the G+ API to efficiently identify a large number of seemingly random users. The function provides a list of up to 1000 users whose name or surname matches a given input keyword. Careful manual inspection of the search results revealed that G+ appears to order the returned list of users based on their level of connectivity and activity, *i.e.,* users with a larger number of connections or a higher level of activity (that are likely to be more interesting) are placed at the top of the returned list. Since searching for popular surnames most likely results in more than 1000 matched users, the subset of returned users represent biased samples since they are more connected and/or active. To avoid this bias, we selected a collection of more than 36K American surnames from the US[5] 2000 census [9] with low to moderate popularity and used the search function of the API to obtain matched G+ users. We consider the list of returned users only if it contains less than 1000 users. These users are assumed to be random samples because G+ must return all matched users for the input surname . Moreover, intuitively there should not be any correlation between surname popularity and the level of connectivity (or activity) of the corresponding users. Table II 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 use the search API to collect more than 140K users in two groups, those whose name match popular and unpopular (*i.e.,* with less than 1000 matching results) surnames in Sep 2012. We focus on samples from each group that are located in the LCC using a complete snapshot of the LCC that serves as the ground truth. In particular, we compare the connectivity of samples from each group that are located in LCC with all users in LCC-Sep12 snapshot.

[4]LCC-Apr12 snapshot was collected before this tuning and therefore took longer.

[5]US is the most represented country in G+ [44], [49]. Furthermore, the high immigration level of US allows to find surnames from different geographical regions.

Figure 1 plots the 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 distribution 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.

Capturing User Activity: We consider user activity as a collection of all posts by individual users and the reaction (*i.e.,* Plusones, Comments and Reshares) from other users to these posts. 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 G+ users. Toward this end, we focus on user activity in the most important element of the network (*i.e.,* the LCC). We leverage the G+ API to collect all the public posts and their associated reactions for all LCC-Jul13 users between G+ release date (Jun 28^{th} 2011) and the date our measurement campaign started (Jul 3^{th} 2013), *i.e.,* roughly 2 years. 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 posts from all users. Note, that G+ API limits the number of daily queries to 10K per registered application. Then, we use 604 accounts to collect the referred data in 94 days. Table III summarizes the main features of the activity dataset. In particular, note that only 32.4M (out of 190M) users in LCC-Jul13 made at least one public post in the analysis period.

Other datasets: There are a few other datasets for Twitter, Facebook and MySpace that we have either collected or obtained from other researchers. Table IV summarizes the main features of these datasets. In the absence of any public dataset for Facebook, we developed our own crawler and collected the profile (FB-Pro), connectivity (FB-Con) and activity (FB-Act) information for random Facebook users. We also collect the profile (TW-Pro) for random Twitter users. In the case of Facebook, we leverage its directory of publicly available profiles[6] to find random samples. This directory is organized in a tree structure per letter. Browsing through the tree structure, we obtain the total number of registered users for each letter and generate an index with all public accounts ranging from 1 (first account in letter A) to N (last account in letter Z). Our random sample of Facebook users is the result of a random selection of users from that index. To identify random samples of Twitter users we use the methodology described in [48]. At the time of our measurement, user IDs for new Twitter users were assigned in a monotonically increasing manner, *i.e.,*account x has a larger ID than account y if x is created after y. We monitor the public time-line, and capture the creation time of accounts associated with sample tweets from the public timeline. We identify an account x that has been created in the last few days, double the x's ID and use it as a conservative estimate for the maximum value of user ID. We then generate random numbers within the identified range of user IDs and check whether an account exist for each randomly selected ID. This allows us to filter out invalid IDs that are associated with the deleted accounts or unassigned numbers. The resulting valid accounts from this process provide random sample of users.

Limitations: Our methodology has some limitation due to the lack of access to the information that is configured as private by some users. In particular, the privacy setting of users could prevent our crawler from capturing the following information: *(i)* If two connected users A and B both set their connecting link as private, we refer to such a link as a *private link*. Private links are not publicly visible and thus can not be captured by our crawler; *(ii)* If a LCC user A sets all of her links as private and all of A's LCC neighbors also set their link to A as private, then our methodology mis-classifies user A as singleton. In fact, 7.5% of the discovered users in our BFS crawl of the LCC have private list of friends and followers, and are discovered through their neighbors. Therefore, we believe this limitation is uncommon and does not lead to a significant error in our captured snapshots of LCC; *(iii)* Private posts of individual users are not captured by our crawlers. We further discuss this issue in Section V. Note that we are not aware of any known technique to overcome these limitations.

IV. Macro-Level Structure & Its Evolution

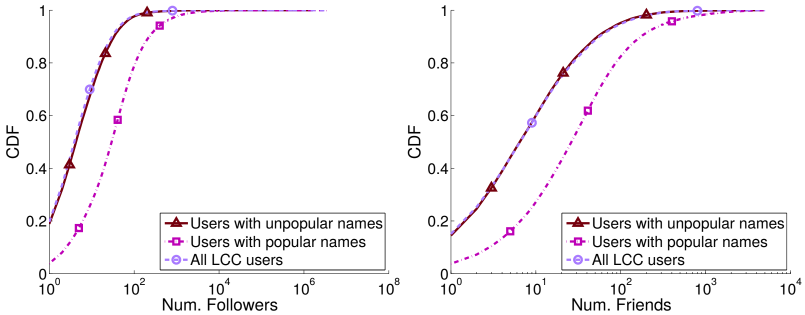
The macro-level connectivity structure among G+ users should intuitively consist of three components: (i) The largest connected component (LCC), (ii) A number of partitions that are smaller than LCC (with at least 2 users), and (iii) Singletons or isolated users. We first examine the temporal evolution of LCC size and then discuss the relative size of different components and their evolution over time.

Evolution of LCC Size: Having multiple snapshots of the LCC at different times enables us to examine the growth in the number of LCC users over time and determine the number of users who depart or arrive between two consecutive snapshots as shown in Figure 2 using log scale for the y axis. This figure illustrates that the overall size of the LCC has increased from 35M to 105M during 2012 at an average growth rate of 176K users per day. This average rate has even increased to 350K users per day during the first half of 2013 resulting on an average growth rate of 263K users per day during the whole studied period (Dec 2011- Jul 2013).

The connectivity of these users to LCC is a clear sign that they have intentionally joined G+ by making the explicit effort to connect to other users (*i.e.,* these are interested users). While the average daily increase of 263K new interested users is impressive, it is 60% smaller than the average ∼650K daily new users registered in G+ between July 2011 and October 2013 that are officially reported by Google [2]. The difference between the rate of growth for the overall system and LCC must be associated with other components of the network (small partitions and singletons) as we explore later in this section.

Figure 2 also shows that LCC users have been departing the LCC at an average rate of 10.1K users per day. We carefully

[6]https://www.facebook.com/directory/people/



(a) Num Followers (b) Num Friends

Fig. 1. Distribution of #followers (a) and #friends (b) for users collected from the search function of G+ API with popular surnames (>1000 users), users collected with unpopular surnames (< 1000 users), and all LCC users.

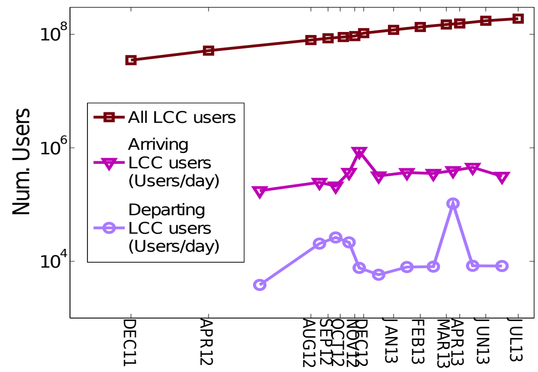


Fig. 2. Evolution of total size and #arriving and #departing LCC users over time

examined these departing users and discovered two points: *(i)* all of the departing users have removed their G+ accounts, and *(ii)* the distribution of #followers, #friends and public attributes of departing users is very similar to all LCC users, however most of them are not active. This seems to suggest that the departing users have lost their interest due to the lack of incentives to actively participate in the system.

