Promotion and resignation in employee networks

Jia Yuan[a], Qian-Ming Zhang[b,c,*], Jian Gao[b,d], Linyan Zhang[d], Xue-Song Wan[f], Xiao-Jun Yu[f], Tao Zhou[b,e]

[a] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[b] Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[c] Center for Polymer Studies, Department of Physics, Boston University, Boston, Massachusetts, United States of America

[d] Chengdu Hire Big Data Co. Ltd., Chengdu, People's Republic of China

[e] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[f] Beijing Strong Union Technology Co. Ltd., Beijing, People's Republic of China

[*] Corresponding: qmzhangpa@gmail.com

ABSTRACT

Enterprises have put more and more emphasis on data analysis so as to obtain effective management advices. Managers and researchers are trying to dig out the major factors that lead to employees' promotion and resignation. Most previous analyses were based on questionnaire survey, which usually consists of a small fraction of samples and contains biases caused by psychological defense. In this paper, we successfully collect a data set consisting of all the employees' work-related interactions (action network, AN for short) and online social connections (social network, SN for short) of a company, which inspires us to reveal the correlations between structural features and employees' career development, namely promotion and resignation. Through statistical analysis and prediction, we show that the structural features of both AN and SN are correlated and predictive to employees' promotion and resignation, and the AN has higher correlation and predictability. More specifically, the in-degree in AN is the most relevant indicator for promotion; while the *k*-shell index in AN and in-degree in SN are both sensitive to resignation. Our results provide a novel and actionable understanding of enterprise management and suggest that to enhance the interplays among employees, no matter work-related or social interplays, can largely improve the loyalty of employees.

1. Introduction

Employees are the core and soul of enterprises, and thus human resources departments are always trying to get comprehensive understanding of employees and provide to managers valuable and effective suggestions. This issue has attracted much attention from interdisciplinary research domains. Some factors related to the employees' performance have been revealed, such as centrality [1,2], self-monitoring orientation [1], sex difference [3], communication patterns [4], and so on. Besides the performance, predicting and controlling resignation in advance are also valuable, since employees' resignation may result in great losses for enterprises. Freely and Barnett showed that people being highly connected and in more central positions of the employee networks are less likely to resign [5]. Ten years later this group put forward that employees who reported a greater number of out-degree links to friends were less likely to resign [6]. Similar indicators include degree, betweenness and closeness [7,8].

Traditional studies were based on the data sets gathered from questionnaire survey that is probably subjective due to the psychological defense. Fortunately, recent IT development brings us more reliable information, such as email data [9], communication records [10], online social relations [11], contact networks built by wearable electronic sensors [4], and so on. In this paper, we collect anonymous employees' work-related interactions and social connections from a social network platform developed and used by a Chinese company consisting of more than a hundred employees, named Beijing Strong Union Technology Co. Ltd. (Strong Union for short). Accordingly, we can build two directed networks, action network (AN) and social network (SN), where nodes represent employees and links indicate the work-related interactions and social connections, respectively. A few studies have already showed that the employees being more central in the networks or with more connections to others were less likely to resign [5,6,7,8,12,13], while managers whose local networks are rich in structural holes could be promoted faster [14]. Here, we take more structural features into consideration to reveal the correlations between topology and career choices. In addition, we would like to show which network is more related to promotion or resignation.

The two networks characterize relationships among employees from different aspects. AN reflects work-related interactions such as downloading working files and working blogs within working groups, while SN reflects social connections among all employees of the company, through which employees can share their family and entertainment activities (usually not related to works) with others. Via statistical analysis, we find that the structural features of both AN and SN are correlated and predictive to employees' promotion and resignation, and the AN has higher correlation and predictability. More specifically, the in-degree in AN is the most relevant indicator for promotion; while the *k*-shell index in AN and in-degree in SN are both sensitive to resignation.

The paper is organized as follows. Section 2 presents the description and basic statistics of the data set. The significance of selected structural features and logistic regression are respectively showed in Section 3 and Section 4. At last we give our discussion and conclusion in Section 5.

