Microblogging in the Enterprise: A few comments are in order

Charalampos CHelmis
Department of Computer Science
University of Southern California
Los Angeles, USA
chelmis@usc.edu

Viktor K. Prasanna
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, USA
prasanna@usc.edu

***Abstract*—Popular social networking sites have revolutionized the way people interact on the Web. Researchers have studied social networks from numerous perspectives, mostly focusing on publicly available social networks and microblogging sites. Enterprises however have recently being adopting and utilizing microblogging services as part of their day to day operations. The goal of this paper is to study the topological properties of a corporate microblogging service, its dynamics and characteristics. Through an extensive analysis of enterprise microblogging data, we provide insights on the structural properties of the extracted network of directed messages sent between users of a corporate microblogging service, as well as the lexical and topical alignment of users. We compare our results to traditional, general purpose, online social networks and discuss the implications of our findings. To the best of our knowledge, this work is the first quantitative study of an enterprise microblogging service, its usage characteristics, and its derived social network based on replies between users.**

***Keywords*-micro-blogging; social networks; social media; enterprise; measurement; analysis;**

I. Introduction

Social Networks have revolutionized the way people communicate and interact, while serving as a platform for information dissemination, content organization and search, expertize identification, and influence discovery. The popularity of online social networks like Facebook and Twitter has given researchers access to massive quantities of data for analysis. Such datasets provide an opportunity to study the characteristics of social networks in order to understand the dynamics of individual and group behavior, underlying structures, and local and global patterns that govern information flow.

Most of the analysis performed thus far has focused on publicly available social networks [1]. However, microblogging capabilities are adopted and used in the enterprise as well [2]. The topological characteristics of enterprise social networks have thus far not been studied, partially due to the lack of available datasets. In this work, we provide an extensive quantitative analysis of enterprise microblogging data, collected from a large, international corporation over a one year period. Specifically, we have extracted and studied the directed network we inferred from @reply messages in a corporate microblogging service, which resembles Twitter. We believe this is the first quantitative examination of structural and topological characteristics of a microblogging service in a coporate setting. Our findings confirm the power-law, small-world, and scale-free properties of the @replies network. We further examine the lexical and topical alignment of users in the @replies network and discover that semantic similarity of users as a function of their distance is significantly higher as compared to online social networks.

In addition to validating structural and semantical properties of the @replies network and comparing our findings to traditional online social networks, we further provide significant insights into the corporate microblogging service. We observe that the world is “smaller” in the corporate environment (as expected), even though the inferred network appears to have similar structure to online networks, with a large, strongly connected core, surrounded by many small clusters of low-degree nodes. This suggests that high-degree nodes in the core exhibit characteristics of expertise, conceptualized by frequent message exchanges with other nodes. Such high-degree nodes are therefore critical for the connectivity and flow of information in the corporate environment.

In this work, we focus on the complete snapshot of a corporate microblogging service. For our analysis, we consider the complete user corpus, instead of only focusing on users belonging to the largest connected component. This includes users who may have contributed to one-to-many conversations, but, who have never sent a directed @reply message. This definition does not include users who tried the service once and never used it, or found it useless. We do not seek to discover or test the perceived benefits and barriers to adoption of microblogging services in enterprise environments. We further do not attempt to examine information flow or temporal evolution of this network. While such aspects are important, they are beyond the scope of this paper.

II. Related Work

The structure and evolution of online social networks has been investigated in detail by Mislove et al. [1] and Kumar et al. [3]. Ahn et al. [4] analyzed Cyworld, MySpace and Orkut. Kumar et al. [3] examined two online social networks and found that both possess a large strongly connected component. Girvan and Newman observed that users in online social networks tend to form tightly knit groups [5]. Amaral et al. [6] and Newman [7] examined the small-world properties (small diameter and high clustering) of different networks, while Kleinberg [8] proposed a model that captures small-world properties. The extracted social network examined in this paper exhibits small-world properties much like other general purpose online social networks. Marlow et al. [9] investigated patterns in Flickr user activity and examined vocabulary overlaps between user pairs. Schifanell et al. [10] focused on topical and lexical alignment among users who lie close to each other in the Flickr social network and exploited this alignment as an indicator of user connectivity.

On the other hand, most of the studies on microblogging networks have focused on Twitter. Krishnamurthy et al. [11] presented a detailed characterization of Twitter, identified distinct classes of Twitter users and their behaviors, geographic growth patterns and current size of the network. Java et al. [12] explored Twitter's topological and geographical properties, analyzed user interactions at the community level and showed how users with similar interests connect to each other. Zhao et al. [13] explored the factors that influence people's tendency to share personal information in Twitter, and examined microblogging's potential impact on informal communication at work. Through a qualitative study, they concluded that microblogging at workplace can assist in building stronger personal bonds between colleagues, rather than being used for professional benefits, even though they hinted that microblogging provides a complementary informal communication channel for coworkers to share and exchange information and ideas. Wu et al. [14] investigated workplace relationships built between coworkers using microblogging services and determined interaction patterns that signal personal versus professional closeness between colleagues.

