

This is not the final version of this paper. You can find the final version on the publisher web page.

Evaluation of Organization Structure based on Email Interactions

Sebastian Palus¹, Piotr Bródka¹, Przemysław Kazienko¹

¹ *Institute of Informatics, Wrocław University of Technology,
Wybrzeże. Wyspianskiego 27, 50-370 Wrocław, Poland
sebastian.palus@gmail.com, piotr.brodka@pwr.wroc.pl,
przemyslaw.kazienko@pwr.wroc.pl*

Abstract

In every company people call each other, send emails, text messages, chat on forum, co-author documents, meet at events, etc. In other words, they collaborate with each other creating and maintaining mutual relationships in one complex social network. We can use this social network to evaluate company structure and based on this evaluation recommend some changes which have to be made within company to improve its structure. The idea for evaluation process together with example of its usage are presented in the paper.

Keywords:

social network, social network analysis, company structure, Enron

1. Introduction

The growing number of opportunities and ways people can communicate and exchange information within an organization provide us with a previously unknown way to evaluate company's structure [6,13]. The data extracted from email services, phone calls, documents co-authoring, and other communication systems or common activities allow to create social networks which contain information about humans interaction and collaboration [7,12]. On the other hand all companies have always sought to obtain the best and most effective structure. By utilizing, already existing within company, social network we can help to archive this goal. In this paper it is presented how to do this. First the analysis of Enron company (Section 2) and its hierarchical structure (Section 3) is presented. Then the Enron e-mail data set (Section 4) together with social network extracted from this set (Section 5) are presented. In Section 6 the social network based approach to evaluate organizational structure is described and next the example of its usage on the Enron company is presented in Section 7. At the end the discussion (Section 8) and conclusion (Section 9) are presented.

2. Enron Case

Enron was formed in 1985 under the direction of Kenneth Lay [3]. The company was based in Houston, Texas. Within 15 years Enron became one of the nation's biggest companies in revenue by buying electricity from generators and selling it to consumers. Enron identified areas where energy needs were higher than energy capacities, built power plants and sold them before their value diminished. In 2002, there were 21,000 people employed in Enron in over 40 countries [5].

This is not the final version of this paper. You can find the final version on the publisher web page.

From 1985 on, Arthur Andersen had been Enron's auditor. Andersen earned millions of dollars from accounting and internal and external consulting services for Enron, which was one of Andersen's largest clients worldwide.

In 1999, Enron officials began to separate losses from equity and derivative trades into "special purpose entities" (SPE) - partnerships that were excluded from the company's net income reports. The systematic omission of negative balance sheets and income statements from SPEs in Enron's reports resulted in an off-balance-sheet-financing system [15].

In December of 2000, president and chief operating officer Jeffrey Skilling took over the position of chief executive from Kenneth Lay. Lay remained chairman while the Enron stock hit a 52-week high of \$84.87. In August 2001, Skilling surprisingly resigned. Lay was named as Enron's chief officer and CEO again in 2001 [11]. In the same month Sherron Watkins, Enron's Vice-President of Corporate Development who became famous as Enron's whistle-blower, wrote an anonymous letter to Lay in which she accused Enron of possible fraud and improprieties such as the SPEs [17]. Andersen knew about these informations.

In October 2001 the losses transferred from Enron to the SPE's exceed \$618 million and Enron publicly reported this amount as net loss for the third quarter. By the end of the year Enron disclosed a reduction of \$1.2 billion in the value of shareholders' stake in the company. One of the people associated with the crash was Andrew Fastow, chief financial officer, who had supported Enron in inflating profits and hiding debts [15].

On October 31, 2001, the Securities and Exchange Commission (SEC) started an inquiry into Enron. Enron announced that the SEC investigation revealed that the amount of losses for the previous five years was actually \$586 million. The market reacted with a fast drop of the value of Enron's shares to levels below \$1 in November 2001. Being forced to transfer stocks in order to satisfy the losses, Enron became insolvent and filed for bankruptcy in December 2001. The fallout and investigations into the Enron collapse continued throughout 2002. Lay resigned as chairman and CEO in January of 2002, and less than two weeks later from the board [1].

Long before Enron's official insolvency, Andersen had possessed knowledge of Enron's organizational situation and financial performance but did not communicate the information to the public [15]. Andersen and Enron intentionally categorized hundreds of millions of dollars of shareholders equity that were a decrease as an increase. Andersen, who did some of Enron's internal bookkeeping, advised Enron not to refer to charges against the third quarter income of 2001 as non-recurring, but did not make this information available for the public. In 2000 Andersen's internal Senior Management already had rated Enron lower than they evaluated the client publicly. Before Enron released its notice of net loss, Andersen retained a New York based law firm from handling further Enron-related issues and took over all legal matters regarding Enron. In late October 2002, Andersen instructed Enron to destroy documentation related to Enron.

Andersen was indicted for altering, destroying and concealing Enron-related material and persuading others to do the same in March 2002, convicted of obstruction in June 2002, and received a probationary sentence and a fine of \$500,000 in October 2002. In 2002 Andersen got banned from auditing public companies.

Lay, Fastow and former top aid Michael Kopper appeared before Congress in February of 2002; all three of them invoked the Fifth Amendment [5]. Skilling testified twice before Congress the same month, stating that he was unaware of any accounting problems. Fastow was indicted in October 2002. Ben Glisan Jr., a former Enron treasurer, pleaded guilty to conspiracy in September 2003, and became the first former Enron executive being imprisoned [1]. Fastow pleaded guilty in January 2004 [5]. His wife, Lea Fastow, and seven former Enron executives also got charged. In February 2004 Skilling got charged with fraud, conspiracy, filing false statements to auditors and insider trading [11]. In July of 2004 Lay surrendered to the FBI and was accused of participating in a conspiracy to

This is not the final version of this paper. You can find the final version on the publisher web page.

manipulate Enron's quarterly financial results, making false and misleading public statements about Enron's financial situation, omitting facts necessary to make financial statements accurate and fair, civil fraud, and insider trading.

