Kindred Spirits: Predicting Fruitful Interactions with Emotional and Social Features

Andrés Gómez Emilsson
nc07agom@stanford.edu

Moritz Sudhof
sudhof@stanford.edu

Jireh Tan
jirehtan@stanford.edu

Abstract

We study the problem of predicting meaningful interactions between users on the Experience Project, an online social network based around life experiences. Although most research on predicting friendships focuses around network structure, we believe that sensitivity to language, especially expressive, emotional language, is required to surface pairs of users that will form deep, meaningful friendships. This intuition guides our development of models to predict whether two users have interacted and to predict the strength of this interaction. We find that network features alone are predictive of interactions, but models based solely on language features approach network-baed performance and models that combine network and language sensitivity appreciably outperform all other models.

1 Introduction

The term "kindred spirit" refers to two individuals who share a deep emotional bond because they have experienced similar events and share similar beliefs. The term has a long cultural history, and equivalents exist in many other natural languages[1], demonstrating the apparently universal appreciation and desire for forming connections with other people who "get" you.

The role that natural language plays in establishing these relationships is evident: as described by Lyons in his overview of semantics (1977), language is used to describe, affect, and socialize. In order to determine whether two individuals are kindred spirits, it is not sufficient to know *who* these individuals are friends with or *what* they have experienced. Instead, these signals must be situated *emotionally* and *socially*. Natural language mediates these experiences with its descriptive, affective, and social functions.

This view of natural language is the basis for our research. Individuals use natural language to share experiences, affirm beliefs, and express sympathy; they also use language to entrench existing differences through insults, excoriation and condemnation. Language that is highly emotional often acquires a social dimension (Potts, 2007). Thus, it is possible to algorithmically discover two individuals who will be "kindred spirits" by taking into account the their experiences, the similarity in their affective states, and the social distance between them.

2 Prior Literature

Much research has focused on modeling expressive language. One approach, sentiment analysis, initially focused on crude polarity classification (assigning a "positive" or "negative" to text) in online reviews (Pang and Lee, 2002) (Turney, 2002). More recent approaches use more sophisticated learning techniques (e.g. recursive autoencoders (Socher, et al., 2011)) and richer training data to obtain high-dimensional sentiment representations of documents. Other research has focused on modeling the expressive content of individual words. For example, Monroe, et al. (2001) ex-

[1]Take, for example, "xīn yǒu líng xī" in Chinese, or any of the variations of the French "âme sœur" in the romance languages

plore methods for identifying and weighting words that encode bias.

However, sentiment analysis and other approaches for modeling expressive language have developed rather independently from social network analysis. Structural information about signed networks (where people can “friend” or “foe”) has been used to predict the sign of relationships between pairs of individuals within the network (Leskovec, et al., 2010). There have been relatively few studies that incorporate both approaches, but these few studies have been quite successful. In particular, Sashan, et al. (2012) improves on community detection algorithms by taking into account not only the structure of the network, but also the text (in particular, the similarity in topics described therein) that is sent between individuals in the network. More recently, unsupervised methods have been applied to discover latent communities in social networks, based on the messages exchanged between users (2010).

It is our goal to address this lacunae in the research by actively incorporating both high-dimensional sentiment analysis as well as social network analysis to predict whether two individuals will fruitfully interact with each other – in other words, whether they have the potential to become “kindred spirits”.

3 Experience Project data

We use data from the Experience Project[2], an anonymous online social network operated by Kanjoya, Inc. The Experience Project focuses on storytelling and experience sharing. Users ranging from teenagers to senior citizens visit the Experience Project to talk about who they are, what theyve experienced, and how they felt about it. They discuss diverse topics ranging from loving coffee to missing a sister who passed away. The language ranges from traditional, proper English to text riddled with emoticons, colloquialisms, and vernacularisms. Data on the Experience Project is well-suited for studying the phenomenon of “kindred spirits” because friendships on the Experience Project do not mirror existing social connections in the real world; rather, friendships develop when users identify with each other’s experiences.

[2]www.experienceproject.com

The Experience Project is organized by “groups”, each of which represents an “experience”. When members join a group, they indicate that they identify with that experience. Members can then also post stories to a group to share their own take on or story about the experience. The stories posted become anchors for interactions between members. Members read each others stories and can rate them, comment on them, or hit a button to thank the author for the story. Through these interactions, members can also decide to become “friends”, enabling direct member-member interactions such as the sending of “gestures” (e.g., a “hug” or a “thumbs-up”), accompanied by a personalized message. Conversely, when a member is annoyed by another member, she can block that member.