Zhang et al. [2] provided a systematic examination of adoption and usage of a microblogging tool in a corporation environment, emphasizing on the perceived benefits of corporate microblogging and barriers to adoption. Ehrlich et al. [15] and DiMicco et al. [16] examined microblogging in workplace with emphasis on content type (percentage of information sharing messages versus questions and status updates) and users microblogging behavior as a function of their motivation. Their main focus was on providing qualitative results and insights on the reasons and ways people are utilizing microblogging for communication in a corporate environment. Guy et al. [17] presented an API for gathering and sharing interpersonal connections across multiple services and demonstrated its potential value with a comprehensive qualitative analysis.

Our study focuses on the social connectedness of an extracted corporate social network, as well as the properties and characteristics of social content, tie formation and information flow, in the context of a corporate microblogging service, with the goal of comparing our findings to traditional, online social networks, identifying similarities and exposing differentiations.

III. Data Set

The dataset for this analysis is a complete snapshot of a corporate microblogging service, which resembles Twitter. It consists of 4,213 unique users, who posted 16,438 messages by the end of August 2011, when we obtained the raw data for this paper. The corporate micro-blogging site does not impose any restrictions on the way people interact or who they chose to follow, much similar to Twitter.

A. Description of the Network

We inferred a directed network of users' interaction flow, mining directed user messages (@replies). We represent the @replies network as a directed graph $G = (V, E)$:

- vertices: $V = \{u_i | i = 1, ..., N\}$, where $N = |V| = 4,213$ is the total number of users,
- edges: $E = \{e_{ij} | i, j = 1, ..., N\}$, where $M = |E| = 4,489$ is the total number of edges,
- $I : E \to V \times V$ defined as follows: an edge e_{ij} exists and points from node i to node j if user i has sent at least one @reply message to user j.

We have chosen this intuitive definition for edges due to the way messages are delivered in the corporate microblogging service. If we were to consider broadcast messages, all users would be connected to everyone else, thus forming a densely connected graph, which would provide little insight. Instead we chose to represent the "transer" of content from user i to user j when user i sends user j a @reply message. An undirected edge e_{ij} between users i and j if either user sent a message to the other would not capture the semantics of directed communication, which may or may not be reciprocal. We considered weighting the edges by the frequency of replies sent from user i to user j. The addition of weights however would have no effect on the structure and properties of the inferred social graph. It would change node rankings in terms of PageRank or similar metrics, but this is not the focus of this paper.

B. High Level Statistics

Table I presents the high level statistics of our dataset. Some comments are in order. First, the average number of messages per thread is 2.02, while the ratio of the broadcast messages to the number of personal replies is ≈ 1.011. Even though these statistics indicate on average shallow conversations, we found that is not the case overall. The mean is so small due to the heavy-tailed distribution of number of messages per user. Further, even though the average number of messages per user is ≈ 4, the average number of replies per user is quite higher ($\approx$ 7.3), indicating users' tendency

Metric	Value
Number of users	4,213
Number of messages	16,438
Number of threads	8,139
Number of broadcast messages	8,174
Number of personal replies	8,264
Number of hashtags	637
Number of groups	88

Table I
HIGH-LEVEL STATISTICS OF THE @REPLIES NETWORK.

to directional communication instead of broadcasting of personal status updates or sharing of news. The study by Zhang et al. [2] reports a 25% average of "conversation seeking" type of messages in an enterprise social network. Assuming that "share news" type of messages also probe some sort of response, the combined average of $\approx 60\%$ in that study aligns quite well with our findings here.

In many online social networks, users with shared interests may create and join groups. In the corporate microblogging service users are able to create and join groups to collaborate with smaller teams. Messages sent within group boundaries are broadcast to group members only, while private message exchanges among group members are also feasible. We found that the average number of messages per group is 24.6, indicating considerably high activity patterns across all groups.

Finally, the sparsity of text in micro blogging social networks has traditionally been a hurdle for researchers interested in performing some type of statistical analysis of micro blogging content. We counted the average number of words per message in this social network and we found it to be ≈ 29.1. This number is quite high, indicating that most messages are adequately descriptive and could be safely used for statistical analysis (e.g. sentiment analysis), like Bayesian inferencing.

On the contrary, the number of hashtags per message is quite low, ≈ 1.6 on average. Tagging, allows users to organize web resources (e.g. photos in Flickr, bookmarks in Delicious or tweets in Twitter). Twitter users adopted hashtags as an attempt to alleviate the significant information overload that the streaming nature of social media impose to users interested in specific topic(s). Huang et al. [18] examined tagging strategies followed by Twitter users for content management and filtering. Rather than using hashtags, users of the corporate microblogging service mostly rely on group membership for content organization.

