Complex Modeling and Analysis of Workplace Collaboration Data

Charalampos Chelmis
Department of Computer Science
University of Southern California
Los Angeles, CA USA
Dissertation Advisor:Viktor K. Prasanna

DOCTORAL DISSERTATION COLLOQUIUM

EXTENDED ABSTRACT

***Abstract*—Complex networks arise everywhere. Online social networks are famous complex networks examples due to (a) revolutionizing the way people interact on the Web, and (b) permitting in practice the study of interdisciplinary theories that arise from human activities, at both micro (i.e. individual) and macro (i.e. community) level. The vast scale (Big–data) of online human interactions impose certain challenges, such as scalable indexing and efficient retrieval of social data, which are by their nature intertwined in multiple dimensions. In our research we focus on modeling such multidimensional data, mining their intra and inter dependencies to uncover hidden structures and emergent knowledge. In particular, we examine informal interactions at the workplace. Through extensive empirical analysis of corporate communication logs we study users' communication behavioral patterns, dynamics and characteristics, statistical properties and complex correlations between social and topical structures. Our modeling and analysis are not limited to enterprise social data, but are extensible and applicable to other domains, offering a unified framework of complex network modeling and analysis, accurately modeling multiple symmetric or asymmetric, explicit and hidden interaction channels between people.**

***Keywords*—Collaborative Knowledge Management; Collaborative Decision Making and Support; Collaboration Technologies in Industry and Businesses; Social Media Applications; Social Multimedia and Networks**

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. 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 mainly focused on publicly available online social networks [1]. However, microblogging capabilities have been adopted by the enterprise as well [2], [3]. Contrary to online social networks, microblogging services for enterprises are primarily designed to support employees in connecting and learning about each other through personal and professional sharing [4]. Organizations hosting internal social networking sites can benefit from mining employees' informal interaction logs, as well as from identifying reliable indicators of expertise.

Enterprise collaboration platforms mainly focus on the business perspective and therefore their content revolves around their main business and work culture, work practices, and everyday problems. Being more focused and less noisy than online social networks, enterprise microblogging platforms offer unique opportunities for identifying experts, as well as studying and understanding the processes of knowledge generation and sharing. Studying collaboration media that are internally used by organizations can result in multiple benefits both for employees and corporations. Employees can "get help or advice, reach opportunities beyond those available through existing ties, discover new routes for potential career development, learn about new projects and assets they can reuse and leverage, connect with subject matter experts and other influential people within the organization, cultivate their organizational social capital, and ultimately grow their reputation and influence within the organization" [5]. Similarly, enterprises can benefit from increased productivity due to reduced time spend in team building and knowledge propagation, as well as expertise mining and preservation [6]. Further, enterprises can utilize the results stemming out of informal interactions analysis, to better understand how their employees work together to complete tasks or produce innovative ideas, reveal trends, identify experts and influential individuals, so as to evaluate and adjust their management strategy, team building and resource allocation policies.

Enterprises have been mainly relying on e-mail traffic to share information among coworkers [7]. Analysis of enterprise communication networks [8], [9] has broadened our understanding of information flow in the enterprise. Media

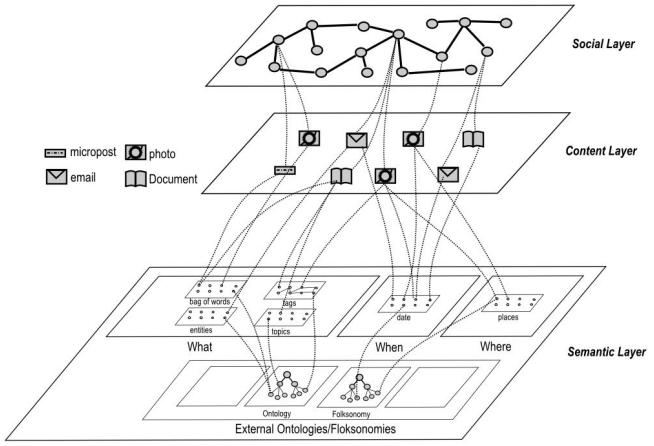

Figure 1. Holistic Modeling of Complex Networks

like SharePoint and Office Communicator are heavily utilized as part of question-answering and problem solving processes, while Active Directory provides a formal structure for employees to comprehend and navigate through the organizational hierarchy, accommodating their need to identify potential collaborators, research teams and business units around the globe, as well as to discover "interesting" projects that others are currently working on. The wealth of information available in the context of enterprises however, is not limited to formal interactions and silos containing structured data. Informal interactions through corporate micro-blogging services span organizational charts, facilitating knowledge generation in the form of direct problem solutions or links to external sources. Knowledge in such cases may not have been formally modeled and captured until a specific question is being asked and an expert answer is provided.

