Collaboration in the Cloud at Google

Yunting Sun, Diane Lambert, Makoto Uchida, Nicolas Remy

Google Inc.

January 8, 2014

Abstract

Through a detailed analysis of logs of activity for all Google employees[1], this paper shows how the Google Docs suite (documents, spreadsheets and slides) enables and increases collaboration within Google. In particular, visualization and analysis of the evolution of Google's collaboration network show that new employees[2], have started collaborating more quickly and with more people as usage of Docs has grown. Over the last two years, the percentage of new employees who collaborate on Docs per month has risen from 70% to 90% and the percentage who collaborate with more than two people has doubled from 35% to 70%. Moreover, the culture of collaboration has become more open, with public sharing within Google overtaking private sharing.

1   Introduction

Google Docs is a cloud productivity suite and it is designed to make collaboration easy and natural, regardless of whether users are in the same or different locations, working at the same or different times, or working on desktops or mobile devices. Edits and comments on the document are displayed as they are made, even if many people are simultaneously writing and commenting on or viewing the document. Comments enable real-time discussion and feedback on the docu-ment, without changing the document itself. Authors are notified when a new comment is made or replied to, and authors can continue a conversation by replying to the comment, or end the discussion by resolving it, or re-start the discussion by re-opening a closed discussion stream. Because documents are stored in the cloud, users can access any document they own or that has been shared with them anywhere, any time and on any device. The question is whether this enriched model of collaboration matters?

There have been a few previous qualitative analyses of the effects of Google Docs on collaboration. For example, the review of Google Docs in [1] suggested that its features should improve collaboration and productivity among college students. A technical report [2] from the University of Southern Queensland, Australia argued that Google Docs can overcome barriers to usability such as difficulty of installation and document version control and help resolve conflicts among co-authors of research papers. There has also been at least one rigorous study of the effect of Google Docs on collaboration. Blau and Caspi [3] ran a small experiment that was designed to compare collaboration on writing documents to merely sharing documents. In their experiment, 118 undergraduate students of the Open University of Israel were randomized to one of five groups in which they shared their written assignments and received feedback from other students to varying degrees, ranging from keeping texts

---

[1] Full-time Google employees, excluding interns, part-times, vendors, etc

[2] Full-time employees who have joined Google for less than 90 days

private to allowing in-text suggestions or allowing in-text edits. None of the students had used Google Docs previously. The authors found that only students in the collaboration group perceived the quality of their final document to be higher after receiving feedback, and students in all groups thought that collaboration improves documents.

This paper takes a different approach, and looks for the effects of collaboration on a large, diverse organization with thousands of users over a much longer period of time. The first part of the paper describes some of the contexts in which Google Docs is used for collaboration, and the second part analyzes how collaboration has evolved over the last two years.

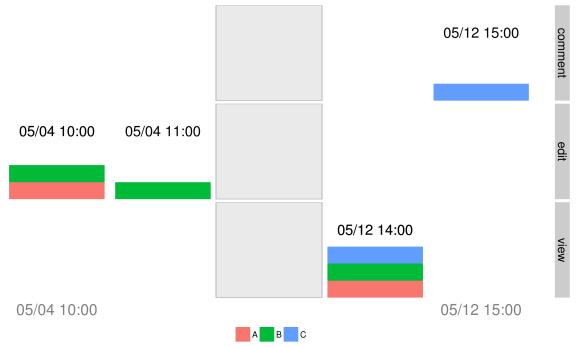## 2   Collaboration Visualization

2.1   The Data

This section introduces a way to visualize the events during a collaboration and some simple statistics that summarize how widespread collaboration using Google Docs is at Google. The graphics and metrics are based on the view, edit and comment actions of all full-time employees on tens of thousands of documents created in April 2013.

2.2   A Simple Example

To start, a document with three collaborators Adam (A), Bryant (B) and Catherine (C) is shown in Figure 1. The horizontal axis represents time during the collaboration. The vertical axis is broken into three regions representing viewing, editing and commenting. Each contributor is assigned a color. A box with the contributor's color is drawn in any time interval in which the contributor was active, at a vertical position that indicates what the user was doing in that time interval. This allows us to see when contributors were active and how often they contributed to the document. Stacking the boxes allows us to show when contributors were acting at

the same time. Only time intervals in which at least one contributor was active are shown, and gaps in time that are shorter than a threshold are ignored. Gray vertical bars of fixed width are used to represent periods of no activity that are longer than the threshold. In this paper, the threshold is set to be 12 hours in all examples.

In Figure 1, an interval represents an hour. Adam and Bryant edited the document together during the hour of 10 AM May 4 and Bryant edited alone in the following hour. The collaboration paused for 8 days and resumed during the hour of 2 pm on May 12. Adam, Bryant and Catherine all viewed the document during that hour. Catherine commented on the document in the next hour. Altogether, the collaboration had two active sessions, with a pause of 8 days between them.

Figure 1: This figure shows an example of the collaboration visualization technique. Each colored block except the gray one represents an hour and the gray one represents a period of no activity. The Y axis is the number of users for each action type. This document has three contributors, each assigned a different color.

