

Air Force Institute of Technology

AFIT Scholar

Theses and Dissertations

Student Graduate Works

3-2006

Detecting Potential Insider Threats through Email Datamining

James S. Okolica

Follow this and additional works at: <https://scholar.afit.edu/etd>



Part of the [Information Security Commons](#)

Recommended Citation

Okolica, James S., "Detecting Potential Insider Threats through Email Datamining" (2006). *Theses and Dissertations*. 3464.

<https://scholar.afit.edu/etd/3464>

This Thesis is brought to you for free and open access by the Student Graduate Works at AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.



DETECTING POTENTIAL INSIDER THREATS
THROUGH EMAIL DATAMINING

THESIS

James Okolica, First Lieutenant, USAF

AFIT/GCS/ENG/06-10

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this Thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

AFIT/GCS/ENG/06-10

DETECTING POTENTIAL INSIDER THREATS THROUGH EMAIL DATAMINING

THESIS

Presented to the Faculty
Department of Electrical and Computer Engineering
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
In Partial Fulfillment of the Requirements for the
Degree of Master of Science (Computer Science)

James Okolica, B.A.
First Lieutenant, USAF

March 2006

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

DETECTING POTENTIAL INSIDER THREATS THROUGH EMAIL DATAMINING

James Okolica, B.A.
First Lieutenant, USAF

Approved:

/signed/	March 2006
_____	_____
Dr. Gilbert Peterson (Chairman)	date
/signed/	March 2006
_____	_____
Dr. Robert Mills (Member)	date
/signed/	March 2006
_____	_____
Lt Col. Summer E. Bartczak PhD (Member)	date

Abstract

Despite a technology bias that focuses on external electronic threats, insiders pose the greatest threat to commercial and government organizations. One means of preventing insider theft is by stopping potential insiders from actually crossing the line. In the overwhelming number of cases, people do not join an organization with the intention of stealing or causing harm. Instead, something or often several things happen while the individual is in the organization that precedes his malevolent actions. One of the traits identified with insiders is their feeling of alienation from the organization. By datamining emails, an employee's interests can be discerned. These interests are then used to construct social networks which are used to identify individuals with interests shared but undiscussed with other members of the organization. These individuals with clandestine interests have the potential to be insider threats. This paper describes the use of Probabilistic Latent Semantic Indexing (PLSI) [?] extended to include users (PLSI-U) and Author Topic [?] extended to include documents to determine topics of interest for employees from their email activity. It then applies PLSI-U and Author Topic to the Enron email corpus. The results show that by comparing the topics of emails that people send internally with the ones sent externally, a small number of employees (0.03% - 1.0%) emerge as having clandestine interests and the potential to become insider threats. Most significantly, one of these individuals is Sherron Watkins, the famous whistleblower in the Enron case.

Acknowledgements

First I would like to thank my dear wife for her support. While I was often in the same physical room with her, I was rarely available. Her understanding and support made this effort possible.

Next, I would like to thank my first advisor, Major Joel Young, for sparking my interest in this topic and for providing my initial direction. I would also like to thank my second and principal advisor, Dr. Gilbert Peterson. When Major Young was called away, Dr. Peterson stepped in despite already having four students of his own. His encouragement and suggestions kept me going and ensured the timely completion of this effort. I would also like to thank my committee members, Dr. Robert Mills and Lieutenant Colonel Summer Bartczak.

Finally, I would like to thank my children. Writing a thesis is a long, lonely task. There are often long periods of wondering whether or not you are making any progress or doing anything right. Being able to spend time with my children, seeing their smiling faces and hearing their infectious laughter are the reasons that this thesis is a success.

James Okolica

Table of Contents

Page

List of Figures

Figure

Page

DETECTING POTENTIAL INSIDER THREATS THROUGH EMAIL DATAMINING

I. Introduction

1.1 *The Insider Threat*

“Espionage is the practice of spying or using spies to obtain information about the plans and activities of a foreign government or a competing company” [?]. While it is possible to insert professional spies into an organization, today, the use of insiders is much more prevalent. Insiders are members of an organization or government who often have a legitimate right to the information that they are accessing. However, they abrogate the trust they have been given by using the information for illegitimate reasons. Since most Cold War espionage was perpetrated by the Soviet Union [?], it was expected that after the end of the Cold War, the spy threat would decrease. This has not been the case. In addition to the increased number of security threats from terrorist organizations, there has been an alarming increase in the number of cases of economic espionage. There is a significant amount of evidence [?] that some nations are funding insider espionage by some of the corporations based in their countries.

When considering who becomes an insider, the different motives of these economic and security insider threats make it important to consider them separately. Insiders who become economic threats are more often motivated solely by financial considerations while those who are security threats may tend to have more divided loyalties. Take for instance Ana Belen Montes, a senior intelligence analyst at the Defense Intelligence Agency, who was arrested on September 21, 2001 for spying for Cuba. In her statement at her sentencing she stated, “I obeyed my conscience rather than the law. I believe our government’s policy towards Cuba is cruel and unfair, profoundly un-neighborly, and I felt morally obligated to help the island defend itself.” She never asked for money and acted as she did only out of ideology [?].

Shaw, et al. [?] describe eight different types of insiders (explorers, good Samaritans, hackers, Machiavellians, exceptions, avengers, career thieves, and moles). The two least common (career thieves and moles) are the only ones who enter a corporation with the intention of stealing from it. In all other cases, improving the pre-screening process is ineffective. When these individuals join an organization they have no intention of spying or stealing from it. What is needed is some way to prevent these individuals from becoming insider threats. The first two cases (explorers and good Samaritans) are innocent individuals who simply stray to areas of computers systems where they are not supposed to. Having adequate safeguards can effectively prevent these two types of insiders. In the other four cases (hackers, Machiavellians, exceptions, and avengers), something happens which turns them. In many cases, this event is some corporate change which results in them being disgruntled. Such events include restructuring, streamlining, being passed over for promotion or simply a bad review. In other cases, this event may have nothing to do with the organization. It may be a personal crisis such as the end of a relationship, the ill health or death of a spouse or child, or a severe financial problem. Faced with these events, individuals often withdraw from the organization and may seek relief in alcohol or drugs. As they withdraw they feel alienated from their organization making it easier to overcome any inhibitions they have about betrayal.

1.2 Preventing the Insider Threat

Since in most cases detection of these warning signs would allow for early intervention and the prevention of any harm such as sabotage and theft, developing methods of detection is vital. A recent report by the U.S. Department of Defense [?] observed that “Nothing can replace first rate management of subordinates, genuine concern for their well being, fairness, and recognition of personal warning signs for mitigating the insider threat”. Unfortunately, while this sentiment is valid, it is stated at the same time that the DoD and every other public and private company is attempting to become more efficient by cutting mid-level management and converting

most of the management that remains into working-managers. At the same time there is a huge increase in both internal turnover and external contractors working in sensitive positions. It is not possible for managers to effectively get to know all of the individuals under their direct supervision to the point that behavioral changes will be noticed. Instead, they must pick a few individuals and focus on getting to know them. What is needed is an effective way for them to pick which individuals to focus on. The DoD report [?] acknowledges this and goes on to recommend the “maximum use of datamining to enable continual online review of personnel security information.”

In today’s Information Age, one of the best sources of personal information available at work is an individuals’ email and internet activity. By datamining an organization’s email, it is possible to learn a lot about not only the organization, but of the individuals within it as well. Datamining can find potential insiders by finding individuals who feel alienated from the organization and/or who have interests contrary to the organization’s well-being. While the goal of this datamining is to detect individuals who wish harm to the organization, the privacy concerns of innocent individuals must be considered. It is possible once personal information is gathered for this information to be used in ways that will cause innocent individuals harm. While this possibility exists, one must recall that most organizations explicitly inform employees that email and internet use is to be restricted to only business purposes. If this restriction is followed, there is little, if any, personal information that can later be used to harm individuals, significantly decreasing any privacy concerns. Probabilistic clustering is a method for dividing information into groups of similar objects by assuming these groups come from a mixture of several populations [?]. The goal in probabilistic clustering is to find parameters for these probability distributions that best fits the data.

Two promising probabilistic clustering algorithms are Probabilistic Latent Semantic Indexing (PLSI) [?] and Author Topic [?]. Each model assumes that documents (or emails) are constructed one word at a time. Before picking each word,

a topic is selected and then the word is selected based on the topic. Author Topic also allows for a similar mechanism that involves the author of the email, allowing some words and topics to be more likely depending on the interests of the author. By applying these models to email (and expanding PLSI to include authors (PLSI-U)), it is possible to extract topics of interest for the organization and its members. In addition, probabilistic clustering can associate individuals employees with those topics that they have the greatest interest in. These topics of interest can then be used to develop social networks. There are several ways to use these social networks to find insiders. First, if there are specific topics of concern, individuals with an interest in these topics can be flagged. Second, if individuals share common interests with known insiders, they can be flagged. Finally, if individuals fit an “insider threat profile” based on their topics of interest, they can be flagged. Such a profile may include feelings of alienation or interests in alcohol, drugs, or financial solutions.

1.3 Using Enron as the Data Source

Ideally, probabilistic clustering and social networking could be tested against a real world dataset of both email traffic and internet activity. However, privacy concerns have prevented this. Most research involving email traffic uses the personal email of the researchers (and possibly a small set of other individuals). It is very difficult to get large organizations to release the private email of their employees for research. The one large corpus of real-world email traffic that is available is a subset of email from Enron. As part of their investigation into Enron, the Federal Energy Regulatory Commission (FERC) seized Enron’s email and made a portion of it publically available. While it only includes the email folders of 151 employees, it still contains over 250,000 email messages. Furthermore, due to the number of individuals the emails were sent to, the resulting corpus has sufficient data on over 34,000 Enron employees to make probabilistic clustering of these individuals possible.

While using the Enron email corpus has some limitations (most noticeably a bias to the 151 employees whose email these folders came from and a lack of any

internet activity data), it does have some benefits. Unlike most organizations, Enron has at least one known insider, Sherron Watkins. Prior to the public disclosure of Enron's questionable accounting, Ms. Watkins, a vice president in Enron's corporate accounting division, sent a letter to Ken Lay, Enron's chairman, detailing the dubious accounting practices and their likely impact on Enron's future. While she did not go outside the corporation, her activities were considered an insider threat by her boss, Andy Fastow, Enron's Chief Financial Officer, who upon finding out that Watkins had wrote the letter demanded that she be fired immediately and her laptop confiscated [?].

Although many would argue that Watkins' actions should characterize her as a whistleblower rather than an insider, the distinction is not clear. At it simplest, what distinguishes a whistleblower is that he (or she) is revealing information about an illegal, or at least unethical, practice and therefore is serving the public good. However, certainly from the organization's point of view, the individual can still be considered an insider threat. Furthermore, if the person's motives are selfish, for instance revenge for a real or imagined slight, the distinction becomes even fuzzier. In the final analysis, whistleblowers still function in the same manner as insiders. It is just that they are revealing information that arguably should be revealed and that they may be doing it for noble reasons.

1.4 The Experiment

The goal of this research is to test the hypothesis that "probabilistic clustering and social networking techniques applied to email are effective at detecting potential insider threats". Consider Figures ?? and ?. First, probabilistic clustering finds the topics for each email (in this case baking and significant others). Next, by looking at all of an individual's email, his topics of interest are determined (in this case, cooking, significant others, accounting, basketball, sports, and partying) and a social network is created. Third, his internal emails are considered to see what topics he

communicates about within the organization (in this case accounting, basketball, and sports) and a second social network is created.

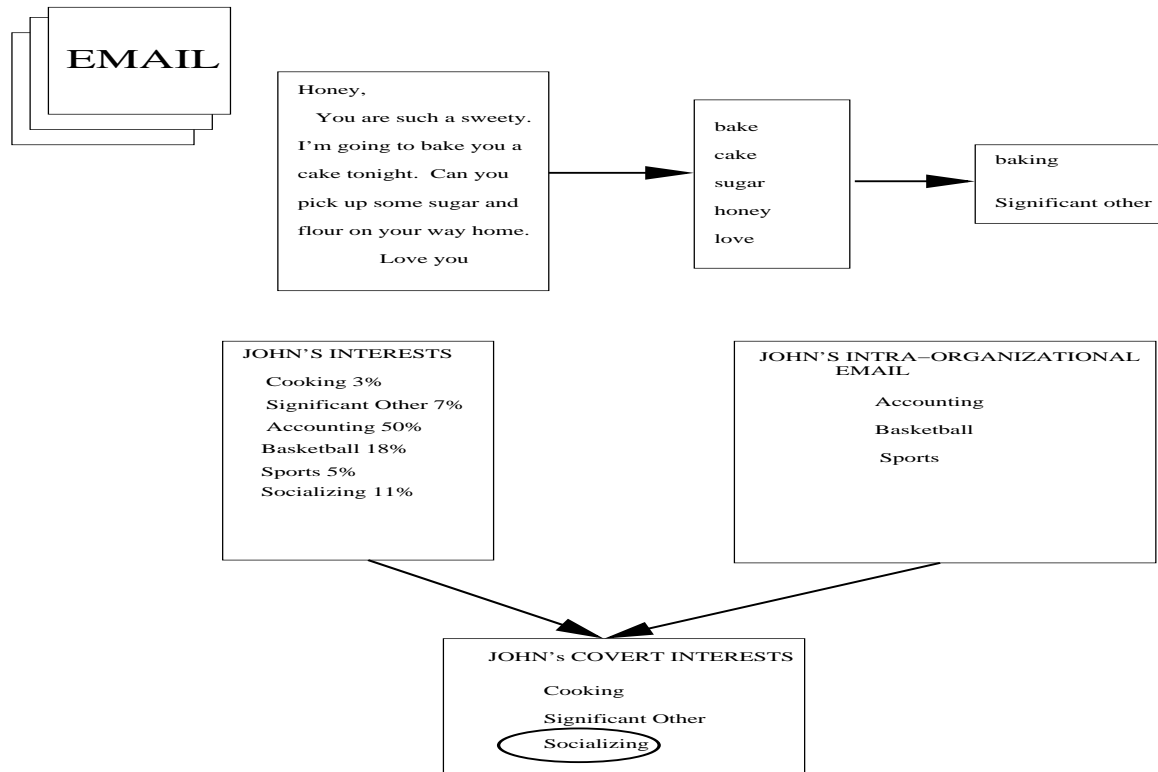


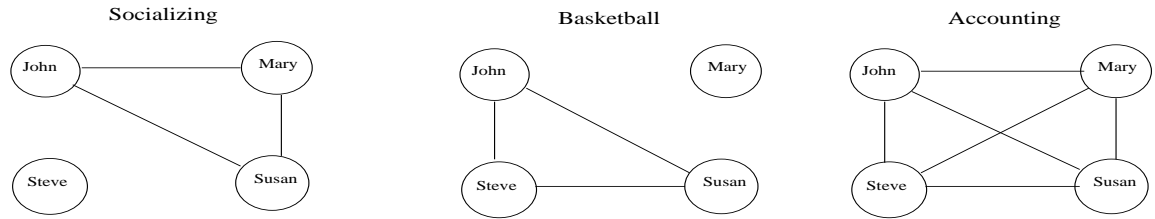
Figure 1.1: Example: An Email John received

The two are compared to determine what, if any, clandestine interests he has (in this case cooking, significant others, and partying). These are interests he has that he does not share with anyone within the organization. Finally, this list of clandestine interests are limited to only sensitive topics (in this case partying). An interest in partying that he doesn't share with people within the organization may indicate that he feels alienated, a warning sign of a potential insider threat. In other cases sensitive topics would be about things the organization does not want revealed. For instance, in the case of Enron, the sensitive topic is about the dubious accounting practices.

1.5 Analysis

After the social networks are created, an analysis of the resulting data is performed. There are three metrics to determine if Author Topic and PLSI-U are appro-

IMPLICIT INTEREST NETWORKS



EXPLICIT INTEREST NETWORKS

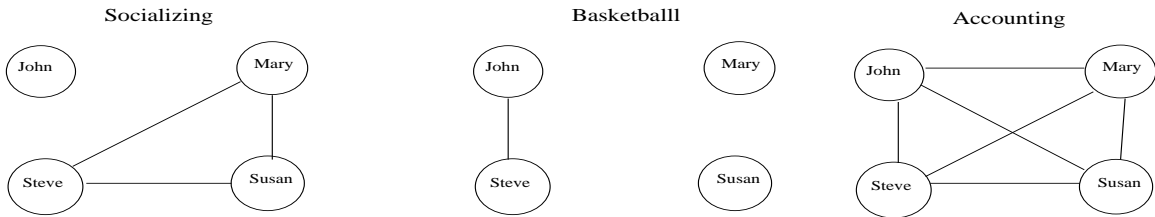


Figure 1.2: Example: Networks that include John

priate for extracting insider threats. They must be timely, useable and valid. Since some topics may only emerge on a sporadic basis, this technique should be performed on at least one to three months of email. If as a result it is only run every one to three months, results that are produced in a seven to ten days can be considered timely. Next, to be useable it must be possible for an administrator to examine the topics produced and quickly be able to identify the underlying topic by examining the most probable words. To measure this, the 25 most probable words and individuals for each topic are checked by the researcher to see if they produce easily identifiable topics.

Finally, the validity of the probabilistic clustering for insider threat detection must be established. For probabilistic clustering to be considered a success, Sherron Watkins must be identified as one of the individuals with a clandestine interest in the topic most strongly related to Enron's questionable accounting practices. This would indicate that she had the potential to become an insider threat.

In addition to analyzing the effectiveness of PLSI-U and Author Topic, the resulting social network data also allows for the calculation of the most central, or important, individuals. Traditional social network analysis (SNA) has metrics to

determine the centrality of individuals and the cohesiveness of the overall networks [?]. By making such calculations, the most probable individuals from Author Topic and PLSI-U can be compared to the most central individuals from SNA. Finding the same individuals appearing in both cases reinforce the fact that Author Topic and PLSI-U are clustering individuals with the appropriate topics.

1.6 Conclusion

Chapter 2 begins with a more detailed discussion identifying threats through likely insider personality traits and motivations. It then reviews the basics of probabilistic clustering as well as an in depth review of Author Topic and Probabilistic Latent Semantic Indexing. The discussion then covers social networks, a technique used to effectively expose individuals alienated from their organization. Chapter 3 reviews the methodology detailing the research questions and the evaluation metrics followed by a description of the data including its preparation, clustering and analysis. It then concludes with two supplementary experiments performed in the hopes of developing additional information. Chapter 4 then discusses the results of the experiments. Both PLSI-U and Author Topic create good clusters of topics appropriate to Enron and associate appropriate individuals with these topics. However, only Author Topic is successful at extracting Sherron Watkins as a potential insider. Chapter 5 concludes with a summary of results as well as some possible next steps.

II. Background

There are estimates that as many as 87% of all electronic thefts come from individuals with legitimate access to the organization [?]. However, in the vast majority of cases, these individuals did not join the organization with the intention to steal [?]. Something changed along the way and they “converted”. This is not a new phenomenon. It has existed since ancient times [?] and, despite the introduction of new high-tech methods for spying, insiders continue to have the ability to steal information not easily accessible by any other means.

Unfortunately, despite the prevalence of insider theft and its cost, there has been only a small amount of analysis of it. Most of that analysis has focused on the character traits of insiders. These insiders generally feel alienated from the organization, possess an ethical flexibility combined with a lack of empathy, and have a strong sense of entitlement [?]. When this lack of social inhibition is combined with an opportunity and a motive, the conditions are ripe for a theft to occur. However, one last ingredient is generally required [?]. There is usually a trigger that sets these individuals off. Triggers include relationship crises, financial crises and even health crises. At such time, behaviors change and individuals begin to contemplate insider theft.

What is required is for managers to notice these times of crisis. Unfortunately, as companies continue to downsize, it is reasonable to expect that fewer managers will be managing more individuals. As a result, there may not be enough time for managers to sufficiently get to know and continually observe their people. Instead, what is needed is an automated way of accomplishing this [?].

By examining the email of an organization, it is possible to perform datamining and find individuals who have the potential to become insider threats. Individuals who feel alienated or who have clandestine interests in sensitive topics can be found by extracting their topics of interest. One means of finding individuals’ topics of interest is probabilistic clustering. Probabilistic clustering divides data into similar groups by assuming that each datum (an email this is a word) comes from a mixture

of several populations whose probability distribution parameters are unknown. The goal of probabilistic clustering is to find values for these parameters that make the probability distributions fit the data. Two popular probabilistic clustering techniques are Probabilistic Latent Semantic Indexing (PLSI) and Latent Dirichlet Allocation (LDA). While PLSI has not been directly applied to the construction of documents involving multiple individuals, LDA has been applied in the form of the Author Topic (AT) model.

Once these techniques have clustered the data into discrete topics, individuals' interests in specific topics are used to construct social networks. These social networks are then used to match people who have some of the tendencies of past insiders. Also, if an insider is known, these networks are used to find other individuals with similar interests who may also be potential insiders.

It is difficult finding data for this type of research. Privacy concerns render most real-world datasets unavailable and artificial datasets are generally too small. One recent dataset that is fast becoming the touchstone of current information retrieval (IR) research is the Enron email corpus. When the Federal Energy Regulatory Commission (FERC) began investigating Enron, it seized a significant amount of email. This email was later released to the public in the form of electronic files. This initial email corpus has since been sanitized, removing private information, and cleaned of data integrity issues. The result contains approximately 250,000 messages involving over 87,000 individuals making it a wonderful dataset for performing email datamining research.

2.1 Defining Insiders

Sun Tzu observed that what enables the wise sovereign and good general to strike and conquer is foreknowledge and this foreknowledge is only obtained from others [?], specifically spies. There are five types of spies: spies inserted into the enemy's local populace, spies inserted into the enemy's government, spies who leak false information to the enemy, spies who bring back news from the enemy, and

converted spies. Converted spies are double agents, spies for the enemy who have been converted to spy on their former employers. Sun Tzu felt that of all of the types of spies, converted spies were the most important because they allowed the discovery and use of the other four types.

Since Sun Tzu's time, more methods have been developed for acquiring information from an enemy or competitor. These methods include signals intelligence (SIGNIT), imagery intelligence (IMINT), Measurement and Signature Intelligence (MASINT), Human-Source Intelligence (HUMINT), Open Source Intelligence (OSINT) and Geospatial Intelligence [?]. While technical means have become more popular recently, there are still many types of information that only HUMINT can obtain including detailed descriptions of underground facilities or facilities under a jungle canopy. HUMINT is also the only source currently available to provide qualitative descriptions of people. Despite the glamour of spies like James Bond, the reality is that most human information obtained comes from insiders. In 2000, the Department of Defense (DoD) produced a report showing that 87% of identified intruders were either employees or other internals to the organizations [?]. Furthermore, in the last fifty years, most insiders in the United States (79%) have sought out foreign agencies to provide information to [?]. Traditionally, it has been more difficult for a foreign service to seek out one of the few cleared Americans willing to betray their country and instead, foreign agencies sit back and wait for Americans to come to them.

To summarize, an insider is someone who “steals, spies, or betrays their loyalty to their country or employer” [?]. Observe that, in addition to theft, this definition also includes an individual who sabotages, extorts or causes some other harm to the organization. In addition to being called insiders, these individuals are also called insider threats and malicious insiders.

2.1.1 Becoming an Insider. In the overwhelming number of cases, individuals do not start in an organization with the intent to do harm. Instead, something happens over time that results in an individual becoming an insider threat. There

are four preconditions that are generally met before a person decides to become an insider threat [?]. First, the individual must have a motive. Second, he must have an opportunity to commit the crime. Third, he must overcome natural inhibitions, e.g. moral values. Finally, a trigger is needed to set the crime in motion. Herbig and Wiskoff [?] in their study of espionage cases found that the number one motivating factor (69%) in their study was money. However, among individuals who sought out foreign governments to sell information to, disgruntlement with the workplace (26%) was also a significant motive. One factor that has significantly increased since the end of the Cold War is divided loyalties with 50% of the spies caught since 1990 citing divided loyalty (34% over all). While money was the number one factor, there are several contributing causes to money. In half of the cases, indebtedness was the reason that money was the top factor. One of the more intuitive reasons, coercion, appears to not have been a significant factor, present in only 5 of the 150 cases.

After having a motive to commit the crime, the next thing necessary is the opportunity. There are two elements to opportunity. The first is the opportunity to steal the information and the second is the opportunity to sell it [?]. As mentioned earlier, in most cases, potential spies had to manufacture the opportunity to sell their information and in many cases were caught as a result. Unlike finding a buyer, which many people put little forethought into, most people do plan how to steal information. While it is difficult for organizations to prevent people from finding buyers, it is within organizations' power to make obtaining information to sell more difficult. As a result, there have been many studies (especially since the dramatic increase in economic espionage and electronic theft) into what factors are most important in protecting information. Shaw, et al. [?] found that vulnerabilities in poor management practices, poor employee screening, incomplete termination procedures, missing warning signs, and not monitoring online communications contributed most to thefts. In contrast, a DoD study [?] found, in most cases, laziness and ignorance of security policies were the causes. ASIS International, largest professional organization for security managers, found theft reductions dependent on limiting information access, making information

and physical security a priority, and ensuring management concern and focus on information loss [?]. Similarly, CSO (Chief Security Officer) Magazine found the best way to deter thefts was to train new employees and have management regularly communicate that security was a priority [?].

Even though many people have motives to cause harm or theft and the opportunity to do it, most do not. In addition to motives and opportunities, in order to steal people must overcome their natural inhibitions to commit an immoral act against their coworkers and country. One motivation that counteracts this is divided loyalties. If a person feels that whichever action they take, they are letting down one of their countries or that they must balance country against family, inhibitions are much decreased. A second factor in lowering inhibitions is substance abuse. In the Herbig and Wiskoff database of 150 spies [?], alcohol abuse was present in 51% of cases for which data was available and drug abuse was present in 53% of cases for which data was available. What these two factors also have in common is that they help individuals feel alienated from their organization. It is much easier to not feel inhibited from an organization if one no longer feels a part of it.

Finally, although substance abuse and alienation can reduce inhibitions, they still exist and still deter most individuals. However, when motivation and opportunity and alienation are combined with a life crisis, people stop thinking. Herbig and Wiskoff [?] report that one-fourth of the spies experienced a life crisis in the months preceding their decision to become a spy (e.g. divorce, death of a loved one, love affair gone bad). Dr. Mike Gelles of the Naval Criminal Investigative service (NCIS) also argues that even narcissists and antisocial individuals require aggravation by a personal, financial, or career crisis which friends, coworkers and supervisors fail to recognize as a serious problem [?]. As a complement to these two traits, Project Slammer (another study of Americans convicted of espionage against the United States) found that there is also a different type of individual who commits espionage, one who is passive, easily influenced, and is lacking in self-esteem. Although most spies fall into the first category, there are several that fall into the second [?].

2.1.2 Insider Traits. Disgruntlement is becoming a much more common reason that individuals become an insider threat. However, there are very few individuals who have never gotten angry at their supervisors or thought their organization, or government, was not making bad choices. If this was all it took to become an insider threat, the bad guys would outnumber the good guys. Therefore, it make sense to look at some of the character traits that increase the chances of someone who is disgruntled becoming an insider threat. Shaw, et al. [?] describe six personality factors that commonly appear in insiders: a history of personal and social frustration, computer dependency, ethical flexibility, reduced loyalty, a sense of entitlement and a lack of empathy.

Dr. Mike Gelles condenses these characteristics down to two: narcissism and antisocial personality disorder [?]. Antisocial people reject the normal rules and standards of society and have no remorse over their actions. They do not form attachments to either people or causes. They are therefore often manipulative, self-serving and place high emphasis on immediate gratification. Narcissists also have strong feeling of entitlement and a lack of empathy for others. When narcissists are criticized (i.e. their high self-image is threatened), they may react viciously, seeking revenge out of proportion with the criticism. They may also seek out other groups that will restore their unrealistic self-image. The basic difference between narcissists and the antisocial personality is that the antisocial personality rejects the rules while the narcissist believes the rules apply to everyone else but not to him or her.

2.1.3 Types of Insiders. Shaw, et al. [?] consider eight types of insider threats: explorers, good Samaritans, hackers, Machiavellians, exceptions, avengers, career thieves, and moles. Explorers are people who are just curious and are looking at places they perhaps shouldn't be. They mean no harm but have the potential to shut a business down by deleting a file that they shouldn't have. Unlike explorers, Good Samaritans have a purpose and wish to fix something that is broken although it is not in their area. Unfortunately, Good Samaritans like explorers often make the

same mistakes and create havoc by “fixing” something that wasn’t broken. Hackers are the worst of the innocents. They also have no wish to harm their company; they just want to prove they can get where they know they aren’t supposed to be, often to impress their friends and receive peer approval. The greatest problem with hackers is that they are often one-upped to the point where they end up doing things that are often destructive and at least annoying in order to prove that they are the best. Machiavellians are the first group to fall into the malicious category. Their motivation is the advancement of their careers and personal goals and see nothing wrong with acquiring unauthorized information in order to do so. Exceptions take this one step further by believing that they are exceptions to the rules and that the rules don’t apply to them. Often this belief has been reinforced by their supervisors who have previously given in to demands because they were too “important” to offend (e.g. offending them may result in the computer system not being available). One subclass of exceptions are the proprietors who believe that the computer system is “theirs ” and they are free to do whatever they think is best. Avengers also feel entitled, but unlike Machiavellians who are trying to advance themselves, avengers are trying to get revenge for some ill done to them. As such, their goal often is to cause damage to the organization. Finally, the last two categories, career thieves and moles, enter into an organization with the explicit goal of theft. While these categories are the least common, they are the most dangerous since every action taken from the beginning is with the long-term goal of stealing from the organization. It is at least a little comforting to note that in the overwhelming number of the cases reported in [?], the traitors did not fall into either of the last two categories, i.e. they did not enter government service with the intent to commit espionage.

2.1.4 Whistleblowers. Observe that whistleblowers are not mentioned above as one of the eight types of insider threats. A whistleblower is an individual who is a current or former member of an organization and who acts with the intention of making information public, either internally or externally, about possible or actual

non-trivial wrongdoing in an organization [?,?]. Whistleblowers can be distinguished from other insiders by the fact that they are revealing information about an activity that violates public laws or public trust. However, while there is no literature that specifies whether a whistleblower should be considered an insider threat, some authors consider whistleblowers to be insiders [?]. Recall that an insider is someone who betrays their loyalty to their organization. It can be argued that a whistleblower does just that. Therefore it is difficult to avoid seeing a similarity between the two.

One possible reason that many people distinguish between whistleblowers and insiders is because they equate the insider threat with the term malicious insider. Using this definition for insider threat (i.e. malicious insider), it is easier to determine if a whistleblowers should be considered an insider threat. In many cases, whistleblowers come forward for revenge. For instance, in one of the Australian Competition and Consumer Commission's (ACCC) cases, the whistleblower was an employee who discovered his wife was having an affair with his boss. He then revealed to ACC that his boss was entering into unlawful arrangements with competitors [?]. In such as case, it is easy to dismiss the whistleblower as an insider threat. However, in other cases, it is not so easy. Consider the seven Hanford pipefitters who were laid off for refusing to install a faulty valve in a system carrying high-level nuclear waste [?] or Russell Tice who was dismissed for raising suspicions that his colleague might be a spy for China [?]. These individuals risked a lot where there is no clear indication of personal gain. These cases point out that if a whistleblower's motives are pure, then he probably is faced with a conflict between his loyalty to his organization and his loyalty to the public at large. In this case, is easy to argue that he is not a malicious insider. However, it is still unclear whether he should be considered an insider threat, defined simply as one who betrays their loyalty to their organization. While many would argue that whistleblowers should be considered different, for this research, whistleblowers are considered insider threats.

2.1.5 Preventing Insider Threats. What is critical to observe is that of the eight types of insiders, only the two least common enter an organization with the intention of becoming insider threats. For the rest something changes along the way. What is needed then is to be vigilant, both to behavioral changes and to the initial tentative fumbblings of betrayal [?].

The Defense Personnel Security Research Center (PERSEREC) and Project Slammer both found that once convicted, many of these spies complained that if someone had stood in their way, asked them what was going on, shown an interest in them before they started or been paying attention after they started, the espionage would have stopped. One of the reasons they began to commit espionage is because they felt no one cared about them and *no one would notice them committing espionage anyway* [?]. In one case, a spy had taken classified documents into another room to photograph them. While he was photographing them, he was interrupted twice by people entering the room. They saw what he was doing but just excused themselves for barging in and left, apparently assuming he had a legitimate reason for his activities and that someone else knew about it. As the DoD Integrated Process Team on Insider Threat observes, “Nothing can replace first rate management of subordinates, genuine concern for their well being, fairness, and recognition of personal warning signs for mitigating the insider threat... Managers must live up to the expectation that they evaluate personnel effectiveness daily, develop the skills to recognize individuals who require special assistance and provide avenues for them to acquire that assistance.” [?]

One place that managers can look for guidance in determining what constitute warning signs Perserec. The “Adjudicative Guidelines for Determining Eligibility for Access to Classified Information” is broken down into thirteen guidelines addressing potential areas of concern. They are: alcohol consumption, allegiance to the United States, criminal conduct, drug involvement, emotional, mental and personality disorders, financial considerations, foreign influence, foreign preference, misuse of information technology systems, outside activities, personal conduct, security violations, and sexual behavior [?]. Unfortunately in most cases, people do not have issues

in these areas when they first go for a security clearance. It is only later that these issues surface making the initial vetting process moot and highlighting the need for an ongoing automated process that would bring to light any developing concerns in these areas. The DoD Insider Threat Report [?] spelled this out with their recommendation 2.7 which advises: “employ maximum use of datamining to enable continual online review of personnel security information.” One method of performing ongoing checks that requires relatively little effort is to do an annual financial review and credit check.

2.2 Detecting Potential Insiders by Datamining

During the first workshop following the Department of Defense’s report [?], the first priority for improving the detection of insider’s misuse was “developing [user] profiling as a technique” [?]. To develop these profiles, the workshop participants proposed using: files and processes normally accessed, periods of time that a user is logged in, and keystroke patterns. By comparing old profiles with current ones, anomalies (e.g. use of administrator or logging commands) are better detected [?]. While this is successful if there is historical data to compare to, the amount of history that is needed is overwhelming.

An alternative means to detect anomalies is to consider an individual’s interests. There are several ways to detect potential insiders based on their interests. First, if an individual has a set of interests that match a known “insider profile”, that person warrants additional attention. However, since no such “insider profiles” exist, other methods must be considered. Second, if an individual’s interests change radically, it may be indicative of a personal crisis that again suggests the need for more personal attention. A third method uses these interests to detect if an individual feels alienated from the organization. If a person has a set of interests but does not share some or all of them with anyone within the organization, this may indicate the person feels alienated. Furthermore, if these interests include sensitive subjects (e.g. terrorism, confidential information, financial issues), this strongly suggests to the researcher that more individual attention is needed.

The first step in developing these interest profiles is to separate the email activity into topics or clusters. Clustering is the division of data into groups of similar objects [?]. The goal of this clustering is the discovery of hidden patterns. Once the data has been clustered, datamining is used to extract desired information. For instance, by clustering people who die of heart attacks, one can identify groups such as smokers and people in high stress jobs. There are several categories of clustering techniques. The first is hierarchical clustering which attempts to build a tree from the data. This is a very general technique and works well for a wide variety of data types and granularity. A second category of clustering is partitioning relocation. While hierarchical clustering sees all of the data as related at some level, partitioning relocation does not make that assumption. It relocates the data into distinct clusters based on some criterion. Unlike hierarchical clustering, partitioning relocation looks at the data multiple times possibly relocating each datum multiple times as it explores more of the data set. Probabilistic clustering is a form of partitioning relocation that assumes that the data comes from a mixture of several populations whose probability distribution parameters and priors it seeks to find. Density-based partitioning is a simple method of clustering that requires the data points to exist in a Euclidean space and then uses distance and density to determine the clusters. While both density-based partitioning and hierarchical clustering are simple to implement and work well, they both suffer from the curse of dimensionality [?] which says that as the number of attributes increase (20 is often considered a reasonable cutoff), the sparsity of data results in the algorithms becoming unstable. While probabilistic clustering also suffers to some extent from the curse of dimensionality, the effect is much less.

2.3 Clustering Email

The first step in determining how to cluster email is to decide how to represent it. There are several different possible information models to consider.

2.3.1 The Vector Model.

Assumption 1: Email can be treated as a “bag-of-words”, i.e. reducing a document down to the words that make it up, without regard to the ordering of those words, still keeps the essential characteristics of the email.

At first glance this assumption appears completely unrealistic. Obviously the order of the words matter. For example the sentences. “the police massacred the protestors” is very different from “the protestors massacred the police”. However, empirically this overly restrictive assumption still performs very well [?].

Definition 1: The collection of email is called a *corpus*.

Definition 2: A *document* (d) represents an email message. There are M documents in the corpus.

Definition 3: Each document is made up of *words*, ($w_1..w_N$). For each document, consider the number of words N_d selected from a Poisson distribution.

Definition 4: Since there are a finite number of documents and each document is composed of a finite number of words, there are a finite number of words, V , contained in the corpus. The collection of all words in the corpus is the *vocabulary* where w_i represents the i^{th} word in the vocabulary.

Definition 5: A document is represented as a vector, with V entries. Entry e_i is the number of times w_i occurs in the document.

With this representation, it is possible to define the similarity between two documents. Since each document is a vector, it is possible to calculate the angle between two vectors, i.e. documents. The closer the angle is to 90 degrees, the closer the cosine of the angle is to 0.0 and the less similar the documents are. Conversely, the closer the angle is to 0 degrees, the closer the cosine of the angle is to 1.0 and

the more similar documents are. Therefore, by calculating the cosine of the angle between two vectors, similarity is measured as a real number between 0.0 and 1.0.

While it is theoretically possible to construct a vector containing every word, considering that there are over hundreds of thousands of words (especially considering the multiplicity of languages), reducing the dimensionality is helpful.

2.3.2 Including Hidden Topics in the Vector Model.

Assumption 2: Emails are created around underlying topic(s).

Assumption 3: The probability of a specific word appearing in an email is conditional on the topic of that email.

Assumption 2 coupled with Assumption 3 ensures that each email has a specific topic and that it is possible to determine what that topic is based on which words appear. Using these assumptions, it is now possible to consider a collection of K topics, $(z_1..z_K)$.

Definition 6: A topic is a latent structure. Certain words occur in a document because the document contains a specific topic. Depending on the model, documents may contain one or more topics. Although there is no way to explicitly count the number of topics in the corpus, some number of topics, K , must be assumed to exist a priori.

Definition 7: A document is also represented as a vector with K entries. Entry e_i is 1 if topic z_i occurs in the document. It is 0 otherwise.

The similarity of documents is computed in an identical manner to the method used for vectors of words. However, while the number of words is at least in the hundreds of thousands, the number of topics is orders of magnitude smaller.

One of the weaknesses in all of the models considered in this paper is in deciding a priori what K is. While it is very simple (if time consuming) to look at every

document to determine the number of words in the vocabulary, there is no objective way to determine the number of topics. However, if too few topics are assumed, different topics are lumped together; where as, if too many topics are assumed, emails that should be of the same topic are split up.

2.3.3 Generative Models. Now that models of representing emails have been discussed, the next step is to consider how an email is constructed. These generative models are then supersets that include the vector models already discussed. To proceed on firm theoretical footing, it is necessary to consider the underlying means by which a new email is constructed. Once this method is known, it is possible to use its characteristics to develop a statistical model. This statistical model is then used to predict the likelihood that a specific email is constructed from a specific topic, and consequently is a member of a particular topic.

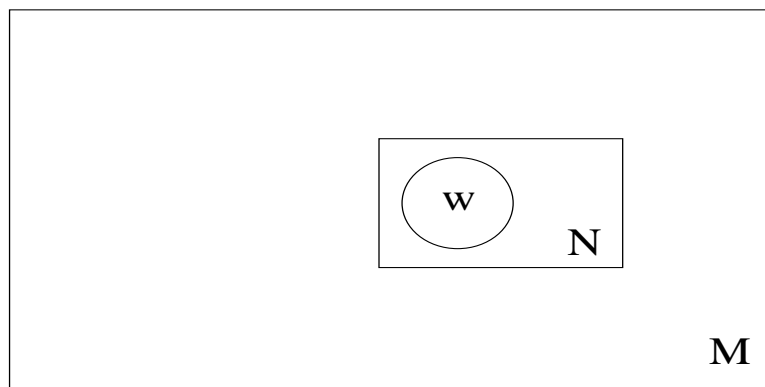


Figure 2.1: Unigram Model

2.3.3.1 Unigram Model. The simplest model to consider is the unigram model shown in Figure ???. This model posits that each word has an underlying probability of occurring in any email. When a new document is constructed, N words are chosen for the email based on an underlying multinomial probability distribution. Therefore, the probability of a particular email occurring is: $\prod_{i=1}^N p(w_i)$. This model assumes that there is no underlying topic and therefore it is equally likely that an email is composed of, for example, the words “baseball, puff pastry, and alcohol abuse”.

2.3.3.2 Mixture of Unigrams.

Assumption 4: Naive Bayes Assumption for words: the existence of a specific word in an email is conditionally independent of the presence of every other word in that email given the topic.

The remaining models all consider the notion of underlying topics, or categories. Notationally, consider z_i to denote a specific topic. The simplest model that includes topics (Figure ??) is one that proposes that an email is constructed by first selecting a topic from a multinomial distribution and then choosing words from a multinomial distribution based on that topic, i.e. choosing words conditioned on the choice of topic. For this model, the probability of a given document is $\sum_{i=1}^k p(z_i) \prod_{j=1}^N p(w_j|z_i)$. While this model is much more expressive than the unigram model, it does still restrict the selection process to a single topic.

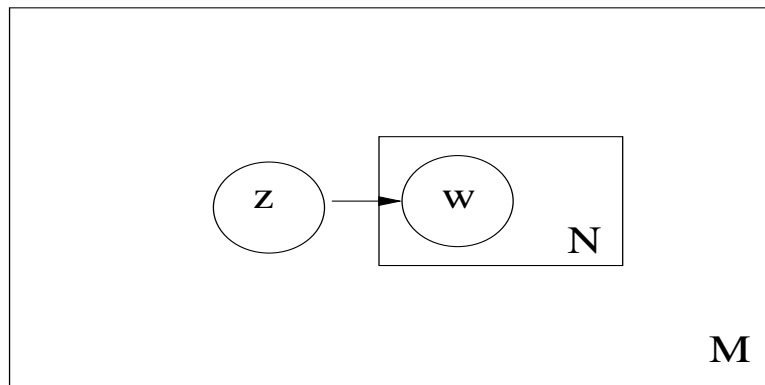


Figure 2.2: Mixture of Unigrams Model

2.3.3.3 Mixture of Words and Topics.

Assumption 5: Naive Bayes Assumption for topics: the existence of a specific topic in an email is independent on the presence of every other topic in that email.

To provide additional flexibility, the mixture of unigrams model is expanded so that prior to *each* word being added, a topic is selected from a multinomial distribution and then the word is selected conditionally given the topic from a multinomial distribution (Figure ??). The two generative models that fall into this category are

Hoffman’s Probabilistic Latent Semantic Indexing (PLSI) [?] and Blei, Ng, and Jordan’s Latent Dirichlet Allocation (LDA) [?].

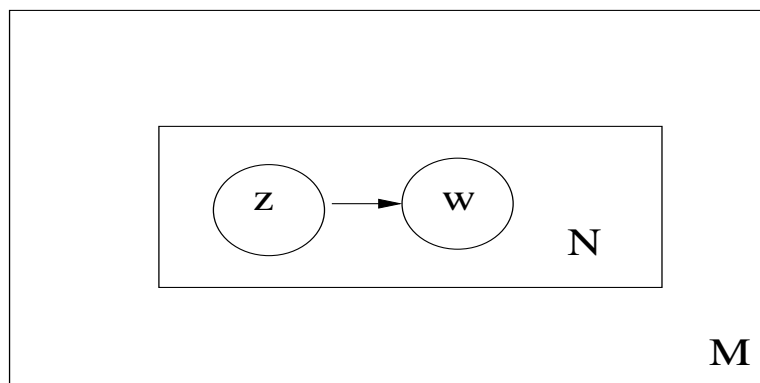


Figure 2.3: Mixture of Words and Topics Model

PLSI assumes the creation of documents just as described above. What is most desired is the joint probability of a word w_i occurring in document d_j which contains topic z_k . However, given the size of the vocabulary, the number of words in the documents and the number of topics, determining this full joint probability is unrealistic. However, it is sufficient to determine the probability of topic z_k for a specific document. Then by looking at the probabilities for all of the topics, one determines which topics the document contains (since they will have the greatest probabilities). Therefore, the goal is to determine $p(z|d)$. However, given the generative model, there is no direct relationship between topics and documents. A topic “produces” words and the collection of words creates the document. Therefore, in order to determine $p(z|d)$, it is first necessary to consider $p(z|d, w)$, i.e. the probability that topic z occurs given document d and word w . By Bayes’ rule,

$$p(z|d, w) = \frac{p(z)p(d|z)p(w|z)}{\sum_{z' \in Z} p(z')p(d|z')p(w|z')} \quad (2.1)$$

In order to come up with the most likely probabilities, PLSI uses Expectation Maximization. It attempts to determine the expected $p(z|d, w)$ by maximizing the values for $p(z)$, $p(w|z)$, and $p(d|z)$ based on the documents in the corpus (i.e. by using the

probabilities that seem most appropriate for the documents in the corpus). Formally,

$$P(w|z) = \frac{\sum_d n(d, w)P(z|d, w)}{\sum_{d, w'} n(d, w')P(z|d, w')} \quad (2.2)$$

$$P(d|z) = \frac{\sum_w n(d, w)P(z|d, w)}{\sum_{d', w} n(d', w)P(z|d', w)} \quad (2.3)$$

$$P(z) = \frac{\sum_{d, w} n(d, w)P(z|d, w)}{\sum_{d, w} n(d, w)} \quad (2.4)$$

$$(2.5)$$

where $n(d, w)$ is the number of times word w occurs in document d .

Expectation Maximization (EM) is a technique for extracting probability distributions that involve latent (i.e. unobserved) variables. For PLSI, at no time are the topics observed. EM assumes that each topic has its own probability distribution (e.g. the probability distribution $p(w|z)$ is potentially different for each value of z). The corpus, then, is a mixture of these different probabilities. For simplicity, EM explicitly assumes that each of these distributions is Gaussian. Although, recent work suggest this is an inappropriate assumption [?] for modeling documents, the empirical results show that EM is effective. EM works by *pretending* that it knows the parameters of each of these distributions. It then infers the probability that each of the observed data points was drawn from a specific distribution. At each step, EM increases the log likelihood (and consequently the likelihood) function. After sufficient iterations, EM is guaranteed to reach a local maxima in likelihood. However, since it is not guaranteed to reach a global maximum, in practice it is often run several times from different starting points [?].

Unfortunately, PLSI is limited by explicitly using the corpus to develop its model. Although, the values it comes up with are reasonable for the corpus, there is no reason to think that they are reasonable for some new, not yet seen, document. In other words, while it is good at *predicting*, it is bad at generating. As a generative model, PLSI implicitly assumes that associated with each topic is a multinomial distribution, specifically the probability of a word being selected given that topic. PLSI

then uses the document to determine what those probabilities are. LDA assumes that there are an infinite number of multinomials that could be associated with picking a topic. The multinomial that actually emerges from the corpus is just one possibility. LDA explicitly “picks” one collection of these topic probabilities by including a parameter, α , selected a priori for a dirichlet distribution (for information on why the dirichlet distribution makes sense, refer to [?,?]). Then θ is drawn from this distribution producing a multinomial probability for topics. This mechanism prevents LDA from being constricted by the corpus. However, if LDA was modified so that α was 1, i.e. the probability of picking any topic was uniform, then LDA becomes PLSI. Said differently, PLSI is just a special case of LDA [?]. In addition to θ , the probability distribution for topics, the conditional probability of picking words given a topic is also required. LDA assumes this joint probability exists a priori and labels it β . The goal, then, for LDA is to determine $p(z, \theta | d, \alpha, \beta)$. This is seen graphically in Figure ??.

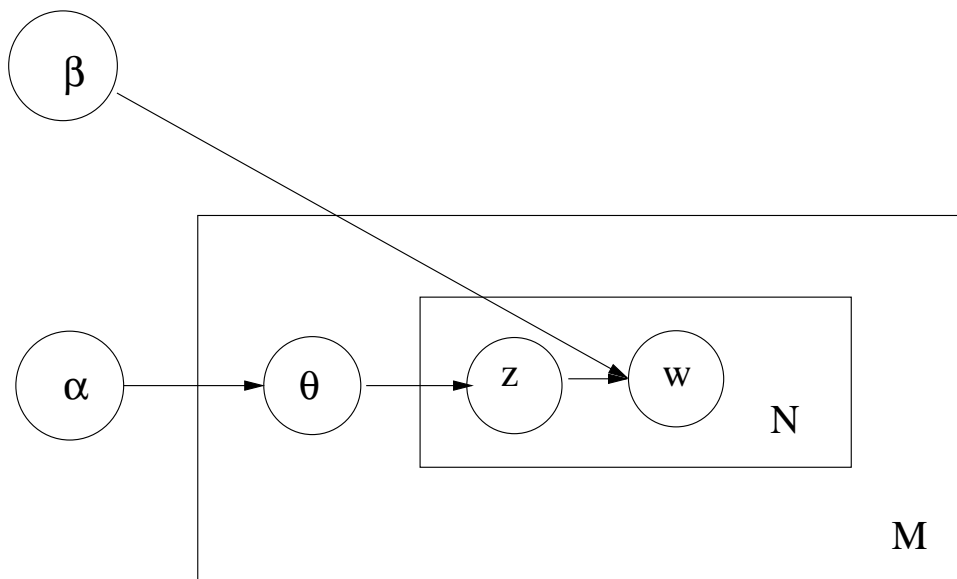


Figure 2.4: Latent Dirichlet Allocation Model

Based on this model, in order to create a document, N words are selected. To pick a word, a topic needs is first selected. Since the probability of picking a topic is conditioned on θ , the probability of picking a topic is $p(z|\theta)$. After the topic

is selected, the probability of picking a word is conditional on the topic, z , and β . Therefore, the probability of selecting a word is $p(w|z, \beta)$. However, since a word can come from any topic, $z_1..z_k$, the probability of each word is actually the probability of that word for topic 1 plus the probability of that word for topic 2 and so on. Similarly the probability of a topic must also be considered for each possible θ . Therefore, the generative equation for a document is:

$$\int p(\theta|\alpha) \left(\prod_{i=1}^N \sum_{j=1}^K p(z_j|\theta) p(w_i|z_j, \beta) \right) d\theta \quad (2.6)$$

and a generative model for the corpus of:

$$\prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{i=1}^{N_d} \sum_{j=1}^K p(z_j|\theta_d) p(w_i|z_j, \beta) \right) d\theta_d \quad (2.7)$$

As with PLSI, the goal is to compute $p(z, \theta|d, \alpha, \beta)$. By using Bayes' Rule, the resulting equation is:

$$p(\theta, z|d, \alpha, \beta) = \frac{p(\theta, z, d|\alpha, \beta)}{p(d|\alpha, \beta)} \quad (2.8)$$

However, this equation is intractable because θ and β are linked through z . To overcome this, Blei, et al. [?] use variational inference to approximate values of α and β .

Variational methods work by taking some function that is either computationally expensive to calculate or, as in the case of LDA, impossible to compute and by introducing additional variables coming up with a function that is computed easily [?]. For instance, the function $\ln x$ is expensive to calculate. However, by introducing an-

other parameter, λ , a linear function, $\lambda x - (\ln \lambda + 1)$, can now provide a tight upper bound on $\ln x$.

This broadens to inference by expanding the concept of functions to probability distributions by introducing an approximating family of conditional probability distributions with variational parameters. For more details, refer to [?], [?].

For LDA, Blei [?] introduces two variational parameters, γ and ϕ for α and β . The distribution family that results is:

$$q(\theta, x | \gamma, \phi) = q(\theta | \phi) \prod_{k=1}^K q(z_k | \phi_k) \quad (2.9)$$

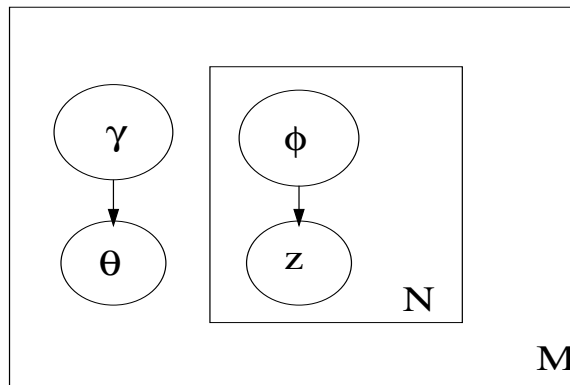


Figure 2.5: Latent Dirichlet Allocation Variational Inference Model

Then by using Expectation Maximization, LDA arrives at values for γ and ϕ of:

$$\phi_{ni} \propto \beta_{iw_n} \exp\{E_q[\ln(\theta_i) | \gamma]\} \quad (2.10)$$

$$\gamma_i = \alpha_i + \sum_{n=1}^N \phi_{ni} \quad (2.11)$$

After calculating values for α and β , there is still the need to consider smoothing the probability distribution. It is still likely that new documents will contain words that are not represented in β . Therefore, variational inference is performed again, adding a prior approximating variable, η for β . Finally after performing Expectation

Maximization again and getting the same equations for γ and ϕ as well as an additional equation, $\lambda_{ij} = \eta + \sum_{d=1}^M \sum_{n=1}^N \phi_{dn} w_{dn}^j$.

There has been recent work to improve upon LDA. One proposed enhancement is to use Gibbs Sampling instead of Variational Inference in order to arrive at the best values of α and β [?]. A second proposal is instead of considering an infinite number of probability distributions for the topics, consider a discrete number of them. Keller and Bengio [?] have proposed a Theme Topic Mixture Model (TTMM) that considers a theme to consist of a finite number of topics. Therefore, given a particular theme, there is a specific probability of a topic occurring. In this case, given a finite number of themes, LDA transforms into a model that is tractable by Expectation Maximization. What makes this model more difficult is that in addition to deciding a priori how many topics there are, to make the model work one must also decide a priori how many themes there are. Finally, Minka and Lafferty have proposed that the variational methods used by Blei, et al. [?] lead to “inaccurate inferences and biased learning”. Instead they propose using expectation-propagation to produce better estimates of α and β [?] which works by iterative applying “deleting/ inclusion” steps on the integral.

2.3.3.4 Extending Mixture of Words and Topics to Users. While the original objective for developing LDA and PLSI was to organize collections of text (i.e. documents) into latent topics, this objective has recently expanded to include attributing people to topics. These user models are required to determine user interests.

2.3.3.5 Author Topic. The first such model is the Author-Topic (AT) model described by Rosen-Zvi, et al. [?]. They introduce a new author variable. There are P authors in the corpus. Each document contains a subset of people who authored the document. AT [?] assumes that there are an infinite number of multinomials that could be associated with picking a topic. The multinomial that actually emerges from the corpus is just one possibility. AT explicitly “picks” one

collection of these topic probabilities by including a parameter, β , selected a priori for a dirichlet distribution. Then ϕ is drawn from this distribution producing a multinomial probability for topics conditioned on users. This mechanism prevents AT from being constricted by the corpus. In addition to β , the probability distribution for topics, a second parameter, α is also selected a priori for a second dirichlet distribution. In a similar manner to topics, θ is then drawn from this distribution to produce a multinomial distribution for each word conditioned on the topic (Figure ??).

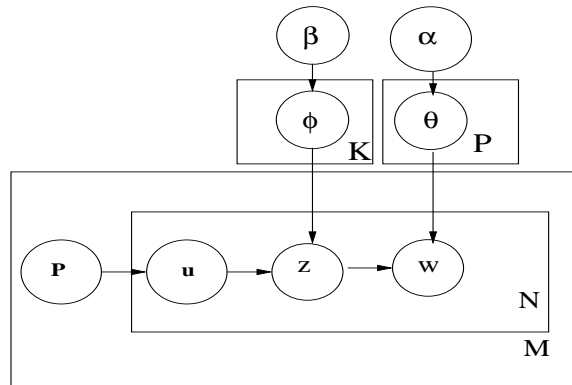


Figure 2.6: Author Topic Model

Based on this model, in order to create an email, N words are selected. To pick a word, an author, u , is chosen uniformly from the population P . Then a topic, z , is selected conditioned on the author chosen. Since the probability of picking a topic is also conditioned on ϕ , the probability of picking a topic is $p(z|u, \phi)$. After the topic is selected, the probability of picking a word is conditional on the topic, z , and θ . Therefore, the probability of selecting a word is $p(w|z, \theta)$. However, since a word could be selected for any topic, the probability of each word is actually the probability of that word for topic 1 plus the probability of that word for topic 2 and so on. Similarly the probability of a topic must also be considered for each possible ϕ . Therefore, the generative equation for an email is:

$$\int p(\phi|\beta) \left(\sum_{r=1}^L p(u_r) \int p(\theta|\alpha) \left(\prod_{i=1}^N \sum_{j=1}^K p(z_j|\phi) p(w_i|z_j, \theta) \right) d\theta \right) d\phi \quad (2.12)$$

and a generative model for the corpus of:

$$\prod_{d=1}^M \int p(\phi_d|\beta) \left(\sum_{r=1}^L p(u_r) \int p(\theta_d|\alpha) \left(\prod_{i=1}^{N_d} \sum_{j=1}^K p(z_j|\phi_d) p(w_i|z_j, \theta_d) \right) d\theta_d \right) d\phi_d \quad (2.13)$$

As with PLSI, the goal is to compute $p(z, \phi, \theta|d, \alpha, \beta)$. By using Bayes' Rule, the resulting equation is:

$$p(\phi, \theta, z|d, \alpha, \beta) = \frac{p(\phi, \theta, z, d|\alpha, \beta)}{p(d|\alpha, \beta)} \quad (2.14)$$

However, this model suffers from the same intractability problem that Blei, et al. suffered from in LDA. However rather than resolving this problem by using variational inference, Rosen-Zvi, et al. solved it by using Gibbs sampling. Gibbs Sampling works by randomly assigning words to users and topics and then finding the resulting conditional probabilities. This process is then repeated until the conditional probabilities converge. Recall that based on the model, a user is selected uniformly. Then a topic is selected conditioned on that user and then a word is selected conditioned on that topic. Therefore, since with a corpus, we start with words and users, it is necessary to work backwards. By looking at the number of times a user has been assigned to a particular topic, one infers the probability of that topic given that user (i.e. $p(z|u) = n(u, z)/n(u)$ where $n(u, z)$ is the number of times topic z was assigned to user u and $n(u)$ is the number of times user u occurs in the corpus (i.e. the number of times user u is assigned to *any* word). Similarly, one infers the number of times a word was chosen for a given topic (i.e. $p(w|z) = n(w, z)/n(z)$ where $n(w, z)$ is the number of times word w was assigned to topic z and $n(z)$ is the number of times topic z occurs in the corpus (i.e. the number of times topic z is assigned to *any* word).

Mathematically, the conditional probabilities are:

$$p(z|u) = \frac{n(u, z) + \alpha}{\sum_{z'} n(u, z') + K\alpha} \quad (2.15)$$

$$p(w|z) = \frac{n(w, z) + \beta}{\sum_{w'} n(w', z) + V\beta} \quad (2.16)$$

$$p(u, z|w) = p(z|u)p(w|z) \quad (2.17)$$

Algorithmically:

1. Assign random probabilities to all conditional probabilities, i.e. $p(z|w)$ and $p(u|w)$, such that they produce a probability distribution (i.e. the probabilities are all non-negative and sum to one).
2. For every word in every document, “determine” what topic and user produced it. To do this, pick a random number between 0 and 1 and see which conditional probability it falls into.
3. Based on the number of times each user and topic was assigned to a word, re-calculate the conditional probabilities.
4. Repeat steps 2 and 3 until convergence.

While it is possible to estimate α and β , following Rosen-Zvi, et al. [?], α is set to $50/K$ and β is set to 0.01.

While Rosen-Zvi, et al. created their generative model to describe the creation of papers for scientific journals, it also applies without adjustment to the creation of emails. If we are only interested in the people who send emails, then the model considers each document as having only a single author. If instead, recipients of emails are also included in the model, than “authors” expand to mean all the people associated with the email, both senders and recipients. Very recently, McCallum, et al. [?] observed that using this model doesn’t differentiate between the senders and recipients. For instance, senior level personnel often have email discussions amongst themselves but blind carbon copy their assistants to perform some related task. In this

case, it is desirable to segregate the different recipients by their roles. McCallum, et al. accomplished this by taking the Author Topic model and separating out recipients from authors (Figure ??). Now each message is composed of only one author but potentially multiple recipients. In addition, the selection of a topic is now based on both the author and the recipient chosen for the given word. A final model also described in [?] includes adding additional variables for roles (Figure ??). In this case each author has one or more roles associated with him. Once a recipient is selected, roles for the author and the recipient are selected conditioned on the author and recipient chosen. These roles are then used to select topics which are then used to select words.

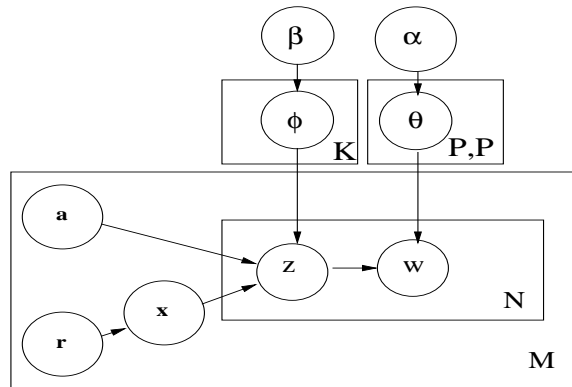


Figure 2.7: Author Recipient Topic Model

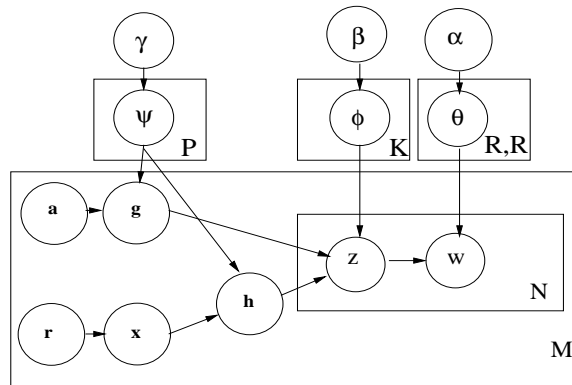


Figure 2.8: Role Author Recipient Topic Model

When considering recipients, depending on the application, one may wish to include or exclude those people carbon copied and/or blind carbon copied. Both the

Author Topic (AT) model and the Author Recipient Topic (ART) model do include recipients. In order to select between them, it is important to consider whether the person who sends the message is qualitatively different from the one who receives it. McCallum, et al. argue that by differentiating between senders and receivers, it is possible to extract the different roles that these people play. While this certainly makes sense, it does not apply to the particular application of only extracting people's interests. Whether a person sends an email about going to a basketball game or receives an email about it, in both cases it is reasonable to assume that the person is interested in basketball. Therefore, for clustering people interests via email the AT model is more appropriate.

2.3.3.6 PLSI-U. While LDA has been expanded to address users through the AT model, PLSI has not. However, the expansion of PLSI to include users is fairly straightforward. To begin, recall that Bayes rule states that $p(a, b) = p(a|b) \cdot p(b) = p(b|a) \cdot p(a)$. Therefore when considering $p(z|u, d, w)$ where u represents a person:

$$p(z|u, d, w)p(u, d, w) = p(u, d, w|z)p(z) \quad (2.18)$$

Now, consider the model in Figure ?? and observe that u , d , and w are all conditionally independent given z (this is a subtle difference from the Author Topic model where z is dependent on u and w is dependent on z). It now follows that:

$$p(z|u, d, w) = \frac{p(u|z)p(d|z)p(w|z)p(z)}{p(u, d, w)} \quad (2.19)$$

But $p(u, d, w)$ is simply $p(u, d, w|z)$ marginalized across all possible z 's. So finally,

$$p(z|u, d, w) = \frac{p(u|z)p(d|z)p(w|z)p(z)}{\sum_{z' \in Z} p(u|z')p(d|z')p(w|z')p(z')} \quad (2.20)$$

In order to evaluate the conditional probabilities in the above equations, consider:

$$p(w|z) = \frac{p(z|w)p(w)}{p(z)} \quad (2.21)$$

By marginalizing across u and d , we get:

$$p(w|z) = \frac{\sum_{u \in U} \sum_{d \in D} p(z|u, d, w)p(w|u, d)}{\sum_{u \in U} \sum_{d \in D} \sum_{w' \in W} p(z|u, d, w')} \quad (2.22)$$

Finally, consider what $p(w|u, d)$ means. This is the probability of a given word occurring for a given document and person. Since the document and person are already specified the probability space is the one document. Therefore the probability is the number of times the word appears in the document divided by the number of words in the document. Therefore:

$$p(w|z) = \frac{\sum_{u \in U} \sum_{d \in D} p(z|u, d, w)n(d, w)}{\sum_{u \in U} \sum_{d \in D} \sum_{w' \in W} p(z|u, d, w')n(d, w)} \quad (2.23)$$

where $n(d, w)$ is the number of times a word occurs in a document. Observe that since a document is the same regardless of which “author” is considered, it is sufficient to specify $n(d, w)$ so long as it is summed across all people. Furthermore, since the denominator sums across all words, the net effect is the quotient described previously. This equation extends naturally to documents and users:

$$p(d|z) = \frac{\sum_{u \in U} \sum_{w \in D} p(z|u, d, w)n(d, w)}{\sum_{u \in U} \sum_{d' \in D} \sum_{w \in W} p(z|u, d', w)n(d, w)} \quad (2.24)$$

$$p(u|z) = \frac{\sum_{d \in D} \sum_{w \in W} p(z|u, d, w)n(d, w)}{\sum_{u' \in U} \sum_{d \in D} \sum_{w \in W} p(z|u', d, w)n(d, w)} \quad (2.25)$$

$$p(z) = \sum_{u \in U} \sum_{d \in D} \sum_{w \in W} p(z|u, d, w) \quad (2.26)$$

These equations now form the expectation (eq. ??) and maximization (eq. ??, eq. ??, eq. ??, eq. ??) equations for Expectation-Maximization (EM). EM alternates two steps:

1. Assign random probabilities to $p(d|z)$, $p(w|z)$, $p(u|z)$, and $p(z)$ such that they produce probability distributions (the probabilities are all non-negative and sum to one).
2. Calculate all of the values for $p(z|u, d, w)$.
3. Using the values from step 2, calculate the new values of $p(d|z)$, $p(w|z)$, $p(u|z)$, and $p(z)$.
4. Repeat steps 2 and 3 until convergence.

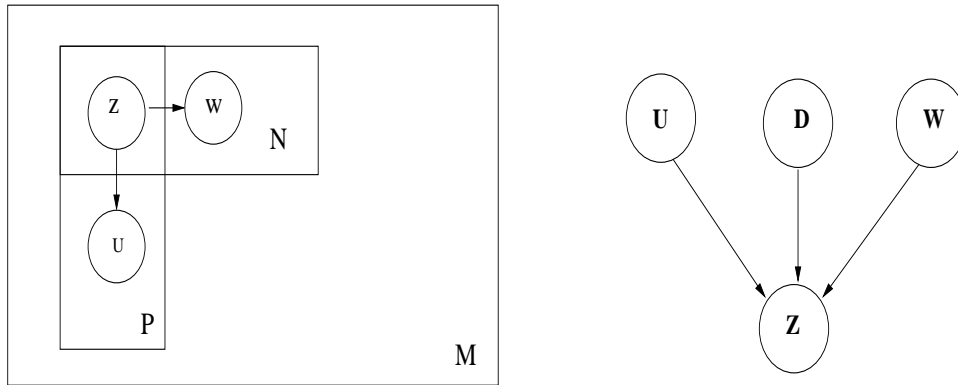


Figure 2.9: PLSI-User Mixture Model

Only one more point needs discussing before moving to the implementation of these models. The Author Topic model focuses on the relationship between authors and categories and categories and words. Other than acknowledgment that these words and authors are combined into documents, documents are not explicitly included in the development of the model. However, since it is necessary to classify the emails in order to create the explicit network, the probability of a category given a document needs development. Observing that for a document to exist, each word in the document must exist, it follows:

$$p(d|z) = \prod_{w \in d} p(w|z) \quad (2.27)$$

The one problem with this is that for documents containing many words, the conditional probability of any document will quickly go to zero. Even if the conditional

probabilities are later normalized to construct a probability distribution, if too much precision is lost first, such a calculation won't work. To avoid this problem, and noticing that the probabilities will be normalized anyway, the logarithm of the function is taken. Furthermore, to prevent the problem of $\log 0$, the function is multiplied by e^N . The result is:

$$p(d|z) \propto \log \sum_{w \in d} (1 + p(w|z)) \quad (2.28)$$

2.4 Examining Social Networks

When the clustering algorithms complete, three things are produced: the most probable words for a topic (highest value of $p(w|z)$), the most probable individuals for a topic (highest value of $p(u|z)$), and the most probable documents for a topic ($p(d|z)$). Topics are then considered as topics of interest for an individual if $p(z|u)$ exceeds a certain threshold. An individual's profile is then the collection of all of his topics of interest.

Once profiles have been generated, the final step is to use these profiles to find likely insider threats. The method used for this research is performed in three steps. The first step is to use these profiles to find individuals that have an interest in sensitive topics. The second and third step depend on using these profiles to develop social networks. The second step uses social networks to find individuals that feel alienated from the organization. Finally, once an insider is known, the third step uses social networks to find other individuals with common interests.

2.4.1 Social Network Attributes. A social network is “a finite set or sets of actors and the relation or relations defined on them” [?]. Actors are either individuals or groups of individuals and the relations between them are any form of social relationship. For this paper, actors refer to individuals and two actors are considered related if they have at least one interest in common. Consider Figure ???. There are several ways to construct this graph. Obviously each vertex represents one actor or person but it is not clear how to represent a shared interest. For example, if John

is very interested in sports and Mary has a small interest in it, do they share this interest? On the other hand, if Susan and Mike share two interests like tennis and skydiving, should they be connected by two edges or should they only be connected by one edge with a much higher weight [?]?

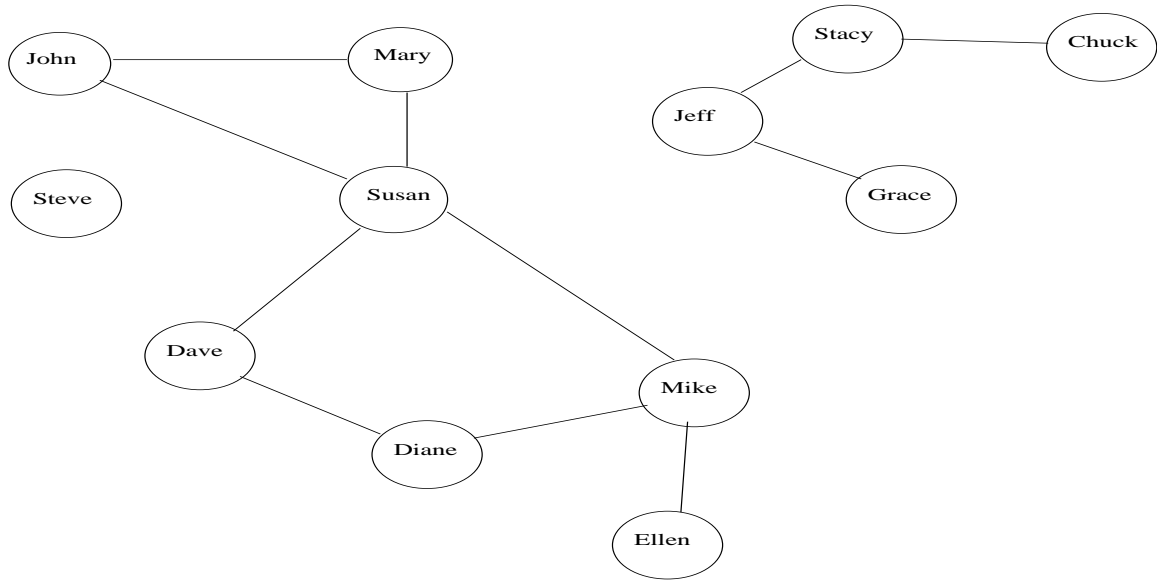


Figure 2.10: Example of a Social Network Graph

Table 2.1: Example of an Adjacency Matrix

	John	Steve	Mary	Susan	Dave	Diane	Mike	Ellen	Jeff	Stacy	Grace	Chuck
John												
Steve												
Mary	1			1								
Susan	1		1		1		2					
Dave				1		1						
Diane					1		2					
Mike				2		2		1				
Ellen							1					
Jeff										5	1	1
Stacy									5			
Grace									1			
Chuck										1		

Once the graph has been constructed, the associated adjacency matrix (Table ??) is used to determine the number of paths between people. The adjacency matrix itself shows the number of paths of length 1 between people, i.e. the number of relationships that people have with each other. This is also the number of edges between vertices. The square of the adjacency matrix (i.e. the adjacency matrix multiplied by itself) shows the number of paths of length 2 between people, i.e. how many

“friends” people have in common. This extends to paths of any length by continuing to multiply the adjacency matrix by itself [?]. The number of paths between people is used for many different purposes. In many applications, this knowledge identifies the “power brokers”, the people on the most paths between people. There are several methods used to measure an actor’s importance.

These individual measurements can then be aggregated to arrive at a measure of the network’s cohesiveness. Consider the three graph types shown in Figure ???. In the Star graph, everyone is clustered around one actor. In addition to this one actor being very important, the overall network is also very effective at disseminating information quickly since everyone is within a distance of at least two of everyone else. Unlike the Star Graph, disseminating information in the Circle and Line graphs may take significantly longer since the distance between two arbitrary people may be quite large. However, while information may take longer to spread in the Line and Circle graph, they do appear more “fair” since roughly everyone is equally well-connected. Also, removing one actor from the Line graph does halt information flow, while removing the wrong actor from the Star graph disconnects everyone.

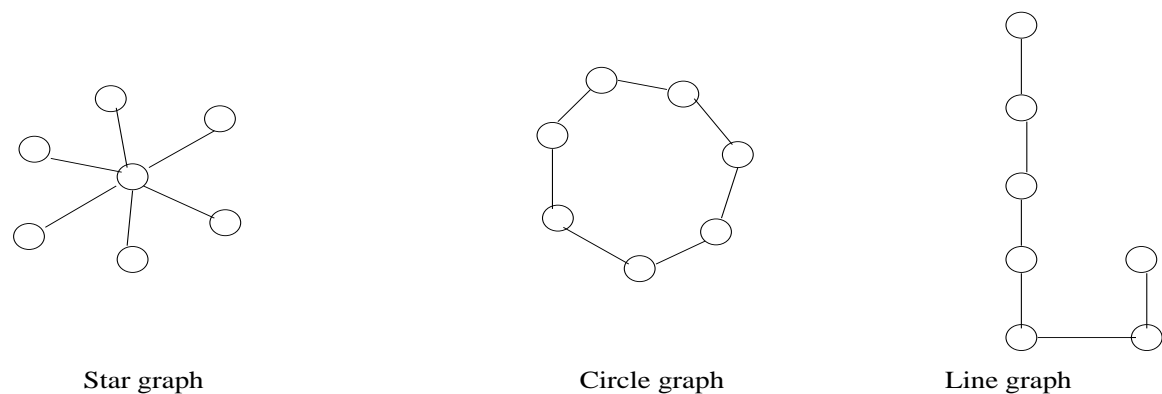


Figure 2.11: Some representative Social Network Graphs

2.4.2 Centrality Measurements. Several measurements that are used in social network analysis to describe an individual’s importance or centrality are: degree, closeness, and betweenness [?]. Degree is the simplest measurement of centrality and assumes that the most central actors must be connected to the most other actors.

The normalized form of this measurement is:

$$C'_d(v_i) = \frac{d(v_i)}{(N - 1)} \quad (2.29)$$

where C'_d is the normalized measurement of degree centralization, v_i is a vertex (actor) in the network, $d(v_i)$ is the degree of the vertex and N is the number of actors in the network. For the Star graph, the resulting values are 1.0 for the central actor and 0.167 for each of the other actors. The values for all of the actors in the Circle graph are 0.333 while the two on the end for the Line graph are 0.167. Actor degree centralization can be extended to the overall network several ways, all of which measure dispersion in some way. One common formula [?] is:

$$C_d = \frac{\sum_{i=1}^N (C_d(v^*) - C_d(v_i))}{(N - 1)(N - 2)} \quad (2.30)$$

where $C_d(v^*)$ is the maximum degree centralization for the network. Another measurement using degree is the density of a network. Density measures how many connections exist between individuals. The more individuals who communicate, the denser the graph is. Density is measured by:

$$density = \frac{\sum C'_d(v_i)}{N} \quad (2.31)$$

The density for the sample graphs are: Star graph: 0.286, Circle graph: 0.333, Line: 0.286. Observe that the density of the complete graph is 1 and the empty graph is 0.

A second measurement of centrality is closeness and measures the distance of an actor from the other actors in the network. It does this by calculating the distance of the shortest path between a vertex (actor) and all of the other vertices in the network. The equation for normalized closeness centralization is:

$$C'_c(v_i) = \frac{N - 1}{\sum_{j=1}^N d(v_i, v_j)} \quad (2.32)$$

where $d(v_i, v_j)$ is the distance of the shortest path from v_i to v_j . Observe that this measurement assumes a connected network since if the network is disconnected then at least one distance is infinite. There are a couple of ways of overcoming this problem. One is to set the distance between two vertices without a path between them to one more than the longest possible shortest path. A second is to perform this measurement on the largest component of the network. While in general this would not make sense, if the network has only one large cluster, e.g. containing 98% of the vertices, then such a measurement is a good heuristic for the closeness measurement. For the Star graph, the closeness centralization of the central actor is 1.0 while it is 0.545 for all of the other actors. For the Circle graph, the measurement of all of the actors is 0.5. Finally, for the Line graph, the measurement is 0.50 for all of the actors except the two on the end for which it is 0.286. In the similar manner as degree, closeness can be extended to the graph as follows:

$$C_c = \frac{\sum_{i=1}^N (C'_c(v^*) - C'_c(v_i))}{(N-2)(N-1)/(2N-3)} \quad (2.33)$$

where $C'_c(v^*)$ is the largest closeness centralization measurement. For the Star graph, this measurement is 1.0 signifying that there is one actor connected to everyone else. Conversely, the Circle graph achieves the measurement of 0 signifying that all shortest path distances are equal. Finally, the Line graph has a relatively small value of 0.277.

The last centralization measurement to discuss is betweenness. This is a very expensive measurement to calculate as it measures the *number* of shortest paths between vertices that pass through a third vertex compared to the total number of shortest paths between vertices. Rather than use the Star, Circle, or Line graphs, consider the graph in Figure ?? . A is on shortest paths from G to F , G to C , G to D , and G to E . However, in each case, there is another shortest path that does not involve A (namely the one through B). Therefore, $C_b(A) = 0.5 + 0.5 + 0.5 + 0.5 = 2$. Finally, this result can be normalized by dividing by the maximum number of shortest

paths through a vertex $((N - 1)(N - 2)/2)$. The equation, then, is:

$$C_b(v_i) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2.34)$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest paths from s to t through v . For the sample graph, the results are: $C_b(A) = 2$, $C_b(B) = 2$, $C_b(C) = 5$, $C_b(D) = 0$, $C_b(E) = 0$, $C_b(F) = 11$, $C_b(G) = 0$. This normalized value of this is:

$$C'_b(v_i) = C_b(v_i)/((N - 1)(N - 2)/2) \quad (2.35)$$

One advantage of this measurement over closeness is that it can be computed exactly even for disconnected graphs. Unfortunately, one disadvantage has been a computation cost of $\theta(n^3)$. However, Brandes [?] has recently developed an algorithm that can compute betweenness for unweighted graphs in $O(NM)$ where M is the number of edges. For dense graphs, where M can be as large as $(N - 1)(N - 2)/2$, this isn't helpful. However, in sparse graphs, this is a huge timesavings. Finally, centralization can also be extended to networks similarly to degree and closeness as:

$$C_b = \frac{\sum_{i=1}^N C'_b(v_i) - C'_b(v_i)}{N - 1} \quad (2.36)$$

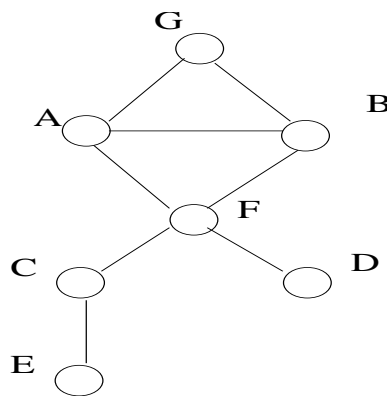


Figure 2.12: An example Social Network Graph to measure Betweenness

2.4.3 Current Social Network Research That Uses Email. It is important to remember that the entire social network is constructed based on perceived common interests. This could be further enhanced by adding edges between people who already know each other. This can be done a priori by an expert and/ or automatically by examining the email headers of each person. While using an expert to do it requires more effort, it does have the added benefit of allowing the shared interest (i.e. why do the two people know each other) encoded in the link. Once this augmented graph has been constructed, it is used to infer other clandestine networks. Liben-Knowell and Kleinberg [?] show that several proximity measurements predict additional links about 40 to 50 times better than pure chance. These measurements include the shortest path between two vertices, the total number of distinct paths between two vertices, the length of a random walk between two vertices, and the degrees of two vertices.

There already exist tools that develop social networks from email information. ReferralWeb [?] uses the co-occurrence of names in close proximity in any document publicly available on the WWW (e.g. journal articles, newsgroups, chatrooms etc.) to denote a close relationship. Lada Adamic has done similar research [?] by using mailing lists and the homepages of students at Stanford and MIT. Since, when people create homepages, they link to their friends' homepages (and ask their friends to link to theirs), she postulated that using homepages would result in an appropriate social network. She also used the text present on the web pages to further predict relationships (i.e. common interests) between people. While she was able to show that the text provided strong indications of friendships between people, it is unclear if this would generalize beyond the rather closed community of a university. Culotta, et al. [?] approached this problem differently but from a more general population. They began by extracting names from email messages. Then they used the WWW to find the person's "web presence" (generally his or her homepage) and used that to describe the person and to find friends of that person. After the network was created, they used graph partitioning algorithms to find highly connected components. While their

dataset was small (53 email correspondents), their results were promising. However, one of the biggest drawbacks was a lack of web presence for many of the correspondents (31 of 53). Although it is likely that over time more and more people will have a web presence (e.g. young adults are much more likely to create home pages than people in their forties and fifties), it will probably still be a while before web presence is a reliable means of predicting relationships.

Since September 11, 2001 there has been increased research in uncovering potential threats through the use of social networks. However, despite several organizations, such as Rand Corporation and Mitre, making proposals for using social networks to detect insider threats [?,?], little public research has been done. Symonenko [?] has generated social networks of intelligence analysts and then used semantic analysis to detect when individuals are showing interests in areas outside of their group. While the results have been promising, the technique requires a large amount of interviews with experts to provide the semantic analysis. In addition, this expert knowledge is then only applicable to the specific group and needs to be repeated each time the application is moved to another organization. Yee [?] has also performed some initial research into generating social networks from email headers for later analysis by social network analysts.

The actual methods used to develop social networks from email activity are discussed in more detail in Chapter ?? and the actual email corpus used in testing is discussed in the following section.

2.5 Using Enron's Email as Source Data

Electronic mail is fast becoming the most common form of communication. By 2006, email traffic is expected to exceed over 60 billion messages daily [?]. While email was always expected to increase communication between people in different cities and countries, it is unexpected how much people in the same building and even the same room communicate via email. Not only is email used to inform others, it is also used to inform one's self. It is becoming more and more common for people to

send email to themselves as reminders. In addition to these true information emails, unwanted email, also known as spam, has evolved from an academic curiosity, to a minor nuisance, to a significant business problem. Spam accounts for over half of all messages sent [?]. In addition to relatively innocuous spam, in 2004, at least eight out of the ten most frequently reported computer worms were delivered via email.