II. Holistic Modeling of Complex Networks

We have proposed a formal model [6], [10] that abstracts the semantics of informal communication into an integrated, context aware, time sensitive, multi-dimensional space, enabling the correlation of seemingly different domains so as to investigate them in conjunction. We introduce a novel social graph representation, shown in Figure 1, which not only contains social links between users but also maintains integrated information regarding users dynamically changing interests and activities, throughout collaboration tools used in working environments.

a) Social Layer: captures users contextual and temporal interactions. Nodes represent users and arcs represent explicit relationships (links) between them. The social layer records user relationships from multiple collaboration platforms: e-mail correspondence, chat networks, blogging activity, shared bookmarks and common ratings. An edge between users is defined by the context under which it is created and has an associated timestamp. A user u may be connected to user v under multiple contexts (e.g sending e-mails and social status updates) in multiple time instances.

In order to process enterprise communication effectively, it is imperative to establish a comprehensive model of contexts, and semantic links to structure them [10]. Various interpretations of captured scalar, hierarchical, and nominal, temporal, or spatial data that differ with context (e.g. point of view) can provide different insights or views (i.e. dimensions). Figure 2 shows various contexts that can be utilized for advanced analysis of corporate communication data.

b) Content Layer: captures published content from multiple sources, including but not limited to resources shared by users (e.g. photos or videos), bookmarked and/or tagged resources (e.g. URLs), users' generated content (e.g. status updates in Facebook), e-mails, chat messages, and blog posts.

c) Semantic Layer: contains meta-information about content, and can be broken into several constituting layers, each containing different metadata about content. This layer may include, but is not limited to, domain ontologies, vocabularies, and folksonomies and taxonomies, external sources of formal knowledge, and linked open data. We use semantic information providers and annotation enablers, such as OpenCalais, AlchemyAPI, and Evri, WordNet, and Freebase, for text analysis and annotation, entity identification and topic discovery. Linked Open Data (LOD) can further be exploited to link to external knowledge repositories.

Our novel model entangles structural information from multiple heterogeneous networks, with arbitrary, multimodal attributes originating from various information sources, and content that has been generated under specific context [6], [10]. Knowledge is discovered, captured and inferred based on such complex information. Our social algebraic operators [6] facilitate dimensional and contextual analysis and mining of social data from multiple views and perspectives, enabling deep exploration and understanding of both explicit and implicit interactions and relations.

Our model is not limited to enterprise social data, but is extensible and applicable to other domains, accurately modeling multiple symmetric or asymmetric, explicit and hidden interaction channels between people. It can capture multimodal, heterogeneous attributes of users and their interrelations, and facilitate complex mining and analysis of evolving, multidimensional, intertwined, auxiliary social networking data.

III. Applications

The full power of our model can be perceived at the context of corporate networks that emerge from explicit formal relations as well as informal collaboration. Towards this direction, we have instantiated our model with exogenous to the social network information, i.e. (a) formal organizational structure, and (b) hints about employees' interests and expertise areas using latent topic descriptors, learned from enterprise communication logs [10]. This facilitates multimodal, semantic, temporal and adaptive analysis of roles along and across dimensions, and permits micro–macro analysis [10].

We demonstrated the effectiveness of our framework on a large scale dataset of corporate communication logs from a Fortune 500 company [10]. Our dataset includes $9,855$