Evolution of the Main Components: To estimate the relative size of individual components and their evolution over time, we determine the mapping of users in a random dataset to the three main components of the G+ structure. The LCC users can be easily detected using the corresponding LCC snapshot for each random data set (*e.g.*, LCC-Oct12 for Rand-Oct12). For all the users outside the 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. Table V(a) presents the relative size of all three components using our random datasets in Apr, Oct and Nov 2012 and Jan, Mar, Apr and Jul 2013. The results show that the relative size of LCC has dropped from 43% (in Apr12) to 27% (in Jul13) while the relative size of singletons has increased from 55% to 69% during the same period. Note that this drop in the relative size of LCC occurs despite the dramatic increase in the absolute size of LCC (as we reported earlier). This simply indicates an even more significant increase in the absolute number of singletons. We believe that this huge increase in the number of singletons is a side effect of the integrated registration procedure that Google has implemented. In this procedure, a new G+ account is implicitly created for any user that creates a new Google account to utilize a specific Google service such as Gmail or YouTube[7]. The implicit addition of these new users to G+ suggests that they may not even be aware of (or do not have any interest in) their G+ accounts.

The relatively small and decreasing size of LCC for G+ network exhibits a completely different characteristic that was reported for LCC of other major OSNs during their growth. For instance, 99.91% of the registered Facebook users were part of LCC as of May 2011 [51] and LCC of Twitter reported to include 94.8% of the users with just 0.2% Singletons in August 2009 [29]. Furthermore, Leskovec et al. [42] showed that the relative size of the LCC of other social networks (*e.g.*, the arXiv citation graph or an affiliation network) typically increases with time until it contains more than 90% of their users. Partitions make up only a small fraction (1.5%) of all G+ users. We identified tens of thousands of such partitions and discovered that 99% of these partitions have less than 4 users in all snapshots. The largest partition was detected in Rand-Apr13 snapshot with 52 users.

Tables V(b) and V(c) present the fraction of all G+ users that have any public posts or provide any public attributes in their profiles and the breakdown of these two groups across different components of G+ network, respectively. We observe that the fraction of all users that generate any post dropped from 10% to 8% during 2012 but remained stable during 2013, and the majority of them are part of LCC. Similarly, the fraction of users with any public attributes have dropped from roughly 30% to 14.2% over the same period. A large but decreasing fraction of active users and users with public attributes are part of LCC and a smaller but growing fraction of them are singletons. While the fraction of active singleton users is much smaller than LCC users, having any activity among singletons is rather intriguing since they do not have any social ties. Our examination revealed that the level of activity among active singletons is very low where 60% of them have published a single post and 95% of them published less than 10 posts since they created their accounts. To gain more insight about the purpose of posts by these accounts, we manually inspected all the posts by the 50 most active singletons. We learned that these accounts are used to record blog entries, uploaded videos to Youtube or even as event pages. Therefore, interested users access this information without establishing any social tie with these accounts. Since the LCC is the well connected component that contains the majority of active users, we focus our remaining analysis only on the LCC.

In summary, the absolute size of LCC in G+ network has been growing by 150-350K users/day while its relative size has been decreasing. This is primarily due to the huge increase in the number of singletons that is caused by the implicit addition of new Google account holders to G+. In July of 2013, the LCC made up 25% and the rest of the network mostly consists of singletons. Around 8% of G+ users generate any post, and

[7]In fact, we examined and confirmed this hypothesis for new Gmail and YouTube accounts.

(a) Fraction of total G+ users in each component

Element	% users						
	Apr12	Oct12	Nov12	Jan13	Mar13	Apr13	Jul13
LCC	43.5	32.3	32.2	28.1	28.0	27.4	26.9
Partitions	1.4	1.7	1.5	3.6	3.1	3.7	4.2
Singletons	55.1	66.0	66.3	68.3	69.0	69.0	68.9
All	100	100	100	100	100	100	100

(b) Fraction of total G+ users who are active in each component

Element	% active users (at least 1 public post)						
	Apr12	Oct12	Nov12	Jan13	Mar13	Apr13	Jul13
LCC	8.9	7.0	6.9	5.8	5.9	5.7	5.7
Partitions	0.1	0.2	0.2	0.4	0.3	0.4	0.4
Singletons	1.4	1.6	1.6	1.6	1.8	1.8	1.9
All	10.4	8.8	8.7	7.8	8.0	7.9	8.0

(c) Fraction of total G+ users with public attributes in each component

Element	% users with at least 1 public attribute						
	Apr12	Oct12	Nov12	Jan13	Mar13	Apr13	Jul13
LCC	27.4	17.9	17.6	13.7	13.0	12.4	11.2
Partitions	0.5	0.6	0.5	0.1	0.7	0.9	0.8
Singletons	1.8	5.7	6.2	3.0	2.75	2.6	2.2
All	29.7	24.2	24.3	17.6	16.4	16.0	14.2

TABLE V

FRACTION OF G+ USERS (A), ACTIVE USERS (B) AND USERS WITH PUBLIC ATTRIBUTES (B) ACROSS G+ COMPONENTS ALONG WITH THE EVOLUTION OF THESE CHARACTERISTICS FROM APRIL 2012 TO JULY OF 2013 (BASED ON THE CORRESPONDING RANDOM DATASETS)

less than 15% provide any public attribute, and a majority of both groups are LCC users.

V. PUBLIC ACTIVITY & ITS EVOLUTION

To investigate user activity, we characterize publicly visible (or in short ”public”) posts by LCC users as well as other users’ reactions (including users outside LCC) to these public posts[8]. An earlier study used ground-truth data to show that more than 30% of posts in G+ were public during the initial phase of the system [39]. However, the proposed setting by Google encourages users to generate public posts and reactions since only these public activities are indexable by search engines (including Google), and thus visible to others (apart from Google) for various marketing and mining purposes [16]. Therefore, characterizing public posts and their reactions provides an important insight about the publicly visible part of G+.

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” (+1) or “reshare” the post with their own followers. To maintain the desired crawling speed for collecting activity information, we decided to only collect the timestamps for individual posts (but not for reactions to each post). Therefore, 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 examined the timestamp of 4M comments associated to 700K posts and observed that more than 80% of the comments occurred within the 24 hours after their corresponding post.

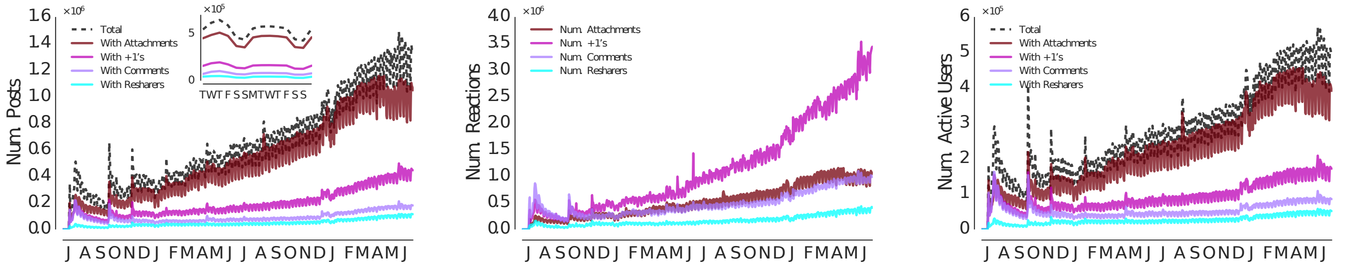
Temporal Characteristics of Public Activity: Having the timestamp for all the posts and their associated reactions enables us to examine the temporal characteristics of all public activity among LCC users during the first 2 years of G+ operation.