Using email as data for research is not new. However, the primary objectives of this specific kind of research have changed. While earlier topics like email categorization [?] and spam filtering [?, ?] continue to be actively researched, new topics have emerged as well. Two years ago a conference (Conference on Email and AntiSpam (CEAS)) was created at Stanford that just focused on email and span (www.ceas.cc). In addition to spam filtering and email categorization, its papers have focused on topics like extracting social networks from email [?, ?], inferring user activity from email [?], and extracting text features [?].

In addition, email is also beginning to emerge as a tool for detecting deceptive communications [?]. While there has been a large amount of research in preventing incoming mail that is deemed suspicious [?], the idea of reading the outgoing mail has not received a lot of activity. This is due in large part to privacy concerns and the lack of large-scale email datasets. While Keila and Skillicorn focus on the use of email features to detect deceptive emails, it is equally reasonable to use the semantic content [?, ?]. Semantic analysis, i.e. extracting meaning from text, has been directly applied to countering insider threats by Symonenko, et al. [?]. They investigated the effectiveness of using natural language processing (NLP) to discover intelligence analysts who were accessing information outside of their community of interest.

2.5.1 Privacy Concerns. The benefits of finding better ways to protect against threats must be balanced with respecting people’s privacy. Despite signing consents to monitoring, people still have the right to expect to come to work without having someone rifle through their desk [?]. In this new electronic age, the same consideration is given to people’s electronic desks (i.e. their computers, personal data

assistants, et cetera) as to their physical ones. This is even more true when the purpose of monitoring is not to detect a real threat but only to perform research. Therefore, before gathering electronic data, investigators must first consider ways to protect people's privacy. One possibility is to have everyone sign a consent form. While this has been done for small scale studies (e.g. 1 - 20 people), it is unrealistic when the goal is to study an organization with personnel in the thousands. The researcher must consider other options.

One method often used to address privacy concerns is stripping out people's names from email headers and replacing them with unique identifiers [?]. In this manner, it is possible to develop social networks without the chance of someone's privacy being violated. Unfortunately, since the research is often based on only the email headers, the same clean up is not done for the email text itself. In many ways this problem is harder since it is possible to refer to someone by many different names (e.g. Robert Smith, Rob, Bob, the boss, the President). One way to address this is to simply remove all proper names from email as well as some titles (e.g. boss, dean, etc.). Unfortunately, this may also strip out words that are helpful to categorize the email. For instance, it may be desirable when looking through email to categorize based on words like "George Bush", "al Qaeda", and "Osama bin Laden".

There is another alternative to artificially generated email and real world email with its associated privacy concerns. During the investigation into the Enron scandal, the Federal Energy Regulatory Commission (FERC) made email from Enron publicly available. As a part of this process, it placed the email of 151, primarily senior, employees of Enron accessible electronically on the World Wide Web (<http://fercic.aspensys.com/members/manager.asp>). This data was later taken by researchers at several institutions and organized and cleansed of some integrity problems before being assembled into a useable format. The data is now available as a series of text documents from Carnegie Melon University [?] and as a MySQL database from University of California at Berkeley [?]. The data (after cleanup) consists of email from 151 people's folders (although it seems that two people appear

twice with different user names and one person only sent automated calendar reminders) comprising 250,484 unique messages. While the data has not been available for long, it is beginning to come to the attention of researchers as a unique source of data. Some of the earliest work using the Enron data includes analysis of automatic email classification into folders [?,?], extracting hidden (i.e. forwarded) emails from emails [?], determining email response times [?] and investigating the structure of words in email [?], as well as developing social networks of Enron [?,?].

In this thesis, the email bodies from the Enron corpus are used to develop user interest profiles. These user profiles are then used to develop an implicit social network based on interests. In addition, a second, explicit, social network is constructed from the email headers. This social network maps the explicit links between people. It is posited that while people may not communicate with all of the people who have the same interests, they will talk to only people with whom they share at least one interest. If people appear to have an interest but do not talk to anyone with a similar interest, this may be suggestive of a hidden interest. While most hidden interests are likely innocuous, their discovery may indicate the need for more specific attention by people like their supervisors or security. The data sources and the techniques used to extract the relevant information is discussed in more detail in Chapter ??.

2.5.2 A Brief Synopsis of Enron. Since this thesis is based on the Enron data, a little background on the events contributing to demise of Enron is helpful. Ken Lay entered into the natural gas industry back in the 1980s because he believed that the industry would soon become deregulated. It did and he was able to create a successful company by riding the wave of natural gas deregulation. In 1990 Lay brought in Jeff Skilling to run a new business within Enron focused on selling long term fixed price contracts for natural gas to regional distributors and purchasing long term fixed price contract from natural gas refineries. The profit came from the difference between the two contracts, modelled in part like bank mortgages. Banks loan money to people (gas refineries) to buy a house and collect monthly payments from

mortgagees and other customers (regional distributors). One of the original things Skilling did was implement “mark-to-market” accounting within this new business. This means that the total profit is recognized at the time the contract is purchased from the refinery by estimating what it will sell for during the next twenty years.

Of course, if the actual profits are different, the additional profit (or loss) must be recognized in full as soon as the company becomes aware of it. Therefore, to avoid reporting losses it is important to lock in the gains. When Skilling first asked Andy Fastow, the Chief Financial Officer, for a way to lock in these gains, Fastow turned to the Research Group for answers. He was told by its head, Vince Kaminski, that there was no way to fully hedge risk. However, Fastow refused to take no for an answer. He continued to investigate and eventually decided that he had discovered a way to lock in gains (although he chose not to ask the Research Group for confirmation) by using off-book-partnerships. The current General Accounting Office (GAO) laws allowed a new company, we will call X, to be considered independent even if 97% of the company is owned by company A (Enron) so long as the last 3% is owned by an independent company. This new company, company X, then provides insurance to company A guaranteeing that company A would never lose any of the profits recognized. This mechanism can also be used to allow company X to purchase assets from company A for more than they are worth. While in reality nothing has changed, on the Profit and Loss statement provided to the shareholders of company A, everything looks excellent.

One thing that made this practice worse was that although company X was supposed to be independent of company A, in Enron’s case it wasn’t. The 3% that was supposed to be independent was owned either by Andy Fastow or another Enron employee, Michael Kopper. They managed the companies (known by various names such as LJM and Raptor) and ensured their stakes were recovered through management fees paid by Enron (along with tens of millions of dollars more unbeknownst to Lay and Skilling). In the end, Enron was the only one with a financial interest in these companies. The final card in the house of cards was how all of this was financed. Lay, Skilling, Fastow, Kopper, and everyone else believed that Enron stock

would continue to go up. They therefore used this “fact” to guarantee the profits by selling Enron stock to company X for capital. If Enron stock ever went below a certain level, company X would be calling on Enron to provide money so that it could pay off on its insurance policy to Enron. However this would be happening at the very time that Enron couldn’t pay company X. This was in fact what happened in late 2001.

Other things contributed to Enron’s downfall including the negative publicity surrounding the California energy crisis of 1999 and 2000. It appears that Enron manipulated demand during this crisis to make huge profits. They received a lot of bad publicity which may have contributed to Wall Street analysts finally beginning to dig into Enron’s financials. Also, the house of cards based on hiding losses began to crumble as the next “big idea” (in this case Enron Broadband and deregulation of electricity sales) didn’t happen. Shortly before this happened, in August of 2001 Sherron Watkins, an employee of Fastow’s, wrote a letter to Ken Lay to explain in detail what Fastow was doing. While this letter wasn’t discovered by Lay until too late (and did not leave Enron until long after Enron declared bankruptcy), this is the best example of an insider within Enron [?, ?].

2.6 Summary

In the post-Cold War, continued cases of state and military espionage have been joined by economic espionage as a matter of national security. Despite the widespread prevalence of the Internet, and the public attention on external individuals hacking into computer systems, survey after survey finds that the greatest threats come from within organizations.

For information to be useful, it must be accessible by the right people at the right time and place. Insiders take advantage of this by using information they may have a legitimate right to for illegitimate reasons. Furthermore, the Information Age with its increased portability, connectivity, and mobility has significantly increased the effectiveness of the insider [?]. In the overwhelming majority of cases, no Ameri-

can citizen convicted of espionage applied for the job with the intention of committing espionage. Events occurred after being employed in a sensitive position that precipitated the actions.

While initial personnel reviews uncover factors that may increase the possibilities of problems in the future, it is the ongoing observations that are critical to preventing not only the commission of acts of sabotage and theft but the creation of insiders. First rate management can discover subordinates in crisis before these personal crises escalate into state crises. By recognizing warning signs, showing their genuine concern and acting on it to provide assistance programs to people in crisis, management has the opportunity to greatly diminish the insider threat. And in cases where people don't respond to these assistance programs, they can be removed from positions where they can cause harm.

Unfortunately, with the increased focus on lean management, there is less time for managers to get to know their people. Instead, there is a greater need to focus on automated techniques to highlight those areas where greater attention may be needed. In today's organizations, email is one of the best indicators of a person's interests and the communities he is a member of. By analyzing it an interest profile is created and if it matches established risk profiles, an alert is generated.

In order to generate these profiles, first the email activity needs categorization. The type of clustering that has shown the greatest promise with text is probabilistic clustering. By incorporating the existence of underlying topics, probabilistic clustering has shown great promise in categorizing text documents. PLSI and Author Topic are two models based on probabilistic clustering. By applying these models to email activity reasonable clusters are developed and from them profiles are generated. Once these profiles have been created, social networks are then used to illustrate common interests and to predict the existence of unseen relationships, as discussed in the following chapters.

Privacy concerns prevent the use of real-world datasets. However, with the public disclosure of Enron's email, researchers have gained a unique source of email data for analysis. Specifically, for this research, it is hoped that whistleblower Sherron Watkins will emerge as a potential insider.

III. Methodology

With 87 percent of all theft attributed to insiders, the need to identify potential insider thieves before they steal is critical. Furthermore, given that the current business model puts more emphasis on working managers, there is a definite need for developing automated tools to at least give managers a direction of where to focus their time. What is needed is to determine which techniques are most effective at detecting potential insider threats. Two techniques that have shown promise at extracting information from documents are Probabilistic Latent Semantic Indexing (PLSI) and Author Topic. By applying them to the Enron email corpus and determining their usefulness at extracting potential insider threats, it is possible to determine how effective they are. They will be successful if they can (1) generate clear topics and (2) find individuals who have clandestine interests in those topics. First, individuals who emerge as hiding their interests in sensitive topics show the potential to be insider threats. Second, individuals who tend not share their interests with people within the organization may feel alienated and have a reduced sense of loyalty to the company. These individuals also have a greater potential to become insider threats.

This chapter begins by covering the specific research question being tested, followed by the evaluation metrics used to determine whether or not the question has been answered affirmatively. Next, the processing of the data is discussed including the data itself as well as its preparation, clustering, and analysis. The chapter concludes with two additional experiments that are run to see if any additional information can be uncovered.

3.1 Research Questions

The original goal of this thesis was to test “Are probabilistic clustering and social networking techniques applied to email and internet activity effective at detecting potential insider threats?”. While this is the question with which the thesis began, the scope of the thesis has been trimmed. The final research question is, “Are probabilistic clustering and social networking techniques applied to email effective at detecting

potential insider threats?” The reason for the scope change is discussed in the final chapter.

3.2 Evaluation Metrics

In order to consider these probabilistic clustering and social networking techniques useful, they must be valid, useable, and timely [?]. In order for the techniques to be valid, they must reveal either potential or actual insiders. This metric is a difficult one to test because there are few real world email datasets where insiders are known.

In the case of Enron, the only known potential insider is Sherron Watkins. Therefore, the first validity metric is whether or not the techniques reveal Sherron Watkins as an insider. This is accomplished in two steps. First, the topic which is most connected to the off-book partnerships is established. Next, the individuals with a clandestine interest in that topic are determined. For the technique to be valid, Sherron Watkins must emerge as being one of those individuals. This same process is then repeated for the topic which is most connected to socializing. It would be ideal if it were possible to establish that other Enron insiders shared similar interests with her. Unfortunately, since there are no other known insiders, this is not possible. Instead, to test whether or not these techniques connect people who should have similar interests, two additional tests are performed. The first test checks if Kenneth Lay (chairman of Enron), Jeffrey Skilling (Enron’s CEO), and Andrew Fastow (Enron’s CFO and a manager of the off-the-book partnerships) share common interests. The second test checks if “LJM” and “raptor” emerge in a category that appears high for Andy Fastow and/ or Michael Kopper (a manager of the off-the-book partnerships).

To test useability, several things must occur. First, the categories that emerge must be understandable, i.e. by looking at the most significant words that make a topic, a general sense of the topic must emerge. Second, the documents that the techniques consider most representative of a topic must actually be representative of that topic. Similarly, the individuals that the techniques consider most associated

with a topic must also actually be representative of that topic. While measuring each of these useability metrics is much more subjective than measuring the two validity metrics, it is possible to make a qualitative analysis of whether or not these techniques emerge with results that “make sense”. In addition to these three metrics, a manageable number of people with clandestine interests must also emerge. If no individuals emerge, then either the technique is running on data from the perfect organization or potential insiders are escaping detection. On the other hand, if ten percent of the employees emerge as potential insiders, the huge number of false positives also makes the technique unusable. Therefore, a final useability metric measures success if the percentage of potential insiders extracted is between 0.1% to 1.0% percent of the population.

Finally, getting valid information is not especially helpful, if it takes months to emerge. At the same time, since potential insiders emerge over time, producing results in hours is not necessary. The techniques are timely if they complete in a time measured in days. Specifically, if the techniques complete within 7 to 10 days, they are considered useful.

3.3 Data Processing

Now that the research question and the metrics have been discussed, the next subject to review is the data that the techniques will be tested on. In 2003, as part of the investigation into Enron, the Federal Energy Regulatory Commission made public Enron’s senior management’s email activity over a period of nine months. In addition to its value in the prosecution of the case against Enron’s senior management, this data has become a touchstone of research into email datamining techniques. The data was originally stored as scanned images as well as .pdf files. The data was purchased by Leslie Kaelbling at MIT who, along with several people at SRI, cleaned up several data integrity issues. After this cleanup, William Cohen at Carnegie Mellon University received the data and organized it into a 400 MB tar file consisting of 500,000 messages from the folders of 150 Enron employees. Andres Corrada-Emmanuel at the University

of Massachusetts performed a hash test on this data and determined that the data actually contained two copies of every message. The newly cleaned data was then retrieved by Jitesh Shetty at UC Berkeley who organized the new, reduced collection of approximately 250,000 messages into a MySQL database. This is the database used for this thesis.

3.3.1 Database. The database consists of the email messages and some files for categorization. Since the categorization files assume a fixed number of categories and are only populated with a small amount of test data, they are excluded. The only files used are the ones directly related to the email messages. This includes a Messages file made up of a unique messageid (MessageID), the date (MessageDt) and time sent (MessageTz), the subject (Subject) and the id of the sender (SenderID); a headers file consisting of multiple records for each header field of the message; and a bodies file consisting of a single record containing the text of the email.

In addition, each sender and recipient is assigned a unique PersonID which is stored in the People file. This PersonID along with the Recipients file allows accessing the messages from either the recipient or the sender (see Figure ??). Finally, a mailgraph file summarizes the number of correspondences between senders and recipients.

While the headers file contains valuable SMTP-type information, the only information used for the clustering algorithms is contained in the People, Recipients, and Messages tables. Furthermore, since this information is in a more concise form with only one record per message, the information is processed much faster by excluding the headers file.

In addition to these files, there are several more files needed to perform the clustering. In short, files are needed to capture the relationship between words and documents. The minimum files needed to perform this analysis concisely are a words file (Dictionary or Words) that creates a unique identifier for each word and tracks the number of times the word appears in the corpus; a second file that tracks how many

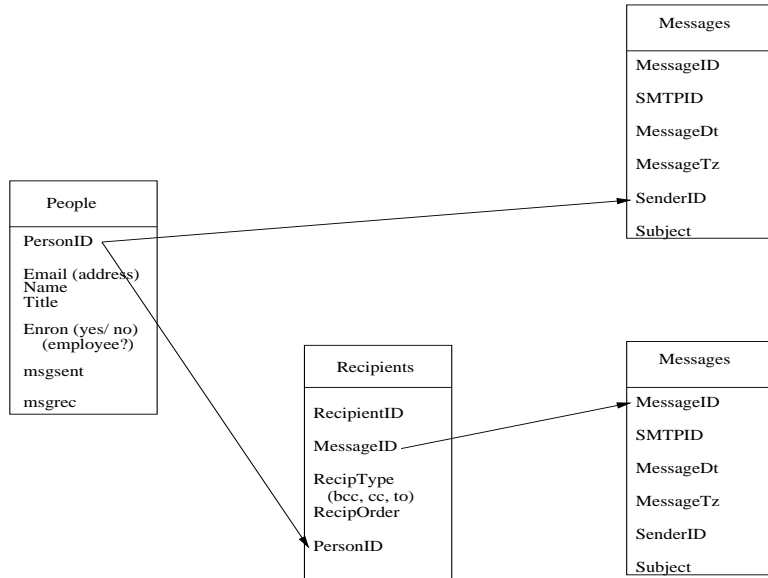


Figure 3.1: Relating People to Messages in the Enron Corpus

words are in each document in order to determine the relative frequency of words in a document (MessageDicts or MessageWords). Observe that words either are created while processing the data files (represented as a collection of letters separated by non-letters constitutes a word) or exist a priori in a dictionary and words in the documents are validated against this list before being added to the database.

3.3.2 Preparation. The process of populating the supplemental files is fairly straightforward. At an abstract level, all that is required is to look at the body of each message, count the number of times each word appears, and load this information in the MessageDicts (MessageWords) file. In addition, by the time the entire corpus is read, the number of times each word appears is stored in the Dictionary (Words) file.

What makes this process a little more difficult is determining what constitutes a word. In previous work, words were defined strictly as collections of letters. This means that phone numbers and dates are not stored. In addition, since email addresses contain non-alphabetic characters, they also are not considered words. While, with a looser definition, these objects can certainly be considered words and although they would serve as excellent discriminatory tools for clustering like messages, this

thesis continues along the same lines of previous research by restricting words to containing only letters. Each time a non-alphabetic character occurs, the current word is considered complete. This approach was selected to decrease the number of words in the corpus so that (1) the algorithms would perform faster and (2) the smaller word space would result in more general clusters. Furthermore, a dictionary [?] was retrieved from the DEC Systems Research Center containing over 104,000 words. The proposed words from the corpus are checked against the dictionary before placement in the database.

A second technique used to decrease the size of the vocabulary is stemming. Stemming removes suffixes from words. For example the words *bake*, *bakes*, *baker*, *baking*, and *baked* all refer to the same thing. Rather than having 5 distinct words in the vocabulary (and consequently more problems clustering), the stem “*bak*” is used for all of them. While this does tend to make some words more difficult to understand, it is expected that the most prominent words in the clusters are complex enough that stemming doesn’t prevent understanding their meaning. The code to perform the stemming was written in C by Martin Porter [?].

Finally, the last technique used to improve the algorithms performance is the removal of the most common words in the English language. Words like “*the*” and “*of*” (these two words alone account for 9% of all written words) are not helpful in developing clusters. In addition, it has been shown [?, ?] that languages follow Zipf’s Law. The frequency distribution of words can be described using an inverse exponential curve. There are a few (approximately 500) *common* words, that appear with very high probability. There are more *average* words (approximately 5,000), that appear with a reasonable frequency. Finally, most words (approximately 50,000) in a language are *rare* and appear with a very small frequency (figure ??). By removing the most common words, the most all-inclusive clusters are removed and more realistic clusters emerge. The list of stop words was retrieved from University of Neuchatel [?].

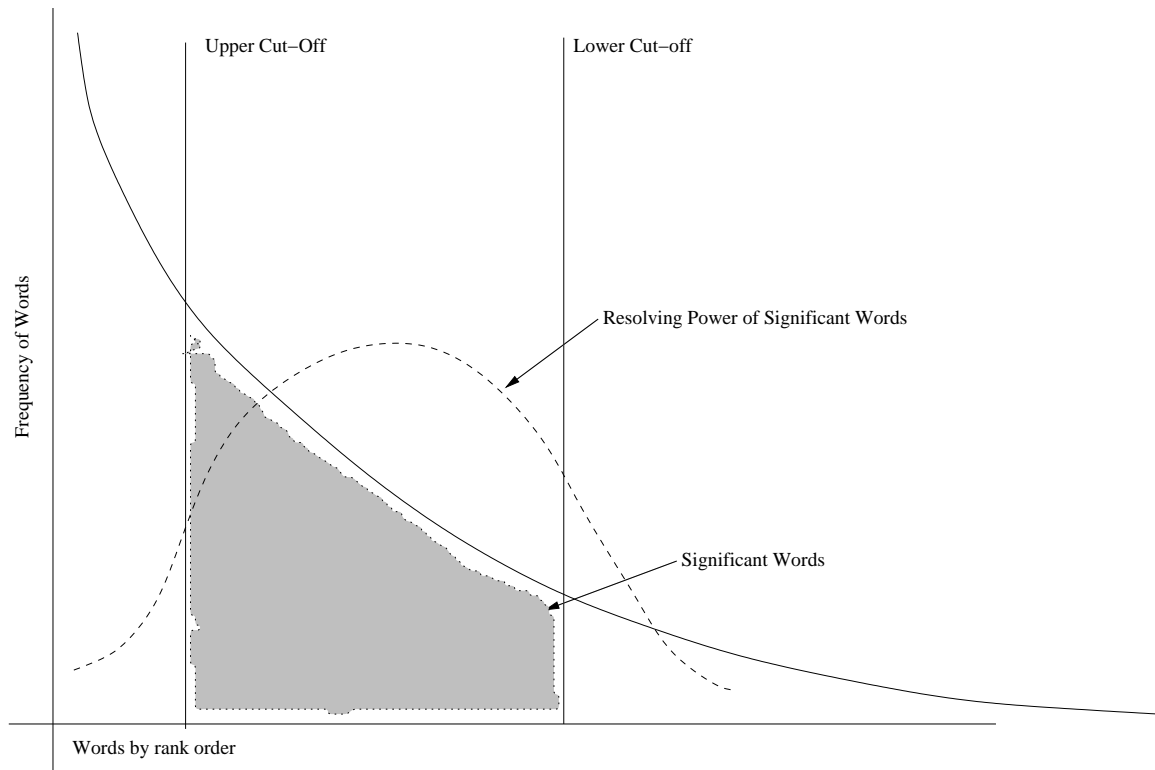


Figure 3.2: Words follow an inverse power law distribution [?]

While it is not the purpose of this research to perform clustering by using tf-idf (term frequency-inverse document frequency) [?], it is possible to use the information extracted to perform this analysis without any additional effort simply by assuming multiple copies of each document corresponding to the number of people associated with it.

3.3.3 Clustering. For PLSI-U, the code was written in C from scratch using the description provided by Hoffman for PLSI [?] and the expectation maximization algorithm described by Moon [?] along with the extensions described in Section ?? to include users. The input parameters include the number of categories, the value of the mean square error that constitutes convergence, and the maximum number of iterations. The number of categories is set to 48 for software and hardware reasons. The maximum number of iterations is initially chosen to be 80. There is no reason to assume that 80 iterations is sufficient. However, after running the algorithm, the data

consistently converged prior to 80 iterations. As a result, 80 is selected as a sufficient number of iterations. One problem with PLSI-U is its long run-time. For instance, the Enron corpus has 254,904 documents, 87,395 users, and using a stemmed dictionary 54,147 words. Combining this with 48 categories, and noting that 4 calculations are performed, means each iteration includes approximately 2×10^{17} calculations. In addition, since every conditional probability ($p(u|z), p(w|z), p(d|z), p(z|d, w, u), p(z)$) needs to be stored, the memory requirements are similarly staggering. Several techniques are used to decrease both the memory and processing requirements. First, by keeping track of the old and new conditional probabilities, it is possible to avoid storing $p(z|d, u, w)$ entirely. The resulting memory savings more than offset the additional processing requirements. In addition, by using the Enron documents to determine what words and users need updating (instead of updating conditional probabilities for which there is no data), the processing requirements are also significantly reduced. Finally, by arranging the categories as the outermost loop, it is possible to *parallelize* the program and process each category for a given iteration in parallel, updating the values between parallel processes before proceeding to the next iteration (needed for the denominators which sum across categories). The need for parallel processing drives the determination of 48 categories. The tests are run on a server farm where only a maximum of sixteen processors are available to a single user. Furthermore, due to the memory allocation on the shared servers, running more than three processes on any server causes the application to crash. As a result, $16 \times 3 = 48$ categories are chosen.

While the Latent Dirichlet Allocation (LDA) code is available for public use from [?], the Author Topic (AT) extensions to it are not. Furthermore, since Rosen-Zvi, et al. suggest using Gibb's sampling while Blei's LDA code used variational inference, the C code for AT was generated from scratch. The inputted parameters include the number of topics and the number of iterations needed for convergence. The number of topics is chosen as 48 to match the PLSI-U tests. The number of iterations selected is 2000 based on Rosen-Zvi's results [?]. The output consists of

two files. One file includes the number of times each word is assigned to each topic and the other file includes the number of times each user is paired with a particular topic. These two pieces of data are sufficient to calculate the desired probabilities.

In order to perform the clustering, the data is extracted from the Enron database and placed into a textfile (Figure ?? A). The textfile has one line for each message and each line is composed of the number of words in the email message. This is followed by the number of people connected to the message. Finally, the line concludes with a list of the identification numbers of the individuals followed by a list of wordId:frequency pairs. From this data, the clustering programs are able to extract clusters and output (1) the probability that a user is in a particular category, (2) a document is in a particular category, (3) a word occurs in a particular document, and (4) a category occurs. They do this iteratively; AT creates a file every 100 iterations, while PLSI-U creates a file every iteration resulting in 20 and 80 files respectively (Figure ?? B). The values in the final iterations file are then loaded back into the database for later analysis (figure ?? C).

The final step is to normalize these probabilities (Figure ?? D). While the conditional probabilities (e.g. $p(d|z)$) were normalized within a category during the PLSI-U iterations, they now are normalized for a given document. In other words, the probability that email message 1 is topic 1, topic 2, ..., topic 48 must sum to 1 (e.g. $p(z|d)$ must be normalized). By Bayes' Rule:

$$p(z|u) = \frac{p(u|z)p(z)}{p(u)} \quad (3.1)$$

Since $p(u|z)$ and $p(z)$ are readily available and $p(u)$ is considered uniform (i.e. the probability of any user being selected is $1 / (\text{Number of Users})$), this calculation is easily performed. A similar calculation is performed to arrive at $p(z|w)$ and $p(z|d)$. Once these calculations are accomplished, it is now possible to observe the social networks that result.

3.3.4 *Analysis.* After data clustering, building the social networks is straightforward. First, an implicit network is constructed from the PLSI-U data. Recall, the implicit network is composed of individuals that share an interest in the same topic. If two people both have an interest in a topic that exceeds a threshold, T , a link is created between those two people (if $p(z = Z_1|u = U_1) > T$ and $p(z = Z_1|u = U_2) > T$, then the link U_1U_2 is created for the implicit PLSI network for category Z_1). This process is repeated for every pair of people for every topic creating separate graphs for each topic. Once the PLSI-U implicit network is formed, the same process using the AT data is used to create an AT implicit network (Figure ?? E).

Once the implicit networks are formed, a final process creates an explicit network based on email data. The explicit network is composed of edges between individuals who have emailed one another. If there is at least one email message for a specific topic between two people, a link is created between them. Mathematically, if $p(z = Z_1|d = D_1) > T$, then $\forall U_1 \in D_1 \forall U_2 \in D_1$ the link U_1U_2 is created for the implicit AT network for category Z_1 . This process is repeated for every topic and every pair of individuals.

What is unclear is how to determine if an email contains the relevant topic. There are several possible ways to do this. One is to set a probability threshold. For example any email or user where the probability of the relevant category given that email or user is above 5% is considered to contain an interest in that topic. However, this gives more weight to popular topics. For instance, the Author Topic algorithm results in one category (Category 0) having a probability for most users of over 90%. Using a straight percentage cutoff, most individuals would only have an interest in that one topic. A second option is to just take the top three categories for all users, regardless of percentages. However, it is easy to construct examples where this also is inappropriate. What seems most appropriate is to take the average probability of a category and set the threshold at a certain number of standard deviations above it. For instance, if, on average, individuals have a 0.5% probability of being interested in category 43 with a standard deviation of 0.7% and someone has a 1.7% probability of

picking that topic, then (assuming a Gaussian distribution) that individual has over a 95% chance of being interested in that topic. For this experiment, 1.64 standard deviations above the mean (resulting in probabilities of at least 95%) is used to determine interest in a topic for both the explicit email network and the implicit interest network.

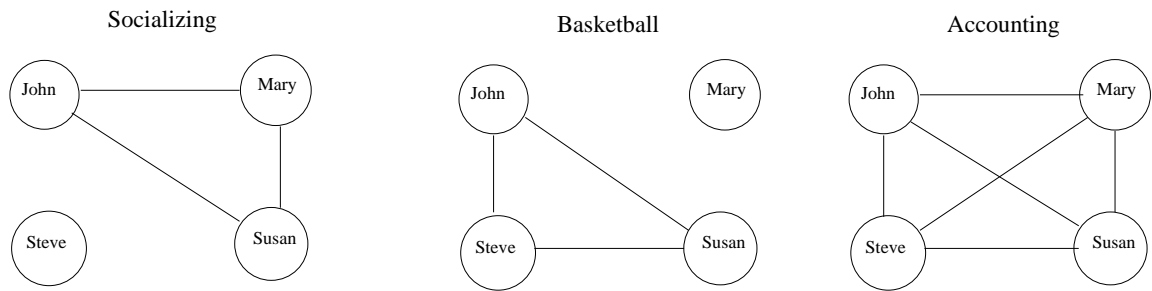
The final step is to examine the implicit and explicit social networks. Each implicit network is compared to the explicit network in turn. If a person has an interest in a topic demonstrated by the presence of links between that person and others in the implicit network but has no links to anyone in the explicit network for that topic, an exception is generated. For instance if John, Mary, Steve, and Diane have an interest in dancing but John has neither sent nor received an email from any of them about dancing, an exception is generated for John and dancing indicating that John has a hidden interest and could be feeling alienated. This process is repeated for every person and for every topic. Figure ?? shows this graphically.

3.4 Additional Experiments

Two additional experiments are discussed in Chapter ?. The first concerns the possibility of extracting additional topics. As observed in Section ?, due to hardware and software constraints, the number of topics that can be clustered is limited to 48. However, limiting the number of topics to 48 may lump several topics into one meta-topic. Therefore, once the data clustering and analysis is performed, additional analysis is performed to extract documents and individuals that appear most probable for a specific topic and perform a second round of probabilistic clustering on them to achieve a finer level of granularity. The ability to quickly drill down and produce a finer level of granularity would overcome the traditional criticism that a limited number of topics can't produce useful topics.

The second concerns the possibility of calculating traditional social network analysis (SNA) metrics from the social networks that are created and using them to

IMPLICIT INTEREST NETWORKS



EXPLICIT INTEREST NETWORKS

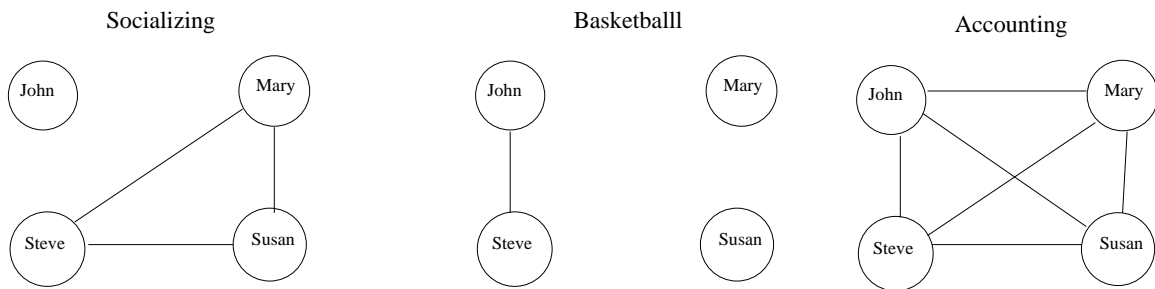


Figure 3.3: An Example of Clandestine Interests (implicit network = external; internal email explicit network = internal email only)

aid in the generation of additional potential insider threats. Once the networks are developed there are several additional questions to answer:

- What are the values of various social network centrality indicators including degree, closeness and betweenness? Are the individuals with high centrality measurements for a given topic “appropriate” for that topic. Said differently, how do centrality measurements compare with the conditional probability of an individuals given a topic? Traditional SNA centrality measurements would provide an additional validation to researchers that individuals who show up both via probabilistic clustering and SNA are worthy of additional attention.
- Can anything about people’s positions be inferred by their email activity? For instance:

- Are there individuals who have more emails sent to and received from people outside of the company rather than inside the company. This may indicate sales people or purchasing agents. At Enron, this may also have indicated people involved in investor relations or public relations.
- Are there some people who appear to have a lot of interests. This may indicate administrative assistants.

If it is found that certain SNA measurements suggest certain job responsibilities, then finding people with those SNA measurements that do not have those job responsibilities may provide another indicator of a potential insider threat.

- Can anything be inferred by comparing the individual explicit networks for each category to a consolidated explicit network composed of all categories? Said differently, what is the relationship between the individual explicit networks and a network that links people if an email passed between them regardless of topic?
- Finally, is it possible to follow a topic as it progresses through the organization? Topics small enough in size to perform this analysis may include Jeff Skilling's resignation and Sherron Watkins' letter to Ken Lay. Such an analysis on a topic of interest may provide an additional means of finding either potential or actual collaborators.

3.5 Summary

In order to find potential insiders, it is first necessary to get to know them. Ideally this is done personally. However, in today's workplace, this may not be possible. The next best thing is to do it electronically. A good way to get to know someone is to learn about their interests and in this new information age, email is one of the best sources of electronic information. By using probabilistic clustering and social networking techniques, specifically Author Topic and PLSI-U, to datamine email, a person's topics of interest emerge. There are several possible uses for these topics

of interests to generate insider threat leads. The first way is to extract people as potential insiders who have an interest in a specific topic but are neither the recipient nor the sender of any internal email that contains that interest. A second way is by finding those people who have an interest in a topic that is itself an indicator of potential insiders (e.g. money, overseas loyalties, substance abuse, etc.). Finally, a third way, not explored in this work, is to track a person's topics of interest over time and watch for any radical changes.

To test the effectiveness of Author Topic and PLSI-U for detecting potential insider threats through email traffic, the Enron email corpus is broken down into a collection of word frequencies and user assignments. These datapoints are then input into Author Topic and PLSI-U to extract topics of interest. These topics of interest are then applied to the emails to find: 1) the words that describe the topics, 2) the documents and people that are the most representative of the topics, 3) and the topics that best describe each individual. Finally, these are reviewed for the potential insiders. The results of these experiments are described in the next chapter.

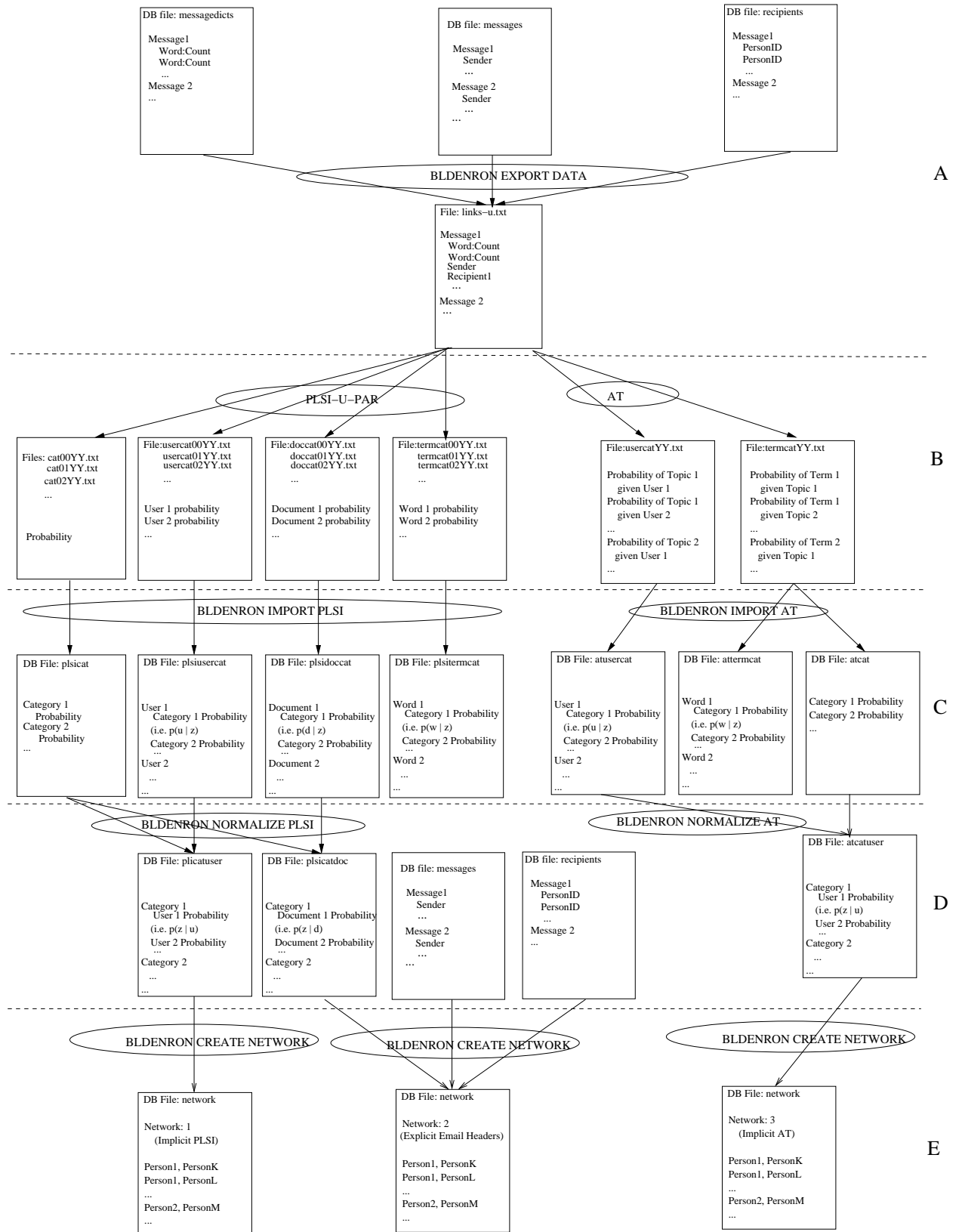


Figure 3.4: Data Clustering Process Flow

IV. Results and Analysis

As discussed previously, the goal of this research is to determine the applicability of probabilistic clustering and social networking techniques applied to email for detecting potential insider threats. This is determined by measuring the following metrics:

1. Useability

- (a) Do the algorithms create clusters of words which provide a clear definition of the topic? An algorithm performs perfectly if all of the most probable words describe a single topic. The greater the number of outliers, words that do not describe the topic described by a majority of words, the poorer the algorithm performed. If no majority of words described a single topic, the algorithm performs poorly.
- (b) Do the algorithms create clusters of individuals which match the topic definitions described by the words? The more individuals who had a role in the topic, the better the algorithm has performed. For instance, if all of the most probable individuals in the California Crisis topic are in the Government Relations group, the algorithm is considered to have performed well.
- (c) Do the algorithms create clusters of documents which match the topic definitions described by the words? In the same manner as the two previous metrics, if the documents are clearly about the topic, the algorithm is considered to have performed well.
- (d) Do the algorithms create a manageable number of individuals with clandestine interests? An individual is considered to have a clandestine interest if he has an interest in a topic but does not send any emails within the company about that topic. The number of individuals is considered manageable if it is between 0.1% and 1.0% of the population.

2. Timeliness - Do the algorithms find potential insider threats in a reasonable period of time (7 - 10 days)?

3. Validity

- (a) Does the analysis reveal Sherron Watkins, the famed Enron whistleblower, as an insider?
- (b) Does the analysis show that the techniques are successful at finding collaborators of known insiders by examining common interests? The first test checks if Ken Lay (Enron’s chairman), Jeff Skilling (Enron’s CEO during most of Enron’s questionable business activities), and Andy Fastow (Enron’s CFO responsible for the off-the-book partnerships that contributed to Enron’s downfall) have common interests? The second test check if Andy Fastow and Michael Kopper (the two principle managers of the off-the-book partnerships) both have a strong interest in the category about “Raptors” (the principle transactions of the off-the-book partnerships)?

After the results of these metrics are discussed, the chapter concludes with the results of the two additional experiments. First, the results of the Two-Tiered Approach are discussed. Finally, the results of the traditional Social Network Analysis (SNA) analysis is reviewed.

The experiment is broken down into eight sub-experiments. The first split is between words in the dictionary and all alphabetic “words” whether or not they are in the dictionary. The second split is between all words and only stemmed words. The final split divided these four datasets between Author Topic and PLSI-U. After running the tests on the stemmed words, it is clear that stemming does not make the identification of words more difficult. In addition, increasing the size of the corpus by not requiring stemmed words, increases the size of the joint probability distribution and makes it more difficult/ time-consuming for the algorithms to converge. As a result, the experiment only checks Author Topic and PLSI-U against stemmed dictionary words and all stemmed words.

The first two experiments are run against the stemmed dictionary word dataset while the third and fourth are run against the stemmed word (no dictionary) dataset.

After the algorithms completes, the clusters are examined to see if they “made sense”. One additional concern when evaluating topics based on the most probable words must be considered. Despite having already eliminated stop words, there are other words that end up appearing in most if not all categories. For instance, in the Enron corpus, power is a very common word (as is Enron). To avoid this problem, when printing the words that describe a category, only words that appeared in at most five categories are included.

One issue that surfaces quickly with the Author Topic data is the existence of a category that most individuals and emails have an overwhelming probability for. When the category is investigated, it emerges as one containing words like meeting, regarding, discuss, issue, and schedule, words that are common to almost any business email (Figure ??). As a result, this category is removed from any subsequent processing and the conditional probabilities for the remaining categories are adjusted to pretend the category does not exist.

For brevity, only four topics are discussed per experiment (Figures ??, ??, ?? and ??). The first topic (Senior Mgmt) is about the senior Enron executives. It is the category that clustered Ken Lay, Jeff Skilling, Greg Whalley, and Jeff McMahon. The second topic (California Crisis) best describes the California energy crisis. The third topic (Research) focuses on the research performed by Vince Kaminski and his group. Finally, the fourth common topic (Info Technology) concerns IT systems and help desks. In addition, Figure ?? shows several additional categories (Scheduling, Sept 11 Attacks, Personal Email, Enron Crisis, Fantasy Football). Although the descriptions of these topics are inferred, they are listed in the the figures for ease of understanding.

4.1 Useability Metrics

In order to be useful, probabilistic clustering and social networking techniques must generate topics that are easily identifiable. For instance when attempting to find insiders, although it is some help to know that John and Mary share in an interest in

topic 17, it is a much greater help if it is known that topic 17 is about the organization’s chief competitor. Therefore, the first three useability metrics examine when Author Topic and PLSI-U create when clustered topics. First, this is checked by looking at the most probable words; second it is checked by the most probable individuals; and third it is checked by the most probable documents. The final useability metric checks the number of individuals that emerge as having clandestine interests. Too many makes the results unusable while too few mean that actual insiders may be overlooked.

Senior Mgmt		California Crisis		Research		Info Technology	
CATEGORY 45		CATEGORY 2		CATEGORY 47		CATEGORY 40	
Video	0.3%	Assembly	0.2%	Research	0.6%	Unify	1.1%
Boardroom	0.1%	AB	0.2%	Model	0.6%	Directory	1.1%
Sherri	0.1%	Crisis	0.2%	Resume	0.2%	Enterprise	1.0%
Task	0.1%	Deregulation	0.2%	Visit	0.2%	Hardware	0.7%
Safety	0.1%	Urgent	0.2%	University	0.2%	Script	0.5%
Peer	0.1%	Declare	0.2%	Finance	0.2%	Logistic	0.5%
Sera	0.1%	Freeze	0.2%	Rice	0.2%	Stage	0.4%
Palmer	0.1%	Legislature	0.1%	Dear	0.2%	Setup	0.3%
Medium	0.1%	Sold	0.1%	Student	0.1%	Solar	0.3%

Figure 4.1: PLSI-U Sample Categories using Stemmed, Dictionary Words

Senior Mgmt		California Crisis		Research		Info Technology	
CATEGORY 9		CATEGORY 43		CATEGORY 30		CATEGORY 28	
Business	1.4%	Electricity	1.8%	Research	2.3%	Access	2.3%
Skilling	1.4%	Commission	1.7%	dear	1.9%	Service	2.1%
Lay	1.4%	State	1.6%	University	1.7%	User	2.1%
Company	1.3%	Utility	1.5%	School	1.0%	Information	1.6%
Year	1.2%	Energy	1.2%	Model	0.9%	System	1.6%
Ken	1.1%	Public	1.1%	Program	0.9%	Contact	1.6%
Opportunity	1.1%	Legislature	1.1%	Resume	0.9%	Manage	1.4%
President	1.0%	Regulatory	1.0%	Interest	0.9%	Server	1.4%
Chairman	1.0%	Senate	0.9%	Finance	0.9%	Thi	1.4%

Figure 4.2: Author Topic Sample Categories using Stemmed, Dictionary Words

4.1.1 Describing Topics with Words. The first metric considered is the most probable words. The topics by most probable words are in Figures ??, ??, ??, ??, and ?. Although complete words are shown, they are extrapolated from the word stems actually produced. Despite initial concerns that stemming might make some of the words difficult to determine (e.g. trying to determine the original word family that stemmed to ‘thi’), the stemmed words that distinguish categories prove easy to identify.

Senior Mgmt		California Crisis		Research		Info Technology	
CATEGORY 11		CATEGORY 0		CATEGORY 44		CATEGORY 7	
PRC	0.3%	Governor	0.3%	Vinc	1.3%	Unify	0.7%
Video	0.2%	Calpin	0.3%	Kaminski	0.7%	SAP	0.4%
Weekly	0.2%	IEP	0.3%	Research	0.4%	Netco	0.3%
Ken	0.2%	Dasovich	0.3%	Model	0.3%	Sitara	0.3%
Dial	0.2%	Edison	0.3%	Shirley	0.2%	Script	0.3%
Kean	0.2%	Gov	0.2%	Rice	0.2%	Class	0.2%
Cindy	0.2%	IEPA	0.2%	Visit	0.2%	Setup	0.2%
VP	0.1%	Duke	0.2%	Crenshaw	0.2%	Path	0.2%
Passcode	0.1%	Mara	0.2%	University	0.2%	Regan	0.2%

Figure 4.3: PLSI-U Sample Categories using Stemmed Words without a Dictionary

Senior Mgmt		California Crisis		Research		Info Technology	
CATEGORY 8		CATEGORY 18		CATEGORY 25		CATEGORY 10	
Develop	0.7%	Jeff	1.7%	Vince	3.8%	Password	1.9%
Opportunity	0.6%	Dasovich	1.7%	Kaminski	3.2%	User	1.9%
Technology	0.5%	Steff	1.0%	Research	1.0%	Access	1.8%
Base	0.5%	Mara	0.8%	Shirley	0.8%	Account	1.5%
Recent	0.5%	Shapiro	0.7%	VKamin	0.7%	ID	0.9%
Success	0.5%	California	0.7%	Universe	0.6%	Login	0.8%
Generate	0.4%	James	0.7%	Crenshaw	0.6%	Center	0.7%
Experience	0.4%	CA	0.6%	Stinson	0.5%	Log	0.7%
Lead	0.4%	Gov	0.5%	Model	0.5%	Respond	0.7%

Figure 4.4: Author Topic Sample Categories using Stemmed Words without a Dictionary

While both Author Topic and PLSI-U produce categories that are understable when only stemmed, dictionary words are used, Author Topic's categories are much easier to identify and do not share many words in common with PLSI-U (see Appendix ?? for the listings of 25 words per category per experiment). This does not persist when stemmed words without a dictionary is the dataset. For the second two experiments, the descriptiveness of PLSI-U is at least equal to Author Topic and possibly more descriptive due to its more extensive use of acronyms. Furthermore, Author

Scheduling		Sept 11 Attacks		Personal Email		Enron Crisis		Fantasy Football	
CATEGORY 0		CATEGORY 27		CATEGORY 22		CATEGORY 38		CATEGORY 41	
Thi	1.3%	Fool	0.2%	Yahoo	2.1%	Fund	0.6%	Joe	0.5%
Subject	1.2%	Attack	0.1%	Mime	1.1%	Consumer	0.5%	League	0.5%
Please	1.1%	Federal	0.1%	Hotmail	1.0%	Stock	0.4%	Fantasy	0.5%
Are	1.1%	Bush	0.1%	Version	0.9%	Lay	0.4%	Team	0.5%
Any	0.8%	Sector	0.1%	Mailer	0.8%	Ken	0.4%	Football	0.5%
Forward	0.6%	Assembly	0.1%	Type	0.7%	Donation	0.4%	Game	0.5%
Call	0.5%	Pressure	0.1%	MSN	0.7%	Retire	0.4%	Commission	0.4%
Attach	0.5%	Flight	0.1%	Return	0.7%	Bankrupty	0.4%	Season	0.4%
Time	0.5%	Terrorist	0.1%	SMTP	0.7%	Declare	0.3%	Player	0.4%

Author Topic w/ Dictionary PLSI-U w/ Dictionary Author Topic No Dictionary PLSI-U No Dictionary Author Topic No Dictionary

Figure 4.5: Additional Categories from Author Topic and PLSI-U by most probable words

Topic does not seem to change much in the descriptive words between dictionary words and all words. Over half (29) of the categories are easily cross-referenced between the stemmed dictionary words and stemmed words (no dictionary) experiments for Author Topic.

In examining the topic Senior Mgmt, recall that it is determined by looking at the preferred topics of Ken Lay, Jeff Skilling and Andy Fastow and selecting the one that is most common to all of them. While there is only a little overlap in the words between the four experiments (Figures ?? - ??), all four clearly give the reader the sense of a senior management topic. This is more true when one knows something about the individuals employed at Enron. Sherri Sera was Jeff Skilling personal assistant. Mark Palmer was head of Corporate Communications. Cindy Olson was head of Human Resources and Steven Kean was the Chief of Staff. All of these individuals were involved in the weekly management meetings. It is interesting to observe that although the first two experiments were restricted to words in the dictionary, some names seeped through if their stemmed base was the same as the stemmed base of a word in the dictionary. Examples of this include Ken (ken) Lay (lay), Skilling (skill), Sherri (sherry) Sera (sera - plural of serum), and Cindy (cinder). In light of this, it emerges that (at least for the first category), there is not much difference in the quality of results between experiments where the words are restricted to the dictionary and experiments where they are not.

Unlike the Senior Mgmt topic, the California Crisis topic emerges strictly by examining the most probable words. This is the first topic where one observes a significant difference in the quality of results between experiments. Interestingly, the highest quality results come from two experiments that are the most different, the Author Topic run restricted to words in the dictionary and the PLSI-U run using all words. Here is where one sees the strengths and weaknesses of not restricting words to the dictionary. When the Author Topic experiment is run restricted to words in the dictionary, a topic clearly emerges. However, when it is run again and includes non-dictionary words, people's names dilute the descriptiveness of the topic. On the

other hand, PLSI-U with only dictionary words barely provides enough information to provide a topic description. But when acronyms are allowed, PLSI-U adds significant descriptiveness by adding the California Planners Information Network (CALPIN), the Independent Energy Producers Association (IEPA) and Jeff Dasovich who was Enron's representative in charge of government affairs in California.

The Research topic at first glance appears to show a mingling of two topics, one of research within Enron and the second involving universities (possibly recruiting). However, upon closer examination, it emerges that Vince Kaminski, head of the Research Group, had a close relationship with the faculty at Rice University (and is currently an adjunct professor there). He and several of his employees often spoke there and/or invited classes to Enron for research projects. As a result, the topic is clearly about Enron's Research Group. Here again, names and email addresses emerge within the most probable words. In addition to Dr. Kaminski, Shirley Crenshaw was the administrative coordinator for the Research Group and Stinson Gibner was a Vice President in the group. With this topic, all of the experiments produce quality results that allow a topic description to easily emerge.

The Information Technology topic is the last that is consistent throughout all of the experiments. While some of the words such as directory, hardware, user, and access are clearly IT-type words, others are not. However, many of the other words are actually the names of computer programs run at Enron. SAP was their Enterprise Resource Planner. Unify was the deal settlement processing system and Sitara was the system used to complete physical gas deals. While all of the experiments produce good results, the experiments not restricted to words in the dictionary appear to perform better since they are able to include the names of software packages. Unlike the previous two topics, only one individual's name appears (Regan Smith was a network administrator) resulting in a poorer quality of results.

Finally, additional categories provide examples of other topics that emerge from these algorithms. Although no pornographic categories emerge, Author Topic extracts

a fantasy football league that appeared fairly active at Enron. There are well over 7,000 internal emails sent between people at Enron on the topic of Fantasy Football. While PLSI-U never extracts a recognizable topic on fantasy football, it does extract a recognizable topic on the September 11 attacks and one on the last days of Enron prior to the bankruptcy. An additional topic that emerges is focused on AOL, Yahoo, and Hotmail, suggesting that many individuals may have accessed their personal emails while at work. Although nothing emerges to suggest industrial espionage, the use of personal emails would have provided an excellent mechanism for moving secrets out of the company.

Both PLSI-U and Author Topic clearly succeed at the useability metric for most probable words. Both PLSI-U and Author Topic work well at extracting coherent topics from the email corpus. In three of the four experiments (the exception being PLSI-U run on only words in the dictionary), it is easy to describe the topics based solely on the twenty-five most probable words for that topic.

4.1.2 Describing Topics with Users. The topics by most probable individuals are in Figures ??, ??, ??, and ?. The first challenge is determining individuals' positions within Enron since it is not contained within the corpus. Luckily, there is a copious amount of information both on the Internet [?, ?, ?, ?] and in books [?, ?] on the rise and fall of Enron. As a result, despite a lack of job information in the actual corpus, it is still possible to determine what many individuals' positions were. One thing that emerges is the number of different email addresses possessed by certain individuals. Vince Kaminski, Managing Director and Head of Research, had at least five. Ken Lay, Chairman, had five personal email addresses and at least as many titular email addresses (e.g. Enron Office of the Chairman, Ken Lay - Office of the Chairman, etc.). This degrades the quality of results in the 10 most probable individuals. Consider the PLSI-U without a dictionary experiment. Vince Kaminski appears in the Research category four times in the top ten individuals under different email addresses for a total probability of 30.04%. For identifying or connecting topics with

CATEGORY 45 SENIOR MGMT			CATEGORY 2 CALIFORNIA CRISIS		
Steven Kean	Chief of Staff, Government Relations Specialist	5.8%	Ken Lay	Chairman of Enron	7.6%
Stanley Horton	Chief Executive – Enron Transportation Group	4.0%	Karen Denne	Vice President of Public Relations	6.7%
Steven Kean	Chief of Staff – Government Relations Specialist	2.5%	Sandra McCubbin	Director of Government Affairs in California	4.9%
Maureen McVicker		2.4%	Paul Kaufman	Director of Government Affairs	3.9%
Rosalee Fleming	Secretary to Enron Chairman Kenneth Lay	2.1%	Jeff Dasovich	Government Affairs Executive	3.8%
Greg Whalley	President of Enron	1.8%	Harry Kingerski		3.6%
Mark Frevert	Vice-Chairman of Enron	1.6%	Steven Kean	Chief of Staff, Government Relations Specialist	3.3%
Kenneth Lay	Chairman of Enron	1.6%	Mark Palmer	Head of Corporate Communications	3.2%
Cindy Olson	Head of Human Resources	1.5%	Susan Mara	Director of Government Affairs in California	3.1%
Jeff McMahon	Chief Financial Officer of Enron	1.3%	James Steffes	Vice President of Government Affairs	2.8%

CATEGORY 47 RESEARCH			CATEGORY 40 INFO TECHNOLOGY		
Vince Kaminski	Managing Director and Head of Research	34.1%	Lisa Kinsey		1.0%
Jeffrey Shankman	Chief Operating Officer for Global Markets	6.2%	Robert Superty	Enron North America – Director Gas Procurement	1.0%
Shirley Crenshaw	Research Group Administrative Coordinator	5.0%	Patti Sullivan		1.0%
Stinson Gibner	Vice President in Quantitative Research Group	4.0%	Daren Farmer	Logistics Manager	0.8%
Vasant Shanbhogue	Vince Kaminski’s Second in Command	1.8%	Victor Lamadrid		0.8%
Tanya Tamarchenko	Director – Value at Risk	1.5%	Darla Saucier		0.8%
Zimin Lu	Director of Valuation and Trading Analytics Group	1.5%	Kirk Lenart		0.7%
Jennifer Burns		1.4%	Tammy Gilmore		0.7%
Grant Masson	Vice President – Research Group	1.2%	Cora Pendergrass		0.7%
Pinnamaneni Krishnarao	Vice President – Research Group	1.2%	Mark Schrab		0.6%

Figure 4.6: PLSI-U Individuals Most Associated with Sample Categories using Stemmed, Dictionary Words

individuals, PLSI-U appears to produce results at least as good as Author Topic. Of the four categories, only in one case did Author Topic perform slightly better.

It is reasonable that the Senior Mgmt topic produces good results since it is created by looking at specific users. The results from PLSI-U bear this out producing Ken Lay (Chairman), Greg Whalley (CEO), and Jeff McMahon (COO). Also very prominent are Enron’s General Counsel, the head of Human Resources, and the head of Investor Relations. While Author Topic also produced good results, it is lessened by the presence of generic emails (ethink?, All Enron Gas Services, and some lower-level individuals). Overall PLSI-U significantly out-performs Author Topic both when words are restricted to a dictionary and when the words are unrestricted.

PLSI-U also produces superior results for the California Crisis topic. It extracts many individuals recognizably involved in the management of public affairs in California. While Author Topic produces some of these individuals as well, there are many individuals that are unidentified. By looking at their emails, they do appear to

CATEGORY 9			SENIOR MGMT	CATEGORY 43			CALIFORNIA CRISIS
Michael Horning			0.05%	'Kevin Fulton			0.1%
Jeff McMahon	Chief Operating Officer		0.05%	Eric Letke	Enron Energy Services		0.1%
Anthony Duenner	Senior Vice Pres Global Assets & Services		0.05%	snovose			0.1%
ethink			0.05%	Robert Frank	State Government Affairs		0.1%
Mitch Meyer			0.05%	Hap Boyd	Enron Wind Corporation		0.1%
All Enron Worldwide			0.05%	.sue			0.1%
Matthew Scrimshaw			0.04%	Tamara Johnson			0.1%
Nate Ellis	Director Enron Energy Services		0.04%	tamara Johnson			0.1%
Mariano Gomez			0.04%	Mark Palmer	Head of Corporate Communications		0.1%
Margaret Carson	Director of Corporate Strategy		0.04%	Becky Merola			0.1%

CATEGORY 30			RESEARCH	CATEGORY 28			INFO TECHNOLOGY
grant Masson	Vice President - Research Group		0.04%	houston.report			0.1%
Kenneth Deng	Manager of Quantitative Research		0.04%	Eric Saibi	Enron Capital & Trading - East Desk		0.1%
Mary Bailey			0.04%	SAP Security			0.1%
Vince Kaminski	Managing Director and Head of Research		0.04%	EES Power Settlements			0.1%
Network Security			0.04%	subscribers@mailman			0.1%
Althea Gordon	Recruiter - Associates/ Analysts Program		0.03%	weatherward@mailman			0.1%
Jason Sokolov	Risk Management Group employee		0.03%	Integrated Solutions Center - I/T Help Desk			0.1%
Lenos Trigeorgis	Risk Management Group employee		0.03%	Jeffrey Jackson			0.1%
Rehman Sharif			0.03%	ISC Systems Notification			0.1%
Nedre Strambler			0.03%	Enron Users			0.1%

Figure 4.7: Author Topic Individuals Most Associated with Sample Categories using Stemmed, Dictionary Words

have been involved in conducting business in California. However, what their exact positions were is unknown. As a result, by examination, it appears that PLSI-U also out-performs Author Topic in this category.

The Research topic differs from the previous two by its limited nature. This topic is focused on a relatively small group within the Enron corporation. As a result, all of the experiments show excellent results. This is despite a mix of small and large email datasets for the top individuals. This suggests that when attempting to find individuals who all participate in a category, if the category is of limited interest, then the results are excellent.

The last category is Information Technology. This category most clearly indicates the problem of attempting to identify individuals without extensive inside knowledge. Only one I/T professional (Regan Smith) is identified. However, Author Topic also extracts email ids associated with software packages such as SAP, ibuyit

CATEGORY 11			SENIOR MGMT		
James Derrick	General Counsel	3.60%	Jeff Dasovich	Enron Government Affairs Executive	1.46%
Cindy Olson	Head of Human Resources	1.98%	James Wright		1.04%
Kay Chapman	Secretary of Management Committee	1.67%	Richard Sanders	VP and Asst General Counsel for Enron Wholesale	0.88%
Mark Koenig	Executive Vice President of Investor Relations	1.67%	Susan Mara	Director of Government Affairs in California	0.84%
Greg Whalley	President of Enron	1.66%	Scott Stoness		0.83%
Steven Kean	Chief of Staff, Government Relations Specialist	1.57%	Dennis benevides	Director of Green Power for Enron Energy in CA	0.80%
Mark Frevert	Vice-Chairman of Enron	1.54%	Sandra McCubbin	Director of Government Affairs in California	0.80%
Jeffrey McMahon	Chief Financial Officer of Enron	1.46%	Richard Shapiro	VP of Regulatory Affairs & principal DC lobbyist	0.80%
Kenneth Lay	Chairman of Enron	1.25%	James Steffes	Vice President of Government Affairs	0.76%
David Delaney	Enron Energy Services CEO	1.19%	Harry Kingerski		0.76%

CATEGORY 44			RESEARCH		
Vince Kaminski	Managing Director and Head of Research	20.88%	Daren Farmer	Logistics Manager	2.37%
Vince Kaminski	Managing Director and Head of Research	5.30%	Robert Superty	Director – Gas Procurement Enron North America	1.82%
Shirley Crenshaw	Research Group Administrative Coordinator	3.49%	Patti Sullivan		1.54%
Vince Kaminski	Manager Director and Head of Research	2.86%	Victor Lamadrid		1.46%
Stinson Gibner	Vice President in Quantitative Research Group	2.40%	Lisa Kinsey		1.38%
Don Baughman	North America Power trader – East Desk	2.00%	Bryce Baxter		1.20%
Vasant Shanbhogue	Kaminski’s second in command	1.52%	Tammy Jaquetr		1.04%
Zimin Lu	Director of Valuation and Trading Analytics Group	1.05%	Clarissa Garcia		0.97%
Eric Bass	trader	1.03%	Regan Smith	Network Administrator	0.89%
Vince Kaminski	Managing Director and Head of Research	1.00%	Kevin Heal		0.87%

CATEGORY 0			CALIFORNIA CRISIS		
Jeff Dasovich	Enron Government Affairs Executive	1.46%	James Wright		1.04%
Richard Sanders	VP and Asst General Counsel for Enron Wholesale	0.88%	Susan Mara	Director of Government Affairs in California	0.84%
Susan Mara	Director of Government Affairs in California	0.84%	Scott Stoness		0.83%
Scott Stoness		0.83%	Dennis benevides	Director of Green Power for Enron Energy in CA	0.80%
Dennis benevides	Director of Green Power for Enron Energy in CA	0.80%	Sandra McCubbin	Director of Government Affairs in California	0.80%
Sandra McCubbin	Director of Government Affairs in California	0.80%	Richard Shapiro	VP of Regulatory Affairs & principal DC lobbyist	0.80%
Richard Shapiro	VP of Regulatory Affairs & principal DC lobbyist	0.80%	James Steffes	Vice President of Government Affairs	0.76%
James Steffes	Vice President of Government Affairs	0.76%	Harry Kingerski		0.76%
Harry Kingerski		0.76%			

Figure 4.8: PLSI-U Individuals Most Associated with Sample Categories using Stemmed Words without a Dictionary

(an Accounts Payable package), and the Enron help desk. For this category, Author Topic outperforms PLSI-U.

Overall, both PLSI-U and Author Topic performed excellently, consistently producing high numbers of relevant individuals. Whether the words are restricted to the dictionary or not, the quality of results is unchanged. Interestingly, despite the greater simplicity of the PLSI-U generative model, it performs better in extracting relevant individuals.

Both PLSI-U and Author Topic succeed at this useability metric. If the goal is to find additional leads given a topic of interest, both provide very appropriate names associated with the topics, although PLSI-U appears slightly better. Both models produce better results when words are not restricted to the dictionary. The only exception to this is Author Topic’s tendency to include names. However, this is easily remedied by excluding names from the analysis. In general this is also desirable from a privacy perspective.

CATEGORY 8	SENIOR MGMT		CATEGORY 18	CALIFORNIA CRISIS	
Enron Office of the Chairman		0.09%	Deborah Whitehead		0.14%
Jeff McMahon	Chief Operating Officer	0.09%	Scott Stoness		0.13%
Ken Lay	Enron Chairman	0.09%	Leasa Lopez	Lawyer for Enron Energy Services	0.13%
All Enron Gas Services		0.07%	Dave Black		0.13%
Enron Operations		0.07%	Ken Gustafson	Enron Wind Corporation	0.13%
Nate Ellis	Director Enron Energy Services	0.07%	JLewis		0.13%
Barabara Taylor		0.06%	Tamara Johnson		0.13%
All Enron Worldwide		0.06%	James Steffes	Vice President of Government Affairs	0.12%
Office of the Chief Executive		0.06%	Terry Donovan		0.12%
Office of the Chairman		0.06%	Tamara Johnson		0.12%

CATEGORY 25	RESEARCH		CATEGORY 10	INFO TECHNOLOGY	
West Desk Support		0.11%	ipayit	Accounts Payable software	0.14%
Julius Zajda		0.11%	SAP Security		0.14%
Brad Romine	Research Group employee	0.09%	enron.payroll		0.14%
Tom Barkley	Manager in Research Group	0.09%	payroll.enron		0.13%
Adam Brulinski		0.08%	ibuyit.payables	E-Procurement software	0.13%
Eloise Meza	Research Group employee	0.08%	payables.ibuyit	E-Procurement software	0.13%
Jason Sokolov	Risk Management Group employee	0.08%	enron.expertfinder	S/W to find subject matter experts with Enron	0.12%
Kenneth Parkhill	Research Group employee	0.07%	mbx_iscinfra		0.12%
Steve Bigalow	Research Group technical analyst	0.07%	Tahnee Stall		0.11%
Bessik Matchavariani	Manager Enron Broadband Services	0.07%	ic		0.11%

Figure 4.9: Author Topic Individuals Most Associated with Sample Categories using Stemmed Words without a Dictionary

4.1.3 *Describing Topics with Documents.* After considering whether topics make sense considering most probable words and most probable users, a final check is to consider the most probable documents. Unfortunately, the results applied to documents are more mixed. Even when considering the most probable documents, some are clearly related to the topic while some are not. This may be due to the relatively small number of topics. As a result, (1) the topics are very general, containing a mix of different sub-topics and (2) documents are composed of multiple topics. Due to space constraints for this thesis, examples of the most probable documents are not included.

4.1.4 *Individuals with Clandestine Interests.* The last metric for the usability of results is the percentage of individuals with clandestine interests. Unfortunately, a problem with this metric quickly emerges when the results are produced. If a person has six emails and the only time a particular topic occurs is on one external email, the

individual is considered to have a clandestine interest. However, this is most likely not the case and is merely the result of the small dataset for that individual. The average number of emails sent and received for Enron employees is 71. So, by examining the distribution of numbers of emails, it is clear that it follows an exponential distribution with $\beta = 70$. Therefore, by including everyone who received at least 12 emails, 85% of the population is included but the small dataset problem is predominately avoided.

PLSI-U produces a total across all 48 categories of 652 individuals with clandestine interests for the dictionary words dataset and 304 individuals with clandestine interests for the all words dataset (no dictionary). This means on average each category has less than 14 (6) individuals with a clandestine interest. On the other hand, Author Topic produces a total across 47 categories of 3,988 individuals for the dictionary words dataset and 1,593 individuals for the all words (no dictionary) dataset. This means, on average, each category has slightly more than 84 (34) individuals with clandestine interests. What makes this curious is that Author Topic seems upon inspection to define the topics more sharply but PLSI-U reveals fewer individuals with clandestine interests despite similar sized interest networks. However, this may simply point to PLSI-U's inability to find clandestine interests because the categories are not finely enough defined. Author Topic finds that, on average, between 0.1% and 0.2% of Enron employees have a clandestine interest in a specific topic while PLSI-U only finds between 0.02% and 0.03% of Enron employees.

Both models produced a manageable number of potential insider threats. What cannot be determined from this data is if PLSI-U produces too few, excluding valid potential insider threats, or Author Topic too many, creating more work for the managers.

4.2 *Timeliness Metric*

In addition to the useability metric, there are two other categories of metrics to consider. The first concerns the timeliness of the results. The algorithms are run on approximately 34,000 employees containing a total of approximately 250,000 emails.

Using a dictionary of stemmed words, Author Topic is able to produce results after 8 days (6 days of the algorithm running and 2 days to analyze the data) while PLSI-U took 9.5 days (7.5 days of the algorithm running and 2 days to analyze the data). In both cases, the results are returned in a period of time that makes monthly or quarterly processing of a company’s email traffic feasible.

4.3 *Validity Metrics*

Finally, a technique that is useable and timely is still useless if the results it produces are not valid. Therefore, the final metric examined checks if Author Topic and PLSI-U produce valid results. First, they are checked to see if Sherron Watkins emerges as an insider and second they are checked to see if common interests emerge between people as expected.

4.3.1 Sherron Watkins as a Potential Insider. To see if Sherron Watkins emerges as a potential insider, she is checked to see if she has a clandestine interest in the off-book-partnerships and if she feel alienated. To see if she feels alienated, she is checked for a clandestine interest in a socializing topic.

4.3.1.1 Off-Book-Partnerships. First consider the Author Topic dataset restricted to dictionary words (Figure ??). The first step an investigator must take is to determine what topic he is interested in. For this investigation, the topic concerns the off-book partnerships. While the initial one was called Rhythms, the later ones were called LJM 1, 2, and 3. The most problematic transactions performed by them were Raptor I, Raptor II, Raptor III, and Raptor IV. In order to allow comparisons between all four datasets, only the word “raptor” is used to find topics (since LJM would only appear in the non-dictionary datasets).

Excluding a general “email” topic, the four topics that the word “raptor” had a non-zero conditional probability for are topics 12 ($p(w = raptor|z = 12) = 0.0011$), 25 ($p(w = raptor|z = 25) = 0.0004$), and 30 ($p(w = raptor|z = 12) = 0.0002$). Observe

that topic 30 is the Research category discussed above. This is very appropriate considering that the Research division was the first to examine and then reject the feasibility of the Raptors. Despite “raptor” not appearing as one of the most probable words, with most probable words like trade, agreement, credit, swap, and financial, this does appear to be a topic related to the Raptors.

STEP 1: DETERMINE THE TOPIC TO INVESTIGATE

Non-zero Raptor probabilities – p(z/w)		Financial Trade Agreements	
Topic 12 Financial Trade Agreements	0.11%	CATEGORY 12	
Topic 25 Financial Risk Management	0.04%	Trade	2.0%
Topic 30: Resarch	0.02%	Copy	1.6%
		Agreement	1.3%
		receive	1.2%
		Executive	1.2%
		click	1.2%
		Credit	1.1%
		Swap	1.1%
		Financial	1.0%

STEP 2: FIND INDIVIDUALS WITH CLANDESTINE INTERESTS IN THE TOPIC

Clandestine Interests – Topic 12 Financial Trade Agreements			
Stacey Ramsey	Angela Liknes	K. Longoria	John Disturnal
Corbin Barnes	Ilan Caplan	Kimberly Hardy	Dave Kistler
Peter Berger	Andrea Reed	Sherron Watkins	Edosa Obayagbona
Trevor Randolph	Frank Lobdell	Mac McLelland	Junellen Pearson
Kelly Lovvorn	Joshua Koenig	Mika Watanabe	John Bottomley
Mark Haedicke	Tori Hayden	Michelle Schultz	Esther Gerratt
Jayanta Sengupta	Nikole Vander	Mchael Nanny	Bryan Garrett
Adam Pollock	Cecil John	Carmella Jones	Victoria McDaniel
Habiba Bayi	Felicia Solis	Anita Grandos	Kimberly Nelson
James Puntumapanitch	Adriana Wynn	Jim Roth	Michael Rump
Melissa Allen	Olivier Herbelot	Nelly Carpenter	Michele Baffer
Katherine Chisley	Laura Johnson	Clay Spears	Patrick Conner
Jeffrey Austin	John Boomer	Tom Halpin	Mary Hubbard
Darla Steffes	Omar Aboudaher	Lena Kasbekar	Peter Traung
James Foster	Gardiner Corby	Robert Pickel	Duncan Croasdale
Peter Maheu	Warren Schick	Joe Hoang	Barbara Hankins
Christi Nicolay	Jay Johnson	Brenda Funk	Fabian Valle
Llewelyn Hughes	Linda Noske	Jesse Alvorado	

Figure 4.10: Investigating if Sherron Watkins is an Insider for Author Topic restricted to Dictionary Words

The next step is checking which individuals have clandestine interests in these topics. Although this investigation needs to be performed for all three topics, only topic 12 is shown here. Observe that Sherron Watkins is one of the 71 individuals who emerge as having a clandestine interest in this topic. Recall that this means that although she has an interest in financial trade agreements, she never sent or received an email from anyone at Enron about them. Observe that in this case the

investigation is being performed after the fact. Therefore, no additional investigation is necessary and Sherron Watkins emerges as a potential insider. If this was being performed to generate potential insider threat leads, the next step would be to talk to the managers of each of the 71 individuals and dig deeper to determine whether these insider threat leads merit further attention.

Unfortunately, when the same steps are performed on the other datasets, the results are not as promising. When the restriction is removed for Author Topic, only two topics emerge as probably related to the word “raptor”: topics 2 ($p(w = raptor|z = 12) = 0.0011$) and 25 ($p(w = raptor|z = 25) = 0.0004$) where topic 25 is again the Research topic and topic 2 appears to be a legal topic. Despite Sherron Watkins being interested in 9 different topics, none of them are either topic 2 or 25. As a result, she does not emerge as a potential insider for this experiment.

After performing the investigation on the Author Topics datasets, the next step is to perform them on the PLSI-U datasets. Raptor does not appear to have a non-zero probability for any topic when PLSI-U is run restricted to dictionary words. However, it does appear when PLSI-U is run without the restriction. The five topics with the highest conditional probabilities are topics 3 ($p(w = raptor|z = 3) = 0.0044$), 33 ($p(w = raptor|z = 33) = 0.0008$), 41 ($p(w = raptor|z = 41) = 0.0005$), 39 ($p(w = raptor|z = 39) = 0.0004$), and 44 ($p(w = raptor|z = 44) = 0.0004$). Observe that category 44 is the Research category discussed above. This is very appropriate considering that the Research division was the first to examine and then reject the feasibility of the Raptors. Topic 3 appears to be about credits and market swaps and topic 39 about trading electricity. Unfortunately, it is difficult to identify any topic descriptions for topics 41 and 33. The next step is checking which individuals have clandestine interests in these topics. While several individuals emerge as having clandestine interests in each of these topics, Sherron Watkins is never one of them.

4.3.1.2 Alienation. Similiar steps are then taken to see if Ms. Watkins appeared to feel alienated at work. For this analysis, there is no clear word

that defines socializing, and so several are used. Appropriate words include *dinner*, *drink*, *fun*, *tonight*, *love*, *weekend*, *family* and *game*. In each case, only one or two topics emerge as having a non-zero probability for each word.

The results appear to parallel those for the off-book partnerships. For the Author Topic dataset restricted to dictionary words, Ms. Watkins emerges as having a clandestine interest in one of the two socializing topics and no interest in the other. Therefore, for this experiment, she emerges (along with 226 other individuals) as possibly feeling alienated and a potential insider threat. When this result is combined with individuals who had clandestine interests in the off-book partnerships, only two other names emerge, Dave Kistler and Llewlyn Hughes. Therefore, if this had been a real world case and the CFO had combined results between these two topics, he could have quickly zeroed in on Watkins as a possible insider threat.

As with the off-book-partnerships, the results from the other datasets are not as promising. Ms. Watkins does emerge as having an interest in a socializing topic when Author Topic is run without the dictionary restriction. However, her interest in this case is not clandestine (she received one email related to this topic). In addition, when the analysis is performed for PLSI-U, she does not emerge as having an interest in the socializing topics at all.

4.3.1.3 Summary. In summary, Sherron only emerges as having a clandestine interest in the “raptor” topic and feeling alienated for Author Topic (dictionary). While it is encouraging that she emerges as a potential insider for Author Topic (dictionary), it is disconcerting that she does not also emerge for Author Topic (no dictionary). One possible explanation for this is the huge number of names that emerge as the most common words for topics. It would be informative to re-run the Author Topic (no dictionary) test with all proper names stripped.

4.3.2 Common Interests. The second metric of validity is checking to see if Ken Lay, Jeff Skilling, and Andy Fastow share similar interests. However, we

have already seen that they do indeed share one, the Senior Mgmt topic, since it was picked by finding a topic of interest they all had in common. This was true for all four experiments. The second test of common interests is to see if Andy Fastow and Michael Kopper emerge as having a shared interest in the “raptor” topic. Unfortunately PLSI generates no interests for Fastow and Kopper when run restricted to dictionary words. However, when run without this restriction, it extracts seven topics of interest for Fastow and two for Kopper, including one in common, topic 24. While topic 24 does not emerge as being related to the Raptors (see Section ??), it does demonstrate a common interest between the two. When Author Topic is checked for common interests between Fastow and Kopper, the results are similar. When run restricted to dictionary words, the only topic they appear to have in common is one related to trading energy. Finally, when run without the restriction, they share an interest in a topic related to home life.

The results for this metric are mixed with promising results for all experiments when considering Lay, Skilling and Fastow and only negative results when considering Kopper and Fastow. However, this may be due to the scarcity of emails for Kopper (only 10) resulting in both techniques having difficulties assigning him to the correct clusters.

4.4 Additional Experimental Results

Now that the probabilistic clustering and social networking techniques have been shown to be useful, there are two final pieces of analysis to perform. The first concerns the number of topics. As discussed in Chapter ??, the number of topics was selected due to hardware and software constraints. A logical question is whether a second iteration of probabilistic clustering can be performed on a single topic and extract 48 additional sub-topics. The second analysis concerns the social networks. While they have already demonstrated their worth by revealing the clandestine interests of various Enron employees, there is still more information that can be extracted from

them. By performing several traditional social network analysis techniques, it may be possible to glean additional useful information from the networks.

4.4.1 Analysis of the Two-Tiered Approach. After performing the above analysis, it is reasonable to wonder if the California Crisis category can be expanded into multiple sub-topics that might allow for a better analysis of Enron's alleged duplicity in California. To test this, the Author Topic and PLSI-U topics from the stemmed word (no dictionary) dataset are considered. The documents and individuals that have a significant probability of this topic are extracted and PLSI-U and Author Topic are performed on them a second time.

The results are not optimistic. When a second level of probabilistic clustering is performed on the PLSI-U results, no additional topics emerge. There is only one recognizable topic that emerges and it is a combination of the California Crisis and the Scheduling topics (Figures ?? and ??). The most probable words for the remaining topics have such a low probability that it suggests that no clustering occurred. While reclustering the Author Topic dataset does produce 48 different identifiable topics, only two topics emerge that are related to the California Crisis. The remaining topics are consistent with the topics that emerged initially when Author Topic was run.

On a positive note, this admittedly limited analysis suggests that the number of topics used for the original analysis is appropriate. If the number of topics had been too small, additional topics unrelated to the original meta-topics would have emerged. That they did not suggests that they are visibly accounted for in the original clustering.

4.4.2 Social Network Analysis. The creation of social networks provides the opportunity to test the effectiveness of traditional social network analysis (SNA) on extracting potential insider threats. First, several centrality measurements are reviewed to see if SNA extracts appropriate individuals as being representative of a particular topic. If so, SNA can also be used to identify individuals central to

(possibly undesirable) topics. Second, the effectiveness of SNA to determine the positions of individuals based on their SNA measurements is considered. Third, an explicit network without topics is reviewed to see if comparing it with the explicit topic networks provides additional insight. Finally, a temporal analysis is performed to see if it is possible to view the movement of a topic across the organization through an analysis of email.

4.4.2.1 Measurements. There are additional methods for considering the explicit social networks generated by the categorized emails. As discussed in Section ??, there are several measurements of centrality that can be performed on these networks. It is of interest to compare the individuals that emerge as the most important by these measurements to those that emerge as being most probable from the implicit social networks generated by the interest profiles. First, consider Figure ?. It is readily apparent that while the most probable individuals and the individuals with the largest centrality measurements are fairly different, there is very little difference between centrality rankings. Between the top three measurements, only five individuals do not appear in at least two of the rankings. This phenomenon repeats for all four topics across all four experiments. Therefore, for brevity, only the top ten most central individuals based on betweenness for each of the four sample topics are shown for each of the four experiments (Figures ?, ?, ?, and ?). Each of these shows a quality of results similar to what is seen with the most probable individuals suggesting that results returned this way may also be used for extracting individuals associated with a (possibly undesirable) topic. As with the most probable individuals, PLSI-U appears better at extracting appropriate individuals as the most central. While there is no efficient way to determine the smallest number of individuals that will most efficiently fracture the network, those individuals deemed the most important can certainly be considered critical to efficient communication within the social network.

The same performance comparison made on individual actors between PLSI-U and Author Topic can also be made on the networks as a whole. By looking at the cohesion measurements (Table ??), one can see that PLSI-U produces clusters with a higher degree cohesiveness (more vertices with the same degrees), closeness cohesiveness and in some cases higher between measurements. This suggests that the PLSI-U clusters may be more cohesive than the Author Topic ones. This may be born out further when one considers the number of components in the social networks among the four experiments (Table ??). PLSI-U run without a dictionary has by far the fewest number of components (in some cases only 1) followed by PLSI-U run with a dictionary. This may be due to the fact that the clusters have far fewer vertices in them. However, since the measurements are normalized to remove graph size from the metric, this is still indicative of more cohesive clusters.