For thematic focus and a manageable dataset size, we restrict our analysis to experiences pertaining to the category Relationships”. The following are some summary statistics about the dataset:

- There were 532 experience “groups” (e.g., “I have sacriced for a relationship”, “I got out of a bad relationship”); these correspond broadly to thematic structure.
- 21,000 members authored the 43,000 stories in this category
- These members were predominantly self-identied as female (12,000), and the remaining members were equally split among males (4,000) and unknown gender (4,000)
- 89,000 pairs of members are friends (a mutual relationship) and 101,000 pairs of members have a fan relationship (asymmetrical; only one member is a fan of the other)
- 9,000 block actions were observed, with 1,600 members having blocked other members, and 3,000 members having been blocked by other members
- 55,000 gestures were sent, with 4,800 members sending gestures and 5,700 receiving
- These members commented on each others stories 95,000 times, averaging 4.5 comments per member and 2.2 comments per story.

4 Feature Extraction

Due to the richness of the Experience Project data, a broad range of informative features are available. For each user-user pair, we extract features that characterize the users and their relative positions in the community of experiences (thematic features), the users' role and connections in the network (network features), and the similarity of the language used by the users to describe their experiences (language features). Features include a mix of absolute features (features that characterize users individually, such as a gender feature or the number of groups a user has joined) and relative features (features that characterize the relationship or similarity between two users, such as the Jaccard index of their friends).

4.1 Thematic Features

Since experience groups represent the basic thematic structure of the Experience Project, the groups that a user has joined are a good summary of the user's experiences. To measure this, for each user-user pair, we extract features for the number of groups each user has joined and the Jaccard similarity of their group memberships.

4.2 Network Features

Users on the Experience Project can choose to "follow" each other, allowing them to keep track of the public activity of the person followed. Following is therefore a sign of personal interest in either the content or the user being followed.

The set of all following relationships is a directed social graph in which an arrow from a node A to a node B means "A is following B" (we say that they are "contacts" of each other when there is at least one arrow from one to the other). When, in addition, there is an arrow from B to A, we say that the two users are friends. Without loss of generality, we will always let A denote the node that has fewer followers.

Previous studies have used network structure to predict the presence of edges, the sign of edges, and other quantitative relations (Gilbert and Karahalios, 2009) (Leskovec, et al., 2010). We use a basic set of network features drawn from previous studies on signed networks[3] and two additional types of features we developed.

The basic feature set characterizes A and B by their absolute network properties (number of friends, contacts, proportion of contacts that are friends, etc.) and their relational properties (common mutual friends, common contacts, fraction of common contacts that are mutual friends, etc.). The basic set also includes the number of common friends that belong to each possible triad (of a total of four possible permutations for each user C).[4]

The two kinds of network features we added to the basic set are (1) revealed interests (user preferences we can deduce from the network) and (2) applications of the surprise metric (Aldecoa and Marn, 2011) for network clustering. For (1) we look at the age and gender distribution of the users that the pair follows. This characterizes the actual "demographic of interest" of each of the users. For (2) we determine how embedded the two users are in the network by measuring how well clustered their mutual friends are. To do so, we compare the expected number of edges between the friends to the actual number of edges and compute a score that represents how unlikely the observed number of edges is when we assume the group is sampled randomly. This can be done with a binomial approximation of the surprise metric. To compensate for the drawbacks of this score at the extremes of the distribution, we also take the proportion of actual to possible edges in this group and use that as a feature.

4.3 Language Features

The experience stories written by users are rich; they are expressive, personal, and emotional. Although leveraging structured data such as the connections between users or the connections between users and groups is certainly useful in predicting whether users will get along, we believe that a truly powerful "kindred spirit" analysis will include a sensitivity to a "spiritual similarity" that we can only measure by

[3] We focus on the features used to predict the sign and presence of directed edges because we care about the polarity and weight, not only about the presence, of an edge.

[4] The four possible triads can be described by the direction of "paths" between A and B. Algebraically, we can describe them by the intersection of the sets of users following and followed by A and B. For example, one of the triads is the intersection of followers of A and people that follow B. Leskovec, et al., (2010) show that these features have significant predictive power in signed networks.

analyzing the unstructured expressions of the users.

