Have You Heard?: How Gossip Flows Through Workplace Email

Tanushree Mitra and Eric Gilbert
School of Interactive Computing & GVU Center
Georgia Institute of Technology
{tmitra3, gilbert}@cc.gatech.edu

Abstract

We spend a significant part of our lives chatting about other people. In other words, we all gossip. Although sometimes a contentious topic, various researchers have shown gossip to be fundamental to social life—from small groups to large, formal organizations. In this paper, we present the first study of gossip in a large CMC corpus. Adopting the Enron email dataset and natural language techniques, we arrive at four main findings. First, workplace gossip is common at all levels of the organizational hierarchy, with people most likely to gossip with their peers. Moreover, employees at the lowest level play a major role in circulating it. Second, gossip appears as often in personal exchanges as it does in formal business communication. Third, by deriving a power-law relation, we show that it is more likely for an email to contain gossip if targeted to a smaller audience. Finally, we explore the sentiment associated with gossip email, finding that gossip is in fact quite often negative: 2.7 times more frequent than positive gossip.

Introduction

Email 1: hey - seems like we aren't the only ones that think kyle is an arrogant asshole anymore – susan told me that on saturday night she had it out with him and doesn't want him around – also, apparently he's been treating dana like shit and it's starting to get noticed by other people – just thought this was an interesting development.

Email 2: Here's the third party assessment of current western supply issues that we're most in agreement with. Sam Van Vactor's group publishes the daily energy market report that's widely read, and Pickel is with Tabors Caramanis, consultants that we have employed on several issues.

In both of these messages, the sender discusses someone who is not on the email. Anthropologists call conversations like these *gossip:* the absence of a third party from the conversation (Besnier 1989; Hannerz 1967). Despite some negative social connotations, gossip is fundamental to healthy societies—from small groups to large, formal organizations (Feinberg et al. 2012). Simply put, we use it to trade social information, information we may find very useful in the future. In fact, Dunbar (1994) goes so far as to suggest that language itself developed so we could gossip about one another. Entire technologies have even arisen to support the practice, such as the dumbwaiter, a rope-and-pulley system designed in the eighteenth century to guard the masters' gossip from nosy servants (Goffman 1959).

Following in this line of scholarship, this paper presents the first study of gossip in a large corpus of computer-mediated communication (CMC). Using the Enron email dataset of 517,431 messages, we look to answer the following research questions. In systems characterized by power and hierarchy—like workplaces—what role does hierarchy play in shaping how people gossip? Going further, can we infer someone's corporate rank from their gossip behavior? Is gossip limited to personal email exchanges, or does it leak into more formal business communication? How does the size of an email's recipient list affect the likelihood of gossip in the body of the message? And finally, do certain phrases and emotions characterize gossip via email?

Though sometimes overlooked in an always-changing internet, email was the internet's first widespread social medium (Henderson Jr. and Myer 1977). Email affords conversations among both small and large groups. Networks of contacts form over time, like Twitter. Unlike Twitter, however, 92% of online adults use email (Purcell 2011). Madden and Jones (2008) recently reported a sharp increase in the number of adults who "constantly" check their work email, a figure that has almost certainly risen as smartphones find their way into more and more pockets. In other words, it may be fair to call email the world's most successful and pervasive type of social media.

With this as a backdrop, we turn to natural language methods—specifically, Named Entity Recognition—to identify gossip in the Enron corpus. We find it present at all levels of the corporate hierarchy. We demonstrate hierarchical signatures of gossip, showing specific pathways for the transmission of gossip via email. People belonging to certain ranks are the major sources of these messages, while other ranks silently receive it. Yet others do both. We find that people gossip most with their peers, indicating their tendency to gossip within their own group, the ones belonging to the same rank. Interestingly, people have a greater likelihood to send gossip messages to smaller audiences: a fact demonstrated by deriving a power law relation between the frequency of gossip email and the number of recipients on an email.

After exploring gossip as framed by hierarchical structure, we take a closer look at the content of gossip messages. Using sentiment analysis, we search for emotional signals in gossip. We explore the sentiment associated with gossip email, finding that gossip is in fact quite often negative: 2.7 times more frequent than positive gossip. We see the primary contribution of this paper as an exploratory study of an important social process, albeit one that is sometimes hidden.

Related Work

Next we give a brief overview of some of the theories in gossip research relevant to our current work. There are two conflicting approaches of examining gossip (Gluckman 1963; Paine 1967). While Gluckman (1963) considers gossip to be a group behavior, providing coercive power, unity and regulation to the group, Paine (1967) considers it to be the outcome of individual self-interest. In this paper, we look at organizational gossip from both these perspectives.