In March of 2003 Enron announced a plan to emerge from bankruptcy as two separate companies. In July the company filed a reorganization plan stating that most creditors would receive about one-fifth of the \$67 billion they were owed.

3. Enron Hierarchy Structure

The Enron Hierarchy Structure is not publicly available. However, there are sources which can provide information about names of job positions of many employees and their department or division. In [18] there is an Microsoft Excel file with a list of over 160 employees and their job title. Many of them do not exist in Enron Corpus, though. Using this list and charts available in [14], four groups from Enron North American West Power Traders were chosen where it was able to distinguish levels of hierarchy by assigning them to job titles. An assumption was made that the job ranking looks as presented in Fig. 1., where Analysts, Specialists and Staff are at the same level. Note the colors assigned here, they will be used later to point out anomalies.

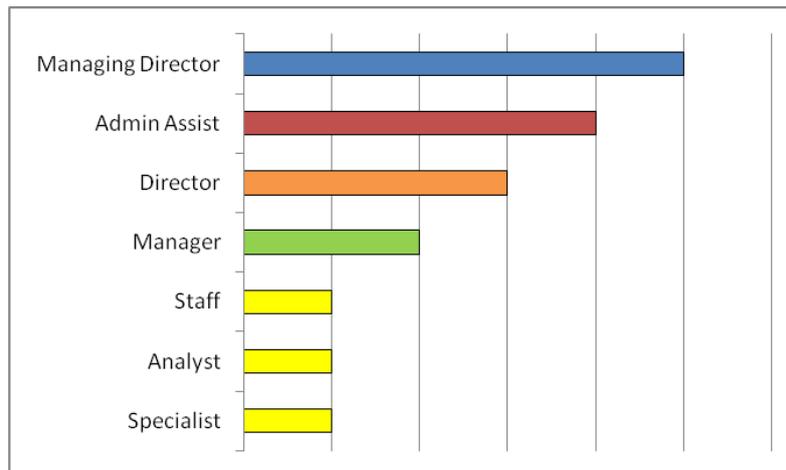


Fig. 1. Job titles hierarchy

It is impossible to find direct relationships of superior-inferior from such limited data. The knowledge of hierarchy levels is however complete enough to perform some tests. The organization chart which was used for analysis is presented on Fig. .

This is not the final version of this paper. You can find the final version on the publisher web page.

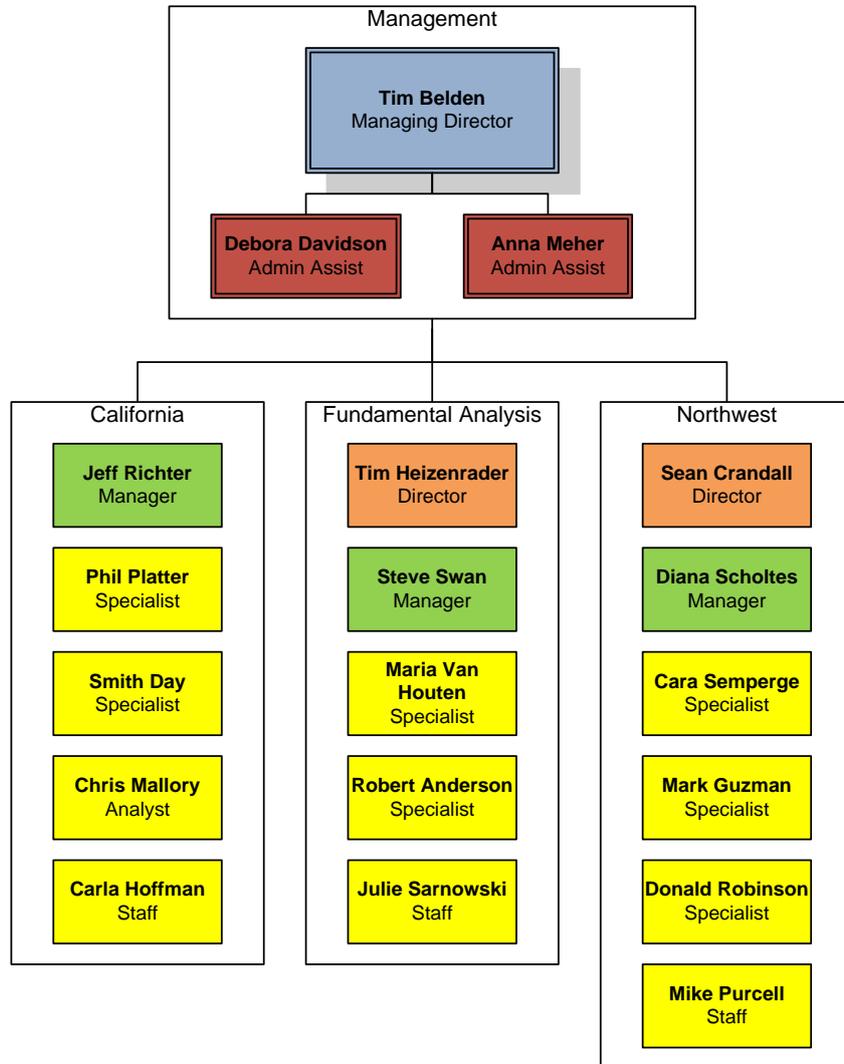


Fig. 2. Part of Enron hierarchy used for analysis

4. Enron Email Dataset Specification

Enron Corpus is a set of mail messages, each email in separate file [2]. The messages are grouped in maildir folders by their owner and organized into folders such as Inbox, Sent, Trash etc. Sample dataset structure is presented in Fig. 3.