To measure "expressive similarity," we consider a variety of vector space models (VSMs): a unigram model, a noun-only model, and an emotional-lexicon model[5]. The motivation for choosing these varieties is the attempt to isolate different aspects of language. The noun-only model approximates a thematic or topical model of the stories; the emotional-lexicon model approximates an emotional or attitudinal model; the unigram model is a mixture of both. The benefit of using a VSM to model the emotional content of a story (rather than just computing sentiment scores) is that similarity measure computed from such a high-dimensional model are better able to leverage nuanced similarities between stories.

For each model, we start by building a term x group matrix (where a "group" is a document made of all the stories submitted to that group). We then transform the term counts using positive pointwise mutual information (PPMI) and retain only the top 2,000 terms that have a positive PMI value for at least one group. This constraint ensures that each term we include in the model is "distinctive" of at least one experience. We then build a term x user matrix using only the terms extracted previously. We then define the language similarity between two users as the cosine similarity of the user columns. For a more complicated model, we can option to also include the user-user overlapping terms as features.

5 The Models

We now build models to predict (1) whether two users have interacted and (2) how strong the interaction between interacting users is.

5.1 Predicting Interactions

To predict whether users have interacted, we consider comments on stories. When a user submits a story to a group, she joins all the other users who have submitted stories in expressing identification with the group's experience. Theoretically, each other story about the experience is an equal candidate for engagement. But everyone's take on an experience can be different, and commenting is not random. Here, we try to predict user pairs that have interacted via commenting.

We take all the pairs of users who submitted stories to the same group and divide them into two camps: the pairs where one user commented on the other's story (interactive), and the ones in which no comments were exchanged between users (non-interactive). To control for the fact that high engagement with the site is highly predictive of commenting, we exclude users who are unusually active on the site, and to avoid over-fitting to the behavior of particular users, we limit the number of times a user can be considered part of a user pair to 100. Overall, this restricted dataset includes 10,900 users. On average, users are part of 9 user-pairs and have submitted 2 entries.

We trained a series of MaxEnt models to predict commenting; results are reported in Table 1.[6] Overall, a pure network-based model outperforms the pure language-based models, but this is unsurprising given the importance of friends and friends-of-friends in the content discovery, and therefore content interaction, process. In fact, it is impressive that our language models approach the network model's performance, suggesting that connections on the Experience Project really do have a foundation in shared experiences and attitudes towards those experiences. The high precision of the language models is also exciting because it suggests a user need not be an existing, embedded member of the Experience Project network in order for us to recommend friends.

All models have similarly high precision but low recall. This tradeoff is acceptable because for the task of recommending stories users will comment on, it is more important to be right about the recommendations made rather than finding all the possible recommendations. The more complicated language models do improve recall without cost to precision, and the combination of the language and network models brings significant recall gains.

5.2 Predicting Interaction Strength

We now turn to predicting how the strength of the interactions that occur, because we are interested not only in *whether* interactions occur but also in the quality of these interactions.

[5] We used an emotional lexicon developed by Kanjoya, Inc.

[6] We report performance on interaction predictions, since we care more about whether users will interact than we do about whether they won't.

Model	P	R	F1
Network	0.97	0.53	0.68
Nouns	0.94	0.40	0.55
Emotions	0.94	0.39	0.55
Unigrams	0.95	0.45	0.61
Unigrams + Emotions	0.96	0.50	0.65
Network + Language	0.97	0.62	0.76

Table 1: Performance of MaxEnt models on predicting positive interactions between two users (defined by the exchange of comments). The test set is balanced over interacting and non-interacting user pairs (2,500 pairs per class), so random guessing yields an accuracy of 50%.

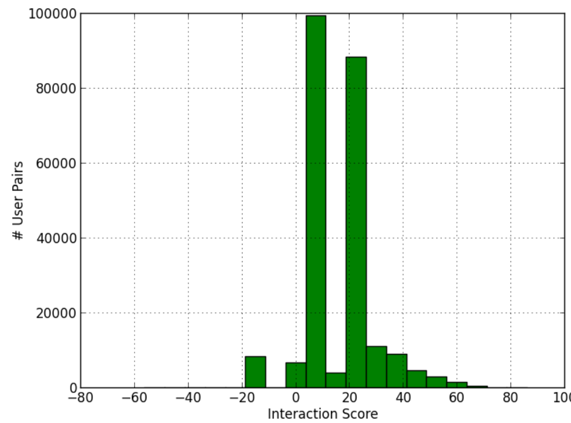

Figure 1: Distribution of interaction strength scores.

