Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2010 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-1-2010

Sentiment Community a New Way to Learn Users' Sentiments in Social Network- A Preliminary Study

Kaiquan Xu

Stephen Liao

Xin Li

Xin Xu

Sentiment Community: a New Way to Learn Users' Sentiments in Social Network ----A Preliminary Study

Kaiquan Xu[a], Stephen Liao[a], Xin Li[a], Xin Xu[b]
[a]Department of Information Systems, City University of Hong Kong
[b]Department of E-Commerce& IM, Southwest Jiaotong University
kaiquan.xu@student.cityu.edu.hk

Abstract

Many enterprises begin to use Social Network Sites (SNS) as an important channel and platform to do online marketing and reputation management, because users' interactions in SNS have more effective impacts on customers' buying decisions and images of enterprises than traditional websites. To do this, the enterprises need to learn and trace users' sentiments on their products/services for designing appropriate business strategies. In this study, the sentiment community is proposed as a method for this. The sentiment communities with different polarities in SNS usually represent groups of users with different preferences, and discovered sentiment communities is very useful for enterprises to do customer segmentation and target marketing. Also the evolvement of sentiment communities is explored, so that enterprises can easily trace users' sentiments and learn their diffusions in SNS. In this paper, a novel method is proposed for discovering users' sentiment communities, and an initial experimental evaluation is executed.

Introduction

In web 2.0, users are permitted to express their sentiments and opinions on products/services through many channels, such as online forums, shopping websites, blogs, and wikis. These sentiment and opinion information always has important impacts on consumers' buying decisions and images of enterprises. Mining these data can bring many benefits [1, 10, 11]: enterprises could better understand the relative strength or weakness of their products; consumers could exercise more informed purchasing decisions.

Recently, the emerging Social Network Sites (SNS) (such as facebook.com, twitter.com, epinions.com) become very popular. These SNS permit users interact with each other and build various social relationships (such as friends, trust, following relationships) for communications and sharing. In these SNS, users' sentiments and opinions always have more effective impacts on consumers' buying decisions and images of enterprises than in previous ones, because people more easily accept and believe the sentiments and opinions of persons having relationships with them. And also the sentiments and opinions in SNS can propagate more quickly and distantly through these relationships.

Many enterprises realize the importance of SNS and begin to use SNS as an important channel and platform to do online marketing and reputation management. To do this, an important task for enterprises is to learn and trace users' sentiments on their products/services, especially the distributions and evolvements of different types of sentiments under various social relationships, in order to do customer segmentation and target marketing. In this study, we propose a new way for this purpose, sentiment community. And the methods of discovering sentiment communities and analyzing their evolvements are also provided. Sentiment community represents a group of users linked through social relationships, and these users have same sentiment polarities on one product/service. The sentiment communities with different sentiment polarities in SNS usually represent groups of users with different preferences. And these sentiment communities can be as an angle for doing customer segmentation and target marketing. The evolvement of sentiment communities can reflect the propagation of users' sentiments and changing trends, which can be used by enterprises in tracing users' sentiments and learning their diffusions in SNS. Although there exists a lot of work on sentiment analysis [1, 10, 11], social community discovery [19] and social network evolving, as we know, this is the first study on sentiment community discovery and analyzing its evolvement.

The paper is organized as: Section 2 introduces the related work; the problem description is presented in Section 3; Section 4 and 5 show the proposed methods for sentiment community discovery and analyzing community evolvement; the initial experimental evaluation is introduced in Section 6; Section 7 summarizes this work and gives future work.

Related Work

Opinion Mining

Much research exists on sentiment analysis of user opinion data (Chau and Xu, 2007; Chen, 2006; Liu, 2006; Pang and Lee, 2008; Raghu and Chen, 2007), which mainly judges the polarities of user reviews. In these studies, sentiment analysis is often conducted at one of three levels: the document level, sentence level, or attribute level. Sentiment analysis at the document level classifies reviews into the types of polarities—positive, negative, or neutral—based on the overall sentiments in the reviews. A number of machine learning techniques have been adopted to classify the reviews (Pang and Lee, 2002). Abbasi and Chen et al. propose the sentiment analysis methodologies for classification of Web forum opinions in multiple languages (Abbasi et. 2008). Sentiment analysis at the sentence level mainly focuses on identifying subjective sentences and judging their polarities. Most of these studies adopted the machine learning methods (Wiebe, 1999; Yu and Hatzivassiloglou 2003). Sentiment analysis at both the document level and sentence level has been too coarse to determine precisely what users like or dislike. In order to address this problem, sentiment analysis at the attribute level is aimed at extracting opinions on products' specific attributes from reviews.