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 combination of attachments, plusones, reshares and comments (*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 G+ users that is significantly lower during the weekend and much higher during weekdays as the smaller plot in Figure 3(a) shows. The timing of most of the observed peaks in this (and other related) figure(s) appears to be perfectly aligned with specific events as follows[9]: *(i)* the peak on Jun. 30 2011 caused by the initial release of the system (by invitation) [3]; *(ii)* the peak on Jul. 11 2011 is due to users reaction to a major failure on Jul 9 when G+ system ran out of disk [4]; *(iii)* the peak on Sep. 20 2011 caused by the public release of the system [3]; *(iv)* the peak on Nov. 7 2011 is due to the release of G+ Pages service [15]; *(v)* the peak on Jan. 17 2012 is caused by the introduction of new functionalities for auto-complete and adding text in photos [5], [6]; and *(vi)* the peak on Apr. 12 2012, caused by a major redesign of G+[17]. 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). We can observe that a significant majority of the posts have attachments but the fraction of posts that trigger any reaction from other users is much smaller and plusones is the most common type of reaction.

Note that Figure 3(a) presents the number of daily posts

[8] We are not aware of any technique to capture private posts in G+ for obvious reasons. It might be feasible to create a G+ account and connect to a (potentially) large number of users in order to collect their private posts. However, such a technique is neither representative nor ethical.

[9] We could not identify any significant event at the time of the peaks on May 3rd, Jun 4th and Aug 7th 2011

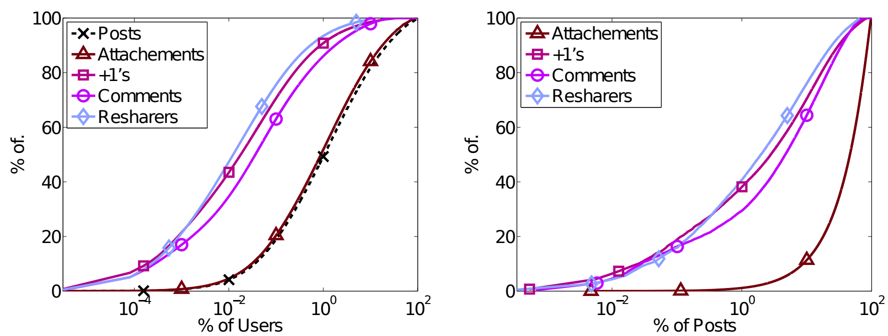


(a) The num. of daily posts, and num. of daily posts with different types of reactions

(b) The num. of daily attachments, plusones, reshares and comments.

(c) The num. of daily active users and num. of daily active users whose posts received each type of reaction

Fig. 3. Evolution of different aspects of public user activity during the 2 years operation of G+ (July 2011 to June 2013)



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

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

Fig. 4. Skewness of actions and reactions contribution per user and post

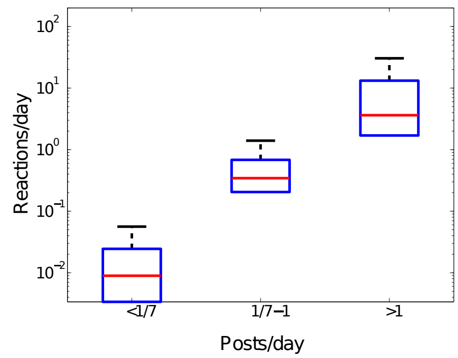


Fig. 5. Post-rate (x axis) vs aggregate reaction rate (y axis) correlation

with attachment 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 the aggregate daily rate of 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 specially plusone reactions have been rapidly growing after the initial phase. Figure 3(b) illustrates that individual posts mostly have single attachment and they are more likely to receive multiple plusones rather than any other type of reaction. 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 whose posts have attachments or triggered at least one type of reaction. This figure reveals that the total number of daily active users with a public post has been steadily growing (after the initial phase) roughly at the rate of 670 users per day. However, this rate of growth in daily active users is significantly (roughly 392 times) lower than the daily rate of new users joining the LCC of G+. While a large fraction 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 below 200K, 100K and 50K, respectively, despite the dramatic growth in the number of LCC users.

Skewness in Activity Contribution: We observed that a relatively small and stable number of users with interesting posts receive most reactions. This raises the question that "how skewed are the distribution of generated posts and associated reactions among users in G+?". Figure 4(a) presents the fraction of all posts in our activity dataset that are generated by the top $x\%$ of LCC users during the life of G+ (the x axis has a log-scale). Other lines in this figure show the fraction of all attachments, plusones, comments and reshares that are associated with the top $x\%$ of LCC users that receive most reactions of each type. 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 82.7% of posts. Furthermore, the distribution of contribution of received reactions to a user's posts is an order of magnitude more skewed than the contribution of total posts per user. In particular, 1% of users receive roughly 86% of comments and 91% of plusones and reshares. These findings offer a strong evidence that *only a very small fraction of the active users (around 5M) create most posts and even a smaller fraction of these users receive most reactions from other users to their posts, i.e., both user action and reaction are centered around a very small fraction of users.* We also repeated a similar analysis at the post level to assess how skewed are the number of reactions to individual posts. Figure 4(b) shows the fraction of attachments, plusones, comments and reshares associated to the top $x\%$ posts. The distribution for attachments is rather homogeneous which indicates that most posts have one or a small number of attachments. For other types of reactions, the distribution is roughly an order of magnitude less skewed than the distribution of reaction across users (Figure 4(a)) .This is a rather expected result since reactions tend to spread across

	All	Top 10%	Top 1%
posts vs. plusones	0.49	0.45	0.30
posts vs. comments	0.39	0.34	0.23
posts vs. reshares	0.4	0.42	0.30
plusones vs. comments	0.55	0.74	0.86
plusones vs. reshares	0.46	0.70	0.85
comments vs. reshares	0.39	0.58	0.78

TABLE VI
RANK CORRELATION BETWEEN ACTIONS (POSTS) AND REACTION (PLUSONES, COMMENTS, RESHARES) AS WELL AS BETWEEN DIFFERENT TYPES OF REACTIONS ASSOCIATED TO ACTIVE USERS FOR ALL THE USERS AND THE TOP 10% AND 1% MOST ACTIVE USERS.

different posts by a user.

Correlation Between User Actions and Reactions: Our analysis so far has revealed that actions and reactions are concentrated on a small fraction of LCC users. However, it is not clear whether users who generate most of the posts are the same users who receive most of the reactions. For example, a celebrity may generate a post infrequently but receives lots of reaction to each post. To answer this question, first we examine the correlation between the rate of posts and the aggregate reactions rate for different groups of users grouped based on their average level of activity as follows:
-*Active* users who post at least once a day ($>$1),
-*Regular* users who post less than once a day but more than once a week ($\frac{1}{7}$-1), and
-*Casual* users who post less than once a week ($<\frac{1}{7}$).
Figure 5 shows the summary distribution of daily reaction rate among users in each one of the described groups using boxplots. This figure reveals that the reaction rate grows exponentially with the user posting rate. This result indicates that *the small group of users that contribute most posts is also receiving the major portion of all reactions.*

To gain further insight in the correlation between users' actions and reactions, the top three rows of Table VI shows the result for the Rank Correlation (RC) [35] between the total number of users' actions (*i.e.,*posts) and the total number of each type of reactions (*i.e.,*plusones, comments, reshares) for all, top 10% and top 1% of users in our activity dataset. RC has a value between -1 (ranks are reversed) and 1 (ranks are the same), and 0 indicates that ranks are independent. Note that due to the large size of our activity dataset the p-value is $\sim$0 in all cases which confirms that there exist a correlation between the studied parameters. Table VI reveals that *there is a notable positive correlation between the action and different types of reactions for all users (0.39-0.49). This correlation slightly changes for the top 10% of users but significantly drops for the top 1% of users.* The bottom three rows of Table VI show the RC between different types of reactions for the same three groups. We observe a moderate correlation between pairs of reactions for all user (0.39-0.55). More interesting, the RC between different types of reactions rapidly increases for the top 10% and 1% of users (0.78-0.86).