5.2.1 The Interaction Strength Score

We use the following formula to determine the interaction strength score. Let N be to the number of words exchanged between the two individuals. Let BLOCK denote if one user had blocked another, FAN denote if one user had become a fan of the user, FRIEND denote if the pair had become friends on the network, and NONE denote that the pair is neither in a block, fan or friend relationship. Then the interaction strength score is given by:

$$y = \begin{cases} -15 - 5 * log_2(N+1), & \text{if BLOCK} \\ 5 * log_2(N+1), & \text{if NONE} \\ 10 + 5 * log_2(N+1), & \text{if FAN} \\ 20 + 5 * log_2(N+1), & \text{if FRIEND} \end{cases}$$

We take the logarithm of N because, like many other types of web data (e.g. product reviews), the values are extremely skewed and follow a power-law distribution; the logarithm normalizes this. We make the simplifying assumption that if one user blocked another, then each word exchanged between them subtracts from their interaction score. These scores are summarized in Figure 1. We observe that the spikes in volume for scores of 10 and 20 represent user pair interactions that solely consist of fan or friendship relationships; in these cases, it is likely that the users liked each others' stories enough to enter into a friend or fan relationship but engaged in no further interaction.

5.2.2 MART

Based on our definition of our dependent variable, y, traditional parametric regression would be inadequate, especially in dealing with the spikes in value that arise from pairs that are simply in a "friend" or "fan" relationship. Therefore, we use a non-parametric regression technique, multiple additive regression trees (MART), to fit the model more flexibly. In this model, gradient boosting is used on regression trees to produce extremely competitive and highly robust results, appropriate for mining data that is non-normal (Jerome Friedman, 2001). In this setup, optimization is performed in function space rather than parameter space: we are interested in additive functions of the form

$$F(\mathbf{x}; \{\beta_m, \mathbf{a}_m\}1^M) = \sum_{m=1}^{M} \beta_m h(\mathbf{x}; \mathbf{a}_m)$$

where each h is a small regression tree, and therefore the parameters of interest, a_m, are the splitting variables, split locations and terminal node means of the individual trees. In MART, these regression trees are repeatedly boosted within a steepest descent framework.

We train our models with 1000 boosting iterations, holding out $1/5$ of the data for testing. To ensure that we were not overfitting on the training data, we test on the held-out data. The results are presented in Figure 2. As is expected of MART models, both the test and training errors show an initial steep learning curve as the algorithm learns from the iterations of boosting. These then taper out, as there can be no more meaningful information gained from the predictor variables. The similarity in the trajectory of training and test errors is evidence that the model did not overfit on the training data.

Our baseline error rate is 11.46 (represented by the dashed line); this baseline is obtained by naively guessing the most frequent value ($y = 20$) for every single observation. We observe that just by using thematic, language and network features in isolation, our model beats this baseline. Including all three types (thematic, language and network) of features leads to training and test errors of 5.574 and 5.747 respectively, indicating that the baseline error rate was reduced by half. Therefore, combining these features can lead to significant improvements in predictive power. Much like in the classification task, the combination of the language and network models brings significant gains in predictive power in the regression task.

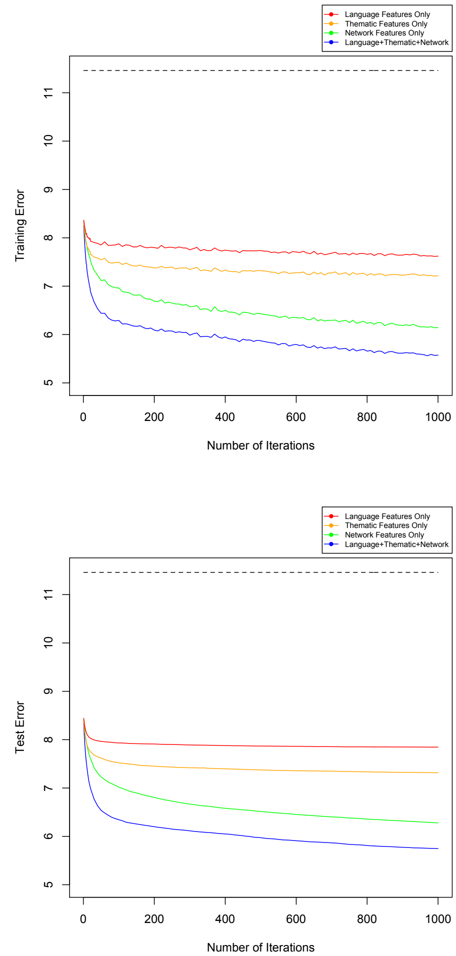

Figure 2: Training and test error at each iteration. The dotted line represents the error rate if the most frequent y-value (=20) is always predicted.