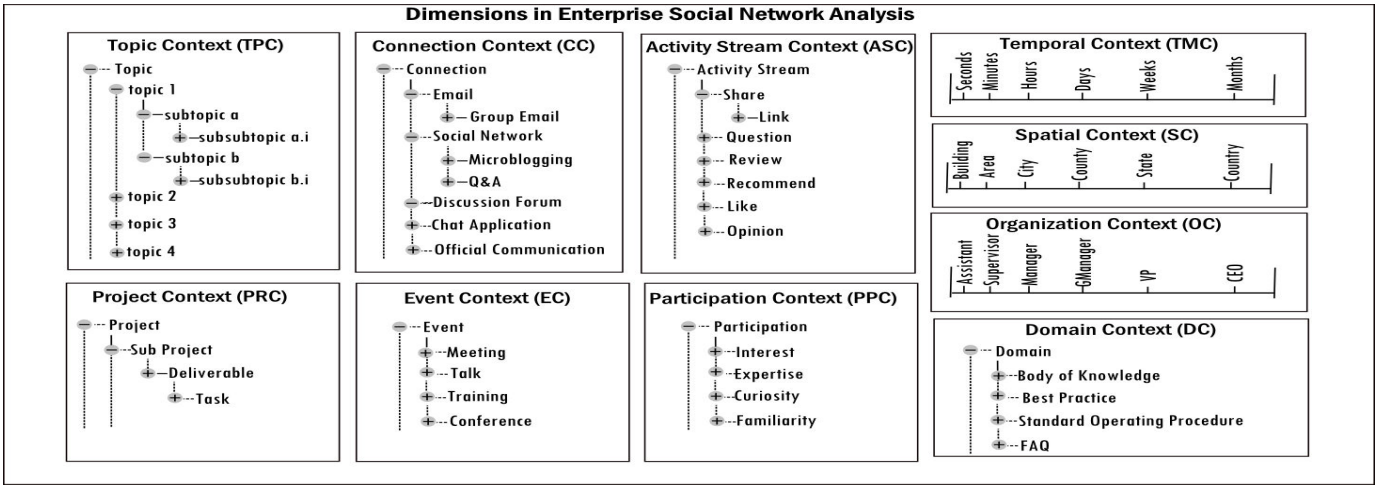

Figure 2. Contexts in informal interactions at the workplace

unique users, and $15,200$ messages with explicit reply links to other messages, over a time span of 1.5 years. The dataset contains users activity (message exchanges) and interactions (e.g. comment/reply, tagging), but lacks explicit social relationships (followee/follower) between users. We compensate the lack of social relationships with a snapshot of the formal organizational hierarchy of users.

A. Enterprise Wisdom Captured Socially

Our model facilitates integration, capturing, search, and retrieval of dynamic and constantly evolving enterprise knowledge captured from informal interactions. In fact, our model enhances collaborative data analysis in the enterprise, revealing latent topics, expertise, and interests, both at micro (single user) and macro (e.g department or business unit) level. We have evaluated our system in the tasks of (a) expert identification, (b) community detection, (c) trends mining, and (d) communication dynamics [10].

1) Multidimensional Expert Identification: Employees' level of expertise may vary from topic to topic and from medium to medium. One might share innovative ideas and contribute to discussions through emails, but not in microblogging sites. We differentiate expertise according to communication channels, time and content type.

2) Context–Sensitive Community Detection: Analysis of employees' interactions reveals hidden dynamics at various granularities. The topic–restrained ego–network of different users may reveal patterns that vary across topics, whereas by clustering users with similar interests, it is possible to detect virtual communities (macro level).

3) Mining of Trends: Discovery of trending topics is a typical application in social network analytics. Typically, trending topics are identified for a single communication channel for single network. Such analysis may be inefficient for an enterprise. Instead, analysis at the macro level can prove to be more useful in discovering, for instance, trending topics and significant contributors in various collaboration platforms, different departments and organizational positions.

4) Communication Dynamics: Organizational hierarchy is static, and dated, whereas communication may create "shortcuts", i.e. collaboration that spans hierarchical levels when seeking for help, or offering guidance, etc. Studying communication may reveal hidden organizational dynamics. Our system enables analysis of information propagation across the rigid structure of formal organizational charts, providing insights on the collaboration processes that employees devise in order to tackle everyday problems and find solutions.

B. Predicting Intention of Communication in Social Media

Social media do not discriminate between users' relationships. Acquaintance relationships invariably model broad relationship types, ranging from trusted friends to total strangers. In reality, relationships fall everywhere along this spectrum and are asymmetric with respect to users. Tie strength is a complicated problem, that depends not only on structural features, but also on user roles in a relationship (e.g. family or friends), context (e.g. time, location) and content (e.g. topic). Further, two users may concurrently maintain multiple relationships (e.g. being friends and coworkers simultaneously). In this sense, tie type and strength is time-varying, and situational, asymmetric and multidimensional. More complex relationships can further been defined, once node types are expanded to include not only people, but entities of any kind.

Our work bridges this gap by integrating various modalities of social interactions and considering such aspects holistically, entangling or disentangling dimensions (i.e. perspectives) when appropriate, in order to perform accurate recommendations. In particular, we examined communication process in social media and the factors that affect this process. We have evaluated our approach, examining its performance in the task of accurate and timely prediction of intention of communication initiation between users in a corporate microblogging service, which resembles Twitter [11]. Our accurate predictor is able to achieve high performance with increasing statistical evidence, leveraging local structural proximity and user–generated content.

