How to Analyze Company Using Social Network?

Sebastian Palus[1], Piotr Bródka[1], Przemysław Kazienko[1]
[1] Wroclaw University of Technology, Wyb. Wyspianskiego 27,
50-370 Wroclaw, Poland
{sebastian.palus, piotr.brodka, przemyslaw.kazienko}@pwr.wroc.pl

Abstract. Every single company or institution wants to utilize its resources in the most efficient way. In order to do so they have to be have good structure. The new way to analyze company structure by utilizing existing within company natural social network and example of its usage on Enron company are presented in this paper.

Keywords: social network, social network analysis, company structure

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 [5]. The data extracted from email services, phone calls, other communication systems or common activities allow to create social networks which contain information about humans interaction and collaboration. 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 achieve this goal. In this paper it is presented how to do this using the Enron company as an example.

2 Enron.

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 [8] there is an 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 [6], 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.

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. .

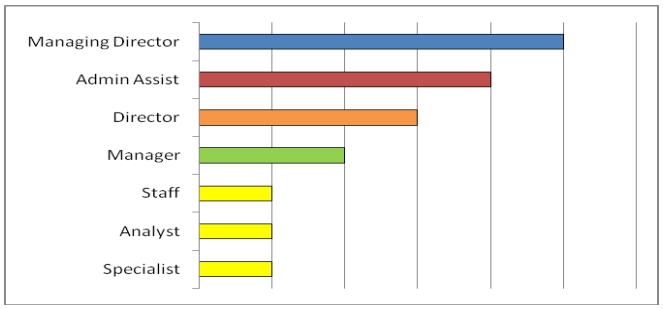

Fig. 1. Job titles hierarchy

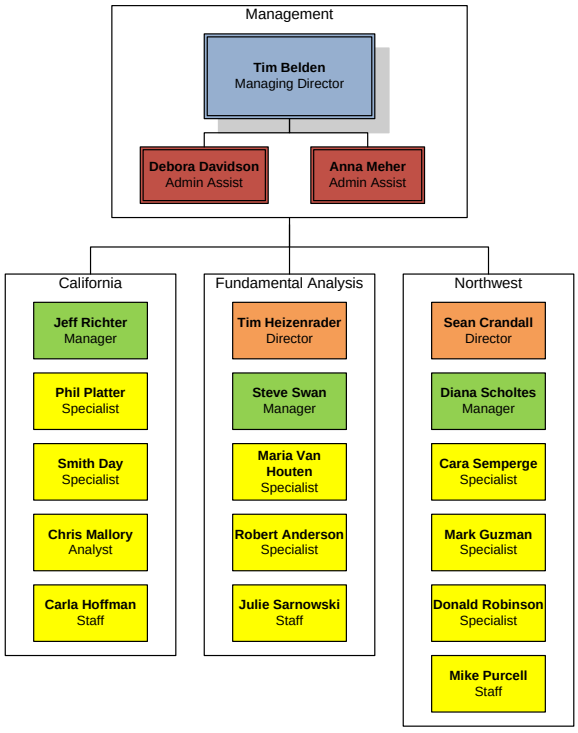

Fig. 2. Part of Enron hierarchy used for analysis

2.1 Enron Email Dataset Specification

Enron Corpus is a set of mail messages, each email in separate file [1]. The messages are grouped in *maildir* folders by their owner and organized into folders such as Inbox, Sent, Trash etc.

Every message is in a standard mail format [9] 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.

2.3 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. Then several measures for each node are applied [7,10]. The approach presented in [6] 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 with 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 [3].

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

3 Organizational Structure Evaluation

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 ui in a company C there is a sum of hierarchical differences D between ui and every user uj in the company divided by the number of other users.

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

The hierarchical difference *D(x,y)* is computed as follows:

$$D(x,y) = \begin{cases} 1, & if\ x\ is\ higher\ in\ the\ hierarchy\ than\ y \\ 0, & if\ x\ and\ y\ are\ at\ the\ same\ level\ of\ the\ hierarchy \\ -1, & if\ x\ is\ lower\ in\ the\ hierarchy\ than\ y \end{cases} \quad (2)$$

At first, the Kendall's rankings comparison method was used [2]. 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 monotonicly 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.

4 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.