IV. ANALYSIS OF NETWORK STRUCTURE

In this section we characterize the structural properties of the @replies network. Let us denote the average in-degree (number of users j who have sent a message to user i) by d_{in} and the out-degree (number of users j to whom user i has sent a message) by d_{out}. Then, $d_{in} = 1.07$ and $d_{out} = 1.07$,

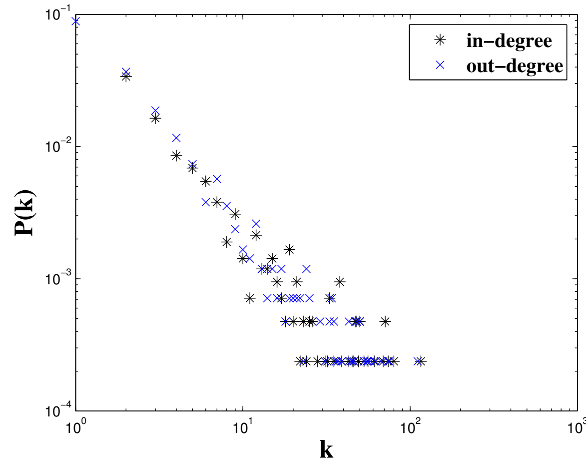

Figure 1. Distribution of users' in-degree and out-degree (both axes are in logarithmic scale).

while average degree is $d = 2.14$. It has been shown that the degree distributions of many complex networks, including social networks, conform to power laws [1]. The @replies network exhibits such characteristics. Figure 1 shows the probability distributions of the number of k neighbors (in-degree and out-degree respectively). For reference Table II reports the mean and variance of in/out-degree.

The existence of edge e_{ij} does not guarantee that the reciprocal edge e_{ji} also exists. Hence the relationship is not symmetric. If user A sends a message to user B, the edge e_{AB} is created, but not vise versa. We call user B the "follower" of user A. If B also replies to A, then they are each other's "mutual followers". Figure 2 shows the scatter plot of the number of followees versus the number of followers. The points are scattered around the diagonal, indicating equal numbers of followers and followees (possibly indicating reciprocal/mutual links). Figure 3 presents the cumulative distribution of the out-degree to in-degree ratio, exhibiting high correlation between in-degree and out-degree. This high correlation could be explained as a result of symmetric links being created due to the tendency of users to reply back when they receive a message from other users. Our analysis of the level of symmetry in the directed @replies network reveals that the degree of symmetry is not as significant as one would expect. Overall, the @replies network exhibits low level of reciprocity with only 21.49% symmetric links, whereas the percentage of symmetric links in the largest connected component is 23.18%. Our results align very well with those reported in [19] for reciprocity in Twitter. Following similar reasoning to [19], we conjecture that users tend to share information with their colleagues in a broadcast manner, rather than exchanging one-to-one messages, even if a conversation is initiated with a directed message between two users. Further validation is out of

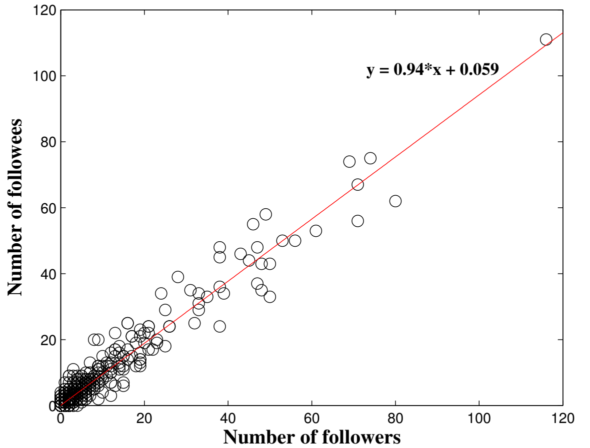

Figure 2. Scatter plot of the number of followers and the number of followees.

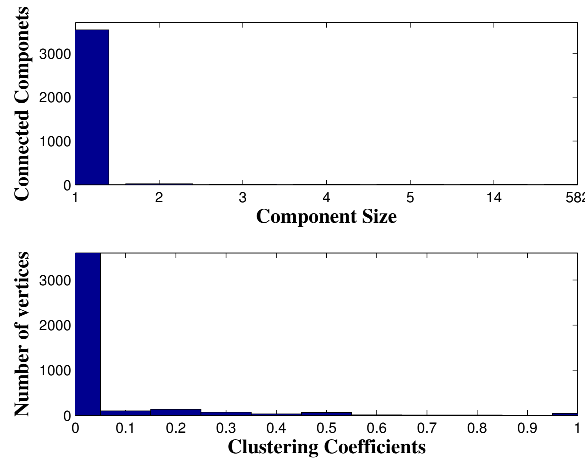

Figure 4. Top: Histogram of connected components sizes. Bottom: Histogram of clustering coefficients.

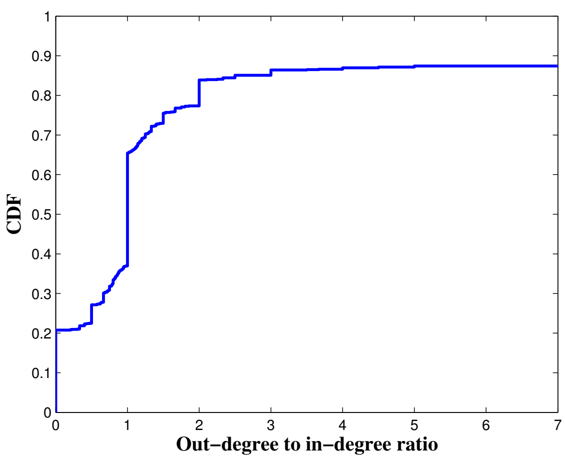

Figure 3. CDF of out-degree to in-degree ratio.