Although we have used color to represent collaborators here, we could instead use color to represent the locations of the collaborators, their organizations, or other variables. Examples with different colorings are given in Sections 2.5 and 2.6.

2.3   Collaboration Metrics

To estimate the percentage of users who concurrently edit a document and the percentage of documents which had concurrent editing, we discretize the timestamps of editing actions into 15 minute intervals and consider editing actions by different contributors in the same 15 minute interval to be concurrent. Two users who edit the same document but always more than 15 minutes apart would not be considered as concurrent, although they would still be considered collaborators. Edge cases in which two collaborators edit the same document within 15 minutes of each other but in two adjacent 15 minute intervals would not be counted as concurrent events.
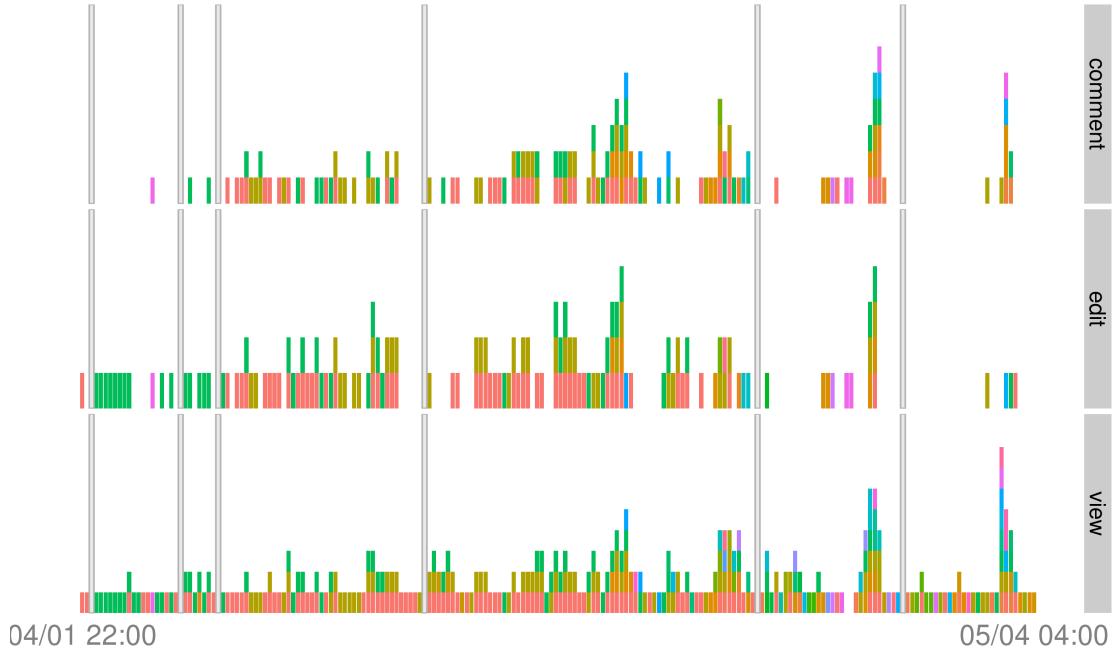
The choice of 15 minutes is arbitrary; however, metrics based on a 15 minute discretization and a 5 minute discretization are little different. The choice of 15 minute intervals makes computation faster. A more accurate approach would be to look for sequences of editing actions by different users with gaps below 15 minutes, but that requires considerably more computing.

2.4   Collaborative Editing

Collaborative editing is common at Google. 53% of the documents that were created and shared in April 2013 were edited by more than one employee, and half of those had at least one concurrent editing session in the following six months. Looking at employees instead of documents, 80% of the employees who edited any document contributed content to a document owned by others and 65% participated in at least one 15 minute concurrent editing session in April 2013. Concurrent editing is sticky, in the sense that 76% of the employees who participate in a 15 minute concurrent editing session in April will do so again the following month.

There are many use cases for collaborative editing, including weekly reports, design documents, and coding interviews. The following three plots show an example of each of these use cases.
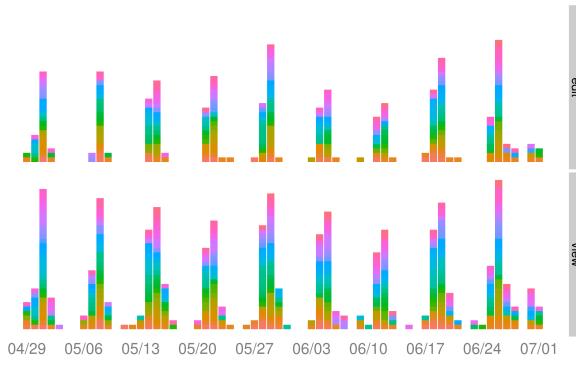

Figure 2: Collaboration activity on a design document. The X axis is time in hours and the Y axis is the number of users for each action type. The document was mainly edited by 3 employees, commented on by 18 and viewed by 50+.

Figure 2 shows the life of a design document created by engineers. The X axis is time in hours and the Y axis is the number of employees working on the document for each action type. The document was mainly edited by three employees, commented on by 18 employees and viewed by more than 50 employees from three major locations. This document was completed within two weeks and viewed many times in the subsequent month. Design documents are common at Google, and they typically have many contributors.