2.2 Community Discovery

There is a lot of work on community discovery [19]. Many principles of community discovery are proposed, and they can be classified into several categories: 1) Multi-criterion scores, which combines the criteria of the number of edges inside and that of the number of edges crossing into a single objective function; 2) Single-criterion scores, which employs only a single of the two criteria. Some algorithms for community discovery are proposed. Some algorithms are based on graph partitions, and some are heuristic approaches. The two well-know graph partitioning algorithms are Spectral Partitioning algorithm and flow-based algorithm.

The main difference of our study with these existing ones is: our work focuses on analyzing sentiment communities in SNS and their evolvement. While the existing work mainly discovers user communities based on the links between them. Our work tries to discover the communities, such that users in one community have same sentiments and are linked through social relationships. This is very useful for enterprises to do customer segmentations and target marketing in SNS.

Problem Description

In SNS, users can link with each others through various relationships (such as friend, trust, and following relationships etc.). Users can hold different sentiment polarities (Positive, Negative and Neutral) on one product/service, and those sentiment polarities always change with time under the influences of opinions. In this study, these information at time t is represented in the undirected graph $G^t = (V, E, S^t)$:

Node i: represents the user. The user set is marked as $V = \{1, \cdots, i, \cdots, n\}$.

Edge (i, j): represents the relationship between i and j. The edge set is marked as $E = \{(i, j) \in V * V \mid i \neq j\}$.

Sentiment polarity s_i^t: represents the sentiment polarity of user i at time t. The sentiment set at time t is marked as $S^t = \{s_i^t \mid i \in V\}$.

Our objectives include

1. Discovering sentiment communities at time t: Partition the user nodes into some clusters, such that if two nodes link with each other and have same sentiments, these two users should be in the same cluster. If they have different sentiments, they should be allocated into different clusters. Thus, one cluster contains the users that link with each other and own the same sentiment. This kind of clusters is called as sentiment communities. This can be described as:

Let $C^t = \{C_1^t, ..., C_k^t, ..., C_K^t\}$ be the discovered clusters at time t, the clusters should follow:

1). If $(i, j) \in E$, $s_i = s_j$, $i \in C_{k1}^t$, and $j \in C_{k2}^t$, then $k1 = k2$

2). If $(i, j) \in E$, $s_i \neq s_j$, $i \in C_{k1}^t$, and $j \in C_{k2}^t$, then $k1 \neq k2$

Here, the sentiment polarity of one sentiment community, $S_{C_k^t}$, is defined as the sentiment polarity that most users hold in that community.

These communities always represent the groups of customers with different preferences, and they can be used for producers to know customer segments, in order to do targeting marketing or design adaptive products for them.

2) Exploring the evolvement of sentiment communities: Compare and analyze the sentiment communities at different time points, and find the evolvement pattern of these sentiment communities. The main evolvement patterns include:

Growth: for the communities C_k^t and C_k^{t+1}, if $C_k^t \subset C_k^{t+1}$ and $|C_k^{t+1}| > |C_k^t|$, C_k^{t+1} is the growth of C_k^t.

Contraction: for the communities C_k^t and C_k^{t+1}, if $C_k^t \supset C_k^{t+1}$ and $|C_k^{t+1}| < |C_k^t|$, C_k^{t+1} is the contraction of C_k^t.

Birth: at time $t+1$, a new community C_k^{t+1} appears.

Death: at time $t+1$, the community C_k^t disappears.

Merging: for the clusters C_{k1}^t, C_{k2}^t and C_{k3}^{t+1}, if $C_{k3}^{t+1} = C_{k1}^t \cup C_{k2}^t$, C_{k3}^{t+1} is the merging of C_{k1}^t and C_{k2}^t.

Splitting: for the communities C_{k1}^t, C_{k2}^{t+1} and C_{k3}^{t+1}, if $C_{k1}^t = C_{k2}^{t+1} \cup C_{k3}^{t+1}$, C_{k2}^{t+1} and C_{k2}^{t+1} are the splitting of C_{k1}^t.

Usually, these evolvement patterns represent the changes of user sentiments on products/services, and can be used to trace users' sentiments and design right marketing strategies.