2. Data Description

The data set is gathered from a social network platform which involves all the 104 employees in Strong Union until the end of 2013. It consists of two types of information, work-related actions and online social connections. The work-related actions of employees include reposting and commenting working blogs, assigning and reporting working tasks, downloading working files, and so on. The online social connections, which are similar to the follower-followee relationships in *www.twitter.com*, are created regardless of working groups, namely everybody can build relationships with others even they don't belong to any common working groups. Accordingly, we can build two directed networks, action network (AN) and social network (SN), where nodes represent employees and links indicate the work-related interactions and social connections, respectively. More specifically, a link from *u* to *v* in AN means the employee *u* has reposted, commented or downloaded *v*'s working files, and a link from u to v in SN means u has followed v. To the date of the data collection, 25 employees have resigned and 12 employees were promoted.

Table 1 summarizes the basic topological features of the two networks. Compared with AN, SN is of higher density and average degree. The two networks are both of high clustering coefficients, quite small average shortest path lengths and very small modularity, since they are so dense that all the nodes are grouped together. Different from most known social networks [15], the two networks are both highly disassortative. It is because many work-related interactions and social connections happening between leaders and ordinary employees: the former are usually of large degrees while the latter are usually less connected.

Table 1. The basic topological features of the two networks. N is the total number of nodes. D denotes the density of directed networks calculated by $|E|/N(N-1)$, where $|E|$ is the number of directed links. Other features are calculated by transferring the directed networks into undirected ones. $\langle k \rangle$ is the average degree, $\langle d \rangle$ is the average shortest path length. C is the clustering coefficient [16] and r is the assortative coefficient [15]. H is the degree heterogeneity quantified by a variant of Gini index [17]: The highe H is, the more heterogeneous the degree distribution is. Q is the modularity of a network calculated by Louvain method [18].

Networks	N	D	$\langle k \rangle$	$\langle d \rangle$	C	r	H	Q
AN	97	0.26	35.73	1.64	0.76	-0.27	0.35	0.09
SN	104	0.29	47.04	1.55	0.81	-0.41	0.32	0.08

3. Significance of Structural Features

Previous studies suggested that some structural features of employees are effective to evaluate the turnover intention [5,6,7,8,12,13]. Here we select three well-known indicators in-degree (k_i), out-degree (k_o) and k-shell index (k_s). Given a network G, k_i is the number of links pointing to a node and k_o is the number of a node's outgoing links. k_s of a node is the largest number such that this node belongs to the k_s-core, which is the maximum subgraph of G in which all nodes have degree no less than k_s [19]. This index defines the employees' positions in the network: the employees with higher k_s are more central in the network, and vice versa. The k-shell index and its variants have been used to identify the influential spreaders in social networks [20-24].

Because the career choices are binary, i.e. promotion vs. non-promotion, and resignation vs. non-resignation, we adopt AUC (area under the receiver operating characteristic curve [25]) value to measure the correlation between structural features and career choices. This metric can be interpreted as the probability that a randomly chosen promoted (non-resigned) employee has a higher value of k_i, k_o, or k_s than a randomly chosen non-promoted (resigned) employee (like the definition in [26]). In the implementation, we compare all the pairs of employees with opposite choices to calculate AUC. Suppose that among n times of comparisons in total, if there are n' times the promoted (non-resigned) employee has higher value than, and n'' times has the equal value to the non-promoted (resigned) employee, the AUC value can be calculated by AUC=(n'+n'')/n.

Promoted employees have obviously higher values of all the three structural features, while resigned employees have much lower values than others. These results strongly imply that the employees who are more central or highly connected are more likely to be promoted and less likely to resign. In addition, there are three secondary results: (i) AN is more related to both promotion and resignation than SN; (ii) in-degree in AN, namely the number of colleagues who have responded to the target employee's working files, is a critical indicator for promotion; (iii) out-degree in AN, indicating the number of employees the target employee follows, is the most relevant indicator for resignation.