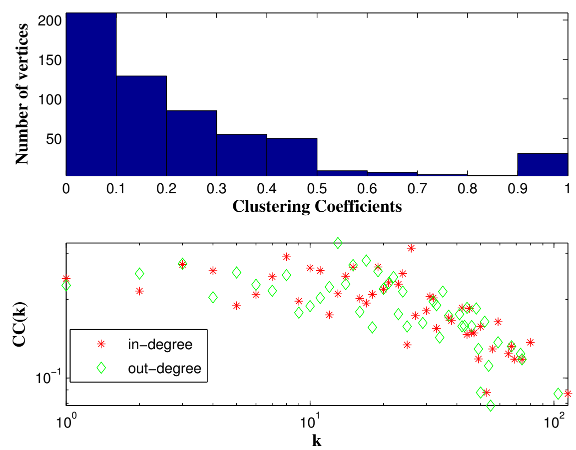

Figure 5. Top: Histogram of clustering coefficients in LCC. Bottom: Average clustering coefficient as a function of degree in LCC.

scope of this paper and we leave it for future work.

We now examine clustering, which quantifies how densely the neighborhood of a node is connected. Not all nodes are connected in one cluster. There are $N_{cc} = 3,570$ connected components, with the largest component encompassing $N_{cc^{max}} = 582$ nodes (13.8% of the network). Figure 4 (top) shows the histogram of connected component sizes. The clustering coefficient of node u, with set S_N of N neighbors, is defined as the number of directed links that exist between nodes in S_N, divided by the number of all possible directed links $N * (N-1)$ between the nodes in S_N. The clustering coefficient of the network, c, is 0.0335. Figure 4 (bottom) shows the histogram of individual clustering coefficients at each node. Due to the fact that we compute this metric over the complete network, the clustering coefficient of the graph is low. However, in a random network with the same number of nodes (N) and degree (d), $c = \frac{d}{N} = 0.0005$ [20]. Figure 5 (top) shows the histogram of connected component sizes for the largest connected component, which contains 582 nodes, with an average node degree $d_{LCC} = 12.97$. The clustering coefficient of the largest connected component is $c_{LCC} = 0.2311 \gg c_{random} = 0.0223$. Figure 5 (bottom) shows how the clustering coefficient of nodes vary as a function of node degree. The average clustering coefficient follows a decreasing trend with increasing node degree. It is higher for nodes of low degree, suggesting significant clustering among low-degree nodes. This evidence of strong local clustering supports the intuition that people tend to be introduced to others via mutual contacts, thus increasing the probability of two neighbors u and v of user z to be

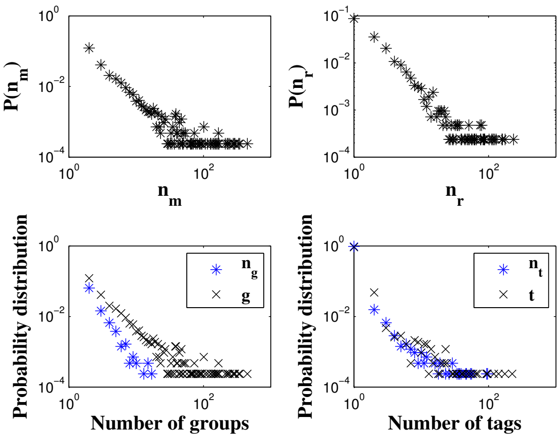

Figure 6. From left to right and top to bottom, distribution of the number of messages n_m per user, the number of replies n_r per user, the number n_g of distinct groups to which a user's messages belong and the number g of group related messages per user, the number n_t of distinct hashtags per user and the number t of hashtag assignments per user.

connected themselves [1], thus exchanging directed @reply messages in this case.

Next, we look at the properties of shortest paths between users in the large weakly connected component. As only 21.49% of links are reciprocal, we expect the average path length between any two users to be longer than other known networks. The average path length is $D_{av} = 3.5677$ and the diameter is $D_{max} = 11$. Even though the graph is directed, these values are remarkably short and quite similar to corresponding values for Flickr and Orkut [1]. In a random network with the same number of nodes (N) and degree (d), $L_{random} = \frac{ln(d)}{ln(N)} = 10.9699$ [20]. The small diameter ($D_{max} \approx L_{random}$) and the strong local clustering of this network ($c \gg c_{random}$) qualify the network of @replies as small-world network [21], and further indicate that the graph has scale-free properties.