Figure 3 shows the life of a weekly report document. Each bar represents a day and the Y axis is the number of employees who edited and viewed the document in a day. This document has the following submission rules:
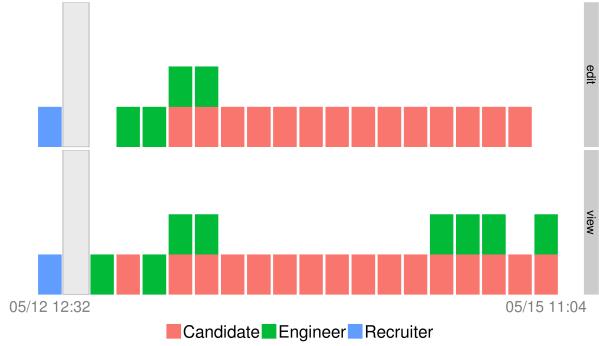
- Wednesday, AM: Reminder for submissions

- Wednesday, PM: All teams submit updates
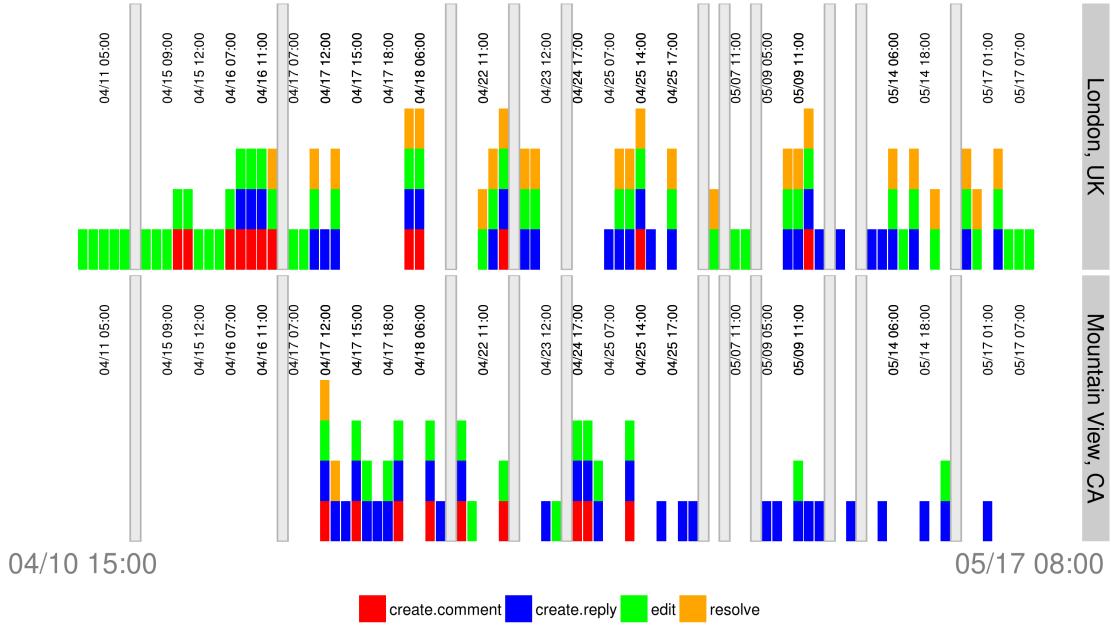
- Thursday, AM: Document is locked

The activities on the document exhibit a pronounced weekly pattern that mirrors the submission rules. Weekly reports and meeting notes that are updated regularly are often used by employees to keep everyone up-to-date as projects progress.

Figure 3: Collaboration on a weekly report. The X axis is time in days and the Y axis is the number of users for each action type. The activities exhibit a pronounced weekly pattern and reflect the submission rules of the document.

Finally, Figure 4 shows the life of a document used in an interview. The X axis represents time in minutes. The document was prepared by a recruiter and then viewed by an engineer. At the beginning of the interview, the engineer edited the document and the candidate then wrote code in the document. The engineer was able to watch the candidate typing. At the end of the interview, the candidate's access to the document was revoked so no further change could be made, and the document was reviewed by the engineer. Collaborative editing allows the coding interview to take place remotely, and it is an integral part of interviews for software engineers at Google.

Figure 4: The activity on a phone interview document. The X axis is time in minutes and the Y axis is the number of users for each action type. The engineer was able to watch the candidate typing on the document during a remote interview.