Feinberg et al. (2012) recently published work on the prosocial benefit of strictly negative gossip. An experimental piece, Feinberg et al. argue that sharing negative information about an absent third party promotes group cooperation and may prevent others from being antisocial. Prior to this, Kurland and Pelled (2000) proposed that positive gossip might build the gossipers reputation and give him "reward power" over the recipients. In this line of reasoning, the recipient perceives him as someone who spreads good news and thus can help him build his reputation and promote his career. Negative gossipers on the other hand will be treated with caution, thus giving the gossiper "coercive power" over the recipient. However, this work did not provide any empirical evidence supporting this hypothesis.

Whether positive or negative, the value of gossip has often been a subject of dispute. Rosnow (1977) uses the marketplace metaphor for human interactions wherein gossip is a valuable social commodity, exchanged in return of more information, entertainment, social control, status and power. In fact, Foster's (2004) literature survey on gossip research summarizes the four main social functions of gossip: information, entertainment, intimacy and influence. Building on these theories, in this paper we closely study the content of workplace gossip so as to search for signs of these categories.

Method

Now we present the dataset used and the steps employed to isolate gossip email, in accordance with the definitions provided earlier (Besnier 1989; Hannerz 1967). Our research is based on four complimentary datasets:

Enron email corpus: This dataset has 517,431 email messages[1], sent by 151 people between 1997 to 2002 (Klimt and Yang 2004; Shetty and Adibi 2004).

[1]http://www.cs.cmu.edu/~enron

Enron job titles dataset: Researchers at USC[2] and John Hopkins gathered the status of 132 employees within Enron and generated a job title dataset for them. Figure 1 shows the job titles assigned to employees (Shetty and Adibi 2004).

Ranks of job titles: We referred to Gilbert's work (2012) to match each job title with a numeric rank relative to its position in the organizational hierarchy. This is, in turn, based on earlier work (Palus, Bródka, and Kazienko 2010). CEOs and Presidents have greatest power in an organization and reside at the top of the hierarchy. They are assigned rank 6, while employees are at the lowest level have rank 0. Vice Presidents and Directors occupy the second-highest level. They are followed by the In-House Lawyers. The subsequent levels are occupied by Managers, followed by Traders, and the second-to-last level belonging to Specialists and Analysts. Figure 1 depicts the relationships.

Personal vs. business email: Jabbari et al. (2006) manually annotated a subset of the CMU Enron email dataset, labeling 11,220 messages as "Business" and 3,598 as "Personal." We use this dataset when we analyze gossip in personal versus business email.

Unit of Analysis: Gossip email messages

We now discuss our computational methods for isolating email messages containing gossip. Using the Enron corpus, we remove all duplicate messages; i.e., we keep only the sender's copy of the message and discard any copies shared by other recipients. We also filter out those messages where the sender is a non-Enron employee or if his rank is unknown. These form the backbone of our filtering process. Figure 2 outlines the steps in the process. Later, we do additional filtering depending on the type of analysis. We cover these in detail in the relevant sections.

After the filtering steps, we scan the body of each email to check if the sender has mentioned a name of a person and has not included him in the recipient list. Email messages satisfying this criteria are termed *gossip email messages*. They form our unit of analysis. We used the Stanford Named Entity Recognition (NER) classifier (Finkel, Grenager, and Manning 2005) to label words in the email body as person or company names. Only those email messages which have person names in the email body are considered as potential gossip. It is a common practice to shorten a person's full name (e.g., Abe for Abraham). NER also labels these nicknames as person names. In order to find the corresponding full names, we borrowed a nickname lookup file from an open source project hosted by the "Web Science and Digital Libraries Research Group" of Old Dominion University[3]. We check if any of the labelled person names are present in the nickname database. We then map any matched nicknames to its corresponding full name. For each email message we call the list of all these full names "in-message names."

[2]http://www.isi.edu/~adibi/Enron/Enron.htm
[3]http://bit.ly/m7tYcC

Rank	Position or Job Title	
6	CEO	President
5	Vice President	Director
4	In-House Lawyer	
3	Manager	
2	Trader	
1	Specialist	Analyst
0	Employee	

Figure 1: Relative ranks of job titles. Figure has been reproduced from earlier work (Gilbert 2012).

Messages in the Enron corpus follow a standard format with well-defined headers: *Message-id, Date, From, To, CC, BCC, Subject* and X-Fields containing Enron Active Directory names (e.g., *X-From, X-To, X-CC*). We use the information in the X-Fields to find the full names of senders and recipients. Next, we check to see if all the in-message names found in the earlier step are present in the recipient list. If not, then the missing names are the ones about whom the sender gossiped in the email. These gossip email messages comprise our corpus.

It is important to note some limitations inherent to our approach. The NER classifier might not be the best classifier to label text in an email message because it has been trained on newswire data. Thus, the classifier may mis-label some names in the email messages, and in the worst-case, may under-estimate the number of gossip email. At present, we cannot see a way around this limitation and build our analyses to deal with it. At the same time, we will likely report few false positives due to over-estimation.