Table 1. Social measures for Enron employees sorted by SocialScore

Name	Surname	Position	Level	HP	Degree	Betweenness	Hubs	Clustering	SocialScore
Tim	Beldon	Managing Director	1	1,00	83	370,35	0,04	0,40	75,68
Debora	Davidson	Admin Assist	2	0,83	66	278,35	0,04	0,41	63,51
Anna	Meher	Admin Assist	2	0,83	62	260,94	0,04	0,42	62,84
Carla	Hoffman	Staff	5	-0,44	55	143,98	0,04	0,49	61,67
Cara	Semperger	Specialist	5	-0,44	63	82,96	0,03	0,52	53,68
Diana	Scholtes	Manager	4	0,33	45	21,44	0,03	0,70	53,31
Sean	Crandall	Director	3	0,61	42	40,04	0,03	0,62	43,64

Tim	Heizenrader	Director	3	0,61	33	19,45	0,02	0,71	35,56
Donald	Robinson	Specialist	5	-0,44	27	6,67	0,02	0,81	33,03
Jeff	Richter	Manager	4	0,33	25	12,80	0,02	0,74	32,53
Julie	Sarnowski	Staff	5	-0,44	28	25,94	0,02	0,63	32,14
Mike	Purcell	Staff	5	-0,44	24	5,02	0,02	0,79	30,36
Chris	Mallory	Analyst	5	-0,44	27	9,92	0,02	0,76	30,19
Phil	Platter	Specialist	5	-0,44	33	34,34	0,02	0,63	27,90
Robert	Anderson	Specialist	5	-0,44	8	0,15	0,01	0,96	20,06
Smith	Day	Specialist	5	-0,44	6	0,00	0,01	1,00	20,00
Mark	Guzman	Specialist	5	-0,44	18	6,84	0,01	0,75	19,97
Steve	Swan	Manager	4	0,33	9	0,20	0,01	0,93	19,55
Maria	VanHouten	Specialist	5	-0,44	7	0,11	0,01	0,95	19,44

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

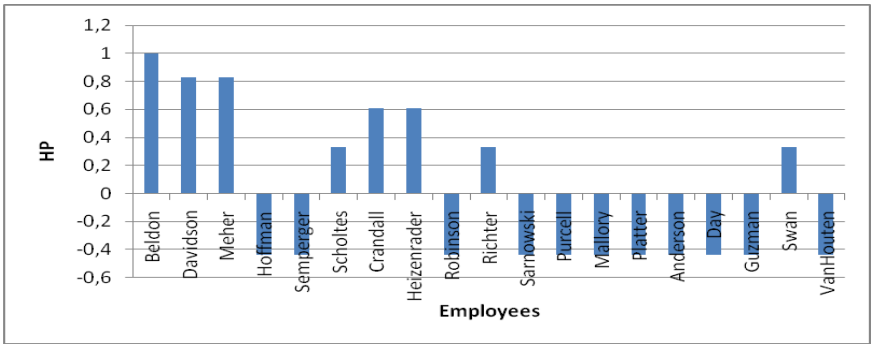

Fig. 4. Hierarchical Positions of Enron employees sorted by Social Score

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.

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

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

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, the analysis shows the rankings are very similar

with Kendall rank over 0.6 with Management department perfectly identical (Kendall rank of 1).

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 [4] 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.

5 Conclusions and Future Work

Managing human resources in a company can be very well supported by the social network approach. However, for the results be more reliable, a perfect company to *calibrate* the system would be needed, where all social leaders are in fact directors. While no perfectly organized company is found, the system have to rely on the experts knowledge and instincts.

In the future we would like to further analyze differences between extracted social network and organizational hierarchy.

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

References

1. Cohen W. (2009) Enron Email Dataset. [Online]. http://www.cs.cmu.edu/~enron/
2. Kendall M.G., Rank Correlation Methods. Oxford: Oxford, 1948.
3. Kleinberg J. "Authoritative sources in a hyperlinked environment". Journal of the ACM, vol. 46, 1999.
4. Levin J., Arluke A. "An exploratory analysis of sex differences in gossip". Sex Roles, vol. 12, no. 3-4, pp. 281-286, February 1985.
5. 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.
6. 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.
7. Scott J., *Social network analysis: A handbook*, 2nd ed. London: Sage, 2000.
8. Shetty J., Adibi J. Ex employee status report. Retrieved from: http://www.isi.edu/~adibi/Enron/Enron_Employee_Status.xls. 2004
9. The Internet Society. RFC 2822 - Internet Message Format. Retrieved from: http://www.faqs.org/rfcs/rfc2822.html. 2001

10. Wasserman S., Faust K., *Social network analysis: Methods and applications*. New York: Cambridge University Press, 1994.