2.5   Commenting

Commenting is common at Google. 30% of the documents created in April 2013 that are shared received comments within six months of creation. 57% of the employees who used Google Docs in April commented at least once in April, and 80% of the users who commented in April commented again in the following month.
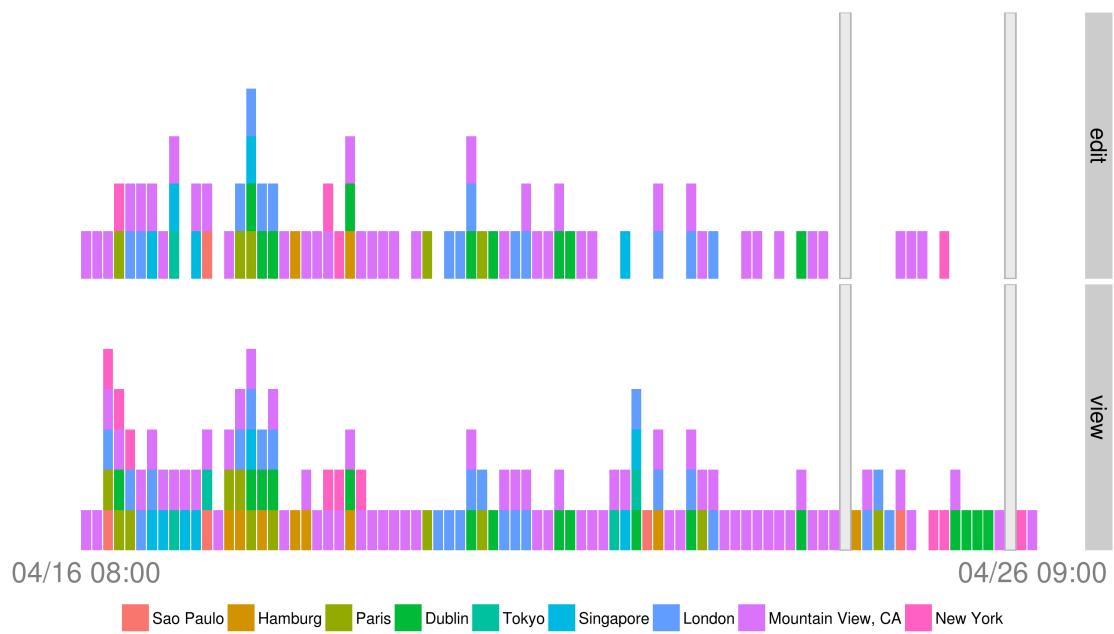
Figure 5: Commenting and editing on a design document. The X axis is time in hours and the Y axis is the number of user actions for each user location. There are four user actions, each assigned a different color. Timestamps are in Pacific time.

Figure 5 shows the life of a design document. Here color represents the type of user action (create a comment, reply to a comment, resolve a comment and edit the document), and the Y axis is split into two locations. The document was written by one engineering team and reviewed by another. The review team used commenting to raise many questions, which the engineering team resolved over the next few days. Collaborators were located in London, UK and Mountain View, California, with a nine hour time zone difference, so the two teams were almost "taking turns" working on the document (timestamps are in Pacific time). There are many similar communication patterns between engineers via commenting to ask questions, have discussions and suggest modifications.
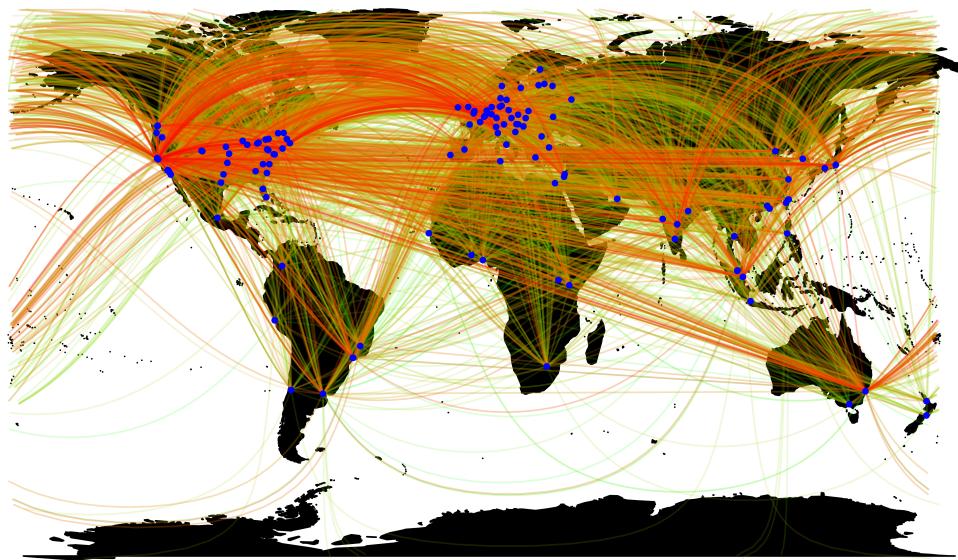
2.6   Collaboration Across Sites

Employees use the Docs suite to collaborate with colleagues across the world, as Figure 6 shows. In that figure, employees working from nine locations in eight countries across the globe contributed to a document that was written within a week. The document was either viewed or edited with gaps of less than 12 hours (the threshold for suppressing gaps in the plot) in the first seven days as people worked in their local timezones. After final changes were made to the document, it was reviewed by people in Dublin, Mountain View, and New York.

