

Multi-Scale Characterization of Social Network Dynamics in the Blogosphere

Munmun De Choudhury Hari Sundaram Ajita John Dorée Duncan Seligmann
Arts Media & Engineering, Arizona State University Collaborative Applications Research, Avaya Labs
Email: {munmun.dechoudhury,hari.sundaram}@asu.edu, {ajita,doree}@avaya.com

ABSTRACT

We have developed a computational framework to characterize social network dynamics in the blogosphere at individual, group and community levels. Such characterization could be used by corporations to help drive targeted advertising and to track the moods and sentiments of consumers. We tested our model on a widely read technology blog called Engadget. Our results show that communities transit between states of high and low entropy, depending on sentiments (positive / negative) about external happenings. We also propose an innovative method to establish the utility of the extracted knowledge, by correlating the mined knowledge with an external time series data (the stock market). Our validation results show that the characterized groups exhibit high stock market movement predictability (89%) and removal of ‘impactful’ groups makes the community less resilient by lowering predictability (26%) and affecting the composition of the groups in the rest of the community.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications – *Data mining*; H.3.5 [Information Storage and Retrieval]: Online Information Services—*Web-based services*; H.5.4 [Information Interfaces and Representation]—*Hypertext/Hypermedia*; J.4 [Computer Applications]: Social and Behavioral Sciences – *Sociology*

General Terms

Algorithms, Experimentation, Human Factors, Verification.

Keywords

Blogosphere, multi-scale characterization, community, social network analysis, stock market movement.

1. INTRODUCTION

This paper deals with the problem of characterizing communication dynamics of the blogosphere at multiple scales – individual, group and community. Secondly, we propose an innovative method to establish the utility of the extracted knowledge, by correlating the mined knowledge with an external time series data (the stock market). The problem is important because multi-scale characterization of social networks dynamics could be used by corporations to help drive targeted advertising and to know how they and their products are being received. Communication characterization within a network could also suggest alternative ranking strategies for ‘impactful’ individuals, groups and communities in social networks.

Our key observation is that a fine-grained, multi-scale, temporal analysis can help us understand the evolutionary aspects

(organization, dynamics) of social networks as *an emergent property of individual and group interaction*. Moreover, we believe that conventional techniques of validating social group extraction do not relate with the usefulness of the extracted properties of the communities. In this paper, we therefore present a novel validation framework which relies on correlation of extracted characteristics with an external source of orthogonal but semantically related data (stock market movement). Our main contributions are as follows:

- a. Characterizing social networks dynamics at individual, group and community levels,
- b. Validating the multi-scale characterization with respect to its compliance to: (i) degree of *predictability* of external events (here we consider predictability of stock market movements of companies by the characterized communities), and (ii) degree of *resilience* of the characterized communities due to removal of impactful groups.

2. MULTI-SCALE CHARACTERIZATION

We define multi-scale characterization to be the ability to capture the diverse aspects of participants in communication at different levels – individuals, groups and communities.

We have chosen to investigate the technology blog, Engadget [1]. It primarily discusses topics related to consumer electronics. These discussions allow us to develop an abstract co-occurrence based social network using its blog posts and comments. Based on this social network, we now present our computational framework of multi-scale characterization.

Individual Characterization: We extract three kinds of features for individuals – activity characteristics, influence on local neighborhood and spatial graph based measures.

We consider a three dimensional activity characterization for individuals. First, the responsivity of an individual is defined as an additive function which depends on *when* she responds and the relative *rank* of her comment among the list of all comments on a blog post. Second, the degree of participation in communication of an individual is the ratio of the number of comments to the number of posts. Third, we assume that a person’s comment on a blog post could incur two kinds of consequences: (a) the person’s comment is followed by a large number of direct replies and / or (b) the mean response time of all comments after her comment is greatly reduced compared to that before her comment. Hence the consequence of participation for an individual is a weighted additive function of these two variables.

Next we characterize people on the basis of how they influence their local neighborhood to communicate. The influence is measured using two attributes: the mean maximal probability path

and the mean minimal delay path of an individual with respect to her neighborhood. We also note that a person comments based on her independent desire when her comment does not follow another comment within a response time difference of δ .

We also characterize people with respect to several traditional graph based properties: degree centrality, betweenness centrality, closeness centrality, eigenvector centrality and clustering coefficient along with intrinsic roles of people – ‘leaders’ and ‘followers’.