V. Content Analysis

In this section we take a close look at the content aspect of the @replies network, focusing on numerous user activities. We further investigate the correlations between such activities. Figure 6 shows the probability distributions of the number of messages n_m and the number of replies n_r per user, the distribution of the number of groups n_g to which a user belongs and the probability of finding a user with a number n_t of distinct hashtags in his vocabulary.

Figure 6 further shows the total number t of hashtag assignments per user (a hashtag used twice is counted twice) and the total number g of group related messages per user (the number of messages sent to a group instead of binary group membership). More precisely, if $f_u(t)$ is the frequency of hashtag t being used by user u, then the total number of hashtag assignments of user u is given by: $t_u = \sum_t f_u(t)$. Similarly, if $f_u(g)$ is the number of times user u has sent a message to group g (either privately to another group member or broadcast to the group), then the total number of group messages of user u is given by: $g_u = \sum_g f_u(g)$.

Activity x	E[x]	Var[x]
in-degree	1.0197	25.2624
out-degree	1.0197	23.6404
n_m	3.9195	321.6220
n_r	1.5196	84.7008
g	3.9067	321.3277
n_g	1.1721	0.6040
t	1.5884	39.7527
n_t	1.2601	9.1692

Table II
Averages and Fluctuations of user activities.

All activities show behavior consistent with power law networks; the majority of users show small activity patterns with few nodes being significantly more active. All distributions are broad, indicating that the activity patterns of users are highly heterogeneous. For reference, Table II reports the mean and variance of different activities. The average number of group related messages (g) seems to be restricted on average, but its variance is quite high. The use of hashtags (t) is relatively low with the average being $\approx$ 1.6 hashtags per message, which makes tagging even scarcer than in traditional social networks like Flickr. Our results have lower values than those of traditional online social networks [1]. This difference can be explained as a result of not restricting our analysis only to users for whom we have messages, hashtags, group participation and contact information, contrary to [10].

A. *Correlations Between Features*

Marlow et al. [9] reported that some Flickr user activities are correlated (e.g. the number of photos uploaded by a user is strongly correlated with the number of hashtags from the same user). We wanted to test the validity of this hypothesis for our @replies network. Figures 7 and 8 respectively show the number of group related messages and the number of hashtag assignments as a function of the number of messages n_m and replies n_r of a user. Clearly, g exhibits strong correlation to the number of messages (replies). Frequent communication between group members is expected, since joining a group is likely driven by a business need. However, we cannot conjecture the same for hashtag assignments. There seems to be a close relationship on a logarithmic scale, but such relationship is not perfectly linear. Even though there exist many users exhibiting high activity patterns with respect to the number of messages (replies) they send, such users do not tend to tag their messages as often. Further, users tend to mostly tag their own content, as in Flickr [22]. It seems reasonable to deduct that group co-membership is

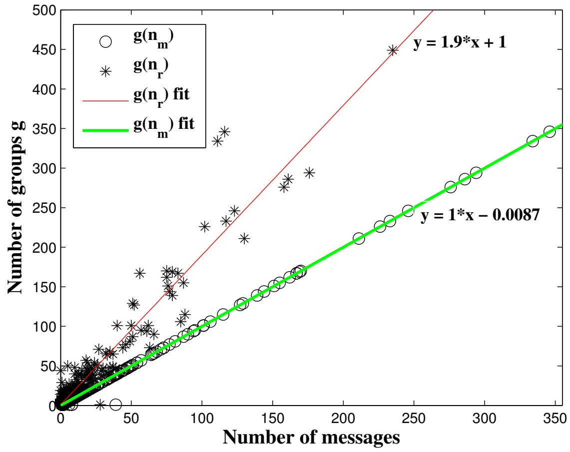

Figure 7. Number of group messages g, as a function of the number of messages n_m and replies n_r of a user.

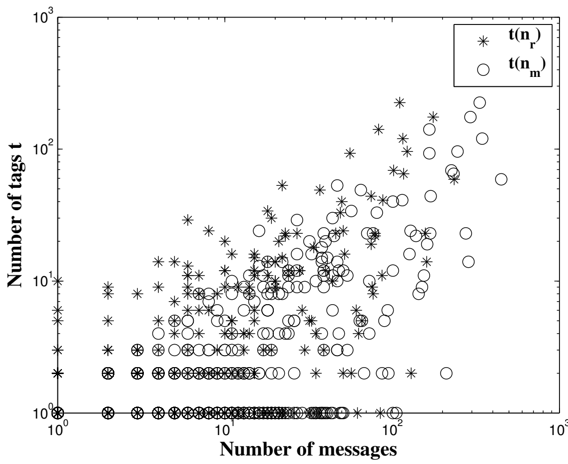

Figure 8. Number of hashtag assignments t, as a function of the number of messages n_m and replies n_r of a user. Both axes are in logarithmic scale.

a natural and direct indicator of shared users' interests in a corporation. Shared hashtags can also be considered as indicator of shared interests, but with some caution.

