

## Research Article

# Discovering Organizational Hierarchy through a Corporate Ranking Algorithm: The Enron Case

Germán G. Creamer <sup>1</sup>, Salvatore J. Stolfo <sup>2</sup>, Mateo Creamer <sup>3</sup>, Shlomo Hershkop <sup>4</sup>,  
and Ryan Rowe <sup>5</sup>

<sup>1</sup>Stevens Institute of Technology, Hoboken, NJ 07030, USA

<sup>2</sup>Department of Computer Science, Columbia University, New York, NY 10027, USA

<sup>3</sup>Stanford Graduate School of Business, Palo Alto, CA 94305, USA

<sup>4</sup>Allure Security, Waltham, MA 02451, USA

<sup>5</sup>Department of Applied Mathematics, Columbia University, New York, NY 10027, USA

Correspondence should be addressed to Germán G. Creamer; [gcreamers@stevens.edu](mailto:gcreamers@stevens.edu)

Received 25 August 2021; Revised 7 December 2021; Accepted 5 January 2022; Published 21 February 2022

Academic Editor: Ning Cai

Copyright © 2022 Germán G. Creamer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes the CorpRank algorithm to extract social hierarchies from electronic communication data. The algorithm computes a ranking score for each user as a weighted combination of the number of emails, the number of responses, average response time, clique scores, and several degree and centrality measures. The algorithm uses principal component analysis to calculate the weights of the features. This score ranks users according to their importance, and its output is used to reconstruct an organization chart. We illustrate the algorithm over real-world data using the Enron corporation's e-mail archive. Compared to the actual corporate work chart, compensation lists, judicial proceedings, and analyzing the major players involved, the results show promise.

## 1. Introduction

A significant challenge in any organization is identifying the underlying organizational structure that might be different from the official version. A simple way to approximate this structure and the corporate hierarchy is based on the pattern of communication between its members. Competitors or clients might use the disclosure of corporate conversations to reveal strategic information such as the development of new products or corporate finance decisions that may have a significant impact on the stock price. However, it is challenging to collect that information, especially considering corporate communications' high level of confidentiality. This paper proposes a method to approximate this organizational structure and hierarchy using a corporate e-mail dataset in an automated fashion. Several bankruptcy scandals in publicly held U.S. companies such as Enron offer a collection of electronic

communication records that could be used to identify its organizational structure.

The Enron Corporation's e-mail collection, described in Section 2, is a publicly available set of private corporate data released during the judicial proceedings against the Enron Corporation. Several researchers have explored it mainly from a natural language processing (NLP) perspective [1–5]. Social network analysis (SNA) examining structural features, Diesner and Carley [6], has also been applied to extract properties of the Enron network and attempts to detect the key players around the time of Enron's crisis: Diesner et al. [7] studied the patterns of communication of Enron employees differentiated by their hierarchical level; Cotterill [8] investigated a set of stylistic language features to predict organizational hierarchy relationships; Danescu-Niculescu-Mizil [9] presented an analysis framework on linguistic coordination to analyze power relationships in static and situational forms; Chundi et al. [1] conducted a time series

analysis of communication patterns from Enron's email; Priebe et al. [10] developed a theory of scan statistics on graphs for anomaly detection using time series; Berry and Browne [11] applied a non-negative matrix factorization to identify and monitor semantic features and message clusters; Keila and Skillicorn [12] interestingly enough, found that word use changed according to the functional position; Padmanabhan et al. [13] conducted a thread analysis to determine employees' responsiveness; Elsayed and Oard [14] presented a method for identity resolution in the Enron e-mail dataset; Bar-Yossef et al. [15] applied a cluster ranking algorithm based on the strength of the clusters to this dataset; Zhang et al. [16] modeled interactions between groups using the semantic content of Enron's email; Menges et al. [17] proposed an agent-based approach to model email-based social networks; Pathak and Srivastava [18] presented a technique to extract concealed relations from social networks; Chapanond et al. [19] used graph theoretical and spectral analysis to discover structures within Enron's organizations; and Shetty and Adibi [20] used an entropy model to identify the most relevant people in the organization.

The problem of organization structure, networks' properties, and hierarchical discovery has also been explored by several researchers without using Enron's database: Hu and Liu [21] reconstructed a multiplex network from an unstructured personal e-mail corpus to analyze social status and roles; Maiya and Berger-Wolf [22] proposed a method to infer a maximum likelihood social hierarchy from weighted social networks; Memon et al. [23] developed an algorithm to capture hidden hierarchies in terrorist networks; Kemp and Tenenbaum [24] implemented a method to discover hierarchies that are optimal for a given dataset; Freeman [25] adapted a method, a canonical analysis of asymmetry, to study hierarchies in organizations; Aral and Van Alstyne [26] analyzed the tradeoff between network diversity and communication channel bandwidth in the case of information diffusion.

Our previous paper was published as follows: Creamer et al. [27] proposed a method to calculate a social score that approximates the organizational hierarchy of Enron using the e-mail Mining Toolkit (EMT) project [28, 29]. Several authors have incorporated our work ([27]) in the area of social hierarchy detection as a basis for their research as follows: Agarwal et al. and Maktoubian et al. [30, 31] incorporated SNA with a simple lower bound instead of an upper bound for social network-based systems; Palus et al. [32] applied the resulting social scores from this research to distinguish levels of hierarchy in the Enron Corpus; Kalia et al. [33] adopted the mentioned social features of degree and betweenness centrality to identify organizational hierarchy, adding an emphasis on content, patterns, and emotions of messages; Nguyen and Zheng [34] referenced this research to assume that a person's influence score is positively correlated with his rank and predict influence spreads in Twitter communities. Others have studied the application of social networks to practical corporate improvement methods as follows: Hovelynck and Chidlovskii [35] adopted commonly used features of nodes to represent

key properties of actors in response to this work, assigning a social score for each node to improve classification performance; Li and Somayaji [36] applied SNA to organizational access control; Michalski et al. [37] matched social network hierarchies in organizations with a stable corporate structure to improve company management; Rivera-Pelayo [38] considered the application of data mining and SNA for its program, ExpertSN, allowing for effective people search in a given work context; Ganjaliyev [39] proposed a new method to identify network communities to enhance social network analysis; and Wang et al. [40] used HumanRank, a method of ranking individuals based on importance using personal electronic interactions.

A possible concern with this line of research is that corporations' emails or electronic communication records are difficult to obtain or that other electronic channels are substituting email systems. However, this line of research is becoming more important as several recent papers have extended our initial research [27] based on e-mail communications or enterprise social networks (ESN; i.e., intranet) to evaluate organizational structures using cluster validation [41]; to infer professional roles in an organization [42]; to extract the social structure and its hierarchy using the flow hierarchy derived from frequent interactions [43, 44]; to evaluate the impact of formal hierarchies on ESN [45, 46]; to classify power relations [47]; to classify texts incorporating social network information [48]; to detect relations and anomalies in text and speech [49]; to detect organizational structure based on machine learning and social network analysis [50]; to identify key actors in an organization [51, 52]; to detect communities [53]; to assign roles and community discovery in social networks [54]; and to visualize a social network with emails' topics [55].

One of the major limitations of our previous research ([27]) is that the calculation of the social score gives equal weights to all the inputs without considering their relative importance to approximate the outcome. In this paper, we expand our previous work [27] and propose the CorpRank algorithm to rank the officers of an organization based on principal component analysis (PCA). We use PCA to weigh several social network and e-mail indicators to calculate each officer's importance and compare our results with several community detection and clustering algorithms.

The rest of the paper is organized as follows: Section 2 describes the dataset; Section 3 presents the CorpRank algorithm; Section 4 describes the research design; Section 5 presents the results; Section 6 discusses the results, and Section 7 concludes the paper.

## 2. Enron Antecedents and Data

In this paper, we will use the well-known Enron e-mail dataset for the period 1998–2002 [6, 56–58]. As a part of Enron's legal proceedings in 2002, the Federal Energy Regulatory Commission (FERC) built a public dataset of 619,449 emails without attachments from 158 Enron employees. The emails used came from about 92% of relevant Enron employees. We use a version of the dataset provided by Shetty and Adibi [58] that has deleted extraneous and

unnecessary emails and has fixed some anomalies in the collection data having to do with empty or illegal user e-mail names and bounced email messages. Any duplicates or blank emails were not included. Our dataset, called the ENRON dataset, has 149 users after cleaning. We also included the position of each officer as provided by Shetty and Adibi [58]. According to this information, 38.5% of the users are classified as “employee” or “N/A.” “employee” only means that the user is working at Enron. We imputed the positions of several employees that were not well classified after reviewing their emails. We looked at signatures, content, internal traders’ list, and documents released during the bankruptcy proceedings [59]. After this review, we classified many “unknown” employees as traders or supporting traders.

We also used a FERC dataset that we call TRADER, with emails of 47 members of the North American West Power Trading division of Enron. We also had access to the organization chart of this division as provided by McCullough [60].

As a proxy of organizational importance, we used executive compensation provided by the United States Congress Joint Committee on Taxation [59] that included reports of outside consultants such as Towers Perkin [61].

### 3. The CorpRank Algorithm

We propose a ranking algorithm to evaluate the importance of the participants of an e-mail network. Even though the application of this algorithm could be used with any social group, we apply it to a corporate environment, as the ranking allows organizations to approximate the organizational hierarchy of a corporation.

The primary input is an e-mail corpus with a fixed number of accounts, each one of individual users.

The algorithm creates an undirected network where every account is a vertex. An edge is made between two vertices only when a minimum number of emails are transmitted between them. The weights of the edges are the number of emails exchanged between each pair of accounts.

The algorithm also finds all subgraphs (maximal complete cliques). It obtains several scores associated with the importance of the cliques according to their size and average response time of the primary account. The assumption is that users associated with a more extensive set and frequency of cliques will be more critical.

The final output of the algorithm is a score that ranks the users according to the mentioned clique scores, a set of social network indicators, the number of emails, the number of responses, and the average response time. This last feature is calculated as the period elapsed between the moment an e-mail is sent, and a response is received within three business days. This limit is established to avoid long response times or unrelated answers after a long time. Average response time and the number of emails are considered indicators of “importance” as probably most relevant messages might have priority over the rest, and a senior management position may require the supervision of a large unit. Therefore, the volume of emails may increase. Another essential set of features is

about the nature of the network’s connections and the social network structure, such as different measures of centrality, average distance, and the importance of hubs. All these social network metrics are indicators of the “importance” of an account in the e-mail network.

We assume that all these features are perfectly reasonable in an equation for “importance.” The weight of each feature may change by situation and organization, and therefore can be optimized with a method that emphasizes the role of each feature to the variations of the dataset such as PCA. For this reason, all calculated features are normalized and combined using PCA, each with an individual contribution to an overall score with which the users are ultimately ranked. The formal definition of the CorpRank algorithm is in Algorithm 1.

Once we obtain the CorpRank score, we can evaluate if the scores are consistent with the organization chart of the corporation under study. Any significant difference may indicate employees playing a central role even though the organization does not correctly recognize them.

### 4. Research Design

We classified the workers into nine organizational categories:

- (1) CEOs-Presidents: includes CEOs, chairman, chief operating officers (COOs), and presidents of Enron’s subsidiaries.
- (2) Executive V.P.s: includes executive vice presidents, functional chief officers (risk, staff, and general counsel), and an assistant to the president. This latter position may qualify as “employees;” however, an employee may communicate with the rest as their immediate subordinates.
- (3) Attorneys-legal asst.: includes lawyers, legal specialists, and general counsel assistant.
- (4) Managing directors
- (5) Vice presidents
- (6) Directors: includes directors and senior directors.
- (7) Managers: includes managers, senior managers, senior specialists, specialists, associates, analysts, functional managers (risk), and contractors.
- (8) Traders: includes analysts, senior specialists, specialists, and associates engaged in or supporting trading activities.
- (9) Employees: includes staff members not well classified in the other categories or with an unknown position.