We begin by first understanding how gossip behavior varies with rank of an employee. Next, we explore if social factors like frequency of interaction through email or the nature of relationships (personal vs. business) affect gossip. Finally, we explore the language behind the content of gossip email.

Analysis & Results

Does position of an employee influence the amount of gossip email he sends and the audience of his gossip messages? In other words, who gossips more, bosses or employees?

Who starts the gossip? We first study the percentages of gossip email originating from each rank. After the basic filtering step, we further restrict our dataset by keeping only those messages where the recipient list contains at least one Enron employee. By doing so, we allow mixtures of ranks in the recipient list, (e.g., a Vice President mailing a trader and an assistant).

After these filtering steps, we were left with a corpus of 49,393 messages of which 7,206 were gossip email. Figure 3 shows the proportion of these email by rank. We see that gossip is a common phenomenon among every rank. One important point to note here is that the percentages are calculated relative to the email traffic originating from each rank. This means that with lower email traffic, a small amount of gossip email would result in higher percentages.

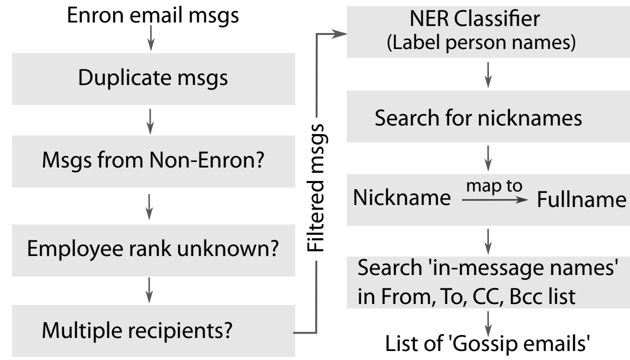

Figure 2: Steps for identifying gossip email from a list of Enron email messages.

Where does gossip go? Now, let's consider the opposite question: who receives the most gossip email? Combined with the result above, this analysis helps us see how gossip flows through the hierarchy. In contrast to the work above, we kept only those message which had exactly one recipient in the *To* list and where we knew the recipient and sender ranks. This resulted in a rather small dataset of 845 email messages. We restricted our analysis to a dataset of single recipients because multiple recipients may belong to different ranks and such a mixture of ranks might be confusing for conclusions about the audience of gossip email and the flow of gossip across ranks. We define "rank difference" as the rank of the recipient minus the rank of the sender. We see a huge peak at rank difference 0 (see Figure 4). This implies that people mostly gossip with their peers (i.e., other employees belonging to the same rank). The next highest peak is at rank difference 1, implying that there is heavy

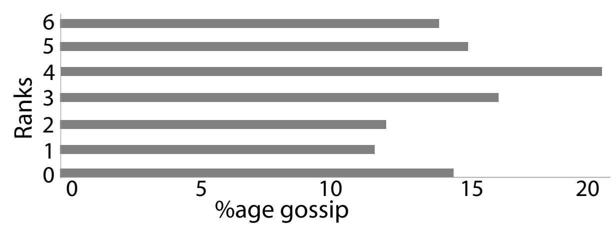

Figure 3: Gossip proportion varying with hierarchical rank.

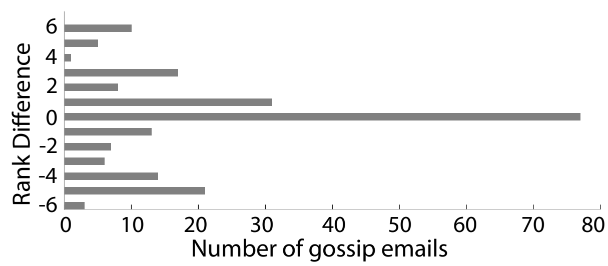

Figure 4: Rank difference versus number of gossip email, where *"Rank Difference = Recipient Rank − Sender Rank"*. Positive differences indicates recipient is of higher rank than the sender, while negative differences indicate that recipient is lower in the hierarchy than the sender.

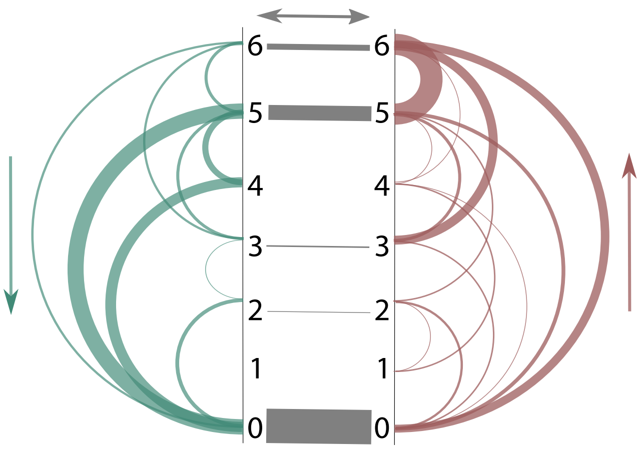

Figure 5: Flow of gossip across ranks. (↑) denotes that gossip email flow up the hierarchy, while (↓) denotes downward flow. (↔) denotes that gossip stays within the same organizational rank.