This is not the final version of this paper. You can find the final version on the publisher web page.

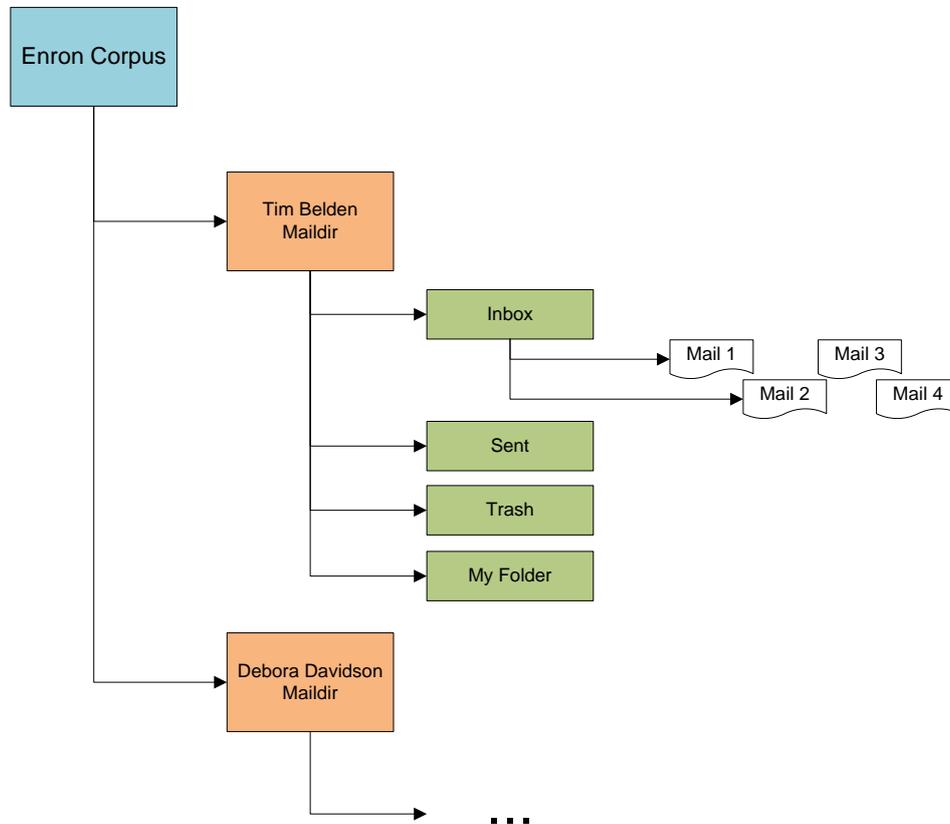


Fig. 3. Enron Corpus Structure

Every message is in a standard mail format [19] and contains elements such as:

- Message-id;
- Date;
- From;
- To (CC, BCC);
- Subject;
- X-Fields with user-friendly Active Directory names (X-To, X-CC etc.);
- Message body;

The whole email dataset was made public after the Federal Energy Regulatory Commission during its investigation on the Enron Scandal. The email dataset had a number of integrity problems which have been corrected by researchers at MIT and SRI International. The dataset does not include attachments, and some messages have been deleted as part of a redaction effort due to requests from affected employees. Invalid email addresses were converted to the form user@enron.com whenever possible (i.e., recipient is specified in some parse-able format like “Doe, John” or “Mary K. Smith”) and to no_address@enron.com when no recipient was specified.

It contains data from 150 Enron employees, mostly Senior Management. There are total number of 517,430 messages.

This is not the final version of this paper. You can find the final version on the publisher web page.

5. Enron Communication Social Network

The social network of Enron is extracted from the email dataset. Using all emails in the dataset, one can construct an undirected graph, where vertices represent accounts and edges represent communication between two accounts [3,4]. Then several measures for each node are applied [16,20]. The approach presented in [14] was used, where social score is computed from:

- a) Emails count – number of email the user has sent and received.
- b) Average response time - the time elapsed between a user sending an email and later receiving an email from that same user. An exchange of this nature is only considered a “response” if a received message succeeds a sent message within three business days.
- c) Response score – a combination of the number of responses and average response time.
- d) Number of cliques – the number of maximal complete subgraphs that the account is contained within.
- e) Raw clique score – a score computed using a size of the given account’s clique set. Bigger cliques are worth more than smaller ones, importance increases exponentially with size.
- f) Weighted clique score – a score computed using the importance of the people in each clique, which is computed strictly from the number of emails and the average response time.
- g) Centrality Degree - count of the number of ties to other actors in the network.
- h) Clustering coefficient - likelihood that two associates of a node are associates themselves.
- i) Mean of shortest path length from a specific vertex to all vertices in the graph.
- j) Betweenness centrality - reflects the number of people who a person is connecting indirectly through their direct links.
- k) „Hubs-and-Authorities” importance – refers to the algorithm proposed in [9].

Above metrics are then weighted and normalized to a [0, 100] scale.