Identity of Users with Most Actions or Most Reactions: We have identified the top 1000 users with the largest number of public posts as well as those who received the largest number of total reactions (of any type) each month. The analysis of the first group did not reveal any clear trend due to the high variation in the characteristics of these users. For the second group, Figure 6 presents the summary distribution of the monthly ranking for 10 users that received the most total reactions during the entire measurement period. Four of these users are well-known individual directly related to the hi-tech industry (Tom Anderson, Sergey Brin, Larry Page and Linus Torvalds). Another user (Matthew Inman) in this group is also an Internet professional since he is the creator of the comic and article webpage theOatmeal.com. However, the ranking of Inman exhibits much wider variations (between 10th and 1000th) over time. Three other users in this top-10 group are different types of celebrities, Dalai Lama is a spiritual leader, Jessi June is a porn star, and Britney Spears is a singer.

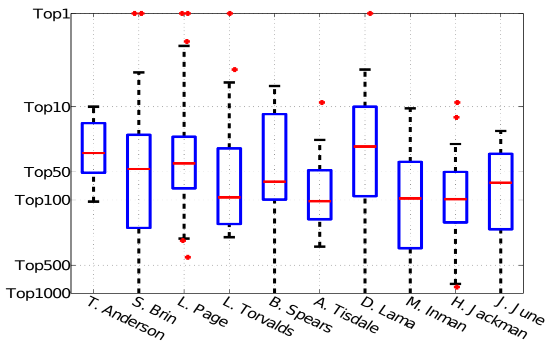
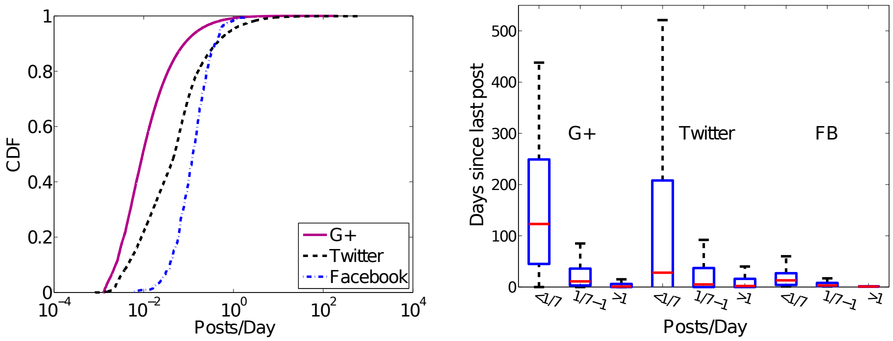


Fig. 6. Summary distribution of monthly ranking of reactions attracted for the main users of G+

Comparison with Other OSNs: We examine a few aspects of user activity (*i.e.,* generating posts or tweets) among G+, Twitter and Facebook users to compare the level of user engagement in these three OSNs. For this comparison, we leverage TW-Act, FB-Act datasets (described in Table IV) that capture activity of random users in the corresponding OSNs. In our analysis, we only consider the active users in each OSN that make up 17%, 35%, and 73% of all users in G+, Facebook and Twitter, respectively.

Activity Rate: Figure 7(a) shows the distribution of average activity rate per user across all active users in each OSN. The activity rate is measured as the total number of posts or tweets divided by the time between the timestamp of a user's first collected action and our measurement time. This figure reveals the following two basic points in about these three OSNs: *(i)* the activity rate among Facebook and G+ users are more homogeneous than across Twitter users, *(ii)* Facebook users are the most active (with the typical rate of 0.19 posts/day) while G+ users exhibit the least activity rate (with the typical rate of 0.06 posts/day).

Recency of Last Activity: An important aspect of user engagement is how often individual users generate a post. We can compute the *recency* of the last post by each active user as the time between the timestamp of last post and our measurement time. The distribution of this metric across a large number of active users provides an insight on how often active users generate a post. Figure 7(b) depicts the distribution of recency of the last post across G+, Twitter and Facebook users. We have divided the users from each OSN into three groups of casual, regular and active users based on their average activity rate ($<\frac{1}{7}$, $\frac{1}{7}$-1, $>$1 post/day) as we



(a) Avg. Post/Tweet rate

(b) Recency of Activity

Fig. 7. Comparison of activity metrics for G+, Twitter and Facebook

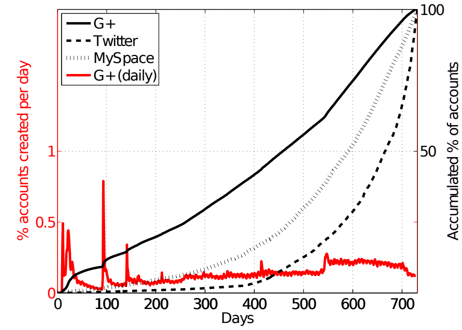


Fig. 8. Relative number of active users in G+, Twitter and MySpace during the first two years of each OSN

described earlier. We observe that among casual users in all three OSNs, Facebook and Twitter users typically generate posts much more frequently (*i.e.,* have lower median recency) than casual G+ users. Regular users in different OSNs exhibit the same relative order in their typical recency of last post. Finally, for active users, it is not surprising to observe that all three OSNs show roughly the same level of recency.

Growth Rate of Active Users: Our TW-Act and MS-Act datasets [48] (presented in Table IV) include information about the evolution of the aggregate number of active users that joined Twitter and MySpace in the two first years after their releases. Hence, comparing these datasets with our G+ activity dataset, we can compare and contrast the growth characteristics of these OSNs. First, if we focus on the total number of active users two years after its release, G+ has 32.4M users that have been active at some point in the network. This value is 2.3 and 8.6 times larger than the number of active users in MySpace and Twitter, respectively.

The left-Y axis in Figure 8 depicts the percentage of new G+ users that were active in each day of our measurement period for the first time whereas the right-Y axis shows the cumulative percentage of active new users over the first two years for G+, Twitter and MySpace. This figure indicates that while G+ has been able to attract active users at a faster average rate, its growth exhibits a very different temporal pattern compared to Twitter and MySpace. More specifically, the slope of growth in Twitter and MySpace steadily increases with time whereas the slope of growth in G+ does not change significantly. In fact, the arrival of new active users in G+ exhibits a bursty pattern (*i.e.,*many users join the system within a short period of time) that appears to be driven by certain events (*e.g.*addition of a new service to G+).

In summary, the analysis of different aspects of user activity in G+ resulted in the following important points: (i) *The number of daily active LCC users has steadily grown but roughly 475 times slower than the whole LCC population.* (ii) *Around 10% of the active LCC users generate a majority of all posts and only 1/10th of these users attract most of all the reactions of any type to their posts (86% of the comments and more than 90% of the plusones and resharers). This is due to the fact that the rate of receiving reaction is correlated with the user posting rate,* (iii) *The comparison of user activity for G+ with Facebook and Twitter reveals that Facebook and Twitter users exhibit a higher rate of generating posts,* (iv) *During the first two years of operation, G+ has attracted more active users than Twitter or MySpace. However, the pace of growth in Twitter and MySpace has been steadily increasing while G+ exhibits rather stable pace of growth with a bursty pattern of arrival for new users.*

VI. Public User Attributes

We compare the willingness of users in different OSNs to publicly share their attributes in their profile. This is an indicator of user engagement and interest in an OSN. Roughly 48% of all the LCC users in G+ provide at least one extra attribute in April 2012 in addition to sex which is a mandatory attribute. This ratio rapidly decreased to 44% at the end of 2012 and eventually reached 30% in our last snapshot in Jul 2013.

