Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Using Organizational Social Networks to Predict Employee Engagement

Shion Guha,[1] Michael Muller,[2] N. Sadat Shami,[3] Mikhil Masli,[3] Werner Geyer[2]
[1]Cornell University, [2]IBM Research, [3]IBM
sguha@cs.cornell.edu, michael_muller@us.ibm.com, sadat@us.ibm.com,
mikhil_masli@us.ibm.com, werner.geyer@us.ibm.com

Abstract

Employee engagement (EE) has been shown to have important implications for the success of organizations. Most researchers have discussed employee engagement in terms of factors in a top-down, hierarchical model of the organization. However, there may also be contributing factors from an employee's intra-organizational social network. (IOSN). In this paper, we show that an employee's social network attributes can contribute to the prediction of engagement, primarily through centrality and homophily in a large, multinational company. Our research expands the range of theoretical factors that can predict employee engagement from a top-down vertical model to a mixed-factor horizontal model. We discuss how this work points toward a richer set of methods to predict engagement, as well as new ways of thinking about organizational networks.

Introduction

Employee engagement has become an important factor in business planning and human resource management (Govendar 2010). Engaged employees are reported to make powerful contributions to desirable organizational outcomes, such as improved financial and operational benefits (Saks 2006). State-of-the-art methods for measuring engagement have involved surveys (Wiley 2010), whose costs limit their frequency of use. Organizations would prefer a measurement of engagement that is closer to real-time. In this paper, we extend earlier research to model and predict engagement based on anonymized employee social media.

In general, organizational models of engagement have emphasized the importance of executive strategies (Venkatesh 2015, Wiley 2010), managerial influences (Luthans & Peterson 2002), and workplace conditions that may be partially determined by executives and managers (Shami et al. 2015, Venkatesh 2015). But, what about employees?

In this paper, we uncover arguably the first statistical evidence of the importance of social network factors to EE in a large corporation. Using two years of survey results and social network relations from approximately 200,000 employees in a large corporation, we show that (a) an employee's network position (centrality) and (b) her neighbors (through homophily) can predict her engagement.

Employee Engagement at IBM

We studied EE at IBM. IBM provides software and consulting services and has offices in over 100 countries. Much of the work is both international and virtualized. Virtual teams are the norm, and can vary from a few people to tens of thousands of people. To support the virtual teams, IBM has implemented a suite of internal social software services for its employees for structured and semi-structured online collaborations in the form of blogs, wikis, communities, and discussion forums. One of these applications is an internal social networking service (IOSN). All of these social media are contained behind the company's firewall. Like many other organizations, IBM measures the engagement of its employees through an annual survey.

Recently, IBM has expanded its efforts from *measuring* engagement to *predicting* engagement, using the survey results as ground truth. Predictions were made *not* for individual employees, but rather for larger groups of employees in what IBM calls an "employee segment," to detect problems that affect many employees and that could be addressed through workplace programs that did not stigmatize individual members of the organization.

We therefore ask:

RQ1: Is EE predicted by node position in organizational social networks?

Variable/Year	2013				2014			
Total number of links	5,535,846				5,951,165			
Total number of nodes	209,471				219,138			
Variable	**Min**	**Max**	**Mean**	**Std. Dev**	**Min**	**Max**	**Mean**	**Std. Dev**
Degree*	0	1	0.452	0.033	0	1	0.476	0.038
Eigenvector*	0	1	0.347	0.081	0	1	0.392	0.089
Clustering Co-efficient*	0	1	0.293	0.036	0	1	0.377	0.045
*all centrality metrics have been normalized								

Table 1. Descriptive Statistics of the IOSN across both years of analyses

RQ2: Is EE predicted by homophily in organizational social networks?

Methods

Data Sources

We combined three large workplace datasets:

- A social network of friendship relationships, recorded from IBM's social networking service (IOSN). We discuss this network in the formal language of social network analysis, where people are "nodes" and friend relationships are "edges" that connect nodes. To protect employee privacy, all identities were anonymized through non-reversible MD5 transformation.
- Demographic data for each employee, provided by IBM's human resources organization, and anonymized in the same way as above. These data included locations, organizations, and human resources historical data such as time-in-title and performance ratings.
- Survey data from two years of EE surveys, provided by the human resources organization, and anonymized in the same way as above. As described above, the survey used three core questions that address pride, satisfaction, and advocacy (likelihood to recommend IBM as a good place to work).

We used the MD5-anonymized employee IDs to align these records. Our resulting dataset contained records for **209,471** employees for 2013 and **219,138** employees for 2014.