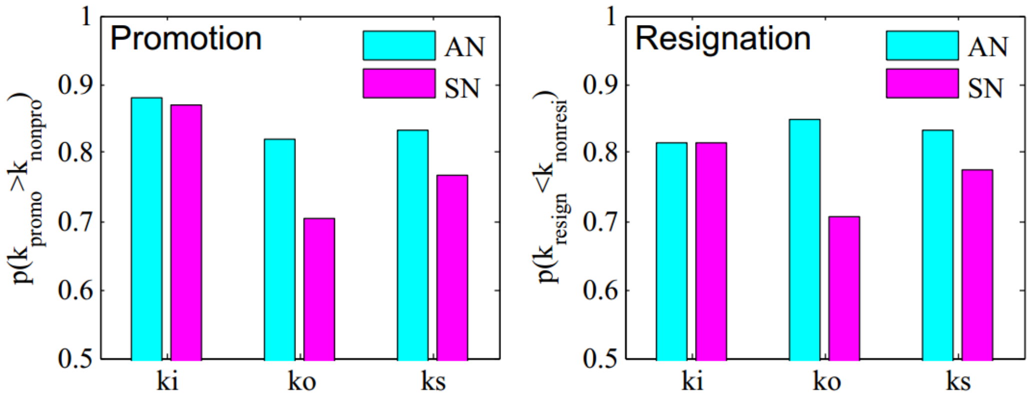

Fig. 1 (Color Online) AUC values for promotion and resignation. Subfigures (a) represents the probability that a randomly chosen promoted employee has a higher value of k_i, k_o, or k_s than a randomly chosen non-promoted employee. Subfigures (b) represents the probability that a randomly chosen resigned employee has a lower value of k_i, k_o, or k_s than a randomly chosen non-resigned employee.

4. Logistic Regression

Since employees are only with two states in each case ("resignation" vs. "non-resignation", and "promotion" vs. "non-promotion"), we adopt logistic regression [27] to identify the resigned employees and promoted employees. It is a type of probabilistic statistical classification model, which is usually used to predict the binomial outcome of a response variable using one or more predictor features (e.g., the three indicators). The binary logistic regression model can be represented by a conditional probability

$$P(1|\vec{x}) = \frac{1}{1+e^{-(b_0+\sum_i^m b_i x_i)}}, \tag{2}$$

where $\vec{x} = (x_1, \ldots, x_m)$ is the vector of features, such as k_i, k_o and k_s, and $b_0, b_1, \ldots, b_m$ are the coefficients we need to estimate based on the data.

In the following, we take resignation as an example to show how binary logistic regression works on classifying employees. First we code the status of employees into numeric values: "resignation" as 1 and "non-resignation" as 0, then map the selected features to a logit function and convert it into odds to get the probability score. Eventually, we set up a cutoff probability to classify employees into two groups, "resignation" or "non-resignation". Then the accuracies of classification can be measured. One basic metric is *precision* which is defined as

$$p = \frac{n}{N}, \quad (3)$$

where N is the total number of employees and n is the number of correctly classified employees. Meanwhile, *recall*, which is known as the sensitivity in binary classification, is also necessary to be considered especially in this case. That is because the numbers of both promoted employees and resigned employees are very small. If we classify all the employees into the “non-promoted” or “non-resigned” class, the *precision* is still very high (up to 88.5% for promotion or 76% for resignation), but no relevant employees (i.e., promoted employees and resigned employees) can be found out. The mathematic form of this metric is

$$recall = \frac{n'}{N'}, \quad (4)$$

where n' is the number of relevant employee who is retrieved, and N' is the number of the relevant employees who has resigned or been promoted.