6 Evaluation of Organizational Structure based on the Social Network

For each employee in the corporate hierarchy it is possible to find people who are higher or lower in the hierarchy. The Hierarchical Position (HP) is a measure that shows the importance of an employee within a company. For each user u_i in a company C there is a sum of hierarchical differences D between u_i and every user u_j in the company divided by the number of other users. This measure can be computed either globally, where all employees in the company are taken into account (1) or at the branch level B , where only a branch of the hierarchy tree where the user is situated is being investigated (2).

$$HP^{global}(u_i) = \frac{\sum_{u_j \in C \wedge u_i \neq u_j} D(u_i, u_j)}{m_C - 1} \quad (1)$$

$$HP^{branch}(u_i) = \frac{\sum_{u_j \in B \wedge u_i \neq u_j} D(u_i, u_j)}{m_B - 1} \quad (2)$$

This is not the final version of this paper. You can find the final version on the publisher web page.

The hierarchical difference $D(x,y)$ can be calculated in two ways: as a sign function (3) or as a direct difference between assigned hierarchy levels $L(u)$, where $L(u) \in N^+$, $L(u)=1$ is the highest level and $L(x)=L(y)+1$ when x is one level higher than y .

$$D^{\text{sgn}}(x, y) = \begin{cases} 1, & \text{if } x \text{ is higher in the hierarchy than } y \\ 0, & \text{if } x \text{ and } y \text{ are at the same level of the hierarchy} \\ -1, & \text{if } x \text{ is lower in the hierarchy than } y \end{cases} \quad (3)$$

$$D^{\text{level}}(x, y) = L(x) - L(y) \quad (4)$$

At first, the Kendall's rankings comparison method was used [5]. To compare two rankings we have to compare the positions in each pair in both rankings. If the position of node A is related to the position of node B in both rankings monotonically in the same direction (lower or higher in the both hierarchies) then this pair is well correlated. It is assumed that when the level in hierarchy is the same within the pair, then it does not matter whether they are in different positions in the second ranking. Kendall's τ rank correlation coefficient is a value from a $[-1,1]$ scale, where 1 means that rankings are perfectly correlated and -1 means that they are completely different. It is impossible to distinguish the importance of departments, e.g. whether the Director of Northwest is higher in the hierarchy than the Director of Fundamental Analysis. Thus, analyses were not performed globally, but locally at department level.

7. Evaluation for the Enron Company

The list of Enron employees sorted by their Social Score is presented on Table 1. The HP measure (see Section 6) and Position column indicates official hierarchy structure. It can be seen very clearly that Social Scores of the Management is far higher than the others.

Name	Surname	Position	Level	HP				Degree	Betweenness	Hubs	Clustering	SocialScore
				sign		level						
				global	branch	global	branch					
Tim	Beldon	Managing Director	1	1.00	1,00	3,28	3,28	83	370.35	0.04	0.40	75.68
Debora	Davidson	Admin Assist	2	0.83	0,83	2,28	2,28	66	278.35	0.04	0.41	63.51
Anna	Meher	Admin Assist	2	0.83	0,83	2,28	2,28	62	260.94	0.04	0.42	62.84
Carla	Hoffman	Staff	5	-0.44	-0,57	-0,94	-1,57	55	143.98	0.04	0.49	61.67
Cara	Semperger	Specialist	5	-0.44	-0,63	-0,94	-1,63	63	82.96	0.03	0.52	53.68
Diana	Scholtes	Manager	4	0.33	0,00	0,11	-0,50	45	21.44	0.03	0.70	53.31
Sean	Crandall	Director	3	0.61	0,25	1,17	0,63	42	40.04	0.03	0.62	43.64

This is not the final version of this paper. You can find the final version on the publisher web page.

Name	Surname	Position	Level	HP				Degree	Betweenness	Hubs	Clustering	SocialScore
				sign		level						
				global	branch	global	branch					
Tim	Heizenrader	Director	3	0.61	0,14	1,17	0,43	33	19.45	0.02	0.71	35.56
Donald	Robinson	Specialist	5	-0.44	-0,63	-0,94	-1,63	27	6.67	0.02	0.81	33.03
Jeff	Richter	Manager	4	0.33	0,14	0,11	-0,43	25	12.80	0.02	0.74	32.53
Julie	Sarnowski	Staff	5	-0.44	-0,71	-0,94	-1,86	28	25.94	0.02	0.63	32.14
Mike	Purcell	Staff	5	-0.44	-0,63	-0,94	-1,63	24	5.02	0.02	0.79	30.36
Chris	Mallory	Analyst	5	-0.44	-0,57	-0,94	-1,57	27	9.92	0.02	0.76	30.19
Phil	Platter	Specialist	5	-0.44	-0,57	-0,94	-1,57	33	34.34	0.02	0.63	27.90
Robert	Anderson	Specialist	5	-0.44	-0,71	-0,94	-1,86	8	0.15	0.01	0.96	20.06
Smith	Day	Specialist	5	-0.44	-0,57	-0,94	-1,57	6	0.00	0.01	1.00	20.00
Mark	Guzman	Specialist	5	-0.44	-0,63	-0,94	-1,63	18	6.84	0.01	0.75	19.97
Steve	Swan	Manager	4	0.33	-0,14	0,11	-0,71	9	0.20	0.01	0.93	19.55
Maria	VanHouten	Specialist	5	-0.44	-0,71	-0,94	-1,86	7	0.11	0.01	0.95	19.44

Table 1. Social measures for Enron employees sorted by SocialScore

The diagram of Hierarchical Position should be descending, but there are deep structural holes in Fig. 4.

This is not the final version of this paper. You can find the final version on the publisher web page.