Preprocessing

We performed two pre-processing steps. First, we calculated several network centrality metrics from the edge-list data to determine the degree of importance of each node (employee) in the IOSN. There are many network centrality metrics; we chose three that we felt best encapsulated the connectedness and importance of a node in our IOSN. Bonacich (Bonacich 2007) recommended the inclusion of eigenvector centrality to measure the relative influence of the position of a node for large, complex networks, while degree centrality is a simple measure of the number of friends of a particular node. Similarly, we chose the clustering co-efficient since prior work (Newman 2001) established that social networks exhibit a phenomenon of clustering based on preferential attachment of nodes. We note that feature selection of the optimum set of network centrality metrics for any given network is a separate, future research project and outside the scope of this particular report.

Second, consistent with current IBM practices (Wiley et al. 2010), we converted the mean score of the three survey items to a binary variable, where mean scores in the range 4-5 (inclusive) were coded as 1 (engaged) and mean scores below 4.0 were coded as 0 (unengaged). This was our ground-truth dependent variable.

Variables and Measurements

Salary Group: IBM has an internal grouping system for employees in different salary groups. This consists of 15 levels and is an ordinal variable.

Performance Review: IBM also has an internal system of rating employees. This consists of 5 levels and is also an ordinal variable.

Time since Last Promotion: This represents the time (in months) since the employee was last promoted at work.

Tenure in the Company: This represents the time (in months) that the employee has been employed at the IBM.

Age: This represents the biological age (in years) of an employee.

Gender: This is a binary variable (male=0/female=1) of an employee's gender. IBM did not provide alternative options to express gender.

Degree centrality: This is a normalized (range = (0, 1)) measure of the number of nodes that a particular node is connected to in the IOSN. (Seidman 1983)

	Year					
	2013			2014		
Variables	**Std. β**	**OR**	**SE**	**Std. β**	**OR**	**SE**
Salary Group	0.11	0.8	0.023	0.06	1.165	0.013
Performance Review	0.09	1.107	0.016	-0.12	1.248	0.028
Time since Last Promotion	- 0.06*	0.859	0.018	-0.09*	0.556	0.018
Tenure in the Company	- 0.14*	0.963	0.024	-0.12*	0.896	0.031
Age	-0.12*	0.717	0.033	-0.13*	0.834	0.022
Gender	- 0.03	0.952	0.014	-0.04	0.914	0.009
Degree	0.11**	1.118	0.001	0.07	1.106	0.002
Eigenvector	0.26***	1.614	0.016	0.28***	1.729	0.019
Clustering Co-efficient	0.09	1.031	0.021	0.19**	1.265	0.017
Model Fit Statistics	AIC = 3142 BIC = 2692			AIC = 3318 BIC = 2958		
***** p < 0.001 ** p < 0.001 * p< 0.05**						

Table 2. Results of the Logistic Regression Model for each of 2013 and 2014.

Eigenvector centrality: This is a normalized (range = (0, 1)) measure of the relative influence of a node, given its position and connectivity in the IOSN. (Bonacich 2007)

Clustering Co-efficient: This is a normalized (range = (0, 1)) measure of how well a node tends to cluster together with immediate neighbors (Watts & Strogatz 1998)

Analysis

First, we used binary logistic regression to estimate the effect on engagement of an individual given her node position and demographical attributes. These results are presented in Table 2. Second, we estimated a social network-based effect by performing a homophily-influence analysis (Ibarra 1992).Therefore, for each node *p* we compute the probability of a net homophily effect for engagement from all her neighbors as follows:

$$\frac{\sum_{i=1}^{u} \frac{c}{t} - \sum_{i=1}^{v} \frac{c}{t}}{n}$$

where:

n = total number of neighbors
u = total number of engaged neighbors
v = total number of disengaged neighbors
c = total number of common demographics with p
t = total number of demographic groups
n = u + v

Thus, the overall prediction of engagement for a given individual (I) as well as network (N) effects is computed by:

$$P(EE) = P(I) + P(N) - P(I).P(N)$$

To assess these predictions, we compared correct predictions vs. incorrect predictions via confusion matrices. The confusion matrix for the computation of *P(I)* as well as *P(EE)* for each respective year of analysis (2013 and 2014) after 10 fold cross validation is presented in Table 3.

Results and Discussion

Table 1 lists summary statistics describing the major properties of the IOSN across both years (2013, 2014). We can get a holistic idea of the structure and size of the IOSN across both years. Comparing the 2013 and 2014 networks, we observed a net increase of 415,319 edges (friend relationships) and 9667 nodes (employees) in the IOSN.