We further examine the distribution of the number of visible attributes across LCC users for different LCC snapshots and compare them with 480K random Facebook users (in FB-Pro dataset from Table IV) in Figure 9. We recall that there are 21 different attributes in both G+ and FB profiles. Figure 9 shows that the distribution for all LCC snapshots is very similar and G+ users publicly share a much smaller number of attributes compare to Facebook users. In particular, half of the users publicly share at least 6 attributes on Facebook while less than 10% of G+ users share 6 or more attributes. Twitter profile only has 6 attributes and 3 of them are mandatory. Examination of TW-Pro dataset shows that 69% and 13% of Twitter users share 0 and 1 non-mandatory attribute, respectively. In short, G+ users appear to share more public and non-mandatory attributes than Twitter users but significantly less than Facebook users.

Table VII presents a more detailed view by showing the fraction of LCC users that provide public information for each specific field of their profile in different snapshots. As we can see, the percentage of users that provide each public attribute generally decreases over time. Only in the case of "Introduction", "Bragging Rights" and "Places Lived" we observe an increment in the percentage of people making them public until Aug 2012 when they also start decreasing following the general trend. Furthermore, users seem to be

attribute	LCC-Dec*	LCC-Apr12	LCC-Aug12	LCC-Nov12	LCC-Mar13	LCC-Jul13
Alias	-%	0.00%	0.01%	0.01%	0.01%	0.01%
Bragging rights	3.90%	3.77%	3.93%	3.14%	2.57%	2.38%
Contact (home)	0.21%	0.21%	0.26%	0.23%	0.44%	0.55%
Contributor to	13.15%	11.95%	11.59%	8.10%	5.85%	4.95%
Gender	97.67%	95.82%	95.76%	94.41%	93.02%	92.08%
Indexable	-%	0.00%	0.00%	0.00%	0.00%	0.00%
Introduction	7.80%	8.42%	9.74%	6.52%	5.03%	4.56%
Links	3.63%	3.26%	3.30%	2.51%	1.88%	1.68%
Looking for	2.74%	2.64%	2.61%	2.03%	1.58%	1.41%
Occupation	13.27%	11.47%	13.32%	8.93%	7.10%	6.40%
Other names	4.39%	4.08%	4.20%	3.34%	2.74%	2.50%
Other Profiles	13.48%	10.70%	10.54%	17.44%	5.87%	5.08%
Places	26.75%	26.98%	28.36%	24.15%	20.83%	19.70%
Relationship	4.31%	3.94%	3.99%	3.11%	2.63%	2.46%
Skills	-%	-%	-%	-%	0.10%	0.46%
Tagline	-%	-%	-%	-%	6.75%	6.12%
Web	-%	1.22%	1.10%	1.15%	0.81%	0.78%
Contact(Work)	0.22%	0.34%	0.16%	0.25%	0.27%	0.28%
Education	27.11%	24.32%	24.72%	20.13%	16.80%	15.67%
Employment	13.27%	11.47%	13.32%	17.36%	14.92%	14.04%

TABLE VII
PERCENTAGE OF LCC USERS THAT MAKE PUBLIC EACH ATTRIBUTE FOR EACH DATASET

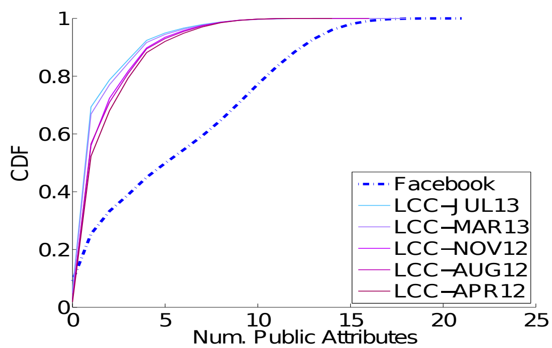
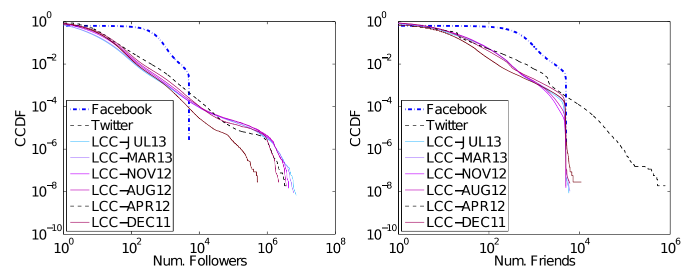


Fig. 9. Distribution of number of public attributes for G+ and Facebook



(a) Num Followers

(b) Num Friends

Fig. 10. Degree Distribution for different snapshots of G+, Twitter and Facebook

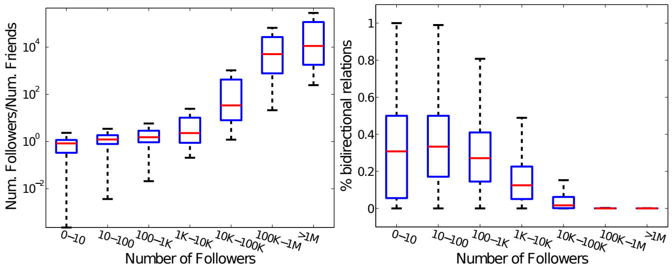
more inclined to share professional attributes (*e.g.*"Studies", "Location", "Profiles" and "Profession") and less willing to share attributes that reveal rather more private aspects of their life such as their relationships (*e.g.*, single, married, friendship, or love). This may be an indication that the fraction of professional users in G+ is larger than other OSNs which makes their attribute more visible. To examine this possible explanation, we have identified the 20 most popular users from Twitter, Facebook and Google+ (*i.e.,* users with the most followers in Twitter and Google+, and Facebook pages with most fans) and manually inspected their professions. We observe that in Twitter and Facebook the Top 20 users are *celebrities* (*e.g.*, politicians, musicians, actors, soccer players) or major companies (*e.g.*, YouTube, Twitter, Facebook). However, in addition to celebrities, there are some professionals from the hi-tech sector (*e.g.*, Google CEO, Virgin CEO, MySpace founder), photographers or even moderately famous Google products (*e.g.*the Google Art project) among the Top 20 G+ users. The presence of these technology-related accounts among the Top 20 G+ users (with many followers) shows that a larger fraction of G+ users are interested in (and thus follow) these technology-related accounts compared to other OSNs. This evidence supports our observation that G+ has a larger fraction of professional users than other OSNs and is aligned with the reported results in [44]. Finally, we observe that the growing trend in all plots of Figure 3 has flattened during the last few months of our measurements. However, we were not able to identify any compelling reason to explain this pronounced change in the observed temporal trends.

VII. LCC CONNECTIVITY & ITS EVOLUTION

In this section, we focus on the evolution of different features of connectivity among LCC users over time as the system becomes more populated, and compare these features with other OSNs [23].

Degree Distribution: The distribution of node degree is one of the basic features of connectivity. Since G+ structure is a directed graph, we separately examine the distribution of the number of followers in Figure 10(a) and friends in Figure 10(b). Each figure shows the corresponding distribution across users in each one of our LCC snapshots, among Twitter users in TW-Con snapshot, and the distribution of neighbors for random Facebook users in FB-Con snapshots[10]. This figure demonstrates a few important points: First, we have performed

[10]Note that Facebook forces bidirectional relationships. Therefore, the distribution for Facebook in both figures is the same.