B. Correlations Between Features and the Network

We now examine correlations between users activities and the structure of the @replies network. Specifically, we investigate if there is a connection between the number of neighbors a user has and the activity patterns of such user (i.e. number of messages, number of replies, group participation and tagging). We characterize the average activity of users with k neighbors (we consider in-degree and out-degree separately), using the following quantities: (i) the average number of messages of users with k neighbors $n_m(k)$, (ii) the average number of replies of users with k neighbors $n_r(k)$, (iii) the average number of distinct groups (similarly for total number of group messages) of users with k neighbors $n_g(k)$, (iv) the average number of distinct hashtags (similarly for total hashtag assignments) of users with k neighbors $n_t(k)$. For example, $n_m(k) = \frac{1}{|u{:}k_u=k|}\sum_{u:k_u=k} n_m(u)$.

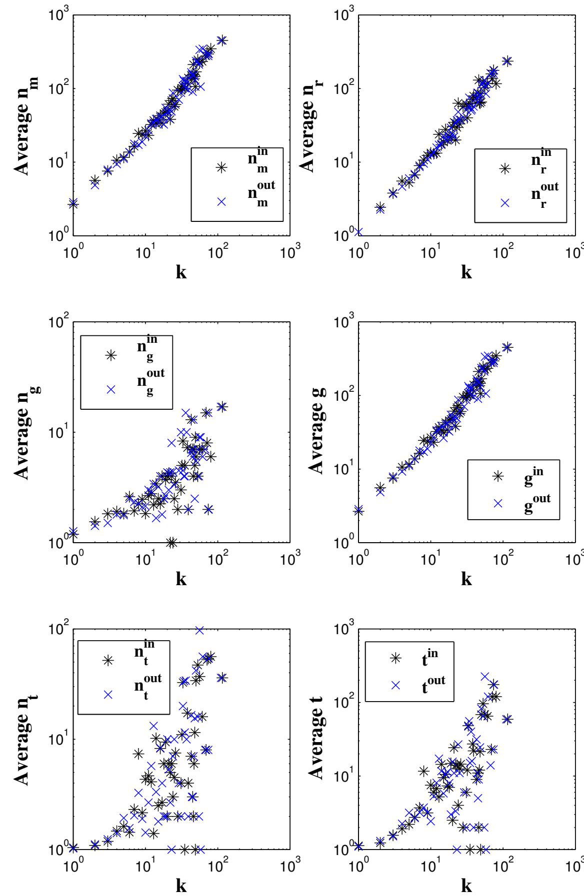

Figure 9. From left to right and top to bottoms, average number of (a) messages n_m, (b) replies n_r, (c) distinct groups n_g and (d) groups g, (e) distinct hashtags n_t and (f) total hashtag assignments t of users having k neighbors in the @replies network.

Figure 9 shows the probability distributions of such quantities. All activities have an increasing trend for increasing values of k (both for in-degree and out-degree). Large fluctuations can be observed for large values of k due to the fewer highly connected users over whom the averages are performed. Notably, the average number of messages and replies are very well correlated to the number of neighbors, as is the average number of distinct groups. The average number of (distinct) hashtag assignments exhibits more

Parameters x, y	$PCC(x,y)$
k_{in}, n_m	0.9779
k_{in}, n_r	0.9608
k_{in}, n_g	0.7543
k_{in}, g	0.9779
k_{in}, n_t	0.7157
k_{in}, n_t	0.6978
k_{out}, n_m	0.9397
k_{out}, n_r	0.9767
k_{out}, n_g	0.6967
k_{out}, g	0.9397
k_{out}, n_t	0.5407
k_{out}, n_t	0.53

Table III
PEARSON CORRELATION COEFFICIENTS.

heterogeneity than the other measures, but still the trend is increasing with increasing values of k. Users with many contacts but using very few hashtags and sending very few group messages can be observed. For reference, Table III reports the Pearson correlation coefficients, measured for k (both for in-degree and out-degree) and all user activities.

C. Lexical Alignment

We now examine user similarity in terms of hashtag usage, with respect to their distance in the @replies network. We argued earlier that users of the corporate microblogging service mostly tag their own content. This observation along with the personal character of tagging make us conjecture that there will be no global hashtag vocabulary across users, or if such a vocabulary exists, it will be extremely incoherent. Hence, we do not anticipate an emergent globally accepted hashtag vocabulary, commonly found in social bookmarking sites [9], [23]. To test the existence of a globally shared vocabulary, we selected pairs of users at random and measured the number of their shared hashtags, which on average is ≈ 1.001.