Individual Level Effects (RQ1)

Demographics, in both years, explain roughly 50% of the overall variance. A close inspection of all the demographic variables starts to reveal a picture of an ideal engaged employee. This hypothetical employee has been with the organization for a short period of time (1-3 years), earns a mid-level salary, was promoted less than a year ago, and usually receives a high performance review. Altogether, this suggests that EE at IBM is driven only partly (~50% of explained variance) by demographic and other workplace related factors. Our question then becomes: can we improve on this prediction by adding network centrality metrics to our predictors? **RQ1** predicted that node position in the network would be an important predictor of EE. In 2013, degree centrality and eigenvector centrality were significant predictors in this model. This result points towards an intuition that employees arrange themselves in specific ways in this IOSN to achieve their enterprise objectives. This situation slightly changes in 2014, where

Confusion Matrix Before Network Effects

Year	2013				2014		
	Observed				Observed		
		0	1			0	1
Predicted	0	**38,819 (44.9%)**	36,421 (31.4%)	Predicted	0	**35,414 (46.2%)**	39,716 (27.9%)
	1	47,637 (45.1%)	**79,569 (69.6%)**		1	41,239 (43.8%)	**102,637 (72.1%)**

Confusion Matrix After Network Effects

Year	2013				2014		
	Observed				Observed		
		0	1			0	1
Predicted	0	**47,896 (55.4%)**	26,330 (22.8%)	Predicted	0	**41,928 (54.7%)**	28,752 (22.2%)
	1	38,560 (44.6%)	**89,660 (77.3%)**		1	34,725 (45.3%)	**113,601 (79.8%)**

Table 3. Confusion Matrices before and after Network Effects

eigenvector and clustering co-efficient are the significant network centrality predictors. This suggests two outcomes of interest: First, relative node position is important in the IOSN. Second, the tightness of node connections is equally important.

Network Level Effects (RQ2)

In the 2013 IOSN, we find that the addition of homophilic influences to individual level effects increases predictive power of true positives by **8.3%** and true negatives by **10.5%**. Similarly, in the 2014 IOSN, we find the predictive power of true positives increased by **7.7%** and true negatives by **8.5%**. These are important points to note for two reasons. First, these increases are moderately high and second, they are in line with effect sizes in existing work (McPherson et al. 2001) arising from homophily. This gives us a relatively high degree of confidence that homophily (or social selection) is an important predictor in estimating EE in IBM. In other words, how you choose your friends in your IOSN matters significantly towards your EE.

RQ2 asked if homophily in IOSNs has some influence on EE. Our homophily analysis suggests that indeed, similarity among neighbors in IOSNs affects EE significantly. The existence and increase in homophily, in significant amounts, could also be one explanation of the statistical significance (Newman & Park 2003) of the clustering co-efficient metric that we found in Table 2. We believe that, in the context of estimating EE, this is a novel and important finding in the organizational network literature which we discuss in greater detail in the next section.

Acknowledgment

We acknowledge the anonymous reviewers and chairs for their reviews as well as Inbal Ronen and Naama Zwerdling for providing friend lists data.

References

Simon L. Albrecht (2010). Handbook of employee engagement: Perspectives, issues, research and practice. Edward Elgar, Cheltenham, UK.

Phillip Bonacich. (2007) "Some unique properties of eigenvector centrality." Social Networks 29.4, 555-564.

Borgatti, S and Foster, P. (2003). The network paradigm in organizational research: A review and typology. Journal of Management 29(6), 991-1013.

Eean R. Crawford, Jeffrey A. LePine, and Bruce Louis Rich (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. J. App. Psych. 95(5), 834-848.

Herminia Ibarra. (1992). "Homophily and differential returns: Sex differences in network structure and access in an advertising firm." Administrative science quarterly, 422-447.

Fred Luthans and Suzanne J. Peterson (2002). Employee engagement and manager self-efficacy. J. Mgmt. Dev. 21(5), 376-387.

Mark E.J. Newman. (2001). "Clustering and preferential attachment in growing networks." Physical Review E 64.2, 025102.

Alan M. Saks (2006). Antecedents and consequences of employee engagement. J. Mgmt. Psych. 71(7), 600-619.

Stephen B. Seidman. (1983) "Network structure and minimum degree." Social networks 5.3, 269-287.

N. Sadat Shami, Michael Muller, Aditya Pal, Mikhil Masli, and Werner Geyer (2015). Inferring employee engagement from social media. Proc. CHI 2015, 3999-4008.

A. Narasima Venkatesh (2015). Employee engagement through leadership. Am. Int. J. Res. Hum. Arts, Soc. Sci. 9(4), 333-336.

Duncan J. Watts, and Steven H. Strogatz. (1998) "Collective dynamics of ‘small-world’networks." nature 393.6684, 440-442.

Jack W. Wiley (2010). The impact of effective leadership on employee engagement. Empl. Rel. Today 37(2), 47-52.