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

Roberto González & Rubén Cuevas, UC3M
Reza Motamedi & Reza Rejaie, Univ. Oregon
Angel Cuevas, Institut Telecom Sud Paris (now UC3M)

Rubén Cuevas
rcuevas@it.uc3m.es
Universidad Carlos III de Madrid

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Motivation

• Social Media market has rapidly grown and reach a maturity
  - Facebook and Twitter have a dominant position
  - Savvy Users

• In this scenario:

  *can a new OSN get a significant piece of the OSN market?*
Motivation

• Google+ (G+) is an interesting candidate to address the previous question

• Some specificities of our case of study:
  - G+ mixes features from both Twitter and Facebook in order to attract users from both OSNs
  - It is supported by a major Internet player (Google)
Our starting point

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Our goal

• Let’s try doing an objective analysis
• i.e., analyze...
  - the evolution of the size of the different components of the network
  - the evolution of the activity in the OSN
  - The evolution of the connectivity properties
• ... over an enough long and representative period of time
Outline

1. Google+ background
2. Measurement Methodology & Datasets
3. Macro-level structure & its evolution
4. Public Activity & its evolution
5. Connectivity Properties & its evolution
6. Conclusion
Google+ Background

- Unidirectional relationships (like TW)
- Control on the visibility of a post (like FB)
  - Post = text + attachments (photo, video)
- Reactions to a post:
  - Comment, Reshare or Plusone (+1)
- Each user a profile with 17 fields
  - Each field can be public, private or empty
- User id space:
  - User-id: 21 integers digit
  - Not clear strategy/Sparsely populated
Measurement Methodology & Datasets
Measurement Methodology
Capturing LCC

• Largest Connected Component (LCC)
• BFS-based
• List of friends, List of followers, Profile
• 21 instances of our crawler + 1 coordinator
  - Each one is responsible for a region of the id-space
  - The coordinator assigns the learnt user-ids to the corresponding crawler instance
• ~ 800K users/hour -> Whole LCC in 7-10 days
Measurement Methodology
Random sample of users

- We leverage the G+ search API
  - Receives a keyword (e.g. surname) as input
  - Returns up to 1000 users including that keyword in its name/surname

- For popular names (> 1000 registered users)
  - Selective answer with well connected and active users

- For mid-popular/unpopular (< 1000 registered users)
  - Return all the users

- We use the US census to provide mid/low popular surnames as input, and only consider as valid those surnames for which the API returns less than 1K users

---

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Measurement Methodology
Capturing Users’ Public Activity

• User’s activity
  - User’s posts
  - Num. attracted reactions per post

• We use the G+ API
  - For all users in LCC Sep 2012
  - User’s activity between G+ release (Jun 28th 2011) and our measurement starting date (Sep 7th 2012) -> 437 days
  - 68 days
### LCC Datasets

<table>
<thead>
<tr>
<th>Name</th>
<th>#nodes</th>
<th>#edges</th>
<th>Start Date</th>
<th>Duration (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC-Dec*</td>
<td>35.1M</td>
<td>575M</td>
<td>11-Nov-2011</td>
<td>46</td>
</tr>
<tr>
<td>LCC-Apr</td>
<td>51.8M</td>
<td>1.1B</td>
<td>15-Mar-2012</td>
<td>29</td>
</tr>
<tr>
<td>LCC-Aug</td>
<td>79.2M</td>
<td>1.6B</td>
<td>20-Aug-2012</td>
<td>4</td>
</tr>
<tr>
<td>LCC-Sep</td>
<td>85.3M</td>
<td>1.7B</td>
<td>17-Sep-2012</td>
<td>5</td>
</tr>
<tr>
<td>LCC-Oct</td>
<td>89.8M</td>
<td>1.8B</td>
<td>15-Oct-2012</td>
<td>5</td>
</tr>
<tr>
<td>LCC-Nov</td>
<td>93.1M</td>
<td>1.9B</td>
<td>28-Oct-2012</td>
<td>6</td>
</tr>
</tbody>
</table>

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Random Samples & Users’ Activity Datasets