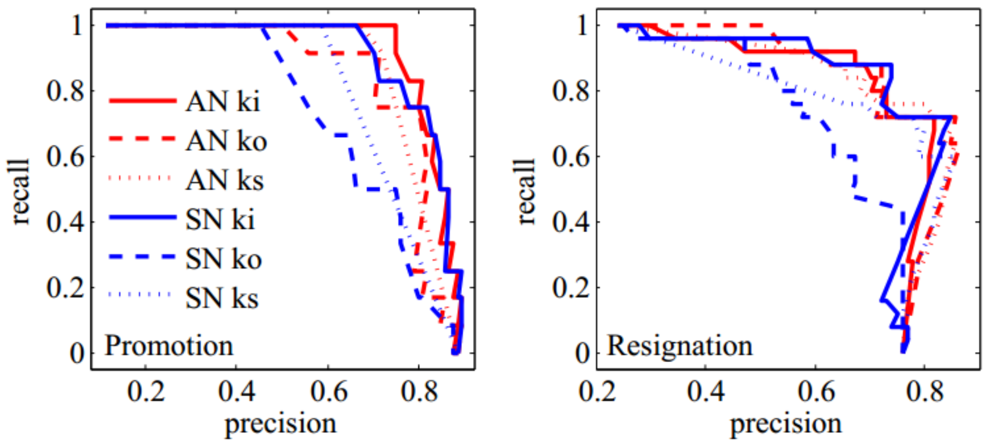

Fig. 2 (Color Online) Precision-recall curve with various cutoff probabilities. AN means action network, while SN means social network. ki, ko and ks correspond to k_i, k_o and k_s respectively, which are the selected variables to classify which employees are promoted or resigned. The significance of chi-square test [28] for the constant b_0 and coefficient b_1 according to the six models for promotion are (0, 0), (0.002, 0), (0.221, 0.208), (0, 0), (0.058, 0), and (0.987, 0.987) respectively. And those for resignation are (0, 0.755), (0, 0.164), (0, 0.022), (0, 0.021), (0.009, 0.054) and (0, 0.007) respectively. In general, the value below 0.05 shows a good fit.

To figure out which feature is most related to promotion/resignation, in each run, we only choose one feature to form the vector $\vec{x}$. The one with higher *precision* and *recall* is considered to be more related. However, the values of *precision* and *recall* depend on the cutoff probability, which is usually set to 0.5 by default. So, if we alter the cutoff probability from 0 to 1, we can get a *precision-recall* curve. As shown in Fig. 2, the most accurate classification can be found according to the most top and right curve. More directly and clearly, we introduce a metric named $F1$ [29,30] to compare the effectiveness by synthesizing *precision* and *recall* in the following style:

$$F1 = \frac{2 * precision * recall}{precision + recall}.$$

It can be interpreted as a weighted average of the *precision* and *recall*. Obviously, an F1 score reaches its best value at 1 and worst score at 0. The alteration of F1 scores can be found in Fig. 3.

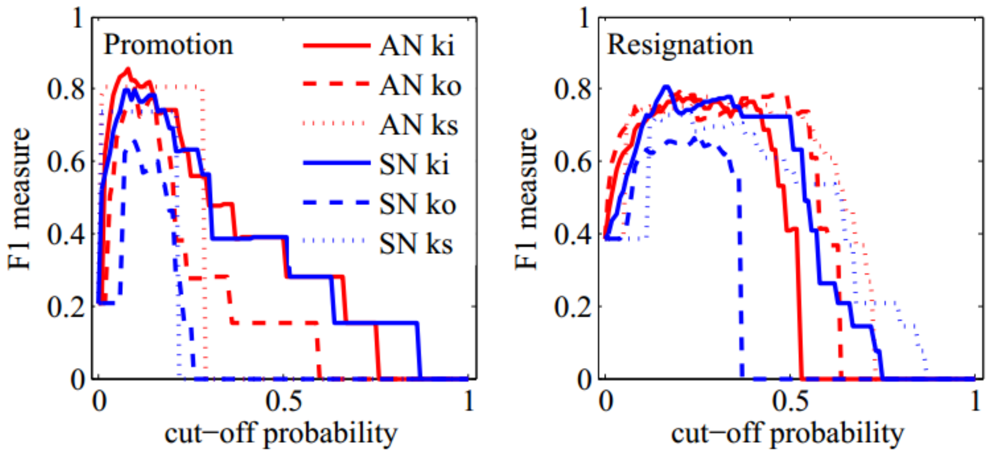

Fig. 3 (Color Online) F1 scores with various cutoff probabilities. AN means action network, while SN means social network. ki, ko and ks mean k_i, k_o and k_s respectively, which are the selected variables to classify which employees are promoted or resigned.