Even though random pairs of users don't have common hashtags, adjacent users in social networks tend to share common interests, a property known as homophily [24], [25] or assortative mixing [26]. We measure user homophily with respect to hashtags as a function of the distance of users in the @replies network. We regard hashtag assignments of user u as a feature vector, whose elements correspond to hashtags and whose entries correspond to frequencies of hashtag usage for user u. Hence, the normalized similarity between two users u and v with respect to their hashtag vectors, $\sigma_{hashtags}(u,v)$ can be computed as follows:

$$\sigma_{hashtags}(u,v) = \frac{\sum_t f_u(t) f_v(t)}{\sqrt{\sum_t f_u(t)^2 \sum_t f_v(t)^2}}, \quad (1)$$

where $f_u(t)$ denotes the number of times user u has used hashtag t. $\sigma_{hashtags}(u,v)$ is equal to 0 if users u and v have no hashtags in common, and 1 if they have used exactly the

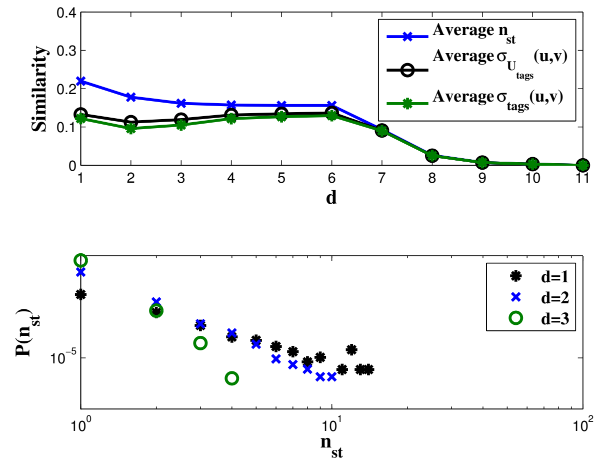

Figure 10. Top: Average number of shared hashtags n_{st}, $\sigma_{hashtags}(u,v)$, and $\sigma_{U_{hashtags}}(u,v)$ of two users as a function of their distance d in the network. Bottom: Probability distribution of the number of shared hashtags n_{st} of two users being at distance d on the network, for $d = 1, 2, 3$.

same hashtags. We further define the normalized similarity between two users u and v with respect to their distinct hashtag usage, as:

$$\sigma_{U_{hashtags}}(u,v) = \frac{\sum_t \delta_u^t \delta_v^t}{\sqrt{n_t(u) n_t(v)}}, \quad (2)$$

where $n_t(u)$ is the total number of distinct hashtags of user u and $\delta_u^t = 1$ if user u has used hashtag t at least once, and 0 otherwise.

To compute averages of the aforementioned similarities we performed an exhaustive investigation of the @replies network up to distance equal to the network diameter (D_{max}). Figure 10 demonstrates the dependency of user similarity on distance, by showing the average number of shared hashtags and the corresponding average cosine similarities of two users as a function of d. The average number of shared hashtags remains almost constant for $d \leq 6$, after which point it drops rapidly. High lexical alignment is observed between neighbors for greater distance than traditional online social networks [10], due to the fact that in a corporate environment users exhibit more focused interests aligned with their discipline, day to day responsibilities and ongoing projects.

We examine user homophily with respect to groups as a function of their distance, following similar reasoning. In particular, we define the normalized similarity between two users u and v with respect to their group participation as:

$$\sigma_{U_{groups}}(u,v) = \frac{\sum_t \delta_u^g \delta_v^g}{\sqrt{n_g(u) n_g(v)}}, \quad (3)$$

where $n_g(u)$ is the number of groups of which user u is

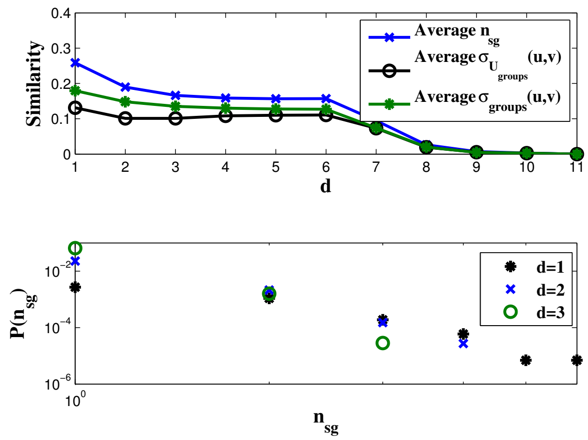

Figure 11. Top: Average number of (i) shared groups n_{sg}, (ii) $\sigma_{groups}(u,v)$, and (iii) $\sigma_{U_{groups}}(u,v)$ of two users as a function of their distance d in the network. Bottom: Probability distribution of the number of shared groups n_{sg} of two users being at distance d on the network, for $d = 1, 2, 3$.

a member and $\delta_u^g = 1$ if user u belongs to group g, and 0 otherwise (a user belongs at most once to a group). We also examine user similarity in terms of messages sent to common groups. We calculate this similarity, as follows:

$$\sigma_{groups}(u,v) = \frac{\sum_g f_u(g) f_v(g)}{\sqrt{\sum_g f_u(g)^2 \sum_g f_v(g)^2}}. \quad (4)$$

Figure 11 demonstrates the dependency of user similarity on distance, allowing us to draw similar conclusions for shared groups, as for shared hashtags.

VI. Conclusion

In this paper we provided an extensive analysis of enterprise microblogging data. The extracted network of directed @-messages sent between users of a corporate microblogging service that we examined is a "smaller" world than online social networks, has a strongly connected core of high-degree nodes, and exhibits strong positive correlation to users degree (both in-degree and out-degree). We further showed that strong correlations exist between user activities, and that users alignment in terms of their hashtag vocabulary and group co-membership is more profound than in online social networks, for greater distances.