6 Feature Analysis

To assess the strength of our features, we look at how the feature values correlate with our interaction score, which is a richer measure of interaction than the binary “comments-exchanged” measure. The correlations of features and the interaction score are reported in Table 2. We also report a “Top 10% Overlap” measure, which is an indicator of the feature’s signal in relation to the top-interacting user pairs rather than all user pairs. This value is the percent of the top 10% of user-pairs by feature value that are also in the top 10% of user-pairs by interaction score. We include this measure because we care most about predicting strong positive interactions rather than medium- or low-level interactions.

As expected, the “thematic similarity” feature, which measures the similarity between two users’ “experience profile” based on the groups they joined, correlates strongest with the interaction score.

6.1 Language Features

Since we hypothesized that sensitivity to the emotions expressed is crucially important in identifying “kindred spirits,” it is heartening that emotional similarity between users is the language feature with the strongest correlation with the interaction score. The correlation is striking especially when compared to the noun similarity feature, which has a correlation almost twice as low as that of emotional similarity. We hypothesize that the strong performance of the unigram similarity feature (which is the simplest language feature considered, as it requires no part-of-speech tagging or lexicon-based tokenization) is due to the fact that the unigrams themselves encode much of the emotional and expressive content of the stories. This occurs for two reasons: (1) the Experience Project is a fundamentally emotional place, and stories have high emotional density to begin with, and (2) our PPMI-based model filters terms that are not distinctive of particular groups, and since the domain of our study is “relationships”, many common topical words (e.g., “relationship” or “boyfriend”) aren’t very distinctive and thus filtered. Finally, the fact that language features correlate more strongly with the interaction score and are more predictive of strong interactions than network features validates our assumption that language sensitivity is crucial to the task of predicting meaningful friendships. Network-only models still outperform language-only models, but we believe this imbalance is due to the range and volume of signal: there are many more network signals, and they encode a diverse range of user characteristics.

6.2 Network Features

The surveyed literature on link prediction (Leskovec, et al., 2010) shows that the degree of any two nodes is a good predictor of the presence of an edge between them. In contrast, our analysis shows that having many connections is negatively correlated with the interaction score: the logarithm of the number of users that either A or B follows has a correlation of -0.15 with the interaction score; this correlation drops to -0.2 when we look at the number of people who follow either individual in the pair. Additionally, pairs with the top number of followers are exceedingly rare at the top of the interaction scores. This strongly suggests that users who follow and are followed by large numbers have a significantly lower proportion of deep interactions. An interesting application of this finding is that, if a service wants to make friendship suggestions that result in meaningful interactions, it may backfire to simply suggest popular or power users. This interpretation of the data is corroborated by the fact that user pairs that have high values for features measuring the number of contacts are significantly less likely to have high interaction scores, suggesting they are negatively indicative of meaningful interactions. Overall, these measures, which include the triad features, do not correlate significantly with the interaction score.[7]

The features that quantify network similarity (e.g., the binomial approximation of surprise and the Jaccard index of contacts of both A and B) are all powerful indicators of the interaction score. Interestingly, surprise values close to zero (indicating no surprise) are much more frequent than expected in the sample of pairs with top interaction scores (20%), which is the opposite of what we anticipated. It turns out that having a diverse set of common mutual friends, who are not friends among themselves, raises the chances of having a deep interaction. We hypothesize that this is an indirect reflection of the fact that people who follow power users are also less likely to interact deeply regardless of their own friend count.

Finally, as expected, our findings show that measures of reciprocity such as the Jaccard index of mutual friends and user demographic similarity (such as similar age, and same gender) are well correlated with the interaction scores. The predictive power of demographic information goes beyond individual users, as the revealed interests are also relevant (e.g. both being female and following many female users predict a higher interaction score).