Figure 7 shows one month of global collaborations for full-time employees using Google Docs. The blue dots show the locations of the employees and a line connects two locations if a document is created in one location and viewed in the other. The warmer the color of the line, moving from green to red, the more documents shared between the two locations.
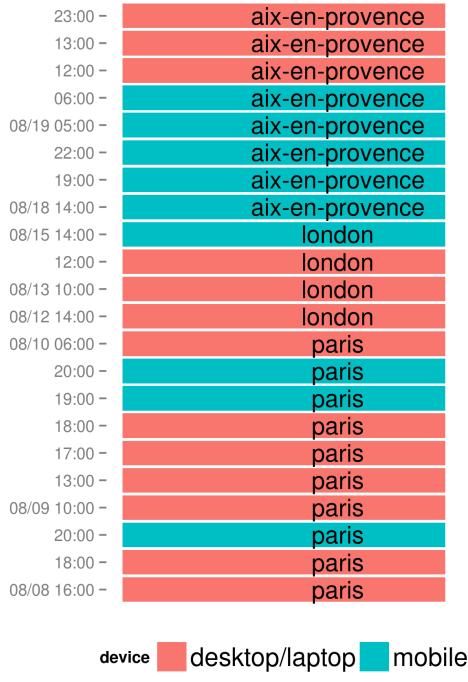
Figure 6: Activity on a document. Each user location is assigned a different color. The X axis is time in hours and the Y axis is the number of locations for each action type. Users from nine different locations contributed to the document.

Figure 7: Global collaboration on Docs. The blue dots are locations and the dots are connected if there is collaboration on Google Docs between the two locations.
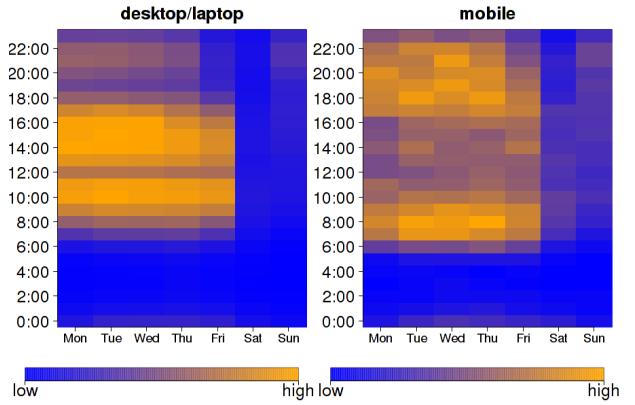
2.7   Cross Device Work

The advantage of cloud-based software and storage is that a document can be accessed from any device. Figure 8 shows one employee's visits to a document from multiple devices and locations. When the employee was in Paris, a desktop or laptop was used during working hours and a mobile device during non-working hours. Apparently, the employee traveled to Aix-En-Provence on August 18. On August 18 and the first part of August 19, the employee continued working on the same document from a mobile device while on the move.

Figure 8: Visits to a document by one user working on multiple devices and from multiple locations.

Not surprisingly, the pattern of working on desktops or laptops during working hours and on mobile devices out of business hours holds generally at Google, as Figure 9 shows. The day of week is shown on the X axis and hour of day in local time on the Y axis. Each pixel is colored according to the average number of employees working in Google Docs in a day of week and time of day slot, with brighter colors represent-

ing higher numbers. Pixel values are normalized within each plot separately. Desktop and laptop usage of Google Docs peaks during conventional working hours (9:00 AM to 11:00 AM and 1:00 PM to 5:00 PM), while mobile device usage peaks during conventional commuting and other out-of-office hours (7:00 AM to 9:00 AM and 6:00 PM to 8:00 PM).

Figure 9: The average number of active users working in Google Docs in each day of week and time of day slot. The X axis is day of the week and the Y axis is time of the day in local time. Desktop/Laptop usage peaks during working hours while mobile usage peaks at out-of-office working hours.

3   The Evolution of Collaboration

3.1   The Data

This section explores changes in the usage of Google Docs over time. Section 2 defined collaborators as users who edited or commented on the same document and used logs of employee editing, viewing and commenting actions to describe collaboration within Google. This section defines collaborators differently using metadata on documents. Metadata is much less rich than the event history logs used in Section 2, but metadata is retained for a much longer period of time.

Document metadata includes the document creation time and the last time that the document

was accessed, but no other information about its revision history. However, the metadata does include the identification numbers for employees who have subscribed to the document, where a subscriber is anyone who has permission to view, edit or comment on a document and who has viewed the document at least once. Here we use metadata on documents, slides and spreadsheets.

We call two employees collaborators (or subscription collaborators to be clear) if one is a subscriber to a document owned by the other and has viewed the document at least once and the document has fewer than 20 subscribers. The owner of the document is said to have shared the document with the subscriber. The number of subscribers is capped at 20 to avoid overcounting collaborators. The more subscribers the document has, the less likely it is that all the subscribers contributed to the document.
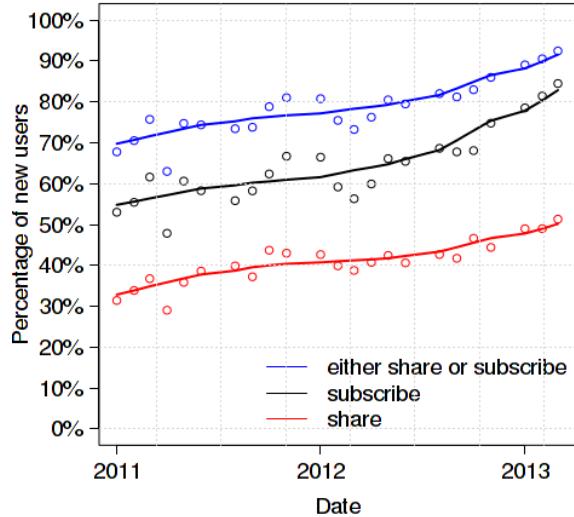
There is no timestamp for when the employee subscribed to the document in the metadata, so the exact time of the collaboration is not known. Instead, the document creation time, which is known, is taken to be the time of the collaboration. An analysis (not shown here) of the event history data discussed in Section 2 showed that most collaborators join a collaboration soon after a document is created, so taking collaboration time to be document creation time is not unreasonable. To make this assumption even more tenable, we exclude documents for which the time of the last view, comment or edit is more than six months after the document was created. This section uses metadata on documents created between January 1, 2011 and March 31, 2013. We say that two employees had a subscription collaboration in July if they collaborated on a document that was created in July.