Group Characterization: Groups are sets of people whose members share similar social network characteristics. The similarity between a pair of individuals is defined using a diffusion kernel. We thereafter deploy an unsupervised clustering based group extraction algorithm [3] known as ‘mutual awareness expansion’ (MAE) to extract groups over each time slice. For each such group, we extract several features. We define the individual characteristics at the group-level so as to characterize the groups at a macro-scale based on a ‘group graph’ which is a social network where each node is a group of individuals. We also define two other group-level characteristics: assortativity and characteristic path length. We now categorize groups into sets (called ‘types’) which share similar group-level characteristics. It is a global characterization, as the group-types are extracted over all time slices. In order to extract the group-types we use the spectral clustering algorithm.

Community Characterization: A community comprises groups of different ‘types’ related a particular topic. In our work, the topics are taken as the tags of the names of technology companies. We define five different measures to characterize the group-types in a community: temporal density, conductance, coverage, group-type entropy and size entropy. *Temporal density* of a group-type in a community determines how many individuals belong to its member groups at each time slice. *Group-type entropy* is used to determine the characteristics of a community at a particular time slice based on the frequency of member groups in its group-type distribution; while *size entropy* is used to determine the characteristics based on the sizes of the member groups.

3. EXPERIMENTAL RESULTS

We have analyzed postings on the widely read technology blog, Engadget (www.engadget.com) related to names of technology companies. In order to understand the dynamics of the multi-scale characterization of these topic communities, we refer to several external events related to the companies. Figure 1 gives a temporal representation of the multi-scale characterization for the topic ‘Apple’, over a span of 20 weeks from Sep 9, 2007 to Jan 20, 2008. The figure shows (a) the groups at each week (as bubbles), and (b) the group-types that each group belongs to (the colors of the bubbles). We denote these non-communicating people as a white bubble. We observe that the community transits between states of high and low entropy, depending on sentiments (positive / negative) about external happenings.

Activity dynamics further show that the community discussing topic Apple is observed to be a moderately responsive community ($r=0.74$) where communication is more editor-driven. While another community discussing topic Google is observed to be a consistent, highly responsive ($r=0.73$), cohesive and high influence community ($\pi=0.69$).

Validation: Our use of company names as topics has enabled us to further validate our findings against stock market movements of the two companies. For the purpose, the stock market returns for the four companies (as per the NASDAQ index) were collected from Google Finance [2]. The validation follows from the fact that the characterized groups exhibit high stock movement predictability (89%) and removal of groups with high predictability makes the communities less resilient by decreasing mean predictability by 26% and yielding low group composition entropy of 0.21.

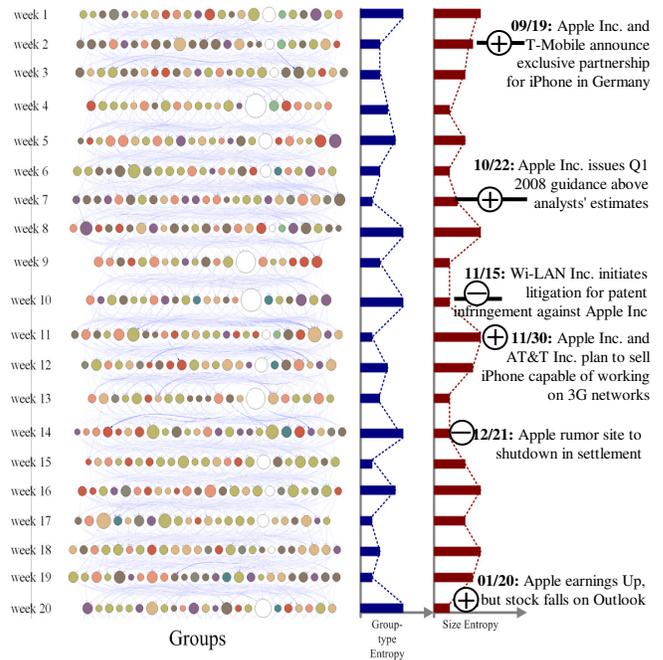


Figure 1: Temporal dynamics of multi-scale characterization for the topic community Apple. Group evolution is shown over 20 weeks and associated are shown external news events (+ / -), group-type and size entropies at each week.

4. CONCLUSIONS

In this paper, we have developed a computational framework to characterize social network dynamics in the blogosphere at multiple scales: individual, group and community levels. We have characterized social networks dynamics at individual, group and community levels. We have excellent results on a gadget-discussing blog Engadget. We further validated the framework using predictability of stock market movements of companies and change of group composition entropy. Our framework and results can serve as a starting point to several interesting future work. The different ‘states’ of the topic communities can also be used as predictors of mood or sentiment in response to external events.

5. REFERENCES

- [1] Engadget <http://www.engadget.com/>.
- [2] Google Finance <http://finance.google.com/finance>.
- [3] Y.-R. LIN, H. SUNDARAM, Y. CHI, et al. (2007). *Blog Community Discovery and Evolution Based on Mutual Awareness Expansion*. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society: 48-56.