We ranked the employees of both datasets ENRON and TRADER using the CorpRank score. We separated both datasets and evaluated our results in three equal-sized clusters: high, medium, and low scores, respectively. Additionally, we grouped Enron workers as senior managers (CEOs-Presidents, Executive V.P.s, Managing directors, and Vice presidents), middle managers (directors and managers), and others. We expect a particular relationship between the cluster and the occupational category of each

Input: a set of corporate emails with  $n$  number of individual accounts.

- (1) Build an undirected graph  $G(V, E)$ , where  $V = v_1, v_2, \dots, v_n$  is the set of vertices that represents the e-mail accounts,  $E$  is the set of edges, and  $e_{ij}$  is the edge between vertices  $v_i$  and  $v_j$  that have exchanged at least  $T$  emails. The value of the edge is the number of emails exchanged between  $v_i$  and  $v_j$ .
  - (2) Find all maximal complete cliques (subgraphs) using a recursive algorithm such as 457 [62].
  - (3) Calculate the adjacency matrix  $A$  and geodesic distance matrix  $D$  (the matrix of all shortest paths between every pair of vertices) for  $G$ .  $a_{ij}$  and  $d_{ij}$  are the elements of  $A$  and  $D$ , respectively. The mean of all the distances  $d_{ij}$  is  $L$ .
  - (4) The following features are calculated for each vertex  $v_i \in V$ :
    - Number of emails (e-mail): total number of emails sent and received.
    - Average response time (AvgTime): average amount of time elapsed between every email sent from  $v_i$  to any other account  $v_j$  and the next email received by  $v_i$  from account  $v_j$ .
    - Number of responses (NResponse): sum of all the responses to emails sent by  $v_i$  to any other accounts  $v_j$ .
    - Number of cliques (Clique): number of all cliques that  $v_i$  is contained within.
    - Raw clique score (RCS):  $R = 2^{u-1}$ , where  $u$  is the number of users in the clique.
    - Weighted clique score (WCS):  $W = t \cdot 2^{u-1}$  raw score weighted by the ‘‘importance’’ of  $v_i$  according to the average response time ( $t = \text{AvgTime}$ ).
    - Degree centrality (Degree):  $\text{deg}(v_i) = \sum a_{ij}$
    - Betweenness centrality (Betweenness):  $B_c(v_i) = \sum_i \sum_j g_{kij} / g_{kj}$ , where  $g_{kij}$  is the number of geodesic paths between vertices  $k$  and  $j$  that include vertex  $i$ , and  $g_{kj}$  is the number of geodesic paths between  $k$  and  $j$  [63].
    - Clustering coefficient (CC):  $C \doteq 1/m \sum_{i=1}^m CC_i$ , where  $CC_i \doteq 2|\{e_{ij}\}| / \text{deg}(v_i)(\text{deg}(v_i) - 1)$ :  $v_j \in M_i, e_{ij} \in E$ . Each vertex  $v_i$  has a neighborhood  $M$  defined by its immediately connected neighbors:  $M_i = \{v_j\}$ :  $e_{ij} \in E$ .
    - Average distance (AvgDistance): mean of the shortest path length from a specific vertex to all vertices in the graph  $G$ :  $L \doteq 1/n \sum_j d_{ij}$ , where  $d_{ij} \in D$ , and  $n$  is the number of vertices in  $G$ .
    - ‘‘Hubs-and-authorities’’ importance (Hubs): calculated with a recursive algorithm as proposed by Kleinberg [64]. ‘‘hub’’ refers to the vertex  $v_i$  that points to many authorities, and ‘‘authority’’ is a vertex  $v_j$  that points to many hubs.
  - (5) Each feature is mapped to a  $[0, 100]$  scale and weighted with the following formula:  $w_x \cdot C_x = w_x \cdot 100 \cdot [x_i - \text{inf}x / \text{sup}x - \text{inf}x] x_i$  is the value of the feature  $x$  for  $v_i$ ,  $w_x$  is the weight for the feature  $x$ ; the supremum  $\text{sup}x$  and infimum  $\text{inf}x$  are computed across all  $v_j$ .
  - (6) Run a principal component analysis (PCA) on all features and select the principal components that explain at least 80% of the variance of the dataset. The weight for each feature  $w_x$  is its normalized contribution to the variations of the selected principal components as follows:  $w_x = \sum_{\text{all } i} P_{x,i} \cdot E_i / \sum_{\text{all } x} \sum_{\text{all } i} P_{x,i} \cdot E_i$ , where  $P_{x,i}$  is the contribution of the feature  $x$  to explaining the variation of principal component  $i$ , and  $E_i$  is the eigenvalue of the principal component  $i$ .
  - (7) The CorpRank score, a ranking score between 0 and 100, is obtained for  $v_i$  as a weighted sum of the indicators:  $CR = \sum_{\text{all } x} w_x \cdot C_x / \sum_{\text{all } x} w_x$
- Output: CorpRank score for each account  $v_i$ .

ALGORITHM 1: The CorpRank algorithm.

employee. For instance, senior managers, middle managers, and employees should be mainly in the first, second, and third segments.

To validate our work, we built contingency tables for ENRON and TRADER with the three clusters of the CorpRank score and the occupational categories. In the case of TRADER, the dataset did not include employees of the first three organizational categories of the above list. We tested the null hypothesis, using the Chi-Square statistics, that either the ENRON’s or TRADER’s contingency table is not different from the expected contingency table. We compared the contingency tables with contingency tables that homogeneously distribute the same number of workers among the three clusters. We also contrast our results with the clusters generated by the following algorithms listed in increasing complexity levels. Besides the community detection algorithms, the rest of the following algorithms generated three clusters to be consistent with our previous classification:

- (1) Community detection algorithms: communities or clusters are detected using the e-mail network.

- (a) Walktrap algorithm: finds an efficient community structure according to a measure of similarity between vertices based on random walks [65].
- (b) Edge-betweenness algorithm: discovers community structures of a network by iteratively removing the edges with the highest betweenness, generating a hierarchical map of the different modules [66].
- (2)  $K$ -means: partitions the data finding a prespecified number of clusters that minimizes the within-cluster variation with the features used to generate the CorpRank score introduced in Algorithm 1.
- (3) PCA hierarchical clustering (PCA HC) uses as inputs the principal components and their coordinates of the Algorithm 1’s features. PCA HC initially treats every observation as a cluster. Iteratively, this method fuses the two clusters with the smallest distance, such as the Euclidean distance, and repeats the same process until all the observations belong to a cluster.

- (4) PCA  $K$ -means: conducts a  $K$ -means using the coordinates of the principal components of the Algorithm 1's features.

We calculated the average bonus of 2001 for each hierarchical category as an additional proxy of organizational importance. The bonus combines the regular performance bonus and a retention bonus paid by Enron in 2001 to maintain its crucial employees (this information is available in [59] and [61]). This bonus is a direct indicator of the importance that each employee acts as a leader or as a profit generator for the company. For instance, some star energy traders received bigger bonuses than some senior managers because of Enron's urgency to generate significant revenues in a short time.

We used  $R$  and EMT [29], a Java-based e-mail analysis engine built on a database backend, for data processing and analysis.

## 5. Results

The first three principal components capture 78.4% and 93% of the variance for the ENRON and TRADER datasets, respectively (Figure 1). For both datasets, the features betweenness centrality and clustering coefficient, and for ENRON alone, the average response time and average distance have a much smaller contribution to the first three principal components than the rest (Table 1). Average response time is associated with the importance of emails; however, this effect is already captured by the number of responses. The rest of the above features is all social network features, where degree, hubs, and clique features are dominant.

The community detection algorithm Walktrap shows the highest modularity for both the ENRON and TRADER datasets (Table 2). Modularity is an indicator of the quality of a partition that evaluates if there are many links connected to every partition above an expected random number of links [66].

Hence, we use Walktrap as our primary benchmark to evaluate the quality of our algorithm. In this respect, the structure of clusters generated by the Walktrap algorithm is not significantly different than the expected value for ENRON (Table 3); while for TRADER, it is significant at the 95% confidence level (Table 4-E). In both datasets, most observations of most occupational categories are concentrated in one segment. Therefore, the algorithm cannot separate observations into clusters related to their organizational importance.

The three segments defined by the CorpRank score have aggregated Enron's employees in a nonrandom pattern (Table 5). The Chi-Square test rejects the null hypothesis of a random assignment with  $p$  value smaller than 0.1% for the ENRON (Table 6-A) and TRADER (Table 4-A) datasets.

To evaluate if the aggregation generated by the CorpRank score also corresponds to the organizational hierarchy, we obtained the average bonus of 2001 for each organizational category (see Tables 7 and 8) The correlation coefficient between the average CorpRank score and the

average bonus for each organizational category is 0.9 and 0.83 for ENRON and TRADER. Even though the number of observations (similar to the number of organizational categories) is small, a high correlation coefficient is a good indicator that the two features move in the same direction. A casual observation of Tables 7 and 8 shows that the average CorpRank score ranks the organizational categories according to their expected importance: senior management, middle management, and traders.

As the traders were the primary profit drivers at Enron, we analyze the organizational hierarchy of traders and managers independently in the next section. There are alternative communication systems for the traders (instantaneous messages, tweets, phones, Bloomberg, or trading terminals). They do not have many emails in our datasets, and they communicate primarily among themselves. Hence, they may have lower CorpRank scores than the rest of the workers. However, in the Enron North American West Power Trading division, the traders are more critical than employees, as this is a specialized trading division. They are mainly distributed between the first and third cluster.

Employees are mainly concentrated in the second cluster for TRADER and in the third cluster for ENRON. According to the emails, some of the employees have a lot of influence and could have been assistants of senior managers or directors. However, the emails studied did not indicate their professional position. For this reason, the average CorpRank score for this group is not the lowest one as expected.

## 6. Discussion

*6.1. Analysis of Complete ENRON Dataset.* Tables 5 and 6 show that most of the senior managers and the legal team (Attorneys-legal asst.) are concentrated in the first segment, which has the most significant CorpRank score. Even though 78% of CEOs and presidents are in this segment, this proportion decreases to 62.5% for the vice presidents. A particular case is the legal team because it does not include any senior managers (the general counsel and other vice presidents involved in legal aspects were classified as senior managers); however, it has about the same average CorpRank score as the executive VPs and managing directors (see Table 7). This can be explained considering the central role that this professional group played while Enron hid its losses using new financial vehicles and then filed for bankruptcy in 2001.

There is a critical jump between middle managers and senior managers. About 46% of the directors are in the second segment according to the CorpRank score (Table 5), while managers have about the same presence in the third segment. The distribution of these two groups reflects their role as middle managers where the directors are at the top of the group and regularly have several managers in their teams. The group of managers is very diverse and may have small team leaders and subject-matter experts with minimal contact with the rest of the organization.