From the peak values of the curves for Promotion, we can see that k_i is the most effective feature, while k_o is the worst. In such small-scale networks, k_i can reflect the employee's direct influence since the leaders are more likely to attract more attention. Following this idea, k_s should be also valid, because the higher k_s an employee has, the more central position he stands. However, k_s does not work so well. It maybe because there are too many employees with the highest k_s, i.e., 39 out of 104 employees with highest AN k_s, and 55 out of 104 employees with highest SN k_s. These employees own quite different values of in-degrees ranging from 33 to 82 for AN, or from 37 to 68 for SN. So the employees are more diacritical from the perspective of k_i than that of k_s. Compared with SN, every feature of AN is better than the corresponding one of SN. Further, let us check the significance of these estimated coefficients (represented in the caption of Fig. 2). Only AN k_i, AN k_o and SN k_i can be accepted. So we can conclude that AN k_i is the most effective and acceptable for promotion.

In the case of Resignation, although the maximal F1 score belongs to SN k_i, the three features of AN are also very competitive. These results make sense because employees who get little attention (low SN k_i and AN k_i), or stands at the periphery of the social network (low k_s) would feel unvalued, under-appreciated and lonely. This negative feeling would make them unwilling to communicate with others (low AN k_o) but they have to create connections to their leaders to complete work. That may be also the reason of bad performance of SN k_o. In turn, employees who own many followers (high k_i), stand at important positions (high k_s), or are the hubs of communication (AN k_o) are less likely to resign. When we check the significance of these estimated coefficients, only AN k_s, SN k_s and SN k_i are acceptable. So we tend to believe AN k_s and SN k_i are better.

Through the above results and analysis, we can conclude that AN k_i is mostly related to promotion, while AN k_s and SN k_i are mostly related to resignation. Our results also suggest that, no matter the information in AN or SN is very valuable to manage employees.

5. Discussion and Conclusion

In this paper, we successfully collect a dataset consisting of all the employees' work-related actions and online social connections from a social network platform developed and used by a company. To reveal the correlation between employees' social/work-related actions and employees' career development, we implement some experiments based on the structural features of those two networks. The correlation results based on AUC show work-related interactions is more sensitive to both promotion and resignation. It is consistent with the classification results through Logistic Regression analysis. We find the employee who can get more attention (higher in-degree k_i) in work-related networks (AN), is more probable to be promoted; while the resigned employees can be indicated by both AN k_s and SN k_i, where k_s is the *k*-shell value and SN means social networks. The results provide a novel and actionable understanding of enterprise management and suggest that to enhance the interplays among employees, work-related interplays can largely improve the loyalty of employee.

In these two employee networks, employees are so highly connected with each other that the networks are not obvious with modules. Maybe that is the nature of small-scale enterprises, but we can imagine that connections might be very sparse and the corresponding network might be organized with community structure in large-scale enterprises. So the conclusions may be limited in small-scale enterprises. Moreover, *k*-shell is a good way of reflecting employees' positions. It makes sense that peripheral employees are more likely to leave. In turn, we may think the central employees are of higher probability to be promoted, but the classification experiments give negative results. We have discussed that this maybe because there are too many employees with the highest *k*-shell value. However, the organizational structures are usually different in different countries [31]. Promotion and resignation are also usually affected by various reasons. For example, maybe the employees in the central position have already been the senior leaders who are restricted by "vacancy chains" [32]. Some employees leave the company just because of the expiry of the employment contract. So, further analysis is very significant to be implemented on more data sets.

Acknowledgments

We are grateful to Jin Yin for his valuable comments and suggestions. This work is partially supported by the National Nature Science Foundation of China under Grand No.11222543. QMZ acknowledges China Scholarship Council and the support from the Program of Outstanding PhD Candidate in Academic Research by UESTC (No. YBXSZC20131034).

References

[1] M.K. Ahuja, D.F. Galletta, K.M. Carley, Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study, Manage. Sci. 49 (2003) 21.

