Structure and Content: Mining and Analysis of Social Interactions and Collaboration

Petko Bogdanov, Kathy Macropol, Nicholas D. Larusso, Misael Mongiovi and Ambuj K. Singh
Department of Computer Science, University of California at Santa Barbara, Santa Barbara, CA 93106-5110
E-mail: petko@cs.ucsb.edu

Abstract: We propose models and develop data mining algorithms tailored to social and collaboration networks, characterized by both topology and content exchange between participating agents. Our analysis of real world systems demonstrate that conversation content is a good predictor of the structure of communication threads, while content evolution can predict controversies in collaboration. In addition, we combine content analysis, social structure and pairwise agent trust to derive a robust social recommendation framework.

Communication graphs such as twitter and email graphs have information content associated with their communication links. Previous research in the area of graph modeling has mostly ignored this additional information, focusing primarily on past graph structure to predict future links. We take a different approach by examining the relationship between information content and graph structure, proposing a new model that combines both. We apply this model to real world communication graphs and demonstrate that it can be used effectively to predict the future structure communication threads. Figure 1 presents the F1 scores of our combined approach in Twitter. Combining structure and content is superior in predicting how a conversation will progress in terms of participants and messages sent.

![Graph and Content Predictions](image1)

Table 1. Prediction of communication links.

We also addresses the problem of quality of collaborative content. We develop a predictive model of editor behavior based on content dynamics throughout the revision history of a Wikipedia article. A central component of our approach is quantifying the level of agreement between contributors based on consecutive revisions. We employ these pairwise editor interactions to predict controversial edits and rank contributors based on their sensitivity to a specific new edit. We present the prediction accuracy (number of correct predictions) for the Wikipedia article on Abortion in Figure 2. The dominating curve is based on the structure entailed by editor interactions with the content. Historical interactions and their strength turn out essential for reactions to future edits by the involved parties.

A different kind of social interactions evolve within micro-blogging services such as Twitter. They combine the ability to share news and engage in focused social conversations. We recommend people to follow based on a content-aware notion of trust. Our recommendation method incorporates the interests of the query user, trustworthiness of candidate users and diversity of their combined content. We derive a topic-based notion of trust based on observed rates of message re-posting.