MOST PROBABLE INDIVIDUALS		
James Derrick	General Counsel	3.60%
Cindy Olson	Head of Human Resources	1.98%
Kay Chapman	Secretary of Management Committee	1.67%
Mark Koenig	Executive Vice President of Investor Relations	1.67%
Greg Whalley	President of Enron	1.66%
Steven Kean	Chief of Staff, Government Relations Specialist	1.57%
Mark Frevert	Vice-Chairman of Enron	1.54%
Jeffrey McMahon	Chief Financial Officer of Enron	1.46%
Kenneth Lay	Chairman of Enron	1.25%
David Delainey	Enron Energy Services CEO	1.19%

DEGREE	
Joannie Williamson	Chief Executive Jeff Skilling's Secretary 0.38
Rosalee Felming	Chairman Ken Lay's Secretary 0.28
Bobbie Power	0.24
Billy Lemmons	0.24
Rebecca Carter	Executive Secretary to Enron Board 0.22
Sherri Sera	Chief Executive Jeff Skilling's Personal Assistant 0.21
Liz Taylor	0.21
Paula Rieker	Deputy Director of Investor Relations 0.21
Billy Dorsey	0.20
David Delainey	Chief Executive of Enron Energy Services 0.20

CLOSENESS		
Joannie Williamson	Chief Executive Jeff Skilling's Secretary	0.060
Rosalee Felming	Chairman Ken Lay's Secretary	0.060
David Delainey	Chief Executive of Enron Energy Services	0.059
Liz Taylor		0.059
Sherri Sera	Chief Executive Jeff Skilling's Personal Assistant	0.059
Kay Chapman	Secretary of Management Committee	0.059
Paula Rieker	Deputy Director of Investor Relations	0.059
Nicki Daw		0.059
Jeff Skilling	Chief Executive	0.059
Cindy Olson	Head of Human Resources	0.059

BETWEENNESS	
Joannie Williamson	Chief Executive Jeff Skilling's Secretary 0.21
Bobbie Power	0.09
Traci Ralston	0.09
Billy Lemmons	0.07
David Delainey	Chief Executive of Enron Energy Services 0.06
Jeff Skilling	Chief Executive 0.06
Cindy Olson	Head of Human Resources 0.06
Rosalee Felming	Chairman Ken Lay's Secretary 0.06
Paula Rieker	Deputy Director of Investor Relations 0.05
Liz Taylor	0.04

Figure 4.11: Senior Mgmt Topic - PLSI-U w/o Dictionary - Email Graph Centrality Measurements

4.4.2.2 Position Classification.

During the development of the research question, it was proposed that by using simple SNA metrics, the positions of

Table 4.1: Group Centrality Measurements

	Topic	Vertices	Group Degree	Graph Density	Group Closeness	Group Betweenness	
SENIOR MGMT							
	PLSI-U Dict	45	4800	0.18445	0.00083	0.00222	0.29961
	Author Topic Dict	9	9817	0.08249	0.00071	0.00075	0.13983
	PLSI-U NoDict	11	606	0.36260	0.02149	0.00466	0.20774
	Author Topic NoDict	8	10862	0.11467	0.00083	0.00053	0.09985
CALIFORNIA CRISIS							
	PLSI-U Dict	2	619	0.19304	0.00647	0.00232	0.27696
	Author Topic Dict	43	7434	0.12082	0.00094	0.00072	0.21977
	PLSI-U NoDict	0	168	0.75254	0.05988	0.34176	0.27509
	Author Topic NoDict	18	2653	0.25123	0.00339	0.00077	0.23940
RESEARCH							
	PLSI-U Dict	47	2855	0.14328	0.00140	0.00328	0.33934
	Author Topic	30	7437	0.10751	0.00081	0.00053	0.15977
	PLSI-U NoDict	44	420	0.49411	0.01909	0.01826	0.43733
	Author Topic NoDict	25	2949	0.16070	0.00136	0.00070	0.26934
INFO TECHNOLOGY							
	PLSI-U Dict	40	532	0.36089	0.00377	0.00151	0.46753
	Author Topic Dict	28	7415	0.10942	0.00094	0.00078	0.16977
	PLSI-U NoDict	7	687	0.38592	0.01020	0.16711	0.54800
	Author Topic Dict	10	9232	0.16899	0.00065	0.00087	0.25984

Table 4.2: Topic Component Statistics

	Topic	Vertices	Vertices in Biggest Component	Percentage	Number of Components
SENIOR MGMT					
	PLSI-U Dict	45	4777	99.52	8
	Author Topic Dict	9	9776	99.58	16
	PLSI-U NoDict	11	591	97.52	3
	Author Topic NoDict	8	10811	99.53	22
CALIFORNIA CRISIS					
	PLSI-U Dict	2	585	94.51	10
	Author Topic Dict	43	7388	99.38	18
	PLSI-U NoDict	0	168	100.00	1
	Author Topic NoDict	18	2600	98.00	17
RESEARCH					
	PLSI-U Dict	47	2835	99.30	9
	Author Topic	30	7383	99.27	18
	PLSI-U NoDict	44	414	98.57	3
	Author Topic NoDict	25	2890	98.00	11
INFO TECHNOLOGY					
	PLSI-U Dict	40	420	78.95	17
	Author Topic Dict	28	7374	99.45	16
	PLSI-U NoDict	7	687	100.00	1
	Author Topic Dict	10	9194	99.59	16

CATEGORY 45		SENIOR MGMT	CATEGORY 2		CALIFORNIA CRISIS
Tracey Kozadinos		0.30	Alan Connes	Director of Government Affairs in California	0.28
Jeff Skilling	Chief Executive of Enron	0.22	Kenneth Lay	Chairman of Enron	0.25
Constance Charles	Human Resources – Associate/ Analyst Program	0.17	Simone La		0.13
Steven Kean	Chief of Staff – Government Relations Specialist	0.15	Clayton Seigle		0.12
Rosalee Fleming	Secretary for Chairman Ken Lay	0.15	Jeff Dasovich	Government Affairs Executive	0.08
Rhonda Denton		0.04	Steven Kean	Chief of Staff – Government Relations Specialist	0.08
Bill Donovan		0.04	Karen Denne	Vice President of Public Relations	0.07
Brian Ripley		0.04	Ginger Dernehl	Admin Assistant – Global Government Affairs	0.07
Janet Butler		0.04	Richard Shapiro	VP of Regulatory Affairs, Chief DC Lobbyist	0.07
Rhonda Denton		0.04	Leonardo Pacheco		0.06

CATEGORY 47		RESEARCH	CATEGORY 40		INFO TECHNOLOGY
Vince Kaminski	Managing Director and Head of Research	0.34	Cheryl Johnson		0.47
Outlook Team		0.15	Outlook Team		0.26
Jewel Meeks		0.11	Emma Welsch		0.15
Kristin Gandy	Associate Recruiter for Enron	0.09	Jim Schwieger	Vice President in Gas Trading Division	0.12
Shirley Crenshaw	Research GRoup Administrative Coordinator	0.09	Julie Meyers		0.10
Jeff Dasovich	Government Affairs Executive	0.08	Darren Vanek	Credit Analyst – Credit Risk Management	0.09
Nicki Daw		0.08	Carolyn Gilley	Enron Networks – Information & Records Mgmt	0.08
Richard Shapiro	VP of Regulatory Affairs, Chief DC Lobbyist	0.07	Geoff Storey		0.08
Ashley Baxter	Recruiter – Global Technology Track	0.07	Kevin Dumas		0.06
Althea Gordon	Recruiter – Associates/ Analyst Program	0.07	Daren Farmer	Logistics Manager	0.05

Figure 4.12: PLSI-U Betweenness Centrality Measurements with Sample Categories using Stemmed, Dictionary Words

individuals within an organization can be determined. For instance, perhaps those individuals with many interests are likely to be administrative assistants (or executives) since they are likely to be pulled in more directions than average employees. The results are unclear (Table ??). While most of the individuals are at least vice presidents, there seems to be a significant bias to individuals involved in the manipulation of California’s deregulation effort to increase Enron profits (e.g. Tim Belden, Christian Yoder, John Lavorato, and David Parquet). This suggests that the nature of the topics rather than individuals’ positions within the company are the determining factor in the number of topics of interest an individual possesses.

Similarly, when looking at individuals with the most external email activity (Table ??) the results are also mixed. In some cases, it is evident that the position dictates the email traffic. The individual most responsible for government affairs has the most emails and the Chairman has the eighth most emails. However, in other cases, it is not as obvious. There are two lawyers as well as the Head of Research also in the top ten. While the lawyers can be explained by considering the trouble Enron

CATEGORY 9 SENIOR MGMT			CATEGORY 43 CALIFORNIA CRISIS		
Jeff Skilling	Chief Executive	0.14	Kenneth Lay	Chairman	0.22
Bob Ambrocik	Technical Consultant	0.12	Jeff Dasovich	Government Affairs Executive	0.09
Sally Beck	Chief Operating Officer of Enron Networks	0.08	Outlook Team		0.06
Rosalee Fleming	Secretary to Chairman Ken Lay	0.07	Veronica Espinoza		0.05
Outlook Team		0.06	Cynthia Morrow		0.05
Omaha Help Desk		0.05	Steven Kean	Chief of Staff, Government Relations Specialist	0.04
Shelley Corman	VP Regulatory/ Gov't Affairs, Asst Gen Counsel	0.04	Ginger Dernehl	Admin Assistance – Global Government Affairs	0.04
Technology.Enron		0.04	Shelley Corman	VP Regulatory/Gov't Affairs, Asst Gen Counsel	0.04
Bodysop		0.04	Nicki Daw		0.04
Jeff Dasovich	Government Affairs Executive	0.03	Maureen McVicker		0.03

CATEGORY 30 RESEARCH			CATEGORY 28 INFO TECHNOLOGY		
Bob Ambrocik	Technical Consultant	0.16	Bob Ambrocik	Technical Consultant	0.17
Outlook Team		0.13	Sally Beck	Chief Operating Officer of Enron Networks	0.09
Vince Kaminski	Managing Director and Head of Research	0.08	ILisa Jones		0.07
Leann Walton		0.06	Technology.Enron		0.06
Sally Beck	Chief Operating Officer of Enron Networks	0.05	Outlook Team		0.05
Jeff Dasovich	Government Affairs Executive	0.04	Rick Buy	Executive Vice President & Chief Risk Officer	0.05
Andrea Richards		0.04	Lillian Carroll		0.05
Shirley Crenshaw	Research Group Administrative Coordinator	0.03	Arfan Aziz		0.05
Vince Kaminski	Managing Director and Head of Research	0.03	Shelley Corman	VP Regulatory/ Gov't Affairs, Asst Gen Counsel	0.04
Shelley Corman	VP Regulatory/ Gov't Affairs, Asst Gen Counsel	0.03	Ted Bland		0.03

Figure 4.13: Author Topic Betweenness Centrality Measurements with Sample Categories using Stemmed, Dictionary Words

was in at the time, one would not presume a priori that the Head of Research would lead the company in external email traffic. Only one of the top ten individuals is a “salesman” (Chris Germany, an Enron trader).

The lack of revealing results from this analysis suggests that a more complex model may be more appropriate. To see how Author Topic can be extended to classification of roles, refer to McCallum, et al. [?].

Table 4.3: Individuals with the most interests

Name	Position	Number of Interests
Greg Wolfe	VP of Marketing	8
William Bradford		8
Fred Lagrasta	VP Risk Management Marketing	8
Scott Neal	VP Enron Capital and Trade Resources - Latin America	8
Leslie Reeves		8
Peggy Hedstrom	VP of Energy Operations - Calgary, CA	8
Christian Yoder	Senior Counsel - Portland, OR	8
Debbie Brackett	Sr. Director - Credit Risk Management	9
Tim Belden	Head of Enron Energy Trading in Portland	9
John Lavorato	CEO of Enron North America	10
David Parquet	Vice President of Enron North America	10

CATEGORY 11		SENIOR MGMT	CATEGORY 0		CALIFORNIA CRISIS
Joannie Williamson	Secretary to CEO Jeff Skilling	0.21	Susan Mara	Director of Government Affairs in California	0.28
Bobbie Power		0.09	Jeff Dasovich	Government Affairs Executive	0.24
Tracy Ralston		0.07	Alan Comnes	Director of Government Affairs	0.12
Billy Lemmons		0.07	Joseph Alamo		0.09
David Delaine	CEO of Enron Energy Services	0.06	Sandra McCubbin	Director of Government Affairs in California	0.09
Jeff Skilling	CEO of Enron	0.06	Dan Leff		0.08
Cindy Olson	Head of Human Resources	0.06	Tamara Johnson		0.06
Rosalee Fleming	Secretary to Chairman Ken Lay	0.06	Michael Tribolet	VP of Underwriting and Investment Valuation	0.06
Paula Rieker	Deputy Director of Investor Relations	0.05	Leticia Botello		0.04
Liz Taylor		0.04	Thomas Bennett		0.02

CATEGORY 44		RESEARCH	CATEGORY 7		INFO TECHNOLOGY
Vince Kaminski	Managing Director and Head of Research	0.44	Cynthia Morrow		0.55
Vince Kaminski	Managing Director and Head of Research	0.18	Regan Smith	Network Administrator	0.17
Shirley Crenshaw	Research Group Administrative Coordinator	0.16	Georgia Ward	QA in Development Support	0.13
Ravi Thuraisingham	Director of Global Bandwidth Risk Management	0.07	Brandee Jackson		0.09
Anjam Ahmad		0.05	Bryce Baxter		0.08
Anita Dupont		0.05	Kenneth Harmon		0.07
Vince Kaminski	Managing Director and Head of Research	0.05	Rita Wynne	Manager for Volume Management Group	0.06
Vasant Shanbhogue	Vince Kaminski's Second in Command	0.04	Brian Ripley		0.05
Zimin Lu	Director of Valuation and Training Analytic Group	0.04	Tony Dugger		0.05
Steven Leppard		0.04	Anwar Melethil		0.04

Figure 4.14: PLSI-U Betweenness Centrality Measurements with Sample Categories using Stemmed Words without a Dictionary

4.4.2.3 *An Explicit Network without Topics.* One additional question is how the explicit networks generated from Author Topic and PLSI-U compare to a traditional social network where two individuals are linked if an email exists between them, regardless of the topic discussed in the email. The graph level metrics are as expected when considers the size of the graph (21,790 vertices). The group degree is 0.068. The graph density is 0.0008. The group closeness is 0.00032 and the group betweenness is 0.060. All of these measurements suggest a sparse graph with a wide variation in the number of edges (emails) connecting different individuals. When one

Table 4.4: Individuals with the most external emails

Name	Position	External Emails	Internal Emails
Jeff Dasovich	Government Affairs Executive	4,104	11,087
Kay Mann	Legal Counsel	2,716	6,417
Matthew Lenhart		2,229	2,193
Vince Kaminski	Managing Director and Head of Research	2,015	8,764
Sara Schackleton	VP and General Counsel - Enron North America	1,971	10,275
Jeff Dasovich	Government Affairs Executive	1,767	337
Chris Germany	Trader for Enron North America - East Desk	1,650	4,797
Ken Lay	Chairman	1,538	20
Gerald Nemeec		1,489	5,633
Tana Jones		1,119	10,799

CATEGORY 8 SENIOR MGMT			CATEGORY 18 CALIFORNIA CRISIS		
Sasly Beck	Chief Operating Officer of Enron Networks	0.10	Jeff Dasovich	Government Affairs Executive	0.24
Technology.Enron		0.10	Mary Hain	Government Affairs lawyer	0.10
Kenneth Lay	Chairman	0.10	Ginger Dernehl	Admin Assistant – Global Government Affairs	0.07
Billy Lemmons		0.05	Paul Kaufman	VP & Western U.S. regulatory & govt lawyer	0.06
Outlook Team		0.05	Susan Mara	Director of Government Affairs in California	0.06
Jeff Skilling	Chief Executive	0.04	Richard Shapiro	VP of Regulatory Affairs, Chief DC lobbyist	0.06
Andrew Wu		0.04	Rhonda Denton		0.06
David Oxley	Human Resources Executive	0.03	Lara Leibman	Manager in Government Affairs	0.06
Louise Kitchen	CEO of Enron Online	0.03	Christi Niclay	Director of Govt Affairs – Electric Power Trading	0.05
David Forster		0.03	Alan Connes	Director of Government Affairs in California	0.04

CATEGORY 25 RESEARCH			CATEGORY 10 INFO TECHNOLOGY		
Vince Kaminski	Managing Director and Head of Research	0.27	Outlook Team		0.26
Khymberly Booth		0.18	Lynette Crawford		0.11
Shirley Crenshaw	Research Group Administrative Coordinator	0.14	Sonya Johnson		0.10
Iris Mack	Manager in Research Group – Broadband Services	0.13	Suzanne Brown		0.09
Cheryl Johnson		0.12	David Forster		0.07
Vince Kaminski	Managing Director and Head of Research	0.10	Julie Clyatt		0.07
Maureen Raymond	Head of Country Risk & Foreign Exchange	0.08	Constance Charles	Human Resources – Associate/ Analyst Program	0.07
Leann Walton		0.07	Paulette Obrecht	Legal Project Coordinator	0.04
Kathie Grabstald	Public Relations for Enron Wholesale Services	0.07	Kay Chapman	Secretary of Management Committee	0.03
Gwyn Koepke		0.06	Khymberly Booth		0.03

Figure 4.15: Author Topic Betweenness Centrality Measurements with Sample Categories using Stemmed Words without a Dictionary

looks at the individuals who are most central to the graph, some of the senior Enron individuals emerge (Figure ??).

Jeff Skilling	Chief Executive	0.06
Outlook Team		0.06
Sasly Beck	Chief Operating Officer of Enron Networks	0.04
David Forster		0.04
Kenneth Lay	Chairman	0.04
Jeff Dasovich	Government Affairs Executive	0.03
Louise Kitchen	CEO of Enron Online	0.02
John Lavorato	CEO of Enron North America	0.02
Jacqueline Coleman		0.02
Technology.Enron		0.02

Figure 4.16: Most central individuals in a traditional Enron social network

4.4.2.4 Temporal Analysis. A final goal of this research is to track a topic as it progresses through the organization’s email. Although several topics were reviewed, including Jeff Skilling’s resignation, Sherron Watkins’ letter to Ken Lay, Jeff Skilling and Rebecca Carter’s marriage, and the renaming of the Astrodome to

Enron Field, none of these topics emerge as moving through the organization by email over time. Given the attractiveness of these rumors, it is suggestive that means of communication other than email were used by Enron employees. These could include either the telephone or talking face-to-face. It is also possible that the reason is the nature of the Enron dataset itself. Since it is extracted from the email folders of only 151 Enron employees, it may be more difficult to perform this temporal analysis.

4.5 Resources

This thesis requires the development and use of multiple computer programs to extract, cluster, and analyze the data. All program development and execution has occurred on a Pentium 4 2.6GHz machine with 512MB RAM running Suse Linux 9.2. The programs were developed in C and C++ using the gcc 3.3.4-11 compiler. In addition, several Unix shell scripts were developed to aid in processing. The database used was MySQL version 4.1.12-max.

4.6 Conclusions

The primary purpose of this research is to determine if probabilistic clustering and social networking techniques applied to email are effective at detecting potential insider threats. Consider the results:

4.6.1 Metric 1: Useability. Both PLSI-U and Author Topic clearly succeed at this metric. The topics are clearly defined by their most probable words and the most probable individuals for the topics are highly appropriate. Both provide appropriate names as additional leads given a topic of interest, although both perform better when words are not restricted to the dictionary. This can be further improved for Author Topic by excluding names from the word lists. The only place where PLSI-U falls short is in the number of individuals extracted with clandestine interests. While Author Topic succeeds at this metric by extracting between 0.1% and 0.2% of

the population as potential insiders, PLSI-U fails to extract enough. This is born out empirically by the discovery of Sherron Watkins by Author Topic and not by PLSI-U.

4.6.2 Metric 2: Timeliness. For PLSI-U, each iteration took almost two hours to perform, resulting in a total run time of almost one week. Similarly, Author Topic took between 6 and 9 hours to perform 100 iterations, resulting in a total run time of also approximately a week. While it is possible to imagine a larger corpus, Enron's 250,000 emails and 35,000 individuals seems on the high side for email data collected on a monthly basis even for organizations that make extensive use of email if broadcast messages are excluded. Although the average individual's email is only in the low 70s, several individuals received emails in the thousands and even tens of thousands. The ability to produce results in a week should result in a system that is sufficiently responsive.

4.6.3 Metric 3: Validity. For Author Topic run restricted to dictionary words, Sherron Watkins emerges as a potential insider. Furthermore for PLSI-U run without being restricted to the dictionary, Sherron Watkins also emerges as having an interest in the Raptor topic. However, in this experiment, she does not appear to have a clandestine interest in it. This is due to the fact that despite the Raptor topic being one of her two principle interests, her interest is not great enough to be considered a member of the implicit network for that topic. The one disconcerting result is that Author Topic run without the dictionary does not cluster Ms. Watkins with a Raptor topic. Perhaps additional experiments run with Author Topic not restricted to the dictionary but with all names excluded will produce better results.

The emergence of common topics of interest for Lay, Skilling, and Fastow provides implicit support that once one insider is known, these techniques can be used to extract additional ones. Finally, despite the fact that neither Kopper nor Fastow emerged as having a common interest in any of the Raptor topics, something interesting does emerge. For each experiment, the Research topic emerged as being a promising topic for the Raptors. Presumably, given the close knitness of the indi-

viduals associated with this topic, that members of the Research Group would have emerged as having a significant interest in the Raptors, thus providing a means, after the fact, of finding people to talk to about them.

4.6.4 Social Network Analysis (SNA). Centrality measurements extract individuals for topic very similar to the most probable individuals extracted by PLSI-U and Author Topic. Interestingly, whether degree, closeness, or betweenness is used, the top ranked individuals vary only slightly. Also, just as with the most probable individuals, PLSI-U produces more cohesive, better identifiable and better clustered results than Author Topic. Where SNA fails to perform is in extracting the positions of individuals within the organization. The two tests performed clearly indicate a more complex model is required. In addition, SNA fails to extract any indications of topics moving through the dataset over time. While this could be a result of the enormous amount of data, the two-tiered approach shows that the 48 topics provide a reasonable amount of granularity for this dataset. Therefore, it is unlikely that increasing the number of topics would improve the results of the SNA analysis. Instead, it is likely that the Enron dataset itself is not conducive to this effort, perhaps due to the dataset being extracted from the email folders of only 151 Enron employees.

V. Discussion

No country or organization can take on the military might of the United States of America head on. The Gulf War of 1991 is the last time, for a while, that a country will go “toe to toe” with the United States in the field. What has occurred and will continue to occur with greater and greater frequency is asymmetric. While the destruction of the World Trade Center and the attack on the Pentagon are the two best known, the asymmetric attacks on America’s economy are no less destructive. Unfortunately, while it is easy to distinguish friend from enemy when it comes to physical attacks, the problem is much more complex for economic attacks. In addition to countries like China, it is considered likely that supposed allies such as France, Germany, and Japan provide active state support for economic espionage [?]. While it is possible to insert traditional spies into organizations, turning those people who were once loyal employees has emerged as the most effective way of stealing an organization’s secrets.

5.1 *Experimental Results*

As such, it is critical to find ways to protect the information. This research proposes one method for detecting potential insider threats by finding individuals who exhibit some of the warning signs. One specific warning sign many individuals demonstrate is a need to separate themselves from the organization prior to betraying it. This separation process may manifest itself by individuals feeling alienated by the organization. By datamining email, individuals’ interests can be extracted. From those interests, one can predict who feels alienated from an organization and who is hiding their interest of sensitive topics. The goal is to test the hypothesis that “probabilistic clustering and social networking techniques applied to email are effective at detecting potential insider threats”. To accept this hypothesis, the techniques must be valid, useable, and timely.

Finding an appropriate corpus to test this hypothesis is difficult. Artificial datasets are too small and real-world datasets are unavailable due to privacy concerns.

Luckily, the release of Enron’s 2000 and 2001 email data by the Federal Energy Regulatory Commission during their investigation has provided a valuable dataset for researchers. By using this dataset, it is possible to test if Sherron Watkins, Enron’s famous whistleblower, emerges as an insider. At the same time, the effectiveness of discovering additional people possibly linked to a known insider can be tested by checking if notables such as Ken Lay (Enron’s chairman), Jeff Skilling (Enron’s CEO), and Andy Fastow (Enron’s CFO) emerge as having similar interests.

The experiments are run for Author Topic and Probabilistic Latent Semantic Index with Users (PLSI-U). Experiments are run on the Enron corpus where words are first restricted to only words in the dictionary and then to all “words” whether or not they are in the dictionary. In all experiments, words are stemmed resulting in a 50% reduction in the number of words considered.

The results show that both Author Topic and PLSI-U produce good results. When the restriction of only dictionary words is removed, both perform excellently at clustering words to produce topics that are easily understandable. Furthermore, both also perform excellently at associating individuals with appropriate topics. As a result, if a topic emerges that suggests a potential insider threat, both techniques do well at uncovering individuals interested in that topic. However, while both perform well at the forensic analysis of finding associates of known insiders, only Author Topic performs well on the Enron dataset at revealing potential insiders through clandestine interests by revealing Sherron Watkins as an insider. She emerges as having a clandestine interest in the off-the-book partnerships and as feeling alienated. PLSI-U finds too few individuals with clandestine interests to be useful. Finally, both techniques produce results in a timely manner, taking less than a week to produce results on a corpus of over 250,000 messages involving over 34,000 employees.

5.2 Future Work

The initial research question was “Are probabilistic clustering techniques applied to email and internet activity effective at detecting potential insider threats?”. This

had to be scaled back due to privacy concerns related to using non-publically available email and internet activity for a research project. However, this resulted in the overloading of email. In this experiment, email is used to both define the topics and to determine who is not revealing their interest in a topic. On the surface this should result in no clandestine interests since the only way someone is considered to have an interest in a topic is if they send or receive an email about it. The only reason this is not the case is because during the definition of topics internal and external emails are considered while during the search for clandestine interests only internal emails are used. By using a different data source for generating topics of interest, more clandestine interests may emerge. Internet history is kept on servers in the same way that email history is. In addition, PLSI-U and Author Topic can easily morph from documents made up of words to web pages made up of hyperlinks [?]. While internet activity was not available for Enron, it is generally available from the same sources that supply email history logs. By excluding the names of individuals from the email bodies and encoding the identities of the individuals in the message and internet activity headers, this experiment could effectively be expanded to include both email and internet activity.

An issue that emerged during analysis was the unknown positions of many of the Enron employees. Although the Internet and books help, the jobs of a large percentage of Enron employees remain unknown. Future research that assembles the positions of all of the Enron employees in the dataset would result in much improved data analysis.

A second problem that emerged during the set up of the experiment was the number of topics. Although the final number was chosen based on hardware and software considerations, it would have been desirable to arrive at an *optimal* number of topics. Teh, et al [?] provide a mechanism for determining the optimal number of topics without having to decide a priori. Although McCallum, et al. [?] suggest that Teh is not effective at performing the actual clustering, McCallum still uses the number of topics suggested by Teh. Performing this analysis on the Enron corpus would

provide a means of finding a better number of topics, possibly resulting in tighter clusters and better results by one or both of the algorithms. An alternative method is to perform a second level of probabilistic clustering on the subset of documents and individuals associated with a specific topic. Unfortunately, when this was performed on the California Crisis topic, no additional information was revealed. However, this was a limited test and it is possible that if performed on all of the topics, some of them might have better results.

One question often asked is whether too much information is lost when context is taken away from emails and replaced with word frequencies. While the results of the experiment show that the simplified model does maintain enough information to be useful, there is an even simpler, more direct method of testing it. By adding in some domain knowledge, it is possible to see whether or not by simply extracting user and word co-occurrence frequencies, certain logical groupings (such as organizational units) occur. This is very similar to the traditional information term-frequency-inverse document frequency model. Since both words and people are represented explicitly in the database, this can be tested directly by database queries without the need for extensive C programming.

Finally, two additional areas of future research emerged during the analysis of results. The performance of Author Topic appeared to degrade when the restriction of only dictionary words was removed. This was due in large part to the prevalence of names as the most probable words for topics. Attempts to exclude all names (including those that are found in the dictionary such as “ken” and “skill”) may produce better results. Secondly, the biggest problem found with PLSI-U was the small size of its implicit and explicit networks. Finding a better way to determine who should be included in the PLSI-U networks would overcome the biggest drawback to this technique.

5.3 Impact

This research has developed a tool that effectively datamines large datasets of email and extracts the interests of individuals. This tool has proven effective at revealing individuals with clandestine interests who may become insider threats and at revealing clusters of people who have the same, possibly questionable, interests. In the short term, this tool can be applied in a real world setting as one of several tools for detecting potential insiders, assisting management, and allocating its limited resources at preventing potential insiders from becoming actual insider threats.

Appendix A. Most Probable Words

A.1 PLSI-U with only Dictionary Words

0		1	
assembl	0.0027719999	alia	0.0399390012
crisi	0.0023409999	unknown	0.0360470004
ab	0.0021980000	varianc	0.0304809995
sue	0.0016960000	detect	0.0237600002
jack	0.0016570000	pars	0.0214889999
ponder	0.0015219999	ancillari	0.0191290006
legislatur	0.0014770000	award	0.0191120002
conserv	0.0014730000	attempt	0.0086970003
republican	0.0014560000	borland	0.0083670001
stout	0.0014140001	engin	0.0083590001
sen	0.0013040000	tie	0.0058269999
megawatt	0.0012260000	manual	0.0051139998
parquet	0.0011549999	intervent	0.0050900001
blackout	0.0010660000	download	0.0050889999
jean	0.0010600000	export	0.0042679999
dean	0.0010370000	insuffici	0.0031940001
carter	0.0010160001	wheel	0.0029290000
freez	0.0010060000	memori	0.0027960001
curt	0.0009570000	interchang	0.0025299999
deregul	0.0009360000	retriev	0.0024200000
lynch	0.0009350000	match	0.0018320000
renew	0.0008830000	lock	0.0010050000
judg	0.0008770000	disk	0.0009740000
grid	0.0008770000	bad	0.0007890000
gate	0.0008750000	mead	0.0007340000
2		3	
assembl	0.0024130000	varianc	0.0329449996
ab	0.0019570000	attempt	0.0251090005
crisi	0.0019150000	engin	0.0249189995
deregul	0.0018990000	borland	0.0248590000
urg	0.0017910000	detect	0.0215210002
declar	0.0016490000	alia	0.0185720008
freez	0.0015300000	pars	0.0151140001
sen	0.0013610000	ancillari	0.0140899997
cent	0.0013300000	award	0.0140580004
retir	0.0012750000	unknown	0.0124930004
donat	0.0012330000	tie	0.0087510003
legislatur	0.0011560000	export	0.0070480001
sold	0.0011280000	insuffici	0.0056960001
relief	0.0010939999	wheel	0.0050360002

lo	0.0010790000	memori	0.0049009998
angel	0.0010400000	retriev	0.0029140001
residenti	0.0010340000	download	0.0022940000
largest	0.0010030000	manual	0.0022690000
commission	0.0009590000	intervent	0.0022489999
megawatt	0.0009500000	disk	0.0020750000
advic	0.0009270000	interchang	0.0018680000
republican	0.0009230000	nob	0.0012240000
hurt	0.0009000000	bad	0.0010390000
write	0.0008950000	empti	0.0010100000
basic	0.0008890000	mead	0.0008050000

4

flash	0.0018510000
troubl	0.0016940000
technocraci	0.0015870000
volatil	0.0013070001
break	0.0013020000
moon	0.0011740000
boi	0.0011670000
realiz	0.0011540001
race	0.0011460000
land	0.0011370000
newspap	0.0011100000
fly	0.0011060000
nobodi	0.0010860000
plane	0.0010790000
neighbor	0.0010770001
broke	0.0010760000
franc	0.0010730000
earth	0.0010670000
editori	0.0010580000
harvest	0.0010570000
gloat	0.0010570000
earthquak	0.0010500000
alwai	0.0009490000
rabbit	0.0009400000
hide	0.0009310000

6

turbin	0.0047860001
exhibit	0.0037930000
interconnect	0.0020490000
tweed	0.0020039999

5

assembl	0.0036800001
legislatur	0.0022140001
republican	0.0015369999
democrat	0.0014880000
sharp	0.0014220000
surcharg	0.0013510000
mitig	0.0012320000
ab	0.0012070000
creditor	0.0011200000
suspend	0.0011150000
compon	0.0011070000
suspens	0.0010640000
judg	0.0010610000
reportedli	0.0010220000
curri	0.0009170000
dedic	0.0009080000
oppos	0.0008980000
attempt	0.0008970000
issuanc	0.0008710000
exempt	0.0008630000
crisi	0.0008600000
neg	0.0008550000
switch	0.0008340000
complaint	0.0008230000
sander	0.0007290000

7

derrick	0.0016060000
hu	0.0011420000
lanc	0.0011350000
telecommun	0.0010410000

abb	0.0019050000	search	0.0010160001
booth	0.0018530000	litig	0.0010010001
equip	0.0017550000	memorandum	0.0009330000
consent	0.0013839999	edg	0.0008540000
dissemin	0.0012900000	block	0.0008160000
heather	0.0011610000	portal	0.0008070000
liabil	0.0009780000	advisor	0.0007850000
construct	0.0009760000	outlin	0.0007360000
warranti	0.0009520000	circul	0.0006980000
studi	0.0009410000	sharp	0.0006810000
shoemak	0.0009360000	consent	0.0006780000
damag	0.0009300000	sander	0.0006540000
escrow	0.0009080000	bar	0.0006110000
dale	0.0008700000	traffick	0.0006060000
assumpt	0.0008700000	narcot	0.0006060000
rose	0.0008580000	wall	0.0005930000
blue	0.0008430000	underli	0.0005720000
exclus	0.0008420000	partner	0.0005640000
northwestern	0.0008070000	merger	0.0005580000
exempt	0.0007650000	watch	0.0005560000
hunger	0.0007570000	conduct	0.0005380000

8

holli	0.0023399999
annex	0.0017400000
worksheet	0.0016850000
taffi	0.0013920000
brant	0.0011800000
diamond	0.0010939999
cordial	0.0010100000
engag	0.0009710000
pinto	0.0008470000
seminar	0.0008300000
threshold	0.0007550000
brokerag	0.0007340000
sack	0.0006980000
glover	0.0006780000
voic	0.0006670000
bear	0.0006620000
appoint	0.0006550000
synchron	0.0006040000
reliant	0.0005640000
jai	0.0005370000
consent	0.0005310000

9

pend	0.0039390000
pep	0.0027760000
verbal	0.0015860000
bridg	0.0012910001
glover	0.0011480000
upstream	0.0011260000
thorn	0.0011260000
templat	0.0010880000
holli	0.0007830000
laurel	0.0007810000
jean	0.0007540000
annex	0.0006790000
cordial	0.0006730000
aris	0.0006150000
payabl	0.0006060000
dissemin	0.0005920000
jai	0.0005860000
destroi	0.0005750000
arriv	0.0005580000
divis	0.0005500000
lost	0.0005490000

perfect	0.0005170000	pat	0.0005330000
enclos	0.0004690000	chase	0.0005290000
doctor	0.0004570000	accordingli	0.0005200000
okai	0.0004490000	synchron	0.0005120000
	10		11
lambert	0.0065680002	turbin	0.0030169999
brant	0.0041580000	holli	0.0021579999
hare	0.0040859999	fisher	0.0017340000
border	0.0037250000	curtail	0.0016830000
amber	0.0032150000	numer	0.0015380000
foreign	0.0030429999	wind	0.0014360000
hardi	0.0029899999	facsimil	0.0013400000
thereund	0.0026370001	fault	0.0010430000
mandola	0.0022030000	divis	0.0009960000
quantiti	0.0021970000	download	0.0009220000
highlight	0.0019030001	restrict	0.0009030000
steel	0.0018480000	beta	0.0008970000
fin	0.0018150000	crude	0.0008510000
curri	0.0017790000	eta	0.0008350000
float	0.0015750000	walker	0.0007860000
cluster	0.0015160000	freight	0.0007740000
profil	0.0015150000	tanker	0.0007390000
campo	0.0015010000	duff	0.0007000000
divis	0.0014980000	sky	0.0006990000
mous	0.0014350000	deni	0.0006730000
merchant	0.0013850000	mesa	0.0006440000
notion	0.0013830001	mainten	0.0006280000
pulp	0.0013750000	diamond	0.0006010000
cross	0.0012370000	sigma	0.0005720000
currenc	0.0011340000	phase	0.0005610000
	12		13
challeng	0.0080220001	sat	0.0050149998
score	0.0059580002	fri	0.0031740000
safeti	0.0043560001	sun	0.0024649999
trimest	0.0042670001	pager	0.0024369999
privaci	0.0042369999	player	0.0021319999
algebra	0.0042320001	buffalo	0.0018900000
speck	0.0034490000	volunt	0.0014550000
studi	0.0033890000	chase	0.0011760000
student	0.0033839999	round	0.0011140000
grade	0.0033770001	kid	0.0009520000
math	0.0029350000	hand	0.0009250000

allianc	0.0027360001	ride	0.0008730000
blizzard	0.0025970000	franchis	0.0008710000
tweed	0.0025919999	babi	0.0008650000
basil	0.0025909999	bowl	0.0008560000
barrow	0.0025820001	children	0.0008350000
similar	0.0025340000	walk	0.0008300000
forum	0.0024750000	unifi	0.0008180000
promot	0.0022080000	stand	0.0007910000
pub	0.0021800001	mile	0.0007650000
maria	0.0021579999	church	0.0007610000
road	0.0020109999	payabl	0.0007540000
annual	0.0019820000	annual	0.0007540000
runner	0.0019789999	shop	0.0007350000
institut	0.0018800000	life	0.0007230000

14

pilot	0.0140190003
supervisor	0.0031840000
rep	0.0029370000
talent	0.0022199999
human	0.0020480000
graduat	0.0019600000
billi	0.0019050000
pep	0.0018300000
elig	0.0018070000
rank	0.0014149999
enjoi	0.0013290000
excit	0.0013090000
guid	0.0012650000
rotat	0.0012460001
campu	0.0012300001
profession	0.0012190000
declin	0.0011440000
choos	0.0010690000
broad	0.0010420000
luncheon	0.0009920000
mailbox	0.0009640000
presenc	0.0009380000
chair	0.0009360000
employ	0.0009290000
philosophi	0.0009220000

16

mailbox	0.0131150000
---------	--------------

15

evid	0.0014900001
reli	0.0014759999
enforc	0.0014570000
bind	0.0014570000
hereto	0.0014540000
club	0.0014130000
hei	0.0013220001
men	0.0012530000
viru	0.0012240000
trip	0.0012210000
brad	0.0011560000
chamber	0.0011480000
bar	0.0011130000
jai	0.0010270000
leas	0.0010019999
wed	0.0010010001
church	0.0009960000
exam	0.0009860001
footbal	0.0009830000
god	0.0009650000
nice	0.0008910000
chart	0.0008860000
sound	0.0008760000
awai	0.0008760000
fun	0.0008400000

17

salli	0.0068250000
-------	--------------

journal	0.0078779999	beck	0.0045540002
configur	0.0055780001	survei	0.0034280000
signer	0.0041439999	beth	0.0019729999
trust	0.0040890002	unifi	0.0014900001
survei	0.0034070001	doorstep	0.0012220000
permiss	0.0033430001	reev	0.0011980000
palm	0.0033320000	consolid	0.0011970000
lotu	0.0033280000	violat	0.0010340000
prompt	0.0033270000	piper	0.0010290000
desktop	0.0033050000	withhold	0.0009320000
convert	0.0031190000	hector	0.0008680000
factor	0.0028919999	captur	0.0008340000
remot	0.0028609999	ap	0.0008250000
prayer	0.0028059999	heather	0.0007870000
task	0.0027920001	logist	0.0007420000
tree	0.0026930000	dedic	0.0007340000
abort	0.0019430000	glover	0.0007280000
failur	0.0019390000	assess	0.0007200000
assort	0.0019290000	mill	0.0007160000
prai	0.0018240000	jenni	0.0006800000
gather	0.0018010000	valid	0.0006600000
manual	0.0017750000	focus	0.0006440000
finish	0.0017470000	audit	0.0006410000
pilot	0.0017460000	roll	0.0006310000

18

channel	0.0040279999
campu	0.0029549999
rice	0.0028440000
univers	0.0028349999
turret	0.0025720000
super	0.0023370001
student	0.0023129999
workshop	0.0021820001
displai	0.0019890000
audio	0.0019550000
wireless	0.0019410000
volunt	0.0019390000
telephoni	0.0019280000
seminar	0.0016990000
guid	0.0016730000
round	0.0016150000
instal	0.0015940000
screen	0.0014130000

19

rotat	0.0066800001
middl	0.0062440000
children	0.0046219998
career	0.0023149999
talent	0.0020740000
paycheck	0.0018260001
hemispher	0.0017370000
focus	0.0017290000
interact	0.0016589999
ongo	0.0016390000
child	0.0016300000
essenti	0.0015670001
sick	0.0015360001
billi	0.0015040000
orient	0.0014530000
simultan	0.0013980001
multipl	0.0013670000
accur	0.0013630000

instant	0.0014060000	task	0.0013570000
coverag	0.0013720000	ascertain	0.0013539999
cultiv	0.0012970000	written	0.0013530001
televis	0.0012890000	verbal	0.0013510000
pep	0.0012560000	fundament	0.0013160000
dolor	0.0012550000	bland	0.0012700000
touch	0.0012520000	motiv	0.0012530000
	20		21
dime	0.0051480001	cougar	0.0089929998
warehous	0.0030960001	alumni	0.0059280000
donat	0.0028760000	channel	0.0040569999
match	0.0018860000	locker	0.0040279999
walk	0.0018740000	garag	0.0039189998
bate	0.0018040000	shop	0.0034709999
fundament	0.0015820001	bodi	0.0033130001
sincer	0.0015790000	membership	0.0032110000
maintain	0.0014900001	studio	0.0029330000
fundrais	0.0014050000	fit	0.0027340001
fire	0.0013150000	turret	0.0026920000
health	0.0012850000	director	0.0026660000
instruct	0.0012530000	heart	0.0024619999
proud	0.0012050000	instal	0.0022650000
map	0.0012020000	walk	0.0022239999
hogan	0.0011930000	upcom	0.0022150001
logist	0.0011390001	enrol	0.0021740000
flight	0.0011110000	machin	0.0021650000
personnel	0.0011100000	screen	0.0021550001
babi	0.0010560000	wireless	0.0020719999
lose	0.0010310001	audio	0.0020510000
volunt	0.0010140000	reconnect	0.0020470000
messeng	0.0010000000	telephoni	0.0020140000
store	0.0009840000	construct	0.0019810000
welcom	0.0009700000	displai	0.0019749999
	22		23
simul	0.0191009995	restrict	0.0046589999
estat	0.0054250001	zone	0.0031560000
restrict	0.0040059998	setup	0.0022970000
mid	0.0034970001	path	0.0021770000
seat	0.0033720001	mainten	0.0020890001
backup	0.0031440000	unifi	0.0020540000
plaza	0.0029249999	stack	0.0020540000
orient	0.0027200000	scenario	0.0020480000

protocol	0.0026509999	victor	0.0019820000
personnel	0.0025609999	quantiti	0.0019360000
payrol	0.0024460000	meter	0.0018940000
profil	0.0023739999	shipper	0.0017570000
regular	0.0022920000	receipt	0.0017560000
speed	0.0021770000	imbal	0.0017130000
split	0.0020369999	station	0.0015760000
verifi	0.0020079999	pipe	0.0014700000
encourag	0.0019499999	personnel	0.0014670000
valuabl	0.0019430000	navig	0.0014320000
divid	0.0019280000	downtim	0.0013680001
compli	0.0018119999	inject	0.0013650000
human	0.0017460000	estat	0.0013500000
gain	0.0015190000	swing	0.0013150000
deduct	0.0014940000	withdraw	0.0012900000
deposit	0.0013800000	upgrad	0.0012630000
setup	0.0013079999	gulf	0.0011760000

24

wild	0.0017000000
goos	0.0016730000
salmon	0.0011610000
fare	0.0011540001
life	0.0011290000
god	0.0010640000
cow	0.0009500000
expans	0.0008620000
alwai	0.0008130000
crosswalk	0.0008070000
true	0.0008020000
subscrib	0.0008020000
newslett	0.0007930000
car	0.0007920000
beer	0.0007900000
fish	0.0007790000
download	0.0007760000
men	0.0007740000
flight	0.0007710000
rice	0.0007670000
hei	0.0007380000
land	0.0007360000
fun	0.0007300000
search	0.0007180000
trip	0.0007070000

25

shipper	0.0019530000
kern	0.0010370000
exhibit	0.0008910000
expans	0.0008280000
drop	0.0008030000
winter	0.0007910000
river	0.0007840000
fundament	0.0007810000
proceed	0.0007320000
video	0.0007220000
cool	0.0007200000
screen	0.0006880000
temperatur	0.0006620000
border	0.0006260000
inject	0.0005540000
receipt	0.0005270000
arriv	0.0005190000
forgot	0.0005050000
jai	0.0004900000
gai	0.0004810000
annual	0.0004770000
rai	0.0004630000
toll	0.0004520000
judg	0.0004450000
extend	0.0004450000

26		27	
dominion	0.0032090000	fool	0.0020620001
ruth	0.0029509999	attack	0.0014470000
robin	0.0026570000	fed	0.0009970000
sold	0.0022700001	bush	0.0009340000
reliant	0.0018119999	sector	0.0008640000
barb	0.0017060000	assembl	0.0008500000
prepai	0.0016960000	pressur	0.0007880000
wire	0.0016530000	font	0.0007680000
hill	0.0016150000	flight	0.0007620000
maria	0.0013810000	face	0.0007600000
loan	0.0012120000	democrat	0.0007490000
forum	0.0011500000	crisi	0.0007460000
troi	0.0010610000	growth	0.0007290000
restaur	0.0010600000	currenc	0.0007210000
island	0.0010479999	inflat	0.0007080000
star	0.0010300000	militari	0.0007050000
ponderosa	0.0009770000	synchron	0.0006940000
temptat	0.0009600000	republican	0.0006590000
lone	0.0009600000	terrorist	0.0006490000
crude	0.0009260000	export	0.0006130000
petit	0.0009150000	airlin	0.0006110000
sabra	0.0009090000	privat	0.0005990000
motion	0.0008620000	health	0.0005970000
receipt	0.0008460000	nuclear	0.0005920000
imbal	0.0008220000	packag	0.0005800000

28		29	
notion	0.0102239996	sander	0.0025220001
extern	0.0061769998	tonn	0.0023670001
trail	0.0051569999	pulp	0.0022450001
gross	0.0046230000	pub	0.0019050000
life	0.0041600000	lost	0.0014680000
metal	0.0025269999	arbitr	0.0014560000
count	0.0024420000	digest	0.0013940000
kingdom	0.0017910000	litig	0.0013580000
barrel	0.0017330000	joint	0.0010319999
percentag	0.0016870000	edit	0.0010130000
difficulti	0.0016790000	cargo	0.0009790000
tab	0.0016370000	mill	0.0009370000
crude	0.0016139999	derrick	0.0008570000
cent	0.0015930000	highlight	0.0008550000
currenc	0.0015890000	metal	0.0008210000

entitl	0.0015840000	usag	0.0007900000
menu	0.0015450000	inventori	0.0007880000
rose	0.0015170000	declin	0.0007040000
audienc	0.0012740000	port	0.0007030000
simon	0.0011750000	consumpt	0.0006920000
coal	0.0010620001	sampl	0.0006780000
slide	0.0009310000	sin	0.0006710000
investor	0.0009220000	damag	0.0006590000
statist	0.0009060000	owner	0.0006040000
ventur	0.0008820000	compar	0.0005940000

30

shipper	0.0019680001
winter	0.0015470000
station	0.0015400000
northern	0.0012170000
drew	0.0012040000
randi	0.0011910000
expans	0.0010490000
interconnect	0.0009520000
engin	0.0009450000
construct	0.0009070000
receipt	0.0008950000
grave	0.0008640000
pipe	0.0008500000
rich	0.0008090000
pressur	0.0007650000
river	0.0007600000
font	0.0007480000
border	0.0007480000
meter	0.0007430000
growth	0.0007250000
jean	0.0007110000
extend	0.0006780000
compressor	0.0006720000
mile	0.0006650000
imbal	0.0005970000

32

sander	0.0013839999
employ	0.0013630000
synchron	0.0013250000
van	0.0011710000
gather	0.0011270000

31

rod	0.0037799999
japan	0.0020490000
ap	0.0015910000
metal	0.0010250000
simon	0.0009280000
eta	0.0008520000
local	0.0008400000
jurisdict	0.0007940000
dale	0.0007910000
blackberri	0.0007630000
chandler	0.0007480000
bandwidth	0.0006870000
foreign	0.0006610000
sin	0.0006510000
claus	0.0006500000
templat	0.0006120000
assess	0.0006010000
advic	0.0005960000
hong	0.0005950000
dilig	0.0005840000
restrict	0.0005820000
incom	0.0005810000
annex	0.0005790000
pa	0.0005770000
settl	0.0005660000

33

affair	0.0032279999
ginger	0.0026310000
sue	0.0017680000
reliabl	0.0009740000
commission	0.0009130000

raptor	0.0011210000	rai	0.0009110000
leas	0.0008650000	proceed	0.0009100000
compress	0.0007580000	studi	0.0008850000
litig	0.0006980000	deregul	0.0008360000
exhibit	0.0006860000	crisi	0.0008250000
construct	0.0006510000	palmer	0.0007880000
interconnect	0.0006220000	jean	0.0007340000
marin	0.0006120000	grid	0.0006800000
dissemin	0.0005950000	interconnect	0.0006370000
consent	0.0005710000	mitig	0.0006270000
dilig	0.0005400000	pat	0.0006240000
compens	0.0005380000	local	0.0006200000
sandi	0.0005260000	folk	0.0006040000
sold	0.0005190000	brother	0.0005470000
sound	0.0005030000	dub	0.0005260000
voic	0.0004980000	chair	0.0004870000
ow	0.0004870000	choic	0.0004830000
enforc	0.0004710000	oppos	0.0004780000
written	0.0004700000	thane	0.0004750000
damag	0.0004670000	bad	0.0004740000

34

temporari	0.0023940001
taffi	0.0019060000
hu	0.0018420001
holli	0.0017630000
van	0.0017350001
mailbox	0.0017070000
pat	0.0014060000
instruct	0.0014020000
maxwel	0.0013580000
lotu	0.0013500000
walk	0.0013180000
welcom	0.0013040000
sandi	0.0012180000
tweed	0.0011770000
desktop	0.0011670000
sander	0.0011540001
lanc	0.0011480000
snow	0.0011239999
evid	0.0010880000
walker	0.0010850000
sweet	0.0010530000
domest	0.0010460000

35

glover	0.0015940000
annex	0.0015870000
chase	0.0011300000
bear	0.0010430000
brokerag	0.0007810000
sack	0.0007580000
trust	0.0007550000
dissemin	0.0005990000
worksheet	0.0005890000
voic	0.0005690000
prime	0.0005550000
currenc	0.0005540000
consent	0.0005400000
pend	0.0005140000
clement	0.0005060000
rod	0.0005010000
stack	0.0004770000
simon	0.0004650000
loan	0.0004540000
regist	0.0004500000
subsidiari	0.0004490000
outstand	0.0004480000

setoff	0.0010260000	specifi	0.0004430000
foreign	0.0010250000	setoff	0.0004430000
secretari	0.0010140000	threshold	0.0004420000

36

congest	0.0015890000
interconnect	0.0014010000
panelist	0.0012550000
hourli	0.0011790000
commission	0.0009970000
er	0.0009710000
oasi	0.0009670000
panel	0.0009260000
reliabl	0.0008680000
dean	0.0008240000
fundament	0.0008120000
proceed	0.0008050000
main	0.0008030000
hydro	0.0007770000
known	0.0007740000
northeast	0.0007620000
imbal	0.0007460000
winter	0.0007370000
grid	0.0007340000
iri	0.0007330000
docket	0.0007230000
scan	0.0006900000
expans	0.0006800000
ancillari	0.0006080000
watt	0.0006050000

38

campaign	0.0024019999
newslett	0.0023560000
dial	0.0021629999
highlight	0.0020659999
pledg	0.0019499999
forum	0.0017580000
video	0.0017320000
bland	0.0011980000
achiev	0.0011840001
encourag	0.0011230001
beck	0.0010530000
conflict	0.0010120000

37

phase	0.0018310000
jai	0.0014880000
estat	0.0014860000
piiper	0.0010450000
beth	0.0010410000
factor	0.0010310001
domest	0.0008670000
rub	0.0008610000
salli	0.0008450000
brad	0.0008190000
jenni	0.0007790000
beck	0.0007590000
port	0.0007110000
infrastructur	0.0006760000
jean	0.0006400000
overview	0.0006330000
blackberri	0.0006290000
outlin	0.0006250000
chief	0.0006240000
owner	0.0005870000
mid	0.0005650000
core	0.0005340000
arriv	0.0005340000
fundament	0.0005270000
faith	0.0005200000

39

bass	0.0052780001
god	0.0024440000
life	0.0022499999
hull	0.0014180000
alwai	0.0013650000
kid	0.0012780000
bless	0.0012490000
mailbox	0.0011970000
nice	0.0011890000
heart	0.0011610000
footbal	0.0011570000
maria	0.0011350000

annual	0.0010050000	size	0.0011280000
hilli	0.0009880000	hand	0.0011089999
peer	0.0009820000	wed	0.0010900000
attende	0.0009700000	timothi	0.0010880000
hot	0.0009630000	prayer	0.0010740000
profil	0.0009490000	drink	0.0010720000
shadow	0.0009290000	tonight	0.0010530000
essenti	0.0009290000	bad	0.0010200000
host	0.0009280000	girl	0.0009940000
supervisor	0.0008870000	funni	0.0009800000
promot	0.0008760000	fun	0.0009620000
salli	0.0008570000	hei	0.0009180000
templat	0.0008530000	walk	0.0009120000

40

unifi	0.0114949998
director	0.0105109997
enterpris	0.0096290000
hardwar	0.0072730002
script	0.0049609998
logist	0.0048600002
stage	0.0040010000
setup	0.0032430000
solar	0.0029460001
menu	0.0025559999
bridg	0.0025520001
escal	0.0025010000
valid	0.0023310001
victor	0.0020969999
sat	0.0017580000
quantiti	0.0017560000
encount	0.0016899999
observ	0.0016200000
pinion	0.0016129999
midnight	0.0014940000
behavior	0.0014790000
abnorm	0.0014730000
volunt	0.0014700000
lab	0.0014640000
willi	0.0014390000

42

hide	0.0038739999
violat	0.0025200001

41

meter	0.0076460000
watch	0.0074330000
inactiv	0.0066920002
personnel	0.0066840001
diamond	0.0065130000
overlai	0.0052009998
brant	0.0049629998
merger	0.0036710000
farmer	0.0035210000
newli	0.0033239999
sort	0.0025620000
tab	0.0019620000
pat	0.0015710000
grave	0.0014060000
preliminari	0.0013130000
bottom	0.0011780000
wellhead	0.0011220000
column	0.0011190000
brows	0.0010190000
usernam	0.0009570000
fetch	0.0009540000
mar	0.0009530000
enabl	0.0008690000
thu	0.0008600000
expir	0.0008550000

43

garag	0.0091720000
bodi	0.0073130000

consolid	0.0016380000	shop	0.0065509998
captur	0.0013140000	garden	0.0060419999
palmer	0.0012900000	celebr	0.0054460000
oracl	0.0012460001	anniversari	0.0046629999
grove	0.0012320000	moodi	0.0043190001
robin	0.0012050000	pipe	0.0036700000
leg	0.0011930000	draw	0.0032790001
reev	0.0011640000	guest	0.0031659999
cross	0.0010960000	pledg	0.0025510001
dawn	0.0010810000	massag	0.0022430001
excel	0.0009820000	enjoi	0.0021559999
brad	0.0009590000	choos	0.0020590001
cad	0.0009190000	winner	0.0020440000
beech	0.0008550000	regist	0.0019620000
screen	0.0008440000	cocktail	0.0019160000
map	0.0007930000	bag	0.0018550000
mill	0.0007400000	registr	0.0018330000
grant	0.0006810000	spa	0.0018030000
stage	0.0006710000	attract	0.0017300000
wire	0.0006640000	bed	0.0016890001
bridg	0.0006590000	congratul	0.0016420000
pend	0.0006580000	seat	0.0016310000
label	0.0006290000	prize	0.0016180000

44

northwest	0.0040139998
southwest	0.0038540000
mid	0.0028540001
op	0.0026630000
grand	0.0022690000
path	0.0015120000
mitig	0.0011220000
tag	0.0010050000
grace	0.0009890000
cob	0.0008200000
rout	0.0007920000
newswir	0.0007160000
foster	0.0006680000
stage	0.0006620000
hourli	0.0006290000
ic	0.0006280000
hood	0.0006190000
holden	0.0005870000
heather	0.0005680000

45

video	0.0026920000
boardroom	0.0013780000
sherri	0.0013700000
task	0.0013090000
safeti	0.0012970000
remot	0.0011860000
peer	0.0011790000
sera	0.0010640000
palmer	0.0009580000
medium	0.0009440000
affair	0.0009110000
portal	0.0008950000
map	0.0008610000
sue	0.0008570000
causei	0.0008400000
rod	0.0007790000
derrick	0.0007450000
extend	0.0007220000
entir	0.0007220000

docket	0.0005640000	leadership	0.0007090000
meter	0.0005500000	billi	0.0007080000
ahead	0.0005490000	rice	0.0006600000
rai	0.0005130000	stark	0.0006340000
annuiti	0.0005030000	partner	0.0006330000
heat	0.0005020000	cliff	0.0005840000
	46		47
renew	0.0029380000	univers	0.0017720000
climat	0.0023320001	rice	0.0015960000
protocol	0.0019920000	student	0.0013400000
reduct	0.0018320000	iri	0.0011710000
complianc	0.0014149999	tropic	0.0010700000
bush	0.0013720000	grant	0.0010629999
pollut	0.0013440000	professor	0.0009450000
sip	0.0012910001	dale	0.0008670000
greenhaus	0.0012790000	engin	0.0008120000
coal	0.0011970000	volatil	0.0008090000
keeler	0.0010450000	studi	0.0007750000
assembl	0.0009560000	sincer	0.0007330000
sector	0.0009540000	valuat	0.0007070000
wind	0.0008380000	trip	0.0006610000
mechan	0.0008340000	speaker	0.0006600000
ey	0.0008210000	factor	0.0006580000
hydro	0.0007780000	storm	0.0006430000
mercuri	0.0007590000	quit	0.0006420000
heather	0.0007560000	mack	0.0006410000
dioxid	0.0007270000	seminar	0.0006100000
treati	0.0007220000	excel	0.0005890000
ring	0.0007220000	write	0.0005650000
adopt	0.0007220000	graduat	0.0005550000
congest	0.0007020000	correl	0.0005440000
carbon	0.0006960000	simul	0.0005390000

A.2 Author Topic with only Dictionary Words

	0		1
thi	0.012870999984443	confidenti	0.022670000791550
subject	0.012153999879956	privileg	0.019942000508308
pleas	0.010761000216007	intend	0.019226999953389
ar	0.010656000114977	copi	0.018130000680685
ani	0.008104000240564	notifi	0.017345000058413
forward	0.006415000185370	sander	0.016126999631524
ha	0.006035000085831	dissemin	0.015985999256372

call	0.005493999924511	delet	0.015341999940574
attach	0.005431000143290	sender	0.014918999746442
time	0.004991000052541	corp	0.014737999998033
wa	0.004912000149488	recipi	0.014657000079751
mai	0.004813000094146	mai	0.014134000055492
work	0.004627000074834	error	0.013973000459373
question	0.004463999997824	immedi	0.013629999943078
thei	0.004294999875128	agreement	0.013499000109732
inform	0.004071999806911	receiv	0.013368999585509
meet	0.003899999894202	inform	0.012795000337064
follow	0.003808999899775	prohibit	0.012201000005007
regard	0.003803000086918	document	0.010982999578118
corp	0.003748999908566	legal	0.009785000234842
discuss	0.003720999928191	letter	0.009693999774754
make	0.003635999979451	messag	0.009131000377238
file	0.003598999930546	notic	0.00899999612570
date	0.003524000057951	distribut	0.008829000405967
chang	0.003388999961317	forward	0.008787999860942

2

eb	0.075423002243042
north	0.063644997775555
smith	0.062176000326872
street	0.058253999799490
phone	0.058189999312162
corp	0.056554999202490
legal	0.052402999252081
pleas	0.029100999236107
depart	0.027504000812769
master	0.022791000083089
agreement	0.021756000816822
carol	0.018600000068545
attach	0.014321000315249
regard	0.013031000271440
draft	0.011726999655366
forward	0.010538999922574
credit	0.010142999701202
thi	0.010067000053823
send	0.009045000188053
confirm	0.008789000101388
execut	0.008264999836683
ani	0.007677999790758
subject	0.007191999815404
review	0.007180000189692

3

request	0.041074000298977
id	0.037032000720501
corp	0.030999999493361
review	0.030887000262737
pleas	0.030837999656796
click	0.030449999496341
asp	0.029705999419093
page	0.028217999264598
resourc	0.028154000639915
date	0.027927000075579
thi	0.026100000366569
type	0.025566000491381
approv	0.024240000173450
creat	0.022687999531627
act	0.021475000306964
ha	0.020245999097824
applic	0.019356999546289
access	0.014844999648631
pend	0.013713000342250
list	0.013616000302136
process	0.013309000059962
dai	0.013147000223398
receiv	0.012920999899507
complet	0.012807999737561

call	0.006988000124693	feedback	0.011189999990165
4		5	
mike	0.057608000934124	thi	0.047834001481533
subject	0.056949999183416	report	0.035441998392344
ar	0.043473999947309	link	0.035016000270844
wa	0.021746000275016	follow	0.034646000713110
work	0.014755000360310	pleas	0.034589000046253
today	0.012818999588490	employe	0.032145000994205
thi	0.012667999602854	expens	0.028108999133110
time	0.012631000019610	due	0.028023999184370
good	0.011823000386357	click	0.026971999555826
leav	0.009623000398278	approv	0.025096999481320
realli	0.009417000226676	total	0.024898000061512
nice	0.008232999593019	amount	0.023561999201775
week	0.008120000362396	chang	0.020322000607848
talk	0.007837999612093	statu	0.019838999956846
sorri	0.007480999920517	payment	0.017450999468565
dai	0.007311999797821	concur	0.015887999907136
becaus	0.007311999797821	paid	0.014382000081241
wai	0.006465999875218	review	0.013898000121117
tomorrow	0.006409000139683	account	0.013813000172377
thing	0.006409000139683	administr	0.011908999644220
lunch	0.006297000218183	send	0.011454000137746
call	0.006089999806136	readi	0.011084999889135
thought	0.005958000198007	inform	0.010289000347257
back	0.005940000060946	question	0.010033000260592
ye	0.005789000075310	complet	0.009748999960721
6		7	
group	0.015916999429464	pleas	0.009333999827504
ha	0.014368000440300	home	0.007546000182629
ani	0.014368000440300	center	0.007532000076026
ar	0.014007999561727	time	0.007354999892414
account	0.014007999561727	employe	0.007164000067860
thi	0.013791999779642	call	0.007081999909133
tax	0.013612000271678	build	0.007040999829769
suggest	0.013287999667227	avail	0.006672999821603
possibl	0.012856000103056	bring	0.006453999783844
time	0.012819999828935	check	0.006291000172496
doe	0.012388000264764	week	0.006209000013769
secur	0.012280000373721	ar	0.006058999802917
recommend	0.011343000456691	join	0.006045000161976
trade	0.011234999634326	american	0.005580999888480

commun	0.011126999743283	event	0.005472000222653
servic	0.011091000400484	inform	0.005404000170529
compani	0.010910999961197	program	0.005307999905199
made	0.010766999796033	find	0.004993999842554
loss	0.010731000453234	volunt	0.004898999817669
abov	0.010587000288069	dai	0.004898999817669
high	0.010514999739826	free	0.004858000203967
member	0.010443000122905	thi	0.004803000018001
short	0.010298999957740	opportun	0.004585000220686
question	0.010298999957740	fill	0.004557999782264
chang	0.010191000066698	present	0.004503000061959

8

pleas	0.049637999385595
question	0.043494999408722
ani	0.032467000186443
ar	0.024071000516415
thi	0.020741000771523
time	0.016493000090122
contact	0.016033999621868
inform	0.015974000096321
work	0.015316000208259
schedul	0.013003000058234
call	0.012803000397980
group	0.012086000293493
locat	0.011846000328660
particip	0.011447000317276
mai	0.010510000400245
regard	0.009812000207603
trade	0.009553000330925
ha	0.009413000196218
manag	0.009313000366092
process	0.008973999880254
assist	0.008895000442863
team	0.008715000003576
offic	0.008715000003576
avail	0.008415999822319
system	0.008177000097930

10

subject	0.042927000671625
book	0.033762998878956
love	0.032283999025822
pleas	0.025389999151230

9

busi	0.014410000294447
skill	0.014147999696434
lai	0.013553000055254
compani	0.013108000159264
year	0.012055999599397
ken	0.011416000314057
opportun	0.011370999738574
presid	0.009816999547184
chairman	0.009713999927044
join	0.009198999963701
veri	0.009165000170469
event	0.008764999918640
director	0.008514000102878
success	0.008457000367343
commun	0.007942000404000
great	0.007714000064880
develop	0.007701999973506
technolog	0.007371000014246
organ	0.007153999991715
dear	0.007017000112683
lead	0.006821999792010
execut	0.006719999946654
board	0.006159999873489
leader	0.006091000046581
manag	0.006068000104278

11

stock	0.006421000231057
invest	0.005491999909282
secur	0.004850000143051
compani	0.004741000011563

forward	0.019432999193668	report	0.004672000184655
thi	0.019204000011086	bui	0.004672000184655
ar	0.015600999817252	advertis	0.004612999968231
deal	0.015162999741733	make	0.004523999989033
frank	0.013829999603331	provid	0.004404999781400
file	0.013621999882162	inform	0.004335999954492
ani	0.012850999832153	market	0.004325999878347
question	0.012684999965131	complet	0.004317000042647
white	0.012206000275910	ani	0.004197999835014
corp	0.011893000453711	research	0.004129000008106
posit	0.011851999908686	thi	0.004019999876618
report	0.011621999554336	sale	0.004000999964774
robin	0.011517999693751	sell	0.003980999812484
curv	0.011164000257850	onli	0.003901999909431
desk	0.010580999776721	give	0.003892000066116
risk	0.010247999802232	subscrib	0.003753999946639
reev	0.010205999948084	mai	0.003723999951035
west	0.009476999752223	alert	0.003704000031576
list	0.009331000037491	gener	0.003693999955431
trader	0.009247999638319	updat	0.003645000047982
east	0.008561000227928	technolog	0.003624999895692

12

trade	0.020122999325395
copi	0.015713000670075
agreement	0.012598999775946
receiv	0.012272000312805
execut	0.012261000461876
click	0.011509999632835
credit	0.011335999704897
swap	0.011160999536514
return	0.010857000015676
note	0.010377000086010
transact	0.010355999693274
abov	0.010050999931991
link	0.009909000247717
financi	0.009767999872565
open	0.009527999907732
indic	0.009103000164032
exchang	0.009060000069439
quot	0.008940000087023
document	0.008896999992430
text	0.008603000082076
web	0.008438999764621

13

derrick	0.031964998692274
attach	0.019671000540257
pleas	0.018203999847174
call	0.015150000341237
ha	0.015070999972522
request	0.014158000238240
held	0.013127000071108
address	0.012254999950528
wall	0.012214999645948
date	0.011858000420034
content	0.011779000051320
reason	0.011660000309348
file	0.011381999589503
believ	0.010986000299454
wa	0.010549000464380
mai	0.010470000095665
thi	0.010431000031531
requir	0.010231999680400
send	0.009955000132322
rob	0.009875000454485
prior	0.009875000454485

address	0.008407000452280	server	0.009557999670506
confirm	0.008275999687612	center	0.009557999670506
notic	0.007981999777257	releas	0.009479000233114
browser	0.007960000075400	system	0.009398999623954

14

travel	0.007017999887466
citi	0.006721999961883
visit	0.006647000089288
ar	0.006043999921530
hotel	0.005853999871761
onli	0.005768999923021
reserv	0.005642000120133
avail	0.005524999927729
includ	0.005408999975771
great	0.005303000099957
make	0.005282000172883
time	0.005187000147998
night	0.005069999955595
thi	0.005028000101447
mai	0.004985999781638
flight	0.004900999832898
special	0.004815999884158
fare	0.004732000175864
book	0.004710000008345
chang	0.004647000227123
stai	0.004625999834388
find	0.004509000107646
san	0.004488000180572
airlin	0.004488000180572
car	0.004466999787837

16

report	0.009363000281155
explor	0.009343000128865
free	0.008847000077367
imag	0.006407000124454
oper	0.006308000069112
download	0.006109999958426
asp	0.005812000017613
control	0.005475000012666
compani	0.005394999869168
energi	0.005078000016510
stori	0.005038000177592

15

follow	0.028179999440908
ani	0.027309000492096
messag	0.026203000918031
avail	0.024829000234604
space	0.023454999551177
current	0.022886000573635
make	0.022717999294400
inform	0.022484000772238
limit	0.022383000701666
item	0.021578999236226
reach	0.021445000544190
client	0.021410999819636
receiv	0.021176999434829
automat	0.020975999534130
ar	0.020607000216842
todai	0.020573999732733
longer	0.020338999107480
thi	0.020004000514746
individu	0.019702000543475
turn	0.019501000642776
date	0.019166000187397
select	0.018664000555873
button	0.018596999347210
folder	0.018463000655174
option	0.017993999645114

17

subject	0.023399999365211
wa	0.022918999195099
thi	0.021668000146747
time	0.019186999648809
ar	0.018417999148369
work	0.016301000490785
dai	0.013709999620914
back	0.012466000393033
good	0.012307999655604
thei	0.012191000394523
make	0.012164000421762

announc	0.004999000113457	talk	0.011690000072122
dai	0.004641999956220	call	0.010013000108302
market	0.004621999803931	sai	0.009984999895096
map	0.004621999803931	give	0.009937000460923
releas	0.004523000214249	veri	0.009820999577641
top	0.004482999909669	thing	0.009374000132084
opportun	0.004482999909669	week	0.009188000112772
manag	0.004323999863118	wai	0.008577000349760
intl	0.004323999863118	today	0.008280999958515
engin	0.004304999951273	befor	0.007689999882132
industri	0.004244999960065	thought	0.007621000055224
design	0.004205000121146	hear	0.007242999970913
search	0.004166000057012	long	0.007153999991715
ga	0.004145999904722	anyth	0.006728000007570

18

pipelin	0.030850000679493
millier	0.023903999477625
capac	0.023812999948859
drew	0.020997999235988
harri	0.020611999556422
rod	0.020181000232697
ga	0.018908999860287
transport	0.015641000121832
blackberri	0.015186999924481
wireless	0.014506000094116
shipper	0.014210999943316
rate	0.012395000085235
terri	0.012213000096381
station	0.010646999813616
net	0.010125000029802
subject	0.0100560000202894
glen	0.009557000361383
rick	0.009488999843597
randi	0.009103000164032
oper	0.009034999646246
northern	0.008808000013232
estim	0.007672999985516
maria	0.007581999991089
expans	0.007490999996662
moor	0.007400000002235

20

salli	0.062577001750469
-------	-------------------

19

note	0.017610000446439
thi	0.016714999452233
call	0.014925000257790
receiv	0.014894000254571
becaus	0.013136000372469
time	0.013104000128806
ar	0.012591999955475
train	0.012432999908924
user	0.012304999865592
center	0.012144999578595
password	0.011570000089705
address	0.011505999602377
payabl	0.011122000403702
forward	0.010994000360370
materi	0.010579000227153
messag	0.010418999940157
launch	0.010162999853492
system	0.009684000164270
action	0.009428000077605
import	0.009236999787390
site	0.008852999657393
dean	0.008821000345051
link	0.008756999857724
box	0.008628999814391
invoic	0.008373999968171

21

thi	0.012626999989152
-----	-------------------

beck	0.050239000469446	ar	0.010595999658108
beth	0.019338000565767	inform	0.009475000202656
meet	0.018765000626445	reserv	0.008402000181377
team	0.016669999808073	post	0.007996000349522
offic	0.013628999702632	send	0.007703000213951
manag	0.013457000255585	chang	0.007476000115275
rick	0.013083999976516	onli	0.007280999794602
week	0.012969000265002	market	0.006484000012279
hall	0.011734999716282	commun	0.006126999855042
oper	0.011591999791563	price	0.006093999836594
patti	0.011447999626398	free	0.005882999859750
risk	0.010731000453234	year	0.005801999941468
busi	0.009754999540746	current	0.005721000023186
bob	0.009267999790609	address	0.005541999824345
work	0.009181999601424	portfolio	0.005508999805897
ted	0.009066999889910	group	0.005444000009447
white	0.009038000367582	talk	0.005396000109613
plan	0.008349999785423	time	0.005152000114322
global	0.008120000362396	question	0.005119000095874
causei	0.007803999818861	copi	0.005053999833763
congratul	0.007604000158608	board	0.004989000037313
attend	0.007431000005454	compani	0.004972999915481
commun	0.007344999816269	latest	0.004941000137478
role	0.007288000080734	call	0.004858999978751

22

attach	0.082843996584415
tom	0.051456000655890
mark	0.044238001108170
subject	0.041002999991179
forward	0.040281999856234
approv	0.035863999277353
sever	0.033264998346567
lambert	0.032283000648022
global	0.030897000804543
lee	0.030203999951482
pleas	0.029510999098420
brant	0.026652000844479
corp	0.025295000523329
file	0.024717999622226
credit	0.024284999817610
hardi	0.018856000155210
frank	0.017788000404835
regard	0.016141999512911

23

subject	0.050204001367092
kitchen	0.041841000318527
laura	0.024723000824451
martin	0.024400999769568
hunter	0.022265000268817
pleas	0.021023999899626
don	0.017485000193119
adam	0.016268000006676
black	0.016061000525951
mark	0.015946000814438
mike	0.014383999630809
meet	0.013143000192940
fletcher	0.011924999766052
cell	0.011741000227630
room	0.011556999757886
assist	0.011420000344515
frank	0.010890999808908
jai	0.010339999571443

hare	0.016141999512911	east	0.009995000436902
center	0.015015999786556	trade	0.009948999620974
bill	0.014697999693453	offic	0.009948999620974
harri	0.011926000006497	peopl	0.009856999851763
amber	0.011753000319004	brad	0.009603999555111
call	0.011060000397265	attend	0.009603999555111
contract	0.010394999757409	forward	0.009580999612808

24

subject	0.051633998751640
thi	0.038672000169754
pleas	0.036132000386715
ar	0.031170999631286
ani	0.030457999557257
ward	0.028250999748707
ga	0.024926999583840
file	0.018493000417948
call	0.016000999137759
august	0.015977000817657
attach	0.015620999969542
thei	0.015596999786794
send	0.012226000428200
wa	0.011703999713063
question	0.010635999962687
price	0.010635999962687
offic	0.009995000436902
address	0.008379999548197
sheet	0.008262000046670
receiv	0.008190000429749
today	0.007857999764383
work	0.007739000022411
daili	0.007431000005454
gui	0.007406999822706
place	0.007360000163317

26

time	0.066546000540257
date	0.066001996397972
calendar	0.064593002200127
detail	0.061627000570297
central	0.057250998914242
entri	0.056212998926640
standard	0.054186001420021
descript	0.053913999348879

25

continu	0.010474000126123
ha	0.009975000284612
addit	0.009701999835670
provid	0.009654999710619
busi	0.009405000135303
relat	0.008608999662101
includ	0.008430999703705
issu	0.008146000094712
compani	0.007849000394344
current	0.007778000086546
oper	0.007588000036776
manag	0.007220000028610
work	0.006579000037163
respons	0.006413000170141
requir	0.006388999987394
effect	0.006293999962509
activ	0.005853999871761
servic	0.005758999846876
inform	0.005617000162601
announc	0.005427000112832
employe	0.005307999905199
ar	0.005260999780148
develop	0.005011000204831
effort	0.004964000079781
market	0.004798000212759

27

subject	0.026696000248194
thi	0.022600000724196
ar	0.021606000140309
bass	0.019857000559568
wa	0.015150999650359
game	0.011570000089705
night	0.009868999943137
good	0.008863000199199

team	0.050923001021147	corp	0.008562999777496
migrat	0.046571999788284	net	0.007389999926090
outlook	0.045880001038313	pleas	0.007354000117630
chairperson	0.044124998152256	weekend	0.007149999961257
appoint	0.042369998991489	work	0.007125999778509
eb	0.017575999721885	messag	0.007017999887466
white	0.012730999849737	ani	0.006767000071704
room	0.011198000051081	origin	0.006742999888957
meet	0.011198000051081	gui	0.006599000189453
confer	0.010827999562025	back	0.006432000081986
offic	0.006254000123590	plai	0.00635999999404
remind	0.005363999865949	make	0.006180000025779
invit	0.005018000025302	call	0.006107999943197
vacat	0.004300999920815	dai	0.006023999769241
weekli	0.003880999982357	mai	0.005857000127435
dai	0.003683000104502	love	0.005857000127435
staff	0.003585000056773	forward	0.005809000227600

28

access	0.022673999890685
servic	0.021262999624014
user	0.020677000284195
inform	0.015816999599338
system	0.015708999708295
contact	0.01568700006437
manag	0.014494000002742
server	0.013907999731600
thi	0.013647999614477
locat	0.012736000120640
follow	0.012129000388086
data	0.011998999863863
pager	0.010804999619722
desk	0.010719000361860
network	0.010675000026822
outag	0.010111000388861
center	0.010111000388861
databas	0.009785999543965
password	0.009634000249207
schedul	0.009611999616027
messag	0.009569000452757
assist	0.009135000407696
onli	0.008960999548435
support	0.008852999657393
custom	0.008549000136554

29

thi	0.012233000248671
industri	0.010920000262558
pleas	0.010819000191987
visit	0.010688999667764
inform	0.010630999691784
view	0.010472999885678
click	0.010356999933720
power	0.009375999681652
daili	0.008468000218272
market	0.008352000266314
energi	0.007703000213951
complet	0.00740000002235
receiv	0.007083000149578
ga	0.007040000054985
edit	0.006635999772698
subscript	0.006606999784708
includ	0.006548999808729
avail	0.006347000133246
price	0.006275000050664
center	0.006275000050664
polici	0.005812999792397
web	0.005510999821126
show	0.005437999963760
analysi	0.005309000145644
newslett	0.005264999810606

30

research 0.023004999384284
 dear 0.018687000498176
 univers 0.017478000372648
 school 0.010286999866366
 model 0.009417000226676
 program 0.009372999891639
 resum 0.008887000381947
 interest 0.008635999634862
 present 0.008561999537051
 manag 0.008518000133336
 financ 0.008445000275970
 sincer 0.008237999863923
 mail 0.008046999573708
 student 0.007176999934018
 web 0.007133000064641
 veri 0.006794000044465
 interview 0.006647000089288
 risk 0.006632000207901
 write 0.006411000154912
 graduat 0.006411000154912
 technolog 0.006337000057101
 professor 0.005762000102550
 visit 0.005688999779522
 confer 0.005630000028759
 paper 0.005615000147372

31

thi 0.017359999939799
 enjoi 0.016250999644399
 person 0.016219999641180
 ani 0.015807999297976
 end 0.013907000422478
 ar 0.013811999931931
 note 0.012101000174880
 howev 0.011880000121891
 mai 0.011531000025570
 contact 0.011531000025570
 servic 0.011277999728918
 address 0.011246000416577
 follow 0.010866000317037
 messag 0.010644000023603
 doe 0.010613000020385
 directli 0.010421999730170
 awai 0.010421999730170
 time 0.010359000414610
 addit 0.010359000414610
 match 0.009757000021636
 work 0.009566999971867
 site 0.009534999728203
 remain 0.009503999724984
 line 0.009472000412643
 real 0.009409000165761

32

click 0.015622000209987
 thi 0.014843000099063
 free 0.014271999709308
 pleas 0.013267000205815
 ar 0.012063000351191
 receiv 0.010755999945104
 messag 0.010738000273705
 mail 0.010174999944866
 offer 0.010027999989688
 remov 0.008867999538779
 imag 0.007751000113785
 net 0.006988999899477
 ani 0.00658999988092
 enter 0.006097000092268
 credit 0.005828000139445

33

offer 0.025985000655055
 repli 0.025698000565171
 sender 0.025150999426842
 sole 0.025085000321269
 prohibit 0.025014000013471
 recipi 0.024856999516487
 strictli 0.024512000381947
 anyon 0.024445999413729
 receiv 0.024374999105930
 affili 0.024133000522852
 intend 0.024023000150919
 accept 0.023860000073910
 confidenti 0.023819999769330
 materi 0.023755000904202
 disclosur 0.023695999756455

wa	0.005820000078529	distribut	0.023656999692321
onli	0.005603000055999	delet	0.023532999679446
time	0.005594000220299	reli	0.023376999422908
id	0.005586000159383	properti	0.023050999268889
servic	0.005464000161737	bind	0.023031000047922
dai	0.005456000100821	creat	0.023025000467896
todai	0.005386999808252	otherwis	0.023011999204755
provid	0.005160999950022	evid	0.022952999919653
guarante	0.005135000217706	administr	0.022940000519156
address	0.005100999958813	hereto	0.022933999076486

34

market	0.025608999654651
price	0.015150000341237
trade	0.013712000101805
year	0.011126999743283
product	0.009843000210822
month	0.008558999747038
week	0.008524999953806
sell	0.007156000006944
end	0.006864999886602
demand	0.0068479998998986
expect	0.006659000180662
produc	0.006591000128537
term	0.006401999853551
industri	0.006300000008196
bui	0.006231000181288
ga	0.006128999870270
natur	0.006008999887854
high	0.005752000026405
short	0.005528999958187
suppli	0.005375000182539
pulp	0.005375000182539
level	0.005340999923646
fundament	0.005307000130415
spread	0.005272999871522
weather	0.005135999992490

36

made	0.014774999581277
thi	0.014592000283301
fund	0.014581999741495
befor	0.014550000429153
plan	0.014538999646902

35

god	0.012435000389814
bless	0.010075000114739
send	0.009266000241041
thi	0.008414000272751
life	0.007474000100046
address	0.007277000229806
list	0.007168000098318
ar	0.007081000134349
thei	0.006643999833614
provid	0.006380999926478
peopl	0.006359999999404
read	0.006337999831885
faith	0.006293999962509
love	0.006271999794990
prayer	0.006184999831021
believ	0.006184999831021
todai	0.006118999794126
daili	0.005878999829292
simpli	0.005857000127435
becaus	0.005834999959916
lord	0.005681999959052
person	0.005594999995083
king	0.005594999995083
wai	0.005551000125706
onli	0.005506999790668

37

ga	0.041397001594305
deal	0.032260999083519
volum	0.030396999791265
subject	0.0290869999595165
farmer	0.022848000749946

effort	0.014484999701381	dai	0.020541999489069
time	0.014464000239968	thi	0.019508000463247
mani	0.014441999606788	forward	0.018068000674248
year	0.014324000105262	pleas	0.017015999183059
ken	0.014324000105262	month	0.01601999980927
reach	0.014259999617934	meter	0.015465999953449
set	0.014228000305593	nomin	0.013565000146627
pleas	0.014228000305593	flow	0.012919000349939
energi	0.014228000305593	deliveri	0.012586999684572
stock	0.014217000454664	contract	0.011350999586284
dure	0.014217000454664	transport	0.010649000294507
monei	0.014174000360072	price	0.010409000329673
lost	0.014131000265479	chang	0.010114000178874
compani	0.014131000265479	invoic	0.009874000214040
pai	0.014120999723673	show	0.008859000168741
state	0.014099000021815	fuel	0.008120999671519
net	0.014088000170887	wa	0.007935999892652
lai	0.014077999629080	corp	0.007880999706686
report	0.014046000316739	ticket	0.007732999976724
million	0.014046000316739	march	0.007695999927819

38

click	0.015893999487162
receiv	0.011936999857426
imag	0.011342000216246
onlin	0.010138999670744
link	0.009661000221968
dear	0.009216999635100
visit	0.009065999649465
custom	0.009022999554873
servic	0.008755000308156
offer	0.008592000231147
special	0.008078999817371
privaci	0.007786000147462
repli	0.006626000162214
address	0.006436000112444
free	0.006246999837458
featur	0.006225000135601
account	0.006202999968082
subscrib	0.006169000174850
regist	0.006091000046581
prefer	0.005764000117779
order	0.005725000053644
list	0.005681999959052