Random Samples

<table>
<thead>
<tr>
<th>Name</th>
<th>#nodes</th>
<th>#edges</th>
<th>Start Date</th>
<th>Duration (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rand-Apr</td>
<td>2.2M</td>
<td>145M</td>
<td>08-Apr-2012</td>
<td>23</td>
</tr>
<tr>
<td>Rand-Oct</td>
<td>5.7M</td>
<td>263M</td>
<td>15-Oct-2012</td>
<td>10</td>
</tr>
<tr>
<td>Rand-Nov</td>
<td>3.5M</td>
<td>157M</td>
<td>28-Oct-2012</td>
<td>13</td>
</tr>
</tbody>
</table>

Users’ Activities

<table>
<thead>
<tr>
<th>Users</th>
<th>Posts</th>
<th>Attachments</th>
<th>Plusones</th>
<th>Comments</th>
<th>Reshares</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.6M</td>
<td>218M</td>
<td>299M</td>
<td>352M</td>
<td>202M</td>
<td>64M</td>
</tr>
</tbody>
</table>
## Other datasets (comparison)

<table>
<thead>
<tr>
<th>Name</th>
<th>OSN</th>
<th>Date</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tw-Pro</td>
<td>Twitter</td>
<td>Jul 2011</td>
<td>Profile (80K rand. Users)</td>
</tr>
<tr>
<td>Tw-Con*</td>
<td>Twitter</td>
<td>Aug 2009</td>
<td>Connectivity (55M users)</td>
</tr>
<tr>
<td>Tw-Act*</td>
<td>Twitter</td>
<td>Jun 2010</td>
<td>Activity (895K rand. Users)</td>
</tr>
<tr>
<td>FB-Pro</td>
<td>Facebook</td>
<td>Jun 2012</td>
<td>Profile (480K rand. Users)</td>
</tr>
<tr>
<td>FB-Con</td>
<td>Facebook</td>
<td>Jun 2012</td>
<td>Connectivity (75K rand. Users)</td>
</tr>
<tr>
<td>FB-Act</td>
<td>Facebook</td>
<td>Sep 2012</td>
<td>Activity (16K rand. Users)</td>
</tr>
</tbody>
</table>
Macro-level structure & its evolution
Macro-level structure & its evolution

• Every OSN is formed by
  - Largest Connected Component (LCC)
  - Partitions (or islands)
    • Connected components smaller than the LCC
  - Singletons
    • Isolated nodes without connections to others
Evolution of LCC size

Avg. daily number of new LCC users
- 150K (Dec 2011-Apr 2012)
- 207K (Apr 2012-Nov 2012)

Impressive...
- but significantly lower than 0.85M-1.8M new registered users reported by Google in the same period

- Why??
  9.6K LCC users leaves the system (in avg.) every day
  - They show a connectivity similar to other LCC users, but they do not have any activity
  - Lack of interest to actively participate in the system

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
## Evolution of the main components

<table>
<thead>
<tr>
<th>Element</th>
<th>Run-Apr</th>
<th>Run-Oct</th>
<th>Run-Nov</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>43.5</td>
<td>32.3</td>
<td>32.2</td>
</tr>
<tr>
<td>Partitions</td>
<td>1.4</td>
<td>1.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Singletons</td>
<td>55.1</td>
<td>66.0</td>
<td>66.3</td>
</tr>
<tr>
<td>All</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

- % singletons (↑), % LCC (↓), % Islands (~)
- LCC in other OSNs → FB (99.91%), TW (94.18%)
- This is a side effect of the integrated registration process imposed by Google
  - e.g., a new gmail (youtube) account automatically generates a G+ accounts
- Singletons may be unaware they are in G+
Public Activity
&
Its evolution
Public activity & its evolution

• Public activity is important
  - It is the one providing more visibility
  - Can be indexed by search engines (including Google)
  - Available to others (excluding Google) for marketing and mining purposes

• An early study using ground truth-data concludes that 30% of posts in G+ are public

• Collecting private posts
  - no representative
  - unethical
Temporal Characteristics of Public Activity (1)

- Steadily increasing rate in # daily posts after initial phase
- Peaks correlated with major events
- Saw-tooth shape due to weekends
- Most posts have attachments but...
- The #posts triggering reactions is significantly smaller