3.2   Collaboration for New Employees

Here we define the new employees for a given month to be all the employees who joined Google no more than 90 days before the beginning of the month and started using Google Docs in the given month. For example, employees called new in the month of January 2011 must have

joined Google no more than 90 days before January 1, 2011 and used Google Docs in January 2011. Each month can include different employees. New employees are said to share a document if they own a document that someone else subscribed to, whether or not the person subscribed to the document is a new employee. Similarly, a new employee is counted as a subscriber, regardless of the tenure of the document creator.
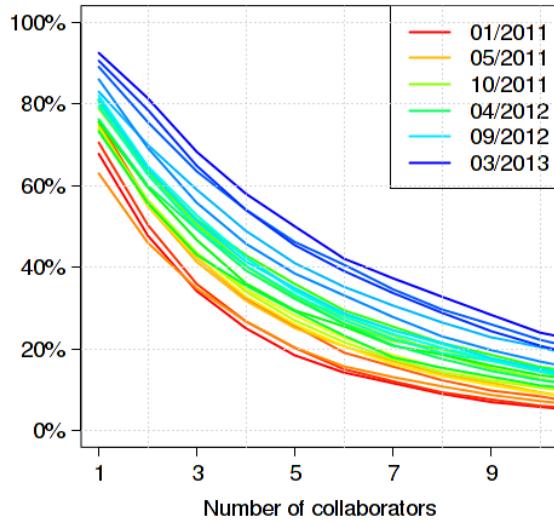
Figure 10 shows that collaboration among new employees has increased since 2011. Over the last two years, subscribing has risen from 55% to 85%, sharing has risen from 30% to 50%, and the fraction of users who either share or subscribe has risen from 70% to 90%. In other words, new employees are collaborating earlier in their career, so there is a faster ramp-up and easier access to collective knowledge.

Figure 10: This figure shows the percentage of new employees who share, subscribe to others' documents and either share or subscribe in each one-month period over the last two years.

Not only do new employees start collaborating more often (as measured by subscription and sharing), they also collaborate with more people. Figure 11 shows the percentage of new employees with at least a given number of collaborators by month. For example, the percentage of

new employees with at least three subscription collaborators was 35% in January 2011 (the bottom red curve) and 70% in March 2013 (the top blue curve), a doubling over two years. It is interesting that the curves hardly cross each other and the curves for the farthest back months lie below those for recent months, suggesting that there has been steady growth in the number of subscription collaborators per new employee over this period.

	25%	50%	75%	90%	95%
January 2011	0	1	4	7	11
January 2013	2	5	10	17	22

Table 1: This table shows the percentile of number of collaborators a new employee have in January 2011 and January 2013. The entire distribution shifts to the right.

3.3   Collaboration in Sales and Marketing

Section 3.2 compared new employees who joined Google in different months. This section follows current employees in Sales and Marketing who joined Google before January 1, 2011. That is, the previous section considered changes in new employee behavior over time and this section considers changes in behavior for a fixed set of employees over time. We only analyze subscription collaborations among this fixed set of employees and collaborations with employees not in this set are excluded.