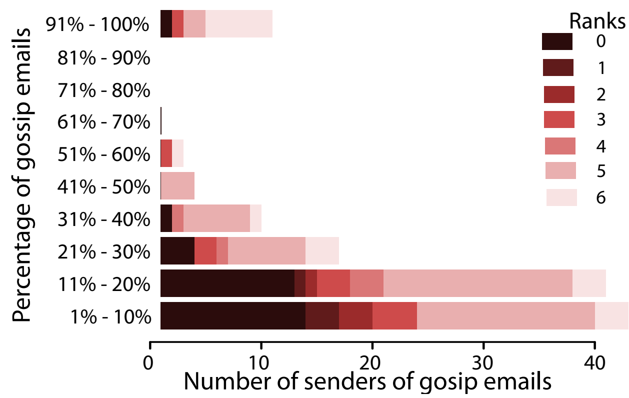

Figure 6: Proportion of gossip email out of total sent mail, broken down by organizational rank. Color code denotes the rank of the sender. The vast majority of people devote 1%–20% of their email to gossip, while a handful gossip nearly all the time (top bar).

flow of gossip messages one level up the hierarchy. There is also significant flow of gossip email four levels down the hierarchy, corresponding to the rank difference of -4.

What's behind these peaks? We produced Figure 5 to answer this question. Each arc in the figure corresponds to the flow of gossip email between any two ranks. The arcs on the right side correspond to flow up (↑) the hierarchy, while those on the left side correspond to flow down (↓) the hierarchy. The thickness of the arc is proportional to the amount of gossip email sent. The thickness gives us a sense of who the major contributors are, moving gossip messages up or down the hierarchy, or keeping them at the same level. The employees with rank 0 gossip the most amongst themselves, compared to how much they push messages up the hierarchy. Vice Presidents and Directors (rank 5) often move gossip email up the hierarchy, i.e., to their immediate next superiors, Enron's CEOs. Their contribution results in the peak at rank difference 1 in Figure 4. They are even influential in the gossip flow down the hierarchy, as depicted by the thick arc from 5 to 0.

In-House Lawyers with rank 4 contribute the second-highest amount of downward-flow gossip. One interesting thing to note from Figure 5 are the distinct "gossip sinks" and "gossip sources" present in either direction. 'Gossip sources' correspond to the ranks which are the major contributors in generating gossip email, while 'gossip sinks' correspond to the ranks which receive most of the gossip. Ranks 6 and 0 are the "gossip sinks" up and down the hierarchy respectively. Ranks 5, 3 and 0, on the other hand, are the major "gossip sources" for gossip flowing up the hierarchy, while the same is true for 5 and 4 for downward flow. This clearly indicates that employees at the lowest level play a prime role in circulating gossip throughout the hierarchy. We interpret these results further in our *Discussion* section.

Individual gossip across ranks. Having explored group-level gossip behavior where a rank denotes a group, we now turn our attention to individual users. We took our original 49,393 message corpus and found each user's rank and the proportion of gossip email each one sends. To get a sense of the distribution of individual user's gossip percentages, we plot a frequency histogram, arbitrarily choosing a bin size of 10%. Figure 6 illustrates the result.

We see that most individuals spend 1% to 20% of their email on gossip. Users belonging to rank 0 have gossip proportions spanning all ranges except 71% to 90%. In-House Lawyers (rank = 4) have a more restricted spread, present in just three bins. The CEOs show an interesting behavior not revealed earlier. There are some CEOs within the corporation whose entire sent email folder contains gossip. This is shown by the rightmost stack in the 91% – 100% bin.

Gossip in personal vs. business email

Researchers have shown that the interpersonal nature of email (personal or business) affects its formality (Peterson, Hohensee, and Xia 2011). We wanted to investigate whether we see a similar pattern in gossip. For this study, we used Jabbari et al.'s (2006) manually annotated 3,598 "Personal" and 11,220 "Business" email dataset. We applied a two step filtering process. First, we removed all duplicate messages similar to what we did for the larger Enron corpus. That is, we keep only one copy of a message. We were left with 9,625 business messages and 3,113 personal messages.