# daily posts

[Graph showing daily posts over time with different categories like Total, Attachments, +1s, Comments, and Reshares]

- Google+ or Google-? Dissecting the evolution of the New OSN in its first year
  - March 2012
Temporal Characteristics of Public Activity (2)

- The number of daily reactions are also steadily increasing after the initial phase
- +1 is the preferred reaction and rapidly growing

![Graph showing daily reactions/attachments over time]

- Google+ or Google-? Dissecting the evolution of the New OSN in its first year
  - March 2012
• Growth rate -> 3K users/day
• ~60 times less than the # new daily LCC registrations
• The comparison of this Fig. with the previous one suggests a clear skewness in the users’ contribution
Skewness in the user’s contribution of posts and attracted reactions

- Top 10% users generate 80% of public posts
- Top 1% users attract:
  - 80% comments
  - 90% +1s and reshares

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
- March 2012
Correlation posting vs reactions

- Defined groups (posts/day):
  - Casual (<1/7)
  - Regular (1/7-1)
  - Active (>1)
- Most active users attract a larger number of reactions
- The public activity (posts + reactions) in G+ happens around a small fraction of active users
Comparison with other OSNs
Activity Rate

- We use our G+, TW and FB activity datasets
- Fraction of active users:
  - FB (73%)
  - TW (35%)
  - G+ (17%)
- Activity rate for active users
  - FB & G+ more homogeneous
  - Median values
    - FB (0.19) vs. G+ (0.08)
Comparison with other OSNs
User’s public attributes

• G+ (17), FB (21), TW (3+3)
• Stability of results across LCC snapshots
• In median FB users make public 6 attributes vs <10% in G+
• In Twitter 69% of users do not make public any non-mandatory attribute and 13% make public 1.
• Level of information sharing:
  – FB > G+ > TW

![Graph showing CDF of num. public attributes across LCC snapshots and OSNs]
Connectivity Properties & its evolution
Connectivity & Its Evolution
Degree Distribution (# followers)

- Stable since Apr 2012
- Power-law ($\alpha = 1.26$)
- Similar to other OSNs (excepting FB)
- Distribution very similar to Twitter!!
Connectivity & Its Evolution
Degree Distribution (# friends)

- Similar results for #friends
- Power-law ($\alpha = 1.39$)
- Distribution very similar to Twitter, but...
- FB & G+ #friends limits
  - 5k

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Connectivity & Its Evolution
Reciprocation

• Aggregate % bidir. relation.
  - Dec 2011 (32%) vs Nov 2012 (21.3%)
  - TW 2009 (22%)
• Again, very similar to TW!!
• Just low popular users (< 1k followers) reciprocate a significant portion of connections (> 30%)
• G+ is used as a broadcast network (similar to TW)

Google+ or Google-? Dissecting the evolution of the New OSN in its first year
March 2012
Conclusion
Conclusion
“Take Aways”

1. **G+** is growing rapidly:
   - 200k new LCC registered users per day (they show interest).
   - However this rate is 1 order of magnitude smaller than the one reported by Google
   - Reason: integrated registration process

2. The number of LCC active users steadily grow (3k per day)
   - But… 60 times less than new LCC registered users per day

3. **G+** activity (posts & reactions) are concentrated around a small fraction of active users
4. Despite of the impressive growth of the LCC, the main connectivity properties have become rather stable. This indicates that the network has reached a mature status.

5. Most key connectivity attributes have a striking similarity with TW and are very different from FB. These attributes suggest that G+ is used for message propagation similar to TW rather than pairwise user interaction like FB.
Conclusion

Answer to the initial question

“Under a mature OSN marketplace where few players (FB, TW) present a dominant position, a new OSN (supported by a major player in the Internet) is able to attract an impressive number of initially interested users (LCC users) but has serious difficulties to get those users actively engaged in the system”
Google+ or Google-?
Dissecting the evolution of the New OSN in its first year

Roberto González & Rubén Cuevas, UC3M
Reza Motamedi & Reza Rejaie, Univ. Oregon
Angel Cuevas, Institut Telecom Sud Paris (now UC3M)

Rubén Cuevas
rcuevas@it.uc3m.es
Universidad Carlos III de Madrid