The top-ranked employees according to the CorpRank score include Liz Taylor, assistant to the president, and COO (proxy of Gregg Whalley and Jeff Skilling); Louise Kitchen,

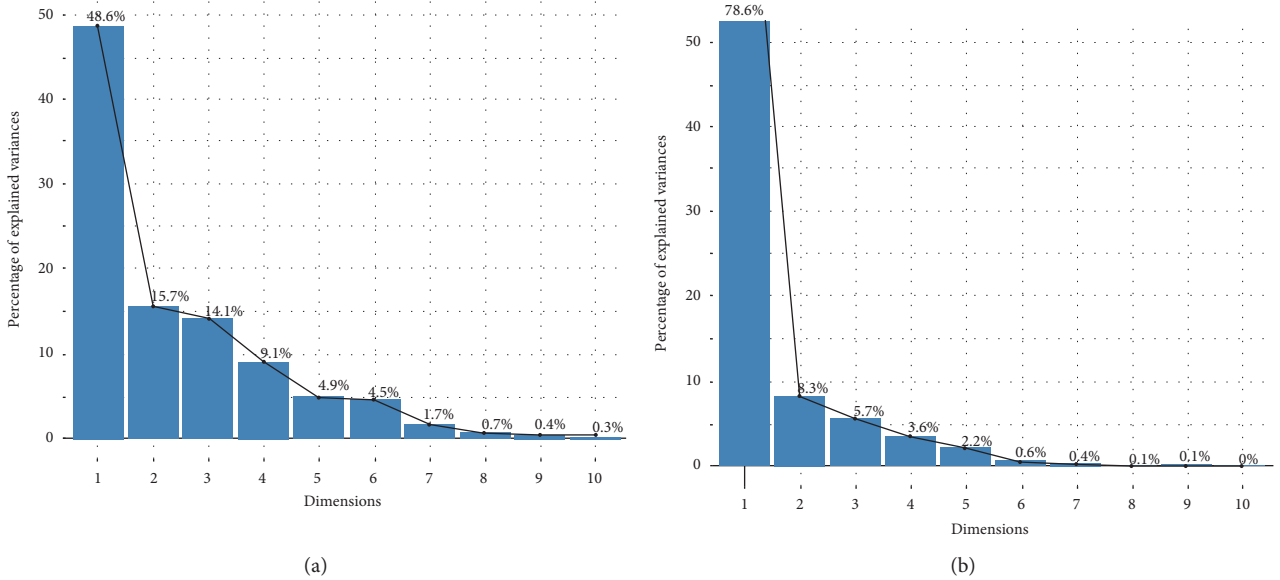


FIGURE 1: Proportion (%) of variance explained by principal components for Enron (ENRON) and the Enron North American West Power Trading division (TRADER). (a) ENRON. (b) TRADER.

TABLE 1: Features’ weights (%) for Enron (ENRON) and Enron North American West Power Trading (TRADER) according to the sum of the product of the contribution of each feature to the first three principal components and their eigenvalues.

	ENRON	TRADER
Number of responses	11.17	9.53
Raw clique score	11.12	9.09
Number of emails	11.11	9.22
Degree centrality	11.00	9.42
Weighted clique score	10.85	9.02
Number of cliques	10.25	9.31
Hubs and authorities	10.24	9.18
Average distance	8.25	9.56
Betweenness centrality	8.19	7.75
Clustering coefficient	7.55	8.26
Average response time	0.26	9.66

TABLE 2: Modularity of the Walktrap [65] and the edge-betweenness algorithms [66] for the ENRON and TRADER datasets.

	Walktrap	Edge betweenness
ENRON	0.64	0.45
TRADER	0.11	0.03

who was the president of Enron Online, had tremendous importance because she implemented the online trading capability; Sally Beck, COO, and Ken Lay, CEO (Table 9). Even though some of the most influential leaders are not at the top of the scale, such as Gregg Whalley and Jeff Skilling, who are in the 18th and 35th places, respectively, their assistant is Liz Taylor (Gregg Whalley replaced Jeff Skilling as president and COO after the latter resigned on August 14, 2001). Her high ranking reflects the high rankings of Gregg Whalley and Jeff Skilling.

The clusters generated by PCA K-means and PCA hierarchical clustering split the occupational categories in a

cluster for the senior managers and another for the rest. Even though most occupational categories are concentrated in two clusters, these clusters are still consistent with the CorpRank score’s clusters. A significant difference is in the structure generated by K-means where most of the observations are concentrated in one cluster. As Figure 2(a) shows, K-means generates very well-differentiated clusters according to the most relevant features evaluated. At the same time, there is a certain overlap among the CorpRank score’s clusters (Figure 2(b)). A similar path is observed by the Enron North American West Power Trading division (although the figures are not included to save space).

TABLE 3: Frequency of Enron’s employees by occupational categories according to the clusters discovered (columns) by the Walktrap algorithm [65]. Chi-square’s  $p$  value compares the results with the expected values assuming that there is independence between the occupational categories and the clusters generated by each algorithm.

	Walktrap								
	1	2	3	4	5	6	7	8	9
Senior managers									
CEOs-President	3	0	4	1	0	0	1	0	0
Executive VP	0	0	2	0	0	0	0	1	0
Managing Director	0	2	2	0	0	0	2	1	0
Vice President	2	6	7	1	4	2	1	1	0
Middle managers									
Director	2	2	12	1	1	2	1	1	0
Manager	3	5	7	1	1	3	4	2	0
Attorneys-legal asst.	2	0	3	1	0	1	1	2	0
Trader	0	7	14	1	2	2	7	1	0
Employee	3	1	6	1	0	2	1	1	1
Total	15	23	57	7	8	12	18	10	1
Chi-sq. $p$ value					0.58				

TABLE 4: Distribution of Enron North American West Power Trading’s employees in occupational categories by A. CorpRank score, B. PCA  $K$ -means, C. PCA Hierarchical cluster (PCA HC), D.  $K$ -means, and E. Walktrap algorithm. Chi-square’s  $p$  value compares the results with the expected values assuming that there is independence between the occupational categories and the clusters generated by each algorithm.

Algorithm	A. CorpRank score			B. PCA $K$ -means			C. PCA HC			D. $K$ -means			E. Walktrap			
	High	Medium	Low	1	2	3	1	2	3	1	2	3	1	2	3	4
Senior managers																
Managing Director	1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0
Vice President	3	1	0	2	0	2	0	2	2	2	2	0	4	0	0	0
Middle managers																
Director	5	5	0	8	2	0	2	8	0	6	0	4	10	0	1	0
Manager	2	3	1	2	3	1	3	2	1	1	1	4	10	1	0	0
Trader	5	3	14	8	13	1	13	8	1	6	1	15	8	13	0	1
Employee	0	3	1	3	1	0	1	3	0	1	0	3	2	3	0	0
Total	16	15	16	23	19	5	19	23	5	16	5	26	35	17	1	1
Chi-sq. $p$ value		$\ll 0.01$			$\ll 0.01$			$\ll 0.01$			$\ll 0.01$			$0.047$		

TABLE 5: Proportion (%) of Enron’s employees in occupational categories by CorpRank score segments.

	CorpRank score segments		
	High	Medium	Low
Senior managers			
CEOs-president	77.78	22.22	0.00
Executive VP	66.67	0.00	33.33
Managing Director	57.14	28.57	14.29
Vice President	62.50	20.83	16.67
Middle managers			
Director	22.73	45.45	31.82
Manager	23.08	30.77	46.15
Attorneys-legal asst.	60.00	40.00	0.00
Trader	5.88	44.12	50.00
Employee	21.43	28.57	50.00

This analysis reflects the flexibility of our approach as the organizational importance of Enron’s employees may lead to situations where a senior manager and a manager may have similar values across different features; however, they still

belong to other clusters. Additionally, the PCA weighting helps identify the most relevant features, and the final ranking of employees is associated with their organizational importance.

TABLE 6: Frequency of Enron’s employees in occupational categories by A. CorpRank score, B. PCA *K*-means, C. PCA hierarchical cluster (PCA HC), and D. *K*-means. Chi-square’s *p* value compares the results with the expected values assuming that there is independence between the occupational categories and the clusters generated by each algorithm.

Algorithm	A. CorpRank score			B. PCA <i>K</i> -means			C. PCA HC			D. <i>K</i> -means		
	High	Medium	Low	1	2	3	1	2	3	1	2	3
Senior managers												
CEOs-President	7	2	0	3	2	4	2	4	3	3	4	2
Executive VP	2	0	1	0	1	2	1	2	0	0	2	1
Managing Director	4	2	1	0	3	4	3	4	0	0	3	4
Vice President	15	5	4	1	8	15	8	15	1	2	7	15
Middle managers												
Director	5	10	7	0	18	4	18	4	0	0	3	19
Manager	6	8	12	0	20	6	20	6	0	0	3	23
Attorneys-legal asst.	6	4	0	0	4	6	4	6	0	0	2	8
Trader	2	15	17	0	29	5	29	5	0	0	0	34
Employee	3	4	7	0	11	3	11	3	0	0	2	12
Total	50	50	49	4	96	49	96	49	4	5	26	118
Chi-sq. <i>p</i> value		≤0.01			≤0.01			≤0.01		≤0.01		≤0.01

TABLE 7: Compensation and average CorpRank score by hierarchical categories for Enron’s employees. Average bonus includes regular performance bonus and a retention bonus given to a selected group of employees in 2001 according to [59]. Avg total is the average total direct compensation. Average bonus and average total direct compensation are in U.S. dollars. NA indicates that information is not available.

	CorpRank score	Avg bonus	Avg total
Senior managers			
CEOs-President	32.03	3,692,171	10,715,788
Executive VP	20.54	900,000	2,798,333
Managing Director	20.66	1,340,000	3,037,501
Vice President	24.08	1,226,864	2,057,013
Middle managers			
Director	18.40	562,300	NA
Manager	16.02	97,406	NA
Attorneys-legal asst.	21.51	125,000	NA
Trader	13.06	54,188	NA
Employee	15.39	NA	NA

TABLE 8: Average bonus for 2001 and average CorpRank score by hierarchical categories and sorted by CorpRank score for Enron North American West Power Trading division. Average bonus includes regular performance bonus and a retention bonus given to a selected group of employees in 2001 according to [59]. Average bonus in U.S. dollars.

	CorpRank score	Avg bonus	Avg. total
Senior managers			
Managing Director	77.13	5,249,999	5,719,962
Vice President	42.46	4,600,000	4,120,000
Middle managers			
Director	31.75	656,250	NA
Manager	35.85	NA	NA
Trader	27.93	30,400	NA
Employee	26.26	NA	NA

6.2. *Analysis of North American West Power Trading Division.* Table 4 shows that the senior managers are concentrated in the first segment (high CorpRank score), middle managers and employees in the second segment, and traders in the third segment (low CorpRank score). Different than what we observed in the previous section, the other cluster algorithms included in Table 4 do not show a consistent pattern to separate the employees by occupational categories. Partially,

this might be explained because TRADER has about a third of ENRON’s observations. Additionally, Table 8 demonstrates that the average CorpRank score and the average bonus follow the same trend.

As seen in Table 10, we can reproduce with high accuracy the very top of the hierarchy of the North American West Power Trading division. Figure 3 shows that Tim Belden, head of the trading operation, and his administrative



TABLE 9: CorpRank scores of employees: Enron. TW: TransWestern pipeline company, N.A.: North America, V.P.: Vice President, E: Enron, RCS: raw clique score, WCS: weighted clique score, Btw: betweenness, Avg. Dist.: average distance, CC: clustering coefficient, Score: CorpRank score. \* \* indicates that the content of emails infers the position.