(a) #followers/#friends (b) % bidirectional relationships

Fig. 11. The level of imbalance and reciprocation for different group of users based on their number of followers.

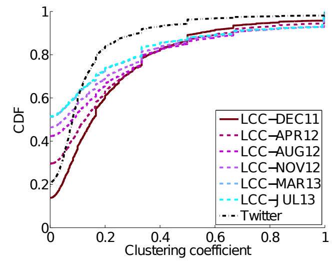


Fig. 12. Clustering Coefficient

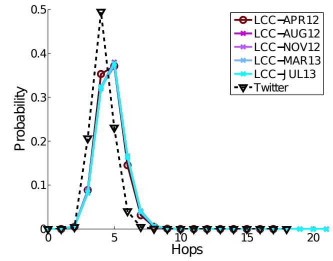


Fig. 13. Average Path Length

the distribution fitting using the method described by Astott et al. [25]. The distribution of the number of followers best fits a lognormal distribution with σ = 0.167 and μ = 2.307 (kolmogorov-smirnov distance = 0.024) whereas the distribution of the number of friends best fits a power law distribution with α = 2.024 (kolmogorov-smirnov distance = 0.006).

Second, comparing the shape of the distribution across different LCC snapshots, we observe that both distributions look very similar for all LCC snapshots. The only exception is the earliest LCC snapshot (LCC-Dec) that has a less populated tail. This comparison illustrates that the shape of both distributions has initially evolved as the LCC became significantly more populated and users with larger degree appear, and then the shape of distributions has stabilized after 14 months since G+ release. Third, interestingly, the shape of the most recent distribution of followers and friends for G+ users is very similar to the corresponding distribution for Twitter users. The only difference is in the tail of the distribution of number of friends which is due to the limit of 5K friends imposed by G+ [10]. *The stability of the distribution of friends and followers for G+ users in recent months coupled with their striking similarity with these features in Twitter indicates that the degree distribution for G+ network appears to have become stable*. Fourth, while the distributions for Facebook are not directly comparable due to its bidirectional nature, Figure 10 shows that the distribution of degree for Facebook users does not follow a power law [51] as they generally exhibit a significantly larger degree than Twitter and G+ users. Specifically, 56% of Facebook users have more than 100 neighbors while only 3.6% (and 0.8%) of the G+ (and Twitter) users maintain that number of friends and followers.

Balanced Connectivity & Reciprocation: Our examination showed that the percentage of bidirectional relationships between LCC users has steadily dropped from 32% (in Dec 2011) and became rather stable in the last month of our study around 22.4% (in Jul 2013). Again, we observe that this feature of connectivity among LCC users in G+ seems to have reached a quasi-stable status after the system have experienced a major growth. Interestingly, Kwak et al. [41] reported a very similar fraction of bidirectional relationships (22%) in their Twitter snapshot from July 2009. This reveals yet another feature of G+ connectivity that is very similar to the Twitter network and very different from the fully bidirectional Facebook network. In order to gain deeper insight on this aspect of connectivity, we examine the fraction of bidirectional connections for individual nodes and its relation with the level of (im)balance between node indegree and outdegree. This in turn provides a valuable clue about the user level connectivity and reveals whether users exchange or simply relay information. To quantify the level of balance in the connectivity of individual nodes, Figure 11(a) plots the summary distribution of the ratio of followers to friends (using boxplots) for different group of users based on their number of followers in our most recent snapshot (LCC-Jul13). This figure demonstrates that only the low degree node (with less than 100 followers) exhibit some balance between their number of followers and friends. Otherwise, the number of friends among G+ users grows much slower than the number of followers.

We calculate the percentage of bidirectional relationships for a node u, called $BR(u)$, as expressed in Equation 1 where Friend(u) and Follower(u) represent the set of friends and followers for u, respectively. In essence, $BR(u)$ is simply the ratio of the total number of bidirectional relationships over the total number of unique relationships for user u.

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

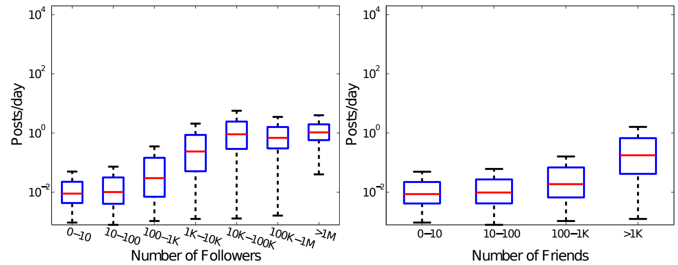
Figure 11(b) presents the summary distribution of $BR(u)$ for different groups of G+ users in LCC based on their number of followers using LCC-Jul13 snapshot. The results for other recent LCC snapshots are very similar. As expected, popular users ($>$ 10k followers) have a very small percentage of bidirectional relationships. As the number of followers decreases, the fraction of bidirectional relationships slowly increases until it reaches around 35% for low-degree users ($<$ 1K followers). In short, even low degree users that maintain a balanced connectivity, do not reciprocate more than 40% of their relationships. Our inspection of 5% of LCC users who reciprocate more than 90% of their edges revealed that 90% of them maintain less than 3 friends/followers and less than 5% of them have any public posts. *These results collectively suggest that G+ users reciprocate a small fraction of their relationships which is often done by very low degree users with no activity.*

Clustering Coefficient: Figure 12 depicts the summary distribution of the undirected version of the clustering coefficient (CC) among G+ users in different LCC snapshots This figure clearly illustrates that during the 18 month period (from Dec 2011 to Jul 2013), the CC among the bottom 90% of users remained below 0.6 and continuously decreased. Moreover, the percentage of users with clustering coefficient 0 has grew from 20% to more than 50% in one year and a half. On the other hand, the CC for the top 10% of users (particularly in the last four snapshots) has become very stable. A similar trend in cluster coefficient has been recently reported for a popular Chinese OSN [56] which indicates such an evolution in CC might be driven by the underlying social forces rather than features of the OSNs. We also noticed that if we remove the growing number of users with CC=0, the distribution of CC among G+ users also exhibit only minor changes between Aug 2012 and Jul 2013 which is another sign of stability in the connectivity features of G+ network. Compared to Twitter network where CC is less than 0.3 for 90% of users, G+ is still more clustered. Furthermore, using the approximation presented by Cha et al. [44], we conclude that just 1% of the nodes in a complete Facebook snapshot [51] collected in May 2011 [51] have a CC larger than 0.2 in comparison with the 16% and 30% in Twitter and G+ (using LCC-Nov13 snapshot). In summary, *the G+ structure has become less clustered as new users joined the LCC over the first 18 months of its operation. Also as the population of G+ has grown, its connectivity has become less clustered but it is still the most clustered network compared to Twitter and Facebook.*

Path Length: Figure 13 plots the probability distribution function for the pairwise path length between nodes in different LCC snapshots for G+ and a snapshot of Twitter (TW-Con). We observe that roughly 99% of the pairwise paths between G+ users are between 2 to 7 hops long and roughly 70% of them are 4 or 5 hops. The diameter of the G+ graph has increased from 17 hops (in April) to 21 hops (in July of 2013). The two visibly detectable changes in this feature of G+ graph as a result of its growth are: *(i)* a small decrease in typical path length (from April 2012 to July 2013), and *(ii)* the increase of its diameter in the same period. Table VIII summarizes the average and mode path length, the diameter and the efficient diameter [42] (*i.e.,* 90 percentile of pairwise path length) for the G+ network (using LCC-Jul13), Twitter (using TW-Con) and a Facebook snapshot from [26]. We observe that G+ and Facebook network exhibit similar average and mode path length but Facebook has a much longer diameter. This could be due to the fact that the size of Facebook network is roughly one order of magnitude larger than G+ LCC. Twitter has the shortest average and mode path length and diameter among the three. We conjecture that this difference is due to the lack of restriction in the maximum number of friends in Twitter that leads to many shortcuts in the Twitter network.