Feature	Correlation	Top 10% Overlap
Thematic similarity	0.38	54.2%
Noun similarity	0.17	20.9%
Unigram similarity	0.25	35.7%
Emotional similarity	0.30	35.8%
Log(num_followers_A)	-0.195	3.2%
Log(num_followers_B)	-0.205	2.3%
Log(num_followed_by_A)	-0.15	1.6%
Log(num_followed_by_B)	-0.156	1.5%
num_mutual_friends	0.06	4.5%
num_mutual_contacts	-0.02	3.1%
Triad_1	0.002	4.4%
Triad_2	0.02	2.4%
Triad_3	0.006	3.1%
Triad_4	0.02	3.6%
Jaccard_A	0.094	14.9%
Jaccard_B	0.228	17.9%
Jaccard for A, B	0.136	14.6%
Binomial Surprise	0.073	18.5%
Fraction(female followers_A)	0.064	17.3%
Fraction(female followers_B)	0.084	21.8%
Squared age difference	-0.118	4.2%

Table 2: Correlation of features with the interaction score.[9] The "Top 10% Overlap" measures the percent of the top 10% of user-pairs by feature value that are also in the top 10% of user-pairs by interaction score. This measure is an indicator of the feature's signal in relation to the top-interacting user pairs rather than all user pairs (if the feature has no relationship to the top-interacting users, this value is 10%). All values reported have p <0.001.

[7]In previous research triads have been very informative. In ours, MART rarely used it as a variable along which to split the regression trees.

7 Conclusion

We study the problem of modeling meaningful relationships between users and find that we can predict the depth of a relationship with reasonable accuracy. Our interaction score, which measures the overall depth of a relationship between users, also allows us to identify the factors that increase the likelihood that users will develop a deep connection.

Our research shows that interaction score prediction is not simply a refinement of link or sign prediction. The key network features that predict interaction stand in stark contrast to those that merely predict links or polarity (e.g., we find that local triad structure is not very informative), and our research suggests that metrics that quantify prosocial behavior (such as features that measure degrees of reciprocity) are very useful in this task.

While a network-based approach has a high predictive value, we show that adding language features enhances the model significantly. In fact, the best features for determining very high levels of interaction are language-based. This makes sense, since demographic and network features do not have the same resolution and specificity as language. The predictive power of language features alone is particularly exciting because it suggests we can recommend friends to new users whose location in the network may not yet be known or established. Submitting a single story is enough to enable us to deliver precise recommendations.

We are optimistic that our approach can yield actionable predictions in an environment like the Experience Project, where users are searching for kindred spirits. We also look forward to a further study of our methods in different domains, such as an emotionally-charged political forum.

References

John Lyons. 1977. *Semantics*, volume 1. Cambridge Univ. Press, Cambridge

Christopher Potts. 2007. The Expressive Dimension *Theoretical Linguistics*, 33(2):255-268.

J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and use of a gold standard data set for subjectivity classifications. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*, pages 246-253. University of Maryland, June.

Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computation Linguistics (ACL). (ACL-02)*, pages 417-424. Philadelphia, July.

Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79-86. Philadelphia, July.

Peter Turney. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning*, pages 491-502. Berlin: Spring-Verlang.

Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2009. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. In *Political Analysis*, 16(4):372403, October.

Noah Goodman and Andreas Stulmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. In *Topics in Cognitive Science*. (5)173184 September.

Eric Gilbert and Karrie Karahalios 2009. Predicting Tie Strength With Social Media In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211-220 , New York.

Mrinmaya Sachan, Danish Contractor, Tanveer A. Faruquie and L. Venkata Subramaniam 2012. Using Content and Interactions for Discovering Communities in Social Networks In *Proceedings of the 21st international conference on World Wide Web*, pages 331-340, May.

Jure Leskovec, Daniel Huttenlocher and Jon Kleinberg 2010. Predicting Positive and Negative Links in Online Social Networks In *Proceedings of the 19th international conference on World wide web* , pages 641-650, April.

Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, Christopher D. Manning 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* , pages 151-161. July.

Ahmed Hassan , Vahed Qazvinian and Dragomir Radev. 2010. What's with the attitude?: identifying sentences with attitude in online discussions In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* , pages 1245-1255, May.

Ahmed Hassan , Amjad Abu-Jbara and Dragomir Radev. 2012. AttitudeMiner: Mining Attitude from Online Discussions In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics* September.

Jerome Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5)1189-1232. September.

Rodrigo Aldecoa and Ignacio Marn 2011. Deciphering Network Community Structure by Surprise In *PLoS ONE 6(9): e24195. doi:10.1371/journal.pone.0024195*