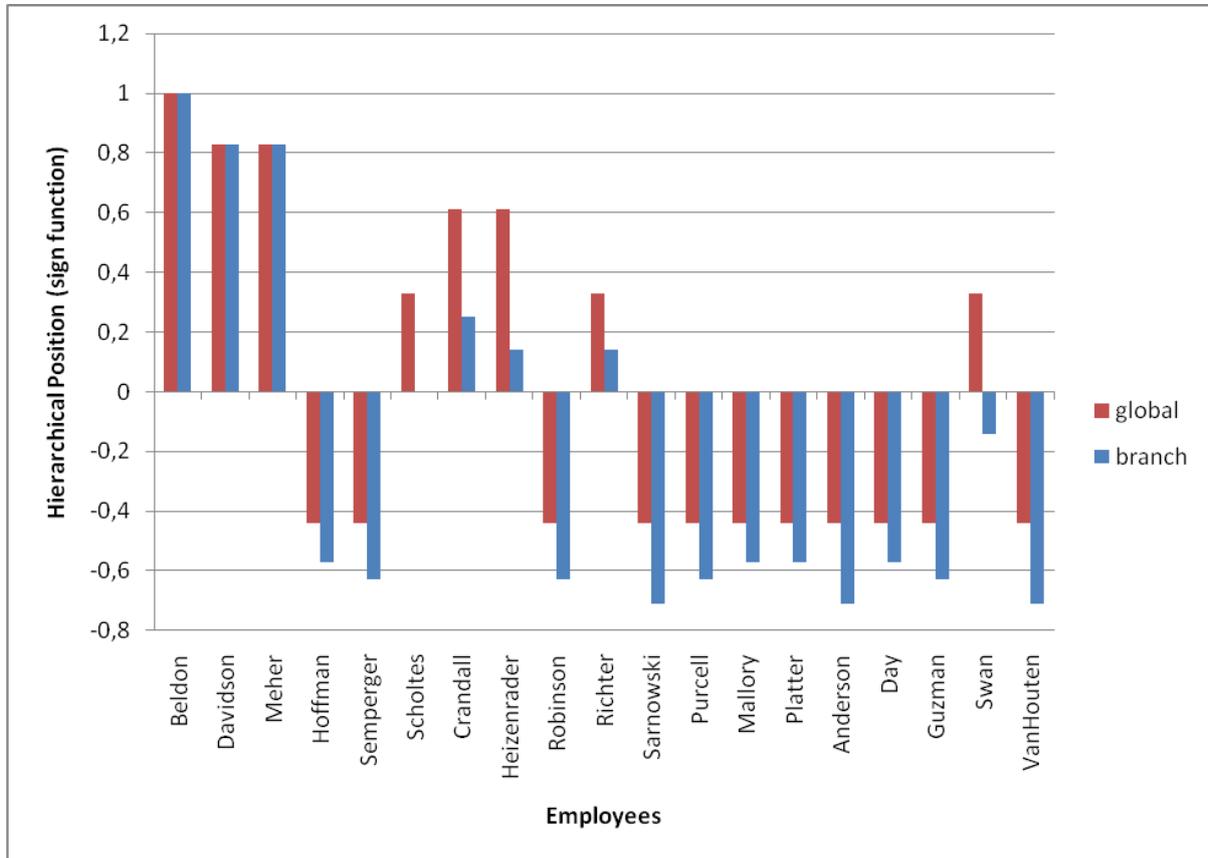


Fig. 4. Hierarchical Positions (sign function) of Enron employees sorted by Social Score

This is not the final version of this paper. You can find the final version on the publisher web page.