Figure 11: This figure shows the proportion of new employees who have at least a given number of collaborators in each one-month period. Each period is assigned a different color. The cooler the color of the curve, moving from red to blue, the more recent the month. The legend only shows the labels for a subset of curves. The percentage of new employees who have at least three collaborators has doubled from 35% to 70%.
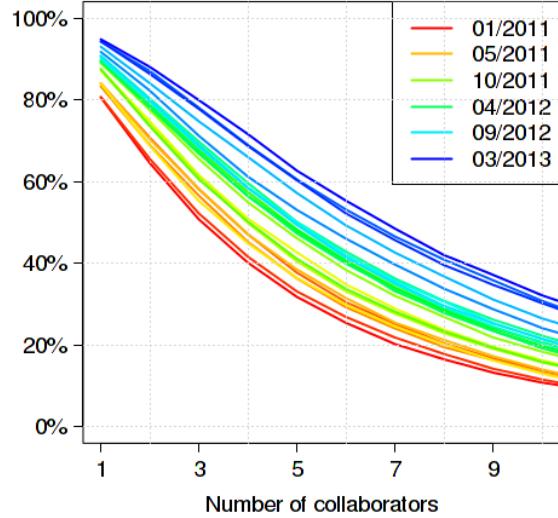
To present the data in Figure 11 in another way, Table 1 shows percentiles of the distribution of the number of subscription collaborators per new employee using Google Docs in January 2011 and in January 2013. For example, the lowest 25% of new employees using Google Docs had no such collaborators in January 2011 and two such collaborators in January 2013.
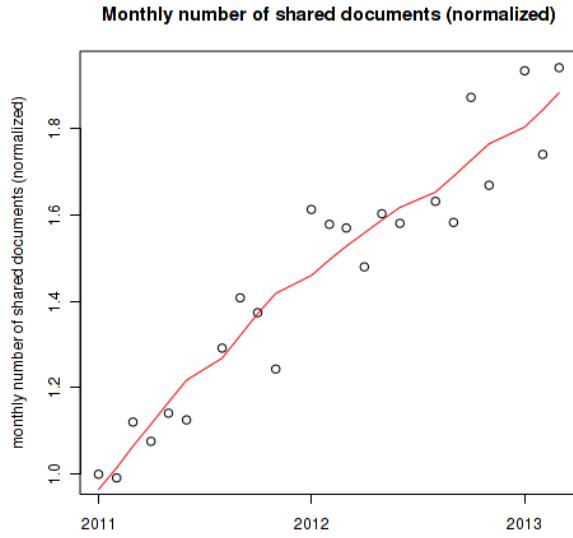

Figure 12: This figure shows the percentage of current employees in Sales and Marketing who have at least a given number of collaborators in each one-month period.

Figure 12 shows the percentage of current employees in Sales and Marketing who have at least

a given number of collaborators at several times in the past. There we see that more employees are sharing and subscribing over time because the fraction of the group with at least one subscription collaborator has increased from 80% to 95%. And the fraction of the group with at least three subscription collaborators has increased from 50% to 80%. It shows that many of the employees who used to have no or very few subscription collaborators have migrated to having multiple subscription collaborators. In other words, the distribution of number of subscription collaborators for employees who have been in Sales and Marketing since January 1, 2011 has shifted right over time, which implies that collaboration in that group of employees has increased over time.

Finally, the number of documents shared by the employees who have been in Sales and Marketing at Google since January 1, 2011 has nearly doubled over the last two years. Figure 13 shows the number of shared documents normalized by the number of shared documents in January, 2011.
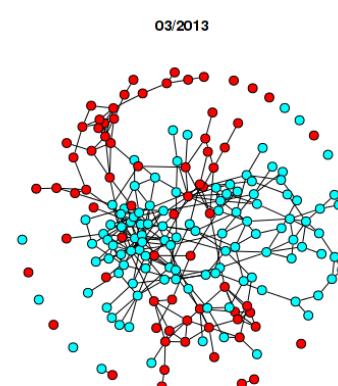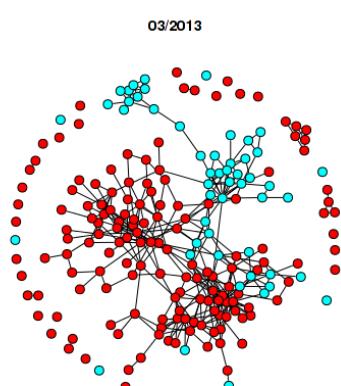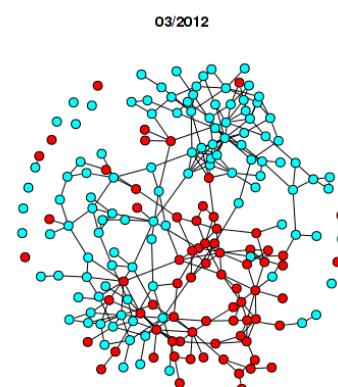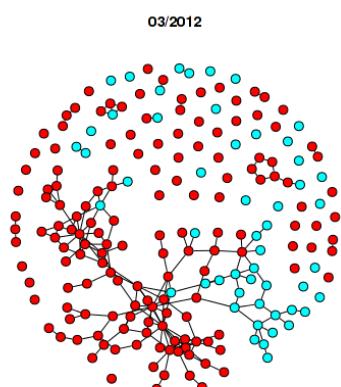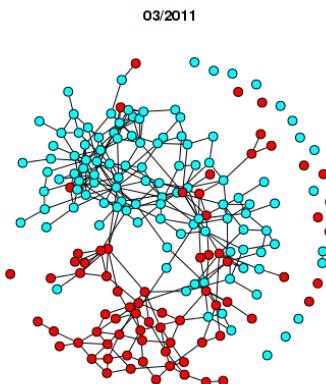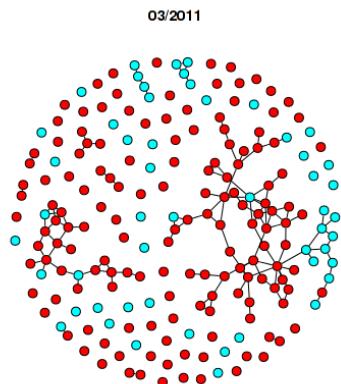

Figure 13: This figure shows the number of shared documents created by employees in Sales and Marketing each month normalized by the number of shared documents in January 2011. The number has almost doubled over the last two years.