39

forward	0.009569000452757
man	0.007660000119358
friend	0.006275000050664
net	0.006033000070602
sai	0.005456000100821
ey	0.005214999895543
woman	0.005204000044614
walk	0.004794999957085
men	0.004637999925762
life	0.004637999925762
stop	0.004375999793410
littl	0.004271000158042
pass	0.004155000206083
door	0.004081999883056
word	0.003914000000805
women	0.003893000073731
care	0.003871999913827
smile	0.003724999958649
world	0.003704000031576
hand	0.003598999930546
live	0.003535999916494
everi	0.003483999986202

web	0.005656000226736	ring	0.003451999975368
select	0.005561000201851	stand	0.003441999899223
messag	0.005198999773711	turn	0.003409999888390
40		41	
pleas	0.011548000387847	park	0.038644999265671
travel	0.010681999847293	subject	0.017247999086976
thi	0.010681999847293	thi	0.013098999857903
matt	0.010413000360131	march	0.012962999753654
ticket	0.010297000408173	ar	0.012176999822259
servic	0.010123999789357	net	0.010115999728441
avail	0.009970000013709	origin	0.009572999551892
smith	0.009930999949574	pleas	0.008949999697506
ar	0.009200000204146	back	0.008759999647737
flight	0.008795999921858	messag	0.006889000069350
call	0.008679999969900	good	0.006860999856144
hotel	0.008352999575436	man	0.006806999910623
fare	0.008198999799788	hunt	0.006725999992341
book	0.007794999983162	night	0.006372999865562
number	0.007410000078380	wa	0.006318999920040
subject	0.007294999901205	work	0.006020999979228
airlin	0.007063999772072	wai	0.005640999879688
free	0.007044000085443	deer	0.005613999906927
mai	0.006833000108600	send	0.005586999934167
itinerari	0.006833000108600	date	0.005396999884397
airport	0.006736000068486	friend	0.005233999807388
chang	0.006525000091642	men	0.005125999916345
arriv	0.006525000091642	ani	0.005125999916345
cancel	0.006370999850333	mai	0.005044000223279
mat	0.006312999874353	call	0.004962999839336
42		43	
veri	0.006866000127047	electr	0.018485000357032
sai	0.006401000078768	commiss	0.016550999134779
presid	0.006248999852687	state	0.016462000086904
seller	0.006066000089049	util	0.015412000007927
million	0.005644000135362	energi	0.012106000445783
past	0.005297999829054	public	0.011222000233829
greet	0.005059999879450	legisl	0.011056000366807
folk	0.004801000002772	regulatori	0.009584999643266
billion	0.004801000002772	senat	0.009043999947608
invest	0.004757000133395	feder	0.008933000266552
perfect	0.004648999776691	affair	0.008689999580383
san	0.004552000202239	regul	0.008523999713361

economi	0.004271000158042	governor	0.008446999825537
alert	0.004087000153959	propos	0.008070999756455
direct	0.003990000113845	sue	0.007938000373542
bond	0.003945999778807	consum	0.007915999740362
copyright	0.003902999917045	gener	0.006390000227839
wrote	0.003892000066116	committe	0.006157999858260
econom	0.003892000066116	rate	0.006014000158757
cut	0.003654999891296	deregul	0.006002999842167
thing	0.003502999898046	press	0.005958999972790
stori	0.003492000047117	vote	0.005638999864459
big	0.003492000047117	rule	0.005594000220299
year	0.003481999970973	commission	0.005539000034332
share	0.003470999887213	bill	0.005516999866813

44

cooper	0.012365999631584
home	0.007422999944538
hous	0.006802999880165
small	0.005903000012040
water	0.005042000208050
food	0.005001999903470
tast	0.004602000117302
cook	0.004422000143677
land	0.00388199989972
sat	0.003682000096887
bottl	0.003522000042722
minut	0.003441999899223
serv	0.003421999979764
qualiti	0.003421999979764
front	0.003402000060305
garden	0.003381999908015
fresh	0.003342000069097
wind	0.003282000077888
build	0.003222000086680
onli	0.003201999934390
high	0.003122000023723
weather	0.003102000104263
american	0.003062000032514
thing	0.003021999960765
red	0.003021999960765

46

thi	0.009634000249207
footbal	0.009565000422299

45

love	0.020480999723077
night	0.018602000549436
hei	0.018373999744654
good	0.018340999260545
hope	0.018232999369502
great	0.016542999073863
home	0.014785000123084
weekend	0.014737999998033
realli	0.013135000132024
fun	0.012249999679625
thing	0.010390999726951
littl	0.010277000255883
friend	0.010142999701202
pretti	0.008995999582112
nice	0.008620000444353
anywai	0.008453000336885
hous	0.008438999764621
big	0.008109999820590
sound	0.007956000044942
dinner	0.007922999560833
sorri	0.007761999964714
gui	0.007613999769092
someth	0.007581000216305
leav	0.007513000164181
tonight	0.007493000011891

47

deal	0.047152001410723
chang	0.022467000409961

ar	0.008351000025868	subject	0.021685000509024
game	0.007868000306189	thi	0.020232999697328
season	0.007302000187337	ha	0.018924999982119
week	0.006914999801666	confirm	0.015095000155270
player	0.006500999908894	bui	0.013435999862850
team	0.006115000229329	check	0.013388000428677
plai	0.006060000043362	price	0.013244000263512
good	0.005549000110477	power	0.012924999929965
pleas	0.005259000230581	trade	0.012717000208795
ha	0.005135000217706	peak	0.012621999718249
fantasi	0.005107000004500	sell	0.011695999652147
start	0.005038000177592	enter	0.011695999652147
year	0.004954999778420	ar	0.011536999605596
onli	0.004817000124604	pleas	0.011505000293255
sign	0.004803000018001	miss	0.011328999884427
dai	0.004445000085980	show	0.010308000259101
wa	0.004348000045866	broker	0.009573999792337
score	0.004306999966502	bill	0.009494000114501
run	0.004265000112355	deliveri	0.008553000167012
big	0.004224000032991	time	0.008392999880016
pick	0.004209999926388	follow	0.008298000320792
save	0.004112999886274	point	0.008266000077128
earli	0.003905999939889	hourli	0.008217999711633

A.3 PLSI-U with all Words (No Dictionary)

0	1		
governor	0.0031310001	dbcap	0.0364030004
calpin	0.0030650001	cpuc	0.0067199999
iep	0.0027960001	pge	0.0057259998
dasovich	0.0025650000	workshop	0.0054040002
edison	0.0025589999	ofc	0.0052580000
gov	0.0024619999	testimoni	0.0051150001
iepa	0.0020570001	socalga	0.0050180000
duke	0.0019930000	gov	0.0045220000
mara	0.0019490001	tp	0.0041040001
kaplan	0.0018820000	exhibit	0.0037829999
cpuc	0.0018240000	accord	0.0037130001
billion	0.0016800000	ii	0.0033090001
dynegi	0.0016470000	oii	0.0030650001
legisl	0.0016200000	tariff	0.0030520000
sce	0.0015350000	sempra	0.0029180001
press	0.0015160000	sce	0.0028009999
smutni	0.0014510000	gmssr	0.0024309999

assembl	0.0013710000	tblload	0.0022940000
ed	0.0013510000	dynegi	0.0022460001
pge	0.0013420000	comprehens	0.0021650000
paula	0.0013120000	core	0.0019729999
jack	0.0012980000	aelaw	0.0019660001
kent	0.0012840000	tblintchg	0.0019290000
jan	0.0012830000	counihan	0.0017560000
kati	0.0012660000	calpin	0.0017410000
2		3	
bpa	0.0085990001	wpd	0.0069550001
gov	0.0064329999	swap	0.0053249998
produc	0.0054600001	pulp	0.0033330000
iep	0.0043360000	forestweb	0.0032690000
pngc	0.0036889999	tonn	0.0032339999
disclos	0.0030189999	paper	0.0029889999
describ	0.0027900001	div	0.0029470001
wp	0.0027230000	trust	0.0027520000
ds	0.0024699999	agmt	0.0027230000
cpuc	0.0021960000	facil	0.0024860001
mr	0.0021929999	andrew	0.0023510000
written	0.0020669999	shackleton	0.0023480000
purpos	0.0019120000	taf	0.0021939999
caiso	0.0018810000	digest	0.0020830000
ci	0.0018720001	userrefer	0.0019540000
ident	0.0018670000	produc	0.0019110000
testimoni	0.0018600000	kurth	0.0018840000
peter	0.0018330000	raptor	0.0018020000
occur	0.0017980000	mcgarret	0.0017659999
jcg	0.0017320000	monika	0.0017610000
crac	0.0017140000	hawaii	0.0016850000
kaplan	0.0016680000	ii	0.0015610000
repres	0.0015870000	certif	0.0015580000
oral	0.0015670001	notifi	0.0015510001
text	0.0015519999	top	0.0015470000
4		5	
haa	0.0040330002	bid	0.0224110000
berkelei	0.0039470000	northwest	0.0058650002
msn	0.0032269999	shipper	0.0058610002
ewc	0.0023429999	button	0.0044359998
mwe	0.0016710000	screen	0.0041089999
chui	0.0016340000	packag	0.0037710001
lng	0.0016320000	mainten	0.0036970000

telex	0.0016100000	submit	0.0035140000
inmarsat	0.0016100000	dth	0.0034159999
swbell	0.0015610000	station	0.0031870001
averag	0.0013670000	invoic	0.0029480001
york	0.0013610000	electron	0.0027739999
hoegh	0.0012180000	nation	0.0025120000
chron	0.0011660000	pacif	0.0024480000
tk	0.0011089999	referenc	0.0023860000
expedia	0.0011010000	receipt	0.0022789999
fare	0.0010470001	replac	0.0022239999
rob	0.0010430000	ebb	0.0021380000
bank	0.0010410000	maximum	0.0020780000
tx	0.0010190000	prearrang	0.0020039999
famili	0.0010090000	cap	0.0018190000
trip	0.0010010001	repres	0.0017930000
rr	0.0009890000	safeti	0.0017890000
consum	0.0009250000	statement	0.0017690000
explor	0.0009240000	central	0.0017680000

6

7

tropic	0.0020830000	unifi	0.0065009999
weather	0.0018130000	sap	0.0041230000
shackleton	0.0016160000	netco	0.0034830000
folder	0.0014930000	sitara	0.0028359999
brazil	0.0014850000	script	0.0026769999
brent	0.0014780000	class	0.0023139999
econom	0.0014720000	setup	0.0022260000
synchron	0.0012660000	path	0.0022040000
low	0.0012400000	cst	0.0021430000
lynn	0.0012370000	regan	0.0021170001
kean	0.0012309999	logist	0.0020570001
educ	0.0012070000	prior	0.0019970001
storm	0.0011970000	scenario	0.0019670001
bruce	0.0010990000	edi	0.0018660000
penn	0.0010900000	server	0.0018170000
central	0.0010850000	invoic	0.0017910000
condit	0.0010720000	integr	0.0017420000
andrea	0.0010260000	tammi	0.0017240000
environment	0.0010220000	terri	0.0016880000
hendri	0.0009980000	wade	0.0016070000
dynegi	0.0009900000	entri	0.0015760000
exist	0.0009730000	estat	0.0015130000
upper	0.0009700000	login	0.0015010000
wind	0.0009460000	stage	0.0014640000

hold	0.0009440000	patti	0.0014330000
	8		9
tana	0.0040020002	children	0.0026499999
ew	0.0032120000	timesheet	0.0022060000
temp	0.0026060001	minut	0.0017300000
cook	0.0026030000	lara	0.0016300000
nda	0.0022900000	janel	0.0015700000
greenberg	0.0021869999	stephen	0.0015160000
carol	0.0020099999	dub	0.0014820000
folder	0.0019080000	lexi	0.0014780000
swap	0.0017980000	maureen	0.0014759999
peter	0.0017860000	mcvicker	0.0014290001
wholesal	0.0016600000	session	0.0014180000
left	0.0016430001	michel	0.0014149999
clair	0.0015930000	sue	0.0014080000
senior	0.0015580000	class	0.0013600000
add	0.0013630000	sa	0.0012940000
lesli	0.0012480000	frank	0.0012800000
enrononlin	0.0012230000	leibman	0.0012350000
temporari	0.0012080000	kean	0.0012030000
calendar	0.0011820000	sarah	0.0011830000
bank	0.0011549999	interview	0.0011820000
provis	0.0011130000	palmer	0.0011730000
nymex	0.0010939999	sfo	0.0011390001
brent	0.0010720000	hu	0.0011370000
heard	0.0010250000	sylvia	0.0010990000
shackleton	0.0009950000	elizabeth	0.0010990000
	10		11
rod	0.0053349999	prc	0.0030720001
hayslett	0.0039769998	video	0.0023230000
traci	0.0039340002	weekli	0.0020699999
eott	0.0032990000	ken	0.0017710000
geaccon	0.0028150000	dial	0.0016400000
stan	0.0022849999	kean	0.0015180000
tw	0.0020160000	cindi	0.0015150000
nng	0.0018480000	vp	0.0014080000
saunder	0.0017340000	passcod	0.0014060000
fgt	0.0015300000	brown	0.0013940000
horton	0.0013330000	lai	0.0013890000
capit	0.0013290000	recommend	0.0013330000
howard	0.0013200000	joanni	0.0013020000
model	0.0012300001	whallei	0.0012710000

blackberri	0.0012200000	boardroom	0.0012540000
danni	0.0011390001	delainei	0.0012400000
budget	0.0011200000	agenda	0.0011950000
clear	0.0011140000	kathi	0.0011940000
expens	0.0011120000	connect	0.0011040000
februari	0.0010900000	memo	0.0010900000
cindi	0.0010500000	conf	0.0010830000
nymex	0.0010280000	derrick	0.0010830000
chandler	0.0010190000	establish	0.0010690000
jr	0.0010140000	summer	0.0010610000
alloc	0.0009270000	peer	0.0010400000

12

gtc	0.0027660001
deriv	0.0027160000
justin	0.0026010000
enrononlin	0.0024029999
weather	0.0018780000
refer	0.0017940000
japan	0.0016830000
forster	0.0016139999
bank	0.0013780000
swap	0.0013670000
boyd	0.0013630000
exchang	0.0013180000
metal	0.0013079999
commod	0.0012730000
descript	0.0012640000
korea	0.0012380000
edmund	0.0012270000
uk	0.0012140000
alan	0.0012070000
determin	0.0011380001
ap	0.0011200000
minn	0.0011070000
japanes	0.0011050000
tana	0.0010950000
louis	0.0010300000

14

cera	0.0037580000
tw	0.0024150000
facil	0.0016230000
el	0.0015340000

13

enrononlin	0.0162870008
notion	0.0083219996
averag	0.0066809999
summari	0.0063689998
extern	0.0049439999
avg	0.0042530000
trail	0.0042409999
gross	0.0037420001
life	0.0035140000
intranet	0.0031000001
weekli	0.0027520000
outsid	0.0027210000
feedback	0.0025899999
ensur	0.0025170001
launch	0.0022239999
approxim	0.0021490001
directli	0.0019759999
billion	0.0019050000
window	0.0018310000
kathi	0.0017620000
embed	0.0016340000
revenu	0.0016000000
experi	0.0015590000
onc	0.0015550000
difficulti	0.0015280000

15

socal	0.0037839999
cera	0.0022060000
bond	0.0017950000
bankruptci	0.0016780000

shipper	0.0014600001	phillip	0.0016400000
paso	0.0014330000	fund	0.0015300000
station	0.0013450000	pira	0.0014299999
winter	0.0013250000	research	0.0013379999
transwestern	0.0012550000	revenu	0.0012420000
eott	0.0011970000	allen	0.0012120000
bcf	0.0011770000	pdf	0.0011399999
drew	0.0011440000	grigsbi	0.0011150000
nysrc	0.0010910000	mou	0.0010430000
permit	0.0010690000	el	0.0010180000
northern	0.0010450000	qf	0.0010130000
expans	0.0010290000	legisl	0.0009900000
larri	0.0009780000	assembl	0.0009780000
fossu	0.0009780000	paso	0.0009730000
miller	0.0009670000	bc	0.0009600000
announc	0.0009630000	xx	0.0009430000
season	0.0009630000	vote	0.0009360000
mmcf	0.0009580000	governor	0.0009240000
averag	0.0009480000	legislatur	0.0008900000
western	0.0009360000	expens	0.0008850000
ot	0.0009140000	burton	0.0008840000

16

analyst	0.0173599999
rotat	0.0041830000
abil	0.0040079998
middl	0.0038870000
prc	0.0036080000
structur	0.0032530001
experi	0.0032430000
skill	0.0032029999
leader	0.0031900001
lead	0.0031280001
session	0.0029219999
recruit	0.0027300001
strong	0.0024659999
western	0.0023890000
talent	0.0021990000
purpos	0.0021200001
submit	0.0020969999
success	0.0020730000
job	0.0019400000
averag	0.0019050000
memo	0.0017530000

17

rto	0.0037030000
pjm	0.0033849999
presto	0.0029130001
black	0.0027190000
jae	0.0025229999
nerc	0.0022350000
east	0.0019990001
tamara	0.0016310000
standard	0.0015580000
design	0.0015219999
panelist	0.0015069999
congest	0.0013250000
msn	0.0012460001
flowgat	0.0011590000
commission	0.0011410000
cell	0.0010960000
roger	0.0010430000
panel	0.0010420000
rika	0.0010380000
outlook	0.0010090000
xl	0.0009800000

repres	0.0017380000	andi	0.0009730000
remain	0.0017350001	cera	0.0009700000
rep	0.0017140000	pira	0.0009500000
mwh	0.0016670000	session	0.0009490000
18		19	
migrat	0.0075320001	lynn	0.0034399999
director	0.0035800000	ce	0.0032649999
enterpris	0.0033940000	tw	0.0026199999
ambrocik	0.0030169999	blair	0.0025670000
sundai	0.0028740000	germani	0.0022559999
weekend	0.0026489999	xl	0.0022020000
saturdai	0.0025490001	columbiaga	0.0021919999
outag	0.0023300000	msn	0.0017760000
hardwar	0.0022760001	dth	0.0017190001
period	0.0021540001	terri	0.0016550000
hous	0.0016740001	nng	0.0015820001
resolut	0.0016450000	nomin	0.0015270000
adam	0.0015350000	nom	0.0014430000
friend	0.0015240000	februari	0.0014230000
approxim	0.0013630000	pec	0.0013680001
forum	0.0013570000	oneok	0.0013440000
leav	0.0013170000	buchanan	0.0013040000
nation	0.0013140000	hpl	0.0013040000
parent	0.0012780000	randi	0.0012850000
salmon	0.0011680000	park	0.0012830000
volunt	0.0011480000	darrel	0.0012610001
appropri	0.0011220000	lee	0.0012159999
environ	0.0011060000	gari	0.0012070000
children	0.0010660000	denni	0.0011920000
life	0.0010570000	kowalk	0.0011920000
20		21	
berkelei	0.0028250001	gerald	0.0042269998
haa	0.0028090000	hpl	0.0031699999
dasovich	0.0020639999	nemec	0.0028260001
seller	0.0018360000	germani	0.0025170001
unsubscrib	0.0016020000	debra	0.0024920001
cameron	0.0014750001	dth	0.0021700000
gif	0.0014600001	perlingier	0.0021289999
msn	0.0014080000	sonat	0.0021050000
visit	0.0013520001	ed	0.0020630001
stock	0.0012960000	park	0.0020190000
school	0.0012840000	invoic	0.0015880000

html	0.0012159999	jr	0.0014240000
student	0.0011840001	dominion	0.0014020000
special	0.0011500000	mcmichael	0.0013610000
friend	0.0011420000	ruth	0.0013140000
internet	0.0011410000	nomin	0.0012910001
subscrib	0.0010990000	kai	0.0012890000
pick	0.0010670000	februari	0.0012390000
game	0.0010460000	gisb	0.0012240000
nanci	0.0010350000	bridgelin	0.0011970000
answer	0.0010230000	dperlin	0.0011410000
schwab	0.0010210000	txu	0.0011290000
san	0.0010100000	hyvl	0.0010820000
perfect	0.0009950000	reliant	0.0010720000
weekend	0.0009520000	balanc	0.0010340000

22

veronica	0.0224249996
espinoza	0.0200679991
watch	0.0166449994
russel	0.0152310003
add	0.0147040002
diamond	0.0145779997
personnel	0.0142759997
insur	0.0142270001
bradford	0.0113430005
brant	0.0100670001
reve	0.0099539999
hotlin	0.0068999999
newli	0.0068799998
jason	0.0039949999
kim	0.0019990001
darren	0.0019560000
vanek	0.0019390000
ad	0.0018210000
steel	0.0017520000
gonzalez	0.0016910000
watchlist	0.0016390000
terra	0.0014750001
canadian	0.0013740000
airlin	0.0012630000
ward	0.0011530000

24

outlook	0.0139669999
---------	--------------

23

kate	0.0091639999
syme	0.0054420000
pdx	0.0047880001
pverd	0.0041060001
tblintchg	0.0038650001
enpow	0.0036700000
path	0.0034070001
cara	0.0028450000
alia	0.0026809999
unknown	0.0026750001
tran	0.0026000000
peak	0.0024339999
portland	0.0024170000
sheet	0.0024029999
cut	0.0021460000
rout	0.0020620001
sp	0.0020369999
short	0.0019340001
hourahead	0.0018910000
varianc	0.0018010000
ep	0.0017370000
iii	0.0017330000
miss	0.0017230000
fcornr	0.0016550000
found	0.0016230000

25

michel	0.0032939999
--------	--------------

migrat	0.0127689997	fool	0.0022920000
pilot	0.0060200002	arnold	0.0012060000
owa	0.0049480000	visit	0.0011910000
mailbox	0.0040300000	shankman	0.0010600000
calendar	0.0036210001	kim	0.0010120000
garag	0.0034429999	roger	0.0009850000
bodi	0.0033320000	special	0.0009760000
shop	0.0030710001	friend	0.0009450000
internet	0.0029050000	dana	0.0009300000
button	0.0028470000	stock	0.0009190000
park	0.0027500000	ben	0.0009030000
connect	0.0026650000	realli	0.0008040000
ye	0.0024949999	jennif	0.0007910000
box	0.0023699999	job	0.0007870000
cougar	0.0022370000	employ	0.0007850000
journal	0.0021740000	benjamin	0.0007850000
onc	0.0020689999	travel	0.0007740000
mainten	0.0020290001	monei	0.0007610000
comput	0.0020180000	weekend	0.0007550000
entri	0.0019749999	littl	0.0007480000
alumni	0.0019140000	donat	0.0007350000
clickathom	0.0018520000	continent	0.0007290000
configur	0.0017630000	answer	0.0007000000
gmt	0.0017400000	leav	0.0006990000

26

outag	0.0029170001
sat	0.0018570001
dorland	0.0014930000
server	0.0014250000
ct	0.0013310000
matt	0.0012970000
pt	0.0012860000
fri	0.0012750000
visit	0.0011730000
special	0.0010380000
travel	0.0010380000
pager	0.0009940000
sun	0.0009840000
environ	0.0009810000
internet	0.0009720000
purpos	0.0009050000
job	0.0008820000
cycl	0.0008820000

27

eei	0.0053780000
cp	0.0048320000
referenc	0.0041769999
stacei	0.0029569999
east	0.0026740001
black	0.0020730000
white	0.0020310001
kayn	0.0018950000
jae	0.0018850000
tamara	0.0018820000
enpow	0.0018429999
citi	0.0017480000
model	0.0015780000
coulter	0.0015219999
lloyd	0.0015170000
doug	0.0014800000
feliccia	0.0014630000
casei	0.0014149999

research	0.0008640000	curv	0.0013410000
skill	0.0008580000	mdea	0.0013300000
train	0.0008530000	wapa	0.0012420000
srr	0.0008300000	larri	0.0012350000
itcapp	0.0008250000	entergi	0.0011710000
password	0.0008180000	known	0.0011320000
emaillink	0.0008160000	baughman	0.0011130000

28

dime	0.0019230000
barri	0.0018520000
tycholiz	0.0014310000
gerald	0.0013920000
cn	0.0013900000
laura	0.0013760000
kim	0.0012919999
whitt	0.0012280000
martin	0.0010680000
held	0.0010260000
donat	0.0009700000
hpl	0.0008930000
dear	0.0008620000
stephani	0.0008610000
nemec	0.0008560000
ina	0.0008520000
basi	0.0008430000
structur	0.0008310000
caus	0.0008290000
ec	0.0008120000
frank	0.0008050000
kimberli	0.0007970000
affili	0.0007950000
ou	0.0007810000
milller	0.0007740000

30

cn	0.0055080000
rto	0.0038280000
sa	0.0034220000
ou	0.0027490000
denron	0.0027160000
dna	0.0026910000
entergi	0.0024880001
drecipi	0.0024309999

29

restrict	0.0054549999
transco	0.0033620000
zone	0.0032780000
columbia	0.0032730000
east	0.0030240000
shipper	0.0024039999
receipt	0.0019349999
navig	0.0019320000
neal	0.0019240000
nomin	0.0018330000
station	0.0017530000
imbal	0.0017130000
nisourc	0.0016520000
germani	0.0016450000
ft	0.0013950000
fyi	0.0013910000
cd	0.0013640000
gulf	0.0013610000
eastern	0.0013379999
quantiti	0.0012930000
wgphou	0.0012650000
customernotic	0.0012640000
ce	0.0012430000
pipe	0.0012080000
victor	0.0012000001

31

short	0.0053949999
pdx	0.0035349999
palo	0.0033630000
mwh	0.0033570000
northwest	0.0031930001
southwest	0.0030169999
sp	0.0027560000
mid	0.0024440000

nyiso	0.0021019999	caiso	0.0023419999
facil	0.0018630000	np	0.0021540001
epa	0.0018380000	op	0.0021520001
environment	0.0017880000	grand	0.0018300000
interconnect	0.0017320000	cap	0.0017920000
pjm	0.0016620000	enpow	0.0017460000
christi	0.0016139999	tim	0.0015620000
renew	0.0015550000	refund	0.0015320000
bid	0.0014370000	portland	0.0015050001
strategi	0.0013880000	dow	0.0014680000
procedur	0.0013689999	dec	0.0014330000
tariff	0.0013660000	bid	0.0013600000
nicolai	0.0013070001	seller	0.0012420000
wholesal	0.0012880000	mitig	0.0011310000
stacei	0.0012700000	bpa	0.0011280000
spp	0.0012500000	nov	0.0010680000
retail	0.0011370000	ub	0.0010650000

32

emiss	0.0031760000
govern	0.0028700000
affair	0.0021259999
legisl	0.0017500001
environment	0.0017300000
ni	0.0017040001
strategi	0.0016370000
regulatori	0.0015590000
kean	0.0013660000
competit	0.0013550000
climat	0.0013060000
sue	0.0012480000
air	0.0012190000
washington	0.0011280000
mtbe	0.0010960000
fee	0.0010880000
epa	0.0010140000
cent	0.0010130000
nation	0.0010120000
shapiro	0.0009910000
articl	0.0009680000
reduct	0.0009520000
announc	0.0009440000
percent	0.0009420000
countri	0.0009360000

33

tana	0.0026670001
carol	0.0020669999
clark	0.0018119999
migrat	0.0017390000
cal	0.0017310000
deb	0.0016920000
outlook	0.0016190000
clair	0.0016020000
janett	0.0015050001
stephani	0.0014730000
kai	0.0014570000
elbertson	0.0013940000
alan	0.0013820000
taffi	0.0013550000
calendar	0.0013320000
linda	0.0013270000
shackleton	0.0013060000
cheryl	0.0012990000
elizabeth	0.0012930000
haedick	0.0012690000
julia	0.0012610001
becki	0.0012420000
lesli	0.0012040000
ann	0.0011840001
bruce	0.0011810000

	34		35
	louis 0.0051859999		salli 0.0060330001
	kitchen 0.0031210000		beck 0.0035120000
	socal 0.0023810000		dpr 0.0024379999
	delainei 0.0020409999		shona 0.0020630001
	lavorato 0.0019150000		beth 0.0016200000
	mm 0.0015770000		hall 0.0015810000
	bond 0.0014720000		commod 0.0015240000
	bankruptci 0.0013740000		wilson 0.0014970000
	assembl 0.0013110000		var 0.0013660000
	edison 0.0012600000		cn 0.0013229999
	vote 0.0011120000		enw 0.0013110000
	tim 0.0010310001		sheila 0.0012280000
	structur 0.0009950000		sap 0.0012150001
	mou 0.0009880000		brent 0.0011930000
	dwr 0.0009400000		memo 0.0010830000
	calger 0.0009200000		abel 0.0010120000
	tammi 0.0009080000		ensur 0.0009560000
	turbin 0.0009050000		summari 0.0009510000
	fund 0.0008800000		white 0.0009480000
	remain 0.0008760000		curv 0.0009340000
	xx 0.0008750000		doorstep 0.0009280000
	budget 0.0008720000		ted 0.0009180000
	summari 0.0008450000		identifi 0.0009080000
	negoti 0.0008300000		statu 0.0009020000
	burton 0.0007810000		egm 0.0008960000
	36		37
	font 0.0050369999		steff 0.0024300001
	shackleton 0.0039989999		dasovich 0.0022350000
	td 0.0038810000		rto 0.0019340001
	tana 0.0036259999		mara 0.0018110001
	br 0.0034700001		affair 0.0017070000
	stephani 0.0034240000		govern 0.0016990000
	clark 0.0027170000		cpuc 0.0016340000
	width 0.0020880001		shapiro 0.0016030000
	tr 0.0020840000		ginger 0.0015950000
	size 0.0020079999		cap 0.0015850000
	transfer 0.0018710000		legisl 0.0014330000
	moran 0.0017870000		dernehl 0.0013720000
	border 0.0017540000		san 0.0012950000
	wendi 0.0016940000		gov 0.0012850000
	samuel 0.0016890001		consum 0.0012550000

tanya	0.0016620000	edison	0.0012300001
face	0.0014940000	governor	0.0011980000
lebrocq	0.0014909999	bracepatt	0.0011980000
lee	0.0014780000	linda	0.0011970000
height	0.0014490000	wholesal	0.0011760000
rohauer	0.0014290001	regulatori	0.0011740000
swap	0.0014200000	puc	0.0011210000
cross	0.0014200000	kaufman	0.0011130000
columbia	0.0013810000	sce	0.0010770001
lesli	0.0013630000	sarah	0.0010710000

38

fund	0.0056670001
consum	0.0050980002
stock	0.0043640002
lai	0.0041680001
ken	0.0037080001
donat	0.0035730000
retir	0.0035550001
bankruptci	0.0035260001
mr	0.0035150000
declar	0.0033869999
urg	0.0033590000
live	0.0022519999
monei	0.0022000000
york	0.0021250001
sincer	0.0020790000
write	0.0020520000
reach	0.0020210000
benefit	0.0019360000
transit	0.0019310000
lost	0.0019060000
worth	0.0018880001
profit	0.0018700000
basic	0.0018690000
dollar	0.0018660000
largest	0.0018470000
unabl	0.0018340000

40

varianc	0.0247840006
alia	0.0247239992
hourahead	0.0217760000
tran	0.0216099992

39

usa	0.0041149999
rac	0.0028979999
memo	0.0028150000
var	0.0027240000
violat	0.0025050000
categori	0.0023779999
notif	0.0022829999
mwh	0.0022610000
mm	0.0021780001
cd	0.0021060000
publish	0.0020820000
kingdom	0.0017590000
commod	0.0017220000
enrononlin	0.0016010000
crude	0.0015790000
erv	0.0014810000
barrel	0.0014759999
model	0.0014660000
consolid	0.0012820000
mccannel	0.0011810000
dpr	0.0011620000
count	0.0011500000
portfolio	0.0011050000
curv	0.0010890000
andi	0.0010860000
tv	0.0010670000

41

interview	0.0071359999
migrat	0.0055849999
outlook	0.0053630001
candid	0.0037250000

unknown	0.0209480003	feedback	0.0028029999
detect	0.0180980004	calendar	0.0023759999
portland	0.0157650001	ye	0.0023220000
westdesk	0.0156640001	recruit	0.0021410000
pars	0.0150509998	analyst	0.0021400000
sc	0.0138259996	pep	0.0020160000
ancillari	0.0135800000	marin	0.0018520000
award	0.0135639999	saturdai	0.0017620000
prefer	0.0122819999	normal	0.0017180000
attempt	0.0121259997	button	0.0015740000
engin	0.0119460002	hunter	0.0014800000
borland	0.0119220000	summer	0.0014430000
occur	0.0119099999	box	0.0014330000
mkt	0.0108340001	prc	0.0014180000
found	0.0101549998	ut	0.0013920000
txt	0.0084690005	weather	0.0013000000
tabl	0.0076110000	super	0.0012830000
engi	0.0067940000	shive	0.0012720000
tie	0.0053679999	aga	0.0012400000
interchg	0.0053539998	track	0.0012309999
ciso	0.0053490000	dinner	0.0012190000

42

maggi	0.0030159999
bass	0.0027079999
michel	0.0022430001
nelson	0.0020780000
friend	0.0018470000
game	0.0018180000
beta	0.0017430000
msn	0.0016400000
jpg	0.0015890000
god	0.0015580000
man	0.0014630000
matthew	0.0014100000
littl	0.0014080000
lenhart	0.0014060000
hotel	0.0013790000
realli	0.0013120000
life	0.0012550000
plai	0.0012309999
prayer	0.0012070000
gore	0.0011970000
saturdai	0.0011590000

43

erm	0.0029559999
meter	0.0029320000
record	0.0027970001
inactiv	0.0026489999
sheet	0.0024770000
overlai	0.0020140000
cheryl	0.0019749999
summari	0.0019240000
sitara	0.0019110000
daren	0.0017659999
gcp	0.0016970000
jan	0.0016899999
period	0.0016630000
short	0.0015450000
merger	0.0015050001
notif	0.0014700000
shortnam	0.0014670000
cp	0.0013100000
publish	0.0013070001
ng	0.0013070001
hpl	0.0013030000

weekend	0.0011510000	tagg	0.0012919999
dinner	0.0011100000	ticket	0.0012700000
big	0.0011010000	farmer	0.0012360000
amanda	0.0010910000	determin	0.0012050000
44		45	
vinc	0.0125489999	kai	0.0075460002
kaminski	0.0074350000	mann	0.0049800002
research	0.0038360001	ge	0.0036130000
model	0.0027170000	ben	0.0019100000
shirlei	0.0024699999	notifi	0.0018960000
rice	0.0016970000	bna	0.0018810000
visit	0.0016600000	delet	0.0017170000
crenshaw	0.0016020000	turbin	0.0016550000
univers	0.0015300000	exhibit	0.0015940000
interview	0.0014720000	prohibit	0.0014400000
financ	0.0013989999	immedi	0.0014240000
stinson	0.0013570000	rto	0.0013040000
html	0.0012780000	disclosur	0.0012600000
dear	0.0012769999	jacobi	0.0012430000
school	0.0011040000	facil	0.0012380000
wharton	0.0010980000	sander	0.0012320000
td	0.0010779999	attorney	0.0011800000
cgi	0.0010370000	claim	0.0011310000
resum	0.0009930000	ps	0.0011250000
travel	0.0009250000	dissemin	0.0011070000
martin	0.0009200000	court	0.0010430000
zimin	0.0009140000	brobeck	0.0010260000
student	0.0009110000	sole	0.0009869999
seminar	0.0008870000	matter	0.0009800000
gibner	0.0008850000	content	0.0009260000
46		47	
ub	0.0127720004	elizabeth	0.0034330001
simul	0.0086939996	sager	0.0028830001
card	0.0057430002	carol	0.0021090000
session	0.0050130002	eei	0.0020089999
februari	0.0041250000	bond	0.0018420001
warburg	0.0033640000	clair	0.0018250000
ensur	0.0028440000	px	0.0016090000
estat	0.0026640000	collater	0.0015990000
drive	0.0026130001	steff	0.0013229999
launch	0.0023220000	traci	0.0011950000
south	0.0022310000	csc	0.0011070000

restrict	0.0018520000	bradford	0.0010650000
ubswenergi	0.0017980000	default	0.0010590000
integr	0.0017030000	bankruptci	0.0010240000
feb	0.0016950000	sce	0.0009570000
mid	0.0016940000	stoklei	0.0009460000
role	0.0016880000	portz	0.0009420000
curv	0.0016640000	mellencamp	0.0009260000
involv	0.0016589999	neil	0.0009070000
pick	0.0015510001	ed	0.0009030000
benefit	0.0015340000	cover	0.0008900000
louis	0.0015170000	christian	0.0008790000
ec	0.0014050000	exposur	0.0008410000
ubsw	0.0013980001	murrai	0.0008400000
variou	0.0013560000	guaranti	0.0008350000

A.4 Author Topic with all Words (No Dictionary)

	0		1
schedul	0.0024183115	tana	0.0270040091
mark	0.0022603997	jone	0.0263075139
request	0.0019460776	taylor	0.0147149982
deal	0.0018795305	counterparti	0.0129796620
agreement	0.0017872652	mark	0.0128616123
final	0.0016853932	mari	0.0105006117
comment	0.0016816906	stephani	0.0101110460
possibl	0.0014882537	lesli	0.0089954734
gener	0.0014797477	moran	0.0085645905
develop	0.0014500266	tom	0.0080982931
respons	0.0014240083	eol	0.0075670676
point	0.0014122999	legal	0.0073309676
document	0.0013497557	approv	0.0073014549
process	0.0013390481	tanya	0.0070358422
david	0.0013299417	master	0.0069532073
sinc	0.0013196344	rohauer	0.0066816919
base	0.0012844095	karen	0.0065341294
jeff	0.0012843094	samuel	0.0063865669
confirm	0.0012639951	lambert	0.0063098343
end	0.0012632946	debbi	0.0062626144
made	0.0012595919	isda	0.0061917845
draft	0.0012572904	lisa	0.0060619293
howev	0.0012140598	david	0.0058848541
hour	0.0012110576	lon	0.0057727066
order	0.0012078554	davi	0.0057313889

	2		3
	eb	0.0202717334	phillip 0.0131834894
	street	0.0195496604	kevin 0.0098652076
	america	0.0193681102	scott 0.0086804442
	north	0.0189678762	salli 0.0078119328
	sara	0.0189678762	book 0.0073776771
	smith	0.0183159467	neal 0.0070378245
	texa	0.0171399992	giron 0.0066460501
shackleton	0.0168635473	love 0.0064761238	
legal	0.0165293310	darron 0.0063675600	
kai	0.0154400319	presto 0.0063533997	
mann	0.0137276864	hunter 0.0063156383	
agreement	0.0129643520	david 0.0062259547	
debra	0.0107527450	beck 0.0059899464	
perlingier	0.0103525100	louis 0.0059285839	
dperlin	0.0081078932	januari 0.0057633780	
depart	0.0079098390	kam 0.0057208962	
draft	0.0073074233	allen 0.0053432826	
master	0.0054217805	keiser 0.0052583194	
document	0.0049720318	gossett 0.0051733563	
carol	0.0048193648	jeffrei 0.0050270311	
confirm	0.0047120852	frank 0.0049326275	
isda	0.0044934005	lavorato 0.0048335041	
clair	0.0043819947	kitchen 0.0045880550	
st	0.0042788414	shive 0.0045078122	
ena	0.0037713270	thoma 0.0043898076	
	4		5
gerald	0.0162994247	request 0.0408676490	
kim	0.0120751252	page 0.0342316255	
nemec	0.0116818016	id 0.0331529304	
arnold	0.0098882439	type 0.0298076123	
ward	0.0091881277	resourc 0.0281008165	
barri	0.0069540469	creat 0.0241820142	
paul	0.0059392713	act 0.0226527266	
mark	0.0057190098	approv 0.0217242297	
septemb	0.0052863532	applic 0.0211507455	
tycholiz	0.0049087624	srr 0.0187612325	
whitt	0.0046963673	auth 0.0183379464	
bt	0.0046649012	itcapp 0.0182014033	
imceanot	0.0045075719	emallink 0.0179965887	
lucci	0.0041771797	pend 0.0125211878	
august	0.0037130571	data 0.0108826645	
ecom	0.0036579918	read 0.0077421609	

gui	0.0035871936	directori	0.0063767242
matt	0.0031938695	altern	0.0062265266
juli	0.0031938695	common	0.0058578588
theresa	0.0031230713	websit	0.0041101002
patric	0.0027297472	cd	0.0038916303
smith	0.0025409518	reject	0.0038643214
cell	0.0025252188	myreq	0.0037550866
charl	0.0025094857	specifi	0.0035775800
sorri	0.0024937529	access	0.0035639254

6

chri	0.0084524360
dutch	0.0069364328
quiglei	0.0061657275
cooper	0.0049969656
dorland	0.0047428869
august	0.0045057465
calgari	0.0044464618
griffith	0.0044041150
januari	0.0041500367
canada	0.0041161594
jon	0.0041076900
jonathan	0.0040399358
mckai	0.0038366728
septemb	0.0033793312
imceanot	0.0032522918
richei	0.0031845374
ca	0.0031083140
zufferli	0.0030151517
rob	0.0029304589
geoff	0.0024985250
june	0.0024815865
gui	0.0024646479
decemb	0.0023968937
xl	0.0022444464
jai	0.0021428149

8

develop	0.0073821032
opportun	0.0062972121
technolog	0.0054445118
base	0.0051162424
recent	0.0046518608
success	0.0045798016

7

post	0.0112593165
client	0.0090507045
particip	0.0090507045
nyiso	0.0076503507
relat	0.0065695406
comment	0.0063345814
found	0.0060432330
schedul	0.0059774444
write	0.0059586475
public	0.0059116557
hour	0.0058646640
websit	0.0058270707
real	0.0057894774
ahead	0.0057800789
tech	0.0057142908
exchang	0.0056954939
correct	0.0052819666
htm	0.0051315930
distribut	0.0050376095
associ	0.0048966343
iso	0.0046992688
begin	0.0045394967
area	0.0043891231
offici	0.0043421313
load	0.0041541643

9

michel	0.0246719364
maggi	0.0138851777
cash	0.0114658074
nelson	0.0101748388
januari	0.0071530170
amanda	0.0061106794

gener	0.0044797193	leav	0.0059289876
experi	0.0042675454	sorri	0.0054030376
lead	0.0041034105	realli	0.0053074104
part	0.0040593743	dont	0.0051735318
presid	0.0040473645	im	0.0045137038
solut	0.0039392756	nice	0.0041598827
organ	0.0038872329	thought	0.0039781909
posit	0.0036750585	ny	0.0037964990
support	0.0036550420	ye	0.0036626209
becom	0.0036230157	thing	0.0035861190
respons	0.0035949927	tomorrow	0.0034809290
relat	0.0033467889	anyth	0.0030888570
global	0.0033107593	lunch	0.0030601688
announc	0.0032587165	feel	0.0029741044
level	0.0032066738	guess	0.0029645415
integr	0.0031546310	told	0.0029454161
ceo	0.0031506277	didn	0.0028976025
effect	0.0031145981	alwai	0.0028784771
leader	0.0030905784	whatev	0.0028689143

10

password	0.0194975734
user	0.0192945674
access	0.0179617833
account	0.0146960206
id	0.0090294806
login	0.0082792379
center	0.0072377245
log	0.0070170648
respond	0.0069729327
onlin	0.0067610997
launch	0.0058608083
submit	0.0057460652
box	0.0056313220
import	0.0053047459
secur	0.0052959197
order	0.0051900027
locat	0.0051811766
enter	0.0048016421
sap	0.0046957252
instruct	0.0045545031
code	0.0045545031
problem	0.0043603228
button	0.0043514962

11

deal	0.0213993974
kate	0.0154789397
mw	0.0124098537
syme	0.0123958075
pdx	0.0123115303
bui	0.0063980967
confirm	0.0058713653
west	0.0058502965
show	0.0057309037
miss	0.0056957887
peak	0.0054429579
enter	0.0054078423
portland	0.0052322652
chri	0.0051269191
real	0.0049794344
enpow	0.0047195805
desk	0.0047055343
sp	0.0044667493
broker	0.0043262877
sean	0.0042560571
diana	0.0042209416
cara	0.0041436879
trader	0.0039681108

train 0.0043514962 schedul 0.0038487182
action 0.0042897118 bill 0.0038276492

12

confidenti 0.0241064131
notifi 0.0228717942
intend 0.0217075013
error 0.0198399425
privileg 0.0190663524
immedi 0.0148702096
recipi 0.0145889036
dissemin 0.0145420199
prohibit 0.0135887060
delet 0.0135340076
individu 0.0131120486
sander 0.0127916727
sender 0.0123306438
distribut 0.0120727811
richard 0.0113070039
entiti 0.0101974094
herebi 0.0091815833
elizabeth 0.0083298525
attorney 0.0081345011
legal 0.0080407327
sager 0.0075562615
strictli 0.0068373694
destroi 0.0064232247
unauthor 0.0060559646
transmit 0.0059621958

14

california 0.0079632951
natur 0.0063743074
public 0.0061551365
commiss 0.0057259272
west 0.0056163422
electr 0.0052784537
pacif 0.0052236612
nation 0.0050592828
south 0.0049496978
jane 0.0049131690
order 0.0048218481
weather 0.0044930917
forecast 0.0040730145

13

lai 0.0158001203
mr 0.0155321658
mani 0.0151533354
ken 0.0151440958
pai 0.0146266678
sincer 0.0145157902
fund 0.0144326324
monei 0.0142940357
made 0.0139798829
bill 0.0139429234
reach 0.0139336837
million 0.0138967251
stock 0.0136472508
worth 0.0132868988
financi 0.0132684195
result 0.0132499402
american 0.0130651444
york 0.0130374255
save 0.0130281849
lost 0.0130097056
bui 0.0129912263
live 0.0129265478
write 0.0128803486
effort 0.0128433900
dollar 0.0127509916

15

polic 0.0066803554
electr 0.0066531999
risk 0.0063363854
pdf 0.0062458669
gener 0.0055579259
analysi 0.0051958524
confer 0.0049061929
regist 0.0045984299
center 0.0043992894
download 0.0041186819
brochur 0.0038561784
edit 0.0037566081
learn 0.0036841934

capac	0.0040638824	subscript	0.0036570376
pg	0.0040456182	full	0.0035212599
mile	0.0038721079	registr	0.0034941044
el	0.0036894658	region	0.0034578971
tholt	0.0036346731	train	0.0033492749
transport	0.0035890124	hotel	0.0033221194
plant	0.0035159555	websit	0.0032949639
transmiss	0.0035068234	present	0.0032678081
east	0.0034976914	util	0.0032497046
area	0.0034794272	rate	0.0032225489
expect	0.0034337665	develop	0.0032225489
north	0.0033241811	option	0.0029781491

16

school	0.0132470103
attend	0.0109742554
univers	0.0101671955
invit	0.0098425159
student	0.0090076262
org	0.0084881391
event	0.0082655018
class	0.0077367383
resum	0.0070131668
associ	0.0069760606
opportun	0.0068090828
director	0.0067627002
mba	0.0067627002
appreci	0.0066421051
haa	0.0063267020
berkelei	0.0062061069
present	0.0061040646
join	0.0059370869
member	0.0054732589
cours	0.0052970047
graduat	0.0052970047
center	0.0049908785
session	0.0045177741
speaker	0.0044992212
mr	0.0041559888

18

jeff	0.0169986300
dasovich	0.0168596078
steff	0.0099042999

17

anyon	0.0218485966
sole	0.0218024366
accept	0.0216537025
otherwis	0.0214947108
confidenti	0.0211049244
disclosur	0.0209664479
author	0.0207818132
strictli	0.0207305253
sender	0.0206125639
recipi	0.0205458887
materi	0.0204484425
parti	0.0203868970
administr	0.0203304812
properti	0.0203099661
prohibit	0.0202638078
distribut	0.0201817472
intend	0.0201663617
delet	0.0200843010
reli	0.0197919607
privileg	0.0197611880
affili	0.0197201576
relev	0.0196534842
evid	0.0194380768
basi	0.0193098560
enforc	0.0191816371

19

process	0.0140070906
feedback	0.0135029834
employe	0.0132989408

mara	0.0078948010	option	0.0128548462
shapiro	0.0073176497	open	0.0121466964
california	0.0073092245	begin	0.0108864289
jame	0.0067362855	respons	0.0106703825
ca	0.0057210042	particip	0.0103823217
gov	0.0049332133	request	0.0094221178
na	0.0049079368	select	0.0091460599
util	0.0048952983	onc	0.0090860473
commiss	0.0048615960	access	0.0089660212
kaufman	0.0048110429	form	0.0089300135
richard	0.0047478508	import	0.0086179478
susan	0.0047183614	london	0.0085219275
electr	0.0046720207	perform	0.0081738532
sue	0.0043265722	desk	0.0079578077
kean	0.0041875504	repres	0.0073816855
jim	0.0040148264	longer	0.0071776421
cpuc	0.0038463150	pep	0.0070456141
legisl	0.0034966536	submit	0.0068775788
ferc	0.0034587388	main	0.0068535735
denn	0.0034292492	immedi	0.0064814948
kingerski	0.0033702701	success	0.0064574894
governor	0.0033281425	goal	0.0062534465

20

travel	0.0124584045
hotel	0.0097735953
airlin	0.0086229630
ticket	0.0080407979
flight	0.0077736867
fare	0.0067463359
citi	0.0066436012
book	0.0061093788
airport	0.0060134926
night	0.0055409116
air	0.0053765355
seat	0.0052738003
tax	0.0049450486
trip	0.0049381992
la	0.0048697093
car	0.0047121821
arriv	0.0046779374
rate	0.0046436922
san	0.0046162964
continent	0.0045683533

21

god	0.0077153584
dana	0.0058440436
bless	0.0053422386
judi	0.0052481499
love	0.0050599729
someon	0.0048613418
believ	0.0046940735
life	0.0046627107
read	0.0043595368
hernandez	0.0043281741
peopl	0.0042863567
davi	0.0039309114
friend	0.0037427344
prayer	0.0034081978
prai	0.0032513838
lord	0.0030841152
andrea	0.0030736611
heart	0.0030422981
reason	0.0030213897
org	0.0029377555

vacat	0.0041574133	jesu	0.0028750298
york	0.0040204329	faith	0.0027600329
destin	0.0040067350	sandra	0.0027182158
rental	0.0039587920	thing	0.0027182158
room	0.0039039999	thought	0.0026032187
22		23	
yahoo	0.0109859817	yahoo	0.0209943801
mime	0.0109363170	tm	0.0069047296
hotmail	0.0097741578	sponsor	0.0065174312
version	0.0094563020	connect	0.0064373007
mailer	0.0075690346	download	0.0062903943
type	0.0074101072	everyon	0.0054623769
msn	0.0073405760	egroup	0.0046076495
fwd	0.0067942617	beta	0.0043138368
return	0.0066452669	left	0.0042871265
smtp	0.0065658032	owner	0.0041802856
plain	0.0057115662	stai	0.0036193705
id	0.0056817676	real	0.0035793053
boundari	0.0054930407	forget	0.0035125297
multipart	0.0052745151	ago	0.0034591092
earthlink	0.0052447161	indic	0.0034323989
esntp	0.0049268603	search	0.0033255580
att	0.0048175976	pick	0.0032854925
path	0.0047778655	consid	0.0032187169
mix	0.0047480669	moment	0.0031385864
rr	0.0047083348	imag	0.0031118761
charset	0.0044898093	advertis	0.0031118761
text	0.0039534280	biggest	0.0030985209
encod	0.0039136959	bottom	0.0030985209
org	0.0038838969	reunion	0.0030851658
jpg	0.0038739638	special	0.0030718108
24		25	
oil	0.0079487469	vinc	0.0380417667
daili	0.0047573945	kaminski	0.0321325436
natur	0.0039149569	research	0.0096860696
petroleum	0.0037167363	shirlei	0.0078447908
stori	0.0037068252	vkamin	0.0071425354
mexico	0.0036473591	univers	0.0058493582
full	0.0034986937	crenshaw	0.0055153584
project	0.0034689605	stinson	0.0048045390
construct	0.0033995833	model	0.0045219245
worldwid	0.0032013627	paper	0.0041879248

crude	0.0031220743	vkaminski	0.0038710537
canada	0.0030725193	present	0.0038453613
headlin	0.0029932309	gibner	0.0037169000
newslett	0.0029734089	monika	0.0036997718
plant	0.0029040317	risk	0.0036826436
global	0.0028941205	financ	0.0036312591
util	0.0028643876	vasant	0.0034771056
sourc	0.0028643876	version	0.0031002855
sector	0.0028445655	shanhogu	0.0030745934
engin	0.0027553663	sincer	0.0030574652
welcom	0.0026859890	zimin	0.0027491581
million	0.0026265227	rice	0.0027063375
lng	0.0026067006	causholli	0.0025864404
annual	0.0025472345	write	0.0025436198
unit	0.0024976793	martin	0.0025093635

26

imag	0.0121091232
er	0.0066877296
se	0.0061155753
es	0.0060217800
ng	0.0059842616
ed	0.0053370711
im	0.0052807936
ne	0.0051494795
ma	0.0049900268
ag	0.0049806470
ge	0.0049431287
al	0.0049337493
nt	0.0048680925
ion	0.0043709748
ing	0.0041552447
fo	0.0041458653
yo	0.0040520695
pr	0.0040239305
de	0.0040145512
polic	0.0039488939
ou	0.0038644779
releas	0.0038082006
om	0.0037894414
parti	0.0036206089
mage	0.0036112294

28

27

love	0.0041144555
perfect	0.0040433290
food	0.0040269154
friend	0.0036001555
women	0.0035727990
wine	0.0033047064
men	0.0030147289
hous	0.0029217170
christma	0.0028834182
gift	0.0027849351
ey	0.0027302224
boi	0.0027247511
eat	0.0027083373
water	0.0025989118
road	0.0025879692
walk	0.0025824979
live	0.0025715553
light	0.0025387276
hand	0.0025332563
car	0.0025168424
door	0.0024840150
beauti	0.0024730724
famili	0.0024347734
music	0.0024293021
kid	0.0023964744

29

ben	0.0123026846	watson	0.0103438850
roger	0.0104535269	kim	0.0098246969
benjamin	0.0094975000	tw	0.0097381659
version	0.0061136670	kimberli	0.0094586033
written	0.0057111294	shellei	0.0079542892
latest	0.0047802608	lynn	0.0075349458
portfolio	0.0044280407	traci	0.0072620395
recipi	0.0042393510	blair	0.0071222577
part	0.0038242340	harri	0.0067561641
sender	0.0037613374	steven	0.0067361952
board	0.0036103858	pipelin	0.0066896016
hard	0.0035852271	transwestern	0.0064499765
master	0.0034845928	corman	0.0064166952
plain	0.0034720134	Michel	0.0061704135
prior	0.0034720134	drew	0.0060838824
deliveri	0.0034720134	rod	0.0058708824
idea	0.0034720134	steve	0.0055181007
fool	0.0034468549	januari	0.0053849756
close	0.0033462204	geaccon	0.0052119130
msg	0.0032833240	fossum	0.0051852879
motlei	0.0032707446	lorrain	0.0051187258
materi	0.0031826894	lokai	0.0050255382
share	0.0031701103	darrel	0.0050188820
direct	0.0031701103	millier	0.0049656318
sponsor	0.0031323722	hayslett	0.0049589756

30

realli	0.0102745108
thing	0.0100450031
love	0.0096607106
night	0.0092790872
weekend	0.0080488175
hei	0.0073656314
littl	0.0072428710
ve	0.0068025361
thought	0.0060793189
didn	0.0060446262
someth	0.0058818357
gui	0.0058177868
pretti	0.0051372689
sound	0.0051105819
hear	0.0050545395
probabl	0.0049691410
fun	0.0049531288

31

total	0.0196107626
amount	0.0180188566
due	0.0169484392
approv	0.0168798212
statu	0.0163995065
expens	0.0161936563
employe	0.0142586678
paid	0.0114316633
jr	0.0111160269
payment	0.0110611338
autom	0.0101691177
jim	0.0098946514
concur	0.0097711412
derrick	0.0086732758
jame	0.0085223196
readi	0.0076851971
administr	0.0071911574

mayb	0.0049371165	deliveri	0.0070539243
leav	0.0048650620	expensexm	0.0067108413
feel	0.0047930069	refer	0.0065461611
anywai	0.0047289585	depart	0.0063814814
nice	0.0046328851	law	0.0059972284
lot	0.0045875171	track	0.0059560584
morn	0.0044300640	tx	0.0058051022
friend	0.0043766904	account	0.0057227621

32

employe	0.0061857468
center	0.0056959158
eb	0.0042075836
support	0.0040380270
schedul	0.0039501088
assist	0.0037554323
join	0.0036549543
build	0.0028950884
risk	0.0028574092
announc	0.0028448494
team	0.0028448494
american	0.0027569311
particip	0.0026752925
facil	0.0024554967
access	0.0024178175
begin	0.0023864179
north	0.0023738581
unit	0.0023236191
open	0.0023173392
desk	0.0022482607
locat	0.0021666221
event	0.0021226630
data	0.0021101031
field	0.0020975433
ee	0.0020787038

34

remov	0.0247787684
line	0.0156588424
word	0.0143632255
field	0.0086240489
enter	0.0062150122
point	0.0056481799
id	0.0046764673

33

expect	0.0063204165
lower	0.0051937182
econom	0.0049828384
end	0.0047358084
higher	0.0043321252
high	0.0042176479
level	0.0041875220
demand	0.0039525423
rise	0.0039103664
economi	0.0037597383
increas	0.0037597383
short	0.0037597383
remain	0.0037356378
fall	0.0037235876
rate	0.0036874367
low	0.0035006579
declin	0.0034946327
billion	0.0034524568
stock	0.0033861806
major	0.0033861806
move	0.0031753012
produc	0.0030608238
point	0.0030367232
weak	0.0029644219
long	0.0029583967

35

invest	0.0094604082
stock	0.0072758151
investor	0.0057509565
copyright	0.0057073892
research	0.0052405959
financi	0.0052032522
bui	0.0050289831

browser	0.0045651253	nasdaq	0.0047800266
privat	0.0043120752	secur	0.0047053397
member	0.0041602449	account	0.0045933090
awai	0.0040489030	advertis	0.0041949786
loan	0.0040185372	alert	0.0039522462
recipi	0.0039982931	articl	0.0039335745
directli	0.0037958529	technolog	0.0039273505
real	0.0037857310	analyst	0.0035041245
bodi	0.0035832908	quot	0.0034854528
profil	0.0035428028	subscrib	0.0033298549
match	0.0033201186	corpor	0.0032738398
someon	0.0032796308	portfolio	0.0032053767
life	0.0032796308	fund	0.0031680332
card	0.0032492648	street	0.0031369138
rate	0.0032087767	access	0.0030933463
automat	0.0031176787	full	0.0030622268
monthli	0.0030569467	perform	0.0030248833
debt	0.0029051166	nyse	0.0029875399

36

past	0.0150199244
return	0.0138555141
open	0.0127338357
browser	0.0125842784
line	0.0125308651
abov	0.0121356063
member	0.0118151270
indic	0.0113237249
press	0.0112062152
quot	0.0111955330
text	0.0110246101
happen	0.0106186690
noth	0.0102661410
cfm	0.0096785948
presid	0.0095076729
exchang	0.0090055875
vice	0.0084821377
log	0.0084287245
nymex	0.0080548311
firm	0.0075100157
clear	0.0070079304
divis	0.0066233547
useremail	0.0064951628
close	0.0061426354

37

size	0.0148552554
face	0.0121037671
text	0.0119427759
gif	0.0116061578
color	0.0116061578
bodi	0.0115768863
font	0.0114890728
href	0.0113280816
br	0.0108890142
tabl	0.0102157779
titl	0.0097620757
height	0.0094839996
head	0.0094108218
center	0.0090595679
border	0.0085473228
arial	0.0085180514
san	0.0083424244
tr	0.0081082555
src	0.0080789840
type	0.0080204420
width	0.0078594508
td	0.0078594508
img	0.0077862726
imag	0.0076984591

effect	0.0060678567	jpg	0.0073911119
38		39	
onlin	0.0081584742	ship	0.0067516831
imag	0.0077572870	softwar	0.0064260443
special	0.0076547940	tax	0.0054382738
subscrib	0.0063926657	comput	0.0049715252
save	0.0052242451	cd	0.0046784505
order	0.0050983252	save	0.0045047766
select	0.0046649259	alwai	0.0043419572
member	0.0045712180	world	0.0042768293
privaci	0.0045390059	limit	0.0042334110
regist	0.0043633035	upgrad	0.0041899923
newslett	0.0043018078	differ	0.0041574286
page	0.0042959512	applic	0.0041357190
shop	0.0042520254	pc	0.0040380275
prefer	0.0042461688	charg	0.0040054638
enjoi	0.0041026785	cgi	0.0039837547
featur	0.0039972570	choos	0.0038426444
ve	0.0039855437	item	0.0038209353
simpli	0.0036253540	store	0.0038100807
gift	0.0035872851	featur	0.0037666622
friend	0.0034232961	purchas	0.0037449528
plu	0.0034115827	displai	0.0035278604
easi	0.0032915196	digit	0.0035170058
promot	0.0032768776	select	0.0033324773
purchas	0.0030660348	flo	0.0032564949
valu	0.0030572498	dvd	0.0031805125
40		41	
msn	0.0197202340	joe	0.0053574485
hotmail	0.0137366047	leagu	0.0050068661
explor	0.0127114905	fantasi	0.0049192202
download	0.0118458383	team	0.0048753973
matthew	0.0082617365	footbal	0.0046891505
lenhart	0.0072669960	game	0.0046891505
intl	0.0068341698	commission	0.0042728339
park	0.0060292655	season	0.0038236501
night	0.0058242427	player	0.0037688715
joe	0.0050497120	plai	0.0036483586
brian	0.0044118632	martin	0.0034621118
chad	0.0042524012	sportslin	0.0033525547
erik	0.0035234310	nfl	0.0033415991
friend	0.0035158375	drop	0.0032539535

weekend	0.0030906051	win	0.0032101306
man	0.0030526379	score	0.0031115292
bmc	0.0029311429	roster	0.0029471938
matt	0.0028779889	stepenovitch	0.0029143267
wollam	0.0028248348	cb	0.0029033709
chet	0.0027716807	injuri	0.0028924153
gui	0.0027109331	smith	0.0028047697
knipe	0.0027033398	sundai	0.0026952126
fenner	0.0026501857	pass	0.0026404341
wed	0.0023768218	run	0.0025966112
game	0.0023540417	rb	0.0025418326

42

word	0.0065916567
world	0.0055805189
org	0.0043517952
nation	0.0041918047
peopl	0.0038462263
polit	0.0036926358
american	0.0036094408
attack	0.0034622499
war	0.0033598563
stori	0.0032446634
countri	0.0032126654
bush	0.0031742677
articl	0.0030526752
subscript	0.0029950789
subscrib	0.0029118839
york	0.0026942973
turn	0.0026431007
washington	0.0025471065
public	0.0025023094
citi	0.0023935160
elect	0.0023807168
terror	0.0023039216
law	0.0022527247
life	0.0022207268
theme	0.0022015278

44

calendar	0.0239226818
team	0.0222685467
standard	0.0189187191
central	0.0185862295

43

suggest	0.0051508322
stop	0.0051137768
letter	0.0051045129
troubl	0.0050304015
washington	0.0048821787
media	0.0048451228
thing	0.0047895396
william	0.0047524837
interview	0.0046969005
event	0.0046876362
mind	0.0045949970
live	0.0045394138
llc	0.0045208856
bureau	0.0044838302
moment	0.0044375104
join	0.0043726629
player	0.0043633990
window	0.0042892876
subscrib	0.0042707599
guest	0.0042522321
experi	0.0042059124
comment	0.0041688569
copyright	0.0041595930
wrap	0.0041225371
cst	0.0040762178

45

eric	0.0237493962
bass	0.0184485298
jason	0.0103926063
na	0.0063704569

descript	0.0185612924	reach	0.0061844615
entri	0.0182537399	automat	0.0060449648
outlook	0.0157101974	turn	0.0059984662
appoint	0.0154275820	ou	0.0058705942
migrat	0.0149787217	space	0.0058357199
chairperson	0.0147626037	de	0.0056962236
shankman	0.0137318866	item	0.0055451025
jeffrei	0.0127011705	longer	0.0054521048
jeff	0.0095508350	gmt	0.0054521048
eb	0.0077304570	warn	0.0054288553
mcconnel	0.0073896553	size	0.0053707319
white	0.0060680108	delet	0.0053242329
confer	0.0056440872	cn	0.0052544847
room	0.0056191506	mailbox	0.0050336150
delainei	0.0049874210	ee	0.0050103660
mtg	0.0049126111	inlin	0.0049289926
kean	0.0047546788	folder	0.0049057435
whallei	0.0047131176	individu	0.0049057435
cindi	0.0045967465	client	0.0048941188
invit	0.0042143837	limit	0.0048592445
horton	0.0041478858	button	0.0048243706

46

chri	0.0159777664
germani	0.0157330483
deal	0.0136963669
xl	0.0124964612
daren	0.0111623565
volum	0.0095756389
farmer	0.0093151331
hpl	0.0074363342
deliveri	0.0065364055
meter	0.0061969585
flow	0.0057627824
capac	0.0050602062
receipt	0.0050049471
nomin	0.0047523356
sitara	0.0046418179
mmbtu	0.0043576299
nom	0.0042155357
ce	0.0041129123
april	0.0040892302
transport	0.0038208300
lisa	0.0037813596

47

game	0.0126587525
plai	0.0072579952
season	0.0062541370
texa	0.0059630182
footbal	0.0059128255
colleg	0.0048788516
team	0.0048386971
saturdai	0.0044672694
austin	0.0041962280
player	0.0040356102
coach	0.0038549160
fan	0.0038248003
big	0.0037143759
sport	0.0034634112
win	0.0033228712
ut	0.0033128324
ticket	0.0032425625
michael	0.0032425625
true	0.0031120609
sundai	0.0029414049
run	0.0029213279

point	0.0035050656	dalla	0.0028912120
confirm	0.0033945481	bowl	0.0028008649
path	0.0033471833	save	0.0028008649
invoic	0.0033156069	st	0.0027406334

Appendix B. Most Probable Users and Explicit Social Network Statistics

B.1 PLSI-U with only Dictionary Words

CATEGORY 0

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 833 COMPONENTS: 4
 LARGEST COMPONENT SIZE: 812 PERCENT OF TOTAL GRAPH: 97.48%
 GROUP DEGREE: 0.25217 GRAPH DENSITY: 0.00361
 GROUP CLOSENESS: 0.00364 GROUP BETWEENNESS: 0.37696
 AVERAGE $p(z|u)$: 0.76 STDEV $p(z|u)$: 0.33

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
0	8431	skean@enron.com.....	Steve Kean.....	0.017804
0	2222	harry.kingerski@enron.com.....	Harry Kingerski.....	0.015902
0	3132	james.wright@enron.com.....	0.014326
0	1475	dennis.benevides@enron.com.....	0.014257
0	34229	mpalmer@enron.com.....	mpalmer@enron.com...	0.013434
0	253	jeff.dasovich@enron.com.....	Jeff Dasovich.....	0.013172
0	1746	scott.stoness@enron.com.....	0.012924
0	9244	richard.sanders@enron.com.....	0.012769
0	181	paul.kaufman@enron.com.....	0.012533
0	8546	sandra.mccubbin@enron.com.....	Sandra McCubbin.....	0.012356
0	801	susan.mara@enron.com.....	Susan Mara.....	0.012340
0	1489	james.steffes@enron.com.....	0.011824
0	28783	roger.yang@enron.com.....	0.011751
0	817	richard.shapiro@enron.com.....	Richard Shapiro.....	0.011482
0	1180	karen.denne@enron.com.....	Karen Denne.....	0.011450
0	213	chris.foster@enron.com.....	Chris H Foster.....	0.011375
0	1016	neil.bresnan@enron.com.....	Neil Bresnan.....	0.011177
0	7213	mike.smith@enron.com.....	Mike Smith.....	0.010670
0	3157	jennifer.rudolph@enron.com.....	0.010646
0	37	tim.belden@enron.com.....	Tim Belden.....	0.010578

CATEGORY 1

EXPLICIT SOCIAL NETWORK STATISTICS

AVERAGE $p(z|u)$: 0.44 STDEV $p(z|u)$: 0.17

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
1	256	pete.davis@enron.com	Pete Davis	0.181147
1	20	geir.solberg@enron.com	Geir Solberg	0.089917
1	19	ryan.slinger@enron.com	Ryan Slinger	0.089820
1	14	mark.guzman@enron.com	Mark Guzman	0.089373
1	12	craig.dean@enron.com	Craig Dean	0.085794
1	8	bill.williams@enron.com	Bill Williams III	0.070370
1	152	john.anderson@enron.com	John Anderson	0.053314
1	219	michael.mier@enron.com	Michael Mier	0.053254
1	108	albert.meyers@enron.com		0.052895
1	15	leaf.harasin@enron.com	Leaf Harasin	0.036543
1	17	bert.meyers@enron.com	Bert Meyers	0.036429
1	79	eric.linder@enron.com	Eric Linder	0.032548
1	11	monika.causholli@enron.com	Monika Causholli	0.025655
1	28280	jbryson@enron.com		0.019979
1	28279	dporter3@enron.com		0.019974
1	24	bill.williams.iii@enron.com	bill.williams.iii	0.019682
1	21	kate.symes@enron.com	Kate Symes	0.016673
1	92	holden.salisbury@enron.com	Holden Salisbury	0.009554
1	89	greg.wolfe@enron.com	Greg Wolfe	0.008244
1	16	steven.merris@enron.com	Steven Merris	0.006087

CATEGORY 2

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 619 COMPONENTS: 10
LARGEST COMPONENT SIZE: 585 PERCENT OF TOTAL GRAPH: 94.51%
GROUP DEGREE: 0.19304 GRAPH DENSITY: 0.00647
GROUP CLOSENESS: 0.00232 GROUP BETWEENNESS: 0.27696
AVERAGE p(z|u): 0.23 STDEV p(z|u): 0.26

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
2	4110	klay@enron.com		0.075640
2	1180	karen.denne@enron.com	Karen Denne	0.066720
2	8546	sandra.mccubbin@enron.com	Sandra McCubbin	0.049375
2	181	paul.kaufman@enron.com		0.039195
2	253	jeff.dasovich@enron.com	Jeff Dasovich	0.038103
2	2222	harry.kingerski@enron.com	Harry Kingerski	0.035919
2	1490	steven.kean@enron.com		0.033102
2	1547	mark.palmer@enron.com		0.032138
2	46859	smara@enron.com	"	0.031253
2	1489	james.steffes@enron.com		0.028124

2	801	susan.mara@enron.com.....	Susan Mara.....	0.025373
2	8431	skean@enron.com.....	Steve Kean.....	0.024463
2	2326	janel.guerrero@enron.com.....	0.023961
2	159	david.parquet@enron.com.....	David Parquet.....	0.020838
2	28654	mona.petrochko@enron.com.....	0.020580
2	4851	peggy.mahoney@enron.com.....	0.018540
2	17095	mary.hain@enron.com.....	0.015173
2	34229	mpalmer@enron.com.....	mpalmer@enron.com...	0.014619
2	46814	jdasovic@enron.com.....	"Jeff Dasovich "....	0.014236
2	3495	nicholas.o'day@enron.com.....	0.013499

CATEGORY 3

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 384	COMPONENTS: 3
LARGEST COMPONENT SIZE: 378	PERCENT OF TOTAL GRAPH: 98.44%
GROUP DEGREE: 0.17030	GRAPH DENSITY: 0.00783
GROUP CLOSENESS: 0.02125	GROUP BETWEENNESS: 0.63193
AVERAGE $p(z u)$: 0.33	STDEV $p(z u)$: 0.19

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
3	256	pete.davis@enron.com.....	Pete Davis.....	0.157820
3	14	mark.guzman@enron.com.....	Mark Guzman.....	0.080374
3	19	ryan.slinger@enron.com.....	Ryan Slinger.....	0.079861
3	20	geir.solberg@enron.com.....	Geir Solberg.....	0.079077
3	12	craig.dean@enron.com.....	Craig Dean.....	0.073113
3	15	leaf.harasin@enron.com.....	Leaf Harasin.....	0.060740
3	17	bert.meyers@enron.com.....	Bert Meyers.....	0.060712
3	79	eric.linder@enron.com.....	Eric Linder.....	0.054025
3	11	monika.causholli@enron.com.....	Monika Causholli....	0.051216
3	24	bill.williams.iii@enron.com.....	bill.williams.iii...	0.042756
3	28279	dporter3@enron.com.....	0.042026
3	28280	jbryson@enron.com.....	0.042018
3	8	bill.williams@enron.com.....	Bill Williams III...	0.037072
3	92	holden.salisbury@enron.com.....	Holden Salisbury....	0.025247
3	89	greg.wolfe@enron.com.....	Greg Wolfe.....	0.022114
3	152	john.anderson@enron.com.....	John Anderson.....	0.020014
3	219	michael.mier@enron.com.....	Michael Mier.....	0.019979
3	108	albert.meyers@enron.com.....	0.019555
3	21	kate.symes@enron.com.....	Kate Symes.....	0.018203
3	16	steven.merris@enron.com.....	Steven Merris.....	0.009220

CATEGORY 4

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1417 COMPONENTS: 12
 LARGEST COMPONENT SIZE: 1383 PERCENT OF TOTAL GRAPH: 97.60%
 GROUP DEGREE: 0.06840 GRAPH DENSITY: 0.00212
 GROUP CLOSENESS: 0.00178 GROUP BETWEENNESS: 0.23742
 AVERAGE $p(z|u)$: 0.80 STDEV $p(z|u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
4	367	w.white@enron.com	Stacey W. White	0.073154
4	268	casey.evans@enron.com	Casey Evans	0.062679
4	2591	jim.schwieger@enron.com	Jim Schwieger	0.047768
4	2587	kevin.ruscitti@enron.com	Kevin Ruscitti	0.042155
4	36263	dan.j.hyvl@enron.com	dan.j.hyvl	0.026494
4	62	john.postlethwaite@enron.com	John Postlethwaite	0.026020
4	360	wayne.vinson@enron.com	Donald Wayne Vinson	0.020790
4	251	andrea.dahlke@enron.com	Andrea Dahlke	0.019940
4	739	andrea.ring@enron.com	Andrea Ring	0.019343
4	53779	center.ets@enron.com	ETS Omaha Solution C	0.015400
4	2548	f.keavey@enron.com	Peter F. Keavey	0.015216
4	85273	tlokey@enron.com	WALTER LOKEY	0.014185
4	302	paul.lewis@enron.com	Jon Paul Lewis	0.013956
4	300	norman.lee@enron.com	Norman Lee	0.012247
4	124	chris.stokley@enron.com	Chris Stokley	0.011201
4	2989	tim.carter@enron.com	Tim Carter	0.010998
4	19764	limor.nissan@enron.com		0.010467
4	447	grant.oh@enron.com	Grant Oh	0.008784
4	1333	zhiyun.yang@enron.com	Zhiyun Yang	0.008420
4	2247	tom.chapman@enron.com	Tom Chapman	0.008346

CATEGORY 5

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1483 COMPONENTS: 7
 LARGEST COMPONENT SIZE: 1497 PERCENT OF TOTAL GRAPH: 99.06%
 GROUP DEGREE: 0.12210 GRAPH DENSITY: 0.00201
 GROUP CLOSENESS: 0.00562 GROUP BETWEENNESS: 0.20771
 AVERAGE $p(z|u)$: 0.40 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
5	701	john.lavorato@enron.com	John Lavorato	0.067369
5	1452	david.delainey@enron.com	David Delainey	0.043118
5	37	tim.belden@enron.com	Tim Belden	0.039772
5	817	richard.shapiro@enron.com	Richard Shapiro	0.028748
5	680	mike.grigsby@enron.com	Mike Grigsby	0.028571
5	782	don.black@enron.com	Don Black	0.026678
5	445	rob.milnthorp@enron.com	Rob Milnthorp	0.026577
5	1453	janet.dietrich@enron.com	Janet Dietrich	0.025799
5	347	d.steffes@enron.com	James D. Steffes	0.024981
5	2823	kevin.presto@enron.com		0.024093
5	253	jeff.dasovich@enron.com	Jeff Dasovich	0.023118
5	10241	phillip.allen@enron.com		0.021235
5	293	louise.kitchen@enron.com	Louise Kitchen	0.020737
5	2318	vicki.sharp@enron.com	Vicki Sharp	0.018829
5	168	christopher.calger@enron.com		0.017476
5	2356	kristin.walsh@enron.com	Kristin Walsh	0.016847
5	1786	michael.tribolet@enron.com		0.015361
5	1489	james.steffes@enron.com		0.015189
5	1456	dan.leff@enron.com	Dan Leff	0.014888
5	1752	mark.tawney@enron.com		0.014572

CATEGORY 6

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1217 COMPONENTS: 3
LARGEST COMPONENT SIZE: 1213 PERCENT OF TOTAL GRAPH: 99.67%
GROUP DEGREE: 0.24263 GRAPH DENSITY: 0.00329
GROUP CLOSENESS: 0.02475 GROUP BETWEENNESS: 0.35841
AVERAGE p(z|u): 0.49 STDEV p(z|u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
6	1093	kay.mann@enron.com	Kay Mann	0.302267
6	1651	ben.jacoby@enron.com		0.058002
6	154	sheila.tweed@enron.com	Sheila Tweed	0.035808
6	17542	roseann.engeldorf@enron.com		0.025677
6	20027	kathleen.carnahan@enron.com		0.023955
6	17261	carlos.sole@enron.com		0.023896
6	2386	heather.kroll@enron.com	Heather Kroll	0.021874
6	6899	fred.mitro@enron.com	Fred Mitro	0.021540
6	17208	chris.booth@enron.com		0.019831
6	1568	lisa.bills@enron.com		0.019195

6	588	reagan.rorschach@enron.com.....	Reagan Rorschach....	0.016202
6	12004	suzanne.adams@enron.com.....	Suzanne Adams.....	0.014484
6	190	dale.rasmussen@enron.com.....	0.013117
6	2245	ozzie.pagan@enron.com.....	Ozzie Pagan.....	0.012718
6	24438	john.schwartzenburg@enron.com.....	0.011801
6	1117	david.fairley@enron.com.....	David Fairley.....	0.011529
6	20018	stuart.zisman@enron.com.....	0.011102
6	24415	scott.dieball@enron.com.....	0.011082
6	3638	gregg.penman@enron.com.....	Gregg Penman.....	0.010857
6	10237	herman.manis@enron.com.....	0.009950

CATEGORY 7

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1401 COMPONENTS: 8
LARGEST COMPONENT SIZE: 1379 PERCENT OF TOTAL GRAPH: 98.43%
GROUP DEGREE: 0.14769 GRAPH DENSITY: 0.00214
GROUP CLOSENESS: 0.00293 GROUP BETWEENNESS: 0.29818
AVERAGE $p(z|u)$: 0.45 STDEV $p(z|u)$: 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
7	4135	mark.haedicke@enron.com.....	0.127026
7	15296	jeffrey.hodge@enron.com.....	0.075508
7	275	e.haedicke@enron.com.....	Mark E. Haedicke....	0.049655
7	155	elizabeth.sager@enron.com.....	Elizabeth Sager....	0.042695
7	1092	travis.mccullough@enron.com.....	Travis McCullough...	0.040162
7	318	marcus.nettelton@enron.com.....	Marcus Nettelton....	0.034889
7	266	janette.elbertson@enron.com.....	Janette Elbertson...	0.020736
7	6015	jordan.mintz@enron.com.....	0.016206
7	3110	julia.murray@enron.com.....	0.015917
7	437	peter.keohane@enron.com.....	Peter Keohane.....	0.014761
7	3100	alan.aronowitz@enron.com.....	0.014708
7	2318	vicki.sharp@enron.com.....	Vicki Sharp.....	0.013622
7	5121	eric.thode@enron.com.....	0.013160
7	17252	james.keller@enron.com.....	0.011810
7	17098	janice.moore@enron.com.....	0.011289
7	3474	.schuler@enron.com.....	legal.....	0.010038
7	14697	barbara.gray@enron.com.....	Barbara Gray.....	0.009889
7	445	rob.milnthorp@enron.com.....	Rob Milnthorp.....	0.009384
7	5621	lance.schuler-legal@enron.com.....	0.009148
7	4940	rob.walls@enron.com.....	0.007870

CATEGORY 8

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3089 COMPONENTS: 11
LARGEST COMPONENT SIZE: 3050 PERCENT OF TOTAL GRAPH: 98.74%
GROUP DEGREE: 0.15864 GRAPH DENSITY: 0.00130
GROUP CLOSENESS: 0.00103 GROUP BETWEENNESS: 0.25931
AVERAGE $p(z|u)$: 0.50 STDEV $p(z|u)$: 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
8	18647	carol.clair@enron.com.....		0.096720
8	280	marie.heard@enron.com.....	Marie Heard.....	0.095487
8	15273	dan.hyvl@enron.com.....	Dan J Hyvl.....	0.071036
8	3101	susan.bailey@enron.com.....		0.065159
8	6815	debra.perlingiere@enron.com.....	Debra Perlingiere...	0.064973
8	2365	mary.cook@enron.com.....	Mary Cook.....	0.058524
8	3098	stephanie.panus@enron.com.....	Stephanie Panus....	0.058013
8	3404	samantha.boyd@enron.com.....		0.036209
8	2390	brent.hendry@enron.com.....	Brent Hendry.....	0.034352
8	20682	cheryl.nelson@enron.com.....	Cheryl Nelson.....	0.029334
8	1100	russell.diamond@enron.com.....	Russell Diamond....	0.028116
8	3113	sara.shackleton@enron.com.....		0.025853
8	5889	frank.sayre@enron.com.....		0.023844
8	3103	robert.bruce@enron.com.....		0.020908
8	14787	jason.williams@enron.com.....	Jason Williams.....	0.015493
8	19792	taffy.milligan@enron.com.....		0.015034
8	9117	susan.flynn@enron.com.....		0.014653
8	5897	mark.taylor@enron.com.....		0.014076
8	9094	stacy.dickson@enron.com.....		0.014028
8	20033	kaye.ellis@enron.com.....		0.013844

CATEGORY 9

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2444 COMPONENTS: 4
LARGEST COMPONENT SIZE: 2433 PERCENT OF TOTAL GRAPH: 99.55%
GROUP DEGREE: 0.33079 GRAPH DENSITY: 0.00082
GROUP CLOSENESS: 0.00786 GROUP BETWEENNESS: 0.56920
AVERAGE $p(z|u)$: 0.56 STDEV $p(z|u)$: 0.44

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
9	355	.taylor@enron.com.....	legal.....	0.187008
9	619	.williams@enron.com.....	credit.....	0.055386
9	1488	perfmgmt@enron.com.....	"Performance Evaluat	0.037612
9	606	s..theriot@enron.com.....	Kim S. Theriot.....	0.029576
9	294	c..koehler@enron.com.....	Anne C. Koehler.....	0.028529
9	613	ellen.wallumrod@enron.com.....	Ellen Wallumrod.....	0.025875
9	3099	laurel.adams@enron.com.....	0.025411
9	8306	n..gray@enron.com.....	Barbara N. Gray.....	0.025397
9	15198	e..dickson@enron.com.....	Stacy E. Dickson....	0.024387
9	3409	stacey.richardson@enron.com.....	0.019492
9	284	t..hodge@enron.com.....	Jeffrey T. Hodge....	0.018528
9	9039	theresa.brogan@enron.com.....	0.014218
9	15204	judy.thorne@enron.com.....	Judy Thorne.....	0.013367
9	14777	kay.young@enron.com.....	0.012102
9	3403	cyndie.balfour-flanagan@enron.com.....	0.011474
9	1058	ann.murphy@enron.com.....	Melissa Ann Murphy..	0.011186
9	783	r..brackett@enron.com.....	Debbie R. Brackett..	0.010655
9	506	susan.elledge@enron.com.....	Susan Elledge.....	0.010414
9	7667	ipayit@enron.com.....	iPayit@Enron.com>@EN	0.008657
9	6591	aneela.charania@enron.com.....	Aneela Charania.....	0.007675