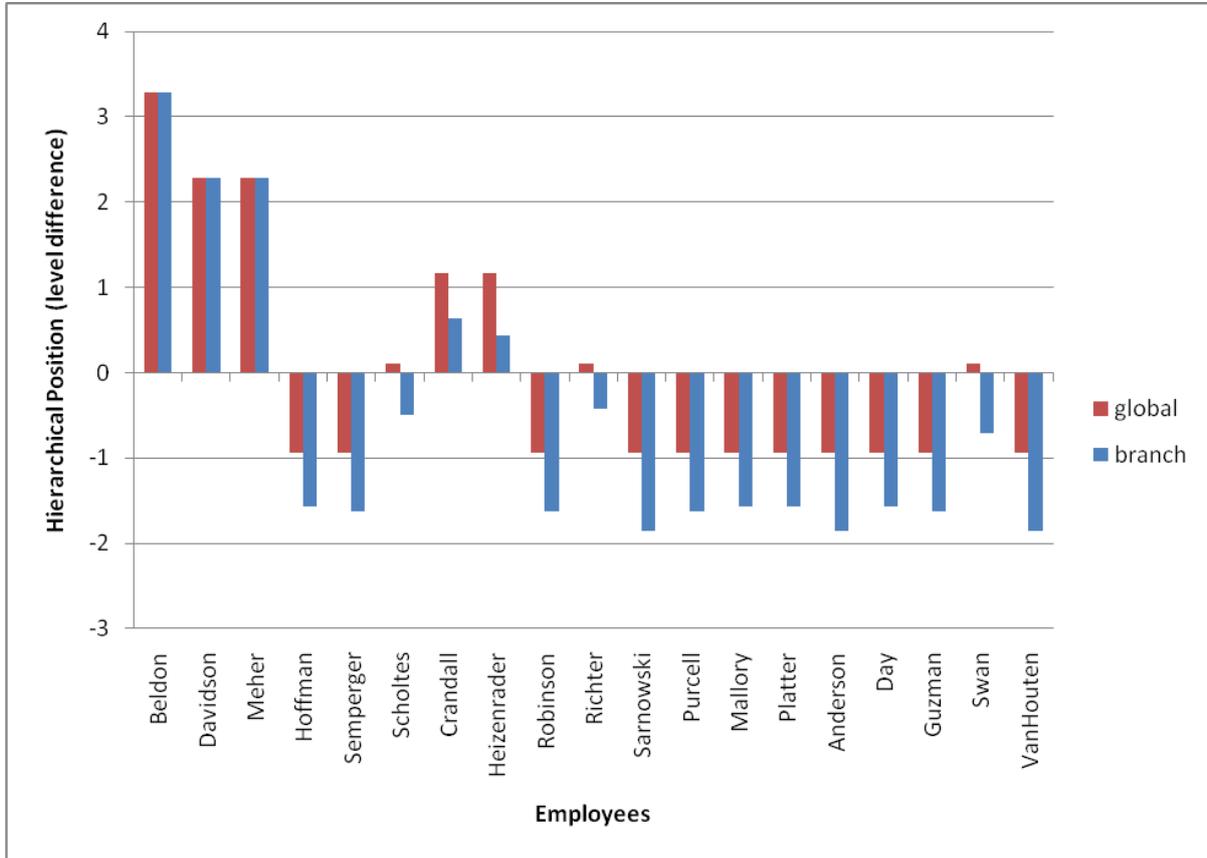


Fig. 5. Hierarchical Positions (level difference) of Enron employees sorted by Social Score

On Table 2 there is a summary of Kendall correlation coefficient between the official hierarchy (ordered by HP) and one derived from social network (ordered by Social Score) for chosen departments.

Department	Kendall's τ
Management (official vs SN)	1,0
California (official vs SN)	0,8
Fundamental Analysis (official vs SN)	0,6
Northwest (official vs SN)	0,6

Table 2. Kendall correlation coefficient for each department between official hierarchy and social network

On Fig. 6. – Fig. 9. there are visible differences between extracted social network and organizational hierarchy. Order of the employees and colors of rectangles reflect the organizational hierarchy, while their width is their Kendall correlation coefficient value.

This is not the final version of this paper. You can find the final version on the publisher web page.

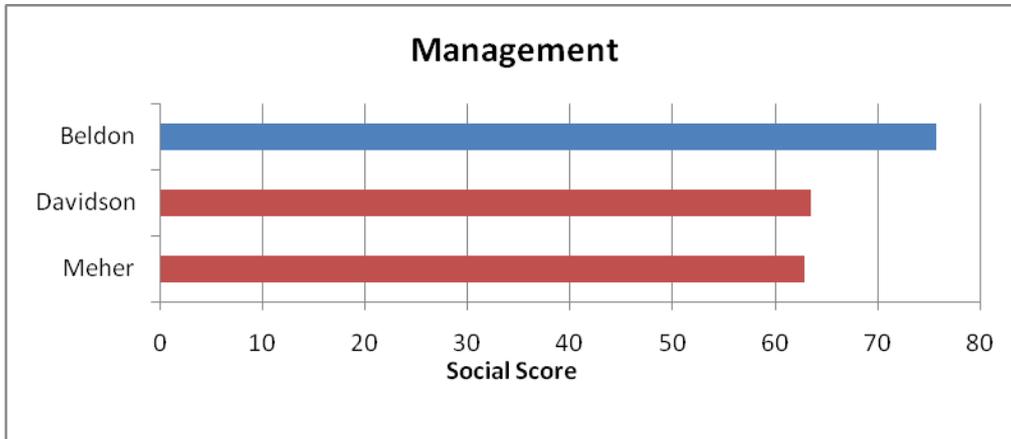


Fig. 6. Social Scores in Management

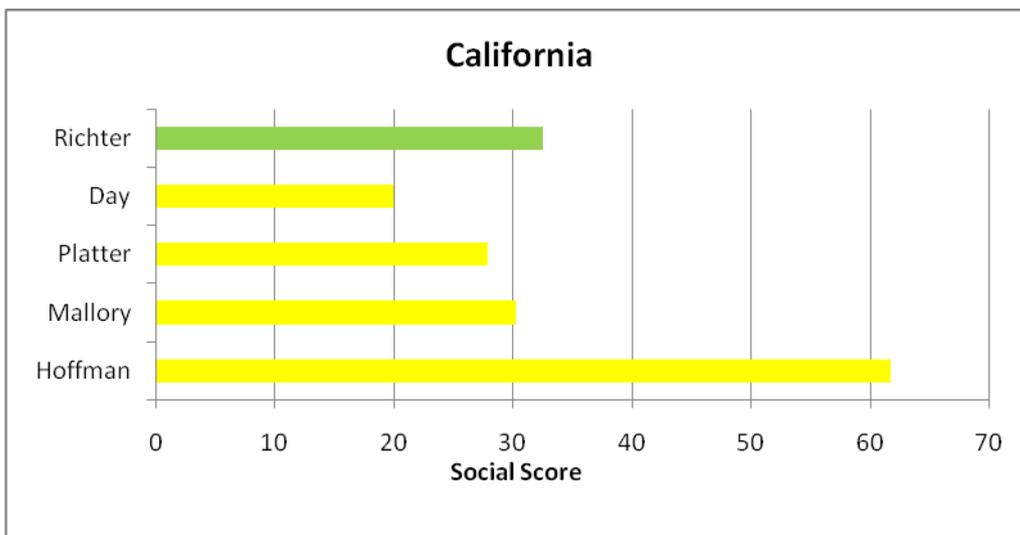


Fig. 7. Social Scores in California

This is not the final version of this paper. You can find the final version on the publisher web page.