	G+	FB	Twitter
Path Length (Avg)	4.75	4.7	4.1
Path Length (Mode)	5	5	4
Eff. Diameter	6	-	4.8
Diameter	21	41	18

TABLE VIII
SUMMARY OF PATH LENGTH AND DIAMETER CHARACTERISTICS FOR G+ (LCC-JUL13 SNAPSHOT), FACEBOOK AND TWITTER



(a) #followers vs Avg. Post Rate (b) #friends vs Avg. Post Rate

Fig. 14. Correlation between Post Rate and Connectivity (#followers and #friends) properties in Google+

In summary, our analysis on the evolution of LCC connectivity led to the following key findings: (i) As the size of LCC significantly increased over the past year, all connectivity features of LCC (except the clustering coefficient) have initially evolved but have become rather stable in recent months despite its continued growth, (ii) Only low degree and non-active users may reciprocate a moderate fraction of their relationships, (iii) Many key features of connectivity for G+ network (e.g., degree distribution, fraction of bidirectional relationships) have striking similarity with the Twitter network and are very different from the Facebook network. The connectivity features of G+ coupled with the fact that a small fraction of users generate most posts and attract more reactions (as we reported in Section V) suggest that G+ is used for one-way message propagation rather than two-way user interaction.

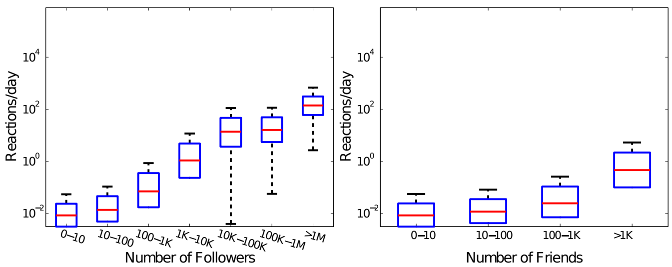
VIII. RELATING USER ACTIVITY & CONNECTIVITY

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. To determine how correlated the connectivity of a user (#followers, #friends) is with different aspects of its activity (#Posts, #Plusones, #Comments, #Reshares), we compute the Rank Correlation (RC) between all 8 pairs of these properties across all, top 10% and top 1% of active users in our last LCC snapshot and summarize the result in Table IX. The results suggest that users' popularity (#followers) is more correlated with two specific types of reactions, #plusones and #of comments (0.33), than with the users direct activity, #posts (0.22). Furthermore, we observe similar results for all the #friends. However, examination of the RC for top 10% and top 1% users show that the RC between the number of followers (friends) and the number of reactions (of any type) increases (decreases) for a smaller group of top users. 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 14 show the summary distribution of posts per day for different groups of users based on #followers, and #friends using log scale for both axis. Figure 14(a) illustrates that the rate of generated posts by users rapidly increases with their number

	All		Top 10%		Top 1%	
	num friends	num followers	num friends	num followers	num friends	num followers
num posts	0.22	0.21	0.25	0.32	0.11	0.28
num attach.	0.17	0.19	0.24	0.31	0.10	0.28
num plusones	0.25	0.33	0.23	0.43	0.05	0.41
num comments	0.34	0.33	0.17	0.37	-0.02	0.39
num reshares	0.16	0.23	0.20	0.40	0.01	0.41

TABLE IX

RANKING CORRELATION AMONG THE CONNECTIVITY AND ACTIVITY PROPERTIES FOR ALL THE USERS AND FOR THE TOP 10% AND TOP 1% USERS WITH MORE FOLLOWERS IN G+



(a) #followers vs Avg. Reaction rate (b) #Friends vs Avg. Reaction rate

Fig. 15. Correlation between Aggregate Reaction Rate and Connectivity (#followers and #friends) properties in Google+

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 14(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 14(a).

Figure 15 presents the summary distribution of average aggregate 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, there is a positive correlation between the #followers and the activity rate or reaction rate of individual users which is more pronounced for users with 100-100K followers and users with more than 100 friends. However, the correlation between #friends and the activity or reaction rate is only visible for users with more than 1K friends.

IX. RELATED WORK

We group related work into three categories as follows:
1) OSN characterization: The importance of OSNs has motivated researchers to characterize different aspects of the most popular OSNs. The graph properties of Facebook [51], [26], Twitter [41], [29] and other popular OSNs [46] 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 G+ graph over a period of one year. In addition, some other works leverage passive (e.g., click streams) [27], [50] or active [55], [36] 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 behavior 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 G+. Ding et al. proposes a collaborative way to obtain big datasets from the OSNs [30]. Finally, few works have also analyzed the users' information sharing through their public attributes in OSNs such as Facebook [47].
2) 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) [40], the growth of an OSN and the evolution of its graph properties [45], [24], [56], [32], [33], [48], [43] or the evolution of the interactions between users [38], [53] and the user availability [28]. In this paper, instead of looking at a specific aspect, we perform a comprehensive analysis to study the evolution of different key aspects of G+ namely, the system growth, the representative of the different network elements, the LCC connectivity and activity properties and the level of information sharing.

3) Google+ Characterization: G+ has recently attracted the attention of the research community. Mango et al. [44] use a BFS-based crawler to retrieve a snapshot of the G+ 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 G+. Schiberg et al. [49] leverage Google's site-maps to gather G+ 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. [34] use a BFS-based crawler to obtain several snapshots of the G+ LCC in its first 100 days of existence. Using this dataset the authors study the evolution of the main graph properties of G+ 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 presents important differences to its early status that is the focus of the previous works. In another interesting, but less related work, Kairam et al. [39] use the complete information for more than 60K G+ users (provided by G+ administrators) and a survey including answers from 300 users to understand the selective sharing in G+. Their results show that public activity represents 1/3 of the G+ activity and that an important fraction of users make public posts frequently. Finally, other papers have study the video telephony system of G+ [54], the public circles feature [31], collaborative privacy management approaches [37] and the new Ripples feature [52].

X. Conclusion

In this paper, we study the evolution of the key features of the last major player released in the OSN market, namely Google+. Toward this end, we capture, to the best of our knowledge, one of the largest collection of datasets used to characterize a specific OSN. These datasets include information related to the connectivity, activity and information sharing properties of Google+ users over a period of two years. Our detailed analysis led to the following main insights:
(i) Contrary to some widespread opinion, G+ is not really a "ghost town". First, the number of interested users who connect to the LCC of the network, is growing at an increasing rate. However, this rate is lower than the one depicted by official reports that most likely include a large number of singletons. These users appear to be automatically registered in G+ after creating a Google account to use other popular Google services. Second, the overall rate of actions and reactions is steadily growing in G+ which is a positive indicator about the level of user engagement.
(ii) Despite the growth in user population and activity, the connectivity and activity features of G+ seem to have reached a statistically stable state after the first year. *(iii)* In this seemingly stable status our detailed analyses of connectivity and activity features reveal that Google+ is used as a broadcast social media system in which a relative small group of popular and very active users contribute most of the posts and attract most users reactions.

XI. Acknowledgements

The authors would like to thank anonymous reviewers for their valuable feedback as well as Meeyoung Cha, Sue Moon, Diego Saez and Haewoon Kwak for sharing information from their datasets with us. This work has been partially supported by the European Union through the FP7 METRICS (607728) and H2020 TYPEs (EU-H2020-653449) Projects, the National Science Foundation (NSF) grant no. IIS-0917381, the Ministerio de Economa y Competitividad of SPAIN through the project BigDatAAM (FIS2013-47532-C3-3-P) and the Regional Government of Madrid through the project BRADE-CM (P2013/ICE-2958). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Funding Agencies.