We found that a significant portion of these messages did not follow the standard format (Group 2001) with well defined X-Field headers. We decided to ignore all such messages for two reasons: to be consistent with the dataset used for earlier analysis and because some of these non-standard messages were system-generated. However unlike the filtering process of the larger Enron email corpus, we do not discard messages where the sender's rank is unknown. Since our gossip analysis in this section is independent of ranks, we decided to keep these messages. After this final filtering step, we were left with 1,618 business messages and 1,613 personal messages. We searched this corpus for gossip email (see Figure 7). We find that the proportion of gossip is independent of whether the email relates to personal matters or business ones, a seemingly counterintuitive result, $\chi^2(1, N = 1618) = 0.1413$, $p = 0.707$.

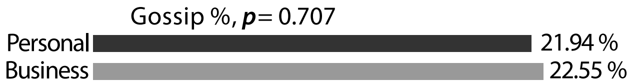

Figure 7: Testing the effect of personal and business email on the proportion of gossip.

Interaction frequency and gossip

Two people who communicate frequently are closer to one another than people with infrequent communication (Gilbert and Karahalios 2009). Peterson et al. (2011) have shown that this closeness affects the rate of formality in email. Does this closeness influence the proportion of gossip email exchanged as well? With respect to the Enron corpus, we measure closeness by the amount of email messages exchanged between exactly two people. Thus we restrict our dataset to a subset of email, where each message had exactly one recipient and both the recipient and the sender are Enron employees. This is the same dataset that we used in our earlier analysis of gossip across ranks. It contains 845 email messages and we scan through them in search of gossip messages.

Table 1 shows the proportion of gossip email partitioned into several buckets. The partitioning is done in a way, so that the number of data points in each bucket is roughly the same. The results are somewhat surprising because we see that the number of gossip email decrease as frequency of contact increases. These results have several interpretations which we will return to later.

Gossip as a function of audience size

We also explore the relationship between the number of gossip messages and the number of recipients on that mail. Our test bed for this analysis consists of all email that had more than one recipient in the *To* list. We were left with a dataset of 16,500 email messages. We searched for gossip email in this corpus and noted the count of its corresponding recipients. Both regular email and gossip email roughly follow a power-law function (see Figure 8). Letting y be frequency and x be the number of recipients on the *To* list, we can model the following relationship: $y \propto x^{-a}$.

The exponent of the fitted line is a = 1.304 for gossip traffic and $a = 1.573$ for overall email traffic of the Enron corpus. These exponents demonstrate that sending email to a

Interaction frequency	Pairs	Total msgs	Gossip msgs (count)	Gossip msgs (%)
1	26	26	26	100%
2	17	34	19	55.9%
3 – 5	24	86	32	37.2%
6 – 11	23	173	42	24.3%
12 – 62	23	526	94	17.9%

Table 1: Influence of interaction frequency on gossip. Here, social contact is measured as the number of messages exchanged between two people. Perhaps surprisingly, gossip decreases as interaction frequency increases.

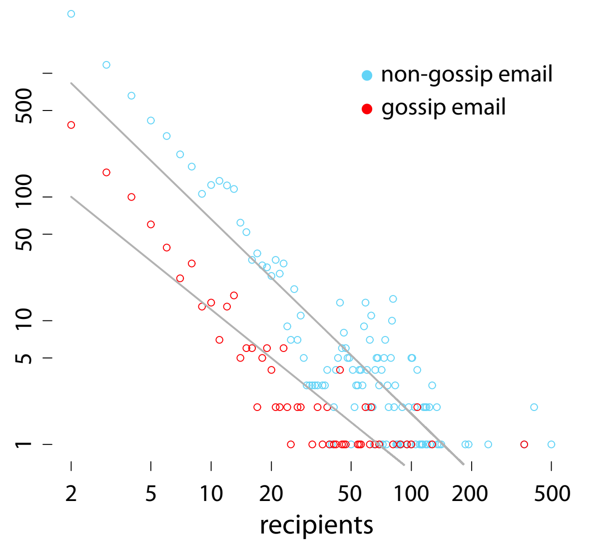

Figure 8: Log-log plot of the number of recipients in the *To* list of an email versus frequency of such an email, shows that the power law relation holds true for both 'gossip email' and overall email traffic.

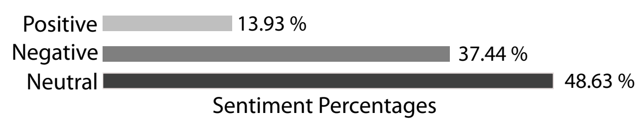

Figure 9: Percentage negative, positive and neutral emotions in the gossip email.

small set of people is more frequent and it is more common to see gossip in messages targeted to a smaller audience.

Content analysis of gossip email

Thus far, we have ignored the content of the messages, focusing on social factors surrounding the people behind the messages. Is there any specific mood attached to gossip email? Gossip is commonly associated with negative emotional valence. Does that come out in the data? Can we identify some commonly occurring phrases? Do the phrases cluster in some meaningful way? Can they reveal the social functions associated with organizational gossip? This section aims to answer these questions.