	Position	# email	Avg time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Liz Taylor	Assistant to president & COO	40.99	27.92	2.58	100	89.73	70.16	100	100	100	0	12.15	63.12
Louise Kitchen	President, Enron online	59.2	27.74	21.02	99.17	100	77.86	71.64	28.68	86.31	11.65	20.82	60.35
Sally Beck	Chief operating officer	26.12	37.4	4.66	78.51	77.14	61.9	88.06	56.99	90.43	7.63	13.93	52.17
Kenneth Lay	CEO, chairman	11.73	53.9	0.69	80.17	91.72	74.32	74.63	38.82	88.04	10.04	18.7	50.65
Kimberly Watson	Director deal origination	42.89	34.84	32.54	10.74	91.28	93.23	23.88	2	14.28	63.45	64.05	44.19
Michelle Lokay	Manager*	39.83	28.48	12.59	12.4	99.19	100	20.9	1.68	12.17	56.63	75	43.08
Mike Grigsby	VP trading, ENA Gas West	82.89	35.5	37.07	67.77	32.44	19.49	47.76	19.78	57.78	18.88	27.63	42.55
Steven Harris	VP TW group	30.02	27.25	7.68	12.4	99.19	100	23.88	2.06	14.9	55.42	69.28	41.55
Lindy Donoho	Employee	23.47	31.8	10.07	12.4	99.19	100	22.39	1.4	13.08	61.45	73.53	41.51
Jeff Dasovich	Director Sr Gov/Reg. Affairs	100	31.27	100	17.36	13.21	10.53	34.33	15.81	37.41	26.51	26.33	39.83
Kevin Hyatt	Director, Pipeline business	14.46	25.13	3.02	9.92	81.33	81.79	20.9	1.68	12.17	56.63	75	34.96
Drew Fossum	VP & Gen counsel	22.31	25.46	7.55	14.05	75.81	76.23	25.37	4.38	17.17	49	53.22	34.5
Lynn Blair	Director	22.94	27.49	7.43	6.61	69.42	70.64	20.9	1.29	12.58	61.85	74.17	33.91
Tana Jones	Senior legal specialist	85.35	24.7	47.39	19.83	9.93	6.6	26.87	6.62	28.66	35.74	35.26	30.74
Scott Neal	VP trading, ENA Gas East	14.07	42.44	1.95	42.98	43.15	30.95	43.28	12.12	55.65	24.1	29.03	30.08
Shelley Corman	VP Gas logistics & comm.	21.66	26.47	7.61	10.74	56.14	55.74	23.88	5.18	18.14	51	55.56	30.03
Sara Shackleton	VP ENA & senior counsel	80	25.06	56.39	10.74	5.02	3.34	14.93	0.68	12.01	59.04	59.09	29.54
Lawrence G. Whalley	President & COO	14.22	10.87	0.44	25.62	52.92	47.68	28.36	2.4	48.39	26.91	47.14	29.4
Teb Lokay	Regulatory affairs manager	7.25	18.59	1.89	4.13	49.57	51.19	16.42	0.09	7.6	91.97	87.18	29.32
Darrell Schoolcraft	Manager*	8.39	49.41	2.39	2.48	43.61	45.39	11.94	0.03	5.78	95.98	95.56	28.39
Roderick Hayslett	Managing Director finance ETS	20	21.69	16.17	6.61	43.61	43.36	22.39	12.44	25.87	32.13	57.35	27.43
Steven Kean	EVP chief of staff	17.28	29.7	3.21	18.18	42.49	39.82	25.37	3.13	40.27	27.31	49.71	26.44
John Arnold	VP trading	20.51	22.14	5.35	33.06	34.49	23.85	26.87	3.26	45.25	29.72	43.16	26.31
Andy Zipper	VP trading	8.94	43.09	1.13	33.06	42.37	30.43	20.9	0.98	39.57	34.54	55.83	26.13
Richard Shapiro	Managing Dir.-Govt. Affairs	62.82	17.27	18.75	14.88	13.71	10.82	23.88	3.58	33.12	33.73	45.1	25.85
Bill Rapp	Attorney	9.08	37.92	4.78	3.31	39.7	41.23	14.93	2	9.01	67.47	78.79	25.11
Hunter Shively	VP trading	10.07	36.27	1.95	26.45	36.72	27.6	26.87	4.18	45.51	30.92	41.05	24.83
Rosalee Fleming	Employee	19.13	33.14	1.89	8.26	35.67	34.48	22.39	0.81	39.98	28.92	61.76	24.65
Phillip Allen	Managing director trading	13.11	6.32	0.44	32.23	30.33	21.85	23.88	1.69	44.88	28.51	50.98	24.14
Stanley Horton	President, Enron gas pipeline	9.33	29.32	1.01	7.44	39.89	39.72	23.88	7.03	33.12	30.52	50.98	23.69
Tracy Geaccone	Sr. Director Consolidations ETS	14.24	13.77	11.71	3.31	23.76	23.83	14.93	0.23	9.76	74.7	83.33	23.61
Stephanie Panus	Employee	43.37	36.02	28.51	23.97	9.37	5.77	22.39	3.96	17.1	47.79	38.24	23.59
Mark McConnell	Manager, TW*	16.12	29.14	7.3	2.48	23.76	23.9	14.93	0.18	7.11	77.51	87.88	23.58
Mark Haedicke	Managing director, legal	9.42	21.91	1.83	23.97	30.96	28.47	31.34	10.52	37.91	25.3	35.97	23.29
Jeffrey Skilling	Enron President & COO	3.49	15.85	0.57	9.09	39.27	38.18	17.91	0.6	34.72	32.93	63.74	23.04

TABLE 9: Continued.

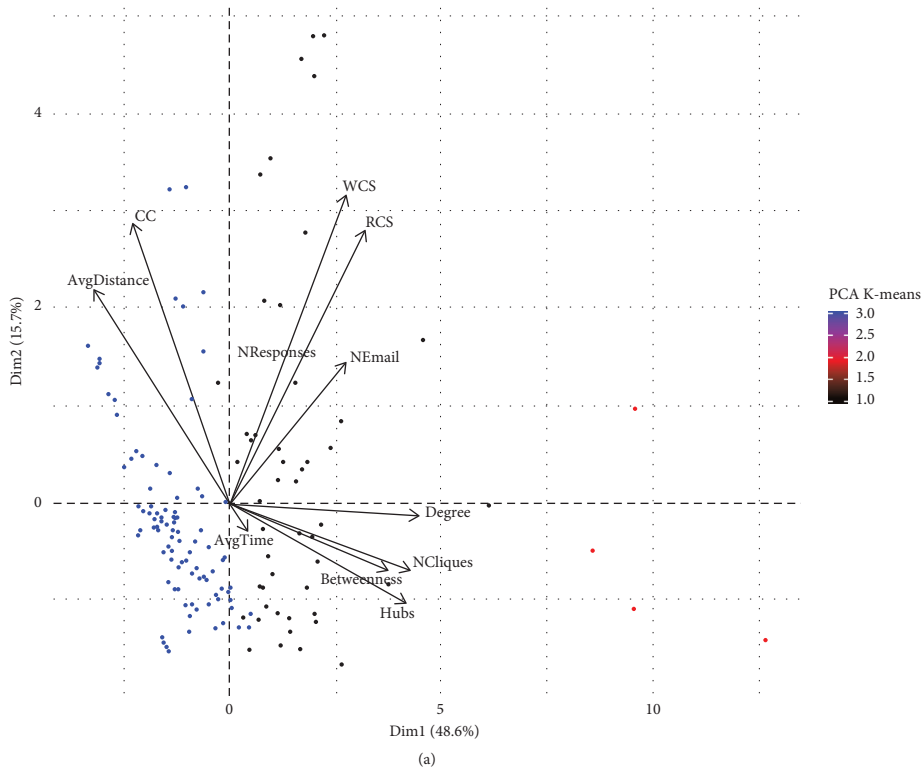
	Position	# email	Avg time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Marie Heard	Lawyer	45.47	31.24	23.41	14.88	8.5	5.68	16.42	1.15	18.23	38.96	64.1	22.65
Paul Barbo	Associate	11.3	35.55	2.08	2.48	31.79	33.98	16.42	3.79	8.45	65.86	67.95	22.6
Rick Buy	EVP chief risk officer	7.88	9.93	0.82	9.09	32.07	31.86	16.42	0.66	34.49	30.92	74.36	22.51
Carol Clair	In-house lawyer, assistant general counsel	43.93	21.44	37.57	8.26	2.92	1.75	10.45	0.72	10.61	57.03	63.89	22.32
Kevin Presto	VP trading, ENA east power	15.33	29.46	7.61	42.15	15.69	8.23	29.85	4.09	39.44	32.13	30.3	22.18
Matthew Lenhart	Trader	26.24	28.12	5.54	33.88	16.38	9.27	26.87	6.19	33.65	28.51	38.42	22.07
Kam Keiser	Manager risk management gas	22.34	17.65	8.87	31.4	12.59	6.81	32.84	6.93	39.41	30.92	29.71	21.89
Susan Scott	Analyst, assistant trader	25.23	34.06	4.97	22.31	12.03	8.35	40.3	25.11	30.27	28.92	13.55	20.97
Jeffrey Shankman	President, Enron global mkts	3.73	17.03	1.07	3.31	25.99	27.05	17.91	0.47	36.93	33.73	71.43	20.71
Susan Bailey	Legal specialist*	38.55	38.79	10.89	7.44	4.9	3.24	11.94	0.14	9.82	65.46	75.56	20.69
James Steffes	VP, government affairs	34.63	20.84	29.01	13.22	5.89	3.74	16.42	2.75	24.81	38.15	43.59	20.57
Thomas Martin	VP	6.02	17.79	0.88	14.88	26.86	21.26	20.9	1.93	37.07	33.33	49.17	20.35
Fletcher Sturm	VP	5.42	24.52	0.88	20.66	20.78	15.84	22.39	1.05	41.27	32.53	52.21	20.31
Barry Tycholiz	VP	27.64	12.88	11.01	20.66	6.76	3.69	20.9	2.88	32.9	29.72	42.5	19.17
Monique Sanchez	Trader*	15.42	34.06	2.27	26.45	10.55	5.67	32.84	6.04	38.04	30.52	26.81	19.09
Phillip Love	Manager risk mgt gas	13.61	21.15	7.87	23.14	7.57	3.74	31.34	9.92	33.63	36.95	26.48	18.82
Jay Reitmeyer	Trader associate	13.4	47.25	1.2	24.79	13.09	7.32	19.4	3.46	31.33	33.73	47.62	18.53
Gerald Nemec	Attorney*	23.35	31.44	13.15	20.66	5.58	2.97	26.87	12.67	23.05	31.33	26.32	18.13
Elizabeth Sager	VP & asst. Gral. Counsel	17.78	12.25	3.4	18.18	8.75	5.82	28.36	7.65	31.14	30.92	33.81	17.9
Debra Perlingiere	Legal specialist*	19.2	29.99	9.75	17.36	6.48	3.78	26.87	6.6	19.63	43.37	28.95	17.48
Jason Williams	Specialist Sr. logistics	18.53	33.57	4.28	22.31	7.2	3.98	22.39	4.34	28.11	31.73	38.97	17.39
Danny McCarty	Managing Director	6.72	20.42	1.26	2.48	18.8	18.88	17.91	3.96	22.62	36.14	58.24	17.32
Mike Maggi	Director	6.72	18.48	0.19	6.61	9.99	7.16	11.94	0.18	25.96	39.76	82.22	16.85
Mark Taylor	VP & Gen counsel	15.81	21.18	7.43	4.96	5.65	4.21	13.43	0.91	18.69	42.57	67.27	16.29
Kim Ward	Trader*	18.8	19.41	9.94	17.36	3.38	1.66	25.37	9.14	28.89	26.91	24.56	16.16
Eric Linder	Trader*	0.39	NA	0	0	0.06	0.02	0	0	0	100	100	15.9
Lysa Akin	Sr Admin asst Gvt Affairs	8.65	30.53	1.89	23.14	6.33	2.9	26.87	11.63	11.91	49.8	25.79	15.82
Lisa Gang	Director	1.2	NA	0	0	0.19	0.09	2.99	0	0.88	92.37	100	15.81
Matt Smith	Trading associate	8.84	39.63	0.82	4.96	2.98	1.73	8.96	0.1	16.94	56.22	82.14	15.76
Phillip Platter	Sr. specialist	0.67	3.82	0.06	0.83	0.43	0.24	2.99	0	1.03	90.76	100	15.72
Errol McLaughlin	Specialist sr. - risk mgt gas	20.77	30.37	2.01	7.44	3.47	1.98	16.42	1.04	25.94	39.76	51.28	15.67
Mark Whitt	VP	16.27	25.15	11.2	6.61	2.67	1.5	16.42	2.15	23.01	39.76	46.15	15.36
Mary Hain	In house lawyer	19.13	18.17	7.49	15.7	3.41	1.59	23.88	8.9	18.69	38.15	22.88	15.31
John Lavorato	CEO, Enron America	6.05	NA	0	19.01	9.86	5.18	20.9	1.4	29.96	40.16	28.33	15.3
Jason Wolfe	Trader analyst	8.89	15.91	0.25	13.22	3.91	2.07	11.94	0.28	26.71	35.74	68.89	15.29
Chris Germany	Trader*	22.87	45.37	3.65	11.57	3.19	1.44	20.9	5.36	13.88	57.43	20.83	15.22
Susan Pereira	Trader*	0	NA	0	0	0.06	0.03	0	0	1.87	89.56	100	15.19
John Griffith	Managing director UK	9.4	27.38	1.13	4.96	6.14	4.08	13.43	0.45	23.45	42.57	63.64	15.1
Cara Semperger	Senior analyst cash	7.49	33.83	2.83	7.44	2.98	1.45	11.94	0.57	3.17	77.51	60	15.09
Tori Kuykendall	Trader	10.72	8.39	0.94	7.44	3.44	1.97	11.94	0.39	14.51	63.05	57.78	15.07
Jane Tholt	Director	1.18	38.34	0.31	0.83	0.12	0.06	1.49	0	4.84	76.71	100	14.9
Dana Davis	VP term	5.35	50.08	0.19	18.18	6.95	3.22	13.43	0.82	28.22	34.94	47.27	14.61
Lawrence May	Director	6.75	42.6	0.82	5.79	6.58	4.57	17.91	2.38	30.14	32.53	51.65	14.6