CATEGORY 10

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1072 COMPONENTS: 4
LARGEST COMPONENT SIZE: 1064 PERCENT OF TOTAL GRAPH: 99.25%
GROUP DEGREE: 0.23147 GRAPH DENSITY: 0.00373
GROUP CLOSENESS: 0.01247 GROUP BETWEENNESS: 0.38826
AVERAGE p(z|u): 0.35 STDEV p(z|u): 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
10	288	tana.jones@enron.com.....	Tana Jones.....	0.048598
10	1102	tom.moran@enron.com.....	Tom Moran.....	0.043781
10	33	stephanie.sever@enron.com.....	Stephanie Sever.....	0.039835
10	551	lisa.lees@enron.com.....	Lisa Lees.....	0.037141
10	1142	karen.lambert@enron.com.....	Karen Lambert.....	0.030246
10	8874	kelly.lombardi@enron.com.....	Kelly Lombardi.....	0.019002
10	5583	frank.davis@enron.com.....	0.017158
10	16417	larry.hunter@enron.com.....	Larry Joe Hunter....	0.015613
10	5897	mark.taylor@enron.com.....	0.015035
10	486	anthony.campos@enron.com.....	Anthony Campos.....	0.014382

10	1101	tanya.rohauer@enron.com.....	Tanya Rohauer.....	0.013296
10	1449	samuel.schott@enron.com.....	Samuel Schott.....	0.013140
10	6098	debbie.brackett@enron.com.....	0.012464
10	571	melissa.murphy@enron.com.....	Melissa Murphy.....	0.012326
10	1044	rhonda.denton@enron.com.....	Rhonda Denton.....	0.012011
10	480	bob.bowen@enron.com.....	Bob Bowen.....	0.010622
10	565	kevin.meredith@enron.com.....	Kevin Meredith.....	0.010263
10	18036	wendi.lebrocq@enron.com.....	Wendi Lebrocq.....	0.009937
10	6706	bernice.rodriguez@enron.com.....	Bernice Rodriguez...	0.009664
10	4854	william.bradford@enron.com.....	0.009508

CATEGORY 11

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1814 COMPONENTS: 7
LARGEST COMPONENT SIZE: 1725 PERCENT OF TOTAL GRAPH: 95.09%
GROUP DEGREE: 0.20814 GRAPH DENSITY: 0.00331
GROUP CLOSENESS: 0.00064 GROUP BETWEENNESS: 0.29916
AVERAGE $p(z|u)$: 0.65 STDEV $p(z|u)$: 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
11	288	tana.jones@enron.com.....	Tana Jones.....	0.508753
11	7158	mark.greenberg@enron.com.....	Mark Greenberg.....	0.064332
11	1005	mark.fisher@enron.com.....	0.040449
11	1019	leslie.hansen@enron.com.....	Leslie Hansen.....	0.032779
11	401	bob.shults@enron.com.....	Bob Shults.....	0.030348
11	4911	hollis.kimbrough@enron.com.....	0.023842
11	23696	mark.walker@enron.com.....	0.015119
11	617	greg.whiting@enron.com.....	Greg Whiting.....	0.010684
11	23690	jeff.duff@enron.com.....	0.010491
11	23703	kurt.anderson@enron.com.....	0.009879
11	18398	thomas.gros@enron.com.....	0.007953
11	23706	jeff.maurer@enron.com.....	0.007285
11	5543	john.allario@enron.com.....	0.006756
11	23699	kevin.cousineau@enron.com.....	0.006632
11	30596	sarah.wesner@enron.com.....	0.005849
11	23724	joe.thorpe@enron.com.....	0.005811
11	29414	denis.o'connell@enron.com.....	0.004942
11	23705	bo.thisted@enron.com.....	0.004807
11	23713	ronald.brzezinski@enron.com.....	0.004792
11	503	daniel.diamond@enron.com.....	Daniel Diamond.....	0.004184

CATEGORY 12

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 204 COMPONENTS: 14
 LARGEST COMPONENT SIZE: 129 PERCENT OF TOTAL GRAPH: 63.24%
 GROUP DEGREE: 0.24411 GRAPH DENSITY: 0.00493
 GROUP CLOSENESS: 0.00297 GROUP BETWEENNESS: 0.24458
 AVERAGE $p(z|u)$: 0.55 STDEV $p(z|u)$: 0.47

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
12	18633	e.brown@enron.com.....	William E. Brown....	0.008556
12	35765	l.johnson@enron.com.....	David L. Johnson....	0.008521
12	6411	kean@enron.com.....	0.007211
12	65725	.chavez@enron.com.....	e-mail.....	0.006382
12	55759	.kean@enron.com.....	e-mail.....	0.005655
12	65733	.speck@enron.com.....	e-mail.....	0.005176
12	65705	.basile@enron.com.....	e-mail.....	0.005060
12	65706	.benzon@enron.com.....	e-mail.....	0.005060
12	65707	.de@enron.com.....	e-mail.....	0.005060
12	65711	.jones@enron.com.....	e-mail.....	0.005060
12	65718	.rainey@enron.com.....	e-mail.....	0.005060
12	65723	.blizzard@enron.com.....	e-mail.....	0.005060
12	65728	.heeg@enron.com.....	e-mail.....	0.005060
12	65729	.imparato@enron.com.....	e-mail.....	0.005060
12	65731	.norris@enron.com.....	e-mail.....	0.005060
12	65732	.salvo-shook@enron.com.....	e-mail.....	0.005060
12	65738	.barrow@enron.com.....	e-mail.....	0.005060
12	65740	.bustillo@enron.com.....	e-mail.....	0.005060
12	65744	.goss@enron.com.....	e-mail.....	0.005060
12	65746	.lehmann@enron.com.....	e-mail.....	0.005060

CATEGORY 13

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4311 COMPONENTS: 16
 LARGEST COMPONENT SIZE: 4267 PERCENT OF TOTAL GRAPH: 98.98%
 GROUP DEGREE: 0.18886 GRAPH DENSITY: 0.00093
 GROUP CLOSENESS: 0.00087 GROUP BETWEENNESS: 0.33955
 AVERAGE $p(z|u)$: 0.73 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
13	734	dutch.quigley@enron.com.....	Dutch Quigley.....	0.136689
13	2781	enron.announcements@enron.com.....	0.132588
13	1078	40enron@enron.com.....	Tracey Ramsey - Glob	0.127196
13	412	no.address@enron.com.....	0.081646
13	2883	all.houston@enron.com.....	0.042026
13	6007	houston.report@enron.com.....	0.024034
13	86	all.worldwide@enron.com.....	All Enron Worldwide.	0.019420
13	4648	amelia.alder@enron.com.....	0.018362
13	84	all_ena_egm_eim@enron.com.....	All_ENA_EGM_EIM....	0.017372
13	7470	kay.quigley@enron.com.....	Kay Quigley.....	0.008875
13	20046	darlene.forsyth@enron.com.....	0.006007
13	6242	all.downtown@enron.com.....	All Enron Downtown..	0.005775
13	19495	lola.willis@enron.com.....	0.005466
13	5588	all.states@enron.com.....	0.005403
13	8401	sarah.wesner-soong@enron.com.....	Sarah Wesner-Soong..	0.004711
13	6226	enron.announcement@enron.com.....	0.004510
13	21232	'.'john@enron.com.....	work.....	0.003539
13	6227	enron.action@enron.com.....	0.003288
13	34617	kmcdani@enron.com.....	0.003020
13	88	ena.employees@enron.com.....	ENA Employees.....	0.002840

CATEGORY 14

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 870 COMPONENTS: 16
LARGEST COMPONENT SIZE: 834 PERCENT OF TOTAL GRAPH: 95.86%
GROUP DEGREE: 0.22795 GRAPH DENSITY: 0.00230
GROUP CLOSENESS: 0.00234 GROUP BETWEENNESS: 0.64604
AVERAGE p(z|u): 0.36 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
14	1419	billy.lemmons@enron.com.....	Billy Lemmons Jr....	0.013132
14	1703	tim.o'rourke@enron.com.....	0.010658
14	2780	kim.melodick@enron.com.....	0.010649
14	5517	sheila.walton@enron.com.....	0.009514
14	5914	kalen.pieper@enron.com.....	0.008566
14	5512	robert.jones@enron.com.....	0.007844
14	222	david.oxley@enron.com.....	David Oxley.....	0.007823
14	1420	traci.warner@enron.com.....	Traci Warner.....	0.007626
14	3454	exec.jones@enron.com.....	0.006929
14	5511	marla.barnard@enron.com.....	0.006869

14	5516	cindy.skinner@enron.com.....	0.006672
14	11380	shanna.funkhouser@enron.com.....	0.006575
14	4934	gary.smith@enron.com.....	0.006044
14	2201	cindy.olson@enron.com..... Cindy Olson.....	0.005948
14	6013	cynthia.barrow@enron.com.....	0.005482
14	18669	kirk.mcdaniel@enron.com.....	0.005385
14	15227	anne.labbe@enron.com.....	0.005378
14	12915	gary.buck@enron.com..... Gary Buck.....	0.004942
14	5653	ted.bland@enron.com.....	0.004861
14	1832	jana.giovannini@enron.com..... Jana Giovannini.....	0.004700

CATEGORY 15

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2757 COMPONENTS: 11
LARGEST COMPONENT SIZE: 2718 PERCENT OF TOTAL GRAPH: 98.59%
GROUP DEGREE: 0.20953 GRAPH DENSITY: 0.00109
GROUP CLOSENESS: 0.00109 GROUP BETWEENNESS: 0.33939
AVERAGE p(z|u): 0.81 STDEV p(z|u): 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
15	703	matthew.lenhart@enron.com.....	Matthew Lenhart.....	0.153960
15	2575	joe.parks@enron.com.....	Joe Parks.....	0.125052
15	2592	m.scott@enron.com.....	Susan M. Scott.....	0.063080
15	453	cooper.richey@enron.com.....	Cooper Richey.....	0.041295
15	2530	chris.germany@enron.com.....	Chris Germany.....	0.039502
15	7568	monique.sanchez@enron.com.....	Monique Sanchez.....	0.031292
15	37122	wollam.erik@enron.com.....	0.020544
15	737	jay.reitmeyer@enron.com.....	Jay Reitmeyer.....	0.020303
15	1612	j.farmer@enron.com.....	0.018191
15	772	laura.vuittonet@enron.com.....	Laura Vuittonet.....	0.011607
15	37145	constantine.brian@enron.com.....	Brian Constantine...	0.010156
15	6943	gregory.schockling@enron.com.....	Gregory Schockling..	0.009379
15	709	a.martin@enron.com.....	Thomas A. Martin....	0.009292
15	73	troy.denetsosie@enron.com.....	Troy Denetsosie.....	0.008784
15	37115	chet.fenner@enron.com.....	0.008253
15	3520	ragan.bond@enron.com.....	Ragan Bond.....	0.008083
15	3580	tiffany.smith@enron.com.....	Tiffany Smith.....	0.007438
15	2563	brad.mckay@enron.com.....	Brad Mckay.....	0.006387
15	37146	fenner.chet@enron.com.....	Chet Fenner.....	0.005871
15	726	michael.olsen@enron.com.....	Michael Olsen.....	0.005364

CATEGORY 16

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1608 COMPONENTS: 16
LARGEST COMPONENT SIZE: 1574 PERCENT OF TOTAL GRAPH: 97.89%
GROUP DEGREE: 0.09905 GRAPH DENSITY: 0.00249
GROUP CLOSENESS: 0.00183 GROUP BETWEENNESS: 0.41824
AVERAGE $p(z|u)$: 0.44 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
16	6031	outlook.team@enron.com.....		0.041492
16	21091	jerry.harkreader@enron.com.....	Jerry Harkreader....	0.002120
16	811	administration.enron@enron.com.....	Enron Messaging Admi	0.002113
16	740	tina.rode@enron.com.....	Tina Rode.....	0.002088
16	32115	mary.moore@enron.com.....		0.001996
16	5068	audrey.robertson@enron.com.....		0.001991
16	8888	geraldine.irvine@enron.com.....	Geraldine Irvine....	0.001977
16	3470	michael.miller@enron.com.....		0.001941
16	2530	chris.germany@enron.com.....	Chris Germany.....	0.001703
16	21092	jorge.olivares@enron.com.....	Jorge Olivares.....	0.001691
16	739	andrea.ring@enron.com.....	Andrea Ring.....	0.001662
16	2563	brad.mckay@enron.com.....	Brad Mckay.....	0.001659
16	14675	susan.pereira@enron.com.....	Susan Pereira.....	0.001656
16	6837	diane.salcido@enron.com.....	Diane Salcido.....	0.001623
16	35727	sigrid.macpherson@enron.com.....	Sigrid Macpherson...	0.001599
16	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.001593
16	5041	jo.matson@enron.com.....		0.001581
16	6018	suzanne.nicholie@enron.com.....		0.001581
16	6992	judy.hernandez@enron.com.....	Judy Hernandez.....	0.001540
16	14706	peter.keavey@enron.com.....	Peter Keavey.....	0.001538

CATEGORY 17

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3525 COMPONENTS: 5
LARGEST COMPONENT SIZE: 3514 PERCENT OF TOTAL GRAPH: 99.69%
GROUP DEGREE: 0.19220 GRAPH DENSITY: 0.00142
GROUP CLOSENESS: 0.00880 GROUP BETWEENNESS: 0.41953
AVERAGE $p(z|u)$: 0.41 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
17	227	sally.beck@enron.com	Sally Beck	0.169318
17	1771	shona.wilson@enron.com		0.025245
17	6399	beth.apollo@enron.com		0.023601
17	10227	brent.price@enron.com		0.020631
17	6526	mike.jordan@enron.com		0.019276
17	334	leslie.reeves@enron.com	Leslie Reeves	0.018080
17	6396	fernley.dyson@enron.com	Fernley Dyson	0.017788
17	3002	m.hall@enron.com	Bob M Hall	0.014155
17	4796	ted.murphy@enron.com		0.013790
17	14781	patti.thompson@enron.com		0.012955
17	5368	brenda.herod@enron.com		0.011778
17	1978	greg.piper@enron.com		0.011435
17	1543	richard.causey@enron.com		0.010914
17	215	stacey.white@enron.com		0.010912
17	15922	bob.hall@enron.com		0.009304
17	19242	mary.solmonson@enron.com		0.008747
17	6398	chris.abel@enron.com		0.008661
17	3402	kristin.albrecht@enron.com		0.008348
17	8847	scott.mills@enron.com	Scott Mills	0.007933
17	788	david.port@enron.com	David Port	0.007924

CATEGORY 18

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 655 COMPONENTS: 21
LARGEST COMPONENT SIZE: 487 PERCENT OF TOTAL GRAPH: 74.35%
GROUP DEGREE: 0.14953 GRAPH DENSITY: 0.00306
GROUP CLOSENESS: 0.00112 GROUP BETWEENNESS: 0.22792
AVERAGE p(z|u): 0.40 STDEV p(z|u): 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
18	1418	lexi.elliott@enron.com	Lexi Elliott	0.013483
18	6032	kevin.moore@enron.com		0.011379
18	2088	paulo.issler@enron.com	Paulo Issler	0.011040
18	2108	william.smith@enron.com	William Smith	0.009490
18	4662	celeste.roberts@enron.com		0.009256
18	4661	charlene.jackson@enron.com		0.009050
18	15958	kevin.kindall@enron.com		0.008596
18	2068	alex.huang@enron.com	Alex Huang	0.007997
18	2071	anita.dupont@enron.com	Anita Dupont	0.007351
18	1831	kristin.gandy@enron.com	Kristin Gandy	0.007163

18	2079	jose.marquez@enron.com.....	Jose Marquez.....	0.007064
18	2072	bob.lee@enron.com.....	Bob Lee.....	0.006873
18	2105	tom.halliburton@enron.com.....	Tom Halliburton.....	0.006776
18	2513	elena.chilkina@enron.com.....	Elena Chilkina.....	0.006719
18	1543	richard.causey@enron.com.....	0.006501
18	8822	kenneth.parkhill@enron.com.....	Kenneth Parkhill....	0.006418
18	2109	zimin.lu@enron.com.....	Zimin Lu.....	0.006278
18	2075	gwyn.koepke@enron.com.....	Gwyn Koepke.....	0.006249
18	1717	maureen.raymond@enron.com.....	0.006239
18	2243	ashley.baxter@enron.com.....	Ashley Baxter.....	0.005935

CATEGORY 19

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 323	COMPONENTS: 16
LARGEST COMPONENT SIZE: 159	PERCENT OF TOTAL GRAPH: 49.23%
GROUP DEGREE: 0.32463	GRAPH DENSITY: 0.00621
GROUP CLOSENESS: 0.00132	GROUP BETWEENNESS: 0.20860
AVERAGE $p(z u)$: 0.47	STDEV $p(z u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
19	11182	marcus.edmonds@enron.com.....	0.002994
19	4754	samuel.pak@enron.com.....	0.002851
19	11	monika.causholli@enron.com.....	Monika Causholli....	0.002790
19	2032	rahul.seksaria@enron.com.....	Rahul Seksaria.....	0.002786
19	11169	brandon.cavazos@enron.com.....	0.002778
19	1874	jody.crook@enron.com.....	Jody Crook.....	0.002744
19	11211	melanie.king@enron.com.....	0.002713
19	2387	tara.piazze@enron.com.....	Tara Piazze.....	0.002677
19	11227	ravi.mujumdar@enron.com.....	0.002657
19	1878	bryan.hull@enron.com.....	Bryan Hull.....	0.002631
19	7125	peter.bennett@enron.com.....	Peter Bennett.....	0.002620
19	120	susan.rance@enron.com.....	Susan Rance.....	0.002619
19	8849	robin.rodrique@enron.com.....	Robin Rodrigue.....	0.002614
19	11203	avinash.jain@enron.com.....	0.002601
19	6909	erin.willis@enron.com.....	Erin Willis.....	0.002597
19	7537	gisselle.rohmer@enron.com.....	Gisselle Rohmer....	0.002591
19	6657	anthony.sexton@enron.com.....	Anthony Sexton.....	0.002586
19	11252	maria.tefel@enron.com.....	0.002558
19	7106	george.thomas@enron.com.....	George Thomas.....	0.002548
19	6649	binh.pham@enron.com.....	Binh Pham.....	0.002533

CATEGORY 20

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 915 COMPONENTS: 16
LARGEST COMPONENT SIZE: 849 PERCENT OF TOTAL GRAPH: 92.79%
GROUP DEGREE: 0.20328 GRAPH DENSITY: 0.00219
GROUP CLOSENESS: 0.00131 GROUP BETWEENNESS: 0.42725
AVERAGE $p(z|u)$: 0.33 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
20	581	vladi.pimenov@enron.com.....	Vladi Pimenov.....	0.017949
20	696	jared.kaiser@enron.com.....	Jared Kaiser.....	0.012259
20	14675	susan.pereira@enron.com.....	Susan Pereira.....	0.009862
20	2501	kimberly.bates@enron.com.....	Kimberly Bates.....	0.008906
20	764	judy.townsend@enron.com.....	Judy Townsend.....	0.008533
20	756	geoff.storey@enron.com.....	Geoff Storey.....	0.008341
20	1676	laura.luce@enron.com.....	0.008216
20	2346	bryant.frihart@enron.com.....	Bryant Frihart.....	0.008215
20	2520	tom.donohoe@enron.com.....	Tom Donohoe.....	0.008072
20	2515	martin.cuilla@enron.com.....	Martin Cuilla.....	0.007621
20	2563	brad.mckay@enron.com.....	Brad Mckay.....	0.007609
20	1713	s..pollan@enron.com.....	0.007258
20	739	andrea.ring@enron.com.....	Andrea Ring.....	0.007102
20	2497	robin.barbe@enron.com.....	Robin Barbe.....	0.007091
20	747	s..shively@enron.com.....	Hunter S. Shively...	0.007038
20	516	chris.gaskill@enron.com.....	Chris Gaskill.....	0.006917
20	719	l..mims@enron.com.....	Patrice L. Mims.....	0.006884
20	2587	kevin.ruscitti@enron.com.....	Kevin Ruscitti.....	0.006723
20	2600	maureen.smith@enron.com.....	Maureen Smith.....	0.006334
20	724	scott.neal@enron.com.....	Scott Neal.....	0.006252

CATEGORY 21

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 724 COMPONENTS: 16
LARGEST COMPONENT SIZE: 646 PERCENT OF TOTAL GRAPH: 89.23%
GROUP DEGREE: 0.14119 GRAPH DENSITY: 0.00277
GROUP CLOSENESS: 0.00145 GROUP BETWEENNESS: 0.39513
AVERAGE $p(z|u)$: 0.30 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
21	2521	david.draper@enron.com	David Draper	0.002853
21	10232	jim.coffey@enron.com		0.002462
21	721	daniel.muschar@enron.com	Daniel Muschar	0.002232
21	5582	body.shop@enron.com		0.002214
21	23807	ge_benefits@enron.com		0.001920
21	15911	heather.johnson@enron.com		0.001906
21	2607	barry.vanderhorst@enron.com	Barry Vanderhorst	0.001873
21	2536	p.hewitt@enron.com	Jess P. Hewitt	0.001866
21	2595	kristann.shireman@enron.com	Kristann Shireman	0.001858
21	73	troy.denetsosie@enron.com	Troy Denetsosie	0.001822
21	671	keith.dziadek@enron.com	Keith Dziadek	0.001700
21	7019	jason.mcnair@enron.com	Jason McNair	0.001642
21	18195	deborah.bubenko@enron.com		0.001577
21	2581	jessica.presas@enron.com	Jessica Presas	0.001533
21	15289	tommy.yanowski@enron.com		0.001531
21	7111	jeffrey.vincent@enron.com	Jeffrey Vincent	0.001489
21	14742	paul.tate@enron.com	Paul Tate	0.001475
21	2381	richard.ring@enron.com	Richard Ring	0.001469
21	24914	robert.humlicek@enron.com		0.001449
21	1401	tina.holcombe@enron.com	Tina Holcombe	0.001443

CATEGORY 22

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 972 COMPONENTS: 15
LARGEST COMPONENT SIZE: 889 PERCENT OF TOTAL GRAPH: 91.46%
GROUP DEGREE: 0.60909 GRAPH DENSITY: 0.00309
GROUP CLOSENESS: 0.00108 GROUP BETWEENNESS: 0.67923
AVERAGE p(z|u): 0.37 STDEV p(z|u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
22	214	dan.dietrich@enron.com	Dan Dietrich	0.009126
22	516	chris.gaskill@enron.com	Chris Gaskill	0.008891
22	262	david.dronet@enron.com	David Dronet	0.007948
22	1757	colin.tonks@enron.com		0.007760
22	751	bruce.smith@enron.com	Bruce Smith	0.005172
22	780	eddie.zhang@enron.com	Eddie Zhang	0.005133
22	2387	tara.piazze@enron.com	Tara Piazze	0.005064
22	2543	chris.hyde@enron.com	Chris Hyde	0.005020
22	223	a.allen@enron.com	Thresa A. Allen	0.004516
22	317	steve.nat@enron.com	Steve Nat	0.004496

22	281	sonia.hennessy@enron.com.....	Sonia Hennessy.....	0.004337
22	370	min.zheng@enron.com.....	Min Zheng.....	0.004261
22	300	norman.lee@enron.com.....	Norman Lee.....	0.004216
22	665	paige.cox@enron.com.....	Paige Cox.....	0.004049
22	671	keith.dziadek@enron.com.....	Keith Dziadek.....	0.003917
22	638	arun.balasundaram@enron.com.....	Arun Balasundaram...	0.003579
22	367	w.white@enron.com.....	Stacey W. White....	0.003496
22	2599	matt.smith@enron.com.....	Matt Smith.....	0.003439
22	708	danielle.marcinkowski@enron.com.....	Danielle Marcinkowsk	0.003402
22	273	sivakumar.govindasamy@enron.com.....	Sivakumar Govindasam	0.003376

CATEGORY 23

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1589	COMPONENTS: 10
LARGEST COMPONENT SIZE: 1557	PERCENT OF TOTAL GRAPH: 97.99%
GROUP DEGREE: 0.16189	GRAPH DENSITY: 0.00252
GROUP CLOSENESS: 0.00175	GROUP BETWEENNESS: 0.30866
AVERAGE $p(z u)$: 0.58	STDEV $p(z u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
23	2530	chris.germany@enron.com.....	Chris Germany.....	0.053649
23	521	scott.goodell@enron.com.....	Scott Goodell.....	0.024669
23	550	victor.lamadrid@enron.com.....	Victor Lamadrid....	0.023634
23	3005	joann.collins@enron.com.....	Joann Collins.....	0.017200
23	724	scott.neal@enron.com.....	Scott Neal.....	0.015794
23	602	robert.superty@enron.com.....	Robert Superty.....	0.015718
23	764	judy.townsend@enron.com.....	Judy Townsend.....	0.015184
23	466	robert.allwein@enron.com.....	Robert Allwein.....	0.014369
23	558	melba.lozano@enron.com.....	Melba Lozano.....	0.013378
23	514	clarissa.garcia@enron.com.....	Clarissa Garcia....	0.012734
23	2537	john.hodge@enron.com.....	John Hodge.....	0.012242
23	604	tara.sweitzer@enron.com.....	Tara Sweitzer.....	0.011871
23	565	kevin.meredith@enron.com.....	Kevin Meredith.....	0.011767
23	14705	dick.jenkins@enron.com.....	Dick Jenkins.....	0.011169
23	22593	joan.veselack@enron.com.....	0.010447
23	6071	steve.gillespie@enron.com.....	0.010423
23	9194	katherine.kelly@enron.com.....	0.009167
23	2578	w.pereira@enron.com.....	Susan W. Pereira....	0.008454
23	1695	torrey.moorer@enron.com.....	0.008301
23	2608	victoria.versen@enron.com.....	Victoria Versen....	0.007969

CATEGORY 24

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3336 COMPONENTS: 10
LARGEST COMPONENT SIZE: 3315 PERCENT OF TOTAL GRAPH: 99.37%
GROUP DEGREE: 0.09085 GRAPH DENSITY: 0.00120
GROUP CLOSENESS: 0.00246 GROUP BETWEENNESS: 0.15928
AVERAGE $p(z|u)$: 0.80 STDEV $p(z|u)$: 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
24	773	.ward@enron.com.....	houston.....	0.061620
24	15669	sscott5@enron.com.....	0.053147
24	427	chris.dorland@enron.com.....	Chris Dorland.....	0.052079
24	4493	lcampbel@enron.com.....	0.043749
24	1115	clint.dean@enron.com.....	Clint Dean.....	0.041200
24	2599	matt.smith@enron.com.....	Matt Smith.....	0.036769
24	3605	t.lucci@enron.com.....	Paul T. Lucci.....	0.030897
24	548	jeff.king@enron.com.....	Jeff King.....	0.021817
24	2515	martin.cuilla@enron.com.....	Martin Cuilla.....	0.018110
24	14660	theresa.staab@enron.com.....	0.016999
24	1769	mark.whitt@enron.com.....	0.012742
24	2553	tori.kuykendall@enron.com.....	Tori Kuykendall.....	0.012416
24	3111	gerald.nemec@enron.com.....	0.012326
24	4111	jskilli@enron.com.....	0.010394
24	3688	.david@enron.com.....	e-mail.....	0.008333
24	16024	mcuilla@enron.com.....	0.008100
24	939	.scott@enron.com.....	e-mail.....	0.007764
24	2410	.mike@enron.com.....	e-mail.....	0.007218
24	45957	j.bump@enron.com.....	Dan J. Bump.....	0.006795
24	3548	tyrell.harrison@enron.com.....	Tyrell Harrison.....	0.006499

CATEGORY 25

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4141 COMPONENTS: 13
LARGEST COMPONENT SIZE: 4096 PERCENT OF TOTAL GRAPH: 98.91%
GROUP DEGREE: 0.08522 GRAPH DENSITY: 0.00097
GROUP CLOSENESS: 0.00079 GROUP BETWEENNESS: 0.16947
AVERAGE $p(z|u)$: 0.51 STDEV $p(z|u)$: 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
25	1559	john.arnold@enron.com		0.070909
25	765	barry.tycholiz@enron.com	Barry Tycholiz	0.069074
25	680	mike.grigsby@enron.com	Mike Grigsby	0.051839
25	724	scott.neal@enron.com	Scott Neal	0.043372
25	717	stephanie.miller@enron.com	Stephanie Miller	0.035511
25	10241	phillip.allen@enron.com		0.034590
25	629	k.allen@enron.com	Phillip K. Allen	0.027811
25	1769	mark.whitt@enron.com		0.025996
25	712	jonathan.mckay@enron.com	Jonathan Mckay	0.025471
25	687	keith.holst@enron.com	Keith Holst	0.018918
25	761	m.tholt@enron.com	Jane M. Tholt	0.018330
25	673	frank.ermis@enron.com	Frank Ermis	0.016955
25	2553	tori.kuykendall@enron.com	Tori Kuykendall	0.016572
25	10240	hunter.shively@enron.com		0.015896
25	710	larry.may@enron.com	Larry May	0.011314
25	80	john.zufferli@enron.com		0.011302
25	757	patti.sullivan@enron.com	Patti Sullivan	0.010944
25	782	don.black@enron.com	Don Black	0.010707
25	735	ina.rangel@enron.com	Ina Rangel	0.010327
25	675	l.gay@enron.com	Randall L. Gay	0.009031

CATEGORY 26

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2169 COMPONENTS: 11
LARGEST COMPONENT SIZE: 2139 PERCENT OF TOTAL GRAPH: 98.62%
GROUP DEGREE: 0.11042 GRAPH DENSITY: 0.00138
GROUP CLOSENESS: 0.00168 GROUP BETWEENNESS: 0.17887
AVERAGE p(z|u): 0.67 STDEV p(z|u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
26	2530	chris.germany@enron.com	Chris Germany	0.106311
26	1688	ed.mcmichael@enron.com		0.054659
26	2202	a.shankman@enron.com	Jeffrey A. Shankman	0.049366
26	2514	ruth.concannon@enron.com	Ruth Concannon	0.032437
26	3540	maria.garza@enron.com	Maria Garza	0.026319
26	1399	eric.boyt@enron.com	Eric Boyt	0.021885
26	1230	phil.polsky@enron.com	Phil Polsky	0.020944
26	2497	robin.barbe@enron.com	Robin Barbe	0.020038
26	8756	margaret.dhont@enron.com	Margaret Dhont	0.016281
26	3021	scott.hendrickson@enron.com	Scott Hendrickson	0.016172

26	2575	joe.parks@enron.com.....	Joe Parks.....	0.015558
26	761	m.tholt@enron.com.....	Jane M. Tholt.....	0.014317
26	2065	mark.breese@enron.com.....	Mark Breese.....	0.014258
26	11389	garrick.hill@enron.com.....	0.014108
26	3619	louis.dicarlo@enron.com.....	Louis Dicarlo.....	0.013896
26	19787	v.weldon@enron.com.....	0.013674
26	774	charles.weldon@enron.com.....	V. Charles Weldon...	0.011411
26	2549	l.kelly@enron.com.....	Katherine L. Kelly..	0.010852
26	3468	doug.leach@enron.com.....	0.009439
26	17618	mike.mazowita@enron.com.....	Mike Mazowita.....	0.009038

CATEGORY 27

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5208 COMPONENTS: 7
LARGEST COMPONENT SIZE: 5170 PERCENT OF TOTAL GRAPH: 99.27%
GROUP DEGREE: 0.19343 GRAPH DENSITY: 0.00096
GROUP CLOSENESS: 0.00102 GROUP BETWEENNESS: 0.30968
AVERAGE p(z|u): 0.78 STDEV p(z|u): 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
27	1654	j.kaminski@enron.com.....	0.092516
27	1145	benjamin.rogers@enron.com.....	Benjamin Rogers.....	0.074724
27	4058	jeff.skilling@enron.com.....	0.046896
27	226	don.baughman@enron.com.....	Don Baughman Jr.....	0.042775
27	3441	kenneth.lay@enron.com.....	0.041885
27	3644	kim.ward@enron.com.....	Kim Ward.....	0.018471
27	4310	ebass@enron.com.....	0.018182
27	11250	e.taylor@enron.com.....	0.017595
27	607	d.thomas@enron.com.....	Paul D. Thomas.....	0.016358
27	488	mike.carson@enron.com.....	Mike Carson.....	0.016266
27	1140	joe.stepenovitch@enron.com.....	Joe Stepenovitch....	0.016048
27	3659	scott.tholan@enron.com.....	0.015570
27	1691	don.miller@enron.com.....	0.015090
27	19919	lavorato@enron.com.....	0.014358
27	3457	gary.hickerson@enron.com.....	0.014333
27	543	robert.johnston@enron.com.....	Robert Johnston.....	0.012577
27	1657	jeff.kinneman@enron.com.....	0.011582
27	2439	john.brindle@enron.com.....	0.011234
27	3656	britt.whitman@enron.com.....	Britt Whitman.....	0.010267
27	1999	jaime.gualy@enron.com.....	Jaime Gualy.....	0.009446

CATEGORY 28

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1210 COMPONENTS: 10
LARGEST COMPONENT SIZE: 1192 PERCENT OF TOTAL GRAPH: 98.35%
GROUP DEGREE: 0.10583 GRAPH DENSITY: 0.00330
GROUP CLOSENESS: 0.00338 GROUP BETWEENNESS: 0.19778
AVERAGE $p(z|u)$: 0.46 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
28	2359	andy.zipper@enron.com.....	Andy Zipper.....	0.049795
28	1477	greg.whalley@enron.com.....	0.026279
28	4132	john.sherriff@enron.com.....	0.022504
28	403	david.forster@enron.com.....	0.022438
28	404	rahil.jafry@enron.com.....	0.021806
28	401	bob.shults@enron.com.....	Bob Shults.....	0.018842
28	1547	mark.palmer@enron.com.....	0.018281
28	293	louise.kitchen@enron.com.....	Louise Kitchen.....	0.015590
28	1200	savita.puthigai@enron.com.....	Savita Puthigai.....	0.014777
28	595	kal.shah@enron.com.....	Kal Shah.....	0.013226
28	3032	sheri.thomas@enron.com.....	Sheri Thomas.....	0.012491
28	4786	dave.samuels@enron.com.....	0.011941
28	3463	joseph.hirl@enron.com.....	0.010571
28	407	jennifer.denny@enron.com.....	0.010295
28	4131	jay.fitzgerald@enron.com.....	0.009952
28	1180	karen.denne@enron.com.....	Karen Denne.....	0.009583
28	3522	michael.bridges@enron.com.....	Michael Bridges.....	0.009314
28	503	daniel.diamond@enron.com.....	Daniel Diamond.....	0.009080
28	2168	meredith.philipp@enron.com.....	Meredith Philipp....	0.008983
28	3041	george.mcclellan@enron.com.....	George McClellan....	0.008740

CATEGORY 29

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2753 COMPONENTS: 7
LARGEST COMPONENT SIZE: 2739 PERCENT OF TOTAL GRAPH: 99.49%
GROUP DEGREE: 0.10474 GRAPH DENSITY: 0.00145
GROUP CLOSENESS: 0.00486 GROUP BETWEENNESS: 0.18919
AVERAGE $p(z|u)$: 0.53 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
29	9244	richard.sanders@enron.com		0.101607
29	1544	rick.buy@enron.com		0.060187
29	4638	james.derrick@enron.com		0.045886
29	11	monika.causholli@enron.com	Monika Causholli	0.022408
29	4940	rob.walls@enron.com		0.019351
29	17434	britt.davis@enron.com		0.017488
29	4796	ted.murphy@enron.com		0.017237
29	2381	richard.ring@enron.com	Richard Ring	0.015333
29	1625	david.gorte@enron.com		0.012697
29	788	david.port@enron.com	David Port	0.012687
29	17433	becky.zikes@enron.com	Becky Zikes	0.011415
29	2370	vladimir.gorny@enron.com	Vladimir Gorny	0.010935
29	3100	alan.aronowitz@enron.com		0.010412
29	2279	andrew.edison@enron.com	Andrew Edison	0.010259
29	1460	c.williams@enron.com	Robert C. Williams	0.009466
29	481	s.bradford@enron.com	William S. Bradford	0.008875
29	20024	linda.guinn@enron.com		0.008756
29	19400	michael.robison@enron.com		0.008493
29	4790	rex.rogers@enron.com		0.008313
29	72	chip.schneider@enron.com	Chip Schneider	0.008296

CATEGORY 30

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8143 COMPONENTS: 16
LARGEST COMPONENT SIZE: 8073 PERCENT OF TOTAL GRAPH: 99.14%
GROUP DEGREE: 0.06028 GRAPH DENSITY: 0.00086
GROUP CLOSENESS: 0.00036 GROUP BETWEENNESS: 0.06977
AVERAGE p(z|u): 0.87 STDEV p(z|u): 0.27

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
30	12134	kevin.hyatt@enron.com	Kevin Hyatt	0.019493
30	22786	kimberly.watson@enron.com	Kimberly Watson	0.018918
30	8308	steven.harris@enron.com	Steven Harris	0.018019
30	43960	lynn.blair@enron.com		0.017466
30	46814	jdasovic@enron.com	"Jeff Dasovich "	0.016116
30	8303	drew.fossum@enron.com	Drew Fossum	0.015461
30	24943	michelle.lokay@enron.com		0.014495
30	2280	shelley.corman@enron.com	Shelley Corman	0.014230
30	14935	susan.scott@enron.com		0.012075
30	9221	lorraine.lindberg@enron.com		0.012052


```

30 24866 lindy.donoho@enron.com..... 0.011587
30 34417 alewis@ect.enron.com..... Andrew Lewis..... 0.011248
30 21117 tk.lohman@enron.com..... TK Lohman..... 0.010361
30 23333 darrell.schoolcraft@enron.com..... 0.010153
30 23672 jeffery.fawcett@enron.com..... 0.009416
30 29453 vkamins@enron.com..... 0.009225
30 2810 larry.campbell@enron.com..... 0.007889
30 46859 smara@enron.com..... ". 0.006995
30 9096 rick.dietz@enron.com..... 0.006940
30 24902 glen.hass@enron.com..... 0.006572

```

CATEGORY 31

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 2950          COMPONENTS: 9
LARGEST COMPONENT SIZE: 2929 PERCENT OF TOTAL GRAPH: 99.29%
GROUP DEGREE: 0.11836   GRAPH DENSITY: 0.00136
GROUP CLOSENESS: 0.00270 GROUP BETWEENNESS: 0.20908
AVERAGE p(z|u): 0.59   STDEV p(z|u): 0.38

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
31	5897	mark.taylor@enron.com.....		0.180582
31	1637	rod.hayslett@enron.com.....		0.047611
31	18299	tracy.geaccone@enron.com.....	Tracy Geaccone.....	0.038249
31	21042	justin.boyd@enron.com.....	Justin Boyd.....	0.026876
31	3100	alan.aronowitz@enron.com.....		0.024302
31	403	david.forster@enron.com.....		0.020308
31	7573	james.saunders@enron.com.....	James Saunders.....	0.017830
31	20030	david.minns@enron.com.....		0.017761
31	3114	paul.simons@enron.com.....		0.015514
31	22321	edmund.cooper@enron.com.....		0.013771
31	1698	susan.musch@enron.com.....		0.010063
31	2390	brent.hendry@enron.com.....	Brent Hendry.....	0.010054
31	20022	jane.mcbride@enron.com.....		0.009282
31	20015	john.viverito@enron.com.....		0.008698
31	573	dale.neuner@enron.com.....	Dale Neuner.....	0.008112
31	5864	janine.juggins@enron.com.....		0.008012
31	18391	mark.evans@enron.com.....		0.007935
31	15310	bob.chandler@enron.com.....		0.007393
31	3046	jonathan.whitehead@enron.com.....	Jonathan Whitehead..	0.006590
31	603	john.suttle@enron.com.....	John Suttle.....	0.006409

CATEGORY 32

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2781 COMPONENTS: 8
LARGEST COMPONENT SIZE: 2749 PERCENT OF TOTAL GRAPH: 98.85%
GROUP DEGREE: 0.10896 GRAPH DENSITY: 0.00216
GROUP CLOSENESS: 0.00157 GROUP BETWEENNESS: 0.26925
AVERAGE $p(z|u)$: 0.53 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
32	3111	gerald.nemec@enron.com.....		0.205329
32	242	michelle.cash@enron.com.....	Michelle Cash.....	0.116556
32	1174	b..sanders@enron.com.....	Richard B. Sanders..	0.067372
32	14696	lisa.mellencamp@enron.com.....	Lisa Mellencamp.....	0.034565
32	1465	twanda.sweet@enron.com.....		0.023181
32	284	t..hodge@enron.com.....	Jeffrey T. Hodge....	0.023067
32	20018	stuart.zisman@enron.com.....		0.018697
32	14697	barbara.gray@enron.com.....	Barbara Gray.....	0.018534
32	14717	eric.gillaspie@enron.com.....	Eric Gillaspie.....	0.018208
32	17250	steve.hooser@enron.com.....		0.014640
32	2357	brian.redmond@enron.com.....	Brian Redmond.....	0.011554
32	2551	mark.knipa@enron.com.....	Mark Knippa.....	0.010383
32	11156	sharon.butcher@enron.com.....	Sharon Butcher.....	0.009703
32	9264	kriste.sullivan@enron.com.....		0.009641
32	9484	shonnie.daniel@enron.com.....		0.009390
32	1460	c..williams@enron.com.....	Robert C. Williams..	0.009278
32	4805	dan.lyons@enron.com.....		0.009216
32	1645	chris.hilgert@enron.com.....		0.009090
32	1786	michael.tribolet@enron.com.....		0.008538
32	24980	lizzette.palmer@enron.com.....		0.008170

CATEGORY 33

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4374 COMPONENTS: 15
LARGEST COMPONENT SIZE: 4331 PERCENT OF TOTAL GRAPH: 99.02%
GROUP DEGREE: 0.11010 GRAPH DENSITY: 0.00137
GROUP CLOSENESS: 0.00094 GROUP BETWEENNESS: 0.18954
AVERAGE $p(z|u)$: 0.85 STDEV $p(z|u)$: 0.29

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
33	253	jeff.dasovich@enron.com	Jeff Dasovich	0.092484
33	817	richard.shapiro@enron.com	Richard Shapiro	0.059408
33	1489	james.steffes@enron.com		0.037731
33	801	susan.mara@enron.com	Susan Mara	0.032505
33	1490	steven.kean@enron.com		0.027141
33	181	paul.kaufman@enron.com		0.026614
33	347	d.steffes@enron.com	James D. Steffes	0.026058
33	813	sarah.novosel@enron.com		0.019082
33	818	linda.robertson@enron.com	Linda Robertson	0.018814
33	36	alan.comnes@enron.com	Alan Comnes	0.016450
33	2222	harry.kingerski@enron.com	Harry Kingerski	0.015687
33	17095	mary.hain@enron.com		0.015139
33	1474	joe.hartsoe@enron.com		0.012779
33	8546	sandra.mccubbin@enron.com	Sandra McCubbin	0.011379
33	1479	leslie.lawner@enron.com		0.010388
33	1180	karen.denne@enron.com	Karen Denne	0.010284
33	28654	mona.petrochko@enron.com		0.009778
33	66	steve.walton@enron.com	Steve Walton	0.009620
33	800	ray.alvarez@enron.com	Ray Alvarez	0.009360
33	812	l.nicolay@enron.com		0.009146

CATEGORY 34

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1063 COMPONENTS: 10
LARGEST COMPONENT SIZE: 1030 PERCENT OF TOTAL GRAPH: 96.90%
GROUP DEGREE: 0.12894 GRAPH DENSITY: 0.00282
GROUP CLOSENESS: 0.00229 GROUP BETWEENNESS: 0.39638
AVERAGE p(z|u): 0.42 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
34	14776	deb.korkmas@enron.com		0.014711
34	3105	nony.flores@enron.com		0.012424
34	3110	julia.murray@enron.com		0.011986
34	2189	wayne.gresham@enron.com	Wayne Gresham	0.011719
34	255	angela.davis@enron.com	Angela Davis	0.010521
34	3100	alan.aronowitz@enron.com		0.010224
34	15220	lou.stoler@enron.com		0.009819
34	3408	mary.ogden@enron.com		0.009693
34	269	genia.fitzgerald@enron.com	Genia Fitzgerald	0.009521
34	17561	matt.maxwell@enron.com		0.009483

34	12004	suzanne.adams@enron.com.....	Suzanne Adams.....	0.009145
34	4805	dan.lyons@enron.com.....	0.009024
34	20036	brenda.whitehead@enron.com.....	0.008511
34	20015	john.viverito@enron.com.....	0.008412
34	14696	lisa.mellencamp@enron.com.....	Lisa Mellencamp.....	0.008388
34	18646	ann.white@enron.com.....	0.008337
34	8238	pat.radford@enron.com.....	0.008268
34	155	elizabeth.sager@enron.com.....	Elizabeth Sager.....	0.008210
34	154	sheila.tweed@enron.com.....	Sheila Tweed.....	0.008055
34	1465	twanda.sweet@enron.com.....	0.008041

CATEGORY 35

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2420 COMPONENTS: 11
LARGEST COMPONENT SIZE: 2400 PERCENT OF TOTAL GRAPH: 99.17%
GROUP DEGREE: 0.15503 GRAPH DENSITY: 0.00207
GROUP CLOSENESS: 0.00301 GROUP BETWEENNESS: 0.27894
AVERAGE p(z|u): 0.66 STDEV p(z|u): 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
35	3113	sara.shackleton@enron.com.....	0.323216
35	155	elizabeth.sager@enron.com.....	Elizabeth Sager.....	0.099711
35	1101	tanya.rohauer@enron.com.....	Tanya Rohauer.....	0.028220
35	206	christian.yoder@enron.com.....	0.027757
35	590	edward.sacks@enron.com.....	Edward Sacks.....	0.020897
35	437	peter.keohane@enron.com.....	Peter Keohane.....	0.020067
35	4854	william.bradford@enron.com.....	0.019964
35	3029	sheila.glover@enron.com.....	Sheila Glover.....	0.019889
35	144	tracy.ngo@enron.com.....	0.019784
35	1081	paul.radous@enron.com.....	Paul Radous.....	0.017521
35	436	greg.johnston@enron.com.....	Greg Johnston.....	0.016513
35	329	david.portz@enron.com.....	David Portz.....	0.015797
35	1175	arsystem@mailman.enron.com.....	ARSystem.....	0.014454
35	481	s..bradford@enron.com.....	William S. Bradford.	0.011189
35	19925	mark.e.haedicke@enron.com.....	0.011033
35	1090	harlan.murphy@enron.com.....	Harlan Murphy.....	0.010609
35	1091	carol.st.@enron.com.....	Carol St. Clair.....	0.010408
35	17099	shari.stack@enron.com.....	0.010257
35	1019	leslie.hansen@enron.com.....	Leslie Hansen.....	0.010129
35	423	sharon.crawford@enron.com.....	Sharon Crawford.....	0.008823

CATEGORY 36

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1871 COMPONENTS: 8
LARGEST COMPONENT SIZE: 1850 PERCENT OF TOTAL GRAPH: 98.88%
GROUP DEGREE: 0.11577 GRAPH DENSITY: 0.00160
GROUP CLOSENESS: 0.00299 GROUP BETWEENNESS: 0.20861
AVERAGE $p(z|u)$: 0.63 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
36	226	don.baughman@enron.com.....	Don Baughman Jr.....	0.018716
36	1126	tom.may@enron.com.....	Tom May.....	0.017882
36	1120	juan.hernandez@enron.com.....	Juan Hernandez.....	0.016727
36	1104	kayne.coulter@enron.com.....	Kayne Coulter.....	0.015195
36	314	jeffrey.miller@enron.com.....	Jeffrey Miller.....	0.014335
36	292	john.kinser@enron.com.....	John Kinser.....	0.011916
36	591	eric.saibi@enron.com.....	Eric Saibi.....	0.011054
36	1140	joe.stepenovitch@enron.com.....	Joe Stepenovitch....	0.010953
36	477	robert.benson@enron.com.....	Robert Benson.....	0.010849
36	267	joe.errigo@enron.com.....	Joe Errigo.....	0.010644
36	618	lloyd.will@enron.com.....	Lloyd Will.....	0.010380
36	1048	m.forney@enron.com.....	John M. Forney.....	0.010335
36	1096	dean.laurent@enron.com.....	Dean Laurent.....	0.009862
36	276	patrick.hanse@enron.com.....	Patrick Hanse.....	0.009811
36	1148	gautam.gupta@enron.com.....	Gautam Gupta.....	0.009177
36	1132	bill.rust@enron.com.....	Bill Rust.....	0.008977
36	1110	rudy.acevedo@enron.com.....	Rudy Acevedo.....	0.008481
36	1138	doug.sewell@enron.com.....	Doug Sewell.....	0.008430
36	1128	steve.olinde@enron.com.....	Steve Olinde Jr.....	0.008328
36	230	corry.bentley@enron.com.....	Corry Bentley.....	0.007989

CATEGORY 37

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3099 COMPONENTS: 3
LARGEST COMPONENT SIZE: 3095 PERCENT OF TOTAL GRAPH: 99.87%
GROUP DEGREE: 0.11962 GRAPH DENSITY: 0.00161
GROUP CLOSENESS: 0.03531 GROUP BETWEENNESS: 0.30933
AVERAGE $p(z|u)$: 0.48 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
37	293	louise.kitchen@enron.com	Louise Kitchen	0.234172
37	701	john.lavorato@enron.com	John Lavorato	0.046942
37	367	w.white@enron.com	Stacey W. White	0.036714
37	222	david.oxley@enron.com	David Oxley	0.029612
37	756	geoff.storey@enron.com	Geoff Storey	0.023718
37	747	s.shively@enron.com	Hunter S. Shively	0.022801
37	1978	greg.piper@enron.com		0.021923
37	1990	harry.arora@enron.com	Harry Arora	0.021570
37	2357	brian.redmond@enron.com	Brian Redmond	0.021237
37	365	jay.webb@enron.com	Jay Webb	0.020057
37	709	a.martin@enron.com	Thomas A. Martin	0.019743
37	34	f.calger@enron.com	Christopher F. Calge	0.019427
37	743	tammie.schoppe@enron.com	Tammie Schoppe	0.019014
37	403	david.forster@enron.com		0.018450
37	2235	beth.perlman@enron.com	Beth Perlman	0.015320
37	80	john.zufferli@enron.com		0.014154
37	1676	laura.luce@enron.com		0.013896
37	14653	stephen.stock@enron.com		0.013537
37	3608	jean.mrha@enron.com		0.013413
37	279	frank.hayden@enron.com	Frank Hayden	0.011336

CATEGORY 38

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1532 COMPONENTS: 13
LARGEST COMPONENT SIZE: 1505 PERCENT OF TOTAL GRAPH: 98.24%
GROUP DEGREE: 0.15242 GRAPH DENSITY: 0.00131
GROUP CLOSENESS: 0.00236 GROUP BETWEENNESS: 0.29807
AVERAGE p(z|u): 0.36 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
38	495	wes.colwell@enron.com	Wes Colwell	0.029936
38	1462	kimberly.hillis@enron.com		0.018186
38	1461	kay.chapman@enron.com		0.018063
38	1689	jennifer.medcalf@enron.com		0.011990
38	1452	david.delainey@enron.com	David Delainey	0.009455
38	608	shirley.tijerina@enron.com	Shirley Tijerina	0.008885
38	1600	jeff.donahue@enron.com		0.008758
38	5073	marsha.schiller@enron.com		0.008513
38	3457	gary.hickerson@enron.com		0.008457
38	5062	cathy.phillips@enron.com		0.007726

38 3041 george.mcclellan@enron.com..... George Mcclellan.... 0.007609
38 701 john.lavorato@enron.com..... John Lavorato..... 0.007603
38 2357 brian.redmond@enron.com..... Brian Redmond..... 0.007557
38 3484 raymond.bowen@enron.com..... 0.007528
38 4654 jennifer.burns@enron.com..... 0.006991
38 1453 janet.dietrich@enron.com..... Janet Dietrich..... 0.006891
38 227 sally.beck@enron.com..... Sally Beck..... 0.006630
38 743 tammie.schoppe@enron.com..... Tammie Schoppe..... 0.006010
38 740 tina.rode@enron.com..... Tina Rode..... 0.006006
38 8357 airam.artega@enron.com..... 0.005964

CATEGORY 39

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3714 COMPONENTS: 14
LARGEST COMPONENT SIZE: 3668 PERCENT OF TOTAL GRAPH: 98.76%
GROUP DEGREE: 0.10989 GRAPH DENSITY: 0.00108
GROUP CLOSENESS: 0.00082 GROUP BETWEENNESS: 0.13935
AVERAGE $p(z|u)$: 0.79 STDEV $p(z|u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
39	642	eric.bass@enron.com.....	Eric Bass.....	0.075491
39	14935	susan.scott@enron.com.....	0.041222
39	795	john.griffith@enron.com.....	John Griffith.....	0.032742
39	707	mike.maggi@enron.com.....	Mike Maggi.....	0.032217
39	6992	judy.hernandez@enron.com.....	Judy Hernandez.....	0.030989
39	621	jason.wolfe@enron.com.....	Jason Wolfe.....	0.028239
39	719	l.mims@enron.com.....	Patrice L. Mims.....	0.020541
39	2347	h.lewis@enron.com.....	Andrew H. Lewis.....	0.019604
39	8773	michelle.nelson@enron.com.....	Michelle Nelson.....	0.017668
39	1181	exchange.administrator@enron.com.....	0.017242
39	1878	bryan.hull@enron.com.....	Bryan Hull.....	0.016953
39	15321	shanna.husser@enron.com.....	Shanna Husser.....	0.013785
39	8917	timothy.blanchard@enron.com.....	0.013342
39	6702	chad.landry@enron.com.....	Chad Landry.....	0.011253
39	22081	dfarmer@enron.com.....	0.010327
39	15272	phillip.love@enron.com.....	Phillip Love.....	0.009138
39	15748	plove@enron.com.....	0.008331
39	19886	leslie.smith@enron.com.....	0.008298
39	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.008284
39	18727	regina.blackshear@enron.com.....	Regina Blackshear...	0.008093

CATEGORY 40

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 532 COMPONENTS: 17
LARGEST COMPONENT SIZE: 420 PERCENT OF TOTAL GRAPH: 78.95%
GROUP DEGREE: 0.36089 GRAPH DENSITY: 0.00377
GROUP CLOSENESS: 0.00151 GROUP BETWEENNESS: 0.46753
AVERAGE $p(z|u)$: 0.53 STDEV $p(z|u)$: 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
40	549	lisa.kinsey@enron.com.....	Lisa Kinsey.....	0.014421
40	602	robert.superty@enron.com.....	Robert Superty.....	0.010874
40	757	patti.sullivan@enron.com.....	Patti Sullivan.....	0.010638
40	1612	j.farmer@enron.com.....	0.008489
40	550	victor.lamadrid@enron.com.....	Victor Lamadrid.....	0.008465
40	6839	darla.saucier@enron.com.....	Darla Saucier.....	0.008191
40	2998	kirk.lenart@enron.com.....	Kirk Lenart.....	0.007214
40	3000	tammy.gilmore@enron.com.....	Tammy Gilmore.....	0.006947
40	579	cora.pendergrass@enron.com.....	Cora Pendergrass....	0.006820
40	593	l.schrab@enron.com.....	Mark L. Schrab.....	0.006068
40	732	richard.pinion@enron.com.....	Richard Pinion.....	0.005953
40	514	clarissa.garcia@enron.com.....	Clarissa Garcia.....	0.005689
40	3023	brandee.jackson@enron.com.....	Brandee Jackson.....	0.005519
40	14674	s.olinger@enron.com.....	Kimberly S. Olinger.	0.005491
40	3003	christina.sanchez@enron.com.....	Christina Sanchez...	0.005468
40	8755	mark.mcclure@enron.com.....	Mark McClure.....	0.005446
40	1400	donna.greif@enron.com.....	Donna Greif.....	0.005408
40	241	suzanne.calcagno@enron.com.....	Suzanne Calcagno....	0.004959
40	660	suzanne.christiansen@enron.com.....	Suzanne Christiansen	0.004888
40	482	kevin.brady@enron.com.....	Kevin Brady.....	0.004703

CATEGORY 41

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2290 COMPONENTS: 8
LARGEST COMPONENT SIZE: 2261 PERCENT OF TOTAL GRAPH: 98.73%
GROUP DEGREE: 0.23811 GRAPH DENSITY: 0.00218
GROUP CLOSENESS: 0.00169 GROUP BETWEENNESS: 0.34927
AVERAGE $p(z|u)$: 0.40 STDEV $p(z|u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
41	15554	daren.farmer@enron.com		0.059135
41	15921	pat.clynes@enron.com		0.015067
41	11079	melissa.graves@enron.com	Melissa Graves	0.012731
41	14684	robert.cotten@enron.com	Robert Cotten	0.009928
41	11108	julie.meyers@enron.com	Julie Meyers	0.008283
41	1777	rita.wynne@enron.com		0.007517
41	11148	george.weissman@enron.com	George Weissman	0.007224
41	1612	j.farmer@enron.com		0.007084
41	9017	lauri.allen@enron.com		0.006958
41	18648	vance.taylor@enron.com		0.006821
41	11138	edward.terry@enron.com	Edward Terry	0.006820
41	11094	gary.lamphier@enron.com	Gary Lamphier	0.006220
41	11063	howard.camp@enron.com	Howard Camp	0.005986
41	18649	donald.reinhardt@enron.com		0.005897
41	14687	edward.gottlob@enron.com	Edward Gottlob	0.005801
41	7637	susan.smith@enron.com	Susan Smith	0.005795
41	11065	clem.cernosek@enron.com	Clem Cernosek	0.005452
41	11128	carlos.rodriguez@enron.com	Carlos Rodriguez	0.005397
41	15753	jackie.young@enron.com		0.005229
41	20240	robert.lloyd@enron.com		0.005144

CATEGORY 42

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2748 COMPONENTS: 10
LARGEST COMPONENT SIZE: 2628 PERCENT OF TOTAL GRAPH: 95.63%
GROUP DEGREE: 0.08948 GRAPH DENSITY: 0.00182
GROUP CLOSENESS: 0.00043 GROUP BETWEENNESS: 0.09911
AVERAGE p(z|u): 0.42 STDEV p(z|u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
42	706	m.love@enron.com	Phillip M. Love	0.056662
42	698	kam.keiser@enron.com	Kam Keiser	0.055671
42	713	errol.mclaughlin@enron.com	Errol McLaughlin Jr.	0.043189
42	14765	darron.giron@enron.com		0.032354
42	678	c.giron@enron.com	Darron C. Giron	0.031285
42	679	c.gossett@enron.com	Jeffrey C. Gossett	0.024133
42	15272	phillip.love@enron.com	Phillip Love	0.021333
42	644	david.baumbach@enron.com	David Baumbach	0.016250
42	8849	robin.rodrique@enron.com	Robin Rodrigue	0.016190
42	718	bruce.mills@enron.com	Bruce Mills	0.012725

42 6266 jeffrey.gossett@enron.com..... Jeffrey C Gossett... 0.012169
42 730 scott.palmer@enron.com..... B. Scott Palmer.... 0.011753
42 3019 anne.bike@enron.com..... Anne Bike..... 0.011640
42 777 d.winfree@enron.com..... Neal D..... 0.011403
42 767 john.valdes@enron.com..... John Valdes..... 0.009893
42 452 kathy.reeves@enron.com..... Kathy Reeves..... 0.008306
42 2520 tom.donohoe@enron.com..... Tom Donohoe..... 0.007935
42 6878 greg.couch@enron.com..... Greg Couch..... 0.007896
42 23910 mary.fischer@enron.com..... Mary Fischer..... 0.007456
42 303 s.lim@enron.com..... Francis S. Lim..... 0.007159

CATEGORY 43

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 517 COMPONENTS: 22
LARGEST COMPONENT SIZE: 418 PERCENT OF TOTAL GRAPH: 80.85%
GROUP DEGREE: 0.36939 GRAPH DENSITY: 0.00388
GROUP CLOSENESS: 0.00165 GROUP BETWEENNESS: 0.55721
AVERAGE $p(z|u)$: 0.28 STDEV $p(z|u)$: 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
43	11132	tom.shelton@enron.com.....	Tom Shelton.....	0.004754
43	14687	edward.gottlob@enron.com.....	Edward Gottlob.....	0.004671
43	18645	steve.schneider@enron.com.....	0.004615
43	11074	michael.eiben@enron.com.....	Michael Eiben.....	0.004143
43	10232	jim.coffey@enron.com.....	0.003795
43	11055	brad.blevins@enron.com.....	Brad Blevins.....	0.003472
43	11077	irene.flynn@enron.com.....	Irene Flynn.....	0.003461
43	4973	lillian.carroll@enron.com.....	0.003259
43	19231	karry.kendall@enron.com.....	0.003234
43	11066	nick.cocavessis@enron.com.....	Nick Cocavessis.....	0.003201
43	1577	carol.carter@enron.com.....	0.003179
43	20310	emma.welsch@enron.com.....	0.003177
43	15687	james.mckay@enron.com.....	0.003099
43	11088	nathan.hlavaty@enron.com.....	Nathan Hlavaty.....	0.003067
43	1777	rita.wynne@enron.com.....	0.002883
43	10242	thomas.martin@enron.com.....	0.002840
43	11072	cheryl.dudley@enron.com.....	Cheryl Dudley.....	0.002760
43	15578	janet.wallis@enron.com.....	0.002752
43	9146	james.haden@enron.com.....	0.002712
43	11064	molly.carriere@enron.com.....	Molly Carriere.....	0.002711

CATEGORY 44

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3600 COMPONENTS: 8
LARGEST COMPONENT SIZE: 3570 PERCENT OF TOTAL GRAPH: 99.17%
GROUP DEGREE: 0.15460 GRAPH DENSITY: 0.00139
GROUP CLOSENESS: 0.00142 GROUP BETWEENNESS: 0.22946
AVERAGE $p(z|u)$: 0.64 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
44	8	bill.williams@enron.com.....	Bill Williams III...	0.040777
44	21	kate.symes@enron.com.....	Kate Symes.....	0.040168
44	37	tim.belden@enron.com.....	Tim Belden.....	0.031482
44	59	diana.scholtes@enron.com.....	Diana Scholtes.....	0.030788
44	124	chris.stokley@enron.com.....	Chris Stokley.....	0.027118
44	42	jeff.richter@enron.com.....	Jeff Richter.....	0.025783
44	54	sean.crandall@enron.com.....	Sean Crandall.....	0.025680
44	47	cara.semperger@enron.com.....	Cara Semperger.....	0.022904
44	57	matt.motley@enron.com.....	Matt Motley.....	0.022533
44	55	robert.badeer@enron.com.....	Robert Badeer.....	0.022190
44	14	mark.guzman@enron.com.....	Mark Guzman.....	0.022096
44	53	mark.fischer@enron.com.....	Mark Fischer.....	0.021644
44	52	tom.alonso@enron.com.....	Tom Alonso.....	0.021624
44	60	mike.swerzbin@enron.com.....	Mike Swerzbin.....	0.021233
44	93	phillip.platter@enron.com.....	Phillip Platter.....	0.020663
44	38	chris.mallory@enron.com.....	Chris Mallory.....	0.018266
44	92	holden.salisbury@enron.com.....	Holden Salisbury....	0.016172
44	175	lisa.gang@enron.com.....	0.015194
44	58	p.o'neil@enron.com.....	Murray P. Neil.....	0.014506
44	145	stewart.rosman@enron.com.....	Stewart Rosman.....	0.010287

CATEGORY 45

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4800 COMPONENTS: 8
LARGEST COMPONENT SIZE: 4777 PERCENT OF TOTAL GRAPH: 99.52%
GROUP DEGREE: 0.18445 GRAPH DENSITY: 0.00083
GROUP CLOSENESS: 0.00222 GROUP BETWEENNESS: 0.29961
AVERAGE $p(z|u)$: 0.53 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
45	1490	steven.kean@enron.com		0.057847
45	4769	stanley.horton@enron.com		0.039712
45	1454	j.kean@enron.com	Steven J. Kean	0.024714
45	1463	maureen.mcvicker@enron.com		0.023896
45	3536	rosalee.fleming@enron.com	Rosalee Fleming	0.021112
45	1477	greg.whalley@enron.com		0.017872
45	3538	mark.frevert@enron.com	Mark Frevert	0.015626
45	3441	kenneth.lay@enron.com		0.015614
45	2201	cindy.olson@enron.com	Cindy Olson	0.014615
45	3444	jeffrey.mcmahon@enron.com		0.013126
45	1543	richard.causey@enron.com		0.011493
45	3443	mark.koenig@enron.com		0.010834
45	4140	cindy.stark@enron.com		0.010079
45	2176	jim.fallon@enron.com	Jim Fallon	0.009690
45	3484	raymond.bowen@enron.com		0.009399
45	4664	sherri.sera@enron.com		0.009346
45	1637	rod.hayslett@enron.com		0.009284
45	1180	karen.denne@enron.com	Karen Denne	0.008729
45	4058	jeff.skilling@enron.com		0.008640
45	2287	mike.mcconnell@enron.com	Mike Mcconnell	0.008524

CATEGORY 46

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2551 COMPONENTS: 9
LARGEST COMPONENT SIZE: 2529 PERCENT OF TOTAL GRAPH: 99.14%
GROUP DEGREE: 0.31489 GRAPH DENSITY: 0.00118
GROUP CLOSENESS: 0.00267 GROUP BETWEENNESS: 0.53932
AVERAGE p(z|u): 0.44 STDEV p(z|u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
46	584	m.presto@enron.com	Kevin M. Presto	0.072968
46	519	doug.gilbert-smith@enron.com	Doug Gilbert-smith	0.040231
46	499	dana.davis@enron.com	Mark Dana Davis	0.037652
46	1642	rogers.herndon@enron.com		0.031286
46	618	lloyd.will@enron.com	Lloyd Will	0.027930
46	2206	stacey.bolton@enron.com	Stacey Bolton	0.023560
46	601	j.sturm@enron.com	Fletcher J. Sturm	0.022593
46	2381	richard.ring@enron.com	Richard Ring	0.017835
46	516	chris.gaskill@enron.com	Chris Gaskill	0.016085
46	63	elliott.mainzer@enron.com	Elliot Mainzer	0.015746

46	14954	mary.schoen@enron.com.....	0.013194
46	2538	kelly.holman@enron.com.....	Kelly Holman.....	0.012776
46	56	tim.heizenrader@enron.com.....	Tim Heizenrader.....	0.010846
46	812	l.nicolay@enron.com.....	0.010817
46	498	mike.curry@enron.com.....	Mike Curry.....	0.010795
46	5956	michael.terraso@enron.com.....	0.010633
46	95	center.dl-portland@enron.com.....	DL-Portland World Tr	0.010536
46	2242	lisa.jacobson@enron.com.....	Lisa Jacobson.....	0.009560
46	37	tim.belden@enron.com.....	Tim Belden.....	0.009261
46	347	d.steffes@enron.com.....	James D. Steffes....	0.008804

CATEGORY 47

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2855 COMPONENTS: 9
LARGEST COMPONENT SIZE: 2835 PERCENT OF TOTAL GRAPH: 99.30%
GROUP DEGREE: 0.14328 GRAPH DENSITY: 0.00140
GROUP CLOSENESS: 0.00328 GROUP BETWEENNESS: 0.33934
AVERAGE p(z|u): 0.59 STDEV p(z|u): 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
47	5335	vince.kaminski@enron.com.....	0.340735
47	4134	jeffrey.shankman@enron.com.....	0.061597
47	2287	mike.mcconnell@enron.com.....	Mike McConnell.....	0.052482
47	2058	shirley.crenshaw@enron.com.....	Shirley Crenshaw....	0.049972
47	4807	stinson.gibner@enron.com.....	0.040387
47	2106	vasant.shanbhogue@enron.com.....	Vasant Shanbhogue...	0.026319
47	30143	vince.j.kaminski@enron.com.....	Vince J" "Kaminski..	0.017587
47	2102	tanya.tamarchenko@enron.com.....	Tanya Tamarchenko...	0.015234
47	2109	zimin.lu@enron.com.....	Zimin Lu.....	0.015077
47	4654	jennifer.burns@enron.com.....	0.014087
47	20277	grant.masson@enron.com.....	0.012089
47	1666	pinnamaneni.krishnarao@enron.com.....	0.012013
47	1388	christie.patrick@enron.com.....	Christie Patrick....	0.011853
47	6033	mike.roberts@enron.com.....	0.010805
47	4664	sherri.sera@enron.com.....	0.009806
47	4785	john.nowlan@enron.com.....	0.009706
47	2069	amitava.dhar@enron.com.....	Amitava Dhar.....	0.008304
47	1748	dale.surbey@enron.com.....	0.007335
47	1755	ravi.thuraisingham@enron.com.....	0.006944
47	15290	molly.magee@enron.com.....	0.006293

B.2 Author Topic with only Dictionary Words

CATEGORY 0

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6273 COMPONENTS: 10
LARGEST COMPONENT SIZE: 6252 PERCENT OF TOTAL GRAPH: 99.67%
GROUP DEGREE: 0.19910 GRAPH DENSITY: 0.00096
GROUP CLOSENESS: 0.00248 GROUP BETWEENNESS: 0.26975
AVERAGE $p(z|u)$: 0.00 STDEV $p(z|u)$: 0.00

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
0	36	alan.comnes@enron.com.....	Alan Comnes.....	0.000274
0	37	tim.belden@enron.com.....	Tim Belden.....	0.000274
0	256	pete.davis@enron.com.....	Pete Davis.....	0.000274
0	403	david.forster@enron.com.....	0.000274
0	1651	ben.jacoby@enron.com.....	0.000274
0	2222	harry.kingerski@enron.com.....	Harry Kingerski.....	0.000274
0	8	bill.williams@enron.com.....	Bill Williams III...	0.000258
0	34	f.calger@enron.com.....	Christopher F. Calge	0.000258
0	41	h.foster@enron.com.....	Chris H. Foster.....	0.000258
0	42	jeff.richter@enron.com.....	Jeff Richter.....	0.000258
0	51	debra.davidson@enron.com.....	Debra Davidson.....	0.000258
0	58	p.o'neil@enron.com.....	Murray P. Neil.....	0.000258
0	62	john.postlethwaite@enron.com.....	John Postlethwaite..	0.000258
0	66	steve.walton@enron.com.....	Steve Walton.....	0.000258
0	67	dave.perrino@enron.com.....	Dave Perrino.....	0.000258
0	72	chip.schneider@enron.com.....	Chip Schneider.....	0.000258
0	89	greg.wolfe@enron.com.....	Greg Wolfe.....	0.000258
0	99	samantha.law@enron.com.....	0.000258
0	124	chris.stokley@enron.com.....	Chris Stokley.....	0.000258
0	144	tracy.ngo@enron.com.....	0.000258

CATEGORY 1

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7078 COMPONENTS: 12
LARGEST COMPONENT SIZE: 7051 PERCENT OF TOTAL GRAPH: 99.62%
GROUP DEGREE: 0.04205 GRAPH DENSITY: 0.00085
GROUP CLOSENESS: 0.00156 GROUP BETWEENNESS: 0.07972
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
1	8494	trushar.patel@enron.com	Trushar Patel	0.000936
1	25328	tim.mckone@enron.com		0.000857
1	33839	mgocker@enron.com		0.000831
1	29433	legal.4@enron.com		0.000808
1	83451	kevin.d.jordan@enron.com		0.000757
1	34227	robert.c.williams@enron.com	robert.c.williams	0.000701
1	41550	anne.c.koehler@enron.com		0.000695
1	1592	charles.delacey@enron.com		0.000647
1	80963	staci_holtzman@enron.com		0.000628
1	47812	clong@enron.com		0.000625
1	41717	brenda.l.funk@enron.com	Brenda L. "Funk"	0.000623
1	15912	russell.kelley@enron.com		0.000578
1	64778	'deberry@enron.com		0.000563
1	77375	tillett@enron.com		0.000551
1	12277	.gerald@enron.com	e-mail	0.000531
1	41565	sstack@enron.com		0.000524
1	41517	jkeller@enron.com		0.000507
1	41581	bdavis@enron.com		0.000445
1	17749	martin.smith@enron.com	Martin Smith	0.000439
1	48505	kpurbhoo@enron.com		0.000439

CATEGORY 2

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5543 COMPONENTS: 8
 LARGEST COMPONENT SIZE: 5526 PERCENT OF TOTAL GRAPH: 99.69%
 GROUP DEGREE: 0.06472 GRAPH DENSITY: 0.00108
 GROUP CLOSENESS: 0.00353 GROUP BETWEENNESS: 0.08963
 AVERAGE p(z|u): 0.03 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
2	6815	debra.perlingiere@enron.com	Debra Perlingiere	0.000935
2	18962	lynn.shivers@enron.com	Lynn Shivers	0.000868
2	15231	joanne.rozycki@enron.com	Joanne Rozycki	0.000809
2	19052	diane.ellstrom@enron.com		0.000790
2	20224	andrea.guillen@enron.com		0.000638
2	7131	bill.bowes@enron.com	Bill Bowes	0.000605
2	29429	majed.nachawati@enron.com		0.000505
2	533	gordon.heaney@enron.com	Gordon Heaney	0.000477

2	42425	allison.mchenry@enron.com.....	0.000445
2	6673	celeste.cisneros@enron.com..... Celeste Cisneros...	0.000434
2	14756	r.williams@enron.com..... Jason R. Williams...	0.000434
2	81534	una.feeley@enron.com..... Una Feeley.....	0.000434
2	3440	esmeralda.gonzalez@enron.com..... Esmeralda Gonzalez..	0.000422
2	7476	andrew.ralston@enron.com..... Andrew Ralston.....	0.000418
2	301	pinto.leite@enron.com..... Francisco Pinto Leit	0.000416
2	81683	james.canney@enron.com..... James Canney.....	0.000402
2	14777	kay.young@enron.com.....	0.000394
2	17737	carol.north@enron.com..... Carol North.....	0.000386
2	10246	steven.kleege@enron.com.....	0.000379
2	18037	nidia.mendoza@enron.com..... Nidia Mendoza.....	0.000351

CATEGORY 3

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6912 COMPONENTS: 15
LARGEST COMPONENT SIZE: 6870 PERCENT OF TOTAL GRAPH: 99.39%
GROUP DEGREE: 0.08246 GRAPH DENSITY: 0.00087
GROUP CLOSENESS: 0.00078 GROUP BETWEENNESS: 0.11972
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
3	1175	arsystem@mailman.enron.com.....	ARSystem.....	0.001519
3	1488	perfmngmt@enron.com.....	"Performance Evaluat	0.001516
3	22319	perfmngmt@ect.enron.com.....	0.001382
3	41760	matt.dawson@enron.com.....	0.001118
3	20117	arsystem@ect.enron.com.....	0.001083
3	14566	approval.eol.gas.traders@enron.com.....	0.001019
3	9320	information.management@enron.com.....	0.000775
3	39760	fletcher.j.sturm@enron.com.....	"fletcher.j.sturm@en	0.000767
3	41075	daemon.extra@enron.com.....	EXTRA Mailer Daemon.	0.000709
3	19486	steve.beck@enron.com.....	0.000629
3	131	maria.van@enron.com.....	Maria Van houten...	0.000448
3	22380	m.hall@enron.com.....	Bob M. Hall.....	0.000419
3	464	sunil.abraham@enron.com.....	Sunil Abraham.....	0.000389
3	32858	arsystem@enron.com.....	0.000377
3	37128	dl-ga-pas@enron.com.....	DL-GA-PAS.....	0.000371
3	30857	erequest@enron.com.....	0.000365
3	15555	neal.d.winfree@enron.com.....	0.000312
3	6070	scott.loving@enron.com.....	0.000306
3	6936	hakeem.ogunbunmi@enron.com.....	Hakeem Ogunbunmi...	0.000291

3 3006 robert.ramirez@enron.com..... Robert Ramirez..... 0.000288

CATEGORY 4

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6247 COMPONENTS: 8
LARGEST COMPONENT SIZE: 6224 PERCENT OF TOTAL GRAPH: 99.63%
GROUP DEGREE: 0.04842 GRAPH DENSITY: 0.00080
GROUP CLOSENESS: 0.00206 GROUP BETWEENNESS: 0.07966
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
4	8773	michelle.nelson@enron.com.....	Michelle Nelson.....	0.001505
4	742	amanda.rybarski@enron.com.....	Amanda Rybarski.....	0.001426
4	707	mike.maggi@enron.com.....	Mike Maggi.....	0.001151
4	8845	sam.leuschen@enron.com.....	Sam Leuschen.....	0.000394
4	14652	amanda.huble@enron.com.....	Amanda Huble.....	0.000387
4	1153	e.kelly@enron.com.....	Mike E. Kelly.....	0.000383
4	6933	gabriel.monroy@enron.com.....	Gabriel Monroy.....	0.000348
4	5123	margaret.allen@enron.com.....	0.000319
4	18956	cecilia.rodriguez@enron.com.....	Cecilia Rodriguez...	0.000270
4	22344	becky.pitre@enron.com.....	0.000261
4	3643	alexandra.villarreal@enron.com.....	Alexandra Villarreal	0.000240
4	24363	roberts@enron.com.....	0.000223
4	21151	steve.bigalow@enron.com.....	Steve Bigalow.....	0.000209
4	38243	jr.martinez@enron.com.....	0.000206
4	12171	.sheila@enron.com.....	e-mail.....	0.000199
4	30688	donna.dye@enron.com.....	0.000196
4	15949	alexandra.saler@enron.com.....	0.000191
4	14710	james.barker@enron.com.....	James Barker.....	0.000189
4	15234	brent.dornier@enron.com.....	Brent Dornier.....	0.000189
4	19898	corey.hollander@enron.com.....	0.000186

CATEGORY 5

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5864 COMPONENTS: 23
LARGEST COMPONENT SIZE: 5792 PERCENT OF TOTAL GRAPH: 98.77%
GROUP DEGREE: 0.06912 GRAPH DENSITY: 0.00085
GROUP CLOSENESS: 0.00039 GROUP BETWEENNESS: 0.12967
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
5	14993	eserver@enron.com.....	eserver@enron.com@EN	0.001461
5	2985	enron.payroll@enron.com.....	"Enron.Payroll@enron	0.001450
5	3652	payroll.enron@enron.com.....	Enron Payroll.....	0.001429
5	1187	confirmit@enron.com.....	Confirmit.....	0.001350
5	18319	tahnee.stall@enron.com.....	0.001239
5	18318	tammy.marcontell@enron.com.....	0.001230
5	18326	mbx_iscinfra@enron.com.....	0.001111
5	1502	ic@enron.com.....	"ic@enron.com".....	0.001071
5	28467	resources@enron.com.....	"human resources@enr	0.001062
5	9987	jderric@enron.com.....	0.000751
5	28781	enronanywhere@enron.com.....	"enronanywhere@enron	0.000751
5	589	jennifer.rosado@enron.com.....	Jennifer Rosado.....	0.000595
5	38978	communications.internal@enron.com.....	Internal Communicati	0.000520
5	6156	j.harris@enron.com.....	0.000508
5	5539	harora@enron.com.....	0.000484
5	409	bwillia5@enron.com.....	William Williams....	0.000474
5	4636	expense.report@enron.com.....	0.000469
5	26756	talent@enron.com.....	"talent@enron.com"..	0.000465
5	16332	communityrelations@enron.com.....	"communityrelations@	0.000458
5	76892	team.oakland@enron.com.....	Team Oakland.....	0.000437

CATEGORY 6

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7767 COMPONENTS: 21
 LARGEST COMPONENT SIZE: 7689 PERCENT OF TOTAL GRAPH: 99.00%
 GROUP DEGREE: 0.07389 GRAPH DENSITY: 0.00077
 GROUP CLOSENESS: 0.00032 GROUP BETWEENNESS: 0.10976
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
6	8780	kensey_subscriber@mailman.enron.com.....	0.001523
6	20045	restricted.list@enron.com.....	0.001002
6	8594	kkeiser@enron.com.....	0.000723
6	16550	clearhead@mailman.enron.com.....	clearhead@mailman.en	0.000448
6	983	westdesksupport@enron.com.....	0.000311
6	37125	communications@enron.com.....	Communications.....	0.000268
6	58211	massage.therapy@enron.com.....	0.000227
6	22827	crodrig@ect.enron.com.....	0.000225

6	25054	enron.gss@enron.com.....	Enron GSS.....	0.000181
6	60043	critical.notice@enron.com.....	0.000181
6	34474	the.daytrader@enron.com.....	0.000180
6	35531	renee.ratcliff@enron.com.....	Renee Ratcliff.....	0.000180
6	13263	robert.gerry@enron.com.....	Robert Gerry.....	0.000173
6	57767	charles.okechukwu@enron.com.....	Charles Okechukwu...	0.000155
6	57370	sshackl@ect.enron.com.....	"Sara Shackelton " ..	0.000144
6	10231	mason.hamlin@enron.com.....	0.000143
6	5737	steven.bailey@enron.com.....	0.000142
6	62459	make.money@mailman.enron.com.....	0.000138
6	15719	liz.hillman@enron.com.....	0.000136
6	7732	gerosimo@enron.com.....	0.000129

CATEGORY 7

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8984 COMPONENTS: 20
LARGEST COMPONENT SIZE: 8926 PERCENT OF TOTAL GRAPH: 99.35%
GROUP DEGREE: 0.09017 GRAPH DENSITY: 0.00078
GROUP CLOSENESS: 0.00045 GROUP BETWEENNESS: 0.12980
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
7	2883	all.houston@enron.com.....	0.001278
7	6242	all.downtown@enron.com.....	All Enron Downtown..	0.001040
7	2590	lauren.schlesinger@enron.com.....	Lauren Schlesinger..	0.000904
7	5582	body.shop@enron.com.....	0.000847
7	61532	runners@enron.com.....	0.000794
7	6227	enron.action@enron.com.....	0.000703
7	12651	susan.poole@enron.com.....	Susan Poole.....	0.000542
7	8630	donna.teal@enron.com.....	Donna Teal.....	0.000520
7	5679	stan.horton@enron.com.....	0.000492
7	20366	enron.houston@enron.com.....	0.000490
7	53565	wanda.chalk@enron.com.....	0.000470
7	65073	public.houston@enron.com.....	0.000470
7	11117	jennifer.pattison@enron.com.....	Jennifer Pattison..	0.000465
7	8645	dl-ga-all_enron_houston@enron.com.....	DL-GA-all_enron_hous	0.000446
7	43107	unspecified-recipients@enron.com.....	unspecified-recipien	0.000437
7	1078	40enron@enron.com.....	Tracey Ramsey - Glob	0.000427
7	54190	gas.houston@enron.com.....	0.000421
7	19793	charla.stuart@enron.com.....	0.000410
7	412	no.address@enron.com.....	0.000400

7 9372 jeffrey.mcclellan@enron.com..... 0.000382

CATEGORY 8

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8795 COMPONENTS: 19
LARGEST COMPONENT SIZE: 8748 PERCENT OF TOTAL GRAPH: 99.47%
GROUP DEGREE: 0.06955 GRAPH DENSITY: 0.00080
GROUP CLOSENESS: 0.00060 GROUP BETWEENNESS: 0.07979
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
8	18047	project.team@enron.com.....		0.001179
8	2625	jose.favela@enron.com.....		0.001013
8	20280	ernie@enron.com.....		0.000905
8	95	center.dl-portland@enron.com.....	DL-Portland World Tr	0.000720
8	19475	gracie.presas@enron.com.....		0.000640
8	211	portland.desk@enron.com.....	Portland West Desk..	0.000637
8	165	desk.portland@enron.com.....	Portland West Desk..	0.000632
8	4981	magdalena.cruz@enron.com.....		0.000619
8	1843	andrea.richards@enron.com.....	Andrea Richards....	0.000564
8	15291	axisteam@enron.com.....	"The Associate and A	0.000536
8	24080	registrar.isc@enron.com.....	ISC Registrar.....	0.000535
8	18046	lee.steele@enron.com.....		0.000525
8	30153	investinme@enron.com.....		0.000485
8	1827	constance.charles@enron.com.....	Constance Charles...	0.000474
8	776	kevin.whitehurst@enron.com.....	Kevin Whitehurst....	0.000460
8	1183	lisa.jones@enron.com.....	Lisa Jones.....	0.000459
8	29402	project.gem@enron.com.....		0.000438
8	127	josie.jarnagin@enron.com.....		0.000429
8	17268	isc.registrar@enron.com.....		0.000426
8	5380	cheryl.kuehl@enron.com.....		0.000413