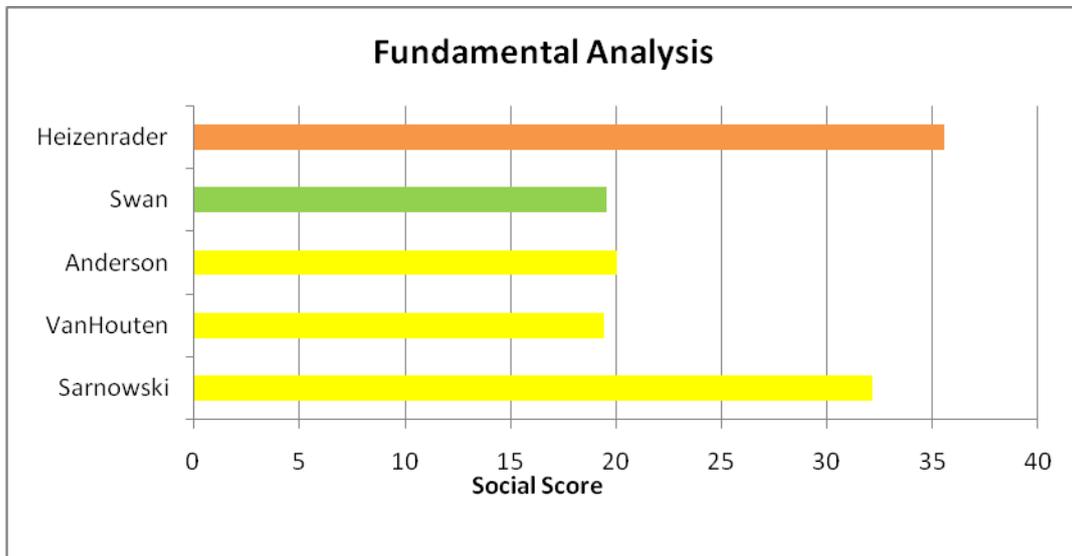


Fig. 8. Social Scores in Fundamental Analysis

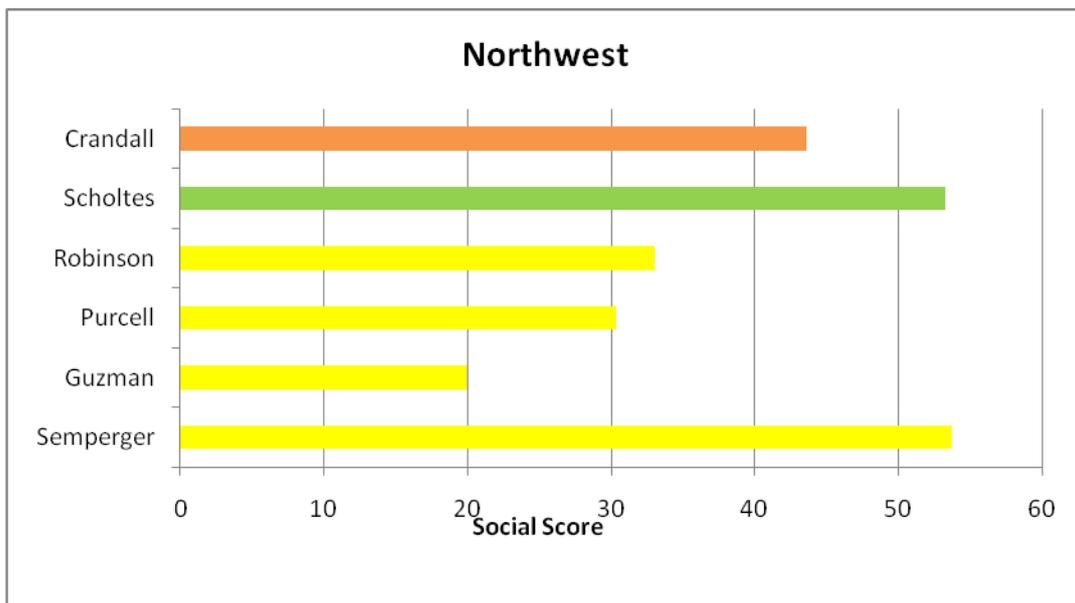


Fig. 9. Social Scores in Northwest

8. Discussion

The main problem with the Enron dataset is the lack of information about direct hierarchy structure. All analysis were performed with assumptions of the structure of levels and departments. However,

This is not the final version of this paper. You can find the final version on the publisher web page.

the analysis shows the rankings are very similar with Kendall rank over 0.6 with Management department perfectly identical (Kendall rank of 1).

By analyzing the differences between social scores within each department/division we can express additional conclusions. In California (Fig. 7) extracted network, a Staff member Carla Hoffman is twice as important as any other employee of this division, even her manager Jeff Richter. In fact, the manager is found to have similar social score to other members who were supposed to be under him in organizational hierarchy. This may mean that Carla is a hidden leader and real manager is not more important than other employees. A proposed solution for the company would be to consider swapping positions of these two.

Even though in Fundamental Analysis (Fig. 8) Kendall's rank is 0.2 lower, the situation is very similar to the one in California. Staff member, Julie Sarnowski has much higher social score than other employees at her level, higher even than manager Steve Swan, who is additionally little lower than other lowest level employee, specialist Robert Anderson. Team director, Tim Heizenrader is on the top of both rankings which means he is a true leader. A proposed solution for the team would be to consider swapping positions of manager Steve Swan and staff member Julie Sarnowski.

In Northwest (Fig. 9) there is a next example of an employee from the bottom of the hierarchy who has much higher importance in terms of social score than other people at the same level. This time, a specialist Cara Semperger with social score of 53.68 is the most important employee in Northwest. Similar score is assigned to the manager, Diana Scholtes. Sean Crandall, the director is a little lower in the ranking, but not as low as other Northwest employees, though. A hierarchy structure which would reflect the social ranking would consist of two directors: Scholtes and Semperger, manager: Sean Crandall and other members with no changes.