Is gossip email mostly positive or negative? Kurland and Pelled (2000) say that an organization has both positive and negative gossip. We wanted to discover their proportions. Therefore, we performed sentiment analysis on the texts of all the gossip email. We extracted the message body from each of 7,206 gossip messages and converted the text to lowercase. We then used the Natural Language Text Processing API provided by `text-processing.com`[4] to perform sentiment analysis. We find that a significant portion of the gossip text has neutral tone. However negative sentiment is predominantly higher compared to positive (see Figure 9).

[4] `http://bit.ly/yGMudg`

Gossip phrase analysis. To answer the second set of questions related to the content of gossip texts, we fetch the phrases from the message bodies. We follow the trigram bag of words model to extract all possible unigrams (single word), bigrams (two words) and trigrams (three words) occurring in every gossip message. We convert the text from the message body to lowercase and discard any phrase solely comprised of stop words. Next we throw away phrases that do not occur in at least fifteen gossip messages. This ensures that we take into account the relatively common English phrases. In this way, we obtained a dataset of 6,778 phrases.

Next, we use penalized logistic regression `glmnet`[5] to help us understand the relative power of the phrases in gossip messages. `glmnet` takes as input a predictive feature vector and predicts a binary response variable, while taking care of highly correlated and sparse inputs. After `glmnet` fits our data we obtain the β coefficients associated with the predictive phrases. These coefficients allow us to investigate the relative power of the phrases in predicting the email to be a "gossip email". One important point to note is that β not only identifies the most distinctive phrases, but also the most inconspicuous ones. In order to control for the second factor we take into account only those phrases which occur in at least fifteen gossip messages. We find that of the 6,778 phrases, 6,527 have β coefficients significantly different from zero at 0.001 level of significance. We extracted the top 100 phrases with the most positive β weights and found that 74 of these phrases had either person names or were specific to Enron (e.g., "operating officer enron", "enron industrial"). Table 2 shows the 26 most predictive gossip phrases.

Can these phrases be grouped under some categories? We used the LIWC program (Pennebaker, Francis, and Booth 2001) to explore this question. After reviewing the phrases in the gossip text and the ones in the LIWC categories, we selected ten LIWC categories: *Social, Present, Inclusion, Cognitive Processes, References to others, Self references, Time, Space, Occupation and Affect*. Since LIWC does simultaneous tests of all these ten categories, we needed a Bonferroni correction, reducing α to 0.05/10 = 0.005. However, all these categories give random results and we find that these phrases cannot be categorized by the LIWC program.

Structure of the predictive phrases. To show how these commonly occurring phrases are used in the message body of gossip email, we present word tree visualizations (Wattenberg and Viégas 2008) using the online site Many Eyes (Viegas et al. 2007). The purpose of doing is to gain insight into the usage of organizational gossip. Figure 10 shows these visualizations of a few phrases. We see that an employee gossips when things are not very pleasant (*am having trouble, am having difficulty, declared bankruptcy*) or when he wants to pass on some new updates or information (*in case you, for your note, commissioner said*), or when he has to respond to something he missed earlier (*in response to, couple of weeks ago*), or when he wants to acknowledge some good work (*getting up to speed, innovative projects*). We have not shown all of these word tree structures in the paper, due to space

Gossip phrases	β	Gossip phrases	β
office of the	0.749	light of	0.401
out of service	0.706	referring	0.398
business controller	0.667	down time	0.386
bunch of	0.639	up to speed	0.386
wholesale power	0.612	in response	0.371
resigned her	0.605	innovative	0.360
board of directors	0.579	couple of weeks	0.358
in the early	0.567	born	0.332
am having	0.487	declared	0.327
has agreed to	0.475	for your note	0.323
commissioner	0.472	the px	0.32
in case you	0.414	i am glad	0.319
not hesitate	0.406	discovered	0.317

Table 2: The top 26 most commonly occurring phrases in gossip email along with their β coefficients.

constraints. The word *born* presents an interesting case study. *born* occurred 19 times in the corpus, with *was born* being the most frequent. People mostly used this word to share personal information about themselves and their families. These results reflect the social functions which gossip plays in an organization. It serves as both entertainment and as information exchange. The role it plays in important exchange might explain why we see a significant neutral tone in these messages. This finding is in line with previous research: Roy (1959) shows that gossip can break the monotony of work.

Discussion

Our study reveals some important characteristics of organizational gossip. First, gossip is present in both personal and business email and across all sections of the hierarchy, which demonstrates its all-pervasive nature in organizations. Next, we showed that the hierarchical position of an employee affects his gossip behavior, both in terms of his frequency of gossip and the audience with whom he gossips. Our results indicate that people are most likely to gossip with their peers. These findings are in line with Gluckman (1963)'s theory: gossip maintains a group's unity and establishes its boundary. Paine (1967) on the other hand, takes an individual perspective towards gossip. Perhaps both theories are true to an extent.