References

[1] http://online.wsj.com/article/SB10001424052970204653604577249341403742390.html?mod=WSJ_hp_LEFTTopStories.
[2] http://google-plus.com/category/statistics/.
[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 blog*. http://googleblog.blogspot.com.es/2013/10/google-hangouts-and-photos-save-some.html.
[12] *Google+ official learn more site*. http://www.google.com/intl/en/+/learnmore/.
[13] *Google+ official support site*. http://support.google.com/plus/.
[14] *Google+ Pages*. http://www.google.com/+/business/.
[15] *Google+ Pages announcement*. http://googleblog.blogspot.com/2011/11/google-pages-connect-with-all-things.html.
[16] *Google Privacy Policy*. http://www.google.com/policies/privacy/.
[17] *Google+ Redesign*. http://mashable.com/2012/04/11/google-plus-redesign/.
[18] *Google Registration. Arstechnica.* http://arstechnica.com/gadgets/2012/01/google-doubles-plus-membership-with-brute-force-signup-process/.
[19] *Google Registration. Blogspot.* http://googlesystem.blogspot.com/2012/01/new-google-accounts-require-gmail-and.html.
[20] *Google wave shut down announcement*. http://support.google.com/bin/answer.py?hl=en&answer=1083134.
[21] *Orkut Official Site*. http://www.orkut.com.
[22] *Orkut Statistics, Wikipedia.* http://en.wikipedia.org/wiki/Orkut.
[23] *Research on online social networks: time to face the real challenges*, Seattle, Washington, June 2009.
[24] 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.
[25] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777, 2014.
[26] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four Degrees of Separation. *CoRR*, abs/1111.4570, 2011.
[27] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *ACM IMC*, 2009.
[28] A. Boutet, A.M. Kermarrec, E. Le Merrer, and A. Van Kempen. On the impact of users availability in osns. In *ACM SNS*, 2012.
[29] M. Cha, H. Haddadi, F. Benevenuto, and K.P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *AAAI ICWSM*, 2010.
[30] C. Ding, Y. Chen, and X. Fu. Crowd crawling: Towards collaborative data collection for large-scale online social networks. COSN, 2013.
[31] L. Fang, A. Fabrikant, and K. LeFevre. Look who i found: Understanding the effects of sharing curated friend groups. In *ACM WebSci*, 2012.
[32] 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.
[33] S. Garg, T. Gupta, N. Carlsson, and A. Mahanti. Evolution of an online social aggregation network: an empirical study. In *ACM IMC*, 2009.
[34] 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.
[35] Z. Govindarajulu. Rank correlation methods. *Technometrics*, 34(1):108–108, 1992.
[36] L. Gyarmati and T.A. Trinh. Measuring user behavior in online social networks. *Network, IEEE*, 24(5):26–31, 2010.
[37] H. Hu, G.J. Ahn, and J. Jorgensen. Enabling collaborative data sharing in google+. In *IEEE Globecom*, 2012.

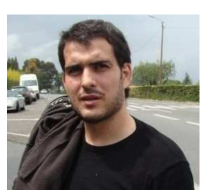
[38] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, and B.Y. Zhao. Understanding Latent Interactions in Online Social Networks. In *ACM IMC*, 2010.
[39] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi. Talking in circles: Selective sharing in google+. In *ACM CHI*, 2012.
[40] R. Kumar, J. Novak, and A Andtomkins. Structure and evolution of online social networks. In *ACM KDD*, 2006.
[41] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *WWW*, 2010.
[42] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *ACM KDD*, 2005.
[43] H. Liu, A. Nazir, J. Joung, and C. Chuah. Modeling/predicting the evolution trend of osn-based applications. In *WWW*, 2013.
[44] G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida. New kid on the block: Exploring the google+ social graph. In *ACM IMC*, 2012.
[45] A. Mislove, H.S. Koppula, K.P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *ACM WOSN*, 2008.
[46] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *ACM IMC*, 2007.
[47] 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.
[48] R. Rejaie, M. Torkjazi, M. Valafar, and W. Willinger. Sizing Up Online Social Networks. *IEEE Network*, 24(5):32 –37, Sept-Oct 2010.
[49] 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.
[50] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In *ACM IMC*, 2009.
[51] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The Anatomy of the Facebook Social Graph. *CoRR*, abs/1111.4503, 2011.
[52] F. Viégas, M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant, and C. Wren. Google+ripples: A native visualization of information flow. In *WWW*, 2013.
[53] L. Weng, J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini. The role of information diffusion in the evolution of social networks. In *ACM KDD*, 2013.
[54] 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.
[55] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *ACM SIGIR*, 2012.
[56] 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.

Roberto Gonzalez received his MSc in Telematic Engineering from Polytechnic University of Cataluña and Carlos III University of Madrid and his PhD in Telematic Engineering from Carlos III University of Madrid in 2011 and 2014, respectively. Since July 2014 he is a research scientist at NEC Laboratories Europe. His research interests are online privacy, social networks and network measurements with a special focus on the data analytics part. He has published more than 10 papers in leading venues in web conferences (WWW, COSN) and networking conferences (TON, IEEE Networks, Computer Networks, GLOBECOM, P2P).

Reza Motamedi is currently a Ph.D. student in the Computer Science Department at the University of Oregon, Eugene. He received his M.Sc. degree in computer science from Sharif University of Technology, Iran, in 2007 and his and B.Sc. from Iran University of Science and Technology in 2002. His research interests include Internet topology mapping, network measurement, and online social network characterization.

Ruben Cuevas obtained his MSc and PhD in Telematics Engineering at University Carlos III of Madrid (Spain) in 2007 and 2010, respectively. He is currently Assistant Professor at the Telematics Engineering Department at University Carlos III of Madrid. He has been Courtesy Assistant Professor at University of Oregon in 2012. He has been the PI in several national and international research projects in the area of content distribution, online social networks and online advertising. Moreover, he is co-author of more than 40 papers in prestigious international journals and conferences such as IEEE/ACM TON, IEEE TPDS, ACM CONEXT, WWW or IEEE INFOCOM. His main research interests include Internet Measurements, Online Social Networks, Online Advertising and Privacy.

Reza Rejaie is currently an Associate Professor at the University of Oregon. From 1999 to 2002, he was a Senior Technical Staff member at AT&T Labs Research in Menlo Park, California. He received a NSF CAREER Award for his work on Peer-to-Peer streaming in 2005. Reza has also been the recipient of the European Union Marie Curie Fellowship in 2009. He received his M.S. and Ph.D. degrees from the University of Southern California in 1996 and 1999, and his B.S. degree from the Sharif University of Technology in 1991. Reza has been a senior member of both the ACM and IEEE since 2006.



Angel Cuevas received his MSc in Telecomunication Engineering, MSc in Telematics Engineering, and Ph.D. in Telematics Engineering from the Universidad Carlos III de Madrid in 2006, 2007, and 2011 respectively. He is currently a tenure-track Assistant Professor in the Department of Telematic Engineering at Unviersidad Carlos III de Madrid and Adjunct Professor at Institut Mines-Telecom SudParis. His research interests focus on On-line Social Networks, P2P Networks, Wireless Sensor Networks and Internet measurements. He is co-author of more than 40 papers in prestigious international journals and conferences such as IEEE/ACM Transactions on Networking, ACM Transactions on Sensor Networks, Elsevier Computer Networks, IEEE Network, IEEE Communications Magazine, WWW, ACM CONEXT, ACM MSWiM and IEEE P2P. He is co-recipient of the Best Paper Award in ACM MSWiM 2010.