TABLE 9: Continued.

	Position	# email	Avg time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Mark Guzman	Trader	1.47	NA	0	0.83	0.68	0.32	5.97	0.04	1.45	87.55	80	14.5
Keith Holst	Director	11.4	51.06	0.31	12.4	6.51	3.62	13.43	1.49	26.64	34.94	45.45	14.45
Steven South	Trader*	0.1	NA	0	0	0	0	0	0	4.08	77.11	100	14.39
Don Baughman	Trader	6.12	46.02	0.88	19.83	5.27	1.96	17.91	3.36	23.86	39.76	35.16	14.34
Kate Symes	Trader*	6.48	24.44	1.89	9.09	3.35	1.58	17.91	1.68	4.8	74.7	39.56	14.22
Frank ermis	Director	7.59	9.73	0.38	10.74	3.85	2.16	14.93	0.53	25.58	36.14	56.06	14.19
Stacy Dickson	Employee	11.73	18.23	1.7	3.31	0.56	0.27	5.97	0.12	4.11	79.92	60	14.18
Jonathan Mckay	Director	4.43	11.4	0.44	15.7	7.75	4.51	14.93	1.22	23.11	39.36	42.42	14.09
Eric Bass	Trader	14.67	17.23	7.99	11.57	2.61	1.31	17.91	4.01	24.97	31.33	31.87	14.03
Randall Gay	Associate	9.3	18.86	0.44	9.09	4.06	2.29	13.43	3.35	22.74	35.34	56.36	14.01
Richard Sanders	VP, Enron wholesale svcs.	5.61	28.72	1.26	8.26	3.1	1.83	10.45	0.34	12.82	61.45	55.56	13.98
Sandra Brawner	Director	3.3	10.93	1.13	5.79	5.4	3.73	10.45	0.27	24.38	39.76	63.89	13.89
Chris Dorland	Manager	4.31	11.92	0.76	22.31	5.71	2.77	20.9	2.27	26.5	32.53	27.5	13.78
Doug Gilbert- smith	Director	5.45	26.98	1.26	10.74	2.92	1.31	14.93	0.87	28.19	38.55	46.97	13.71
Theresa Staab	Employee	3.95	42.65	2.01	0.83	0.31	0.15	4.48	0.06	5.14	69.48	80	13.7
Dan Hyvl	Employee	9.37	26.89	5.41	4.96	1.36	0.77	10.45	1.13	9.32	60.64	52.78	13.64
Bill Williams	Trader analyst	8.17	40.95	1.38	7.44	2.67	1.32	17.91	5.23	4.71	68.27	32.97	13.36
Michelle Cash	Gen counsel assistant	2.77	34.06	0.82	5.79	3.41	2.13	10.45	0.34	14.94	54.22	58.33	13.27
Paul Lucci	Director	5.57	20.98	3.02	3.31	1.8	1.1	10.45	0.43	13.72	49.8	63.89	13.19
Dutch Quigley	Trader	10.05	21.56	2.52	7.44	2.23	1.2	14.93	1.23	23.73	37.75	42.42	13.08
Jeff King	Manager	1.71	NA	0	8.26	1.43	0.4	4.48	0.02	13.78	46.18	80	13.06
David Delainey	CEO, Enron N.A. & E.energy	1.52	NA	0	1.65	1.92	1.13	7.46	0.11	21.18	38.55	80.95	13.03
Vince Kaminski	Risk management head	2.94	13.37	0.57	1.65	4.19	3.87	7.46	0.68	17.34	46.59	66.67	13.01
Darron Giron	Trader associate	4.39	23.48	2.33	0.83	0.06	0.03	1.49	0.01	5.48	76.31	66.67	12.95
Harry Arora	VP	1.95	5.44	0.25	5.79	2.05	1.06	8.96	0.2	21.15	48.19	60.71	12.92
Martin Cuilla	Manager	4.63	26.02	0.5	14.05	2.42	0.97	16.42	3.61	22.19	41.37	34.62	12.85
Kay Mann	Employee	8.46	22.67	2.89	9.92	2.45	1.25	14.93	5.85	9.47	51	36.36	12.79
Stacey White	Director risk management	4	17.39	0.82	9.92	2.3	1.04	10.45	0.67	22.29	40.16	52.78	12.75
James Derrick	EVP general counsel	5.13	29.66	0.44	3.31	4.84	3.94	11.94	1.48	23.73	33.73	53.33	12.67
Robert Benson	Director	1.86	NA	0	4.96	1.67	0.6	5.97	0.06	18.07	42.97	73.33	12.63
Brad Mckay	Contractor	2.34	NA	0	7.44	2.67	1.05	10.45	0.69	17.6	46.18	55.56	12.52
Diana Scholtes	Trader	6.02	39.39	1.32	7.44	2.67	1.38	13.43	4.92	9.77	43.78	50.91	12.46
Charles Weldon	Trader associate	1.08	0.02	0.19	3.31	0.43	0.18	2.99	0.01	12.14	46.99	83.33	12.29
Andrew Lewis	Director, trader	1.23	NA	0	1.65	0.43	0.16	4.48	0.04	15.64	44.58	80	12.27
John Zufferli	VP	3.93	14.93	0.25	9.09	1.67	0.68	8.96	0.35	19.68	40.96	53.57	12.15
Vladi Pimenov	Trading associate	1.49	31.33	0.19	4.13	2.42	1.25	5.97	0.12	15.92	44.58	66.67	12.1
Jim Schwieger	VP	3.81	36.44	0.63	10.74	5.33	3.13	11.94	0.72	18.47	42.17	35.56	12.04
Sean Crandall	Director, trader	4.6	32.22	0.38	9.09	2.85	1.32	10.45	2.25	8.69	44.58	52.78	11.91
Holden Salisbury	Cash analyst	2.63	17.94	0.25	3.31	1.05	0.52	8.96	3.68	6.79	49.8	64.29	11.82
Scott Hendrickson	Trader*	4.02	27.61	0.38	1.65	0.31	0.09	2.99	0.12	9.12	56.63	66.67	11.75
Larry Campbell	Senior specialist	0.22	NA	0	0	0.06	0.01	1.49	0.03	2.83	74.7	66.67	11.73
Geir Solberg	Trader analyst	2.12	28.57	0.13	3.31	0.93	0.42	7.46	1.93	6.45	50.6	66.67	11.65
Mike Carson	Employee	2.14	10.62	0.06	6.61	1.67	0.78	5.97	0.15	19.14	38.55	60	11.56
Judy Townsend	Trader*	7.23	19.73	0.31	0.83	0.56	0.34	2.99	0.24	9.77	49	66.67	11.49
John Hodge	Managing director	2.22	37.51	0.38	1.65	0.5	0.2	4.48	0.36	6.81	61.85	60	11.48
Mike Swerzbin	VP trading	3.2	22.23	0.06	8.26	2.17	0.94	13.43	4.22	22.09	34.14	38.18	11.39

TABLE 9: Continued.

	Position	# email	Avg time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Robin Rodrigue	Analyst*	0.65	35.86	0.13	0.83	0.25	0.12	4.48	0.18	10.52	51	70	11.37
Matt Motley	Director	1.61	NA	0	4.13	1.24	0.49	5.97	0.7	7.32	60.24	53.33	11.34
Juan Hernandez	Senior specialist logistics	2.29	40.86	0.5	3.31	0.56	0.15	5.97	0.17	12.35	48.59	60	11.3
Benjamin Rogers	Associate	1.42	5.31	0.5	2.48	0.56	0.22	4.48	0.12	11.43	52.61	60	11.11
Robert Badeer	Director	9.18	8.86	0.19	4.13	0.53	0.22	7.46	1.17	8.56	54.62	42.86	11.11
Eric Saibi	Trader	3.45	55.72	0.13	3.31	0.74	0.31	5.97	0.23	14.85	40.96	60	11.09
Peter Keavey	Employee	0.75	0	0.06	0.83	0.22	0.12	2.99	0.02	7.4	72.29	50	11.04
Joe Stepenovitch	Specialist Sr.trader support	2.14	100	0.13	1.65	0.31	0.07	4.48	0.11	8.08	53.82	60	11
Jerry Farmer	Logistics manager	1.11	13.93	0.31	1.65	0.31	0.15	4.48	0.21	9.17	53.01	60	10.77
John Forney	Manager, real time trading	4.53	57.91	1.13	8.26	1.09	0.28	8.96	0.45	12.38	47.39	35.71	10.66
Craig Dean	Trader	1.16	NA	0	0.83	0.06	0.01	1.49	0.27	5.07	55.02	66.67	10.57
Ryan Slinger	Trader	2.39	20.66	0.19	3.31	0.74	0.3	7.46	5.63	5.8	51.81	47.62	10.53
Geoffrey Storey	Director, trader	3.57	4.4	0.63	5.79	1.18	0.59	8.96	0.75	20.83	39.76	35.71	10.42
Kevin Ruscitti	Trader	1.23	17.34	0.25	1.65	0.19	0.07	2.99	0.27	2.81	78.31	33.33	10.02
Andrea Ring	Trader*	2.29	NA	0	1.65	0.28	0.09	5.97	1.74	11.97	44.98	46.67	9.86
Paul Thomas	Trader associate	1.08	84.69	0.06	4.13	0.87	0.25	8.96	1.49	11.96	44.58	39.29	9.85
Cooper Richey	Manager	1.64	NA	0	3.31	0.5	0.14	7.46	0.91	14.67	40.96	42.86	9.71
Richard Ring	Employee	0.43	17.43	0.19	0.83	0.09	0.05	1.49	0.11	2.93	76.31	33.33	9.5
Joe Parks	Trader*	2.34	44.92	0.19	1.65	0.37	0.16	4.48	0.57	8.43	50.6	40	9.21
Patrice Mims	Employee*	2.58	7.41	0.76	0.83	0.56	0.33	4.48	0.41	13.31	41.77	40	8.93
Chris Stokley	Employee*	1.08	69.58	0.06	3.31	0.34	0.11	5.97	0.95	13.13	40.96	33.33	8.65
Tom Donohoe	Employee*	0.19	23.12	0.13	0.83	0.03	0.01	0	0.08	4.73	61.85	0	5.77
Monika Causholli	Analyst risk management	0.17	NA	0	0.83	0.03	0	0	0.16	4.51	59.44	0	5.55
Joe Quenet	Trader	0.34	NA	0	1.65	0.09	0.01	0	0.02	5.56	55.42	0	5.43
Albert Meyers	Specialist, trader	0.34	14.55	0.13	0.83	0	0	0	0.27	4.75	55.42	0	5.25



(a)  
FIGURE 2: Continued.