The primary focus of corporate microblogging services is narrower than traditional social network sites like Twitter, Facebook, and Flicker. Corporate microblogging services mostly emphasize on the business perspective and therefore their content revolves around the work culture, work practices, and everyday problems (technical or otherwise related to business). Discussions are often about problem solving, relevant emerging techniques, applications and technologies, trends, etc. Usually, some users lead the overall discussion by expressing their opinion on a matter, which then triggers replies. The existence of high-degree nodes that we observed in the @replies network confirms this behavior and suggests that such high-degree nodes are critical for the connectivity and flow of information in this context. On the other hand, information searches that exploit the social structure rapidly reach the core. The design of algorithms for information search or expertise identification in this context should consider this observation.

Conversely to general purpose online social networks, trust is a minor issue in corporate microblogging services, where malicious users cannot penetrate the core due to restricted access. Trust is often used as an indicator of expertise, hence newly hired employees may pursue highly ranked positions in the @replies network by providing unconvincing and/or unhelpful responses (i.e. a form of "spamming"), contributing to discussions nonetheless. We conjecture that a "new" user should be highly trusted not only if multiple short disjoint paths to the user can be discovered [1] but also if the overall impact and positive sentimental response that her replies trigger are sufficiently large.

One possible criticism of our study is that it does not account for network evolution. Our dataset spans a time period between July, 2010 and August, 2011. During this time frame, the network grows rapidly. However, our observations remain valid throughout this time period, indicating that the basic network structure does not drastically change over time.

In conclusion, we have presented our quantitative study of enterprise microblogging data at scale, where we examined 1) the network structural characteristics and 2) users alignment with respect to content. We concluded by discussing the implications of our findings. However, examination of multiple other corporate social networks is required to further confirm our findings.

Acknowledgment

We are thankful to Roger Cutler, Vikram Sorathia, and Hao Wu for constructive feedback on this work. This work is supported by Chevron Corp. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

References

[1] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, ser. IMC '07. New York, NY, USA: ACM, 2007, pp. 29–42.

[2] J. Zhang, Y. Qu, J. Cody, and Y. Wu, "A case study of microblogging in the enterprise: use, value, and related issues," in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 123–132.

[3] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '06. New York, NY, USA: ACM, 2006, pp. 611–617.

[4] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of the 16th international conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 835–844.

[5] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[6] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11 149–11 152, October 2000.

[7] M. E. J. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. pp. 404–409, 2001.

[8] J. Kleinberg, "The small-world phenomenon: an algorithm perspective," in *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, ser. STOC '00. New York, NY, USA: ACM, 2000, pp. 163–170.

[9] C. Marlow, M. Naaman, D. Boyd, and M. Davis, "Ht06, tagging paper, taxonomy, flickr, academic article, to read," in *Proceedings of the seventeenth conference on Hypertext and hypermedia*, ser. HYPERTEXT '06. New York, NY, USA: ACM, 2006, pp. 31–40.

[10] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer, "Folks in folksonomies: social link prediction from shared metadata," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 271–280.

[11] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about twitter," in *Proceedings of the first workshop on Online social networks*, ser. WOSN '08. New York, NY, USA: ACM, 2008, pp. 19–24.

[12] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, ser. WebKDD/SNA-KDD '07. New York, NY, USA: ACM, 2007, pp. 56–65.

[13] D. Zhao and M. B. Rosson, "How and why people twitter: the role that micro-blogging plays in informal communication at work," in *Proceedings of the ACM 2009 international conference on Supporting group work*, ser. GROUP '09. New York, NY, USA: ACM, 2009, pp. 243–252.

[14] A. Wu, J. M. DiMicco, and D. R. Millen, "Detecting professional versus personal closeness using an enterprise social network site," in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1955–1964.

[15] K. Ehrlich and N. Shami, "Microblogging inside and outside the workplace," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '10. AAAI, 2010.

[16] J. DiMicco, D. R. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller, "Motivations for social networking at work," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 711–720.

[17] I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, and S. Farrell, "Harvesting with sonar: the value of aggregating social network information," in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1017–1026.

[18] J. Huang, K. M. Thornton, and E. N. Efthimiadis, "Conversational tagging in twitter," in *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, ser. HT '10. ACM, 2010, pp. 173–178.

[19] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 591–600.

[20] A.-L. Barabsi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[21] D. J. Watts and S. H. Strogatz, "Collective dynamics of /'small-world/' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[22] T. Vander Wal, "Explaining and showing broad and narrow folksonomies," Personal InfoCloud.com, February 2005, www.personalinfocloud.com/2005/02/explaining_and_html.

[23] P. Mika, "Ontologies are us: A unified model of social networks and semantics," *Semantic Web*, vol. 5, pp. 5–15, March 2007.

[24] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 261–270.

[25] M. McPherson, L. S. Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[26] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E*, vol. 67, p. 026126, Feb 2003, http://link.aps.org/doi/10.1103/PhysRevE.67.026126.