CATEGORY 9

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 9817 COMPONENTS: 16
LARGEST COMPONENT SIZE: 9776 PERCENT OF TOTAL GRAPH: 99.58%
GROUP DEGREE: 0.08249 GRAPH DENSITY: 0.00071
GROUP CLOSENESS: 0.00075 GROUP BETWEENNESS: 0.13983
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
9	5600	michael.horning@enron.com		0.000549
9	2368	coo.jeff@enron.com	Jeff McMahon - Presi	0.000511
9	1603	anthony.duenner@enron.com		0.000493
9	1176	ethink@enron.com	ethink	0.000489
9	34025	mitch.meyer@enron.com		0.000470
9	86	all.worldwide@enron.com	All Enron Worldwide.	0.000454
9	5381	matthew.scrimshaw@enron.com		0.000442
9	65258	nate.ellis@enron.com		0.000425
9	34043	mariano.gomez@enron.com	Mariano Gomez	0.000405
9	29742	mcarson@enron.com		0.000403
9	43993	norm.ruiz@enron.com		0.000385
9	54281	new.jun-sept@enron.com		0.000384
9	4088	dorothy.dalton@enron.com		0.000378
9	6320	gail.whipple@enron.com		0.000376
9	19727	john.tollefsen@enron.com		0.000360
9	5620	curly.baca@enron.com		0.000346
9	44158	mike.teal@enron.com		0.000345
9	12527	rodney.derbigny@enron.com	Rodney Derbigny	0.000338
9	51410	rmukherjee@enron.com		0.000335
9	1894	chairman.enron@enron.com	Enron Office Of The	0.000332

CATEGORY 10

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7170 COMPONENTS: 15
 LARGEST COMPONENT SIZE: 7107 PERCENT OF TOTAL GRAPH: 99.12%
 GROUP DEGREE: 0.12595 GRAPH DENSITY: 0.00098
 GROUP CLOSENESS: 0.00047 GROUP BETWEENNESS: 0.18976
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
10	6683	carole.frank@enron.com	Carole Frank	0.001058
10	6752	chance.rabon@enron.com	Chance Rabon	0.000945
10	718	bruce.mills@enron.com	Bruce Mills	0.000893
10	311	hal.mckinney@enron.com	Hal McKinney	0.000845
10	694	brad.jones@enron.com	Brad Jones	0.000845
10	360	wayne.vinson@enron.com	Donald Wayne Vinson	0.000757
10	6579	andres.balmaceda@enron.com	Andres Balmaceda	0.000687
10	730	scott.palmer@enron.com	B. Scott Palmer	0.000665

10	500	sherry.dawson@enron.com.....	Sherry Dawson.....	0.000638
10	594	amanda.schultz@enron.com.....	Amanda Schultz.....	0.000637
10	274	sanjeev.gupta@enron.com.....	Sanjeev Gupta.....	0.000615
10	8763	tiffany.miller@enron.com.....	Tiffany Miller.....	0.000546
10	23681	brian.kristjansen@enron.com.....	0.000521
10	20242	delma.salazar@enron.com.....	0.000511
10	649	randy.bhatia@enron.com.....	Randy Bhatia.....	0.000509
10	754	cathy.sprowls@enron.com.....	Cathy Sprowls.....	0.000505
10	3025	shifali.sharma@enron.com.....	Shifali Sharma.....	0.000475
10	3019	anne.bike@enron.com.....	Anne Bike.....	0.000467
10	452	kathy.reeves@enron.com.....	Kathy Reeves.....	0.000408
10	14656	lee.fascetti@enron.com.....	0.000397

CATEGORY 11

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7614 COMPONENTS: 20
LARGEST COMPONENT SIZE: 7556 PERCENT OF TOTAL GRAPH: 99.24%
GROUP DEGREE: 0.07235 GRAPH DENSITY: 0.00092
GROUP CLOSENESS: 0.00046 GROUP BETWEENNESS: 0.12978
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
11	34417	alewis@ect.enron.com.....	Andrew Lewis.....	0.000879
11	29798	investor@mailboy.enron.com.....	0.000612
11	8694	investor@mailman.enron.com.....	0.000478
11	62034	jwillia@enron.com.....	0.000469
11	29847	hot39d@mailman.enron.com.....	0.000454
11	16172	dbaughm@ect.enron.com.....	0.000446
11	20449	press.release@enron.com.....	0.000398
11	6221	larimore@enron.com.....	0.000391
11	2548	f.keavey@enron.com.....	Peter F. Keavey.....	0.000383
11	34375	alewis@enron.com.....	0.000365
11	34732	andrew.h.lewis@enron.com.....	andrew.h.lewis.....	0.000336
11	6166	glenn.dubin@enron.com.....	0.000320
11	21296	money.in.motion@mailman.enron.com.....	0.000313
11	82051	'appling@enron.com.....	0.000298
11	68426	bmckay@ect.enron.com.....	".....	0.000275
11	41521	brapp@enron.com.....	0.000257
11	19928	pallen@enron.com.....	0.000254
11	36645	sneal@ei.enron.com.....	0.000251
11	57370	sshackl@ect.enron.com.....	"Sara Shackelton".....	0.000220

11 1999 jaime.gualy@enron.com..... Jaime Gualy..... 0.000219

CATEGORY 12

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7712 COMPONENTS: 20
LARGEST COMPONENT SIZE: 7656 PERCENT OF TOTAL GRAPH: 99.27%
GROUP DEGREE: 0.06178 GRAPH DENSITY: 0.00078
GROUP CLOSENESS: 0.00052 GROUP BETWEENNESS: 0.07976
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
12	3053	john.wilson@enron.com.....	John Wilson.....	0.000774
12	14663	bill.briggs@enron.com.....	0.000725
12	14664	phil.clifford@enron.com.....	0.000722
12	30596	sarah.wesner@enron.com.....	0.000432
12	512	jason.fischer@enron.com.....	Jason Fischer.....	0.000336
12	41351	nymex.list@enron.com.....	0.000324
12	3113	sara.shackleton@enron.com.....	0.000323
12	532	reginald.hart@enron.com.....	Reginald Hart.....	0.000316
12	41717	brenda.l.funk@enron.com.....	Brenda L. "Funk".....	0.000311
12	288	tana.jones@enron.com.....	Tana Jones.....	0.000293
12	6673	celeste.cisneros@enron.com.....	Celeste Cisneros....	0.000293
12	10238	mary.ruffer@enron.com.....	0.000286
12	15914	kim.stanley@enron.com.....	0.000265
12	81667	brewer@enron.com.....	0.000249
12	1396	vicsandra.trujillo@enron.com.....	Vicsandra Trujillo..	0.000247
12	8409	trevor.randolph@enron.com.....	Trevor Randolph....	0.000246
12	37136	counsel.dave@enron.com.....	Assistant General Co	0.000227
12	46280	john.west@enron.com.....	0.000207
12	17096	kimberly.allen@enron.com.....	0.000199
12	8953	jason.moore@enron.com.....	Jason Moore.....	0.000193

CATEGORY 13

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7018 COMPONENTS: 18
LARGEST COMPONENT SIZE: 6973 PERCENT OF TOTAL GRAPH: 99.36%
GROUP DEGREE: 0.11584 GRAPH DENSITY: 0.00086
GROUP CLOSENESS: 0.00073 GROUP BETWEENNESS: 0.19976
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
13	3973	admin.enron@enron.com	Enron MailSweeper Ad	0.001527
13	3780	enron.mailsweeper.admin@enron.com		0.001500
13	4987	billy.dorsey@enron.com		0.001323
13	1285	suzanne.danz@enron.com	Suzanne Danz	0.001204
13	16110	dbaughm@notes.enron.com		0.001154
13	31730	enron.messaging.administration@enron.com		0.000990
13	35216	greg.gonzales@enron.com	Greg Gonzales	0.000774
13	35359	mmaggi@notes.enron.com		0.000761
13	3357	adam.senn@enron.com	Adam Senn	0.000735
13	29682	vkamins@notes.enron.com		0.000714
13	25042	victoria.wilbeck@enron.com		0.000696
13	60310	mlehnart@notes.enron.com		0.000665
13	35004	plove@notes.enron.com		0.000646
13	81364	rsander@notes.enron.com		0.000563
13	24890	melody.gray@enron.com		0.000551
13	85760	mhain@ect.enron.com	"	0.000534
13	40561	swhite@notes.enron.com		0.000482
13	3781	crandall@notes.enron.com		0.000480
13	4116	katherine.brown@enron.com		0.000475
13	46388	ed.cattigan@enron.com	Ed Cattigan	0.000448

CATEGORY 14

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7882 COMPONENTS: 20
 LARGEST COMPONENT SIZE: 7831 PERCENT OF TOTAL GRAPH: 99.35%
 GROUP DEGREE: 0.05171 GRAPH DENSITY: 0.00076
 GROUP CLOSENESS: 0.00055 GROUP BETWEENNESS: 0.06977
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
14	60189	pkeavey@ect.enron.com		0.000846
14	81466	mheard@ect.enron.com		0.000810
14	24256	cgerman@ect.enron.com		0.000788
14	77175	esager2@ect.enron.com		0.000702
14	15669	sscott5@enron.com		0.000682
14	60973	kruscit@ect.enron.com	Kevin	0.000611
14	37552	pplatte@ect.enron.com		0.000604
14	71045	mcuilla@ect.enron.com	Martin Cuilla	0.000545

14	71324	tkuyken@ect.enron.com.....	0.000482
14	83277	mtaylor1@ect.enron.com.....	0.000458
14	25054	enron.gss@enron.com..... Enron GSS.....	0.000434
14	73030	jshankm@ect.enron.com..... Wharton Alumni.....	0.000434
14	78508	kholst@enron.com.....	0.000421
14	31824	joseph.lippeatt@enron.com..... Joseph Lippeatt.....	0.000416
14	78538	kholst@ect.enron.com.....	0.000378
14	24019	backroads.travel.update@mailman.enron.co	0.000367
14	41627	scorman@enron.com.....	0.000367
14	71848	ddavis@ect.enron.com.....	0.000365
14	4310	ebass@enron.com.....	0.000355
14	79405	gblair@enron.com.....	0.000330

CATEGORY 15

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5448 COMPONENTS: 18
LARGEST COMPONENT SIZE: 5411 PERCENT OF TOTAL GRAPH: 99.32%
GROUP DEGREE: 0.04648 GRAPH DENSITY: 0.00092
GROUP CLOSENESS: 0.00364 GROUP BETWEENNESS: 0.37696
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
15	3430	system.administrator@enron.com.....		0.001450
15	1181	exchange.administrator@enron.com.....		0.001211
15	29154	postmaster@enron.com.....		0.001062
15	23209	ect.admin@enron.com.....		0.000369
15	68893	idrc.houston.chapter@mailman.enron.com..		0.000261
15	24888	david.glessner@enron.com.....		0.000258
15	30984	lopez@enron.com.....		0.000247
15	21066	rumaldo.lopez@enron.com.....	Rumaldo Lopez.....	0.000210
15	35804	everyone_in_ect_calgary@enron.com.....		0.000201
15	20489	crandal.hardy@enron.com.....		0.000179
15	38713	sumey@enron.com.....		0.000174
15	37125	communications@enron.com.....	Communications.....	0.000172
15	15319	sharon.peace@enron.com.....		0.000170
15	6075	debra.young@enron.com.....		0.000167
15	8409	trevor.randolph@enron.com.....	Trevor Randolph.....	0.000162
15	3000	tammy.gilmore@enron.com.....	Tammy Gilmore.....	0.000161
15	7658	jennifer.oliver@enron.com.....	Jennifer Oliver.....	0.000159
15	39836	ibuyit.approvers@enron.com.....		0.000153
15	19970	rob.bakondy@enron.com.....		0.000148

15 14908 rajesh.chettiar@enron.com..... 0.000142

CATEGORY 16

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7500 COMPONENTS: 25
LARGEST COMPONENT SIZE: 7429 PERCENT OF TOTAL GRAPH: 99.05%
GROUP DEGREE: 0.10801 GRAPH DENSITY: 0.00093
GROUP CLOSENESS: 0.00036 GROUP BETWEENNESS: 0.16978
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
16	1105	j..broderick@enron.com.....	Paul J. Broderick...	0.001183
16	459	garrett.tripp@enron.com.....	Garrett Tripp.....	0.000728
16	23333	darrell.schoolcraft@enron.com.....	0.000672
16	15399	tmartin@enron.com.....	0.000595
16	71298	rgay@enron.com.....	0.000575
16	34375	alewis@enron.com.....	0.000391
16	56955	alexandre.bueno@enron.com.....	0.000352
16	72491	jtholt@enron.com.....	0.000348
16	40227	maurice.gilbert@enron.com.....	Maurice Gilbert.....	0.000311
16	11170	gabriel.chavez@enron.com.....	0.000293
16	23675	roger.westfall@enron.com.....	0.000278
16	11392	john.millar@enron.com.....	0.000270
16	52383	jeff_dasovich@ees.enron.com.....	0.000258
16	11391	james.bryja@enron.com.....	0.000249
16	42065	bill.mangels@enron.com.....	0.000249
16	45145	jsteffe@enron.com.....	0.000244
16	15648	farzad.farhangnia@enron.com.....	0.000242
16	15007	security.sap@enron.com.....	SAP Security.....	0.000231
16	20345	gas.operations@enron.com.....	0.000226
16	11176	milagros.daetz@enron.com.....	0.000210

CATEGORY 17

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8409 COMPONENTS: 16
LARGEST COMPONENT SIZE: 8374 PERCENT OF TOTAL GRAPH: 99.58%
GROUP DEGREE: 0.10491 GRAPH DENSITY: 0.00083
GROUP CLOSENESS: 0.00100 GROUP BETWEENNESS: 0.17981
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
17	3015	ted.evans@enron.com	Ted Evans	0.001210
17	11297	jason.jennaro@enron.com		0.000787
17	18956	cecilia.rodriguez@enron.com	Cecilia Rodriguez	0.000775
17	11322	elizabeth.peters@enron.com		0.000729
17	7076	zachary.mccarroll@enron.com	Zachary McCarroll	0.000569
17	5645	wilson.kriegel@enron.com		0.000535
17	6933	gabriel.monroy@enron.com	Gabriel Monroy	0.000530
17	29138	douglas.nichols@enron.com		0.000527
17	14878	li.sun@enron.com		0.000508
17	3617	chris.cramer@enron.com	Chris Cramer	0.000505
17	36661	donnis.traylor@enron.com		0.000505
17	2197	ted.noble@enron.com	Ted Noble	0.000486
17	34254	donald.miller@enron.com		0.000484
17	2592	m.scott@enron.com	Susan M. Scott	0.000481
17	80350	'nielsen@enron.com		0.000481
17	35941	jmckay2@ect.enron.com	Jon McKay	0.000479
17	14744	zarin.imam@enron.com	Zarin Imam	0.000476
17	3643	alexandra.villarreal@enron.com	Alexandra Villarreal	0.000469
17	9446	susan.weison@enron.com		0.000469
17	26591	ora.cross@enron.com	Ora Cross	0.000469

CATEGORY 18

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7325 COMPONENTS: 18
 LARGEST COMPONENT SIZE: 7282 PERCENT OF TOTAL GRAPH: 99.41%
 GROUP DEGREE: 0.06819 GRAPH DENSITY: 0.00082
 GROUP CLOSENESS: 0.00074 GROUP BETWEENNESS: 0.12972
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
18	63604	lhs-gas.kvammen@enron.com	Kjell - LHS-GAS Kvam	0.001045
18	63612	lhc-gas.kvammen@enron.com	Kjell - LHC-GAS Kvam	0.001022
18	63315	gas.lhc@enron.com	LHC GAS	0.000690
18	63316	hfs.reite@enron.com	NILS - B. Superinten	0.000676
18	3000	tammy.gilmore@enron.com	Tammy Gilmore	0.000587
18	76971	dl-etsgascontrollers@enron.com	DL-ETS Gas Controlle	0.000583
18	44089	ranelle.paladino@enron.com		0.000556
18	53773	controllers.dl-ets@enron.com	DL-ETS Gas Controlle	0.000555

```

18 19807 alma.carrillo@enron.com..... 0.000513
18 7005 jane.joyce@enron.com..... Jane Joyce..... 0.000437
18 53722 angela.white@enron.com..... Angela White..... 0.000430
18 54051 jim.fernie@enron.com..... Jim Fernie..... 0.000377
18 54135 bullets@enron.com..... 0.000367
18 44209 jan.moore@enron.com..... 0.000366
18 76960 pipeline.team@enron.com..... Team Pampa Pipeline,0.000357
18 15299 theresa.branney@enron.com..... 0.000336
18 24804 kelly.allen@enron.com..... 0.000330
18 19826 kim.perez@enron.com..... 0.000325
18 53749 v.dickerson@enron.com..... Steve V Dickerson... 0.000325
18 24885 ava.garcia@enron.com..... 0.000318

```

CATEGORY 19

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 7159          COMPONENTS: 8
LARGEST COMPONENT SIZE: 7145 PERCENT OF TOTAL GRAPH: 99.80%
GROUP DEGREE: 0.05925   GRAPH DENSITY: 0.00084
GROUP CLOSENESS: 0.00460 GROUP BETWEENNESS: 0.10972
AVERAGE p(z|u): 0.02   STDEV p(z|u): 0.01

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
19	7667	ipayit@enron.com.....	iPayit@Enron.com>@EN	0.001524
19	11479	ibuyit.payables@enron.com.....	iBuyit.Payables@Enro	0.001384
19	3602	robert.jones@mailman.enron.com.....		0.001375
19	6520	payables.ibuyit@enron.com.....	iBuyit Payables....	0.001243
19	73730	mariachi.el@enron.com.....	El Mariachi.....	0.001086
19	21165	carolyn.graham@enron.com.....	Carolyn Graham.....	0.000801
19	2629	bbutler2@enron.com.....		0.000701
19	1115	clint.dean@enron.com.....	Clint Dean.....	0.000438
19	8342	ibuyit@enron.com.....		0.000382
19	17268	isc.registrar@enron.com.....		0.000242
19	29191	vanessa.griffin@enron.com.....	Vanessa Griffin....	0.000175
19	30857	erequest@enron.com.....		0.000168
19	39836	ibuyit.approvers@enron.com.....		0.000167
19	15060	ashu.tewari@enron.com.....	Ashu Tewari.....	0.000163
19	14905	bradley.stewart@enron.com.....		0.000147
19	79464	quickplace@nahou-lnw01.ots.enron.com...	"customerservice"...	0.000145
19	15556	sap.coe@enron.com.....		0.000142
19	822	portland.dl-ubsw@enron.com.....	DL-UBSW Energy Port1	0.000132
19	17913	deborah.heath@enron.com.....	Deborah Heath.....	0.000127

19 27028 john.chambers@enron.com..... 0.000127

CATEGORY 20

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7086 COMPONENTS: 11
LARGEST COMPONENT SIZE: 7065 PERCENT OF TOTAL GRAPH: 99.70%
GROUP DEGREE: 0.12786 GRAPH DENSITY: 0.00085
GROUP CLOSENESS: 0.00285 GROUP BETWEENNESS: 0.23978
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
20	4320	erwin.landivar@enron.com.....		0.000409
20	20271	north.america@enron.com.....		0.000383
20	36085	north.america_europe@enron.com.....		0.000358
20	19300	coralie.evans@enron.com.....		0.000332
20	19088	neil.tarling@enron.com.....	Neil Tarling.....	0.000323
20	227	sally.beck@enron.com.....	Sally Beck.....	0.000314
20	7163	holly.heath@enron.com.....	Holly Heath.....	0.000295
20	1844	corina.taylor@enron.com.....	Corina Taylor.....	0.000277
20	63664	enw.all@enron.com.....		0.000271
20	15536	jay.smith@enron.com.....	Jay Smith.....	0.000260
20	18400	bill.gulyassy@enron.com.....		0.000252
20	19258	egm.employees@enron.com.....		0.000246
20	20456	peter.ghavami@enron.com.....		0.000242
20	7005	jane.joyce@enron.com.....	Jane Joyce.....	0.000239
20	14726	george.hope@enron.com.....	George Hope.....	0.000227
20	56873	elizabeth.serralheiro@enron.com.....		0.000226
20	23554	enw.piper@enron.com.....		0.000208
20	6879	geynille.dillingham@enron.com.....	Geynille Dillingham.	0.000204
20	18607	sally_beck@enron.com.....		0.000198
20	19116	sbeck2@enron.com.....		0.000182

CATEGORY 21

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6667 COMPONENTS: 13
LARGEST COMPONENT SIZE: 6617 PERCENT OF TOTAL GRAPH: 99.25%
GROUP DEGREE: 0.08622 GRAPH DENSITY: 0.00090
GROUP CLOSENESS: 0.00058 GROUP BETWEENNESS: 0.12973
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
21	1145	benjamin.rogers@enron.com	Benjamin Rogers	0.000686
21	84535	rring@ees.enron.com	NYISO TIE List	0.000629
21	54044	'ball@enron.com		0.000520
21	6324	mauboussin@enron.com		0.000345
21	30924	maria.garcia@enron.com		0.000259
21	76603	benjamin.rogers@ect.enron.com	"	0.000253
21	4182	brogers2@enron.com		0.000244
21	34474	the.daytrader@enron.com		0.000204
21	607	d.thomas@enron.com	Paul D. Thomas	0.000202
21	21345	subscriber@mailboy.enron.com		0.000181
21	69527	stock.option.grant.list@enron.com		0.000179
21	41601	rshapiro@enron.com		0.000175
21	85	ebiz@enron.com	eBiz	0.000170
21	6859	j.vitrella@enron.com	David J. Vitrella	0.000153
21	15719	liz.hillman@enron.com		0.000144
21	75988	brogers2@ect.enron.com		0.000141
21	37501	julio.guzman@enron.com	Julio Guzman	0.000136
21	37149	union.credit@enron.com	Credit Union	0.000126
21	19371	danny.wilson@enron.com		0.000124
21	2980	mcurry@enron.com		0.000120

CATEGORY 22

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4952 COMPONENTS: 12
 LARGEST COMPONENT SIZE: 4926 PERCENT OF TOTAL GRAPH: 99.47%
 GROUP DEGREE: 0.07514 GRAPH DENSITY: 0.00121
 GROUP CLOSENESS: 0.00197 GROUP BETWEENNESS: 0.12954
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
22	19980	harry.collins@enron.com		0.001380
22	22502	gary.a.hanks@enron.com	gary.a.hanks	0.001323
22	18664	brian.lindsay@enron.com	Brian Lindsay	0.001213
22	15567	earl.tisdale@enron.com		0.001114
22	37910	amy.heffernan@enron.com		0.001111
22	2995	juana.fayett@enron.com	Juana Fayett	0.001107
22	250	cynthia.clark@enron.com	Cynthia Clark	0.001098
22	18180	sonya.clarke@enron.com		0.001034

22	6265	jana.morse@enron.com.....	Jana Morse.....	0.000976
22	17783	paul.maley@enron.com.....	Paul Maley.....	0.000941
22	2186	robby.rossi@enron.com.....	Robby Rossi.....	0.000880
22	2993	trang.le@enron.com.....	Trang Le.....	0.000857
22	19791	nicole.hunter@enron.com.....	0.000844
22	19629	albert.escamilla@enron.com.....	0.000810
22	8876	tandra.coleman@enron.com.....	Tandra Coleman.....	0.000769
22	18181	tim.davies@enron.com.....	0.000745
22	22068	enron.counterparty@enron.com.....	0.000702
22	29291	.cooper@enron.com.....	ebs.....	0.000643
22	1094	karen.o'day@enron.com.....	Karen day.....	0.000641
22	1202	center.eol@enron.com.....	0.000625

CATEGORY 23

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8278 COMPONENTS: 16
LARGEST COMPONENT SIZE: 8212 PERCENT OF TOTAL GRAPH: 99.20%
GROUP DEGREE: 0.08240 GRAPH DENSITY: 0.00085
GROUP CLOSENESS: 0.00039 GROUP BETWEENNESS: 0.14980
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
23	20676	transportation.parking@enron.com.....	Parking & Transporta	0.000669
23	15274	parking.transportation@enron.com.....	Parking & Transporta	0.000557
23	2363	morris.larubbio@enron.com.....	Morris Larubbio.....	0.000539
23	20547	lance.jameson@enron.com.....	0.000530
23	772	laura.vuittonet@enron.com.....	Laura Vuittonet.....	0.000507
23	2501	kimberly.bates@enron.com.....	Kimberly Bates.....	0.000497
23	8764	jennifer.mendez@enron.com.....	Jennifer Mendez.....	0.000486
23	23522	john.o'conner@enron.com.....	John Conner.....	0.000480
23	650	jae.black@enron.com.....	Tamara Jae Black....	0.000467
23	64095	devries@enron.com.....	0.000440
23	11174	lindsay.culotta@enron.com.....	0.000438
23	779	becky.young@enron.com.....	Becky Young.....	0.000427
23	7028	joseph.nieten@enron.com.....	Joseph Nieten.....	0.000426
23	8357	airam.arteaga@enron.com.....	0.000424
23	655	karen.buckley@enron.com.....	Karen Buckley.....	0.000417
23	353	mark.symms@enron.com.....	Mark Symms.....	0.000413
23	64088	mcmichael@enron.com.....	0.000393
23	15546	laura.harder@enron.com.....	0.000376
23	2188	s.gartner@enron.com.....	Julie S. Gartner....	0.000358

23 6097 simone.rose@enron.com..... 0.000352

CATEGORY 24

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7781 COMPONENTS: 19
LARGEST COMPONENT SIZE: 7727 PERCENT OF TOTAL GRAPH: 99.31%
GROUP DEGREE: 0.11346 GRAPH DENSITY: 0.00077
GROUP CLOSENESS: 0.00054 GROUP BETWEENNESS: 0.20979
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
24	7009	jay.knoblauch@enron.com.....	Jay Knoblauch.....	0.000818
24	15206	jeff.stephens@enron.com.....	Jeff Stephens.....	0.000746
24	1599	john.disturnal@enron.com.....	0.000741
24	2505	mara.bronstein@enron.com.....	Mara Bronstein.....	0.000708
24	8829	jeffery.stephens@enron.com.....	Jeffery Stephens....	0.000613
24	52900	'williams@enron.com.....	0.000534
24	14777	kay.young@enron.com.....	0.000531
24	3520	ragan.bond@enron.com.....	Ragan Bond.....	0.000524
24	35820	greg.frers@enron.com.....	Greg Frers.....	0.000454
24	6943	gregory.schockling@enron.com.....	Gregory Schockling..	0.000421
24	465	dipak.agarwalla@enron.com.....	Dipak Agarwalla....	0.000389
24	72408	'fildes@enron.com.....	0.000347
24	60731	turner@enron.com.....	0.000344
24	48843	roger_yang@enron.com.....	0.000313
24	81783	woodruff@enron.com.....	0.000302
24	2548	f.keavey@enron.com.....	Peter F. Keavey....	0.000287
24	81795	'bevans@enron.com.....	0.000278
24	773	.ward@enron.com.....	houston.....	0.000273
24	16324	tracey.kari@enron.com.....	0.000272
24	71735	'hadix@enron.com.....	0.000272

CATEGORY 25

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8513 COMPONENTS: 19
LARGEST COMPONENT SIZE: 8459 PERCENT OF TOTAL GRAPH: 99.37%
GROUP DEGREE: 0.04440 GRAPH DENSITY: 0.00094
GROUP CLOSENESS: 0.00052 GROUP BETWEENNESS: 0.06979
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
25	1895	dl-ga-all_enron_worldwide2@enron.com....	DL-GA-all_enron_worl	0.000964
25	1889	dl-ga-all_enron_worldwide1@enron.com....	DL-GA-all_enron_worl	0.000859
25	6180	enron.expertfinder@enron.com.....		0.000859
25	1914	chairman.ken@enron.com.....	Ken Lay - Office of	0.000818
25	1894	chairman.enron@enron.com.....	Enron Office Of The	0.000807
25	17603	dl-ga-all_enron_worldwide5@enron.com....	DL-GA-ALL_enron_worl	0.000785
25	2377	dl-ga-all_enron_worldwide@enron.com....	DL-GA-all_enron_worl	0.000732
25	1497	legalonline-compliance@enron.com.....	Office of the Chairm	0.000714
25	6226	enron.announcement@enron.com.....		0.000693
25	87	enron.chairman@enron.com.....	Enron Americas - Off	0.000674
25	54525	dl-ga-all_egs@enron.com.....	DL-GA-all_egs.....	0.000659
25	4762	office.chairman@enron.com.....		0.000656
25	17421	dl-ga-all_enron_worldwide4@enron.com....	DL-GA-all_enron_worl	0.000640
25	13404	deane.pierce@enron.com.....	Deane Pierce.....	0.000634
25	5688	ken.skilling@enron.com.....		0.000618
25	15799	all.employees@enron.com.....		0.000618
25	88	ena.employees@enron.com.....	ENA Employees.....	0.000581
25	17604	dl-ga-all_enron_worldwide6@enron.com....	DL-GA-ALL_enron_worl	0.000579
25	55972	lprior@enron.com.....	"......	0.000578
25	2368	coo.jeff@enron.com.....	Jeff McMahon - Presi	0.000577

CATEGORY 26

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4645 COMPONENTS: 16
 LARGEST COMPONENT SIZE: 4609 PERCENT OF TOTAL GRAPH: 99.22%
 GROUP DEGREE: 0.12181 GRAPH DENSITY: 0.00086
 GROUP CLOSENESS: 0.00099 GROUP BETWEENNESS: 0.19950
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
26	6031	outlook.team@enron.com.....		0.001577
26	215	stacey.white@enron.com.....		0.001282
26	7112	zionette.vincent@enron.com.....	Zionette Vincent....	0.000583
26	34175	kathryn.thomas@enron.com.....	Kathryn Thomas.....	0.000353
26	5377	sean.long@enron.com.....		0.000286
26	6083	mike.thomas@enron.com.....		0.000262
26	39763	valeria.a.hope@enron.com.....		0.000231
26	39764	roxann.salina.enronxgate@enron.com.....		0.000229

26	21095	renee.pena@enron.com.....	Renee Pena.....	0.000188
26	21091	jerry.harkreader@enron.com.....	Jerry Harkreader....	0.000180
26	20005	cuthbert.roberts@enron.com.....	0.000176
26	17830	cheryl.oliver@enron.com.....	0.000172
26	6077	john.reese@enron.com.....	0.000165
26	39765	vance.bates@enron.com.....	0.000157
26	20013	david.terlip@enron.com.....	0.000154
26	19903	cooper@enron.com.....	0.000151
26	30858	enron.customers@enron.com.....	0.000144
26	11357	ching.lun@enron.com.....	0.000139
26	1782	joe.zhou@enron.com.....	0.000136
26	1116	todd.decook@enron.com.....	Todd Decook.....	0.000135

CATEGORY 27

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6516 COMPONENTS: 15
LARGEST COMPONENT SIZE: 6478 PERCENT OF TOTAL GRAPH: 99.42%
GROUP DEGREE: 0.13563 GRAPH DENSITY: 0.00077
GROUP CLOSENESS: 0.00102 GROUP BETWEENNESS: 0.25973
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
27	14855	luis.mena@enron.com.....	0.000894
27	703	matthew.lenhart@enron.com.....	Matthew Lenhart....	0.000597
27	642	eric.bass@enron.com.....	Eric Bass.....	0.000520
27	1878	bryan.hull@enron.com.....	Bryan Hull.....	0.000459
27	62798	matt.sample@enron.com.....	Matt Sample.....	0.000442
27	1383	nick.hiemstra@enron.com.....	Nick Hiemstra.....	0.000415
27	16849	dbaughm@enron.com.....	0.000414
27	8917	timothy.blanchard@enron.com.....	0.000400
27	14834	micah.hatten@enron.com.....	0.000391
27	62801	allan.elliott@enron.com.....	Allan Elliott.....	0.000378
27	62784	michael.cherry@enron.com.....	Michael Cherry.....	0.000377
27	1881	greg.martin@enron.com.....	Greg Martin.....	0.000360
27	2614	christa.winfrey@enron.com.....	Christa Winfrey....	0.000321
27	8771	jackson.logan@enron.com.....	Jackson Logan III...	0.000317
27	66671	moscoso@enron.com.....	0.000278
27	37969	molnar.mark@enron.com.....	0.000274
27	15272	phillip.love@enron.com.....	Phillip Love.....	0.000266
27	8776	thomas.underwood@enron.com.....	Thomas Underwood....	0.000260
27	37170	'allison'@enron.com.....	Allison.....	0.000250

27 19276 nicholas.stephan@enron.com..... 0.000247

CATEGORY 28

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7415 COMPONENTS: 16
LARGEST COMPONENT SIZE: 7374 PERCENT OF TOTAL GRAPH: 99.45%
GROUP DEGREE: 0.10942 GRAPH DENSITY: 0.00094
GROUP CLOSENESS: 0.00078 GROUP BETWEENNESS: 0.16977
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Table with 4 columns: Topic#, ID#, Email Address, Name, p(z|u). Lists 28 users with their respective probabilities.

CATEGORY 29

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8140 COMPONENTS: 13
LARGEST COMPONENT SIZE: 8100 PERCENT OF TOTAL GRAPH: 99.51%
GROUP DEGREE: 0.09964 GRAPH DENSITY: 0.00086
GROUP CLOSENESS: 0.00079 GROUP BETWEENNESS: 0.14979
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
29	22613	tdonoho@enron.com		0.001113
29	30920	kwatson@enron.com		0.000933
29	35461	tmartin@ect.enron.com		0.000874
29	39519	jquenet@enron.com	Quenet	0.000835
29	82058	mwhitt@ect.enron.com		0.000827
29	77620	fermis@ect.enron.com		0.000745
29	59958	tdonoho@ect.enron.com		0.000729
29	58588	"undisclosed-recipient"@enron.com	"Undisclosed-Recipie	0.000652
29	11411	phyllis.anzalone@enron.com		0.000646
29	17265	fsturm@enron.com	fsturm	0.000632
29	44859	mwhitt@enron.com		0.000609
29	24866	lindy.donoho@enron.com		0.000561
29	4556	mlehart@enron.com		0.000535
29	2258	rita.hartfield@enron.com	Rita Hartfield	0.000467
29	35562	esource@enron.com	eSource	0.000373
29	79405	gblair@enron.com		0.000363
29	2453	undisclosed-recipients@enron.com	undisclosed-recipien	0.000332
29	29712	vkamins@ect.enron.com		0.000331
29	17341	rbuy@enron.com		0.000328
29	41633	fking@enron.com		0.000315

CATEGORY 30

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7437 COMPONENTS: 18
 LARGEST COMPONENT SIZE: 7383 PERCENT OF TOTAL GRAPH: 99.27%
 GROUP DEGREE: 0.10751 GRAPH DENSITY: 0.00081
 GROUP CLOSENESS: 0.00053 GROUP BETWEENNESS: 0.15977
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
30	31503	grant_masson@ei.enron.com		0.000438
30	2082	kenneth.deng@enron.com	Kenneth Deng	0.000435
30	19763	mary.bailey@enron.com		0.000361
30	30940	vince_j_kaminski@enron.com	"Vince J Kaminski"	0.000353
30	32865	network.security@enron.com		0.000352
30	10959	althea.gordon@enron.com		0.000346
30	2078	jason.sokolov@enron.com	Jason Sokolov	0.000341
30	31619	lenos.trigeorgis@enron.com		0.000330

30	9251	rehman.sharif@enron.com.....	0.000327
30	19479	nedre.strambler@enron.com.....	0.000326
30	32307	andrew.s.fastow@enron.com..... Andrew.S.Fastow....	0.000320
30	30722	benjamin.parsons@enron.com.....	0.000316
30	8822	kenneth.parkhill@enron.com..... Kenneth Parkhill....	0.000305
30	8909	a.cote@enron.com..... John A. Cote.....	0.000299
30	30143	vince.j.kaminski@enron.com..... Vince J" "Kaminski..	0.000289
30	18538	stephanie.taylor@enron.com..... Stephanie Taylor....	0.000283
30	53111	keffer.lesley@enron.com..... Lesley Keffer.....	0.000278
30	29453	vkamins@enron.com.....	0.000274
30	29971	kamins@enron.com..... kamins@enron.com....	0.000267
30	20956	maggie.li@enron.com.....	0.000260

CATEGORY 31

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6597 COMPONENTS: 19
LARGEST COMPONENT SIZE: 6546 PERCENT OF TOTAL GRAPH: 99.23%
GROUP DEGREE: 0.12270 GRAPH DENSITY: 0.00076
GROUP CLOSENESS: 0.00062 GROUP BETWEENNESS: 0.21973
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
31	2005	thomas.lowell@enron.com.....	Thomas Lowell.....	0.000976
31	77707	plucci@enron.com.....	Paul Lucci.....	0.000781
31	323	seung-taek.oh@enron.com.....	Seung-Taek Oh.....	0.000579
31	620	ryan.williams@enron.com.....	Ryan Williams.....	0.000393
31	494	kevin.cline@enron.com.....	Kevin Cline.....	0.000376
31	15540	natalie.wells@enron.com.....	Natalie Wells.....	0.000212
31	3605	t.lucci@enron.com.....	Paul T. Lucci.....	0.000175
31	11422	jim.braniff@enron.com.....	0.000175
31	15042	danny.lee@enron.com.....	Danny Lee.....	0.000173
31	44231	sap.hotline@enron.com.....	0.000157
31	5807	general.announcement@enron.com.....	0.000148
31	55961	fermis@enron.com.....	".....	0.000148
31	32548	john.kiani@enron.com.....	0.000144
31	9360	andrew.zabriskie@enron.com.....	0.000136
31	77887	pourchot@enron.com.....	0.000136
31	37504	sidrac.flores@enron.com.....	Sidrac Flores.....	0.000131
31	11181	sarah.driscoll@enron.com.....	0.000130
31	12468	richard.orellana@enron.com.....	Richard Orellana....	0.000124
31	64983	jackson.vo@enron.com.....	0.000124

31 33734 sarah.zarkowsky@enron.com..... Sarah Zarkowsky..... 0.000123

CATEGORY 32

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7636 COMPONENTS: 14
LARGEST COMPONENT SIZE: 7596 PERCENT OF TOTAL GRAPH: 99.48%
GROUP DEGREE: 0.11694 GRAPH DENSITY: 0.00079
GROUP CLOSENESS: 0.00089 GROUP BETWEENNESS: 0.20977
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
32	8722	home.owner@mailman.enron.com.....		0.001089
32	8629	undisclosed.recipients@mailman.enron.com.....		0.000943
32	22643	pmims@enron.com.....		0.000837
32	81740	plucci@ect.enron.com.....		0.000801
32	33798	us.home.owner@mailman.enron.com.....		0.000708
32	8779	valued.client@mailman.enron.com.....		0.000704
32	2347	h.lewis@enron.com.....	Andrew H. Lewis.....	0.000661
32	22403	dfarmer@ect.enron.com.....		0.000636
32	62384	mwoodson@enron.com.....		0.000530
32	70877	list.subscriber@mailman.enron.com.....		0.000530
32	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.000516
32	79414	lblair@enron.com.....		0.000512
32	29515	postmaster@mailboy.enron.com.....		0.000496
32	70836	schlenker@mailman.enron.com.....		0.000487
32	29641	aol.users@mailman.enron.com.....		0.000479
32	8759	valued.home.owner@mailman.enron.com.....		0.000474
32	8799	event@mailman.enron.com.....		0.000466
32	71892	extramoney@mailman.enron.com.....		0.000466
32	77844	jreitme@enron.com.....		0.000461
32	82621	valued.customer@mailman.enron.com.....		0.000457

CATEGORY 33

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7619 COMPONENTS: 18
LARGEST COMPONENT SIZE: 7568 PERCENT OF TOTAL GRAPH: 99.33%
GROUP DEGREE: 0.10614 GRAPH DENSITY: 0.00079
GROUP CLOSENESS: 0.00054 GROUP BETWEENNESS: 0.15977
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
33	37111	.stephens@enron.com.....	bridgeline.....	0.001161
33	37443	'fenner@enron.com.....	0.000901
33	3580	tiffany.smith@enron.com.....	Tiffany Smith.....	0.000875
33	3354	tom.ward@enron.com.....	TOM WARD.....	0.000854
33	4581	mday@enron.com.....	0.000836
33	52959	'thompson@enron.com.....	0.000798
33	62800	beth.cherry@enron.com.....	Beth Cherry.....	0.000772
33	31826	'sbigalow"@enron.com.....	"sbigalow".....	0.000748
33	24509	'proctor@enron.com.....	0.000744
33	68290	a.fishkin@enron.com.....	Charles A Fishkin...	0.000738
33	85652	vo.hoang@enron.com.....	Hoang Vo.....	0.000725
33	81781	'jernigan@enron.com.....	0.000723
33	38921	jennifer.ballas@enron.com.....	Jennifer Ballas....	0.000720
33	37222	k.c@enron.com.....	0.000693
33	20246	valerie.curtis@enron.com.....	0.000682
33	52960	'lipper@enron.com.....	0.000669
33	78906	'ketcherside@enron.com.....	0.000652
33	46483	'ward@enron.com.....	0.000632
33	87152	'rector@enron.com.....	0.000624
33	53983	'thomas@enron.com.....	0.000613

CATEGORY 34

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7887 COMPONENTS: 22
 LARGEST COMPONENT SIZE: 7794 PERCENT OF TOTAL GRAPH: 98.82%
 GROUP DEGREE: 0.06480 GRAPH DENSITY: 0.00089
 GROUP CLOSENESS: 0.00026 GROUP BETWEENNESS: 0.08978
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
34	7514	craig.rickard@enron.com.....	Craig Rickard.....	0.001114
34	18854	teresa.aguilera-peon@enron.com.....	Maria Teresa Aguiler	0.001076
34	7528	t..robinson@enron.com.....	Richard T. Robinson.	0.000975
34	14946	cecil.stapley@enron.com.....	0.000967
34	3079	r..conner@enron.com.....	Andrew R. Conner....	0.000897
34	18901	dirk.dimitry@enron.com.....	Dirk Dimitry.....	0.000877
34	26919	janet.bowers@enron.com.....	Janet Bowers.....	0.000875
34	14574	mike.barry@enron.com.....	Mike Barry.....	0.000746

34	36470	larry.swett@enron.com.....	0.000745
34	44124	lisa.valley@enron.com.....	0.000703
34	15268	eileen.peebles@enron.com.....	0.000683
34	17270	tim.johanson@enron.com.....	0.000682
34	18877	greg.bruch@enron.com..... Greg Bruch.....	0.000627
34	43984	loren.penkava@enron.com.....	0.000626
34	11195	morela.hernandez@enron.com.....	0.000607
34	18923	elizabeth.hutchinson@enron.com..... Elizabeth Hutchinson	0.000606
34	1364	laura.lantefield@enron.com..... Laura Lantefield....	0.000600
34	1617	jennifer.fraser@enron.com.....	0.000562
34	41985	allen.cohrs@enron.com.....	0.000542
34	43946	larry.pavlou@enron.com.....	0.000516

CATEGORY 35

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6630 COMPONENTS: 17
LARGEST COMPONENT SIZE: 6569 PERCENT OF TOTAL GRAPH: 99.08%
GROUP DEGREE: 0.14227 GRAPH DENSITY: 0.00091
GROUP CLOSENESS: 0.00051 GROUP BETWEENNESS: 0.23975
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
35	548	jeff.king@enron.com.....	Jeff King.....	0.001014
35	20483	amber.limas@enron.com.....		0.000963
35	22685	lmay2@enron.com.....		0.000958
35	16217	rose.botello@enron.com.....		0.000802
35	53196	davette.warren@enron.com.....		0.000703
35	6837	diane.salcido@enron.com.....	Diane Salcido.....	0.000687
35	35590	gurley@enron.com.....		0.000526
35	36452	juantongia.calvin@enron.com.....		0.000481
35	18727	regina.blackshear@enron.com.....	Regina Blackshear...	0.000479
35	19951	amber.ebow@enron.com.....		0.000428
35	53747	l.miller@enron.com.....	Chris L Miller.....	0.000417
35	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.000411
35	19277	benjamin.freeman@enron.com.....		0.000406
35	81470	.carroll@enron.com.....	e-mail.....	0.000393
35	9260	don.stevens@enron.com.....		0.000384
35	16087	mudd'.lisa@enron.com.....	Lisa Mudd.....	0.000374
35	86643	kevin_hyatt@enron.com.....		0.000355
35	6992	judy.hernandez@enron.com.....	Judy Hernandez.....	0.000354
35	71827	ddavis@enron.com.....		0.000347

35 19800 marilyn.rivera@enron.com..... 0.000345

CATEGORY 36

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7447 COMPONENTS: 16
LARGEST COMPONENT SIZE: 7407 PERCENT OF TOTAL GRAPH: 99.46%
GROUP DEGREE: 0.06138 GRAPH DENSITY: 0.00094
GROUP CLOSENESS: 0.00083 GROUP BETWEENNESS: 0.07973
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
36	4110	klay@enron.com.....		0.001302
36	14521	qwest.net@mailman.enron.com.....	"smoray!qwest.net"..	0.000498
36	14112	lfan@mailman.enron.com.....	"lfan".....	0.000305
36	18017	ruth.mann@enron.com.....		0.000236
36	1911	resources.human@enron.com.....	Human Resources....	0.000214
36	57681	munder.mail.list@mailman.enron.com.....		0.000143
36	12940	mike.underwood@enron.com.....	Mike Underwood.....	0.000140
36	13411	rebecca.longoria@enron.com.....	Rebecca Longoria....	0.000140
36	12993	cecil.stinemetz@enron.com.....	Cecil Stinemetz....	0.000139
36	419	steven.burnham@enron.com.....	Steven Burnham.....	0.000131
36	7634	s.smith@enron.com.....	Robert S. Smith....	0.000130
36	49937	rsanders@enron.com.....		0.000129
36	24910	celestine.hollan@enron.com.....		0.000124
36	9171	sally.hsieh@enron.com.....		0.000117
36	21873	jeff.borg@enron.com.....	Jeff Borg.....	0.000117
36	12952	johnnie.nelson@enron.com.....	Johnnie Nelson.....	0.000115
36	18587	patricia.weatherspoon@enron.com.....	Patricia Weatherspool	0.000114
36	32139	jnorden@enron.com.....		0.000113
36	7008	jona.kimbrough@enron.com.....	Jona Kimbrough.....	0.000109
36	21872	dave.ellis@enron.com.....	Dave Ellis.....	0.000108

CATEGORY 37

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7197 COMPONENTS: 16
LARGEST COMPONENT SIZE: 7145 PERCENT OF TOTAL GRAPH: 99.28%
GROUP DEGREE: 0.07675 GRAPH DENSITY: 0.00083
GROUP CLOSENESS: 0.00052 GROUP BETWEENNESS: 0.10975
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
37	22503	carlos.j.rodriguez@enron.com	carlos.j.rodriguez	0.001368
37	22763	briley@enron.com		0.001246
37	18724	kevin.alvarado@enron.com	Kevin Alvarado	0.000992
37	11081	rebecca.griffin@enron.com	Rebecca Griffin	0.000922
37	6067	alvin.thompson@enron.com		0.000914
37	6070	scott.loving@enron.com		0.000890
37	11116	megan.parker@enron.com	Megan Parker	0.000877
37	20241	susan.hadix@enron.com		0.000772
37	22446	juliann.kemp@enron.com		0.000739
37	35828	greg.mann@enron.com	Greg Mann	0.000721
37	3006	robert.ramirez@enron.com	Robert Ramirez	0.000695
37	22443	benjamin.schoene@enron.com	Benjamin Schoene	0.000695
37	14684	robert.cotten@enron.com	Robert Cotten	0.000693
37	8968	h.fletcher@enron.com		0.000691
37	11087	katherine.herrera@enron.com	Katherine Herrera	0.000653
37	9211	aimee.lannou@enron.com		0.000632
37	3007	l.dinari@enron.com	Sabra L. Dinari	0.000628
37	14685	kate.fraser@enron.com	Kate Fraser	0.000621
37	20244	laurie.ellis@enron.com		0.000619
37	6068	joe.casas@enron.com		0.000596

CATEGORY 38

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7983 COMPONENTS: 13
 LARGEST COMPONENT SIZE: 7942 PERCENT OF TOTAL GRAPH: 99.28%
 GROUP DEGREE: 0.10007 GRAPH DENSITY: 0.00088
 GROUP CLOSENESS: 0.00076 GROUP BETWEENNESS: 0.15980
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
38	79501	gblair@ei.enron.com		0.000872
38	26360	dgiron@enron.com		0.000727
38	56591	tracy_geaccone@enron.com		0.000724
38	35021	plove@ect.enron.com		0.000699
38	68835	jarnold@ect.enron.com		0.000673
38	69656	lcampbel@ect.enron.com		0.000664
38	22681	mmotley@enron.com		0.000659
38	85541	tgeacco@enron.com		0.000639

```

38 82262 pallen@ect.enron.com..... 0.000627
38 4493 lcampbel@enron.com..... 0.000624
38 37545 pplatte@enron.com..... 0.000622
38 4479 kruscit@enron.com..... 0.000610
38 22678 jforney@enron.com..... 0.000610
38 37735 kpresto@ect.enron.com..... 0.000602
38 3779 chuck.randall@enron.com..... 0.000553
38 36000 emclaug@enron.com..... 0.000550
38 68870 jarnold@ei.enron.com..... 0.000547
38 60337 mlenhart@ect.enron.com..... Matt Lenhart..... 0.000522
38 26485 dgiron@ect.enron.com..... 0.000519
38 30015 kamins@ect.enron.com..... kamins@ect.enron.com 0.000516

```

```

*****
CATEGORY 39

```

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 8313          COMPONENTS: 19
LARGEST COMPONENT SIZE: 8253 PERCENT OF TOTAL GRAPH: 99.28%
GROUP DEGREE: 0.07858   GRAPH DENSITY: 0.00084
GROUP CLOSENESS: 0.00042 GROUP BETWEENNESS: 0.13980
AVERAGE p(z|u): 0.02   STDEV p(z|u): 0.03

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
39	7571	maria.sandoval@enron.com.....	Maria Sandoval.....	0.000804
39	44164	jon.trevelise@enron.com.....		0.000754
39	44102	christi.culwell@enron.com.....		0.000717
39	19465	catherine.dumont@enron.com.....		0.000714
39	5076	cindy.shaffer@enron.com.....		0.000677
39	778	ashley.worthing@enron.com.....	Ashley Worthing....	0.000676
39	5023	anne.jolibois@enron.com.....		0.000615
39	71921	warren.perry@enron.com.....	Warren Perry.....	0.000609
39	44142	ron.beidelman@enron.com.....		0.000602
39	7570	rebecca.sanchez@enron.com.....	Rebecca Sanchez....	0.000598
39	16179	pearson.ken@enron.com.....	KEN PEARSON.....	0.000589
39	18727	regina.blackshear@enron.com.....	Regina Blackshear...	0.000582
39	44275	paul.pfeffer@enron.com.....		0.000576
39	9021	richard.babin@enron.com.....		0.000573
39	6948	gary.stadler@enron.com.....	Gary Stadler.....	0.000568
39	69041	joanie.h.ngo@enron.com.....		0.000547
39	2349	craig.taylor@enron.com.....	Craig Taylor.....	0.000543
39	37179	fenner'. 'molly@enron.com.....	Molly Fenner.....	0.000540
39	16840	'trio@enron.com.....		0.000527

39 909 .jennifer@enron.com..... e-mail..... 0.000526

CATEGORY 40

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7067 COMPONENTS: 17
LARGEST COMPONENT SIZE: 7028 PERCENT OF TOTAL GRAPH: 99.45%
GROUP DEGREE: 0.11454 GRAPH DENSITY: 0.00071
GROUP CLOSENESS: 0.00088 GROUP BETWEENNESS: 0.18976
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
40	41733	dperlin@enron.com.....	"dperlin@enron.com".	0.001173
40	22614	pkeavey@enron.com.....		0.001028
40	60202	peter.f.keavey@enron.com.....	peter.f.keavey.....	0.000773
40	22317	heidi.dubose@enron.com.....		0.000763
40	2599	matt.smith@enron.com.....	Matt Smith.....	0.000637
40	65	anna.mehrer@enron.com.....		0.000633
40	14706	peter.keavey@enron.com.....	Peter Keavey.....	0.000622
40	81838	tstaab@enron.com.....	Theresa Staab.....	0.000589
40	83912	mark.e.taylor@enron.com.....	".....	0.000513
40	1530	sandy.morris@enron.com.....	Sandy Morris.....	0.000490
40	686	d.hogan@enron.com.....	Irena D. Hogan.....	0.000472
40	18884	nyree.chanaba@enron.com.....	Nyree Chanaba.....	0.000436
40	7112	zionette.vincent@enron.com.....	Zionette Vincent....	0.000431
40	56219	jane.m.tholt@enron.com.....	Jane Tholt.....	0.000420
40	682	claudia.guerra@enron.com.....	Claudia Guerra.....	0.000387
40	7482	tracy.ramsey@enron.com.....	Tracy Ramsey.....	0.000387
40	15105	trey.cash@enron.com.....	Trey Cash.....	0.000336
40	72458	jtholt@ect.enron.com.....		0.000273
40	8644	holly.keiser@enron.com.....		0.000259
40	78941	kim.watson@enron.com.....		0.000240

CATEGORY 41

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6821 COMPONENTS: 15
LARGEST COMPONENT SIZE: 6772 PERCENT OF TOTAL GRAPH: 99.28%
GROUP DEGREE: 0.11865 GRAPH DENSITY: 0.00088
GROUP CLOSENESS: 0.00064 GROUP BETWEENNESS: 0.19975
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
41	37217	c_r_zander@enron.com		0.001341
41	37146	fenner.chet@enron.com	Chet Fenner	0.001258
41	37115	chet.fenner@enron.com		0.001226
41	37220	erwollam@enron.com		0.001176
41	37151	knipe'. 'chad@enron.com	chad knipe	0.001126
41	37218	feder'. 't@enron.com		0.001108
41	37114	wollam'. 'erik@enron.com		0.001067
41	37159	mccomb.keith@enron.com	Keith McComb	0.001066
41	37223	chet_fenner@enron.com		0.001062
41	37122	wollam.erik@enron.com		0.001054
41	29102	chambers.john@enron.com	John Chambers	0.001044
41	37216	feder.t@enron.com		0.001040
41	37161	mccomb.chris@enron.com	Chris McComb	0.000961
41	37157	constantine'. 'brian@enron.com		0.000940
41	37152	corrier.brad@enron.com	Brad Corrier	0.000709
41	37147	knipe.chad@enron.com	chad knipe	0.000678
41	37145	constantine.brian@enron.com	Brian Constantine	0.000664
41	37189	mccomb'. 'chris@enron.com	Chris McComb	0.000576
41	36662	sneal@ect.enron.com		0.000380
41	37190	mccomb'. 'keith@enron.com	Keith McComb	0.000355

CATEGORY 42

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7713 COMPONENTS: 22
 LARGEST COMPONENT SIZE: 7651 PERCENT OF TOTAL GRAPH: 99.20%
 GROUP DEGREE: 0.12212 GRAPH DENSITY: 0.00091
 GROUP CLOSENESS: 0.00046 GROUP BETWEENNESS: 0.18979
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
42	2374	kristin.quinn@enron.com	Kristin Quinn	0.000567
42	12449	dave.lawlor@enron.com		0.000500
42	18877	greg.bruch@enron.com	Greg Bruch	0.000459
42	2129	.chad@enron.com	e-mail	0.000415
42	15719	liz.hillman@enron.com		0.000394
42	13283	michelle.foust@enron.com	L Michelle Foust	0.000378
42	13263	robert.gerry@enron.com	Robert Gerry	0.000351
42	20070	pavel.zadorozhny@enron.com		0.000340

42	34229	mpalmer@enron.com.....	mpalmer@enron.com...	0.000331
42	11222	david.marye@enron.com.....	0.000297
42	12474	john.kemp@enron.com.....	John Kemp.....	0.000272
42	82051	'appling@enron.com.....	0.000272
42	428	dan.dorland@enron.com.....	Dan Dorland.....	0.000271
42	30902	christopher.long@enron.com.....	0.000266
42	11017	the.globalist@enron.com.....	0.000257
42	49848	ca.team@enron.com.....	0.000257
42	29903	ze.powergroup.inc.@mailman.enron.com....	0.000245
42	33746	louise@enron.com.....	0.000226
42	76534	gwoulfe@enron.com.....	0.000224
42	12475	laura.ewald@enron.com.....	Laura Ewald.....	0.000212

CATEGORY 43

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7434 COMPONENTS: 18
LARGEST COMPONENT SIZE: 7388 PERCENT OF TOTAL GRAPH: 99.38%
GROUP DEGREE: 0.12082 GRAPH DENSITY: 0.0004
GROUP CLOSENESS: 0.00072 GROUP BETWEENNESS: 0.21977
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
43	50081	dfulton@enron.com.....	0.001103
43	50873	eletke@enron.com.....	0.001050
43	41598	snovose@enron.com.....	0.000965
43	41583	rfrank@enron.com.....	0.000957
43	49892	hap_boyd@enron.com.....	"Hap Boyd ".....	0.000945
43	24207	.sue@enron.com.....	e-mail.....	0.000877
43	49932	tjohnso8@enron.com.....	0.000875
43	52415	johnson.tamara@enron.com.....	0.000820
43	34229	mpalmer@enron.com.....	mpalmer@enron.com...	0.000795
43	2276	becky.merola@enron.com.....	Becky Merola.....	0.000786
43	34210	shapiro.rick@enron.com.....	0.000766
43	8474	david.nutt@enron.com.....	David Nutt.....	0.000716
43	50605	james_trudeau@enron.com.....	"Jim Trudeau ".....	0.000715
43	21134	corey.wilkes@enron.com.....	Corey Wilkes.....	0.000704
43	49848	ca.team@enron.com.....	0.000701
43	49937	rsanders@enron.com.....	0.000693
43	49829	bhawkin@enron.com.....	0.000638
43	24810	rita.bahner@enron.com.....	0.000623
43	37928	dblack@enron.com.....	0.000613

43 48839 smara@ect.enron.com..... Sue" "Mara..... 0.000606

CATEGORY 44

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6800 COMPONENTS: 18
LARGEST COMPONENT SIZE: 6747 PERCENT OF TOTAL GRAPH: 99.22%
GROUP DEGREE: 0.08448 GRAPH DENSITY: 0.00088
GROUP CLOSENESS: 0.00056 GROUP BETWEENNESS: 0.13972
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
44	20010	cstclai@enron.com.....		0.000522
44	15530	.mom@enron.com.....	e-mail.....	0.000517
44	7507	crystal.reyna@enron.com.....	Crystal Reyna.....	0.000506
44	6277	teresa.mcomber@enron.com.....	Teresa McOmber.....	0.000451
44	10561	fahd.lodi@enron.com.....		0.000435
44	29310	claudia.santos@enron.com.....	Claudia Santos.....	0.000416
44	1040	'kuehn@enron.com.....		0.000414
44	43991	judy.kudym@enron.com.....		0.000404
44	77911	walt.serrano@enron.com.....	Walt Serrano.....	0.000404
44	510	israel.estrada@enron.com.....	Israel Estrada.....	0.000391
44	1219	.gregory@enron.com.....	e-mail.....	0.000379
44	62765	.zeke@enron.com.....	e-mail.....	0.000377
44	19950	michelle.laurant@enron.com.....		0.000372
44	7146	kristi.demaiolo@enron.com.....	Kristi Demaiolo.....	0.000369
44	43107	unspecified-recipients@enron.com.....	unspecified-recipien	0.000353
44	64938	dawson@enron.com.....		0.000349
44	15340	lenine.jeganathan@enron.com.....		0.000340
44	83620	ludwig@enron.com.....		0.000340
44	3021	scott.hendrickson@enron.com.....	Scott Hendrickson...	0.000338
44	51739	antoine.duvauchelle@enron.com.....		0.000336

CATEGORY 45

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8824 COMPONENTS: 16
LARGEST COMPONENT SIZE: 8770 PERCENT OF TOTAL GRAPH: 99.39%
GROUP DEGREE: 0.09991 GRAPH DENSITY: 0.00079
GROUP CLOSENESS: 0.00050 GROUP BETWEENNESS: 0.16982
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
45	15949	alexandra.saler@enron.com		0.000694
45	3643	alexandra.villarreal@enron.com	Alexandra Villarreal	0.000660
45	10757	alma.martinez@enron.com		0.000585
45	2060	pete.heintzelman@enron.com	Pete Heintzelman	0.000557
45	61637	charlotte.kraham@enron.com		0.000503
45	75867	'jewell@enron.com		0.000503
45	66817	thomas.'paul@enron.com		0.000468
45	7470	kay.quigley@enron.com	Kay Quigley	0.000446
45	22	crystal.hyde@enron.com	Crystal Hyde	0.000444
45	25314	veselack@enron.com		0.000444
45	26591	ora.cross@enron.com	Ora Cross	0.000439
45	29138	douglas.nichols@enron.com		0.000426
45	75299	'erb@enron.com		0.000422
45	82927	'peters@enron.com		0.000408
45	2609	alex.villarreal@enron.com	Alex Villarreal	0.000404
45	6857	denae.umbower@enron.com	Denae Umbower	0.000398
45	15263	patrice.mims@enron.com		0.000391
45	104	laura.wente@enron.com		0.000386
45	77998	'mack@enron.com		0.000376
45	25315	germany.jr@enron.com		0.000373

CATEGORY 46

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7360 COMPONENTS: 25
 LARGEST COMPONENT SIZE: 7298 PERCENT OF TOTAL GRAPH: 99.16%
 GROUP DEGREE: 0.10930 GRAPH DENSITY: 0.00068
 GROUP CLOSENESS: 0.00042 GROUP BETWEENNESS: 0.16976
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
46	15748	plove@enron.com		0.000665
46	70	mark.brand@enron.com	Mark Brand	0.000638
46	673	frank.ermis@enron.com	Frank Ermis	0.000552
46	734	dutch.quigley@enron.com	Dutch Quigley	0.000492
46	44193	l.foust@enron.com		0.000476
46	11221	aaron.martinsen@enron.com		0.000475
46	488	mike.carson@enron.com	Mike Carson	0.000422
46	1129	andy.pace@enron.com	Andy Pace	0.000414

46	38006	servello.anthony@enron.com.....	0.000410
46	1140	joe.stepenovitch@enron.com.....	Joe Stepenovitch....	0.000392
46	2629	bbutler2@enron.com.....	0.000365
46	1115	clint.dean@enron.com.....	Clint Dean.....	0.000326
46	3426	mswerzb@ect.enron.com.....	0.000300
46	82632	lester.terry@enron.com.....	0.000295
46	53621	duke.kyle@enron.com.....	Kyle Duke.....	0.000291
46	1952	valerie.ramsower@enron.com.....	Valerie Ramsower....	0.000280
46	37391	.judy@enron.com.....	e-mail.....	0.000275
46	761	m.tholt@enron.com.....	Jane M. Tholt.....	0.000268
46	60337	mlenhart@ect.enron.com.....	Matt Lenhart.....	0.000268
46	7096	mike.sheedy@enron.com.....	Mike Sheedy.....	0.000259

CATEGORY 47

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6248 COMPONENTS: 19
LARGEST COMPONENT SIZE: 6198 PERCENT OF TOTAL GRAPH: 99.20%
GROUP DEGREE: 0.08176 GRAPH DENSITY: 0.00096
GROUP CLOSENESS: 0.00062 GROUP BETWEENNESS: 0.10967
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
47	566	evelyn.metoyer@enron.com.....	Evelyn Metoyer.....	0.001523
47	1056	kerri.thompson@enron.com.....	Kerri Thompson.....	0.001469
47	582	stephanie.piwetz@enron.com.....	Stephanie Piwetz....	0.001131
47	175	lisa.gang@enron.com.....	0.001077
47	24113	lgang@enron.com.....	0.000870
47	138	kysa.alport@enron.com.....	0.000707
47	796	shift.dl-portland@enron.com.....	DL-Portland Real Tim	0.000694
47	8792	sitara@enron.com.....	Sitara.....	0.000674
47	490	sharen.cason@enron.com.....	Sharen Cason.....	0.000656
47	2009	alexander.mcelreath@enron.com.....	Alexander McElreath.	0.000600
47	92	holden.salisbury@enron.com.....	Holden Salisbury....	0.000531
47	1051	judy.dyer@enron.com.....	Judy Dyer.....	0.000433
47	17096	kimberly.allen@enron.com.....	0.000432
47	1082	billy.braddock@enron.com.....	Billy Braddock.....	0.000410
47	981	portland.shift@enron.com.....	Portland Shift.....	0.000393
47	20	geir.solberg@enron.com.....	Geir Solberg.....	0.000390
47	9032	jennifer.blay@enron.com.....	0.000380
47	972	shift.portland@enron.com.....	Portland Shift.....	0.000370
47	21	kate.symes@enron.com.....	Kate Symes.....	0.000359

B.3 PLSI-U with all Words (No Dictionary)

CATEGORY 0

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 168	COMPONENTS: 1
LARGEST COMPONENT SIZE: 168	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.75254	GRAPH DENSITY: 0.05988
GROUP CLOSENESS: 0.34176	GROUP BETWEENNESS: 0.27509
AVERAGE $p(z u)$: 0.56	STDEV $p(z u)$: 0.34

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
0	253	jeff.dasovich@enron.com.....	Jeff Dasovich.....	0.014615
0	3132	james.wright@enron.com.....	0.010392
0	9244	richard.sanders@enron.com.....	0.008802
0	801	susan.mara@enron.com.....	Susan Mara.....	0.008433
0	1746	scott.stoness@enron.com.....	0.008333
0	1475	dennis.benevides@enron.com.....	0.007993
0	8546	sandra.mccubbin@enron.com.....	Sandra McCubbin.....	0.007983
0	817	richard.shapiro@enron.com.....	Richard Shapiro.....	0.007954
0	1489	james.steffes@enron.com.....	0.007633
0	2222	harry.kingerski@enron.com.....	Harry Kingerski.....	0.007553
0	181	paul.kaufman@enron.com.....	0.007373
0	3140	marty.sunde@enron.com.....	0.007369
0	2318	vicki.sharp@enron.com.....	Vicki Sharp.....	0.007312
0	1016	neil.bresnan@enron.com.....	Neil Bresnan.....	0.007169
0	1180	karen.denne@enron.com.....	Karen Denne.....	0.007139
0	3152	wanda.curry@enron.com.....	0.006854
0	7213	mike.smith@enron.com.....	Mike Smith.....	0.006835
0	1456	dan.leff@enron.com.....	Dan Leff.....	0.006735
0	8431	skean@enron.com.....	Steve Kean.....	0.006728
0	802	gordon.savage@enron.com.....	Gordon Savage.....	0.006718

CATEGORY 1

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 19	COMPONENTS: 2
LARGEST COMPONENT SIZE: 17	PERCENT OF TOTAL GRAPH: 89.47%
GROUP DEGREE: 0.42810	GRAPH DENSITY: 0.11111

GROUP CLOSENESS: 0.07210 GROUP BETWEENNESS: 0.50944
 AVERAGE p(z|u): 0.37 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
1	28280	jbryson@enron.com.....		0.026801
1	256	pete.davis@enron.com.....	Pete Davis.....	0.008654
1	46814	jdasovic@enron.com.....	"Jeff Dasovich "....	0.008066
1	47085	jfawcet@enron.com.....		0.005827
1	14935	susan.scott@enron.com.....		0.005779
1	2941	sscott3@enron.com.....		0.004511
1	20	geir.solberg@enron.com.....	Geir Solberg.....	0.004187
1	14	mark.guzman@enron.com.....	Mark Guzman.....	0.004186
1	19	ryan.slinger@enron.com.....	Ryan Slinger.....	0.004186
1	12	craig.dean@enron.com.....	Craig Dean.....	0.004000
1	28809	pchoi2@enron.com.....		0.003113
1	8	bill.williams@enron.com.....	Bill Williams III...	0.003057
1	108	albert.meyers@enron.com.....		0.002382
1	152	john.anderson@enron.com.....	John Anderson.....	0.002317
1	219	michael.mier@enron.com.....	Michael Mier.....	0.002317
1	253	jeff.dasovich@enron.com.....	Jeff Dasovich.....	0.001929
1	15	leaf.harasin@enron.com.....	Leaf Harasin.....	0.001854
1	17	bert.meyers@enron.com.....	Bert Meyers.....	0.001850
1	79	eric.linder@enron.com.....	Eric Linder.....	0.001659
1	48157	bgaillar@enron.com.....	".....	0.001439

CATEGORY 2

EXPLICIT SOCIAL NETWORK STATISTICS

AVERAGE p(z|u): 0.28 STDEV p(z|u): 0.26

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
2	46814	jdasovic@enron.com.....	"Jeff Dasovich "....	0.011197
2	46859	smara@enron.com.....	".....	0.007691
2	30959	jhartso@enron.com.....		0.003643
2	85760	mhain@ect.enron.com.....	".....	0.003588
2	48638	susan_j_mara@enron.com.....		0.003288
2	46847	mpetroch@enron.com.....	"Mona Petrochko "...	0.002137
2	48611	dparque@ect.enron.com.....	"Dave Parquet".....	0.002054
2	23755	bob.gates@enron.com.....		0.001577
2	48573	rboyd@enron.com.....	"Hap Boyd".....	0.001487
2	49892	hap_boyd@enron.com.....	"Hap Boyd ".....	0.001315

2	166	lysa.akin@enron.com.....	0.001259
2	28868	mhain@enron.com.....	0.001066
2	49932	tjohnso8@enron.com.....	0.001005
2	46826	jalamo@enron.com..... "	0.000912
2	10551	hap.boyd@enron.com.....	0.000824
2	48738	smccubbi@enron.com.....	0.000725
2	52415	johnson.tamara@enron.com.....	0.000659
2	48090	jeff_dasovich@enron.com.....	0.000655
2	48839	smara@ect.enron.com..... Sue" "Mara.....	0.000610
2	48736	hkingers@enron.com.....	0.000573

CATEGORY 3

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 652	COMPONENTS: 1
LARGEST COMPONENT SIZE: 652	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.55654	GRAPH DENSITY: 0.00614
GROUP CLOSENESS: 0.24734	GROUP BETWEENNESS: 0.72785
AVERAGE p(z u): 0.43	STDEV p(z u): 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
3	3113	sara.shackleton@enron.com.....		0.072098
3	11	monika.causholli@enron.com.....	Monika Causholli....	0.047180
3	17807	clement.abrams@enron.com.....	Clement Abrams.....	0.017906
3	8379	clint.freeland@enron.com.....	Clint Freeland.....	0.014831
3	18882	e..carter@enron.com.....	Karen E. Carter.....	0.013859
3	18855	david.allan@enron.com.....	David Allan.....	0.013510
3	3079	r..conner@enron.com.....	Andrew R. Conner....	0.013198
3	7528	t..robinson@enron.com.....	Richard T. Robinson.	0.013069
3	11391	james.bryja@enron.com.....		0.012995
3	18901	dirk.dimitry@enron.com.....	Dirk Dimitry.....	0.012590
3	1589	bob.crane@enron.com.....		0.012588
3	18877	greg.bruch@enron.com.....	Greg Bruch.....	0.012500
3	7514	craig.rickard@enron.com.....	Craig Rickard.....	0.012178
3	7029	jeff.nogid@enron.com.....	Jeff Nogid.....	0.011857
3	8497	joel.ephross@enron.com.....	Joel Ephross.....	0.010795
3	41477	gareth.bahlmann@enron.com.....		0.009969
3	4637	jim.armogida@enron.com.....		0.009464
3	18926	ayasha.kanji@enron.com.....	Ayasha Kanji.....	0.009102
3	2365	mary.cook@enron.com.....	Mary Cook.....	0.008981
3	255	angela.davis@enron.com.....	Angela Davis.....	0.008858

 CATEGORY 4

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 353 COMPONENTS: 88
 LARGEST COMPONENT SIZE: 158 PERCENT OF TOTAL GRAPH: 44.76%
 GROUP DEGREE: 0.20608 GRAPH DENSITY: 0.01136
 GROUP CLOSENESS: 0.00041 GROUP BETWEENNESS: 0.09861
 AVERAGE $p(z|u)$: 0.84 STDEV $p(z|u)$: 0.32

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
4	25048	paul.y'barbo@enron.com.....		0.016515
4	1005	mark.fisher@enron.com.....		0.011929
4	4911	hollis.kimbrough@enron.com.....		0.007286
4	16327	dan.masters@enron.com.....	Dan Masters.....	0.007236
4	27797	wayne.perry@enron.com.....		0.006743
4	53483	tony.galt@enron.com.....		0.005676
4	23696	mark.walker@enron.com.....		0.005020
4	54051	jim.fernie@enron.com.....	Jim Fernie.....	0.004713
4	81838	tstaab@enron.com.....	Theresa Staab.....	0.003723
4	2317	greg.curran@enron.com.....	Greg Curran.....	0.003678
4	23690	jeff.duff@enron.com.....		0.003603
4	6061	mariella.mahan@enron.com.....		0.003376
4	60973	kruscit@ect.enron.com.....	Kevin.....	0.003130
4	71045	mcuilla@ect.enron.com.....	Martin Cuilla.....	0.002903
4	7611	rick.sierra@enron.com.....	Rick Sierra.....	0.002624
4	23703	kurt.anderson@enron.com.....		0.002590
4	63242	miguel.maltes@enron.com.....	Miguel Maltes.....	0.002573
4	63248	federico.haeussler@enron.com.....		0.002330
4	22614	pkeavey@enron.com.....		0.002308
4	23699	kevin.cousineau@enron.com.....		0.002291

 CATEGORY 5

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 519 COMPONENTS: 7
 LARGEST COMPONENT SIZE: 286 PERCENT OF TOTAL GRAPH: 55.11%
 GROUP DEGREE: 0.25873 GRAPH DENSITY: 0.00386
 GROUP CLOSENESS: 0.00054 GROUP BETWEENNESS: 0.15861
 AVERAGE $p(z|u)$: 0.74 STDEV $p(z|u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
5	78508	kholst@enron.com		0.009242
5	45145	jsteffe@enron.com		0.007759
5	462	ryan.watt@enron.com	Ryan Watt	0.005368
5	417	stephane.brodeur@enron.com	Stephane Brodeur	0.005330
5	55961	fermis@enron.com	"	0.005200
5	55981	sbrewer@enron.com	"	0.004987
5	22612	ssouth@enron.com		0.004691
5	55972	lprior@enron.com	"	0.004691
5	54135	bullets@enron.com		0.003601
5	43980	sharris1@enron.com		0.002856
5	78538	kholst@ect.enron.com		0.002797
5	3392	johnson@enron.com		0.001948
5	22786	kimberly.watson@enron.com	Kimberly Watson	0.001907
5	155	elizabeth.sager@enron.com	Elizabeth Sager	0.001861
5	77861	wood@enron.com		0.001642
5	54121	market.team@enron.com		0.001592
5	3497	stuart.rexrode@enron.com		0.001385
5	72534	tomlinson@enron.com		0.001373
5	11699	evans@enron.com		0.001338
5	35765	l.johnson@enron.com	David L. Johnson	0.001333

CATEGORY 6

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 327 COMPONENTS: 5
LARGEST COMPONENT SIZE: 239 PERCENT OF TOTAL GRAPH: 73.09%
GROUP DEGREE: 0.29613 GRAPH DENSITY: 0.00920
GROUP CLOSENESS: 0.00213 GROUP BETWEENNESS: 0.30785
AVERAGE p(z|u): 0.48 STDEV p(z|u): 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
6	1454	j.kean@enron.com	Steven J. Kean	0.134864
6	2390	brent.hendry@enron.com	Brent Hendry	0.055713
6	3113	sara.shackleton@enron.com		0.036853
6	1179	.palmer@enron.com	pr	0.036653
6	1180	karen.denne@enron.com	Karen Denne	0.019328
6	20382	lynn.aven@enron.com		0.018104
6	741	jenny.rub@enron.com	Jenny Rub	0.015713
6	3450	bill.donovan@enron.com		0.015673
6	8519	steve.hotte@enron.com	Steve Hotte	0.013934
6	20029	andrea.calo@enron.com		0.011996

6	1634	bruce.harris@enron.com.....	0.011846
6	2439	john.brindle@enron.com.....	0.011136
6	1463	maureen.mcvicker@enron.com.....	0.010873
6	711	bob.mcauliffe@enron.com..... Bob McAuliffe.....	0.008630
6	56819	andrea.bertone@enron.com.....	0.008554
6	17612	scott.abshire@enron.com..... Scott Abshire.....	0.008536
6	671	keith.dziadek@enron.com..... Keith Dziadek.....	0.008170
6	17563	barton.clark@enron.com.....	0.007417
6	8420	randy.petersen@enron.com..... Randy Petersen.....	0.007152
6	9039	theresa.brogan@enron.com.....	0.007129

CATEGORY 7

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 687 COMPONENTS: 1
LARGEST COMPONENT SIZE: 687 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.38592 GRAPH DENSITY: 0.01020
GROUP CLOSENESS: 0.16711 GROUP BETWEENNESS: 0.54800
AVERAGE p(z|u): 0.39 STDEV p(z|u): 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
7	1612	j.farmer@enron.com.....		0.023699
7	602	robert.superty@enron.com.....	Robert Superty.....	0.018165
7	757	patti.sullivan@enron.com.....	Patti Sullivan.....	0.015374
7	550	victor.lamadrid@enron.com.....	Victor Lamadrid.....	0.014556
7	549	lisa.kinsey@enron.com.....	Lisa Kinsey.....	0.013849
7	645	bryce.baxter@enron.com.....	Bryce Baxter.....	0.011966
7	691	tammy.jaquet@enron.com.....	Tammy Jaquet.....	0.010387
7	514	clarissa.garcia@enron.com.....	Clarissa Garcia.....	0.009664
7	344	m.smith@enron.com.....	Regan M. Smith.....	0.008862
7	8931	kevin.heal@enron.com.....	Kevin Heal.....	0.008748
7	11116	megan.parker@enron.com.....	Megan Parker.....	0.007795
7	1777	rita.wynne@enron.com.....		0.007718
7	593	l.schrab@enron.com.....	Mark L. Schrab.....	0.007635
7	327	matt.pena@enron.com.....	Matt Pena.....	0.007385
7	732	richard.pinion@enron.com.....	Richard Pinion.....	0.007278
7	1400	donna.greif@enron.com.....	Donna Greif.....	0.006221
7	632	sherry.anastas@enron.com.....	Sherry Anastas.....	0.006147
7	637	natalie.baker@enron.com.....	Natalie Baker.....	0.006055
7	333	ramesh.rao@enron.com.....	Ramesh Rao.....	0.006026
7	14674	s.olinger@enron.com.....	Kimberly S. Olinger.	0.005569

CATEGORY 8

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 309 COMPONENTS: 2
LARGEST COMPONENT SIZE: 307 PERCENT OF TOTAL GRAPH: 99.35%
GROUP DEGREE: 0.36347 GRAPH DENSITY: 0.01948
GROUP CLOSENESS: 0.05294 GROUP BETWEENNESS: 0.31571
AVERAGE $p(z|u)$: 0.37 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
8	288	tana.jones@enron.com.....	Tana Jones.....	0.129389
8	355	.taylor@enron.com.....	legal.....	0.064858
8	2365	mary.cook@enron.com.....	Mary Cook.....	0.039520
8	5897	mark.taylor@enron.com.....	0.033920
8	7158	mark.greenberg@enron.com.....	Mark Greenberg.....	0.032178
8	437	peter.keohane@enron.com.....	Peter Keohane.....	0.027741
8	1019	leslie.hansen@enron.com.....	Leslie Hansen.....	0.022414
8	280	marie.heard@enron.com.....	Marie Heard.....	0.019159
8	436	greg.johnston@enron.com.....	Greg Johnston.....	0.017884
8	1092	travis.mccullough@enron.com.....	Travis McCullough...	0.015881
8	284	t..hodge@enron.com.....	Jeffrey T. Hodge....	0.015827
8	5889	frank.sayre@enron.com.....	0.014296
8	1091	carol.st.@enron.com.....	Carol St. Clair.....	0.014112
8	3113	sara.shackleton@enron.com.....	0.013717
8	2390	brent.hendry@enron.com.....	Brent Hendry.....	0.013341
8	294	c.koehler@enron.com.....	Anne C. Koehler.....	0.012106
8	401	bob.shults@enron.com.....	Bob Shults.....	0.011187
8	8306	n..gray@enron.com.....	Barbara N. Gray....	0.010455
8	423	sharon.crawford@enron.com.....	Sharon Crawford....	0.009768
8	503	daniel.diamond@enron.com.....	Daniel Diamond.....	0.009520

CATEGORY 9

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2146 COMPONENTS: 2
LARGEST COMPONENT SIZE: 2137 PERCENT OF TOTAL GRAPH: 99.58%
GROUP DEGREE: 0.19058 GRAPH DENSITY: 0.00186
GROUP CLOSENESS: 0.00847 GROUP BETWEENNESS: 0.32903
AVERAGE $p(z|u)$: 0.29 STDEV $p(z|u)$: 0.34

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
9	2155	lara.leibman@enron.com	Lara Leibman	0.007156
9	2209	sue.nord@enron.com	Sue Nord	0.005958
9	5569	donald.lassere@enron.com		0.005814
9	253	jeff.dasovich@enron.com	Jeff Dasovich	0.005405
9	2185	michelle.hicks@enron.com	Michelle Hicks	0.004549
9	47920	mike.dahlke@enron.com		0.004313
9	2156	ginger.dernehl@enron.com	Ginger Dernehl	0.003892
9	6112	jane.wilson@enron.com		0.003818
9	2186	robbi.rossi@enron.com	Robbi Rossi	0.003701
9	2784	ron.mcnamara@enron.com		0.003684
9	23995	james.ginty@enron.com		0.003598
9	2184	cynthia.harkness@enron.com	Cynthia Harkness	0.003557
9	5018	sylvia.hu@enron.com		0.003521
9	2187	wayne.gardner@enron.com	Wayne Gardner	0.003429
9	9488	hardie.davis@enron.com		0.003417
9	28735	xi.xi@enron.com		0.003387
9	2237	geriann.warner@enron.com	Gerriann Warner	0.003285
9	2312	jan.haizmann@enron.com	Jan Haizmann	0.003271
9	5034	marcia.linton@enron.com		0.003265
9	1463	maureen.mcvicker@enron.com		0.003241

CATEGORY 10

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 689 COMPONENTS: 1
LARGEST COMPONENT SIZE: 689 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.36647 GRAPH DENSITY: 0.00727
GROUP CLOSENESS: 0.19886 GROUP BETWEENNESS: 0.42712
AVERAGE p(z|u): 0.53 STDEV p(z|u): 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
10	1637	rod.hayslett@enron.com		0.089934
10	18299	tracy.geaccone@enron.com	Tracy Geaccone	0.064229
10	4769	stanley.horton@enron.com		0.053066
10	7573	james.saunders@enron.com	James Saunders	0.030331
10	5480	danny.mccarty@enron.com		0.029677
10	15310	bob.chandler@enron.com		0.014505
10	1647	a..howard@enron.com		0.013025
10	2280	shelley.corman@enron.com	Shelley Corman	0.011683
10	18986	john.cobb@enron.com		0.011220
10	4140	cindy.stark@enron.com		0.010721

```

10 18987 jerry.peters@enron.com..... 0.009553
10 5490 julie.armstrong@enron.com..... 0.008680
10 17653 james.centilli@enron.com..... James Centilli..... 0.008609
10 6084 morris.brassfield@enron.com..... 0.008229
10 5489 kathy.campos@enron.com..... 0.008104
10 8308 steven.harris@enron.com..... Steven Harris..... 0.007407
10 7953 dave.neubauer@enron.com..... Dave Neubauer..... 0.007347
10 11386 dan.fancler@enron.com..... 0.007161
10 31667 steve.gilbert@enron.com..... 0.006977
10 8643 john.keiser@enron.com..... John Keiser..... 0.006889