Figure 6 shows individual gossip behavior within a group. The restricted spread of gossip percentages for In-House lawyers, might be indicative of a more coherent group, compared to the diverse spread found in the other groups. Building off this finding, perhaps using gossip as an indicator, an organization could build self-aware applications to spot peer groups and groups which have diverse interests.

We also show that organizational gossip is a social process. Some people are actively involved in generating gossip messages ("gossip source"), while others are silent readers of the messages ("gossip sink"'), and there are some who play both roles. Acting as a conduit of information, identifying

[5] `http://bit.ly/zNmLNM`

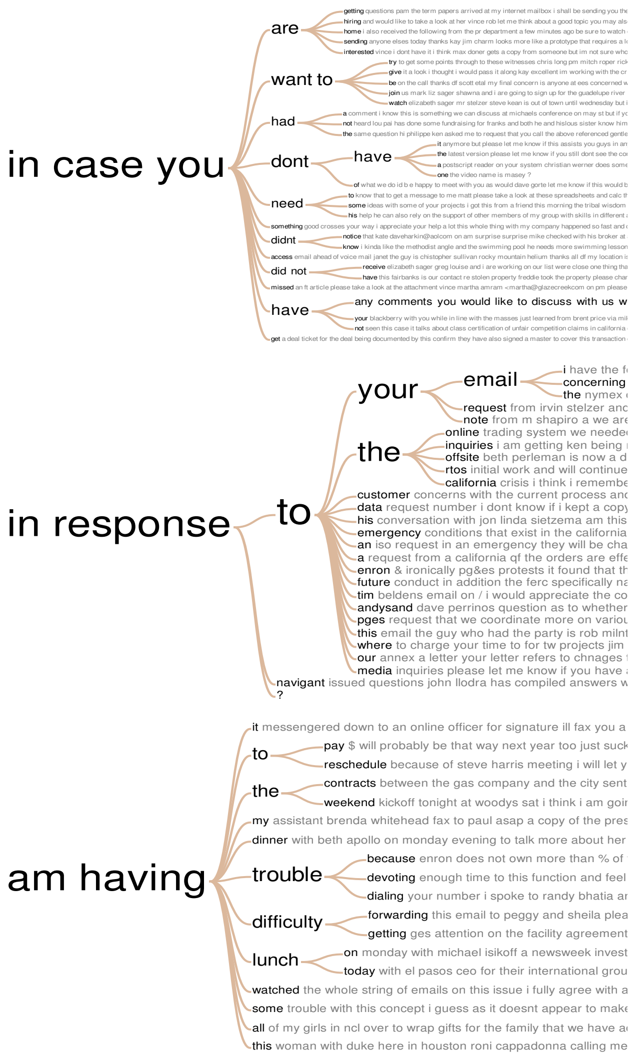

Figure 10: Word tree visualizations showing how some phrases are used in gossip email. A search for *in case you* gives 33 hits, while *in response to* occurs 25 times in the entire gossip text corpus. *am having* has 18 occurrences with *am having trouble* being the most likely trigram. The font size is proportional to the frequency of occurrence of the word/phrase in the corpus.

gossip sources and sinks may help an organization locate its information hot spots.

Table 1 reveals a surprising finding: frequent dyadic email interactions do not show an increase in gossip email. This fact raises more questions than answers. It might be the case that social contact between two people in an organization is not well captured by email exchanges; there may be other channels of communication. It is more natural for two colleagues who frequently meet in the office cafeteria to gossip while in the cafe. Also, different dynamics come into action if two people knew each other prior to email interactions or if they interact outside work. These conditions qualify a person to be a close contact, a strong tie. In other words, one's gossip behavior could be different for his strong and weak ties (Granovetter 1973), and hence could be used as an indicator of tie strength between the communicating parties.

The results of our work are heavily dependent on the way we have defined gossip. The only criterion we have used to qualify an email as gossip, is the absence of third parties from the conversation. On the other hand, Gluckman (1963) says that people even gossip in the presence of the subject. Adhering to Gluckman's definition of gossip might change our current results. Eggins (1997) says "in gossip the events are not experientially unusual but interpersonally unacceptable." Going by her negative connotation of gossip texts, we might end up with different results. However, her statement may explain another unexpected result: an increase in the gossip percentages in the year 2000 and 2001 (not shown graphically in this paper, due to space constraints). These years witnessed several difficult-to-accept events: an energy crisis, a world economic recession and Enron's own downfall.