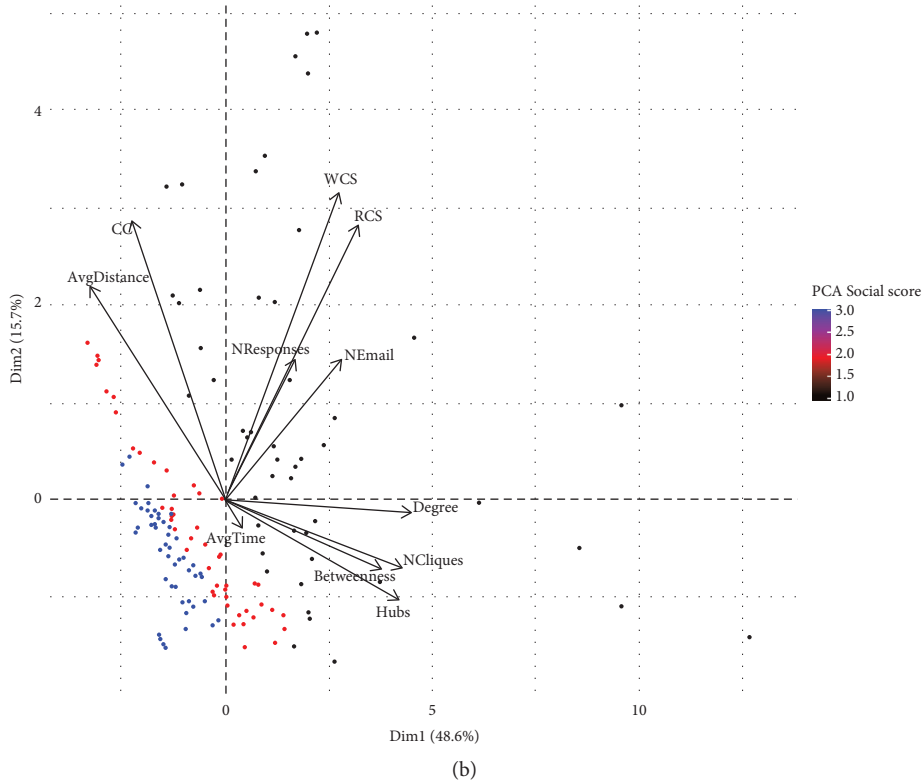


FIGURE 2: Clusters and features by two main principal components according to  $K$ -means (a) and the PCA CorpRank score (b) for Enron. The distance between each feature and the origin is proportional to the quality of its representation on the factor map. Feature definition is in Algorithm 1. (a) Clusters by  $K$ -means. (b) Clusters by PCA CorpRank score.

TABLE 10: CorpRank scores of employees: Enron North American subsidiary. RCS: raw clique score, WCS: weighted clique score, Btw: betweenness, Avg. dist.: average distance, CC: clustering coefficient, score: CorpRank score.

	Position	# e-mail	Avg. time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Tim Belden	Managing Director	100	47.83	100	100	100	100	100	100	100	0	0	77.05
Anna Meher	Vice President	42.33	35.08	7.46	97.87	99.69	99.81	72.73	70.46	98.53	4.17	3.74	57.19
Carla Hoffman	Manager	57.91	25.15	5.97	93.62	99.24	99.51	63.64	38.88	94.67	14.58	15.13	55.37
Debra Davidson	Vice President	41.77	0	2.99	99.57	100	100	77.92	75.16	99.78	2.08	2.31	54.17
Cara Semperger	Trader	54.23	51.96	58.21	57.87	66.58	68.11	74.03	22.4	81.01	27.08	20.31	53.55
Diana Scholtes	Manager	36.26	46.86	38.81	52.34	81.04	84.11	50.65	5.79	65.95	47.92	49.95	51.4
Sean Crandall	Director	31.79	41.22	14.93	38.3	50.53	52.09	46.75	10.81	63.25	45.83	36.38	39.65
Holden Salisbury	Trader	30.83	18.04	43.28	34.89	41.48	42.18	55.84	10.07	65.72	43.75	34.99	38.69
Mark Fischer	Director	26.52	30.19	11.94	31.49	50.15	53.61	36.36	4.2	59.08	54.17	53.12	37.6
Paul Choi	Director	11.42	NA	0	38.3	51.8	49.53	49.35	12.97	70.96	41.67	34.23	37.48
Heather Dunton	Trader	25.16	48.54	22.39	25.11	35.33	36.69	48.05	13.92	67.62	43.75	32.01	36.66
Bill Williams	Trader	19.41	62.53	17.91	20.43	32.41	34.66	38.96	6.78	61.13	52.08	46.37	36.17
Chris Mallory	Trader	13.26	NA	0	23.4	31.34	30.4	27.27	2.68	52.83	60.42	60.26	31.82
Jeff Richter	Director	15.5	84.53	1.49	14.04	17.29	15.88	24.68	3.46	48.16	62.5	55.99	31.73
Greg Wolfe	Vice President	8.15	NA	0	24.68	32.61	30.83	37.66	6.97	58.19	52.08	41.08	31.32

TABLE 10: Continued.

	Position	# e-mail	Avg. time	# Responses	# Cliques	RCS	WCS	Degree	Btw	Hubs	Avg. dist.	CC	Score
Tom Alonso	Director	23	NA	0	17.45	26.9	28.71	25.97	2.67	52.49	60.42	59.05	31.26
Matt Motley	Director	16.69	NA	0	10.64	22.42	23.17	22.08	0.82	45.34	68.75	79.42	30.29
Phil Platter	Trader	15.26	NA	0	22.55	26.48	25	35.06	9.27	53.28	54.17	37.97	29.93
Chris Foster	Director	15.65	30.81	16.42	19.15	23.33	22.47	37.66	6.26	54.97	54.17	43.65	29.81
John Forney	Manager	3.91	100	2.99	13.62	5.37	3.75	29.87	6.5	46.59	58.33	35.74	28.58
Tim Heizenrader	Director	20.29	15.96	5.97	20.85	22.67	23.65	35.06	5.25	49.43	60.42	52.36	28.5
Stacy Runswick	Employee	13.9	54.47	5.97	10.21	9.57	10.68	33.77	3	45.63	64.58	57.3	28.46
John Mallowny	Manager	10.06	NA	0	6.38	16.6	17.82	32.47	1.75	46.57	66.67	69.06	28.2
Julie Sarnowski	Employee	21.57	43.89	2.99	17.87	17.11	17.44	28.57	7	47.25	58.33	37.95	27.63
Kourtney Nelson	Trader	12.22	NA	0	17.02	14.34	13.97	29.87	5.89	49.03	58.33	39.06	27.55
Kim Ward	Manager	10.62	75.02	7.46	1.28	0.29	0.19	9.09	0.1	13.77	89.58	90.93	27.39
Mike Swerzbis	Vice President	20.37	33.51	7.46	9.36	14.59	15.5	32.47	4	46.29	62.5	50.73	27.25
Donald Robinson	Trader	15.97	28.38	2.99	9.36	13.8	16.12	27.27	1.8	41.52	68.75	68.34	26.87
Stewart Rosman	Director	8.31	26.44	5.97	8.09	16.1	16.33	25.97	1.85	40.4	68.75	67.63	26.09
Mike Purcell	Employee	9.98	25.52	26.87	4.26	6.1	7.26	23.38	1.36	35.28	72.92	65.24	25.5
Holli Krebs	Director	2	NA	0	0.43	0.09	0.07	3.9	0.02	8.94	93.75	94.06	24.67
Geir Solberg	Trader	9.03	60.65	1.49	7.66	2.28	1.89	22.08	2.05	30.85	72.92	55.46	24.61
Jeremy Morris	Trader	4.15	NA	0	2.13	0.4	0.28	7.79	0.23	16.54	87.5	81.87	24.55
Robert Anderson	Trader	2.4	NA	0	0.43	0.09	0.07	2.6	0.04	7.15	93.75	94.06	24.43
Smith day	Trader	0	NA	0	0	0	0	0	0	0	100	100	24.34
Lisa Gang	Trader	8.47	NA	0	4.68	5.19	5.17	20.78	1.99	37.14	70.83	59.18	24.27
Caroline emmert	Trader	2.48	NA	0	0.85	0.4	0.32	7.79	0.23	14.53	87.5	84.89	24.23
Steve Swan	Manager	1.12	NA	0	0.43	0.06	0.04	3.9	0.05	8.01	93.75	88.13	24.01
Maria Van Houten	Trader	0.48	NA	0	0.43	0.04	0.03	1.3	0.03	4.17	95.83	92.09	23.88
Monika Causholli	Trader	2.4	NA	0	4.68	1.21	0.86	12.99	0.6	25.52	81.25	76.26	23.86
Jesse Bryson	Trader	4.55	44.99	5.97	6.81	2.66	2.37	22.08	2	33.01	72.92	61.12	23.77
Les Rawson	Trader	5.19	NA	0	6.38	2.64	2.1	22.08	2.06	31.96	72.92	56.55	23.69
Stanley Cocke	Trader	5.19	51.61	1.49	8.51	5.95	4.94	25.97	4.9	43.12	62.5	40.22	23.51
Kelly Axford	Employee	4.95	NA	0	1.28	1.01	1.08	11.69	0.45	22.57	83.33	78.69	23.45
Mark Guzman	Trader	11.58	NA	0	5.53	2.08	1.58	15.58	1.85	28.01	75	58.45	23.42
Cooper Richey	Trader	1.76	NA	0	2.55	0.6	0.39	10.39	0.72	18.97	83.33	70.17	22.64
Ryan Slinger	Trader	7.75	21.9	4.48	3.4	0.52	0.43	15.58	1.02	23.75	81.25	63.59	20.46

assistants appear on the top of our list and the organization chart. Most of the directors and a critical number of managers and specialists are in the first fifteen positions. It is

more difficult to differentiate between employees at the lower level of the organization chart as they may have similar communication patterns. However, it is possible to build a