[2] A. Mehra, M. Kilduff, D.J. Brass, The social networks of high and low self-monitors: Implications for workplace performance, Admi. Sci. Quart. 46 (2001) 121.

[3] N. Bu, J.P. Roy, Career success networks in China: sex differences in network composition and social exchange practices, Asia Pac. J.Manage. 22 (2005) 381.

[4] A.S. Pentland, The new science of building great teams, Harvard Bus. Rev. 90 (2012) 1.

[5] T.H. Feeley, G.A. Barnett, Predicting employee turnover from communication networks, Hum. Commun. Res. 23 (1997) 370.

[6] T.H. Feeley, J. Hwang, G.A. Barnett, Predicting employee turnover from friendship networks, J. Appl. Commu. Res. 36 (2008) 56.

[7] B. Mullen, C. Johnson, E. Salas, Effects of communication networks structure: Components of positional centrality, Soc. Networks 13 (1991) 169.

[8] K.W. Mossholder, R.P. Settoon, S.C. Henagan, A relational perspective on turnover: Examining structural, attitudinal, and behavioral predictors, Acad. Manage. J. 48 (2005) 607.

[9] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, A. Arenas, Self-similar community structure in a network of human interactions, Phys. Rev. E 68 (2003) 065103.

[10] N. Eagle, A.S. Pentland, Reality mining: sensing complex social systems, Pers. Ubiquit. Comput. 10 (2006) 255.

[11] K. Lewis, M. Gonzalez, J. Kaufman, Social selection and peer influence in an online social network, Proc. Natil. Acad. Sci. USA 109 (2012) 68.

[12] J.M. McPherson, M. Popielarz, S. Drobnic, Social networks and organizational dynamics, Am. Sociol. Rev. 57 (1992) 153.

[13] D.P. Moynihan, S.K. Pandey, The ties that bind: Social networks, person-organization value fit, and turnover intention, J. Public. Admin. Res. Theory 18 (2008) 205.

[14] R.S. Burt, Structural holes: The structural holes of competition, Cambridge, MA: Harvard University Press, 1992.

[15] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.

[16] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440.

[17] H.-B. Hu, X.-F. Wang, Unified index to quantifying heterogeneity of complex networks, Physica A 387 (2008) 3769.

[18] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech - Theory E 10 (2008) P10008.

[19] S.B. Seidman, Network structure and minimum degree, Soc. Networks 5 (1983) 269.

[20] M. Kitsak, L.K. Gallos, S. Havlin, F.Liljeros, L. Muchnik, H.E. Stanley, H.A. Makes, Identification of influential spreaders in complex networks, Nat. Phys. 6 (2010) 888.

[21] J.-G. Liu, Z.-M. Ren, Q. Guo, Ranking the spreading influence in complex networks, Physica A 392 (2013) 4154.

[22] A. Zeng, C.-J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377 (2013) 1031.

[23] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Physica A 391 (2012) 1777.

[24] Y. Liu, M. Tang, T. Zhou, Y. Do, Core-like groups resulting in invalidation of k-shell decomposition analysis, arXiv: 1409.5187.[25] J.A. Hanely, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29.
[26] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (2012) 1150.
[27] C.M. Bishop, Pattern Recognition and Machine Learning, Vol. 1. New York: springer, 2006.
[28] F. Yates, Contingency table involving small numbers and the χ^2 test, Supplement to the Journal of the Royal Statistical Society 1 (1934) 217.
[29] C. J. van Rijsbergen, Information Retrieval, Butterworths, London, second edition, 1979.
[30] D. D. Levis, W. A. Gale, A sequential algorithm for training text classifiers, in: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '94), W. Bruce Croft and C. J. van Rijsbergen (Eds.). Springer-Verlag, New York, Inc., New York, NY, USA, pp. 3-12.
[31] P. Hedstrom, Organizational differentiation and earnings dispersion, Am. J. Sociol. 97 (1991) 96.
[32] I. D. Chase, Vacancy Chains, Annu. Rev. Sociol. 17 (1991) 133.