IV. Related Work

Online social network have been well studied, leading to a vast literature around this area [12]. Here, we focus on prior work that is mostly related to collaboration media at the workplace. [3] discussed various challenges and solutions in conducting social network analysis in the enterprise. [2] provided a systematic examination of adoption and usage of microblogging in a corporation environment, emphasizing on the perceived benefits of corporate microblogging and barriers to adoption. [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. [4] investigated workplace relationships built between coworkers using microblogging services and determined interaction patterns that signal personal versus professional closeness between colleagues. [14] and [15] examined users microblogging behavior at work as a function of their motivation.

V. RESEARCH STATUS

Building models of complex networks, understanding their rich properties, hidden structures and dimensional interdependencies are not trivial tasks. Our work has made several steps towards these broad directions. We now have a robust model that integrates multiple networks, each with rich properties and hidden dynamics, facilitating multimodal analysis of time–varying, complex social networking data [6], [10]. We also have a better understanding of microblogging behavior in the workplace, and the effect of formal structure in informal communication [11], [16]. Last but not least, we have better models for prediction and recommendation tasks in social media [11].

However many research problems still remain open, and offer prominent future directions. Information diffusion and propagation of topics has been well studied in online social media (e.g. [17]), but remains unexplored in the context of workplace. Communication dynamics at the workplace are relatively unknown. Knowledge capturing in the form of best practices has not been automatically achieved. We plan to address these issues in our future work, while at the same time incorporating qualitative studies and expanding our work to multiple datasets.

Biography

Charalampos Chelmis is a a PhD candidate in the Department of Computer Science at University of Southern California. He received his M.S. from the Computer Science Department, University of Southern California in 2010. He received his Bachelors and M.S. in Computer Engineering and Informatics from Computer Engineering and Informatics Departments, University of Patras, Greece in 2007. His research interests include Complex Social Networking Data Mining and Analysis, Semantic Enrichment of Social Network Data, Collective Intelligence, Social Computing, Information Networks, Big Data Analytics and Enterprise Microblogging.

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 micro-blogging in the enterprise: use, value, and related issues," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 123–132. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753346

[3] C.-Y. Lin, L. Wu, Z. Wen, H. Tong, V. Griffiths-Fisher, L. Shi, and D. Lubensky, "Social network analysis in enterprise," *Proceedings of the IEEE*, vol. 100, no. 9, pp. 2759–2776, sept. 2012.

[4] 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.

[5] I. Guy, S. Ur, I. Ronen, A. Perer, and M. Jacovi, "Do you want to know?: recommending strangers in the enterprise," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, ser. CSCW '11. New York, NY, USA: ACM, 2011, pp. 285–294.

[6] C. Chelmis, V. Sorathia, and V. K. Prasanna, "Enterprise wisdom captured socially," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Istanbul, Turkey, August 26-29 2012.

[7] T. Burkhart, D. Werth, and P. Loos, "Context-sensitive business process support based on emails," in *Proceedings of the 21st international conference companion on World Wide Web*, ser. WWW '12 Companion. New York, NY, USA: ACM, 2012, pp. 851–856.

[8] P. A. Gloor, R. Laubacher, S. B. C. Dynes, and Y. Zhao, "Visualization of communication patterns in collaborative innovation networks - analysis of some w3c working groups," in *Proceedings of the twelfth international conference on Information and knowledge management*, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 56–60.

[9] J. Diesner, T. L. Frantz, and K. M. Carley, "Communication networks from the enron email corpus "it's always about the people. enron is no different"," *Comput. Math. Organ. Theory*, vol. 11, no. 3, pp. 201–228, Oct. 2005. [Online]. Available: http://dx.doi.org/10.1007/s10588-005-5377-0

[10] C. Chelmis, H. Wu, V. Sorathia, and V. K. Prasanna, "Semantic social network analysis for the enterprise," *Computing and Informatics–Special Issue on Computational Intelligence for Business Collaboration*, 2013, in press.

[11] C. Chelmis and V. K. Prasanna, "Predicting communication intention in social networks," in *2012 ASE/IEEE Fourth International Conference on Social Computing (SocialCom)*, Amsterdam, The Netherlands, September 2012.

[12] ——, "Social networking analysis: A state of the art and the effect of semantics," in *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*, October 2011.

[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] W. S. Chow and L. S. Chan, "Social network, social trust and shared goals in organizational knowledge sharing," *Information and Management*, vol. 45, no. 7, pp. 458–465, Nov. 2008.

[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] C. Chelmis and V. K. Prasanna, "Microblogging in the enterprise: A few comments are in order," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Istanbul, Turkey, August 26-29 2012.

[17] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, ser. SOCIALCOM '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 177–184.