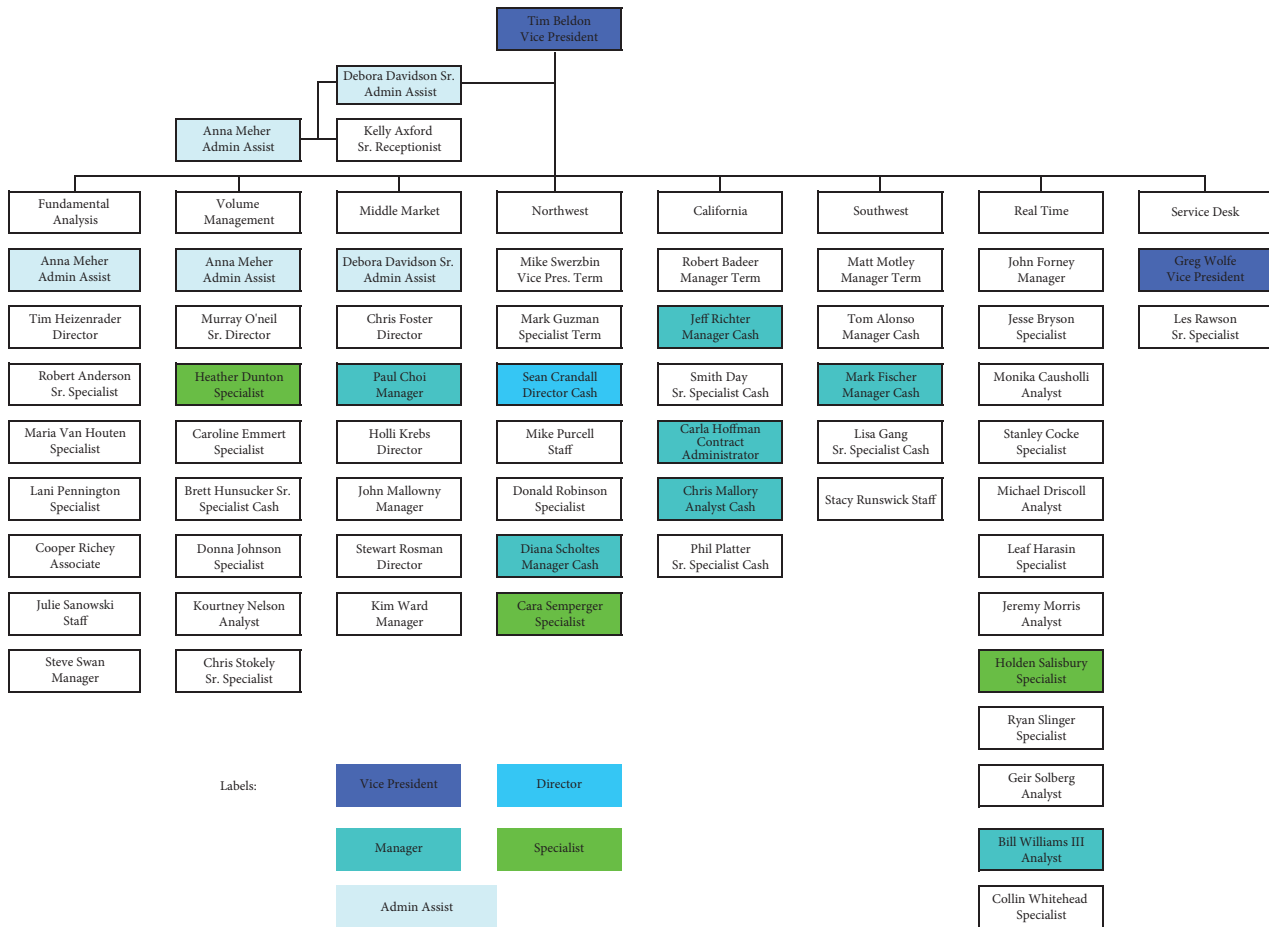


FIGURE 3: Network chart with highlighted results. Graph highlights the fifteen employees (about a third) with the highest CorpRank score.

hierarchy from small groups up as the algorithm recognizes the two or three most essential individuals in every segment (interested researchers can find organization charts of public companies in their annual reports).

### 7. Conclusions and Future Work

This paper shows the CorpRank algorithm’s capacity to capture the organizational hierarchy in a corporation, which may differ from the formal organizational structure. However, this variation may indicate different degrees of leadership or communication that senior managers may use to recognize informal leaders and influencers in an organization. We tested this algorithm on the Enron dataset; however, it can be extended to other corporations or social groups. We also show that our method provides more information than different well-known clustering algorithms.

The next step in this research is to explore organizational changes in a dynamic framework using the Enron dataset. By varying the feature weights, it is possible to use the mentioned parameters to identify the most critical officers in an organization, cluster individuals by their social attributes, organizational characteristics, or compensation, and draw a chart of the actual organizational hierarchy in question.

We could also include some variations in our algorithm to improve our calculations. First, the average response time

can be defined by the order of responses like in [67]. Another approach is to take into account the e-mail usage pattern for each officer and adjust the received time of email to the beginning of the next common e-mail usage time. Consequently, this may improve the average response time calculation as people have different work schedules. A third improvement is aggregating users by percentile or standard deviations of common distributions. Furthermore, rather than ignoring the clique connections, the graph edges could help group users according to their social attributes, organizational characteristics, or compensation.

Future work would extend our model using dynamic processes in social networks [22, 68] and analyze individual business units as we did in the case of the Enron North American West Power Trading division. Chung et al. [68] shows that the temporal signatures of an e-mail thread by officers of a trading unit are consistent with their hierarchical relationship with the president of the division. These results are similar to the results presented in this paper, where we automatically extracted “informal” relationships among employees based on communication patterns. Another direction would be to independently evaluate the hierarchical subsets of senior managers, middle managers, traders, or Enron’s central business units (transportation, wholesale, energy, and broadband services) and integrate them into a single organizational dynamic process. Such

efforts may require reorganizing the e-mail dataset and including new players that initially were eliminated because of quality concerns or ignorance of their roles in the organization. Kan et al. [69] paved the way for the possibility to extend this research to querying evolving graphs, showing that spatiotemporal patterns can be used to find evolving subgraphs that can be related to known real-world events or hierarchies. Dynamic network analysis can be further expanded through the generation of stochastic-based simulated networks, a concept explored by Menges et al. [17] based on social network analysis.

Additionally, our analysis can be extended to other databases of electronic communications such as Twitter or electronic bulletin boards. In forensic analysis, this technology plays a central role as it would be able to detect the importance and influence of different actors. In several criminal cases such as in the case of Enron, the senior managers claim that they were only concentrated on the high-level decisions and policies and that they were not responsible for the implementation of these policies. However, the forensic analysis may reveal their direct or indirect involvement with all parts of the organization.

Another critical area of application is forensic finance. According to the SEC's report [70], the significant increase of the price and volatility of GameStop (GME), a meme stock, in January 2021 was driven by a large number of investors that were investing in this stock rather than by a "short squeeze." Many of these individual investors exchanged the information that led to this rally through WallStreetBets, an electronic forum of Reddit specialized in trading and finance. Using our methodology, it could be possible to identify the leaders that moved the stocks' prices. For instance, Keith Gill played a significant role in GameStop's retail-trading frenzy as many investors followed him at Reddit's WallStreetBets forum [71]. Analysis of Reddit's messages may identify him and his main followers that influenced GME's stock price.

The public availability of a significant amount of documentation collected during the judicial proceedings regarding the Enron managers or other cases offers new insights into the organizational structure previously reserved for insiders. This new information and the generation of specialized models for dynamic social networks could help uncover the relationships that are the backbone of any modern corporation. The automatic application of these models can protect corporations' assets and aid corporations in the early detection of information abuse and misuse.

## Data Availability

The Enron dataset used to support the findings of this study are available from the corresponding author upon request. We used a version of the Enron dataset provided by J. Shetty and J. Adibi. The Enron e-mail dataset database schema and brief statistical report, 2004.

## Disclosure

This paper is an extended version of our paper published in the 9th International Workshop on Knowledge Discovery on the Web, WebKDD 2007 and 1st International Workshop on Social Networks Analysis, SNA-KDD 2007, San Jose, CA, 2007 [27].

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors thank both the National Science Foundation and DARPA for funding this work, particularly, the NSF grant (e-mail Mining Toolkit Supporting Law Enforcement Forensic Analysis from the Digital Government research program, No. 0429323). The authors would like to thank Sinan Aral, participants of the Eastern Economics Association meeting 2019, the 2nd Stevens Internal Workshop on Cognition-Centric Enterprises 2011, and the 9th International Workshop on Knowledge Discovery on the Web 2007.

## References

- [1] P. Chundi, M. Subramaniam, and D. K. Vasireddy, "An approach for temporal analysis of email data based on segmentation," *Data & Knowledge Engineering*, vol. 68, no. 11, pp. 1253–1270, 2009.
- [2] E. Gilbert, "Phrases that signal workplace hierarchy," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 1037–1046, ACM, Washington, DC, USA, February 2012.
- [3] B. Klimt and Y. Yang, "The Enron corpus: a new dataset for email classification research," in *Proceedings of the European Conference on Machine Learning*, Pisa, Italy, September 2004.
- [4] A. McCallum, A. Corrada-Emmanuel, and X. Wang, "The author-recipient-topic model for topic and role discovery in social networks: Experiments with Enron and academic email," in *Proceedings of the NIPS'04 Workshop on 'Structured Data and Representations in Probabilistic Models for Categorization'*, Whistler, Canada, July 2004.
- [5] J.-Y. Yeh and A. Harnly, "Email thread reassembly using similarity matching," in *Proceedings of the Third Conference on Email and Anti-Spam*, Mountain View, CA, USA, July 2006.
- [6] J. Diesner and K. M. Carley, "Exploration of communication networks from the Enron email corpus," in *Proceedings of the Workshop on Link Analysis, Counter Terrorism and Security, SIAM International Conference on Data Mining 2005*, Newport Beach, CA, USA, April 2005.
- [7] J. Diesner, T. L. Frantz, and K. M. Carley, "Communication networks from the enron email corpus "It's Always about the people. Enron is no different"," *Computational and Mathematical Organization Theory*, vol. 11, no. 3, pp. 201–228, 2005.
- [8] R. Cotterill, "Using stylistic features for social power modeling," *Computación Y Sistemas*, vol. 17, no. 2, pp. 219–227, 2013.
- [9] C. Danescu-Niculescu-Mizil, *A computational approach to linguistic coordination*, PhD Thesis, Cornell University, New York, NY, USA, 2012.