Discovering Sentiment Community

The objective of discovering sentiment communities is: partition users into clusters, such that the users within one cluster have same sentiment polarities. The community discovery bases on the similarities of the users' sentiment polarities, and the principle is to discover the clusters with maximum agreement on sentiment polarities.

1. Building sentiment similarity graph:

In $G^t = (V, E, S^t)$, for each edge, a weight is assigned according the similarity of two users' sentiment polarities. If the two users have the same sentiment polarity, the weight is set as +1, otherwise -1. Now, the updated information can be represented as $G^t = (V, E, S^t, W^t)$

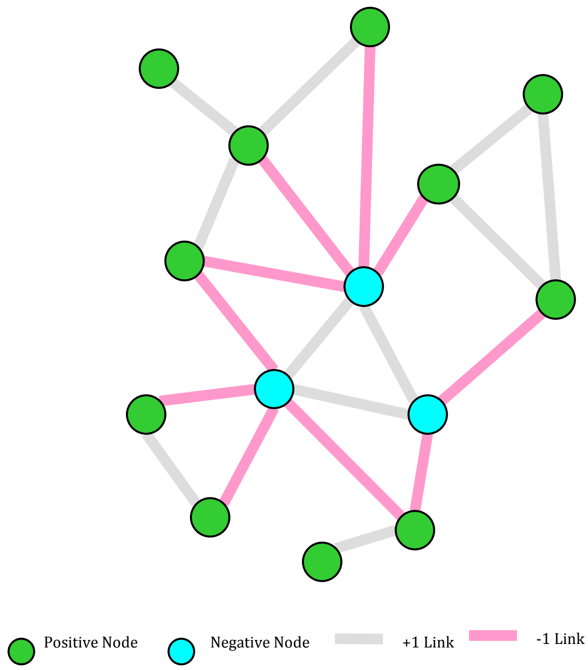


Fig 1. Sentiment similarity graph

2. Graph-Based Clustering:

Since the principle of discovering sentiment communities is: maximum agreement on sentiment polarities in one clusters. This equals to maximize the number of +1 edges inside clusters plus the number of -1 edges between clusters in the $G^t = (V, E, S^t, W^t)$.

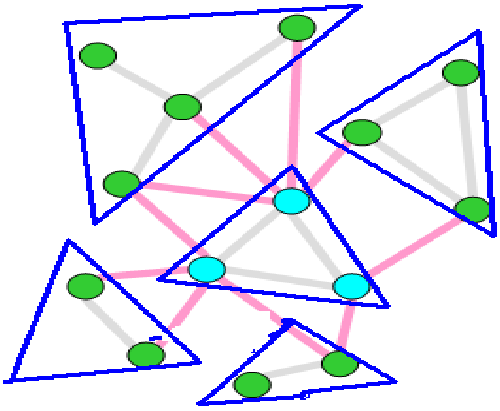


Fig 2. Sentiment communities

This is a typical correlation clustering problem [16], which is NP-hard. The efficient Semidefinite Programming(SDP) -based rounding method [17] can be used here:

Let the vector, e_i, represent a possible cluster. If $x_u = e_i$, node u is assigned to the cluster C_i. The maximum agreement on sentiment polarities can be represented as a SDP problem:

$$\max \sum_{W_{u,v}^t=+1} x_u \cdot x_v + \sum_{W_{u,v}^t=-1} (1 - x_u \cdot x_v)$$

$$\text{subject to } x_u \cdot x_u = 1 \quad \forall u \in V$$

$$x_u \cdot x_v > 0 \quad \forall u, v \in V, u \neq v$$

The random rounding method is used to put close x_u and x_v in a cluster, far x_u and x_v in separate clusters.

Analyzing the Evolvement of Sentiment Communities

After discovering sentiment communities at different time points, the evolvement of sentiment communities can be explored. Some typical evolvement patterns can be captured. These typical patterns represent special meanings for users' sentiments in SNS.