```

CATEGORY 11

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 606           COMPONENTS: 3
LARGEST COMPONENT SIZE: 591 PERCENT OF TOTAL GRAPH: 97.52%
GROUP DEGREE: 0.36260   GRAPH DENSITY: 0.02149
GROUP CLOSENESS: 0.00466 GROUP BETWEENNESS: 0.20774
AVERAGE p(z|u): 0.39   STDEV p(z|u): 0.38

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
11	4638	james.derrick@enron.com.....		0.036013
11	2201	cindy.olson@enron.com.....	Cindy Olson.....	0.019847
11	1461	kay.chapman@enron.com.....		0.016723
11	3443	mark.koenig@enron.com.....		0.016652
11	1477	greg.whalley@enron.com.....		0.016576
11	1490	steven.kean@enron.com.....		0.015718
11	3538	mark.frevert@enron.com.....	Mark Frevert.....	0.015387
11	3444	jeffrey.mcmahon@enron.com.....		0.014574
11	3441	kenneth.lay@enron.com.....		0.012490
11	1452	david.delainey@enron.com.....	David Delainey.....	0.011898
11	222	david.oxley@enron.com.....	David Oxley.....	0.011356
11	1543	richard.causey@enron.com.....		0.011350
11	1544	rick.buy@enron.com.....		0.011146
11	3484	raymond.bowen@enron.com.....		0.010597
11	1463	maureen.mcvicker@enron.com.....		0.010338
11	1724	paula.rieker@enron.com.....		0.010201
11	6156	j.harris@enron.com.....		0.010005
11	4058	jeff.skilling@enron.com.....		0.009845
11	4130	kevin.hannon@enron.com.....		0.009549
11	2345	liz.taylor@enron.com.....	Liz Taylor.....	0.009328

CATEGORY 12

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 376 COMPONENTS: 9
 LARGEST COMPONENT SIZE: 342 PERCENT OF TOTAL GRAPH: 90.96%
 GROUP DEGREE: 0.31566 GRAPH DENSITY: 0.01333
 GROUP CLOSENESS: 0.00322 GROUP BETWEENNESS: 0.54592
 AVERAGE $p(z|u)$: 0.45 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
12	5897	mark.taylor@enron.com.....		0.202823
12	403	david.forster@enron.com.....		0.045717
12	21042	justin.boyd@enron.com.....	Justin Boyd.....	0.036726
12	3100	alan.aronowitz@enron.com.....		0.025537
12	20030	david.minns@enron.com.....		0.024279
12	3114	paul.simons@enron.com.....		0.021258
12	22321	edmund.cooper@enron.com.....		0.019739
12	1698	susan.musch@enron.com.....		0.016082
12	404	rahil.jafry@enron.com.....		0.014950
12	5864	janine.juggins@enron.com.....		0.013825
12	573	dale.neuner@enron.com.....	Dale Neuner.....	0.012468
12	31733	jeff.blumenthal@enron.com.....	Jeff Blumenthal.....	0.011894
12	20015	john.viverito@enron.com.....		0.011029
12	20022	jane.mcbride@enron.com.....		0.010295
12	18391	mark.evans@enron.com.....		0.010200
12	3046	jonathan.whitehead@enron.com.....	Jonathan Whitehead..	0.009499
12	3475	bryan.seyfried@enron.com.....		0.009468
12	18402	stephen.douglas@enron.com.....	Stephen H Douglas...	0.009409
12	6098	debbie.brackett@enron.com.....		0.009278
12	4786	dave.samuels@enron.com.....		0.009147

CATEGORY 13

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1123 COMPONENTS: 1
 LARGEST COMPONENT SIZE: 1123 PERCENT OF TOTAL GRAPH: 100%
 GROUP DEGREE: 0.45866 GRAPH DENSITY: 0.00535
 GROUP CLOSENESS: 0.21732 GROUP BETWEENNESS: 0.60890
 AVERAGE $p(z|u)$: 0.30 STDEV $p(z|u)$: 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
13	3041	george.mcclellan@enron.com	George McClellan	0.016337
13	403	david.forster@enron.com		0.010709
13	1718	daniel.reck@enron.com		0.009831
13	4134	jeffrey.shankman@enron.com		0.009321
13	401	bob.shults@enron.com	Bob Shults	0.009060
13	3032	sheri.thomas@enron.com	Sheri Thomas	0.009008
13	2359	andy.zipper@enron.com	Andy Zipper	0.008698
13	1200	savita.puthigai@enron.com	Savita Puthigai	0.008645
13	1462	kimberly.hillis@enron.com		0.008540
13	1687	kevin.mcgowan@enron.com		0.007943
13	227	sally.beck@enron.com	Sally Beck	0.007812
13	293	louise.kitchen@enron.com	Louise Kitchen	0.007649
13	595	kal.shah@enron.com	Kal Shah	0.007441
13	1752	mark.tawney@enron.com		0.007295
13	1695	torrey.moorer@enron.com		0.007059
13	3538	mark.frevert@enron.com	Mark Frevert	0.006945
13	407	jennifer.denny@enron.com		0.006601
13	3476	stuart.staley@enron.com		0.006433
13	577	leonardo.pacheco@enron.com	Leonardo Pacheco	0.006428
13	4785	john.nowlan@enron.com		0.005572

CATEGORY 14

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 949 COMPONENTS: 2
LARGEST COMPONENT SIZE: 947 PERCENT OF TOTAL GRAPH: 99.79%
GROUP DEGREE: 0.21779 GRAPH DENSITY: 0.01160
GROUP CLOSENESS: 0.04811 GROUP BETWEENNESS: 0.16829
AVERAGE p(z|u): 0.78 STDEV p(z|u): 0.34

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
14	12134	kevin.hyatt@enron.com	Kevin Hyatt	0.027850
14	8308	steven.harris@enron.com	Steven Harris	0.025017
14	8303	drew.fossum@enron.com	Drew Fossum	0.023359
14	22786	kimberly.watson@enron.com	Kimberly Watson	0.018504
14	14935	susan.scott@enron.com		0.016858
14	24943	michelle.lokay@enron.com		0.016388
14	9221	lorraine.lindberg@enron.com		0.016099
14	23672	jeffery.fawcett@enron.com		0.014024
14	24866	lindy.donoho@enron.com		0.013937
14	21117	tk.lohman@enron.com	TK Lohman	0.012491

14	2810	larry.campbell@enron.com.....	0.010799
14	24902	glen.hass@enron.com.....	0.009459
14	39747	mary.miller@enron.com.....	0.008694
14	43927	rich.jolly@enron.com.....	0.008478
14	8320	louis.soldano@enron.com..... Louis Soldano.....	0.007676
14	2280	shelley.corman@enron.com..... Shelley Corman.....	0.007379
14	20524	lorna.brennan@enron.com.....	0.007022
14	43926	maria.pavlou@enron.com.....	0.006968
14	7953	dave.neubauer@enron.com..... Dave Neubauer.....	0.006550
14	8476	john.shafer@enron.com.....	0.005879

CATEGORY 15

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 232	COMPONENTS: 14
LARGEST COMPONENT SIZE: 180	PERCENT OF TOTAL GRAPH: 77.59%
GROUP DEGREE: 0.21397	GRAPH DENSITY: 0.02165
GROUP CLOSENESS: 0.00348	GROUP BETWEENNESS: 0.23424
AVERAGE $p(z u)$: 0.50	STDEV $p(z u)$: 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
15	680	mike.grigsby@enron.com.....	Mike Grigsby.....	0.100385
15	10241	phillip.allen@enron.com.....	Phillip K. Allen....	0.056589
15	629	k.allen@enron.com.....	Phillip K. Allen....	0.041173
15	761	m.tholt@enron.com.....	Jane M. Tholt.....	0.034956
15	3659	scott.tholan@enron.com.....	Scott Tholan.....	0.031703
15	516	chris.gaskill@enron.com.....	Chris Gaskill.....	0.030711
15	1721	claudio.ribeiro@enron.com.....	Claudio Ribeiro.....	0.026102
15	80	john.zufferli@enron.com.....	John Zufferli.....	0.023783
15	687	keith.holst@enron.com.....	Keith Holst.....	0.023231
15	56	tim.heizenrader@enron.com.....	Tim Heizenrader....	0.020200
15	2553	tori.kuykendall@enron.com.....	Tori Kuykendall....	0.018342
15	2356	kristin.walsh@enron.com.....	Kristin Walsh.....	0.018124
15	673	frank.ermis@enron.com.....	Frank Ermis.....	0.017959
15	737	jay.reitmeyer@enron.com.....	Jay Reitmeyer.....	0.015372
15	1673	james.lewis@enron.com.....	James Lewis.....	0.013522
15	445	rob.milnthorp@enron.com.....	Rob Milnthorp.....	0.011750
15	14875	jeffrey.snyder@enron.com.....	Jeffrey Snyder.....	0.010901
15	2575	joe.parks@enron.com.....	Joe Parks.....	0.009810
15	703	matthew.lenhart@enron.com.....	Matthew Lenhart....	0.009466
15	675	l.gay@enron.com.....	Randall L. Gay.....	0.009367

CATEGORY 16

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2426 COMPONENTS: 3
LARGEST COMPONENT SIZE: 2413 PERCENT OF TOTAL GRAPH: 99.46%
GROUP DEGREE: 0.36036 GRAPH DENSITY: 0.00289
GROUP CLOSENESS: 0.00543 GROUP BETWEENNESS: 0.29942
AVERAGE $p(z|u)$: 0.33 STDEV $p(z|u)$: 0.34

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
16	6909	erin.willis@enron.com.....	Erin Willis.....	0.002479
16	1832	jana.giovannini@enron.com.....	Jana Giovannini.....	0.002378
16	11182	marcus.edmonds@enron.com.....	0.002111
16	1874	jody.crook@enron.com.....	Jody Crook.....	0.001996
16	2032	rahul.seksaria@enron.com.....	Rahul Seksaria.....	0.001984
16	1130	joseph.piotrowski@enron.com.....	Joseph Piotrowski...	0.001974
16	11262	james.wininger@enron.com.....	0.001927
16	11172	christopher.chenoweth@enron.com.....	0.001856
16	6702	chad.landry@enron.com.....	Chad Landry.....	0.001826
16	1586	mark.courtney@enron.com.....	0.001810
16	7106	george.thomas@enron.com.....	George Thomas.....	0.001801
16	6657	anthony.sexton@enron.com.....	Anthony Sexton.....	0.001784
16	1758	carl.tricoli@enron.com.....	0.001778
16	8849	robin.rodrique@enron.com.....	Robin Rodrigue.....	0.001772
16	1878	bryan.hull@enron.com.....	Bryan Hull.....	0.001771
16	7614	michael.simmons@enron.com.....	Michael Simmons.....	0.001771
16	11	monika.causholli@enron.com.....	Monika Causholli...	0.001759
16	11227	ravi.mujumdar@enron.com.....	0.001739
16	7583	ethan.schultz@enron.com.....	Ethan Schultz.....	0.001714
16	11402	ron.bertasi@enron.com.....	0.001713

CATEGORY 17

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 336 COMPONENTS: 1
LARGEST COMPONENT SIZE: 336 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.50323 GRAPH DENSITY: 0.02090
GROUP CLOSENESS: 0.24403 GROUP BETWEENNESS: 0.36570
AVERAGE $p(z|u)$: 0.36 STDEV $p(z|u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
17	584	m.presto@enron.com	Kevin M. Presto	0.072435
17	601	j.sturm@enron.com	Fletcher J. Sturm	0.026405
17	607	d.thomas@enron.com	Paul D. Thomas	0.026170
17	499	dana.davis@enron.com	Mark Dana Davis	0.025981
17	519	doug.gilbert-smith@enron.com	Doug Gilbert-smith	0.025086
17	1642	rogers.herndon@enron.com		0.019672
17	1990	harry.arora@enron.com	Harry Arora	0.019444
17	650	jae.black@enron.com	Tamara Jae Black	0.016925
17	618	lloyd.will@enron.com	Lloyd Will	0.015534
17	226	don.baughman@enron.com	Don Baughman Jr	0.012883
17	477	robert.benson@enron.com	Robert Benson	0.012035
17	538	rika.imai@enron.com	Rika Imai	0.011523
17	1126	tom.may@enron.com	Tom May	0.010276
17	1103	d.baughman@enron.com	Edward D. Baughman	0.009412
17	1105	j.broderick@enron.com	Paul J. Broderick	0.009133
17	812	l.nicolay@enron.com		0.008751
17	314	jeffrey.miller@enron.com	Jeffrey Miller	0.008388
17	257	l.day@enron.com	Smith L. Day	0.008188
17	1111	f.campbell@enron.com	Larry F. Campbell	0.007795
17	1148	gautam.gupta@enron.com	Gautam Gupta	0.007759

CATEGORY 18

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2284 COMPONENTS: 5
LARGEST COMPONENT SIZE: 2252 PERCENT OF TOTAL GRAPH: 98.60%
GROUP DEGREE: 0.35760 GRAPH DENSITY: 0.00088
GROUP CLOSENESS: 0.00165 GROUP BETWEENNESS: 0.67926
AVERAGE p(z|u): 0.71 STDEV p(z|u): 0.41

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
18	911	.john@enron.com	e-mail	0.006060
18	3688	.david@enron.com	e-mail	0.005111
18	2410	.mike@enron.com	e-mail	0.004360
18	3719	.michael@enron.com	e-mail	0.003620
18	907	.jeff@enron.com	e-mail	0.003619
18	939	.scott@enron.com	e-mail	0.003591
18	2431	.tom@enron.com	e-mail	0.003547
18	883	.dan@enron.com	e-mail	0.003233
18	3715	.mark@enron.com	e-mail	0.003123
18	875	.bob@enron.com	e-mail	0.003113

```

18      889 .eric@enron.com..... e-mail..... 0.003055
18      874 .bill@enron.com..... e-mail..... 0.002984
18     3228 .jim@enron.com..... e-mail..... 0.002980
18      913 .kevin@enron.com..... e-mail..... 0.002876
18     3229 .joe@enron.com..... e-mail..... 0.002865
18      947 .steve@enron.com..... e-mail..... 0.002734
18     2130 .chris@enron.com..... e-mail..... 0.002645
18     4058 jeff.skilling@enron.com..... 0.002447
18      933 .robert@enron.com..... e-mail..... 0.002406
18      895 .gary@enron.com..... e-mail..... 0.002283

```

CATEGORY 19

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 525           COMPONENTS: 4
LARGEST COMPONENT SIZE: 505 PERCENT OF TOTAL GRAPH: 96.19%
GROUP DEGREE: 0.36501   GRAPH DENSITY: 0.02290
GROUP CLOSENESS: 0.00377 GROUP BETWEENNESS: 0.33710
AVERAGE p(z|u): 0.77   STDEV p(z|u): 0.37

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
19	43960	lynn.blair@enron.com.....		0.042597
19	2530	chris.germany@enron.com.....	Chris Germany.....	0.037474
19	23333	darrell.schoolcraft@enron.com.....		0.020053
19	22786	kimberly.watson@enron.com.....	Kimberly Watson.....	0.019329
19	2280	shelley.corman@enron.com.....	Shelley Corman.....	0.018253
19	2575	joe.parks@enron.com.....	Joe Parks.....	0.015446
19	6966	john.buchanan@enron.com.....	John Buchanan.....	0.014915
19	44099	terry.kowalke@enron.com.....		0.013305
19	9096	rick.dietz@enron.com.....		0.012458
19	24943	michelle.lokay@enron.com.....		0.010309
19	8528	mark.mcconnell@enron.com.....		0.008525
19	25049	raetta.zadow@enron.com.....		0.007601
19	3005	joann.collins@enron.com.....	Joann Collins.....	0.007235
19	53745	steve.january@enron.com.....	Steve January.....	0.006966
19	521	scott.goodell@enron.com.....	Scott Goodell.....	0.006878
19	22593	joan.veselack@enron.com.....		0.006711
19	23666	sheila.nacey@enron.com.....		0.006084
19	21117	tk.lohman@enron.com.....	TK Lohman.....	0.005919
19	466	robert.allwein@enron.com.....	Robert Allwein.....	0.005870
19	19794	ramona.betancourt@enron.com.....		0.005386

 CATEGORY 20

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 76 COMPONENTS: 8
 LARGEST COMPONENT SIZE: 38 PERCENT OF TOTAL GRAPH: 50.00%
 GROUP DEGREE: 0.17658 GRAPH DENSITY: 0.04000
 GROUP CLOSENESS: 0.00547 GROUP BETWEENNESS: 0.15440
 AVERAGE $p(z|u)$: 0.85 STDEV $p(z|u)$: 0.32

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
20	253	jeff.dasovich@enron.com.....	Jeff Dasovich.....	0.051477
20	34417	alewis@ect.enron.com.....	Andrew Lewis.....	0.032326
20	734	dutch.quigley@enron.com.....	Dutch Quigley.....	0.020997
20	15669	sscott5@enron.com.....	0.016981
20	46814	jdasovic@enron.com.....	"Jeff Dasovich "....	0.015975
20	2592	m.scott@enron.com.....	Susan M. Scott.....	0.012083
20	4493	lcampbel@enron.com.....	0.010704
20	41627	scorman@enron.com.....	0.010053
20	41601	rshapiro@enron.com.....	0.009135
20	48090	jeff_dasovich@enron.com.....	0.008397
20	77707	plucci@enron.com.....	Paul Lucci.....	0.007649
20	2347	h.lewis@enron.com.....	Andrew H. Lewis....	0.006508
20	14935	susan.scott@enron.com.....	0.005442
20	548	jeff.king@enron.com.....	Jeff King.....	0.005317
20	2453	undisclosed-recipients@enron.com.....	undisclosed-recipien	0.004478
20	41521	brapp@enron.com.....	0.004313
20	60089	shendri@ect.enron.com.....	scott hendrickson...	0.004136
20	2515	martin.cuilla@enron.com.....	Martin Cuilla.....	0.004060
20	2587	kevin.ruscitti@enron.com.....	Kevin Ruscitti.....	0.004009
20	69405	khyatt@enron.com.....	0.003721

 CATEGORY 21

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 380 COMPONENTS: 1
 LARGEST COMPONENT SIZE: 380 PERCENT OF TOTAL GRAPH: 100%
 GROUP DEGREE: 0.38471 GRAPH DENSITY: 0.01319
 GROUP CLOSENESS: 0.23608 GROUP BETWEENNESS: 0.45623
 AVERAGE $p(z|u)$: 0.32 STDEV $p(z|u)$: 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
21	3111	gerald.nemec@enron.com		0.131946
21	2530	chris.germany@enron.com	Chris Germany	0.047453
21	15273	dan.hyvl@enron.com	Dan J Hyvl	0.043104
21	6815	debra.perlingiere@enron.com	Debra Perlingiere	0.042656
21	1688	ed.mcmichael@enron.com		0.031111
21	2514	ruth.concannon@enron.com	Ruth Concannon	0.020229
21	2575	joe.parks@enron.com	Joe Parks	0.019008
21	20018	stuart.zisman@enron.com		0.018083
21	3540	maria.garza@enron.com	Maria Garza	0.015160
21	14717	eric.gillaspie@enron.com	Eric Gillaspie	0.014902
21	1399	eric.boyt@enron.com	Eric Boyt	0.013725
21	14697	barbara.gray@enron.com	Barbara Gray	0.012778
21	1230	phil.polsky@enron.com	Phil Polsky	0.012274
21	8756	margaret.dhont@enron.com	Margaret Dhont	0.010542
21	17250	steve.hooser@enron.com		0.010413
21	2497	robin.barbe@enron.com	Robin Barbe	0.010170
21	2357	brian.redmond@enron.com	Brian Redmond	0.009285
21	1100	russell.diamond@enron.com	Russell Diamond	0.008737
21	11389	garrick.hill@enron.com		0.008569
21	1093	kay.mann@enron.com	Kay Mann	0.008517

CATEGORY 22

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 651 COMPONENTS: 2
LARGEST COMPONENT SIZE: 649 PERCENT OF TOTAL GRAPH: 99.69%
GROUP DEGREE: 0.77675 GRAPH DENSITY: 0.00462
GROUP CLOSENESS: 0.07842 GROUP BETWEENNESS: 0.73914
AVERAGE p(z|u): 0.31 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
22	3644	kim.ward@enron.com	Kim Ward	0.051340
22	14787	jason.williams@enron.com	Jason Williams	0.027822
22	213	chris.foster@enron.com	Chris H Foster	0.022375
22	1100	russell.diamond@enron.com	Russell Diamond	0.014306
22	20487	recipients@enron.com		0.011722
22	654	craig.breslau@enron.com	Craig Breslau	0.009517
22	14665	veronica.espinoza@enron.com	Veronica Espinoza	0.009159
22	480	bob.bowen@enron.com	Bob Bowen	0.007059
22	2573	lucy.ortiz@enron.com	Lucy Ortiz	0.006014
22	1102	tom.moran@enron.com	Tom Moran	0.005816

22	4854	william.bradford@enron.com.....	0.005410
22	6098	debbie.brackett@enron.com.....	0.005187
22	14716	linda.ewing@enron.com..... Linda Ewing.....	0.005035
22	1668	fred.lagrasta@enron.com.....	0.004921
22	206	christian.yoder@enron.com.....	0.004746
22	3542	lisa.gillette@enron.com..... Lisa Gillette.....	0.004723
22	1117	david.fairley@enron.com..... David Fairley.....	0.004711
22	14668	kim.theriot@enron.com..... Kim Theriot.....	0.004640
22	269	genia.fitzgerald@enron.com..... Genia Fitzgerald....	0.004556
22	19867	james.shirley@enron.com.....	0.004538

CATEGORY 23

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 115	COMPONENTS: 1
LARGEST COMPONENT SIZE: 115	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.69523	GRAPH DENSITY: 0.05263
GROUP CLOSENESS: 0.33713	GROUP BETWEENNESS: 0.66430
AVERAGE $p(z u)$: 0.29	STDEV $p(z u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
23	21	kate.symes@enron.com.....	Kate Symes.....	0.250872
23	8	bill.williams@enron.com.....	Bill Williams III...	0.129422
23	47	cara.semperger@enron.com.....	Cara Semperger.....	0.062312
23	345	will.smith@enron.com.....	Will Smith.....	0.026454
23	566	evelyn.metoyer@enron.com.....	Evelyn Metoyer.....	0.025118
23	19	ryan.slinger@enron.com.....	Ryan Slinger.....	0.023141
23	156	david.poston@enron.com.....	David Poston.....	0.022346
23	20	geir.solberg@enron.com.....	Geir Solberg.....	0.020995
23	14	mark.guzman@enron.com.....	Mark Guzman.....	0.020132
23	17	bert.meyers@enron.com.....	Bert Meyers.....	0.019710
23	15	leaf.harasin@enron.com.....	Leaf Harasin.....	0.017548
23	230	corry.bentley@enron.com.....	Corry Bentley.....	0.017445
23	12	craig.dean@enron.com.....	Craig Dean.....	0.016386
23	79	eric.linder@enron.com.....	Eric Linder.....	0.016381
23	306	duong.luu@enron.com.....	Duong Luu.....	0.015535
23	18	v.porter@enron.com.....	David V. Porter.....	0.012784
23	1158	vishwanatha.venkataswami@enron.com.....	Vishwanatha Venkatas	0.012631
23	78	todd.bland@enron.com.....	Todd Bland.....	0.012504
23	1057	kimberly.hundl@enron.com.....	Kimberly Hundl.....	0.012452
23	582	stephanie.piwetz@enron.com.....	Stephanie Piwetz....	0.012158

CATEGORY 24

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4231 COMPONENTS: 6
LARGEST COMPONENT SIZE: 4220 PERCENT OF TOTAL GRAPH: 99.74%
GROUP DEGREE: 0.20961 GRAPH DENSITY: 0.00047
GROUP CLOSENESS: 0.00563 GROUP BETWEENNESS: 0.28964
AVERAGE $p(z|u)$: 0.43 STDEV $p(z|u)$: 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
24	6031	outlook.team@enron.com.....		0.012505
24	9321	susan.lopez@enron.com.....		0.002325
24	1493	april.hrach@enron.com.....		0.002086
24	41163	gwhalle@ect.enron.com.....		0.002050
24	228	michael.belmont@enron.com.....	Michael Belmont.....	0.001982
24	283	bob.hillier@enron.com.....	Bob Hillier.....	0.001928
24	2469	vicky.ha@enron.com.....	Vicky Ha.....	0.001902
24	2466	victor.browner@enron.com.....	Victor Browner.....	0.001881
24	18593	scott.williamson@enron.com.....	Scott Williamson....	0.001849
24	2019	r.harrington@enron.com.....		0.001796
24	19381	jim.fussell@enron.com.....		0.001770
24	24665	stephen.harrington@enron.com.....		0.001770
24	785	jarod.jenson@enron.com.....	Jarod Jenson.....	0.001750
24	787	kevin.montagne@enron.com.....	Kevin Montagne.....	0.001737
24	10552	dan.bruce@enron.com.....		0.001732
24	406	michael.guadarrama@enron.com.....		0.001730
24	1213	backbone.ens@enron.com.....		0.001727
24	114	gray.calvert@enron.com.....		0.001717
24	641	michael.barber@enron.com.....	Michael Barber.....	0.001714
24	6891	gail.kettenbrink@enron.com.....	Gail Kettenbrink....	0.001712

CATEGORY 25

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 398 COMPONENTS: 18
LARGEST COMPONENT SIZE: 266 PERCENT OF TOTAL GRAPH: 66.83%
GROUP DEGREE: 0.23041 GRAPH DENSITY: 0.00756
GROUP CLOSENESS: 0.00155 GROUP BETWEENNESS: 0.31710
AVERAGE $p(z|u)$: 0.71 STDEV $p(z|u)$: 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
25	242	michelle.cash@enron.com	Michelle Cash	0.073555
25	1145	benjamin.rogers@enron.com	Benjamin Rogers	0.057591
25	1559	john.arnold@enron.com		0.043092
25	4134	jeffrey.shankman@enron.com		0.028072
25	773	.ward@enron.com	houston	0.020803
25	499	dana.davis@enron.com	Mark Dana Davis	0.018514
25	621	jason.wolfe@enron.com	Jason Wolfe	0.017101
25	4654	jennifer.burns@enron.com		0.015920
25	1140	joe.stepenovitch@enron.com	Joe Stepenovitch	0.013753
25	1465	twanda.sweet@enron.com		0.013010
25	2202	a..shankman@enron.com	Jeffrey A. Shankman	0.013010
25	19919	lavorato@enron.com		0.012111
25	1691	don.miller@enron.com		0.012004
25	11250	e.taylor@enron.com		0.010706
25	1689	jennifer.medcalf@enron.com		0.008170
25	7568	monique.sanchez@enron.com	Monique Sanchez	0.007925
25	32271	jarnold@enron.com	jarnold	0.007245
25	11156	sharon.butcher@enron.com	Sharon Butcher	0.006716
25	9264	kriste.sullivan@enron.com		0.006530
25	1617	jennifer.fraser@enron.com		0.005907

CATEGORY 26

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 408 COMPONENTS: 11
LARGEST COMPONENT SIZE: 352 PERCENT OF TOTAL GRAPH: 86.27%
GROUP DEGREE: 0.21545 GRAPH DENSITY: 0.00491
GROUP CLOSENESS: 0.00230 GROUP BETWEENNESS: 0.37435
AVERAGE p(z|u): 0.66 STDEV p(z|u): 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
26	2781	enron.announcements@enron.com		0.069997
26	1078	40enron@enron.com	Tracey Ramsey - Glob	0.067741
26	4058	jeff.skilling@enron.com		0.063704
26	427	chris.dorland@enron.com	Chris Dorland	0.061735
26	412	no.address@enron.com		0.044696
26	2599	matt.smith@enron.com	Matt Smith	0.043858
26	1175	arsystem@mailman.enron.com	ARSystem	0.028197
26	2883	all.houston@enron.com		0.021788
26	86	all.worldwide@enron.com	All Enron Worldwide	0.019729
26	6007	houston.report@enron.com		0.014175

26	22081	dfarmer@enron.com.....	0.013624
26	84	all_ena_egm_eim@enron.com.....	All_ENA_EGM_EIM.....	0.011844
26	710	larry.may@enron.com.....	Larry May.....	0.010501
26	1757	colin.tonks@enron.com.....	0.009049
26	428	dan.dorland@enron.com.....	Dan Dorland.....	0.008364
26	262	david.dronet@enron.com.....	David Dronet.....	0.006543
26	4762	office.chairman@enron.com.....	0.006421
26	516	chris.gaskill@enron.com.....	Chris Gaskill.....	0.005673
26	2387	tara.piazze@enron.com.....	Tara Piazze.....	0.005357
26	6226	enron.announcement@enron.com.....	0.004674

CATEGORY 27

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 405	COMPONENTS: 2
LARGEST COMPONENT SIZE: 399	PERCENT OF TOTAL GRAPH: 98.52%
GROUP DEGREE: 0.66146	GRAPH DENSITY: 0.01980
GROUP CLOSENESS: 0.01750	GROUP BETWEENNESS: 0.47792
AVERAGE $p(z u)$: 0.33	STDEV $p(z u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
27	367	w.white@enron.com.....	Stacey W. White.....	0.064266
27	268	casey.evans@enron.com.....	Casey Evans.....	0.025626
27	62	john.postlethwaite@enron.com.....	John Postlethwaite..	0.020753
27	226	don.baughman@enron.com.....	Don Baughman Jr....	0.020010
27	1104	kayne.coulter@enron.com.....	Kayne Coulter.....	0.018961
27	1120	juan.hernandez@enron.com.....	Juan Hernandez.....	0.017268
27	588	reagan.rorschach@enron.com.....	Reagan Rorschach....	0.016150
27	618	lloyd.will@enron.com.....	Lloyd Will.....	0.015298
27	1126	tom.may@enron.com.....	Tom May.....	0.013930
27	1044	rhonda.denton@enron.com.....	Rhonda Denton.....	0.012500
27	1140	joe.stepenovitch@enron.com.....	Joe Stepenovitch....	0.012058
27	1117	david.fairley@enron.com.....	David Fairley.....	0.011785
27	1110	rudy.acevedo@enron.com.....	Rudy Acevedo.....	0.010945
27	292	john.kinser@enron.com.....	John Kinser.....	0.010925
27	1096	dean.laurent@enron.com.....	Dean Laurent.....	0.010373
27	2810	larry.campbell@enron.com.....	0.009907
27	267	joe.errigo@enron.com.....	Joe Errigo.....	0.009482
27	2740	chad.starnes@enron.com.....	0.009398
27	2813	miguel.garcia@enron.com.....	0.009299
27	314	jeffrey.miller@enron.com.....	Jeffrey Miller.....	0.008834

CATEGORY 28

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 385 COMPONENTS: 3
LARGEST COMPONENT SIZE: 381 PERCENT OF TOTAL GRAPH: 98.96%
GROUP DEGREE: 0.43526 GRAPH DENSITY: 0.01302
GROUP CLOSENESS: 0.03227 GROUP BETWEENNESS: 0.57669
AVERAGE $p(z|u)$: 0.34 STDEV $p(z|u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
28	765	barry.tycholiz@enron.com.....	Barry Tycholiz.....	0.078402
28	1769	mark.whitt@enron.com.....	0.048881
28	717	stephanie.miller@enron.com.....	Stephanie Miller....	0.036904
28	712	jonathan.mckay@enron.com.....	Jonathan Mckay.....	0.030210
28	709	a..martin@enron.com.....	Thomas A. Martin....	0.028498
28	619	.williams@enron.com.....	credit.....	0.028301
28	2591	jim.schwieger@enron.com.....	Jim Schwieger.....	0.025808
28	719	l..mims@enron.com.....	Patrice L. Mims.....	0.019376
28	773	.ward@enron.com.....	houston.....	0.016969
28	14660	theresa.staab@enron.com.....	0.014697
28	3605	t..lucci@enron.com.....	Paul T. Lucci.....	0.013274
28	774	charles.weldon@enron.com.....	V. Charles Weldon...	0.012376
28	2515	martin.cuilla@enron.com.....	Martin Cuilla.....	0.012031
28	1676	laura.luce@enron.com.....	0.010473
28	3548	tyrell.harrison@enron.com.....	Tyrell Harrison.....	0.010358
28	2362	gary.bryan@enron.com.....	Gary Bryan.....	0.010155
28	2520	tom.donohoe@enron.com.....	Tom Donohoe.....	0.010003
28	581	vladi.pimenov@enron.com.....	Vladi Pimenov.....	0.009959
28	747	s..shively@enron.com.....	Hunter S. Shively...	0.009723
28	3111	gerald.nemec@enron.com.....	0.009570

CATEGORY 29

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 223 COMPONENTS: 2
LARGEST COMPONENT SIZE: 213 PERCENT OF TOTAL GRAPH: 95.52%
GROUP DEGREE: 0.29045 GRAPH DENSITY: 0.02703
GROUP CLOSENESS: 0.01056 GROUP BETWEENNESS: 0.29059
AVERAGE $p(z|u)$: 0.29 STDEV $p(z|u)$: 0.34

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
29	724	scott.neal@enron.com	Scott Neal	0.084950
29	2530	chris.germany@enron.com	Chris Germany	0.053341
29	764	judy.townsend@enron.com	Judy Townsend	0.032323
29	521	scott.goodell@enron.com	Scott Goodell	0.029347
29	2537	john.hodge@enron.com	John Hodge	0.024274
29	739	andrea.ring@enron.com	Andrea Ring	0.023856
29	550	victor.lamadrid@enron.com	Victor Lamadrid	0.023635
29	602	robert.superty@enron.com	Robert Superty	0.023183
29	1048	m.forney@enron.com	John M. Forney	0.019497
29	2563	brad.mckay@enron.com	Brad Mckay	0.019347
29	2608	victoria.versen@enron.com	Victoria Versen	0.017748
29	2578	w.pereira@enron.com	Susan W. Pereira	0.016941
29	2548	f.keavey@enron.com	Peter F. Keavey	0.016556
29	652	f.brawner@enron.com	Sandra F. Brawner	0.015440
29	14705	dick.jenkins@enron.com	Dick Jenkins	0.014574
29	770	frank.vickers@enron.com	Frank Vickers	0.014021
29	514	clarissa.garcia@enron.com	Clarissa Garcia	0.013714
29	2350	chuck.ames@enron.com	Chuck Ames	0.013374
29	2349	craig.taylor@enron.com	Craig Taylor	0.013243
29	2604	colleen.sullivan@enron.com	Colleen Sullivan	0.012638

CATEGORY 30

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 482 COMPONENTS: 1
LARGEST COMPONENT SIZE: 482 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.43150 GRAPH DENSITY: 0.01040
GROUP CLOSENESS: 0.18911 GROUP BETWEENNESS: 0.41663
AVERAGE p(z|u): 0.48 STDEV p(z|u): 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
30	2381	richard.ring@enron.com	Richard Ring	0.031736
30	1416	christi.nicolay@enron.com	Christi Nicolay	0.021109
30	2823	kevin.presto@enron.com		0.016200
30	2206	stacey.bolton@enron.com	Stacey Bolton	0.014342
30	817	richard.shapiro@enron.com	Richard Shapiro	0.013577
30	1489	james.steffes@enron.com		0.012671
30	813	sarah.novosel@enron.com		0.012512
30	1642	rogers.herndon@enron.com		0.011073
30	63	elliott.mainzer@enron.com	Elliot Mainzer	0.010418
30	519	doug.gilbert-smith@enron.com	Doug Gilbert-smith	0.009125

30	618	lloyd.will@enron.com.....	Lloyd Will.....	0.009028
30	2229	janine.migden@enron.com.....	Janine Migden.....	0.008796
30	2382	mark.bernstein@enron.com.....	Mark Bernstein.....	0.008443
30	2259	thane.twiggs@enron.com.....	Thane Twiggs.....	0.007970
30	2809	edward.baughman@enron.com.....	0.007922
30	1553	jeff.ader@enron.com.....	0.007858
30	413	john.llodra@enron.com.....	John Llodra.....	0.007837
30	5568	joe.kishkill@enron.com.....	0.007809
30	2784	ron.mcnamara@enron.com.....	0.007789
30	1485	donna.fulton@enron.com.....	0.007367

CATEGORY 31

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 302	COMPONENTS: 1
LARGEST COMPONENT SIZE: 302	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.38487	GRAPH DENSITY: 0.05316
GROUP CLOSENESS: 0.19020	GROUP BETWEENNESS: 0.11591
AVERAGE $p(z u)$: 0.53	STDEV $p(z u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
31	37	tim.belden@enron.com.....	Tim Belden.....	0.042943
31	59	diana.scholtes@enron.com.....	Diana Scholtes.....	0.036263
31	54	sean.crandall@enron.com.....	Sean Crandall.....	0.030541
31	42	jeff.richter@enron.com.....	Jeff Richter.....	0.029644
31	55	robert.badeer@enron.com.....	Robert Badeer.....	0.027046
31	57	matt.motley@enron.com.....	Matt Motley.....	0.026607
31	60	mike.swerzbin@enron.com.....	Mike Swerzbin.....	0.024860
31	52	tom.alonso@enron.com.....	Tom Alonso.....	0.024538
31	53	mark.fischer@enron.com.....	Mark Fischer.....	0.024408
31	93	phillip.platter@enron.com.....	Phillip Platter.....	0.023640
31	38	chris.mallory@enron.com.....	Chris Mallory.....	0.021094
31	92	holden.salisbury@enron.com.....	Holden Salisbury....	0.019483
31	175	lisa.gang@enron.com.....	0.017121
31	8	bill.williams@enron.com.....	Bill Williams III...	0.016884
31	124	chris.stokley@enron.com.....	Chris Stokley.....	0.013516
31	36	alan.comnes@enron.com.....	Alan Comnes.....	0.012209
31	14	mark.guzman@enron.com.....	Mark Guzman.....	0.012191
31	110	heather.dunton@enron.com.....	0.011767
31	115	m.driscoll@enron.com.....	Michael M. Driscoll.	0.011648
31	41	h.foster@enron.com.....	Chris H. Foster.....	0.011597

CATEGORY 32

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 448 COMPONENTS: 2
 LARGEST COMPONENT SIZE: 443 PERCENT OF TOTAL GRAPH: 98.88%
 GROUP DEGREE: 0.24667 GRAPH DENSITY: 0.01566
 GROUP CLOSENESS: 0.02283 GROUP BETWEENNESS: 0.25615
 AVERAGE $p(z|u)$: 0.48 STDEV $p(z|u)$: 0.40

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
32	1490	steven.kean@enron.com.....		0.080563
32	817	richard.shapiro@enron.com.....	Richard Shapiro.....	0.045726
32	1463	maureen.mcvicker@enron.com.....		0.017581
32	1547	mark.palmer@enron.com.....		0.016680
32	2209	sue.nord@enron.com.....	Sue Nord.....	0.015046
32	5416	mark.schroeder@enron.com.....		0.014551
32	1779	lisa.yoho@enron.com.....		0.014144
32	5956	michael.terraso@enron.com.....		0.013448
32	28729	scott.bolton@enron.com.....		0.012918
32	818	linda.robertson@enron.com.....	Linda Robertson.....	0.012390
32	253	jeff.dasovich@enron.com.....	Jeff Dasovich.....	0.012386
32	2256	marchris.robinson@enron.com.....	Marchris Robinson...	0.012175
32	2157	elizabeth.linnell@enron.com.....	Elizabeth Linnell...	0.011677
32	2214	stephen.burns@enron.com.....	Stephen Burns.....	0.011380
32	1570	rob.bradley@enron.com.....		0.010403
32	2220	chris.long@enron.com.....	Chris Long.....	0.010320
32	2155	lara.leibman@enron.com.....	Lara Leibman.....	0.010191
32	28730	susan.landwehr@enron.com.....		0.010097
32	14954	mary.schoen@enron.com.....		0.010035
32	9314	jeffrey.keeler@enron.com.....		0.009731

CATEGORY 33

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 316 COMPONENTS: 1
 LARGEST COMPONENT SIZE: 316 PERCENT OF TOTAL GRAPH: 100%
 GROUP DEGREE: 0.43672 GRAPH DENSITY: 0.06667
 GROUP CLOSENESS: 0.21378 GROUP BETWEENNESS: 0.09600
 AVERAGE $p(z|u)$: 0.33 STDEV $p(z|u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
33	3100	alan.aronowitz@enron.com		0.034792
33	15296	jeffrey.hodge@enron.com		0.033737
33	18647	carol.clair@enron.com		0.024891
33	19980	harry.collins@enron.com		0.024811
33	288	tana.jones@enron.com	Tana Jones	0.020152
33	5897	mark.taylor@enron.com		0.019317
33	9094	stacy.dickson@enron.com		0.019237
33	3113	sara.shackleton@enron.com		0.018068
33	9244	richard.sanders@enron.com		0.016180
33	266	janette.elbertson@enron.com	Janette Elbertson	0.015597
33	1019	leslie.hansen@enron.com	Leslie Hansen	0.015299
33	19792	taffy.milligan@enron.com		0.015121
33	12004	suzanne.adams@enron.com	Suzanne Adams	0.014889
33	3103	robert.bruce@enron.com		0.013211
33	4135	mark.haedicke@enron.com		0.012553
33	20015	john.viverito@enron.com		0.011790
33	20033	kaye.ellis@enron.com		0.011554
33	17099	shari.stack@enron.com		0.011208
33	155	elizabeth.sager@enron.com	Elizabeth Sager	0.011204
33	19400	michael.robison@enron.com		0.011097

CATEGORY 34

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 256 COMPONENTS: 2
LARGEST COMPONENT SIZE: 254 PERCENT OF TOTAL GRAPH: 99.22%
GROUP DEGREE: 0.50996 GRAPH DENSITY: 0.01569
GROUP CLOSENESS: 0.06689 GROUP BETWEENNESS: 0.43588
AVERAGE p(z|u): 0.29 STDEV p(z|u): 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
34	293	louise.kitchen@enron.com	Louise Kitchen	0.198510
34	701	john.lavorato@enron.com	John Lavorato	0.126343
34	1452	david.delainey@enron.com	David Delainey	0.061198
34	37	tim.belden@enron.com	Tim Belden	0.036809
34	445	rob.milnthorp@enron.com	Rob Milnthorp	0.035902
34	1453	janet.dietrich@enron.com	Janet Dietrich	0.030647
34	34	f.calger@enron.com	Christopher F. Calge	0.026040
34	2823	kevin.presto@enron.com		0.024960
34	1477	greg.whalley@enron.com		0.020042
34	618	lloyd.will@enron.com	Lloyd Will	0.018848

34	222	david.oxley@enron.com	David Oxley	0.017276
34	168	christopher.calger@enron.com		0.016651
34	275	e.haedicke@enron.com	Mark E. Haedicke	0.016644
34	743	tammie.schoppe@enron.com	Tammie Schoppe	0.015659
34	782	don.black@enron.com	Don Black	0.014802
34	2357	brian.redmond@enron.com	Brian Redmond	0.014781
34	495	wes.colwell@enron.com	Wes Colwell	0.014399
34	60	mike.swerzbin@enron.com	Mike Swerzbin	0.010634
34	3483	joseph.deffner@enron.com		0.010007
34	3608	jean.mrha@enron.com		0.009001

CATEGORY 35

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 535	COMPONENTS: 1
LARGEST COMPONENT SIZE: 535	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.47224	GRAPH DENSITY: 0.01311
GROUP CLOSENESS: 0.27967	GROUP BETWEENNESS: 0.46781
AVERAGE $p(z u)$: 0.36	STDEV $p(z u)$: 0.36

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
35	227	sally.beck@enron.com	Sally Beck	0.184642
35	1771	shona.wilson@enron.com		0.027381
35	6399	beth.apollo@enron.com		0.023193
35	10227	brent.price@enron.com		0.021862
35	6526	mike.jordan@enron.com		0.018646
35	14781	patti.thompson@enron.com		0.016208
35	334	leslie.reeves@enron.com	Leslie Reeves	0.016130
35	6396	fernley.dyson@enron.com	Fernley Dyson	0.015534
35	3029	sheila.glover@enron.com	Sheila Glover	0.014537
35	3002	m.hall@enron.com	Bob M Hall	0.012954
35	4796	ted.murphy@enron.com		0.012286
35	1978	greg.piper@enron.com		0.010775
35	6398	chris.abel@enron.com		0.010679
35	3402	kristin.albrecht@enron.com		0.010008
35	19242	mary.solmonson@enron.com		0.009424
35	15922	bob.hall@enron.com		0.009192
35	3455	barry.pearce@enron.com		0.008548
35	8963	hector.mcloughlin@enron.com	Hector McLoughlin	0.008350
35	1292	cassandra.schultz@enron.com	Cassandra Schultz	0.008282
35	3032	sheri.thomas@enron.com	Sheri Thomas	0.008241

CATEGORY 36

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 367 COMPONENTS: 10
 LARGEST COMPONENT SIZE: 357 PERCENT OF TOTAL GRAPH: 97.28%
 GROUP DEGREE: 0.51432 GRAPH DENSITY: 0.02732
 GROUP CLOSENESS: 0.00888 GROUP BETWEENNESS: 0.33721
 AVERAGE $p(z|u)$: 0.71 STDEV $p(z|u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
36	3113	sara.shackleton@enron.com		0.123371
36	288	tana.jones@enron.com	Tana Jones	0.073892
36	280	marie.heard@enron.com	Marie Heard	0.023954
36	1102	tom.moran@enron.com	Tom Moran	0.018733
36	3101	susan.bailey@enron.com		0.018431
36	3098	stephanie.panus@enron.com	Stephanie Panus	0.017748
36	1101	tanya.rohauer@enron.com	Tanya Rohauer	0.016405
36	33	stephanie.sever@enron.com	Stephanie Sever	0.015137
36	551	lisa.lees@enron.com	Lisa Lees	0.013782
36	22643	pmims@enron.com		0.012239
36	18647	carol.clair@enron.com		0.011726
36	1142	karen.lambert@enron.com	Karen Lambert	0.011010
36	3404	samantha.boyd@enron.com		0.009015
36	1449	samuel.schott@enron.com	Samuel Schott	0.008093
36	20682	cheryl.nelson@enron.com	Cheryl Nelson	0.007243
36	1099	brant.reves@enron.com	Brant Reves	0.007189
36	8874	kelly.lombardi@enron.com	Kelly Lombardi	0.006891
36	1181	exchange.administrator@enron.com		0.006090
36	3029	sheila.glover@enron.com	Sheila Glover	0.006041
36	5583	frank.davis@enron.com		0.005975

CATEGORY 37

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 625 COMPONENTS: 2
 LARGEST COMPONENT SIZE: 623 PERCENT OF TOTAL GRAPH: 99.68%
 GROUP DEGREE: 0.42207 GRAPH DENSITY: 0.01923
 GROUP CLOSENESS: 0.06416 GROUP BETWEENNESS: 0.32801
 AVERAGE $p(z|u)$: 0.77 STDEV $p(z|u)$: 0.35

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
37	253	jeff.dasovich@enron.com	Jeff Dasovich	0.076816
37	817	richard.shapiro@enron.com	Richard Shapiro	0.049471
37	1489	james.steffes@enron.com		0.038013
37	801	susan.mara@enron.com	Susan Mara	0.035751
37	181	paul.kaufman@enron.com		0.029256
37	347	d.steffes@enron.com	James D. Steffes	0.029135
37	2222	harry.kingerski@enron.com	Harry Kingerski	0.021314
37	1490	steven.kean@enron.com		0.018507
37	36	alan.comnes@enron.com	Alan Comnes	0.018317
37	813	sarah.novosel@enron.com		0.017805
37	17095	mary.hain@enron.com		0.015869
37	818	linda.robertson@enron.com	Linda Robertson	0.015279
37	8546	sandra.mccubbin@enron.com	Sandra McCubbin	0.012470
37	1474	joe.hartsoe@enron.com		0.012146
37	1479	leslie.lawner@enron.com		0.012061
37	1180	karen.denne@enron.com	Karen Denne	0.011729
37	28654	mona.petrochko@enron.com		0.008652
37	812	l.nicolay@enron.com		0.008379
37	800	ray.alvarez@enron.com	Ray Alvarez	0.008375
37	2326	janel.guerrero@enron.com		0.008131

CATEGORY 38

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 182 COMPONENTS: 13
LARGEST COMPONENT SIZE: 141 PERCENT OF TOTAL GRAPH: 77.47%
GROUP DEGREE: 0.38956 GRAPH DENSITY: 0.01105
GROUP CLOSENESS: 0.00442 GROUP BETWEENNESS: 0.40365
AVERAGE p(z|u): 0.70 STDEV p(z|u): 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
38	4110	klay@enron.com		0.114435
38	3441	kenneth.lay@enron.com		0.088106
38	3536	rosalee.fleming@enron.com	Rosalee Fleming	0.051847
38	4664	sherri.sera@enron.com		0.041385
38	4058	jeff.skilling@enron.com		0.019147
38	4104	joannie.williamson@enron.com		0.016989
38	14	mark.guzman@enron.com	Mark Guzman	0.016475
38	14935	susan.scott@enron.com		0.009385
38	4116	katherine.brown@enron.com		0.006860
38	80	john.zufferli@enron.com		0.005168

```

38 11168 tobin.carlson@enron.com..... 0.004727
38   12 craig.dean@enron.com..... Craig Dean..... 0.004022
38 6948 gary.stadler@enron.com..... Gary Stadler..... 0.004008
38 2288 misha.siegel@enron.com..... Misha Siegel..... 0.003858
38 5485 tori.wells@enron.com..... 0.003184
38 4063 sherri.reinartz@enron.com..... 0.002397
38 24446 executive.office@enron.com..... Office of the Chief 0.002256
38 5026 sally.keepers@enron.com..... 0.002184
38 19276 nicholas.stephan@enron.com..... 0.002027
38 5645 wilson.kriegel@enron.com..... 0.001937

```

CATEGORY 39

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 293           COMPONENTS: 2
LARGEST COMPONENT SIZE: 291 PERCENT OF TOTAL GRAPH: 99.32%
GROUP DEGREE: 0.32581   GRAPH DENSITY: 0.01370
GROUP CLOSENESS: 0.06182 GROUP BETWEENNESS: 0.35442
AVERAGE p(z|u): 0.3    STDEV p(z|u): 0.36

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
39	1544	rick.buy@enron.com.....		0.079570
39	2287	mike.mcconnell@enron.com.....	Mike Mcconnell.....	0.071530
39	2359	andy.zipper@enron.com.....	Andy Zipper.....	0.047493
39	4134	jeffrey.shankman@enron.com.....		0.038991
39	1477	greg.whalley@enron.com.....		0.037847
39	4132	john.sherriff@enron.com.....		0.028263
39	1978	greg.piper@enron.com.....		0.027670
39	788	david.port@enron.com.....	David Port.....	0.025199
39	4796	ted.murphy@enron.com.....		0.024818
39	2202	a.shankman@enron.com.....	Jeffrey A. Shankman.	0.022218
39	365	jay.webb@enron.com.....	Jay Webb.....	0.017257
39	1625	david.gorte@enron.com.....		0.015062
39	2370	vladimir.gorny@enron.com.....	Vladimir Gorny.....	0.011995
39	4133	philippe.bibi@enron.com.....		0.010761
39	481	s.bradford@enron.com.....	William S. Bradford.	0.010596
39	2383	michael.brown@enron.com.....	Michael Brown.....	0.010569
39	5062	cathy.phillips@enron.com.....		0.010036
39	1200	savita.puthigai@enron.com.....	Savita Puthigai.....	0.009958
39	1206	brad.richter@enron.com.....		0.009650
39	293	louise.kitchen@enron.com.....	Louise Kitchen.....	0.009375

CATEGORY 40

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 31 COMPONENTS: 1
 LARGEST COMPONENT SIZE: 31 PERCENT OF TOTAL GRAPH: 100%
 GROUP DEGREE: 0.49540 GRAPH DENSITY: 0.10000
 GROUP CLOSENESS: 0.16908 GROUP BETWEENNESS: 0.56133
 AVERAGE p(z|u): 0.56 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
40	256	pete.davis@enron.com.....	Pete Davis.....	0.177489
40	14	mark.guzman@enron.com.....	Mark Guzman.....	0.088616
40	19	ryan.slinger@enron.com.....	Ryan Slinger.....	0.088490
40	20	geir.solberg@enron.com.....	Geir Solberg.....	0.088011
40	12	craig.dean@enron.com.....	Craig Dean.....	0.082880
40	8	bill.williams@enron.com.....	Bill Williams III...	0.056518
40	15	leaf.harasin@enron.com.....	Leaf Harasin.....	0.049016
40	17	bert.meyers@enron.com.....	Bert Meyers.....	0.048908
40	79	eric.linder@enron.com.....	Eric Linder.....	0.043338
40	108	albert.meyers@enron.com.....	0.039818
40	219	michael.mier@enron.com.....	Michael Mier.....	0.039728
40	152	john.anderson@enron.com.....	John Anderson.....	0.039724
40	11	monika.causholli@enron.com.....	Monika Causholli...	0.039187
40	24	bill.williams.iii@enron.com.....	bill.williams.iii...	0.030895
40	28279	dporter3@enron.com.....	0.030679
40	92	holden.salisbury@enron.com.....	Holden Salisbury....	0.015729
40	89	greg.wolfe@enron.com.....	Greg Wolfe.....	0.013466
40	16	steven.merris@enron.com.....	Steven Merris.....	0.008022
40	28280	jbryson@enron.com.....	0.004246
40	95	center.dl-portland@enron.com.....	DL-Portland World Tr	0.002170

CATEGORY 41

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 870 COMPONENTS: 15
 LARGEST COMPONENT SIZE: 803 PERCENT OF TOTAL GRAPH: 92.30%
 GROUP DEGREE: 0.18464 GRAPH DENSITY: 0.00230
 GROUP CLOSENESS: 0.00124 GROUP BETWEENNESS: 0.27743
 AVERAGE p(z|u): 0.35 STDEV p(z|u): 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
41	756	geoff.storey@enron.com.....	Geoff Storey.....	0.021761
41	747	s..shively@enron.com.....	Hunter S. Shively...	0.021107
41	735	ina.rangel@enron.com.....	Ina Rangel.....	0.021059
41	10240	hunter.shively@enron.com.....	0.020921
41	1488	perfmgmt@enron.com.....	"Performance Evaluat	0.020366
41	1559	john.arnold@enron.com.....	0.013471
41	1418	lexi.elliott@enron.com.....	Lexi Elliott.....	0.013286
41	19787	v.weldon@enron.com.....	0.012157
41	655	karen.buckley@enron.com.....	Karen Buckley.....	0.011389
41	8357	airam.artea@enron.com.....	0.011050
41	23910	mary.fischer@enron.com.....	Mary Fischer.....	0.010572
41	1419	billy.lemmons@enron.com.....	Billy Lemmons Jr....	0.010545
41	795	john.griffith@enron.com.....	John Griffith.....	0.010305
41	2780	kim.melodick@enron.com.....	0.009985
41	710	larry.may@enron.com.....	Larry May.....	0.009911
41	4661	charlene.jackson@enron.com.....	0.007626
41	750	jeanie.slone@enron.com.....	Jeanie Slone.....	0.007531
41	1543	richard.causey@enron.com.....	0.006660
41	4662	celeste.roberts@enron.com.....	0.006151
41	798	announcements.enron@enron.com.....	Enron General Announ	0.005602

CATEGORY 42

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 423 COMPONENTS: 7
LARGEST COMPONENT SIZE: 382 PERCENT OF TOTAL GRAPH: 90.31%
GROUP DEGREE: 0.26931 GRAPH DENSITY: 0.00711
GROUP CLOSENESS: 0.00268 GROUP BETWEENNESS: 0.35488
AVERAGE p(z|u): 0.62 STDEV p(z|u): 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
42	642	eric.bass@enron.com.....	Eric Bass.....	0.042792
42	703	matthew.lenhart@enron.com.....	Matthew Lenhart....	0.035787
42	707	mike.maggi@enron.com.....	Mike Maggi.....	0.023799
42	6992	judy.hernandez@enron.com.....	Judy Hernandez.....	0.018926
42	8773	michelle.nelson@enron.com.....	Michelle Nelson....	0.014069
42	1115	clint.dean@enron.com.....	Clint Dean.....	0.013477
42	795	john.griffith@enron.com.....	John Griffith.....	0.010839
42	1878	bryan.hull@enron.com.....	Bryan Hull.....	0.010542
42	706	m..love@enron.com.....	Phillip M. Love....	0.010374
42	678	c..giron@enron.com.....	Darron C. Giron....	0.010239

42	15272	phillip.love@enron.com.....	Phillip Love.....	0.009753
42	14765	darron.giron@enron.com.....	0.009049
42	8917	timothy.blanchard@enron.com.....	0.008525
42	453	cooper.richey@enron.com.....	Cooper Richey.....	0.008415
42	15321	shanna.husser@enron.com.....	Shanna Husser.....	0.008368
42	6702	chad.landry@enron.com.....	Chad Landry.....	0.006848
42	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.005651
42	19886	leslie.smith@enron.com.....	0.005561
42	742	amanda.rybarski@enron.com.....	Amanda Rybarski.....	0.005443
42	18727	regina.blackshear@enron.com.....	Regina Blackshear...	0.005079

CATEGORY 43

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1256 COMPONENTS: 3
LARGEST COMPONENT SIZE: 1246 PERCENT OF TOTAL GRAPH: 99.20%
GROUP DEGREE: 0.28213 GRAPH DENSITY: 0.00637
GROUP CLOSENESS: 0.00884 GROUP BETWEENNESS: 0.26829
AVERAGE p(z|u): 0.50 STDEV p(z|u): 0.39

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
43	15554	daren.farmer@enron.com.....	0.045088
43	698	kam.keiser@enron.com.....	Kam Keiser.....	0.021511
43	713	errol.mclaughlin@enron.com.....	Errol McLaughlin Jr.	0.017938
43	15921	pat.clynes@enron.com.....	0.012968
43	706	m..love@enron.com.....	Phillip M. Love.....	0.012744
43	558	melba.lozano@enron.com.....	Melba Lozano.....	0.010818
43	14765	darron.giron@enron.com.....	0.010146
43	644	david.baumbach@enron.com.....	David Baumbach.....	0.009285
43	565	kevin.meredith@enron.com.....	Kevin Meredith.....	0.009056
43	11079	melissa.graves@enron.com.....	Melissa Graves.....	0.008244
43	612	chris.walker@enron.com.....	Chris Walker.....	0.007684
43	15272	phillip.love@enron.com.....	Phillip Love.....	0.007458
43	1777	rita.wynne@enron.com.....	0.007307
43	1695	torrey.moorer@enron.com.....	0.006879
43	15318	robert.cass@enron.com.....	0.006732
43	718	bruce.mills@enron.com.....	Bruce Mills.....	0.006592
43	14684	robert.cotten@enron.com.....	Robert Cotten.....	0.006567
43	11108	julie.meyers@enron.com.....	Julie Meyers.....	0.006533
43	604	tara.sweitzer@enron.com.....	Tara Sweitzer.....	0.006436
43	679	c..gossett@enron.com.....	Jeffrey C. Gossett..	0.005868

CATEGORY 44

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 420 COMPONENTS: 3
LARGEST COMPONENT SIZE: 414 PERCENT OF TOTAL GRAPH: 98.57%
GROUP DEGREE: 0.49411 GRAPH DENSITY: 0.01909
GROUP CLOSENESS: 0.01826 GROUP BETWEENNESS: 0.43733
AVERAGE $p(z|u)$: 0.49 STDEV $p(z|u)$: 0.42

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
44	5335	vince.kaminski@enron.com.....		0.208832
44	1654	j.kaminski@enron.com.....		0.052962
44	2058	shirley.crenshaw@enron.com.....	Shirley Crenshaw....	0.034859
44	29453	vkamins@enron.com.....		0.028588
44	4807	stinson.gibner@enron.com.....		0.024029
44	226	don.baughman@enron.com.....	Don Baughman Jr....	0.019995
44	2106	vasant.shanbhogue@enron.com.....	Vasant Shanbhogue...	0.015230
44	2109	zimin.lu@enron.com.....	Zimin Lu.....	0.010503
44	4310	ebass@enron.com.....		0.010253
44	30143	vince.j.kaminski@enron.com.....	Vince J" "Kaminski..	0.009942
44	2102	tanya.tamarchenko@enron.com.....	Tanya Tamarchenko...	0.009657
44	34375	alewis@enron.com.....		0.008964
44	1666	pinnamaneni.krishnarao@enron.com.....		0.008216
44	60189	pkeavey@ect.enron.com.....		0.008159
44	488	mike.carson@enron.com.....	Mike Carson.....	0.007894
44	1388	christie.patrick@enron.com.....	Christie Patrick....	0.007204
44	6033	mike.roberts@enron.com.....		0.007103
44	20277	grant.masson@enron.com.....		0.006883
44	6417	kaminski@enron.com.....		0.004629
44	1748	dale.surbey@enron.com.....		0.004280

CATEGORY 45

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 363 COMPONENTS: 4
LARGEST COMPONENT SIZE: 351 PERCENT OF TOTAL GRAPH: 96.69%
GROUP DEGREE: 0.37839 GRAPH DENSITY: 0.01105
GROUP CLOSENESS: 0.00827 GROUP BETWEENNESS: 0.53580
AVERAGE $p(z|u)$: 0.75 STDEV $p(z|u)$: 0.37

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
45	1093	kay.mann@enron.com	Kay Mann	0.130320
45	9244	richard.sanders@enron.com		0.036335
45	1174	b.sanders@enron.com	Richard B. Sanders	0.020703
45	1651	ben.jacoby@enron.com		0.020693
45	29416	richard.b.sanders@enron.com		0.015770
45	154	sheila.tweed@enron.com	Sheila Tweed	0.014446
45	17542	roseann.engeldorf@enron.com		0.011599
45	4638	james.derrick@enron.com		0.011240
45	20027	kathleen.carnahan@enron.com		0.010589
45	17261	carlos.sole@enron.com		0.010262
45	4940	rob.walls@enron.com		0.008472
45	1568	lisa.bills@enron.com		0.008230
45	17208	chris.booth@enron.com		0.007672
45	6899	fred.mitro@enron.com	Fred Mitro	0.007671
45	17434	britt.davis@enron.com		0.007496
45	1460	c.williams@enron.com	Robert C. Williams	0.006835
45	14935	susan.scott@enron.com		0.006371
45	2279	andrew.edison@enron.com	Andrew Edison	0.006344
45	24438	john.schwartzenburg@enron.com		0.006127
45	2386	heather.kroll@enron.com	Heather Kroll	0.006031

CATEGORY 46

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 978 COMPONENTS: 1
LARGEST COMPONENT SIZE: 978 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.61492 GRAPH DENSITY: 0.00716
GROUP CLOSENESS: 0.266322 GROUP BETWEENNESS: 0.23861
AVERAGE p(z|u): 0.35 STDEV p(z|u): 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
46	3457	gary.hickerson@enron.com		0.025693
46	1657	jeff.kinneman@enron.com		0.019869
46	543	robert.johnston@enron.com	Robert Johnston	0.017357
46	3659	scott.tholan@enron.com		0.014595
46	279	frank.hayden@enron.com	Frank Hayden	0.013467
46	14653	stephen.stock@enron.com		0.012490
46	366	zhiyong.wei@enron.com	Zhiyong Wei	0.011913
46	293	louise.kitchen@enron.com	Louise Kitchen	0.010144
46	2235	beth.perlman@enron.com	Beth Perlman	0.009677
46	3462	eric.gonzales@enron.com		0.009453

46	756	geoff.storey@enron.com.....	Geoff Storey.....	0.008877
46	1737	per.sekse@enron.com.....	0.008811
46	1627	john.greene@enron.com.....	0.008799
46	1571	michael.bradley@enron.com.....	0.008786
46	3475	bryan.seyfried@enron.com.....	0.008119
46	2202	a..shankman@enron.com.....	Jeffrey A. Shankman.	0.008106
46	23304	michelle.cisneros@enron.com.....	0.007940
46	1999	jaime.gualy@enron.com.....	Jaime Gualy.....	0.007823
46	3535	markus.fiala@enron.com.....	Markus Fiala.....	0.007595
46	1711	paul.pizzolato@enron.com.....	0.007570

CATEGORY 47

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 237	COMPONENTS: 1
LARGEST COMPONENT SIZE: 237	PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.29692	GRAPH DENSITY: 0.02119
GROUP CLOSENESS: 0.16754	GROUP BETWEENNESS: 0.29250
AVERAGE $p(z u)$: 0.32	STDEV $p(z u)$: 0.38

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
47	155	elizabeth.sager@enron.com.....	Elizabeth Sager.....	0.139767
47	4135	mark.haedicke@enron.com.....	0.066912
47	206	christian.yoder@enron.com.....	0.038500
47	124	chris.stokley@enron.com.....	Chris Stokley.....	0.036201
47	14696	lisa.mellencamp@enron.com.....	Lisa Mellencamp.....	0.029949
47	481	s..bradford@enron.com.....	William S. Bradford.	0.026618
47	590	edward.sacks@enron.com.....	Edward Sacks.....	0.025208
47	144	tracy.ngo@enron.com.....	0.022714
47	4854	william.bradford@enron.com.....	0.022092
47	9244	richard.sanders@enron.com.....	0.020364
47	329	david.portz@enron.com.....	David Portz.....	0.019284
47	318	marcus.nettelton@enron.com.....	Marcus Nettelton....	0.019239
47	2318	vicki.sharp@enron.com.....	Vicki Sharp.....	0.019177
47	1786	michael.tribolet@enron.com.....	0.018787
47	347	d..steffes@enron.com.....	James D. Steffes....	0.017370
47	58	p..o'neil@enron.com.....	Murray P. Neil.....	0.016995
47	15296	jeffrey.hodge@enron.com.....	0.016055
47	18647	carol.clair@enron.com.....	0.013327
47	1090	harlan.murphy@enron.com.....	Harlan Murphy.....	0.010520
47	269	genia.fitzgerald@enron.com.....	Genia Fitzgerald....	0.009790

B.4 Author Topic with all Words (No Dictionary)

CATEGORY 0

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2635 COMPONENTS: 9
LARGEST COMPONENT SIZE: 2606 PERCENT OF TOTAL GRAPH: 98.90%
GROUP DEGREE: 0.17486 GRAPH DENSITY: 0.00114
GROUP CLOSENESS: 0.00173 GROUP BETWEENNESS: 0.28903
AVERAGE $p(z|u)$: 0.13 STDEV $p(z|u)$: 0.25

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
0	256	pete.davis@enron.com.....	Pete Davis.....	0.000224689115565
0	5335	vince.kaminski@enron.com.....	0.000224629534140
0	2222	harry.kingerski@enron.com.....	Harry Kingerski.....	0.000224151636700
0	3132	james.wright@enron.com.....	0.000224056885215
0	36	alan.comnes@enron.com.....	Alan Comnes.....	0.000223919554515
0	1651	ben.jacoby@enron.com.....	0.000223898827630
0	3100	alan.aronowitz@enron.com.....	0.000223896134608
0	8308	steven.harris@enron.com.....	Steven Harris.....	0.000223838281534
0	37	tim.belden@enron.com.....	Tim Belden.....	0.000223834623853
0	1688	ed.mcmichael@enron.com.....	0.000223694291968
0	813	sarah.novosel@enron.com.....	0.000223666544479
0	403	david.forster@enron.com.....	0.000223605891197
0	818	linda.robertson@enron.com.....	Linda Robertson.....	0.000223533527968
0	168	christopher.calger@enron.com.....	0.000223458217152
0	7158	mark.greenberg@enron.com.....	Mark Greenberg.....	0.000223410680637
0	1474	joe.hartsoe@enron.com.....	0.000223400873207
0	812	l.nicolay@enron.com.....	0.000223341104209
0	401	bob.shults@enron.com.....	Bob Shults.....	0.000223222289714
0	2810	larry.campbell@enron.com.....	0.000223194609193
0	154	sheila.tweed@enron.com.....	Sheila Tweed.....	0.000223119660127

CATEGORY 1

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2709 COMPONENTS: 15
LARGEST COMPONENT SIZE: 2654 PERCENT OF TOTAL GRAPH: 97.97%
GROUP DEGREE: 0.18232 GRAPH DENSITY: 0.00258
GROUP CLOSENESS: 0.00075 GROUP BETWEENNESS: 0.17932
AVERAGE $p(z|u)$: 0.23 STDEV $p(z|u)$: 0.04

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
1	19980	harry.collins@enron.com		0.001437257393948
1	18664	brian.lindsay@enron.com	Brian Lindsay	0.001352654135122
1	37910	amy.heffernan@enron.com		0.001314934606771
1	6265	jana.morse@enron.com	Jana Morse	0.001290984787760
1	2995	juana.fayett@enron.com	Juana Fayett	0.001269252153405
1	18180	sonya.clarke@enron.com		0.001255208014767
1	17783	paul.maley@enron.com	Paul Maley	0.001202681298278
1	1094	karen.o'day@enron.com	Karen day	0.001197873773149
1	18181	tim.davies@enron.com		0.001187950996886
1	81490	ngo@enron.com		0.001162785669110
1	29291	.cooper@enron.com	ebs	0.001161458455270
1	250	cynthia.clark@enron.com	Cynthia Clark	0.001157936702115
1	2186	robbi.rossi@enron.com	Robbi Rossi	0.001142328065856
1	8876	tandra.coleman@enron.com	Tandra Coleman	0.001137397289212
1	19629	albert.escamilla@enron.com		0.001124752596785
1	18031	lesli.campbell@enron.com	Lesli Campbell	0.001110176158394
1	2993	trang.le@enron.com	Trang Le	0.001090336692365
1	19627	julie.brewer@enron.com		0.001087762015664
1	1202	center.eol@enron.com		0.001084833158682
1	17577	alison.keogh@enron.com	Alison Keogh	0.001066101166139

CATEGORY 2

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2613 COMPONENTS: 4
 LARGEST COMPONENT SIZE: 2606 PERCENT OF TOTAL GRAPH: 99.73%
 GROUP DEGREE: 0.16524 GRAPH DENSITY: 0.00230
 GROUP CLOSENESS: 0.01478 GROUP BETWEENNESS: 0.17925
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
2	6815	debra.perlingiere@enron.com	Debra Perlingiere	0.001219539848980
2	56883	robin.decker@enron.com		0.001117162899362
2	56873	elizabeth.serralheiro@enron.com		0.000851194436798
2	29433	legal.4@enron.com		0.000796814254506
2	81683	james.canney@enron.com	James Canney	0.000726944434387
2	15231	joanne.rozycki@enron.com	Joanne Rozycki	0.000723907351742
2	20033	kaye.ellis@enron.com		0.000695117396012
2	7476	andrew.ralston@enron.com	Andrew Ralston	0.000685078306837

2	19802	sheri.cromwell@enron.com.....	0.000665121467956
2	56908	lina.jimenez@enron.com.....	0.000660426109974
2	12226	carolyn.george@enron.com.....	0.000624006796842
2	29409	ned.cradyc@enron.com.....	0.000595200797749
2	15265	paul.burgener@enron.com.....	0.000572896648689
2	15246	margo.terrell@enron.com.....	0.000537901138048
2	14777	kay.young@enron.com.....	0.000529420421996
2	42425	allison.mchenry@enron.com.....	0.000519970465448
2	41717	brenda.l.funk@enron.com..... Brenda L. "Funk....	0.000508707286008
2	17585	j.simmons@enron.com.....	0.000504685196113
2	19054	jorge.garcia@enron.com.....	0.000494286468186
2	1093	kay.mann@enron.com..... Kay Mann.....	0.000494261619411

CATEGORY 3

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4698 COMPONENTS: 6
LARGEST COMPONENT SIZE: 4685 PERCENT OF TOTAL GRAPH: 99.72%
GROUP DEGREE: 0.07049 GRAPH DENSITY: 0.00170
GROUP CLOSENESS: 0.00510 GROUP BETWEENNESS: 0.06960
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.04

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
3	2363	morris.larubbio@enron.com.....	Morris Larubbio.....	0.001020780948689
3	718	bruce.mills@enron.com.....	Bruce Mills.....	0.000980615289663
3	594	amanda.schultz@enron.com.....	Amanda Schultz.....	0.000949058445429
3	16342	amy.ochoa@enron.com.....	Amy Ochoa.....	0.000948042922951
3	6752	chance.rabon@enron.com.....	Chance Rabon.....	0.000903863120773
3	15536	jay.smith@enron.com.....	Jay Smith.....	0.000872105001813
3	4320	erwin.landivar@enron.com.....	Erwin Landivar.....	0.000850443826055
3	6579	andres.balmaceda@enron.com.....	Andres Balmaceda....	0.000839203276462
3	23522	john.o'conner@enron.com.....	John Conner.....	0.000822004114002
3	237	charles.brewer@enron.com.....	Charles Brewer.....	0.000803686250356
3	353	mark.symms@enron.com.....	Mark Symms.....	0.000800571151252
3	6683	carole.frank@enron.com.....	Carole Frank.....	0.000792833595108
3	655	karen.buckley@enron.com.....	Karen Buckley.....	0.000778043460693
3	2501	kimberly.bates@enron.com.....	Kimberly Bates.....	0.000750125656713
3	64095	devries@enron.com.....	Devries.....	0.000740211436400
3	14765	darron.giron@enron.com.....	Darron Giron.....	0.000733296006186
3	16333	enw-employees@enron.com.....	enw-employees.....	0.000729033912227
3	63997	lagrasta@enron.com.....	Lagrasta.....	0.000716721930510
3	20686	rosalinda.zermeno@enron.com.....	Rosalinda Zermeno...	0.000693197947004

3 7163 holly.heath@enron.com..... Holly Heath..... 0.000688478749333

CATEGORY 4

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3157 COMPONENTS: 9
LARGEST COMPONENT SIZE: 3136 PERCENT OF TOTAL GRAPH: 99.33%
GROUP DEGREE: 0.09114 GRAPH DENSITY: 0.00127
GROUP CLOSENESS: 0.00289 GROUP BETWEENNESS: 0.19926
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
4	2505	mara.bronstein@enron.com.....	Mara Bronstein.....	0.001187997088076
4	20150	philip.polsky@enron.com.....	0.000818567856795
4	773	.ward@enron.com.....	houston.....	0.000751953743619
4	14660	theresa.staab@enron.com.....	0.000725819772445
4	18754	valerie.vela@enron.com.....	Valerie Vela.....	0.000654024597874
4	465	dipak.agarwalla@enron.com.....	Dipak Agarwalla.....	0.000612009047108
4	3605	t.lucci@enron.com.....	Paul T. Lucci.....	0.000536658305317
4	52900	'williams@enron.com.....	0.000527977577506
4	8819	d.mcilvoy@enron.com.....	Karen D. McIlvoy....	0.000515260857618
4	715	shelly.mendel@enron.com.....	Shelly Mendel.....	0.000509582900182
4	6635	arvel.martin@enron.com.....	Arvel Martin.....	0.000477629632255
4	625	p.adams@enron.com.....	Jacqueline P. Adams.	0.000464865109127
4	40426	s.ward@enron.com.....	Kim S. Ward.....	0.000460023685377
4	32548	john.kiani@enron.com.....	0.000427630112348
4	1769	mark.whitt@enron.com.....	0.000401837665184
4	19098	amy.felling@enron.com.....	Amy Felling.....	0.000396195180604
4	719	l.mims@enron.com.....	Patrice L. Mims.....	0.000386317255399
4	774	charles.weldon@enron.com.....	V. Charles Weldon...	0.000373450799570
4	338	bryce.schneider@enron.com.....	Bryce Schneider.....	0.000373107573232
4	43902	chike.okpara@enron.com.....	Chike Okpara.....	0.000362428886254

CATEGORY 5

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2385 COMPONENTS: 13
LARGEST COMPONENT SIZE: 2343 PERCENT OF TOTAL GRAPH: 98.24%
GROUP DEGREE: 0.20805 GRAPH DENSITY: 0.00126
GROUP CLOSENESS: 0.00107 GROUP BETWEENNESS: 0.37902
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
5	1175	arsystem@mailman.enron.com.....	ARSystem.....	0.001574130836517
5	41760	matt.dawson@enron.com.....	0.001341456308938
5	311	hal.mckinney@enron.com.....	Hal McKinney.....	0.001330269634127
5	20117	arsystem@ect.enron.com.....	0.001228208602414
5	360	wayne.vinson@enron.com.....	Donald Wayne Vinson.	0.001210361972996
5	14566	approval.eol.gas.traders@enron.com.....	0.001159613137693
5	15555	neal.d.winfree@enron.com.....	0.000952146810863
5	14602	kay.classen@enron.com.....	Kay Classen.....	0.000916469529903
5	39760	fletcher.j.sturm@enron.com.....	"fletcher.j.sturm@en	0.000907536017550
5	9320	information.management@enron.com.....	0.000636907455690
5	131	maria.van@enron.com.....	Maria Van houten....	0.000610300955835
5	22380	m.hall@enron.com.....	Bob M. Hall.....	0.000516527037409
5	41075	daemon.extra@enron.com.....	EXTRA Mailer Daemon.	0.000496930219667
5	17260	shemeika.landry@enron.com.....	0.000436118753118
5	32858	arsystem@enron.com.....	0.000419123791477
5	694	brad.jones@enron.com.....	Brad Jones.....	0.000359644926769
5	30857	erequest@enron.com.....	0.000358074536540
5	6784	daryll.fuentes@enron.com.....	Daryll Fuentes.....	0.000339869838226
5	1332	michael.kass@enron.com.....	Michael Kass.....	0.000312726975176
5	14798	sonya.johnson@enron.com.....	Sonya Johnson.....	0.000278234712176

CATEGORY 6

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2856 COMPONENTS: 9
 LARGEST COMPONENT SIZE: 2819 PERCENT OF TOTAL GRAPH: 98.70%
 GROUP DEGREE: 0.05589 GRAPH DENSITY: 0.00140
 GROUP CLOSENESS: 0.00135 GROUP BETWEENNESS: 0.12907
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
6	427	chris.dorland@enron.com.....	Chris Dorland.....	0.000829621769665
6	453	cooper.richey@enron.com.....	Cooper Richey.....	0.000769113638628
6	38243	jr.martinez@enron.com.....	0.000748348974864
6	13416	gerri.gosnell@enron.com.....	Gerri Gosnell.....	0.000747754316165
6	712	jonathan.mckay@enron.com.....	Jonathan Mckay.....	0.000744392012942
6	35820	greg.frers@enron.com.....	Greg Frers.....	0.000741072769627
6	80	john.zufferli@enron.com.....	0.000726454459572
6	6765	carlos.torres@enron.com.....	Carlos Torres.....	0.000691880065337

6	462	ryan.watt@enron.com.....	Ryan Watt.....	0.000612055728614
6	21948	michelle.wells@enron.com.....	Michelle Wells.....	0.000577940195825
6	37969	molnar.mark@enron.com.....	0.000561595583295
6	1189	dl-ga-canada_calgary@enron.com.....	DL-GA-Canada_Calgary	0.000556375086526
6	734	dutch.quigley@enron.com.....	Dutch Quigley.....	0.000526694264641
6	795	john.griffith@enron.com.....	John Griffith.....	0.000524559329619
6	3617	chris.cramer@enron.com.....	Chris Cramer.....	0.000523267256669
6	444	angela.mcculloch@enron.com.....	Angela McCulloch....	0.000520137853778
6	35828	greg.mann@enron.com.....	Greg Mann.....	0.000511308688617
6	38598	f.wong@enron.com.....	Michael F Wong.....	0.000478494144063
6	63973	milnthorp@enron.com.....	0.000478287851721
6	2563	brad.mckay@enron.com.....	Brad Mckay.....	0.000444208485577

CATEGORY 7

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6623 COMPONENTS: 12
LARGEST COMPONENT SIZE: 6572 PERCENT OF TOTAL GRAPH: 99.23%
GROUP DEGREE: 0.08334 GRAPH DENSITY: 0.00106
GROUP CLOSENESS: 0.00060 GROUP BETWEENNESS: 0.09969
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
7	1017	settlements.ees@enron.com.....	EES Power Settlement	0.001409769839186
7	84535	rring@ees.enron.com.....	NYISO TIE List.....	0.001169214415918
7	6998	jeffrey.jackson@enron.com.....	Jeffrey Jackson.....	0.000963320018556
7	59116	weekly.report@enron.com.....	0.000826629825354
7	21830	paul.rodger@enron.com.....	Paul Rodger.....	0.000781719530528
7	11307	jebong.lee@enron.com.....	0.000724479013296
7	2352	fabian.taylor@enron.com.....	Fabian Taylor.....	0.000711825762056
7	438	sean.lalani@enron.com.....	Sean Lalani.....	0.000639518998844
7	443	mike.macphee@enron.com.....	Mike Macphee.....	0.000591816044036
7	35738	penny.mccarran@enron.com.....	Penny McCarran.....	0.000565921355502
7	21829	paul.dunsmore@enron.com.....	Paul Dunsmore.....	0.000556405798017
7	844	robert'.harshbarger@enron.com.....	Robert Harshbarger..	0.000549804819768
7	2380	tom.dutta@enron.com.....	Tom Dutta.....	0.000540184041164
7	58211	massage.therapy@enron.com.....	0.000533279294119
7	49422	scott.kauffman@enron.com.....	0.000524067551177
7	53751	vicki.berg@enron.com.....	Vicki Berg.....	0.000512397293925
7	26919	janet.bowers@enron.com.....	Janet Bowers.....	0.000478241192143
7	194	julie.sarnowski@enron.com.....	0.000472234530220
7	8792	sitara@enron.com.....	Sitara.....	0.000407644180612

7 324 juan.padron@enron.com..... Juan Padron..... 0.000407453318653

CATEGORY 8

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 237 COMPONENTS: 1
LARGEST COMPONENT SIZE: 237 PERCENT OF TOTAL GRAPH: 100%
GROUP DEGREE: 0.11467 GRAPH DENSITY: 0.00083
GROUP CLOSENESS: 0.00053 GROUP BETWEENNESS: 0.09985
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
8	1894	chairman.enron@enron.com.....	Enron Office Of The	0.000927562299313
8	2368	coo.jeff@enron.com.....	Jeff McMahon - Presi	0.000914154964422
8	1914	chairman.ken@enron.com.....	Ken Lay - Office of	0.000909037256528
8	54525	dl-ga-all_egs@enron.com.....	DL-GA-all_egs.....	0.000712308411108
8	44086	enron.operations@enron.com.....	0.000678703259497
8	65258	nate.ellis@enron.com.....	0.000674942026342
8	58495	barbara.taylor@enron.com.....	0.000647457698254
8	1895	dl-ga-all_enron_worldwide2@enron.com....	DL-GA-all_enron_worl	0.000619311533344
8	24446	executive.office@enron.com.....	Office of the Chief	0.000597890616624
8	4762	office.chairman@enron.com.....	0.000593806777944
8	4884	enron.worldwide@enron.com.....	0.000591586827695
8	630	jason.althaus@enron.com.....	Jason Althaus.....	0.000584571775075
8	21296	money.in.motion@mailman.enron.com.....	0.000583867882416
8	58487	mhaedic@ect.enron.com.....	"Mark Haedicke J.D "	0.000580002758654
8	41153	esa_employees@enron.com.....	0.000577065975487
8	3609	milagros.velasquez@enron.com.....	Milagros Velasquez..	0.000576277848309
8	18400	bill.gulyassy@enron.com.....	0.000569971766168
8	13404	deane.pierce@enron.com.....	Deane Pierce.....	0.000564355366972
8	20456	peter.ghavami@enron.com.....	0.000556176469265
8	5688	ken.skilling@enron.com.....	0.000552705327190