- [10] C. E. Priebe, J. M. Conroy, D. J. Marchette, and Y. Park, "Scan statistics on enron graphs," *Computational and Mathematical Organization Theory*, vol. 11, no. 3, pp. 229–247, 2005.
- [11] M. W. Berry and M. Browne, "Email surveillance using non-negative matrix factorization," *Computational and Mathematical Organization Theory*, vol. 11, no. 3, pp. 249–264, 2005.
- [12] P. S. Keila and D. B. Skillicorn, "Structure in the Enron email dataset," *Computational and Mathematical Organization Theory*, vol. 11, no. 3, pp. 183–199, 2005.
- [13] D. Padmanabhan, D. Garg, and V. Varshney, "Analysis of Enron email threads and quantification of employee responsiveness," in *Proceedings of the Text Mining and Link Analysis Workshop on International Joint Conference on Artificial Intelligence*, Hyderabad, India, August 2007.
- [14] T. Elsayed and D. W. Oard, "Modeling identity in archival collections of email: a preliminary study," in *Proceedings of the Third Conference on Email and Anti-spam (CEAS)*, Mountain View, CA, July 2006.
- [15] Z. Bar-Yossef, I. Guy, R. Lempel, Y. S. Maarek, and V. Soroka, "Cluster ranking with an application to mining mailbox networks," in *Proceedings of the ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, pp. 63–74, IEEE Computer Society, Washington, DC, USA, December 2006.
- [16] W. Zhang, A. Ahmed, J. Yang, V. Josifovski, and A. J. Smola, "Annotating Needles in the Haystack without looking: product information extraction from emails," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2257–2266, ACM, Sydney, Australia, August 2015.
- [17] F. Menges, B. Mishra, and G. Narzisi, "Modeling and simulation of e-mail social networks: a new stochastic agent-based approach," in *Proceedings of the 40th Conference on Winter Simulation*, pp. 2792–2800, Miami, Florida, December 2008.
- [18] N. Pathak and J. Srivastava, "Automatic extraction of concealed relations from email logs," in *Proceedings of the International Workshop/School on Network Science*, Bloomington, Indiana, May, 2006.
- [19] A. Chapanond, M. S. Krishnamoorthy, and B. Yener, "Graph theoretic and spectral analysis of enron email data," *Computational and Mathematical Organization Theory*, vol. 11, no. 3, pp. 265–281, 2005.
- [20] J. Shetty and J. Adibi, "Discovering important nodes through graph entropy: the case of Enron email database," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, IL, USA, August 2005.
- [21] X. Hu and H. Liu, "Social status and role analysis of palin's email network," in *Proceedings of the 21st international conference companion on World Wide Web*, pp. 531–532, ACM, Lyon France, April 2012.
- [22] A. S. Maiya and T. Y. Berger-Wolf, "Inferring the maximum likelihood hierarchy in social networks," vol. 4, pp. 245–250, in *Proceedings of the IEEE CSE'09, 12th IEEE International Conference on Computational Science and Engineering*, vol. 4, IEEE, Vancouver, Canada, August 2009.
- [23] N. Memon, H. L. Larsen, D. L. Hicks, and N. Harkiolakis, "Retracted: Detecting hidden hierarchy in terrorist networks: some case Studies," *Intelligence and Security Informatics*, Springer-Verlag, in *Proceedings of the PAISI, PACCF and SOCO '08: Proceedings of the IEEE ISI 2008 PAISI, PACCF and SOCO international workshops on Intelligence and Security Informatics*, pp. 477–489, June 2008.
- [24] C. Kemp and J. B. Tenenbaum, "The discovery of structural form," *Proceedings of the National Academy of Sciences*, vol. 105, no. 31, pp. 10687–10692, 2008.
- [25] L. C. Freeman, "Uncovering organizational hierarchies," *Computational & Mathematical Organization Theory*, vol. 3, no. 1, pp. 5–18, 1997.
- [26] S. Aral and M. W. Van Alstyne, "The diversity-bandwidth trade-off," *American Journal of Sociology*, SSRN eLibrary, vol. 117, no. 1, , pp. 90–171, 2011.
- [27] G. Creamer, R. Rowe, S. Hershkop, and S. Stolfo, "Segmentation and automated social hierarchy detection through email network analysis," in *Proceedings of the Advances in Web Mining and Web Usage Analysis - 9th WEBKDD and 1st SNA-KDD Workshop at KDD 2007, Lecture Notes in Computer Science*, Springer-Verlag, San Jose, CA, USA, August 2007.
- [28] S. J. Stolfo, G. Creamer, and S. Hershkop, "A temporal based forensic discovery of electronic communication," in *Proceedings of the National Conference on Digital Government Research*, San Diego, CA, USA, May 2006.
- [29] S. J. Stolfo, S. Hershkop, C.-W. Hu, W.-J. Li, O. Nimeskern, and K. Wang, "Behavior-based modeling and its application to email analysis," *ACM Transactions on Internet Technology*, vol. 6, no. 2, pp. 187–221, 2006.
- [30] A. Agarwal, A. Omuya, A. Harnly, and O. Rambow, "A comprehensive gold standard for the enron organizational hierarchy," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, ACL '12*, pp. 161–165, Association for Computational Linguistics, Stroudsburg, PA, USA, July 2012.
- [31] J. Maktoubian, M. Noori, M. Amini, and M. Ghasempour-Mouziraji, "The hierarchy structure in directed and undirected signed networks," *International Journal of Communications, Network and System Sciences*, vol. 10, no. 10, pp. 209–222, 2017.
- [32] S. Palus, P. Bródka, and P. Kazienko, "How to analyze company using social network?" in *Knowledge Management, Information Systems, E-Learning, and Sustainability Research* Springer, New York, NY, USA, 2010.
- [33] A. K. Kalia, N. Buchler, D. Ungvarsky, R. Govindan, and M. P. Singh, "Determining team hierarchy from broadcast communications," in *Social Informatics* Springer, New York, NY, USA, 2014.
- [34] H. Nguyen and R. Zheng, "A data-driven study of influences in Twitter communities," in *Proceedings of the Communications (ICC), 2014 IEEE International Conference on Communications*, pp. 3938–3944, IEEE, Sydney, Australia, June 2014.
- [35] M. Hovelynck and B. Chidlovskii, "Multi-modality in one-class classification," in *Proceedings of the 19th international conference on World wide web*, pp. 441–450, ACM, Raleigh Convention Center in Raleigh, NC, USA, April 2010.
- [36] Y. Li and A. Somayaji, *Fine-grained Access Control Using Email Social Network*, CA Labs Research, Chennai, India, 2012.
- [37] R. Michalski, S. Palus, and P. Kazienko, "Matching organizational structure and social network extracted from email communication," in *Business Information Systems* Springer, New York, NY, USA, 2011.
- [38] V. Rivera-Pelayo, S. Braun, U. V. Riss, H. F. Witschel, and B. Hu, *Building Expert Recommenders from Email-Based Personal Social Networks*, Springer, New York, NY, USA, 2013.
- [39] F. Ganjaliyev, "New method for community detection in social networks extracted from the Web," in *Proceedings of the Problems of Cybernetics and Informatics (PCI), 2012 IV*

- International Conference*, pp. 1-2, IEEE, Baku, Azerbaijan, September 2012.
- [40] Y. Wang, M. Iliofotou, M. Faloutsos, and B. Wu, "Analyzing interaction communication networks in enterprises and identifying hierarchies," in *Proceedings of the Network Science Workshop (NSW)*, pp. 17–24, IEEE, New York, NY, USA, June 2011.
- [41] V. Boeva, L. Lundberg, S. M. H. Kota, and L. Sköld, "Evaluation of organizational structure through cluster validation analysis of email communications," *Journal of Computational Social Science*, vol. 1, no. 2, pp. 327–347, 2018.
- [42] D. Jin, M. Heimann, T. Safavi et al., "Smart roles: Inferring professional roles in email networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2923–2933, Anchorage, Alaska, August 2019.
- [43] T. M. G. Tennakoon, *Knowledge discovery from social networks using interaction frequency and user hierarchy*, PhD thesis, Brisbane, Australia, 2020.
- [44] T. M. G. Tennakoon and R. Nayak, "A concise social network representation with flow hierarchy using frequent interactions," in *Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 631–638, Volos, Greece, November 2018.
- [45] S. Behrendt, J. Klier, M. Klier, A. Richter, and K. Wiesneth, "The impact of formal hierarchies on enterprise social networking behavior," in *Proceedings of the International Conference on Information Systems (ICIS) Proceedings*, p. 12, Fort Worth, TX, USA, December 2015.
- [46] J. Klier, M. Klier, A. Richter, and K. Wiesneth, "Two sides of the same coin?-the effects of hierarchy inside and outside enterprise social networks," in *Proceedings of the International Conference on Information Systems (ICIS)*, Singapore, December 2017.
- [47] P. Raut, R. Chawhan, T. Joshi, and P. Kasle, "Classification of power relations based on email exchange," in *Proceedings of the 2020 IEEE International Conference on Computing*, pp. 486–489, Power and Communication Technologies (GUCON), New Delhi, India, December 2020.
- [48] S. Alkhereyf, *Text Classification: Exploiting the Social Network*, Columbia University, New York, NY, USA, 2021.
- [49] O. Rambow, M. Diab, J. Hirschberg, K. McKeown, S. Muresan, and M. Ostendorf, *Detecting relations and anomaly in Text and speech (drats)*, Technical report, Columbia University, New York, NY, USA, 2018.
- [50] M. Nurek and R. Michalski, "Combining machine learning and social network analysis to reveal the organizational structures," *Applied Sciences*, vol. 10, no. 5, p. 1699, 2020.
- [51] D. Barbucha and P. Szyman, "Detecting communities in organizational social network based on e-mail communication," in *Intelligent Decision Technologies* Springer, New York, NY, USA, 2021.
- [52] D. Barbucha and P. Szyman, "Identifying key actors in organizational social network based on e-mail communication," in *Proceedings of the International Conference on Computational Collective Intelligence*, pp. 3–14, Springer, Rhodes, Greece, October 2021.
- [53] T. Magelinski and K. M. Carley, "Community-based time segmentation from network snapshots," *Applied Network Science*, vol. 4, no. 1, pp. 1–19, 2019.
- [54] G. Costa and R. Ortale, "Mining overlapping communities and inner role assignments through bayesian mixed-membership models of networks with context-dependent interactions," *ACM Transactions on Knowledge Discovery from Data*, vol. 12, no. 2, 2018.
- [55] C. Kalinowski and M. Z. KURDI, "A topic-based forensic analysis and visualization of an email network: application to the enron dataset," *The Islamic University Journal of Applied Sciences*, vol. 1, no. 2, pp. 1–22, 2019.
- [56] W. Cohen, "Enron data set," 2004, <https://hdl.loc.gov/loc.gdc/gdcdatasets.2018487913>.
- [57] B. Klimt and Y. Yang, *Introducing the Enron Corpus* CEAS, Cincinnati, Ohio, 2004.
- [58] J. Shetty and J. Adibi, *The Enron email dataset database schema and brief statistical report*, Citeseerx, Princeton, NJ, USA, 2004.
- [59] United States Congress Joint Committee On Taxation, *Report of investigation of Enron Corporation and related entities regarding federal Tax and Compensation issues, and policy recommendations*, 108th Cong 1st. sess. 3 vols. Appendix D, Washington, DC, USA, 2003.
- [60] R. McCullough, "Memorandum related to reading enron's scheme accounting materials," 2004, <http://www.mresearch.com/pdfs/89.pdf>.
- [61] United States Congress Joint Committee On Taxation, "Enron corp. Executive compensation stress test finding," in *Report of Investigation of Enron corporation and related entities regarding federal tax and compensation Issues, and policy recommendations. 108th Cong 1st. sess. 3 vols. Appendix D*, T. Perrin, Ed., GPO, Washington, DC, USA, 2003.
- [62] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, 1973.
- [63] L. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–39, 1979.
- [64] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, 1999.
- [65] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *Journal of Graph Algorithms and Applications*, vol. 10, no. 2, pp. 191–218, 2006.
- [66] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 69, Article ID 026113, 2004.
- [67] S. Hershkop, "Behavior-based email analysis with application to Spam detection," PhD thesis, Columbia University, New York, NY, USA, 2006.
- [68] W. Chung, R. Savell, J. peter Schütt, and G. Cybenko, "Identifying and tracking dynamic processes in social networks," in *Proceedings of the SPIE Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V*, pp. 1–12, Orlando, FL, USA, May 2006.
- [69] A. Kan, J. Chan, J. Bailey, and C. Leckie, "A query based approach for mining evolving graphs," in *Proceedings of the Eighth Australasian Data Mining Conference-*, 101, pp. 139–150, Australian Computer Society, Inc, Melbourne, Australia, December 2009.
- [70] U.S. Securities and Exchange Commission, *Staff report on equity and options market structure Conditions in early 2021*, 2021.
- [71] J.-A. Verlaine and G. Banerji, *Keith Gill drove the GameStop Reddit mania. He talked to the journal*, Wall Street Journal, New York, NY, USA, 2021.