We have studied gossip behavior in organizational email only. It would be interesting to perform the study in other communication media such as instant messaging or Facebook. While instant messaging is a purely dyadic private communication channel, where the third party has no knowledge about the interaction, gossip on a Facebook wall has every chance to be noticed by the third party. It would be interesting to see if and when gossip percolates from one social circle to another and what triggers this process. What would happen if, in group-level gossip, multiple people present conflicting facts? How would the listeners react to such conflicting information? Will it be detrimental to the unity of the group? Are gossipers eager to confirm the information from multiple social interconnections? Does the reputation of the gossiper (gossips too much or too little; gossips about positive things) determine his trustworthiness? Does the type of information (entertaining, concerning) determine their willingness to confirm? The process of confirmation might in turn cause the information to flow to other social connections. It would be interesting to see if and when the information gets garbled while cascading to different levels. More work needs to be done to explore these deep questions.

Conclusion

We believe our work addresses one of the most pervasive social activities in an organization and highlights its significance. Although our results are tied to the Enron corpus, we believe these are important for the following reasons. First, we provide a detailed study, with empirical evidence, of an under-researched yet important societal phenomenon. Second, we provide empirical insights by testing gossip theories originating from anthropology on a real world large email dataset. Third, our study also raises new questions about organizational gossip and email behavior. We hope our work motivates researchers to address these questions on different datasets and in different contexts.

Acknowledgements

We would like to thank Clayton J. Hutto, Abraham Doris-Down, Beki Grinter and our entire comp.social group at Georgia Tech for their valuable feedback on early versions of this work.

References

Besnier, N. 1989. Information withholding as a manipulative and collusive strategy in nukulaelae. *Language in Society* 18:315–341.

Dunbar, R. 1994. *Grooming, Gossip and the Evolution of Language.* London, Faber and Faber.

Eggins, S.and Slade, D. 1997. *Analyzing casual conversation.* London, Cassell.

Feinberg, M.; Willer, R.; Stellar, J.; and Keltner, D. 2012. The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*.

Finkel, J. R.; Grenager, T.; and Manning, C. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 363–370. Ann Arbor, Michigan: Association for Computational Linguistics.

Foster, E. K. 2004. Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology* 8(2):78–99.

Gilbert, E., and Karahalios, K. 2009. Predicting tie strength with social media. In *Proc. CHI*, 211–220.

Gilbert, E. 2012. Phrases that signal workplace hierarchy. In *CSCW*.

Gluckman, M. 1963. Papers in honor of melville j. herskovits: Gossip and scandal. *Current Anthropology* 4(3):307–316.

Goffman, E. 1959. *The Presentation of Self in Everyday Life*. New York, Doubleday.

Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology* 78(6):1360–1380.

Group, N. W. 2001. The internet society. rfc 2822 - internet message format.

Hannerz, U. 1967. Gossip, networks and culture in a black american ghetto*. *Ethnos* 32(1–4):35–60.

Jabbari, S.; Allison, B.; Guthrie, D.; and Guthrie, L. 2006. Towards the orwellian nightmare: separation of business and personal emails. In *Proceedings of the COLING/ACL on Main conference poster sessions*, 407–411. Morristown, NJ, USA: Association for Computational Linguistics.

Henderson Jr., D., and Myer, T. 1977. Issues in message technology. In *Proceedings of the fifth symposium on Data communications*, 6–9.

Klimt, B., and Yang, Y. 2004. Introducing the Enron corpus. In *First Conference on Email and Anti-Spam (CEAS)*.

Kurland, N. B., and Pelled, L. H. 2000. Passing the word: Toward a model of gossip and power in the workplace. *The Academy of Management Review* 25(2):pp. 428–438.

Madden, M., and Jones., S. 2008. Networked workers. Technical report, Pew Internet and American Life Project.

Paine, R. 1967. What is gossip about? an alternative hypothesis. *Man* 2(2):278–285.

Palus, S.; Bródka, P.; and Kazienko, P. 2010. How to analyze company using social network? In *WSKS (1)*, 159–164.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. Linguistic inquiry and word count: Liwc 2001. *Word Journal Of The International Linguistic Association*.

Peterson, K.; Hohensee, M.; and Xia, F. 2011. Email formality in the workplace: a case study on the enron corpus. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, 86–95. Stroudsburg, PA, USA: Association for Computational Linguistics.

Purcell, K. 2011. Search and email still top the list of most popular online activities. Technical report, Pew Internet and American Life Project.

Rosnow, R. L. 1977. Gossip and marketplace psychology. *Journal of Communication* 27(1):158–163.

Roy, D. F. 1959. Banana time: Job satisfaction and informal interaction. *Human Organization* 18(04):158–168.

Shetty, J., and Adibi, J. 2004. The enron email dataset database schema and brief statistical report.

Viegas, F. B.; Wattenberg, M.; van Ham, F.; Kriss, J.; and McKeon, M. 2007. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics* 13:1121–1128.

Wattenberg, M., and Viégas, F. 2008. The Word Tree, an Interactive Visual Concordance. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*.