CATEGORY 9

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3091 COMPONENTS: 8
LARGEST COMPONENT SIZE: 3075 PERCENT OF TOTAL GRAPH: 99.48%
GROUP DEGREE: 0.08164 GRAPH DENSITY: 0.00097
GROUP CLOSENESS: 0.00465 GROUP BETWEENNESS: 0.23917
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
9	8773	michelle.nelson@enron.com	Michelle Nelson	0.001507114155269
9	742	amanda.rybarski@enron.com	Amanda Rybarski	0.001394458259131
9	707	mike.maggi@enron.com	Mike Maggi	0.001278302835769
9	12142	adriana.wynn@enron.com	Adriana Wynn	0.000739838987366
9	14652	amanda.huble@enron.com	Amanda Huble	0.000642563356775
9	532	reginald.hart@enron.com	Reginald Hart	0.000636878764667
9	20547	lance.jameson@enron.com		0.000555991530417
9	2185	michelle.hicks@enron.com	Michelle Hicks	0.000426219171297
9	14874	reginald.smith@enron.com		0.000419969124568
9	8988	s.presas@enron.com	Gracie S. Presas	0.000406706548263
9	83584	adam.giannone@enron.com	Adam Giannone	0.000329105902344
9	7619	vikram.singh@enron.com	Vikram Singh	0.000312798538870
9	2603	julia.sudduth@enron.com	Julia Sudduth	0.000302450459812
9	243	jim.cashion@enron.com	Jim Cashion	0.000298734579388
9	2609	alex.villarreal@enron.com	Alex Villarreal	0.000298599782582
9	1810	wendy.fischer@enron.com	Wendy Fincher	0.000297432909315
9	663	julie.clyatt@enron.com	Julie Clyatt	0.000294151169298
9	7661	marc.graubart@enron.com	Marc Graubart	0.000287774474574
9	21096	rita.houston@enron.com	Rita Houston	0.000284827677990
9	8400	robin.veariel@enron.com	Robin Veariel	0.000274705443903

CATEGORY 10

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 9232 COMPONENTS: 16
 LARGEST COMPONENT SIZE: 9194 PERCENT OF TOTAL GRAPH: 99.59%
 GROUP DEGREE: 0.16899 GRAPH DENSITY: 0.00065
 GROUP CLOSENESS: 0.00087 GROUP BETWEENNESS: 0.25984
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
10	7667	ipayit@enron.com	iPayit@Enron.com>@EN	0.001408019131315
10	797	sap_security@enron.com		0.001376348794641
10	2985	enron.payroll@enron.com	"Enron.Payroll@enron	0.001370844057360
10	3652	payroll.enron@enron.com	Enron Payroll	0.001339999077432
10	11479	ibuyit.payables@enron.com	iBuyit.Payables@Enro	0.001322737866577
10	6180	enron.expertfinder@enron.com		0.001277069974023
10	6520	payables.ibuyit@enron.com	iBuyit Payables	0.001224519134846
10	3602	robert.jones@mailman.enron.com		0.001206765097339

10	18326	mbx_iscinfra@enron.com.....	0.001114268782204
10	18319	tahnee.stall@enron.com.....	0.001110400348778
10	1502	ic@enron.com.....	"ic@enron.com".....	0.001050188442557
10	28467	resources@enron.com.....	"human resources@enr	0.001014836571422
10	1187	confirmit@enron.com.....	Confirmit.....	0.000989726115596
10	18318	tammy.marcontell@enron.com.....	0.000952825592129
10	21165	carolyn.graham@enron.com.....	Carolyn Graham.....	0.000823777057485
10	464	sunil.abraham@enron.com.....	Sunil Abraham.....	0.000815470882058
10	28781	enronanywhere@enron.com.....	"enronanywhere@enron	0.000769680891276
10	17156	clickathomepilot3@enron.com.....	"ClickAtHomePilot3@e	0.000756431101787
10	8342	ibuyit@enron.com.....	0.000738355106937
10	34952	isc.groups@enron.com.....	0.000650148600542

CATEGORY 11

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 1966 COMPONENTS: 14
LARGEST COMPONENT SIZE: 1934 PERCENT OF TOTAL GRAPH: 98.37%
GROUP DEGREE: 0.10475 GRAPH DENSITY: 0.00305
GROUP CLOSENESS: 0.00176 GROUP BETWEENNESS: 0.18861
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
11	566	evelyn.metoyer@enron.com.....	Evelyn Metoyer.....	0.001526876706561
11	1056	kerri.thompson@enron.com.....	Kerri Thompson.....	0.001495630223664
11	582	stephanie.piwetz@enron.com.....	Stephanie Piwetz....	0.001246524881606
11	21	kate.symes@enron.com.....	Kate Symes.....	0.001071523355340
11	138	kysa.alport@enron.com.....	0.000970956469983
11	796	shift.dl-portland@enron.com.....	DL-Portland Real Tim	0.000959432896811
11	20	geir.solberg@enron.com.....	Geir Solberg.....	0.000837022274781
11	126	theresa.villegiante@enron.com.....	0.000823167032216
11	127	josie.jarnagin@enron.com.....	0.000808053992738
11	40	kara.ausenhus@enron.com.....	Kara Ausenhus.....	0.000804592961174
11	490	sharen.cason@enron.com.....	Sharen Cason.....	0.000729561254883
11	2009	alexander.mcelreath@enron.com.....	Alexander McElreath.	0.000718309066716
11	1051	judy.dyer@enron.com.....	Judy Dyer.....	0.000717313008419
11	1953	kayla.harmon@enron.com.....	Kayla Harmon.....	0.000692622393449
11	219	michael.mier@enron.com.....	Michael Mier.....	0.000658581398353
11	981	portland.shift@enron.com.....	Portland Shift.....	0.000656633930748
11	17	bert.meyers@enron.com.....	Bert Meyers.....	0.000647197008757
11	620	ryan.williams@enron.com.....	Ryan Williams.....	0.000638369003422
11	972	shift.portland@enron.com.....	Portland Shift.....	0.000621784981891

11 123 chris.mumm@enron.com..... Chris Mumm..... 0.000620099069163

CATEGORY 12

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4630 COMPONENTS: 13
LARGEST COMPONENT SIZE: 4578 PERCENT OF TOTAL GRAPH: 98.88%
GROUP DEGREE: 0.09902 GRAPH DENSITY: 0.00108
GROUP CLOSENESS: 0.0068 GROUP BETWEENNESS: 0.15957
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
12	25328	tim.mckone@enron.com.....		0.001112582690605
12	51741	ramon.alvarez@enron.com.....		0.000875863487561
12	17749	martin.smith@enron.com.....	Martin Smith.....	0.000661615718698
12	80962	mlawles@enron.com.....		0.000552769019784
12	41651	tlehan@enron.com.....		0.000543875920904
12	41578	ccheek@enron.com.....		0.000540950394832
12	80963	staci_holtzman@enron.com.....		0.000537241130249
12	83451	kevin.d.jordan@enron.com.....		0.000525195133049
12	81353	jazayeri.peter@enron.com.....		0.000486844092153
12	82077	'bump@enron.com.....		0.000474503349677
12	17623	duncan.dave@enron.com.....	Dave Duncan.....	0.000458159956047
12	82078	dbump@ect.enron.com.....		0.000458070562629
12	30929	john.hopley@enron.com.....		0.000441690483965
12	2232	david.reinfeld@enron.com.....	David Reinfeld.....	0.000439909549706
12	35852	'rogers@enron.com.....		0.000439313499888
12	43924	filterpst@enron.com.....	FILTERPST.....	0.000438781215502
12	4370	hcubill@enron.com.....		0.000417858611289
12	6344	erica.harris@enron.com.....		0.000414267975706
12	6351	fernando.parra@enron.com.....		0.000411151153344
12	58471	nicola.sanders@enron.com.....		0.000397580267494

CATEGORY 13

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6933 COMPONENTS: 18
LARGEST COMPONENT SIZE: 6873 PERCENT OF TOTAL GRAPH: 99.13%
GROUP DEGREE: 0.08059 GRAPH DENSITY: 0.00087
GROUP CLOSENESS: 0.00050 GROUP BETWEENNESS: 0.14974
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
13	4110	klay@enron.com		0.001601547055236
13	14112	lfan@mailman.enron.com	"lfan"	0.000275370341249
13	3156	steve.walker@enron.com		0.000237467793189
13	10603	colin.skellett@enron.com		0.000228362440052
13	21872	dave.ellis@enron.com	Dave Ellis	0.000183015269560
13	14521	qwest.net@mailman.enron.com	"smoray!qwest.net"	0.000181180119423
13	1911	resources.human@enron.com	Human Resources	0.000175591724167
13	12993	cecil.stinemetz@enron.com	Cecil Stinemetz	0.000174198878161
13	12940	mike.underwood@enron.com	Mike Underwood	0.000166045381950
13	13411	rebecca.longoria@enron.com	Rebecca Longoria	0.000158315996298
13	1816	josh.duncan@enron.com	Josh Duncan	0.000146302011811
13	21873	jeff.borg@enron.com	Jeff Borg	0.000145705460432
13	13343	delia.walters@enron.com	Delia Walters	0.000144502959270
13	41375	steve.dahnke@enron.com		0.000140111382700
13	18017	ruth.mann@enron.com		0.000134505674594
13	1891	dl-ga-all_enron_worldwide3@enron.com	DL-GA-all_enron_worl	0.000124086451007
13	7942	timothy.hubbard@enron.com		0.000123929995919
13	50042	psmith3@enron.com		0.000121327168634
13	56739	jim.barnes@enron.com		0.000120099584951
13	62265	corn1983@enron.com		0.000120099584951

CATEGORY 14

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4888 COMPONENTS: 11
 LARGEST COMPONENT SIZE: 4856 PERCENT OF TOTAL GRAPH: 99.35%
 GROUP DEGREE: 0.06139 GRAPH DENSITY: 0.00143
 GROUP CLOSENESS: 0.00127 GROUP BETWEENNESS: 0.08957
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
14	83555	weatherwarn@mailman.enron.com		0.001378078541067
14	83556	subscribers@mailman.enron.com		0.001347729515068
14	8676	troy'.brothers@enron.com		0.001120182691548
14	3000	tammy.gilmore@enron.com	Tammy Gilmore	0.001040966817847
14	1364	laura.lantefield@enron.com	Laura Lantefield	0.001019182917989
14	14572	jeff.nielsen@enron.com	Jeff Nielsen	0.001013928475712
14	22443	benjamin.schoene@enron.com	Benjamin Schoene	0.000996013660566
14	14742	paul.tate@enron.com	Paul Tate	0.000965088538894

14	11187	elberg.gelin@enron.com.....	0.000923310865935
14	21134	corey.wilkes@enron.com.....	Corey Wilkes.....	0.000916634570509
14	24546	dscott4@enron.com.....	0.000825500585103
14	15268	eileen.peebles@enron.com.....	0.000792167307948
14	54135	bullets@enron.com.....	0.000756070089463
14	549	lisa.kinsey@enron.com.....	Lisa Kinsey.....	0.000748912162941
14	53749	v.dickerson@enron.com.....	Steve V Dickerson...	0.000724070652120
14	14816	sebastian.corbacho@enron.com.....	0.000656492998498
14	44117	carolyn.descoteaux@enron.com.....	0.000627984004197
14	36476	paul.miller@enron.com.....	0.000620904414153
14	53750	michael.loeffler@enron.com.....	Michael Loeffler....	0.000619806090657
14	761	m.tholt@enron.com.....	Jane M. Tholt.....	0.000607471228890

CATEGORY 15

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8081 COMPONENTS: 17
LARGEST COMPONENT SIZE: 8034 PERCENT OF TOTAL GRAPH: 99.42%
GROUP DEGREE: 0.06638 GRAPH DENSITY: 0.00087
GROUP CLOSENESS: 0.00060 GROUP BETWEENNESS: 0.08977
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
15	11411	phyllis.anzalone@enron.com.....	0.001253298613513
15	22613	tdonoho@enron.com.....	0.000914945170041
15	30920	kwatson@enron.com.....	0.000774913707859
15	17265	fsturm@enron.com.....	fsturm.....	0.000749642270803
15	58588	"undisclosed-recipient"@enron.com.....	"Undisclosed-Recipie	0.000672556823129
15	39519	jquenet@enron.com.....	Quenet.....	0.000654113616571
15	77620	fermis@ect.enron.com.....	0.000639238376987
15	59958	tdonoho@ect.enron.com.....	0.000558169368411
15	35461	tmartin@ect.enron.com.....	0.000487853254502
15	14579	michael.schilmoeller@enron.com.....	Michael SCHILMOELLER	0.000442609712329
15	41633	fking@enron.com.....	0.000409298705275
15	29712	vkamins@ect.enron.com.....	0.000406623618470
15	4556	mlehart@enron.com.....	0.000375768458329
15	85723	blichte@enron.com.....	0.000307549527753
15	2618	rick.wurlitzer@enron.com.....	Rick Wurlitzer.....	0.000267904255165
15	8792	sitara@enron.com.....	Sitara.....	0.000259920051887
15	18943	james.monroe@enron.com.....	James Monroe.....	0.000246386620151
15	32484	"the.desk":@enron.com.....	0.000237989286848
15	6322	rick.guttruff@enron.com.....	0.000232005155999

15 24011 announcement.ees@enron.com..... EES Product Announce 0.000225283228657

CATEGORY 16

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 10117 COMPONENTS: 13
LARGEST COMPONENT SIZE: 10085 PERCENT OF TOTAL GRAPH: 99.68%
GROUP DEGREE: 0.09216 GRAPH DENSITY: 0.00069
GROUP CLOSENESS: 0.00117 GROUP BETWEENNESS: 0.10983
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
16	6169	dottie.kerr@enron.com.....		0.000580233467676
16	15138	kirk.neuner@enron.com.....	Kirk Neuner.....	0.000558905753396
16	83222	pickel.robert@enron.com.....		0.000508826209101
16	30930	david.martin@enron.com.....		0.000473606941108
16	1838	paul.lebeau@enron.com.....	Paul Lebeau.....	0.000467157456745
16	17950	tracey.kozadinos@enron.com.....	Tracey Kozadinos....	0.000459044558479
16	32307	andrew.s.fastow@enron.com.....	Andrew.S.Fastow....	0.000453735287827
16	24080	registrar.isc@enron.com.....	ISC Registrar.....	0.000451068744026
16	19236	lisa.polk@enron.com.....		0.000401456948249
16	29742	mcarson@enron.com.....		0.000360312055141
16	10959	althea.gordon@enron.com.....		0.000346135857069
16	9372	jeffrey.mcclellan@enron.com.....		0.000344556711531
16	10779	michael.rosen@enron.com.....		0.000340814768847
16	1827	constance.charles@enron.com.....	Constance Charles...	0.000340010446842
16	29741	jballen@enron.com.....		0.000334910619895
16	20337	ginger.gamble@enron.com.....		0.000317164129345
16	19662	riccardo.bortolotti@enron.com.....		0.000311737273871
16	15911	heather.johnson@enron.com.....		0.000308307655930
16	6165	ypo_international@enron.com.....		0.000308089644486
16	23987	sammi@enron.com.....	Sammi.....	0.000301927804327

CATEGORY 17

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4851 COMPONENTS: 11
LARGEST COMPONENT SIZE: 4817 PERCENT OF TOTAL GRAPH: 99.30%
GROUP DEGREE: 0.09475 GRAPH DENSITY: 0.00103
GROUP CLOSENESS: 0.00109 GROUP BETWEENNESS: 0.14956
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
17	37111	.stephens@enron.com.....	bridgeline.....	0.001323511196538
17	3354	tom.ward@enron.com.....	TOM WARD.....	0.000966067937626
17	37443	'fenner@enron.com.....	0.000937569931055
17	81781	'jernigan@enron.com.....	0.000928733539493
17	52959	'thompson@enron.com.....	0.000920837164148
17	4581	mday@enron.com.....	0.000882216853996
17	31826	"sbigalow"@enron.com.....	"sbigalow".....	0.000852736525420
17	24509	'proctor@enron.com.....	0.000850029138921
17	87152	'rector@enron.com.....	0.000740618599562
17	62800	beth.cherry@enron.com.....	Beth Cherry.....	0.000710503914088
17	20246	valerie.curtis@enron.com.....	0.000700356319002
17	85652	vo.hoang@enron.com.....	Hoang Vo.....	0.000694711807182
17	38921	jennifer.ballas@enron.com.....	Jennifer Ballas.....	0.000671776000082
17	82927	'peters@enron.com.....	0.000638237810676
17	52960	'lipper@enron.com.....	0.000613008799940
17	64777	'keffer@enron.com.....	0.000581600982873
17	87149	'gapinski@enron.com.....	0.000573763927601
17	81538	kalembka"."lech@enron.com.....	0.000567123206439
17	52997	assoc.'.'california@enron.com.....	California Cast Meta	0.000563394640988
17	24662	.germany@enron.com.....	wd.....	0.000560715397035

CATEGORY 18

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2653 COMPONENTS: 17
 LARGEST COMPONENT SIZE: 2600 PERCENT OF TOTAL GRAPH: 98.00%
 GROUP DEGREE: 0.25123 GRAPH DENSITY: 0.00339
 GROUP CLOSENESS: 0.00077 GROUP BETWEENNESS: 0.23940
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.04

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
18	3126	deborah.whitehead@enron.com.....	0.001354693023873
18	49935	sstoness@enron.com.....	0.001343278555921
18	3136	leasa.lopez@enron.com.....	0.001339989957675
18	37928	dblack@enron.com.....	0.001334889634941
18	24207	.sue@enron.com.....	e-mail.....	0.001329385040080
18	1629	ken.gustafson@enron.com.....	0.001307291111144
18	51122	jlewis@enron.com.....	".....	0.001284769312297
18	49932	tjohnso8@enron.com.....	0.001275922795514

```

18 48832 james_d_steffes@enron.com..... James D Steffes..... 0.001245590861660
18 28468 terry.donovan@enron.com..... 0.001190484546632
18 52415 johnson.tamara@enron.com..... 0.001164952932672
18 49988 clarian.vondrak@enron.com..... 0.001072665655453
18 50118 fvickers@enron.com..... 0.001048300576940
18 34229 mpalmer@enron.com..... mpalmer@enron.com... 0.001047373129123
18 1228 mark.fillinger@enron.com..... Mark Fillinger..... 0.001027154035031
18 52531 savage.gordon@enron.com..... Gordon Savage..... 0.001002542002956
18 52809 sf'.sue@enron.com..... Sue Mara at Enron SF 0.001002248494077
18 52530 coffing.timothy@enron.com..... Timothy Coffing.... 0.000969899406956
18 46859 smara@enron.com..... ". 0.000945990094958
18 3228 .jim@enron.com..... e-mail..... 0.000936966686254

```

CATEGORY 19

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 7223          COMPONENTS: 13
LARGEST COMPONENT SIZE: 7164 PERCENT OF TOTAL GRAPH: 99.18%
GROUP DEGREE: 0.09008   GRAPH DENSITY: 0.00083
GROUP CLOSENESS: 0.00047 GROUP BETWEENNESS: 0.17975
AVERAGE p(z|u): 0.02   STDEV p(z|u): 0.02

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
19	1488	perfmgmt@enron.com.....	"Performance Evaluat	0.001525791142995
19	22319	perfmgmt@ect.enron.com.....		0.001371552789879
19	3580	tiffany.smith@enron.com.....	Tiffany Smith.....	0.001153296600028
19	1599	john.disturnal@enron.com.....		0.001062847787452
19	7009	jay.knoblauh@enron.com.....	Jay Knoblauh.....	0.001029857296022
19	6943	gregory.schockling@enron.com.....	Gregory Schockling..	0.000984788500517
19	15206	jeff.stephens@enron.com.....	Jeff Stephens.....	0.000978016231526
19	19486	steve.beck@enron.com.....		0.000842856199101
19	8829	jeffery.stephens@enron.com.....	Jeffery Stephens....	0.000831147356346
19	3520	ragan.bond@enron.com.....	Ragan Bond.....	0.000820963992964
19	46295	derrick.jr.@enron.com.....		0.000560675749269
19	18047	project.team@enron.com.....		0.000526954152307
19	6170	henry.emery@enron.com.....		0.000507744129306
19	2625	jose.favela@enron.com.....		0.000477683957123
19	60151	hendrickson@enron.com.....		0.000454311071428
19	11553	hodge@enron.com.....		0.000411327981537
19	27157	griffith@enron.com.....		0.000397450507797
19	7147	cullen.duke@enron.com.....	Cullen Duke.....	0.000384117123224
19	15291	axisteam@enron.com.....	"The Associate and A	0.000373160231628

19 37128 dl-ga-pas@enron.com..... DL-GA-PAS..... 0.000358662874751

CATEGORY 20

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7203 COMPONENTS: 19
LARGEST COMPONENT SIZE: 7140 PERCENT OF TOTAL GRAPH: 99.13%
GROUP DEGREE: 0.08949 GRAPH DENSITY: 0.00069
GROUP CLOSENESS: 0.00047 GROUP BETWEENNESS: 0.11971
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
20	1530	sandy.morris@enron.com.....	Sandy Morris.....	0.000874483033774
20	22317	heidi.dubose@enron.com.....	0.000830742698895
20	1213	backbone.ens@enron.com.....	0.000729138965996
20	228	michael.belmont@enron.com.....	Michael Belmont....	0.000687949885586
20	19381	jim.fussell@enron.com.....	0.000663413538894
20	40864	murray.bridgmout@enron.com.....	Murray Bridgmout...	0.000633919853904
20	15669	sscott5@enron.com.....	0.000611384940309
20	682	claudia.guerra@enron.com.....	Claudia Guerra.....	0.000570851660099
20	65	anna.mehrer@enron.com.....	0.000554254779448
20	60202	peter.f.keavey@enron.com.....	peter.f.keavey.....	0.000538657148020
20	40865	datacomms.european@enron.com.....	European DataComms..	0.000497040933653
20	24256	cgerman@ect.enron.com.....	0.000492514042629
20	6861	david.wile@enron.com.....	David Wile.....	0.000491017748451
20	6891	gail.kettenbrink@enron.com.....	Gail Kettenbrink....	0.000470825018992
20	686	d..hogan@enron.com.....	Irena D. Hogan.....	0.000470467464701
20	83277	mtaylo1@ect.enron.com.....	0.000439404166548
20	25054	enron.gss@enron.com.....	Enron GSS.....	0.000420930480870
20	83912	mark.e.taylor@enron.com.....	".....	0.000417120023755
20	37242	bryan.vaclavik@enron.com.....	Bryan Vaclavik.....	0.000412479769939
20	52919	'pp'@enron.com.....	PP.....	0.000409820643319

CATEGORY 21

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5312 COMPONENTS: 15
LARGEST COMPONENT SIZE: 5262 PERCENT OF TOTAL GRAPH: 99.06%
GROUP DEGREE: 0.16252 GRAPH DENSITY: 0.00075
GROUP CLOSENESS: 0.00077 GROUP BETWEENNESS: 0.36966
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
21	3780	enron.mailsweeper.admin@enron.com.....		0.001361717288649
21	3973	admin.enron@enron.com.....	Enron MailSweeper Ad	0.001355363291674
21	57681	munder.mail.list@mailman.enron.com.....		0.001339082411155
21	6580	angela.barnett@enron.com.....	Angela Barnett.....	0.001095645310547
21	16217	rose.botello@enron.com.....		0.001085409345096
21	18727	regina.blackshear@enron.com.....	Regina Blackshear...	0.001036375541931
21	6837	diane.salcido@enron.com.....	Diane Salcido.....	0.001014008469762
21	20483	amber.limas@enron.com.....		0.001006886782093
21	7571	maria.sandoval@enron.com.....	Maria Sandoval.....	0.000940728480294
21	6992	judy.hernandez@enron.com.....	Judy Hernandez.....	0.000914536743364
21	16110	dbaughm@notes.enron.com.....		0.000868739973280
21	74130	shauncy.mathews@enron.com.....		0.000848616431053
21	19886	leslie.smith@enron.com.....		0.000816267499590
21	6977	jennifer.cutiaia@enron.com.....	Jennifer Cutaia....	0.000799789279406
21	31730	enron.messaging.administration@enron.com		0.000768852347271
21	74198	yvonne.acosta@enron.com.....		0.000729258614691
21	20467	shirlet.williams@enron.com.....		0.000680051378254
21	18364	nicole.mendez@enron.com.....		0.000660028430904
21	3357	adam.senn@enron.com.....	Adam Senn.....	0.000654768934867
21	53196	davette.warren@enron.com.....		0.000646320487144

CATEGORY 22

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4561 COMPONENTS: 11
 LARGEST COMPONENT SIZE: 4515 PERCENT OF TOTAL GRAPH: 98.99%
 GROUP DEGREE: 0.07743 GRAPH DENSITY: 0.00088
 GROUP CLOSENESS: 0.00077 GROUP BETWEENNESS: 0.14952
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
22	15169	jeff.pearson@enron.com.....	Jeff Pearson.....	0.001253668216739
22	69639	spendegr@enron.com.....		0.000989841761548
22	4989	debbie.doyle@enron.com.....		0.000869871909463
22	18735	lois.ford@enron.com.....	Lois Ford.....	0.000836060891752
22	28817	postmaster@corp.enron.com.....		0.000832470263197
22	24009	landry.pamela@enron.com.....	pamela landry.....	0.000808897513137
22	8399	jimmy.simien@enron.com.....	Jimmy Simien.....	0.000797366149122
22	39389	cyntia.distefano@enron.com.....	Cyntia DiStefano....	0.000792161720882

22	1365	annette.glod@enron.com.....	Annette Glod.....	0.000781796475648
22	15192	mike@enron.com.....	Mike.....	0.000766113550658
22	69562	larry.f.campbell@enron.com.....	".....	0.000751543355103
22	2188	s.gartner@enron.com.....	Julie S. Gartner....	0.000738060823994
22	69632	j.kinser@enron.com.....	0.000737706838406
22	82885	theresa_staab@enron.com.....	0.000704316279000
22	49937	rsanders@enron.com.....	0.000660653664704
22	44165	robin.border@enron.com.....	0.000637761479596
22	79464	quickplace@nahou-lnw01.ots.enron.com...	"customerservice"...	0.000621914392699
22	26498	scott.crowell@enron.com.....	0.000607660447529
22	1801	leesa.white@enron.com.....	Leesa White.....	0.000599852978852
22	8931	kevin.heal@enron.com.....	Kevin Heal.....	0.000596159519573

CATEGORY 23

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6040 COMPONENTS: 18
LARGEST COMPONENT SIZE: 5990 PERCENT OF TOTAL GRAPH: 99.17%
GROUP DEGREE: 0.10322 GRAPH DENSITY: 0.00066
GROUP CLOSENESS: 0.00064 GROUP BETWEENNESS: 0.14962
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
23	79501	gblair@ei.enron.com.....	0.000462412953153
23	21413	ken@enron.com.....	Ken.....	0.000323841152123
23	21422	hill.heather@enron.com.....	0.000286077056466
23	14561	hshivel@enron.com.....	0.000269967911928
23	59931	mogel@enron.com.....	0.000243749594960
23	52372	nahigian@enron.com.....	0.000192334598779
23	62694	jason.r.williams@enron.com.....	0.000187377638054
23	52357	counihan@enron.com.....	0.000171950572198
23	52378	violette@enron.com.....	0.000167424947171
23	11194	omar.hasan@enron.com.....	0.000160489024376
23	26485	dgiron@ect.enron.com.....	0.000155796766887
23	71916	fatimah.ducros@enron.com.....	Fatimah Ducros.....	0.000153229223724
23	49830	theizenrader@enron.com.....	0.000146050067599
23	22671	dsmith3@enron.com.....	0.000145812358196
23	25760	.pam@enron.com.....	e-mail.....	0.000142357027217
23	59955	america.dl-outlook@enron.com.....	DL-Outlook Users Nor	0.000129385918782
23	39271	gstorey@enron.com.....	0.000128688760621
23	32865	network.security@enron.com.....	0.000127291074582
23	13429	cynthia.boseman-harris@enron.com.....	Cynthia Boseman-Harr	0.000123471212527

23 58138 julie.a.gomez@enron.com..... 0.000120555589723

CATEGORY 24

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6534 COMPONENTS: 16
LARGEST COMPONENT SIZE: 6492 PERCENT OF TOTAL GRAPH: 99.36%
GROUP DEGREE: 0.06928 GRAPH DENSITY: 0.00092
GROUP CLOSENESS: 0.00076 GROUP BETWEENNESS: 0.09970
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Table with 4 columns: Topic#, ID#, Email Address, Name, p(z|u). Lists 24 users with their respective IDs and email addresses, such as ted.robinson@enron.com and alexandre.bueno@enron.com.

CATEGORY 25

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2949 COMPONENTS: 11
LARGEST COMPONENT SIZE: 2890 PERCENT OF TOTAL GRAPH: 98.00%
GROUP DEGREE: 0.16070 GRAPH DENSITY: 0.00136
GROUP CLOSENESS: 0.00070 GROUP BETWEENNESS: 0.26934
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
25	983	westdesksupport@enron.com		0.001101294436163
25	31991	julius.zajda@enron.com		0.001085519268517
25	6707	brad.romine@enron.com	Brad Romine	0.000907610059737
25	2104	tom.barkley@enron.com	Tom Barkley	0.000896814524694
25	33702	adam.brulinski@enron.com		0.000861365322836
25	21046	jarek.dybowski@enron.com	Jarek Dybowski	0.000819074111278
25	30684	eloise.meza@enron.com		0.000789624728705
25	2078	jason.sokolov@enron.com	Jason Sokolov	0.000769174246179
25	8822	kenneth.parkhill@enron.com	Kenneth Parkhill	0.000722874376853
25	21151	steve.bigalow@enron.com	Steve Bigalow	0.000700520452098
25	15129	bessik.matchavariani@enron.com	Bessik Matchavariani	0.000676921169226
25	6991	john.henderson@enron.com	John Henderson	0.000577687162124
25	32739	wsmith2@enron.com		0.000571572152068
25	32865	network.security@enron.com		0.000541798179713
25	6032	kevin.moore@enron.com		0.000535716372260
25	19763	mary.bailey@enron.com		0.000519439944775
25	29987	youyi.feng@enron.com		0.000516879223629
25	2092	rakesh.bharati@enron.com	Rakesh Bharati	0.000497251843575
25	1728	mark.ruane@enron.com		0.000480913131608
25	31231	katja.schilling@enron.com		0.000476298585543

CATEGORY 26

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6559 COMPONENTS: 14
 LARGEST COMPONENT SIZE: 6527 PERCENT OF TOTAL GRAPH: 99.51%
 GROUP DEGREE: 0.07382 GRAPH DENSITY: 0.00091
 GROUP CLOSENESS: 0.00124 GROUP BETWEENNESS: 0.09968
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
26	3783	douglas.smith@enron.com	Douglas SMITH	0.000588198618238
26	12183	boardroom@enron.com	Boardroom	0.000400120120558
26	39745	joao.albuquerque@enron.com		0.000396507373111
26	53169	.arm@enron.com	e-mail	0.000362655683908
26	21296	money.in.motion@mailman.enron.com		0.000337474714293
26	587	lindsay.renaud@enron.com	Lindsay Renaud	0.000335546829323
26	18361	s.yao@enron.com	Anne S. Yao	0.000334001683022
26	1706	gary.peng@enron.com		0.000314439436021

26	18460	jens.gobel@enron.com.....	Jens Gobel.....	0.000305669278358
26	17645	cassandra.chinkin@enron.com.....	Cassandra Chinkin...	0.000297263581907
26	3502	e.jones@enron.com.....	Karen E. Jones.....	0.000292377166431
26	17733	padmesh.thuraisingham@enron.com.....	Padmesh Thuraisingha	0.000290774455809
26	48719	the.distribution@enron.com.....	0.000284647663389
26	9525	lkitchen@enron.com.....	0.000272276242132
26	2251	kikumi.kishigami@enron.com.....	Kikumi Kishigami...	0.000269348319116
26	80335	leonardo.cardoso@enron.com.....	Leonardo Cardoso...	0.000254792475714
26	82299	open2win.0111.net@mailman.enron.com.....	0.000253328876855
26	416	jason.biever@enron.com.....	Jason Biever.....	0.000240887029362
26	12155	david.tonsall@enron.com.....	David Tonsall.....	0.000235305103431
26	56715	.boe@enron.com.....	lawrence.....	0.000230033120632

CATEGORY 27

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7261 COMPONENTS: 15
LARGEST COMPONENT SIZE: 7220 PERCENT OF TOTAL GRAPH: 99.44%
GROUP DEGREE: 0.29692 GRAPH DENSITY: 0.02119
GROUP CLOSENESS: 0.16754 GROUP BETWEENNESS: 0.29250
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
27	17493	.linda@enron.com.....	e-mail.....	0.000849482133330
27	44102	christi.culwell@enron.com.....	0.000831723790035
27	7507	crystal.reyna@enron.com.....	Crystal Reyna.....	0.000827492263327
27	24848	maria.cisneros@enron.com.....	0.000715916931542
27	77911	walt.serrano@enron.com.....	Walt Serrano.....	0.000696540149539
27	20010	cstclai@enron.com.....	0.000677666209391
27	16130	mills.bret@enron.com.....	Bret Mills.....	0.000659217051214
27	2828	jay.wills@enron.com.....	0.000643678177557
27	4511	lgillet@enron.com.....	0.000637170888441
27	13179	.sally@enron.com.....	e-mail.....	0.000627451721759
27	19631	sandra.mcnichols@enron.com.....	0.000613256029816
27	37179	fenner', 'molly@enron.com.....	Molly Fenner.....	0.000613180814743
27	778	ashley.worthing@enron.com.....	Ashley Worthing....	0.000602175621182
27	29310	claudia.santos@enron.com.....	Claudia Santos.....	0.000593610671816
27	44164	jon.trevelise@enron.com.....	0.000590106209773
27	1040	'kuehn@enron.com.....	0.000581110745711
27	1264	rose.rivera@enron.com.....	Rose Rivera.....	0.000579042137249
27	969	.tim@enron.com.....	e-mail.....	0.000554628316058
27	769	laura.vargas@enron.com.....	Laura Vargas.....	0.000530206305260

27 44142 ron.beidelman@enron.com..... 0.000522829926453

CATEGORY 28

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3175 COMPONENTS: 11
LARGEST COMPONENT SIZE: 3137 PERCENT OF TOTAL GRAPH: 98.80%
GROUP DEGREE: 0.06818 GRAPH DENSITY: 0.00069
GROUP CLOSENESS: 0.00081 GROUP BETWEENNESS: 0.13974
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Table with 5 columns: Topic#, ID#, Email Address, Name, p(z|u). Lists 28 users with their respective IDs and scores.

CATEGORY 29

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3242 COMPONENTS: 20
LARGEST COMPONENT SIZE: 3191 PERCENT OF TOTAL GRAPH: 98.43%
GROUP DEGREE: 0.29692 GRAPH DENSITY: 0.02119
GROUP CLOSENESS: 0.16754 GROUP BETWEENNESS: 0.29250
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.04

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
29	53773	controllers.dl-ets@enron.com	DL-ETS Gas Controlle	0.001424273287272
29	76971	dl-etsgascontrollers@enron.com	DL-ETS Gas Controlle	0.001257231555973
29	17619	beverly.miller@enron.com	Beverly Miller	0.001211426693265
29	19807	alma.carrillo@enron.com		0.001136626725753
29	53722	angela.white@enron.com	Angela White	0.001117578736012
29	7005	jane.joyce@enron.com	Jane Joyce	0.000894385936708
29	24885	ava.garcia@enron.com		0.000878997392153
29	54567	renee.perry@enron.com	Renee Perry	0.000839418723067
29	19826	kim.perez@enron.com		0.000815505379513
29	54138	jerry.wilkens@enron.com		0.000809453965912
29	18192	sandy.sheffield@enron.com		0.000807162895742
29	42059	alicia.lenderman@enron.com		0.000782401198273
29	41978	dan.bunch@enron.com		0.000753560473298
29	24804	kelly.allen@enron.com		0.000708985549170
29	35481	randy.bryan@enron.com	Randy Bryan	0.000661569025376
29	18640	valerie.giles@enron.com	Valerie Giles	0.000656297229909
29	23680	rosemary.gracey@enron.com		0.000634957388910
29	43983	joni.bollinger@enron.com		0.000626956792132
29	24829	sharon.brown@enron.com		0.000616777142454
29	43959	jenerso@ei.enron.com		0.000612136420482

CATEGORY 30

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3984 COMPONENTS: 6
 LARGEST COMPONENT SIZE: 3971 PERCENT OF TOTAL GRAPH: 99.67%
 GROUP DEGREE: 0.05718 GRAPH DENSITY: 0.00100
 GROUP CLOSENESS: 0.00554 GROUP BETWEENNESS: 0.10933
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
30	15949	alexandra.saler@enron.com		0.001095899895894
30	6933	gabriel.monroy@enron.com	Gabriel Monroy	0.001034703115733
30	3015	ted.evans@enron.com	Ted Evans	0.000948143362125
30	5645	wilson.kriegel@enron.com		0.000859855612774
30	29138	douglas.nichols@enron.com		0.000851231371145
30	61637	charlotte.kraham@enron.com		0.000838904049729
30	22344	becky.pitre@enron.com		0.000815326423115
30	18956	cecilia.rodriguez@enron.com	Cecilia Rodriguez	0.000796196467072

```

30 10757 alma.martinez@enron.com..... 0.000790564297686
30 2197 ted.noble@enron.com..... Ted Noble..... 0.000768644742282
30 30824 joanne.smith@enron.com..... JoAnne Smith..... 0.000745825496676
30 66817 thomas.'paul@enron.com..... 0.000713062366461
30 11322 elizabeth.peters@enron.com..... 0.000711554608652
30 22 crystal.hyde@enron.com..... Crystal Hyde..... 0.000686471982273
30 44299 lnemec@ect.enron.com..... "Lisa Nemec"..... 0.000681339667818
30 11168 tobin.carlson@enron.com..... 0.000671220769090
30 19602 dianne.swiber@enron.com..... 0.000667135943555
30 3643 alexandra.villarreal@enron.com..... Alexandra Villarreal 0.000658175910057
30 104 laura.wente@enron.com..... 0.000647290680664
30 2060 pete.heintzelman@enron.com..... Pete Heintzelman.... 0.000643179613152

```

CATEGORY 31

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 4177          COMPONENTS: 20
LARGEST COMPONENT SIZE: 4113 PERCENT OF TOTAL GRAPH: 98.47%
GROUP DEGREE: 0.06371   GRAPH DENSITY: 0.00120
GROUP CLOSENESS: 0.00055 GROUP BETWEENNESS: 0.08940
AVERAGE p(z|u): 0.02   STDEV p(z|u): 0.02

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
31	14993	eserver@enron.com.....	eserver@enron.com@EN	0.001535657667116
31	38660	esaibi@enron.com.....		0.001381493992136
31	9987	jderric@enron.com.....		0.001109771432530
31	46388	ed.cattigan@enron.com.....	Ed Cattigan.....	0.000979791192791
31	6156	j.harris@enron.com.....		0.000614701006350
31	1396	vicsandra.trujillo@enron.com.....	Vicsandra Trujillo..	0.000368189267180
31	58222	status_updates@enron.com.....		0.000354788031081
31	4638	james.derrick@enron.com.....		0.000304719850918
31	4636	expense.report@enron.com.....		0.000289673888247
31	41562	esager2@enron.com.....		0.000281996079695
31	76758	cms.router@enron.com.....	CMS Router.....	0.000268302778046
31	35486	h.fields@enron.com.....	Sharon H Fields....	0.000263009472195
31	69556	enron.general.announcements.enronxgate@e		0.000222602733185
31	41645	lena.kasbekar@enron.com.....		0.000218030097477
31	592	paul.schiavone@enron.com.....	Paul Schiavone.....	0.000201423346559
31	64778	'deberry@enron.com.....		0.000190172446330
31	5671	kelly.johnson@enron.com.....		0.000175098701755
31	24913	gary.hugo@enron.com.....		0.000172751462835
31	17913	deborah.heath@enron.com.....	Deborah Heath.....	0.000169617204329

31 29445 all.users@enron.com..... 0.000169571501677

CATEGORY 32

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 11931 COMPONENTS: 4
LARGEST COMPONENT SIZE: 11895 PERCENT OF TOTAL GRAPH: 99.70%
GROUP DEGREE: 0.07389 GRAPH DENSITY: 0.00059
GROUP CLOSENESS: 0.00079 GROUP BETWEENNESS: 0.08986
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Table with 4 columns: Topic#, ID#, Email Address, Name, p(z|u). Contains 25 rows of user data.

CATEGORY 33

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6300 COMPONENTS: 8
LARGEST COMPONENT SIZE: 6285 PERCENT OF TOTAL GRAPH: 99.76%
GROUP DEGREE: 0.07417 GRAPH DENSITY: 0.00095
GROUP CLOSENESS: 0.00432 GROUP BETWEENNESS: 0.13968
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
33	7514	craig.rickard@enron.com	Craig Rickard	0.001222342981691
33	18854	teresa.aguilera-peon@enron.com	Maria Teresa Aguilera	0.001161656534928
33	14946	cecil.stapley@enron.com		0.001158314626099
33	7528	t.robinson@enron.com	Richard T. Robinson	0.001055521301486
33	3079	r.conner@enron.com	Andrew R. Conner	0.001003383371760
33	18877	greg.bruch@enron.com	Greg Bruch	0.001002888020087
33	43984	loren.penkava@enron.com		0.000940013378977
33	11222	david.marye@enron.com		0.000893487419038
33	17270	tim.johanson@enron.com		0.000832301378341
33	18901	dirk.dimitry@enron.com	Dirk Dimitry	0.000766225906147
33	6221	larimore@enron.com		0.000702638068746
33	18923	elizabeth.hutchinson@enron.com	Elizabeth Hutchinson	0.000694783161391
33	17598	olivier.herbelot@enron.com	Olivier Herbelot	0.000646370839464
33	13263	robert.gerry@enron.com	Robert Gerry	0.000630807518148
33	1999	jaime.gualy@enron.com	Jaime Gualy	0.000605399939833
33	11254	ami.thakkar@enron.com		0.000602358283799
33	11207	danilo.juvane@enron.com		0.000589967045480
33	68620	harora@ect.enron.com	"harpreet"	0.000588929000308
33	41985	allen.cohrs@enron.com		0.000586888788662
33	4925	lou.potempa@enron.com		0.000583993987068

CATEGORY 34

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 8465 COMPONENTS: 14
 LARGEST COMPONENT SIZE: 8420 PERCENT OF TOTAL GRAPH: 99.47%
 GROUP DEGREE: 0.07955 GRAPH DENSITY: 0.00059
 GROUP CLOSENESS: 0.00071 GROUP BETWEENNESS: 0.15979
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
34	8722	home.owner@mailman.enron.com		0.000919970216830
34	77707	plucci@enron.com	Paul Lucci	0.000914890164452
34	8629	undisclosed.recipients@mailman.enron.com		0.000787836984093
34	33798	us.home.owner@mailman.enron.com		0.000555623544131
34	8799	event@mailman.enron.com		0.000466134097510
34	70956	home.refinace@mailman.enron.com		0.000450228711128
34	29515	postmaster@mailboy.enron.com		0.000424305594125
34	8749	user@mailman.enron.com		0.000407537199994

34	8779	valued.client@mailman.enron.com.....	0.000391976300601
34	62311	zone34@mailman.enron.com.....	0.000332889620982
34	2347	h.lewis@enron.com.....	Andrew H. Lewis.....	0.000286785135164
34	2453	undisclosed-recipients@enron.com.....	undisclosed-recipien	0.000281521579392
34	8654	help.3a12_@mailman.enron.com.....	0.000278870249796
34	29948	valuableclients@mailman.enron.com.....	0.000269589948415
34	33801	b3@mailman.enron.com.....	0.000261015534416
34	6734	calvin.lee@enron.com.....	Calvin Lee.....	0.000246829661997
34	29920	wwenger@enron.com.....	0.000246085990346
34	30158	ndisclosed.recipients@mailman.enron.com.	ndisclosed.Recipient	0.000243912617602
34	37149	union.credit@enron.com.....	Credit Union.....	0.000237150152396
34	8759	valued.home.owner@mailman.enron.com....	0.000232339513153

CATEGORY 35

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7568 COMPONENTS: 23
LARGEST COMPONENT SIZE: 7464 PERCENT OF TOTAL GRAPH: 98.63%
GROUP DEGREE: 0.16509 GRAPH DENSITY: 0.00079
GROUP CLOSENESS: 0.00024 GROUP BETWEENNESS: 0.23979
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
35	8780	kensey_subscriber@mailman.enron.com....	0.001506825392421
35	8594	kkeiser@enron.com.....	0.000891309311095
35	47904	dlassere@enron.com.....	0.000557234183126
35	57370	sshackl@ect.enron.com.....	"Sara Shackelton "	0.000510391836097
35	34417	alewis@ect.enron.com.....	Andrew Lewis.....	0.000494223752658
35	41564	sshackl@enron.com.....	0.000474014224776
35	62034	jwillia@enron.com.....	0.000446540030464
35	34375	alewis@enron.com.....	0.000446372437523
35	60089	shendri@ect.enron.com.....	scott hendrickson...	0.000436527685215
35	41521	brapp@enron.com.....	0.000375700502252
35	34474	the.daytrader@enron.com.....	0.000351200037885
35	16172	dbaughm@ect.enron.com.....	0.000317865806771
35	21140	david.mayeux@enron.com.....	0.000299770428530
35	29847	hot39d@mailman.enron.com.....	0.000296390397385
35	19928	pallen@enron.com.....	0.000294949309447
35	6355	amosconi@enron.com.....	0.000245879955333
35	56804	ect.security@enron.com.....	0.000229024856121
35	15719	liz.hillman@enron.com.....	0.000221079706631
35	47358	slandwe@enron.com.....	0.000220409623075

35 6360 shwab@enron.com..... 0.000206628310012

CATEGORY 36

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6758 COMPONENTS: 15
LARGEST COMPONENT SIZE: 6719 PERCENT OF TOTAL GRAPH: 99.42%
GROUP DEGREE: 0.12011 GRAPH DENSITY: 0.00089
GROUP CLOSENESS: 0.00086 GROUP BETWEENNESS: 0.19975
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
36	14664	phil.clifford@enron.com.....		0.001117969429071
36	3053	john.wilson@enron.com.....	John Wilson.....	0.001052278465631
36	14663	bill.briggs@enron.com.....		0.001045070013300
36	30596	sarah.wesner@enron.com.....		0.000732370764522
36	10238	mary.ruffer@enron.com.....		0.000715440982101
36	6842	don.schroeder@enron.com.....	Don Schroeder Jr....	0.000440779983176
36	41574	jarmogi@enron.com.....		0.000426096256999
36	617	greg.whiting@enron.com.....	Greg Whiting.....	0.000393924165010
36	15183	l.wilson@enron.com.....	John L. Wilson.....	0.000359504521779
36	18654	c.griffin@enron.com.....		0.000255931690006
36	41351	nymex.list@enron.com.....		0.000249573820242
36	15581	brian_hoskins@enron.com.....	"Brian Hoskins"....	0.000232340671915
36	37136	counsel.dave@enron.com.....	Assistant General Co	0.000225210149311
36	20070	pavel.zadorozhny@enron.com.....		0.000216111924819
36	68958	pmg@enron.com.....		0.000212073342463
36	6142	esop.america@enron.com.....		0.000210246004258
36	41581	bdavis@enron.com.....		0.000201070751783
36	24452	mailing.dl-ga-all_special@enron.com....	DL-GA-all_special ma	0.000181896550497
36	6228	evite@enron.com.....		0.000181855283811
36	84400	mike.harper.enronxgate@enron.com.....		0.000175735047509

CATEGORY 37

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6063 COMPONENTS: 29
LARGEST COMPONENT SIZE: 5991 PERCENT OF TOTAL GRAPH: 98.81%
GROUP DEGREE: 0.13825 GRAPH DENSITY: 0.00082
GROUP CLOSENESS: 0.00041 GROUP BETWEENNESS: 0.26972
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
37	61754	jschwie@ect.enron.com	JSCHWIE	0.000696842147218
37	71959	aa@mailman.enron.com		0.000693156788563
37	34651	philbi1@ect.enron.com		0.000620391215386
37	15381	ebass@ect.enron.com		0.000583114156419
37	62384	mwoodson@enron.com		0.000487794592246
37	24019	backroads.travel.update@mailman.enron.co		0.000472587110233
37	44859	mwhitt@enron.com		0.000467571333293
37	70877	list.subscriber@mailman.enron.com		0.000411887926942
37	70827	start.the.new.year.with.a.clean.slate@ma		0.000388682712186
37	70836	schlenker@mailman.enron.com		0.000366137384702
37	70797	barr@mailman.enron.com		0.000350957031855
37	70959	valuable.customer@mailman.enron.com		0.000316567669689
37	22613	tdonoho@enron.com		0.000300560372321
37	70910	subscriber@mailman.enron.com		0.000258586675413
37	34866	tkuyken@enron.com	tkuyken	0.000237788283081
37	29915	carter@mailman.enron.com		0.000227801729953
37	87104	hi@mailman.enron.com		0.000222889717125
37	8490	info@enron.com	Enron Consumer Affai	0.000216785852683
37	81740	plucci@ect.enron.com		0.000212485211463
37	70929	marrow@mailman.enron.com		0.000202602422447

CATEGORY 38

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 9483 COMPONENTS: 25
 LARGEST COMPONENT SIZE: 9382 PERCENT OF TOTAL GRAPH: 98.93%
 GROUP DEGREE: 0.14278 GRAPH DENSITY: 0.00063
 GROUP CLOSENESS: 0.00022 GROUP BETWEENNESS: 0.23984
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
38	22643	pmims@enron.com		0.001214757260154
38	22403	dfarmer@ect.enron.com		0.001190479703897
38	4493	lcampbel@enron.com		0.001150836312456
38	78538	kholst@ect.enron.com		0.001076378797115
38	79414	lblair@enron.com		0.001024856235163
38	72458	jtholt@ect.enron.com		0.000992002873818
38	85541	tgeacco@enron.com		0.000985890599981
38	36000	emclaug@enron.com		0.000958813419644

```

38 68835 jarnold@ect.enron.com..... 0.000935988817177
38 81740 plucci@ect.enron.com..... 0.000921286502272
38 60973 kruscit@ect.enron.com..... Kevin..... 0.000910087386993
38 74278 kward@ect.enron.com..... "Kim Ward"..... 0.000904572493622
38 77844 jreitme@enron.com..... 0.000874476844228
38 78508 kholst@enron.com..... 0.000868779182142
38 71045 mcuilla@ect.enron.com..... Martin Cuilla..... 0.000859396701212
38 36250 pmims@ect.enron.com..... 0.000844357143993
38 35021 plove@ect.enron.com..... 0.000837649877088
38 71324 tkuyken@ect.enron.com..... 0.000837299636707
38 34866 tkuyken@enron.com..... tkuyken..... 0.000831368009449
38 4310 ebass@enron.com..... 0.000830791850316

```

CATEGORY 39

EXPLICIT SOCIAL NETWORK STATISTICS

```

VERTICES: 7900          COMPONENTS: 23
LARGEST COMPONENT SIZE: 7839 PERCENT OF TOTAL GRAPH: 99.23%
GROUP DEGREE: 0.18034   GRAPH DENSITY: 0.00063
GROUP CLOSENESS: 0.00045 GROUP BETWEENNESS: 0.29982
AVERAGE p(z|u): 0.02   STDEV p(z|u): 0.01

```

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
39	41733	dperlin@enron.com.....	"dperlin@enron.com".	0.000903739067878
39	22614	pkeavey@enron.com.....		0.000789044685087
39	81838	tstaab@enron.com.....	Theresa Staab.....	0.000583248027857
39	48549	mike.hernandez@enron.com.....		0.000418949006629
39	26360	dgiron@enron.com.....		0.000398157756866
39	82421	jmoore@enron.com.....		0.000334568870084
39	69405	khyatt@enron.com.....		0.000295634334730
39	53412	siegel.avram@enron.com.....		0.000270611524906
39	53015	siegel'. 'avram@enron.com.....		0.000268987115314
39	36409	newport-news.com@mailman.enron.com.....		0.000268667343561
39	24830	eileen.buerkert@enron.com.....		0.000268428184913
39	17165	clickathome.mailout@enron.com.....		0.000260881875838
39	33916	todd.bowen@enron.com.....	Todd Bowen.....	0.000256691757385
39	31719	erisk@enron.com.....		0.000255897991206
39	68037	gary.kane@enron.com.....		0.000247546227294
39	12151	diana.peters@enron.com.....	Diana Peters.....	0.000225040004984
39	37735	kpresto@ect.enron.com.....		0.000221535967321
39	55204	home.dvd@mailman.enron.com.....		0.000214326914687
39	63803	tamayo@enron.com.....		0.000214326914687

39 24038 robert.pechar@enron.com..... Robert Pechar..... 0.000195555424072

CATEGORY 40

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3198 COMPONENTS: 13
LARGEST COMPONENT SIZE: 3151 PERCENT OF TOTAL GRAPH: 98.53%
GROUP DEGREE: 0.15501 GRAPH DENSITY: 0.00094
GROUP CLOSENESS: 0.00091 GROUP BETWEENNESS: 0.30925
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
40	37217	c_r_zander@enron.com.....		0.001362062064994
40	37114	wollam'.erik@enron.com.....		0.001352652606597
40	37115	chet.fenner@enron.com.....		0.001333974284042
40	37146	fenner.chet@enron.com.....	Chet Fenner.....	0.001324966830009
40	37122	wollam.erik@enron.com.....		0.001286542806865
40	37220	erwollam@enron.com.....		0.001217350755984
40	29102	chambers.john@enron.com.....	John Chambers.....	0.001192661493249
40	37151	knipe'.chad@enron.com.....	chad knipe.....	0.001191418349095
40	37157	constantine'.brian@enron.com.....		0.001141971905505
40	37152	corrier.brad@enron.com.....	Brad Corrier.....	0.001123786642376
40	37218	feder'.t@enron.com.....		0.001098026456433
40	37216	feder.t@enron.com.....		0.001086028832547
40	37147	knipe.chad@enron.com.....	chad knipe.....	0.001016546916833
40	37223	chet_fenner@enron.com.....		0.001015054490788
40	37159	mccomb.keith@enron.com.....	Keith McComb.....	0.000995266434670
40	61559	jarrod.haughn@enron.com.....		0.000858525615543
40	703	matthew.lenhart@enron.com.....	Matthew Lenhart.....	0.000802244258104
40	37161	mccomb.chris@enron.com.....	Chris McComb.....	0.000778106690150
40	14855	luis.mena@enron.com.....		0.000636448337936
40	14863	purvi.patel@enron.com.....		0.000619514335550

CATEGORY 41

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3755 COMPONENTS: 17
LARGEST COMPONENT SIZE: 3671 PERCENT OF TOTAL GRAPH: 97.76%
GROUP DEGREE: 0.05087 GRAPH DENSITY: 0.00107
GROUP CLOSENESS: 0.00045 GROUP BETWEENNESS: 0.11925
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
41	73730	mariachi.el@enron.com	El Mariachi	0.001367308136551
41	15748	plove@enron.com		0.000762304310293
41	15729	gary.w.lamphier@enron.com		0.000431011007515
41	71023	wkaseme@enron.com		0.000396938966542
41	16893	carter.ellis@enron.com	Carter Ellis	0.000358006784754
41	2561	reagan.mathews@enron.com	Reagan Mathews	0.000348886178625
41	1140	joe.stepenovitch@enron.com	Joe Stepenovitch	0.000303042576406
41	1129	andy.pace@enron.com	Andy Pace	0.000249534277156
41	1952	valerie.ramsower@enron.com	Valerie Ramsower	0.000245887393921
41	22519	mike.morris@enron.com		0.000229771976309
41	15722	william.kasemervisz@enron.com		0.000214208468158
41	1881	greg.martin@enron.com	Greg Martin	0.000212687621422
41	11111	michael.morris@enron.com	Michael Morris	0.000204999101831
41	52720	'.'andre@enron.com	e-mail	0.000204009703757
41	61577	dl-erc@enron.com	DL-ERC	0.000191261744055
41	38006	servello.anthony@enron.com		0.000188452133106
41	15516	o'neal.winfree@enron.com		0.000185441179200
41	53731	shannon.ed@enron.com	Ed Shannon	0.000184879075621
41	52729	kenny'.'denis@enron.com		0.000176946199448
41	2515	martin.cuilla@enron.com	Martin Cuilla	0.000176037943189

CATEGORY 42

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 7411 COMPONENTS: 11
 LARGEST COMPONENT SIZE: 7389 PERCENT OF TOTAL GRAPH: 99.70%
 GROUP DEGREE: 0.15184 GRAPH DENSITY: 0.00081
 GROUP CLOSENESS: 0.00263 GROUP BETWEENNESS: 0.30977
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
42	24810	rita.bahner@enron.com		0.000805608717791
42	9021	richard.babin@enron.com		0.000797487326104
42	2990	warrick.franklin@enron.com	Warrick Franklin	0.000732665859683
42	16301	.marc@enron.com	e-mail	0.000732426166615
42	15530	.mom@enron.com	e-mail	0.000719260565466
42	8838	r..lilly@enron.com	Kyle R. Lilly	0.000713713041196
42	81467	.genia@enron.com	e-mail	0.000713374066709
42	3788	.erik@enron.com	e-mail	0.000707024570697

42 12449 dave.lawlor@enron.com..... 0.000670644318918
42 9044 susan.bulgawicz@enron.com..... 0.000650229323052
42 12447 jim.roth@enron.com..... Jim Roth..... 0.000650060151848
42 51739 antoine.duvauchelle@enron.com..... 0.000643280571798
42 16082 thiem'. 'matt@enron.com..... Matt Thiem..... 0.000629237284975
42 61565 'hotmail.com@enron.com..... 0.000627229037215
42 12448 dan.botsch@enron.com..... 0.000615150204166
42 16087 mudd'. 'lisa@enron.com..... Lisa Mudd..... 0.000610940473321
42 37659 'tconl.com@enron.com..... 0.000590838897926
42 19277 benjamin.freeman@enron.com..... 0.000567691789362
42 85338 .mordente@enron.com..... e-mail..... 0.000561229286806
42 41104 allison.healy-poe@enron.com..... Allison Healy-Poe... 0.000544563372251

CATEGORY 43

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 6715 COMPONENTS: 17
LARGEST COMPONENT SIZE: 6655 PERCENT OF TOTAL GRAPH: 99.11%
GROUP DEGREE: 0.09178 GRAPH DENSITY: 0.00089
GROUP CLOSENESS: 0.00048 GROUP BETWEENNESS: 0.14972
AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
43	44150	jim.peterson@enron.com.....		0.000249575678766
43	2453	undisclosed-recipients@enron.com.....	undisclosed-recipien	0.000225172162911
43	76756	anngtc.dl-ets@enron.com.....	DL-ETS ANNGTC.....	0.000181271087772
43	55936	steven.p.south@enron.com.....	Steve South.....	0.000180816666327
43	85	ebiz@enron.com.....	eBiz.....	0.000167467120787
43	62868	bruce.bowden@enron.com.....		0.000165167437456
43	78819	richard.gilliland@enron.com.....	Richard Gilliland...	0.000160509581501
43	35833	colin.poon@enron.com.....	Colin Poon Tip.....	0.000156107235208
43	41550	anne.c.koehler@enron.com.....		0.000139679454219
43	55710	itrezzo.agent@enron.com.....	Itrezzo Agent.....	0.000136195817496
43	81356	sheryl'. 'gussett@enron.com.....		0.000135645126650
43	6036	maritta.mullet@enron.com.....		0.000131422272010
43	20326	katrina.chapman@enron.com.....		0.000129431166380
43	41075	daemon.extra@enron.com.....	EXTRA Mailer Daemon.	0.000124903059073
43	19898	corey.hollander@enron.com.....		0.000124859520390
43	26463	boudreaux.john@enron.com.....	John Boudreaux.....	0.000123440754475
43	26414	newsome.linda@enron.com.....		0.000120691890243
43	799	dl-ga-all_domestic@enron.com.....	DL-GA-all_domestic..	0.000119694054478
43	5242	jay.patel@enron.com.....		0.000119494829467

43 11176 milagros.daetz@enron.com..... 0.000113096329756

CATEGORY 44

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 4757 COMPONENTS: 6
LARGEST COMPONENT SIZE: 4745 PERCENT OF TOTAL GRAPH: 99.75%
GROUP DEGREE: 0.27845 GRAPH DENSITY: 0.00105
GROUP CLOSENESS: 0.00647 GROUP BETWEENNESS: 0.45969
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.03

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
44	6031	outlook.team@enron.com.....		0.001457639238429
44	4987	billy.dorsey@enron.com.....		0.001448022707578
44	1285	suzanne.danz@enron.com.....	Suzanne Danz.....	0.001390803746101
44	38627	mcmahon@enron.com.....		0.001075689070940
44	215	stacey.white@enron.com.....		0.001035505396680
44	4116	katherine.brown@enron.com.....		0.001001321457678
44	12042	georgene.moore@enron.com.....	Georgene Moore.....	0.000704291034546
44	17437	elaine.overturf@enron.com.....		0.000633403923622
44	7112	zionette.vincent@enron.com.....	Zionette Vincent....	0.000605428407510
44	4963	hilda.bourgeois-galloway@enron.com.....		0.000595142103078
44	5484	vanessa.groscrand@enron.com.....		0.000571125577302
44	6097	simone.rose@enron.com.....		0.000559112815457
44	10552	dan.bruce@enron.com.....		0.000524759469048
44	34175	kathryn.thomas@enron.com.....	Kathryn Thomas.....	0.000514232504702
44	1244	barbara.hooks@enron.com.....	Barbara Hooks.....	0.000506891760699
44	4063	sherri.reinartz@enron.com.....		0.000504281954880
44	57673	legal.7@enron.com.....		0.000466587322574
44	5060	jana.paxton@enron.com.....		0.000466130256002
44	18048	rick.carson@enron.com.....		0.000450157316280
44	35725	daniel.lyons@enron.com.....	Daniel Lyons.....	0.000423396381388

CATEGORY 45

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 3351 COMPONENTS: 6
LARGEST COMPONENT SIZE: 3340 PERCENT OF TOTAL GRAPH: 99.67%
GROUP DEGREE: 0.06876 GRAPH DENSITY: 0.00090
GROUP CLOSENESS: 0.00745 GROUP BETWEENNESS: 0.14918
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.02

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
45	1181	exchange.administrator@enron.com		0.001517057163547
45	3430	system.administrator@enron.com		0.001380664474799
45	642	eric.bass@enron.com	Eric Bass	0.000836370999018
45	29154	postmaster@enron.com		0.000829706720554
45	8776	thomas.underwood@enron.com	Thomas Underwood	0.000782691820933
45	81667	brewer@enron.com		0.000670754388186
45	66671	moscoso@enron.com		0.000593532000242
45	1383	nick.hiemstra@enron.com	Nick Hiemstra	0.000566085668242
45	65104	meredith@enron.com	"	0.000454517520423
45	7122	rafael.avila@enron.com	Rafael Avila	0.000414596161193
45	11353	mark.morrow@enron.com		0.000409627634599
45	19276	nicholas.stephan@enron.com		0.000362676763839
45	46356	aaron.klemm@enron.com		0.000359370679031
45	5727	rob.gay@enron.com		0.000351821963634
45	16538	.russell@enron.com	e-mail	0.000348453239846
45	2614	christa.winfrey@enron.com	Christa Winfrey	0.000333860854958
45	1878	bryan.hull@enron.com	Bryan Hull	0.000330947117717
45	24005	suzanne.russell@enron.com	Suzanne Russell	0.000330474580949
45	501	misti.day@enron.com	Misti Day	0.000321193293553
45	9128	randall.gay@enron.com		0.000316053128149

CATEGORY 46

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 2525 COMPONENTS: 18
 LARGEST COMPONENT SIZE: 2483 PERCENT OF TOTAL GRAPH: 98.34%
 GROUP DEGREE: 0.13137 GRAPH DENSITY: 0.00238
 GROUP CLOSENESS: 0.00112 GROUP BETWEENNESS: 0.22905
 AVERAGE p(z|u): 0.02 STDEV p(z|u): 0.04

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	p(z u)
46	22503	carlos.j.rodriguez@enron.com	carlos.j.rodriguez	0.001361925244531
46	22502	gary.a.hanks@enron.com	gary.a.hanks	0.001358652912680
46	14685	kate.fraser@enron.com	Kate Fraser	0.001250193273273
46	6068	joe.casas@enron.com		0.001169417430292
46	14755	cindy.vachuska@enron.com	Cindy Vachuska	0.001142188670334
46	18724	kevin.alvarado@enron.com	Kevin Alvarado	0.001130869990743
46	63604	lhs-gas.kvammen@enron.com	Kjell - LHS-GAS Kvam	0.001123994126733
46	6070	scott.loving@enron.com		0.001108816235622

46	6578	briant.baker@enron.com.....	Briant Baker.....	0.001088074634254
46	15567	earl.tisdale@enron.com.....	0.001017305226123
46	20267	sabra.dinari@enron.com.....	0.001016906056959
46	22831	stone.charlie@enron.com.....	0.001015577070221
46	63612	lhc-gas.kvammen@enron.com.....	Kjell - LHC-GAS Kvam	0.001000535662754
46	1422	jessie.patterson@enron.com.....	Jessie Patterson....	0.000863189169090
46	11081	rebecca.griffin@enron.com.....	Rebecca Griffin....	0.000848303716935
46	6067	alvin.thompson@enron.com.....	0.000846458822023
46	2999	margie.straight@enron.com.....	Margie Straight....	0.000842621902831
46	22832	avila.david@enron.com.....	0.000786294011057
46	20241	susan.hadix@enron.com.....	0.000767909391233
46	63316	hfs.reite@enron.com.....	NILS - B. Superinten	0.000754027412323

CATEGORY 47

EXPLICIT SOCIAL NETWORK STATISTICS

VERTICES: 5813 COMPONENTS: 14
LARGEST COMPONENT SIZE: 5772 PERCENT OF TOTAL GRAPH: 99.29%
GROUP DEGREE: 0.11283 GRAPH DENSITY: 0.00069
GROUP CLOSENESS: 0.00087 GROUP BETWEENNESS: 0.22968
AVERAGE $p(z|u)$: 0.02 STDEV $p(z|u)$: 0.01

MOST PROBABLE USERS

Topic#	ID#	Email Address	Name	$p(z u)$
47	2629	bbutler2@enron.com.....	0.001256182039730
47	70	mark.brand@enron.com.....	Mark Brand.....	0.001010339147226
47	1115	clint.dean@enron.com.....	Clint Dean.....	0.000847763531732
47	44193	l.foust@enron.com.....	0.000531142021457
47	11221	aaron.martinsen@enron.com.....	0.000530399196322
47	53620	culbertson.david@enron.com.....	David Culbertson....	0.000505657927192
47	26363	giron.kristi@enron.com.....	0.000499362665570
47	37391	.judy@enron.com.....	e-mail.....	0.000447030997618
47	7076	zachary.mccarroll@enron.com.....	Zachary McCarroll...	0.000416268350065
47	1881	greg.martin@enron.com.....	Greg Martin.....	0.000408819462068
47	11289	sarah.goodpastor@enron.com.....	0.000393161899312
47	38011	wiengerski.dan@enron.com.....	0.000380934354936
47	83584	adam.giannone@enron.com.....	Adam Giannone.....	0.000368199021783
47	1129	andy.pace@enron.com.....	Andy Pace.....	0.000354615277351
47	7614	michael.simmons@enron.com.....	Michael Simmons....	0.000353895939361
47	3426	mswerzb@ect.enron.com.....	0.000344006437805
47	41170	enron.above@enron.com.....	0.000328388216605
47	2389	robert.vargas@enron.com.....	Robert Vargas.....	0.000320739204939
47	53621	duke.kyle@enron.com.....	Kyle Duke.....	0.000320082224495

47 38027 'kevin'@enron.com..... Kevin..... 0.000317172268516

Bibliography

1. Adamic, L. and E. Adar. “Friends and Neighbors on the Web”, 2001. URL citeseer.ist.psu.edu/article/adamic01friends.html.
2. Adibi, Jafar and Jitesh Shetty. “Enron Ex-Employee Status Report”. Retrieved 15-Nov-2005 from http://www.isi.edu/~adibi/Enron/Enron_Employee_Status.xls.
3. Air War College, . “Military Intelligence, Intelligence Operations, National Intelligence, Intelligence Analysis, Gateway to Intelligence”. Air War College, Air University. Retrieved on 15 June, 2005 from <http://www.au.af.mil/au/awc/awcgate/awc-ntel.htm#types>.
4. ASIS Foundation, PriceWaterhouse Coopers, and U.S. Chamber of Commerce (editors). *Trends in Proprietary Information Loss*. ASIS Foundation, September 2002.
5. Baeza-Yates, Ricardo and Berthier Ribeiro-Neto. *Modern Information Retrieval*. ACM Press Books, New York, NY, 1999.
6. Bekkerman, Ron, Andrew McCallum, and Gary Huang. *Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora*. Technical Report UMass CIIR Technical Report IR-418, University of Massachusetts Amherst, 2005.
7. Berkhin, Pavel. *Survey Of Clustering Data Mining Techniques*. Technical report, Accrue Software, San Jose, CA, 2002. URL citeseer.ist.psu.edu/berkhin02survey.html.
8. Berry, Michael W. and Murray Brown. “Email Surveillance Using Nonnegative Matrix Factorization”. *Proc of Workshop of Link Analysis, Counterterrorism and Security at SIAM International Conference on Data Mining 2005*.
9. Blei, D.M. “C Implementation of Latent Dirichlet Allocation”. Retrieved 1-Jun-2005 from <http://www.cs.berkeley.edu/~blei/lda-c/>.
10. Blei, D.M., A.Y. Ng, and M.I. Jordan. “Latent Dirichlet Allocation”. *Journal of Machine Learning*, 2(5):993–1022, 2003.
11. Brandes, Ulrik. “A Faster Algorithm for Betweenness Centrality”. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
12. Burke, Stephanie. “The Collapse of Enron: A Bibliography of Online Legal, Government, and Legislative Resources”. Retrieved 15-Nov-2005 from <http://www.llrx.com/features/enron.htm#enron>.
13. Carenini, Giuseppe, Raymond Ng, and Xiaodong Zhou. “Scalable Discovery of Hidden Emails from Large Folders”. *Proc. 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005)*. 2005.

14. Carnegie Mellon University, . “Enron Email Dataset”, 30 June 2005. URL <http://www-2.cs.cmu.edu/enron>.
15. Carr, Rebecca. “In Federal Job: Blow Whistle, Get Boot”, 2005. Retrieved 25-Jan-2006 from <http://www.statesman.com/news/content/news/stories/nation/12/11whistleblower.html>.
16. Cohen, William. “Learning Rules that Classify E-Mail”. *Proceedings of AAAI Spring Symposium on Machine Learning and Information Retrieval, 1996*. 1996.
17. Cohn, David and Huan Chang. “Learning to Probabilistically Identify Authoritative Documents”. *Proc. 17th International Conf. on Machine Learning*, 167–174. Morgan Kaufmann, San Francisco, CA, 2000. URL citeseer.ist.psu.edu/cohn00learning.html.
18. CSO magazine, U.S. Secret Service, and CERT Coordination Center (editors). *2004 eCrime Watch Survey*. CSO Magazine, May 2004.
19. Culotta, Aron, Ron Bekkerman, and Andrew McCallum. “Extracting social networks and contact information from email and the Web”. *First Conference on Email and Anti-Spam (CEAS)*. Mountain View, CA, 2004. URL <http://www.cs.umass.edu/~culotta/pubs/ceas04.pdf>.
20. DEC Systems Research Center, . “DEC-SRC-92-04-05 word list”. Retrieved 13-Aug-2005 from <ftp://nic.funet.fi/pub/unix/security/dictionaries/DEC-collection>.
21. Diesner, Jana and Kathleen Carley. “Exploration of Communication Networks from the Enron Email Corpus”. *Proc of Workshop of Link Analysis, Counterterrorism and Security at SIAM International Conference on Data Mining 2005*.
22. DoD Insider Threat Integrated Process Team, . *DoD Insider Threat Mitigation: Final Report of Insider Threat Integrated Process Team*. Department of Defense, April 2000. Retrieved 30-May-2005 from https://acc.dau.mil/simplify/ev.php?ID=64069_201&ID2=DO_TOPIC.
23. Eichenwald, Kurt. *Conspiracy of Fools: a True Story*. L Rienner Publishers, Boulder, CO, 2002.
24. Electronic Privacy Information Center, . “Workplace Privacy”. Retrieved 26-Jan-2006 from <http://www.epic.org/privacy/workplace/>.
25. Fels, Allan. “Whistleblowing: Opportunity or Threat”, 2003. Retrieved 25-Jan-2006 from www.transparency.org.au/wb/afspeech310703.pdf.
26. Fiore, Andrew and Jeff Heer. “UC Berkeley Enron Email Analysis”, 30 June 2005. URL http://bailando.sims.berkeley.edu/enron_email.html.
27. Fischer, Lynn. “Espionage: Why Does It Happen”. DoD Security Institute, Security Awareness Bulletin 1-94, 1994.

28. Friedman, Nir and Moises Goldszmidt. *Learning in Graphical Methods*, chapter Learning Bayesian Networks with Local Structure. Kluwer Academic Publishers, 1998.
29. Gelles, Dr. Mike. “Exploring the Mind of a Spy”. Defense Personnel Security Research Center (PERSEREC), 2001. Retrieved on 15 June, 2005 from [http://www.dss.mil/training/csg/security/Treason/Mind.htm#Exploring The Mind](http://www.dss.mil/training/csg/security/Treason/Mind.htm#Exploring%20The%20Mind).
30. Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/ CRC, Boca Raton, FL.
31. Giles, Lionel. *Sun Tzu On The Art of War: The Oldest Military Treatise in the World*. Project Gutenberg, 1910. Retrieved on 15 June from <http://www.au.af.mil/au/awc/awcgate/artofwar.htm#13>.
32. Girolami, M. and A. Kaban. “On an equivalence between PLSI and LDA”, 2003. URL citeseer.ist.psu.edu/girolami03equivalence.html.
33. Griffiths, Thomas L. and Mark Steyvers. “Finding Scientific Topics”. *PNAS Colloquium Papers*, volume 101-1. April 2004.
34. Hanneman, Robert. “Introduction to Social Network Methods”. Online Textbook, University of California, Riverside, 2001. Retrieved 30-May-2005 from.
35. Herbig, Katherine L. and Martin F. Wiskoff. *Espionage Against the United States by American Citizens 1947 - 2001*. Technical report, Defense Personnel Security Research Center (PERSEREC), 2002. Retrieved on 15 June, 2005 from <http://www.fas.org/sgp/library/spies.pdf>.
36. Heuer, Richards and Katherine Herbig. “The InsiderEspionageThreat”. Defense Personnel Security Research Center (PERSEREC), 2001. Retrieved on 15 June, 2005 from [http://www.dss.mil/training/csg/security/Treason/Insider.htm#The Insider](http://www.dss.mil/training/csg/security/Treason/Insider.htm#The%20Insider).
37. Hoffman, Thomas. “Probabilistic Latent Semantic Indexing”. *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*. 1999.
38. Houston Chronicle, . “Special Report: Enron”. Retrieved 15-Nov-2005 from <http://www.chron.com/content/chronicle/special/01/enron/index.html>.
39. Huang, Yifen, Dinesh Govindaraju, Tom Mitchell, Vitor Rocha de Carvalho, and William Cohen. “Inferring Ongoing Activities of Workstation Users by Clustering Emails”. *First Conference on Email and Anti-Spam (CEAS)*. Mountain View, CA, 2004.
40. Johnson, Roberta Ann. *Whistleblowing: When it Works and Why*. Broadway Books, New York, NY, 2005.

41. Jordan, Michael, Zoubin Ghahramani, Tommi Jaakkola, and Lawrence Saul. *Learning in Graphical Methods*, chapter An Introduction to Variational Methods for Graphical Models. Kluwer Academic Publishers, 1998.
42. Kalman, Yoram and Sheizaf Rafraeli. “Email Chronemics: Unobtrusive Profiling of Response Times”. *Proceedings of the 38th Hawaii International Conference on Systems Sciences - 2005*. 2005.
43. Kautz, Henry, Bart Selman, and Mehul Shah. “Referral Web: Combining Social Networks and Collaborative Filtering”. *Communications of the ACM*, 40(3):63–65, 1997. URL citeseer.ist.psu.edu/kautz97referralweb.html.
44. Keila, Parambir and David Skillicorn. *Detecting Unusual and Deceptive Communication in Email*. Technical report, Queen’s University, Kingston, Ontario, Canada, June 2005.
45. Keila, Parambir and David Skillicorn. “Structure in the Enron Email Dataset”. *Proc of Workshop of Link Analysis, Counterterrorism and Security at SIAM International Conference on Data Mining 2005*. 2005.
46. Keller, M. and S. Bengio. “Theme Topic Mixture Model: A Graphical Model for Document Representation”. *PASCAL Workshop on Learning Methods for Text Understanding and Mining*. 2004.
47. Kipp, Steven. “Espionage and the Insider”. SecurityDocs.com, April 2004. Retrieved 30-May-2005 from <http://www.securitydocs.com/library/2032>.
48. Klimt, Bryan and Yiming Yang. *Lecture Notes in Computer Science*, chapter The Enron Corpus: A new Dataset for Email Classification Research. Springer-Verlag GmbH, 2004.
49. Li. “Random Texts Exhibit Zipf’s Law-Like Word Frequency Distribution”. *IEEE Transactions on Information Theory*, 38, 1992. URL citeseer.ist.psu.edu/li92random.html.
50. Liben-Nowell, David and Jon Kleinberg. “The Link Prediction Problem for Social Networks”. *Proceeding of the Twelfth Annual ACM International Conference on Information and Knowledge Management (CIKM’03)*, 556–559. November 2003.
51. Madsen, Rasmus, David Kauchak, and Charles Elkan. “Modeling Word Burstiness Using the Dirichlet Distribution”. *Proceedings of the 22nd International Conference on Machine Learning*. 2005.
52. Martin, Steve, Anil Sewani, Blaine Nelson, Karl Chen, and Anthony Joseph. “Analyzing Behavioral Features for Email Classification”. *Second Conference on Email and Anti-Spam (CEAS)*. Mountain View, CA, 2005.
53. McCallum, Andrew, Andres Corrada-Emmanuel, and Xuerui Wang. “Topic and Role Discovery in Social Networks”. *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*.

54. McLean, Bethany and Peter Elkind. *The Smartest Guys in the Room*. Penguin Group (USA), New York, NY, 2003.
55. Merriam-Webster Collegiate Dictionary, . “Espionage”, 10 May 2005. URL <http://www.m-w.com/cgi-bin/dictionary>.
56. Micelli, Marcia and Janet Near. *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees*. Lexington Books, New York, NY, 1992.
57. Minka, Thomas and John Lafferty. “Expectation-Propagation for the Generative Aspect Model”. *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*. 2002.
58. Mitre. “Detecting Insider Threat Behavior”, 2004. Retrieved 25-Jan-2006 from www.mitre-corporation.com/news/events/tech04/briefings/1344.pdf.
59. Moon, Todd. “The Expectation-Maximization Algorithm”. *IEEE Signal Processing Magazine*, 47–60, November 1996.
60. Mooney, Raymond J. “Intelligent Information Retrieval and Web Search”. Retrieved 10-Aug-2005 from <http://www.cs.utexas.edu/users/mooney/ir-course/slides/TextProperties.ppt>.
61. University of Neuchatel, . “CLEF and Multilingual Information Retrieval”. Retrieved 16-Jul-2005 from <http://www.unine.ch/info/clef/>.
62. News, FindLaw Legal and . Commentary. “Enron Investigation News”. Retrieved 15-Nov-2005 from <http://news.findlaw.com/legalnews/lit/enron/>.
63. O'Brien, C. and C. Vogel. “Exploring the Subject of Email Filtering”. submitted for review, 2004.
64. Porter, Martin. “The Porter Stemming Algorithm”. Retrieved 16-Jul-2005 from <http://www.tartarus.org/~martin/PorterStemmer>.
65. Project, Government Accountability. “The Hanford Pipefitters Story”, 2005. Retrieved 25-Jan-2006 from http://www.whistleblower.org/template/page.cfm?page_id=134.
66. RAND. “Research and Development Initiatives Focused on Preventing, Detecting, and Responding to Insider Misuse of Critical Defense Information Systems”, 1999. Retrieved 24-Aug-2005 from <http://www.rand.org/publications/CF/CF151/CF151.pdf>.
67. RAND. “Understanding the Insider Threat: Proceedings of a March 2004 Workshop”, 2004. Retrieved 25-Jan-2006 from www.rand.org/pubs/conf_proceedings/2005/RAND_CF196.pdf.
68. Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. “The Author-Topic Model for Authors and Documents”. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 487–494. 2004.

69. Russell, Stuart J. and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2003.
70. Sahami, M., S. Dumais, D. Heckerman, and E. Horvitz. "A Bayesian Approach to Filtering Junk E-mail". *Proceedings of Workshop on Learning for Text Categorization*. 1998.
71. Shaw, Eric, Jerrold Post, and Keven Ruby. "Inside the Mind of the Insider". *Security Management*, 43:34–41.
72. Shetty, Jitesh and Jafar Adibi. "The Enron Dataset Schema and Brief Statistical Report". Retrieved 30-Jul-2005 from http://www.isi.edu/~adibi/Enron/Enron_Dataset_Report.pdf.
73. Skillicorn, D.B. "Beyond Keyword Filtering for message and conversation detection". *Lecture Notes in Computer Science*, 3495:231–243, 2005.
74. Stolfo, Salvatore, Shlomo Hershkop, Ke Wang, Olivier Nimeskern, and Chia-Wei Hu. "A Behavior-based approach to Securing Email Systems". 2003.
75. Symonenko, Svetlana, Elizabeth D. libby, Ozgur Yilmazel, Robert Del Zoppo, Eric Brown, and Matt Downey. "Semantic Analysis for Monitoring Insider Threats". *Second Symposium on Intelligence and Security Informatics (ISI 2004)*. 2004.
76. Teh, Yee Whye, Michael I. Jordan, Matthew J. Beal, and David M. Blei. "Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes". Lawrence K. Saul, Yair Weiss, and Léon Bottou (editors), *Advances in Neural Information Processing Systems 17*, 1385–1392. MIT Press, Cambridge, MA, 2005.
77. Verton, Dan. *The Insider: A True Story*. Llumina Press, Coral Springs, FL, 2005.
78. Wasserman, Stanley and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, NY, 1994.
79. Yee, Jason Wei Sung. *Efficient Generation of Social Network Data from Computer-Mediated Communication Logs*. Master's thesis, Air Force Institute of Technology, 2005.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 074-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 23-03-2006		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From - To) Aug 2004-Mar 2006	
4. TITLE AND SUBTITLE Detecting Potential Insider Threats Through Email Datamining				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Okolica, James S., First Lieutenant, USAF				5d. PROJECT NUMBER If funded, enter ENR #	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GCS/ENG/06-10	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Despite a technology bias that focuses on external electronic threats, insiders pose the greatest threat to commercial and government organizations. One means of preventing insider theft is by stopping potential insiders from actually crossing the line. In the overwhelming number of cases, people do not join an organization with the intention of stealing or causing harm. Instead, something or often several some things happen while the individual is in the organization that precedes his malevolent actions. One of the traits identified with insiders is their feeling of alienation from the organization. By data mining emails, an employee's interests can be discerned. These interests are then used to construct social networks which are used to identify individuals with interests shared but undiscussed with other members of the organization. These individuals with clandestine interests have the potential to be insider threats. This paper describes the use of Probabilistic Latent Semantic Indexing (PLSI) extended to include users (PLSI-U) and Author Topic extended to include documents to determine topics of interest for employees from their email activity. It then applies PLSI-U and Author Topic to the Enron email corpus. The results show that by comparing the topics of emails that people send internally with the ones sent externally, a small number of employees (0.03%-1.0%) emerge as having clandestine interests and the potential to become insider threats. Most significantly, one of these individuals is Sherron Watkins, the famous whistleblower in the Enron case.					
15. SUBJECT TERMS Insider, Threat, Data Mining, Clustering, Social Networks					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 324	19a. NAME OF RESPONSIBLE PERSON Gilbert L. Peterson, PhD (ENG)
REPORT U	ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (937) 255-6565 ext 4281; email: gilbert.peterson@afit.edu