An interesting fact is that all employees who are lower in the hierarchy than it comes from social network are women. There are 7 women among 19 analyzed employees and there are 5 female workers in the top 6 of the social ranking (Table 1), while 4 have been classified as lowest level employees according to the hierarchy. There are three possible reasons of such case. First is that a wrong assumption has been made while ranking job titles. Secondly, there can be a simple but important reason that women are underestimated and should occupy higher company positions. Last, but not less probable, is that women are more likely than men to gossip [10] and this fact is disrupting the process of proper social network extraction. It is very likely that the real reason is combined of these three.

9. Conclusions

Social network approach to manage described problem is capable to significantly improve the Human Resources efficiency by detection of hidden anomalies in the corporate hierarchy and making communication between employees more effective and easier. By analyzing semantics of email messages exchanged within the corporate network we are able to identify individuals with "the hidden knowledge".

Overall, social network approach to the problem of corporate management appears to be very helpful, however, all analysis need to be well interpreted to improve performance and social health of the company. Actually, all interpretations of comparison results can be dependent on company character, nationality of employees or the environment. For the results be reliable, a perfect company to calibrate the system would be needed. A company, where all social leaders are directors and there are no hidden knowledge flow between groups and employees. However, in real life this is not possible. Hence, while no perfectly organized company is found, the system have to rely on the

This is not the final version of this paper. You can find the final version on the publisher web page.

experts knowledge and instincts. This is only a tool. Still, human resources have to be managed by humans.

A collaborative environment with corporations should be created to develop a comprehensive solution to make management of human resources in huge companies easier and more efficient. More real-world data is needed to ensure the correctness of performed analyses, not forgetting about the privacy issue. Fresh and professional opinions of experienced HR specialists will be necessary to make more professional interpretations of social measures and comparisons.

A lot of work is needed to make this system fully functional and appropriately helpful for the corporate management. However, it is an excellent starting point for further analyses and implementations. Hopefully, it will become an essential tool for the big companies HR departments and will significantly improve the internal knowledge flows and employees' efficiency.

Acknowledgments.

This work was supported by The Polish Ministry of Science and Higher Education, the development project, 2009-11.

10. References

- [1] A Chronology of Enron Corp. NewsMax Wires. Retrieved from: <http://www.newsmax.com/archives/articles/2004/7/8/110332.shtml>. 2004
- [2] Cohen W. (2009) Enron Email Dataset. [Online]. <http://www.cs.cmu.edu/~enron/>
- [3] Diesner J., Carley K. "Exploration of communication networks from the enron email corpus". Proceedings of Workshop on Link Analysis, Counterterrorism and Security, 2005.
- [4] Diesner J., Frantz T.L., Carley K. "Communication networks from the enron email corpus". *Journal of Computational and Mathematical Organization Theory*, no. 11, pp. 201:228, 2005.
- [5] Enron Scandal at a Glance. BBC News. Retrieved from: <http://news.bbc.co.uk/1/hi/business/1780075.stm>. 2002
- [6] Kascarone R., Paauwe K., Zupan N. "HR Practices, Interpersonal relations and intrafirm knowledge transfer in knowledge-intensive firms: a social network perspective". *Human Resource Management*, vol. 48, no. 4, pp. 615-639, 2009.
- [7] Kazienko P., Musiał K., Zgrzywa A. "Evaluation of Node Position Based on Email Communication". *Control and Cybernetics*, vol. 38, no. 1, pp. 67-86, 2009.
- [8] Kendall M.G., Rank Correlation Methods. Oxford: Oxford, 1948.
- [9] Kleinberg J. "Authoritative sources in a hyperlinked environment". *Journal of the ACM*, vol. 46, 1999.
- [10] Levin J., Arluke A. "An exploratory analysis of sex differences in gossip". *Sex Roles*, vol. 12, no. 3-4, pp. 281-286, February 1985.
- [11] Lights out at Enron. CBSNews.com. Retrieved from: <http://www.cbsnews.com/stories/2003/02/06/60minutes/main539719.shtml>. 2003
- [12] Musiał K., Juszczyszyn K. "A method for evaluating organizational structure on the basis of social network analysis". *Foundations of Control and Management Sciences*, vol. 9, pp. 97-108, 2008.
- [13] Palus S., Kazienko P., "Social Network Analysis in Corporate Management" in *Multimedia & Network Information Systems MISSI '10*, Wrocław, Advances in Intelligent and Soft Computing, Springer, 2010, accepted.
- [14] Rowe R., Creamer G., Hershkop S., Stolfo S.J. "Automated Social Hierarchy Detection through Email Network Analysis". Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 109-117, 2007.
- [15] Sanborn R. Enron. Retrieved from: <http://www.hoylecpa.com/cpe/lesson001/Lesson.htm>. 2004
- [16] Scott J., *Social network analysis: A handbook*, 2nd ed. London: Sage, 2000.
- [17] Sherron Watkins eMail to Enron Chairman Kenneth Lay. Retrieved from: www.itmweb.com/f012002.htm. 2002

This is not the final version of this paper. You can find the final version on the publisher web page.

- [18] Shetty J., Adibi J. Ex employee status report. Retrieved from:
http://www.isi.edu/~adibi/Enron/Enron_Employee_Status.xls. 2004
- [19] The Internet Society. RFC 2822 - Internet Message Format. Retrieved from:
<http://www.faqs.org/rfcs/rfc2822.html>. 2001
- [20] Wasserman S., Faust K., Social network analysis: Methods and applications. New York: Cambridge University Press, 1994.