3.4   Collaboration Between Organizations

Collaboration between organizations has increased over time. To show that, we consider hundreds of employees in nine teams within the Sales and Marketing group and the Engineering and Product Management group who joined Google before January 1, 2011, were still active in March 31, 2013 and used Google Docs in that period. Figure 14 represents the Engineering and Product Management employees as red dots and the Sales and Marketing employees as blue dots. The same dots are included in all three plots in Figure 14 because the employees included in this analysis do not change. A line connects two dots if the two employees had at least one subscription collaboration in the month shown. The denser the lines in the graph, the more collaboration, and the more lines connecting red and blue dots, the more collaboration between organizations. Clearly, subscription collaboration has increased both within and across organizations in the past two years. Moreover, the network shows more pronounced communities (groups of connected dots) over time. Although there are nine individual teams, there seems to be only three major communities in the network. Figure 14 indicates that teams can work closely with each other even though they belong to separate departments.

We also sampled 187 teams within the Sales and Marketing group and the Engineering and Product Management group. Figure 15 represents teams in Engineering and Product Management as red dots and teams in Sales and Marketing as blue dots. Two dots are connected if the two teams had a least one subscription collaboration between their members in the month. Figure 15 shows that the collaboration between those teams has increased and the interaction between the two organizations has becomed stronger over the past two years.
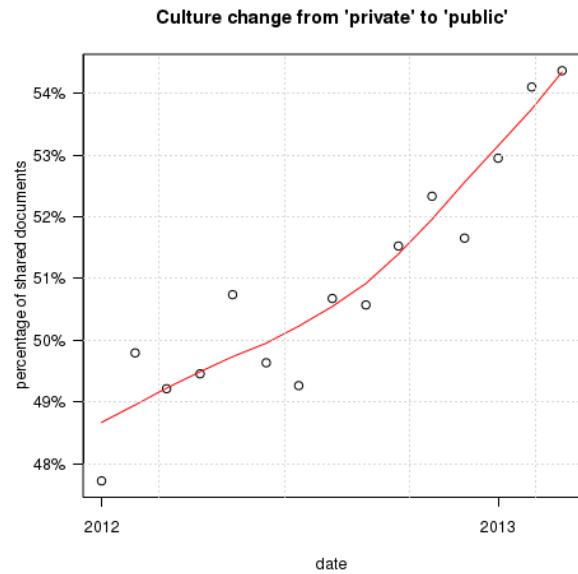
Figure 14: An example of collaboration across organizations. Red dots represent employees in Engineering and Product Management and blue dots represent employees in Sales and Marketing

Figure 15: An example of collaboration between teams. Red dots represent teams in Engineering and Product Management and blue dots represent teams in Sales and Marketing

3.5 Cultural Changes in Collaboration

Google Docs allows users to specify the access level (visibility) of their documents. The default access level in Google Docs is private, which means that only the user who created the document or the current owner of the document can view it. Employees can change the access level on a document they own and allow more people to access it. For example, the document owner can specify particular employees who are allowed to access the document, or the owner can mark the document as public within Google, in which case any employee can access the document. Clearly, not all documents created in Google can be visible to everyone at Google, but the more documents are widely shared, the more open the environment is to collaboration.

Figure 16: This figure shows the percentage of shared documents that are "public within Google" created in each month. Public sharing is overtaking private sharing at Google.

Figure 16 shows the percentage of shared documents in Google created each month between January 1, 2012 and March 31, 2013 that are public within Google. The red line, which is a curve fit to the data to smooth out variability, shows that the percentage has increased about 12% from 48% to 54% in the last year alone. In that sense, the culture of sharing is changing in Google from private sharing to public sharing.

4 Conclusions

We have examined how Google employees collaborate with Docs and how that collaboration has evolved using logs of user activity and document metadata. To show the current usage of Docs in Google, we have developed a visualization technique for the revision history of a document and analyzed key features in Docs such as collaborative editing, commenting, access from anywhere and on any device. To show the evolution of collaboration in the cloud, we have analyzed new employees and a fixed group of employees in Sales and Marketing, and computed collaboration network statistics each month. We find that employees are engaged in using the Docs suite, and collaboration has grown rapidly over the last two years.

It would also be interesting to conduct a similar analysis for other enterprises and see how long it would take them to reach the benchmark Google has set for collaboration on Docs. Not only has the collaboration on Docs changed at Google, the number of emails, comments on G+, calender meetings between people who work together has also had significant changes over the past few years. How those changes reinforce each other over time would also be an interesting topic to study.

Acknowledgements

References

[1] Dan R. Herrick (2009). *Google this!: using Google apps for collaboration and productivity.* Proceeding of the ACM SIGUCCS fall conference (pp. 55-64).

[2] Stijn Dekeyser, Richard Watson (2009). *Extending Google Docs to Collaborate on Research Papers.* Technical Report, The University of Southern Queensland, Australia.

[3] Ina Blau, Avner Caspi (2009). *What Type of Collaboration Helps? Psychological Ownership, Perceived Learning and Outcome Quality of Collaboration Using Google Docs.* Learning in the technological era: Proceedings of the Chais conference on instructional technologies research (pp. 48-55).