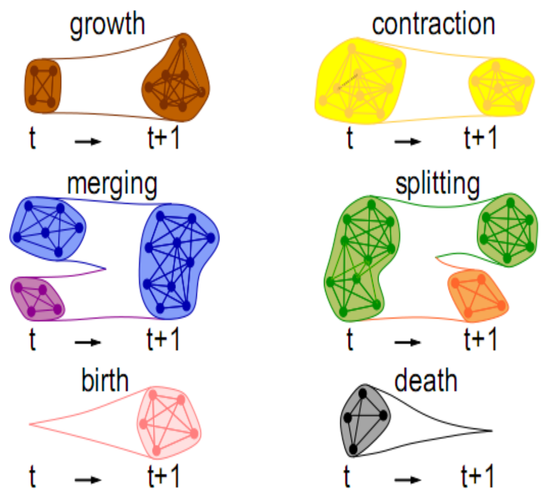


Fig 3: Evolvement patterns

Initial Experimental Evaluation

Dataset

The dataset for the experiment is obtained from a famous opinion site, Epinions (www.epinions.com). In Epinions, users are allowed to rate products, write text reviews, and build trust relationships based on his/her previous experiences. For one review, other users can rate it as helpful or not.

In this experiment, the trust relationship is extracted as social relationship between users. The product ratings given by users are taken as the sentiment polarities of users on that product. The dates of rating product are taken as the time points. Since the number of product ratings is limited, the rating information on reviews is also utilized to infer users' sentiment polarities on that product. If one user rates one review as helpful, this user is taken as having the same sentiment polarity as the user writing this review. In this initial experiment, the data on RIM BlackBerry 8100 and Apple iPhone 3GS (16 GB) smart phones are collected.

Setting

In this experiment, users' sentiments are categorized into three polarities: Positive, Negative, and Neutral. If the product rating (scale from 1 to 5) given by one user is below 3, the sentiment polarity of this user is Negative; if it is above 4, the sentiment polarity is Positive; if it is 4, or the user does not rate the product, the sentiment polarity is Neutral.

For solving the SDP problem, the SDP solver, SDPT3[18], is used. Since of the constrain of computing capability, only very small part of users in Epinions are used. In the experiment, the maximum number of clusters is set as 4.

Initial Results

The following figures show the initial analyzing results. (In these figures, a red node represents a user with negative sentiment; a green node represents one with positive sentiment, a gray node represents one with neutral sentiment).

Result on iPhone:

For iPhone, we mainly analyze the sentiment communities of users. Fig 4 shows the social relationships between users and their sentiment polarities at the end of June 2010. Figure 5 shows the two discovered communities with different sentiment polarities.

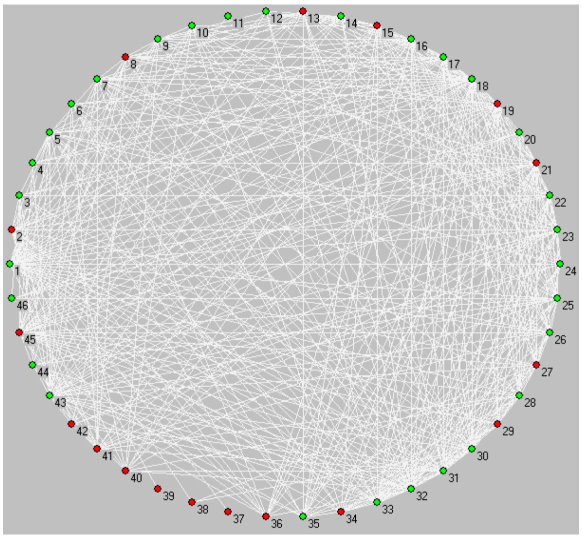


Fig 4: Sentiment similarity graph on iPhone

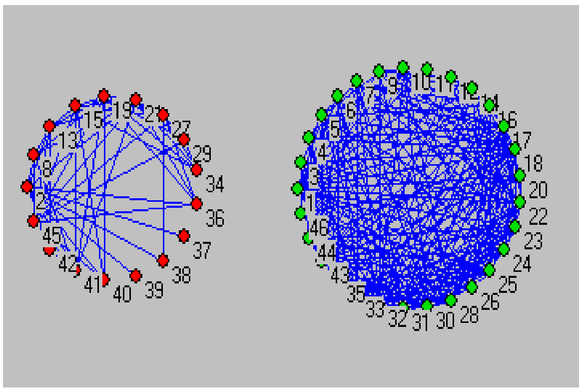


Fig 5. Sentiment communities on iPhone

Result on BlackBerry 8100:

For BB8100, we mainly analyze the evolvement of users' sentiment communities. Figure 6 shows the social relationships of users on BB8100, and Figure 7 shows the discovered sentiment communities at the end of 2006, 2007 and 2008.

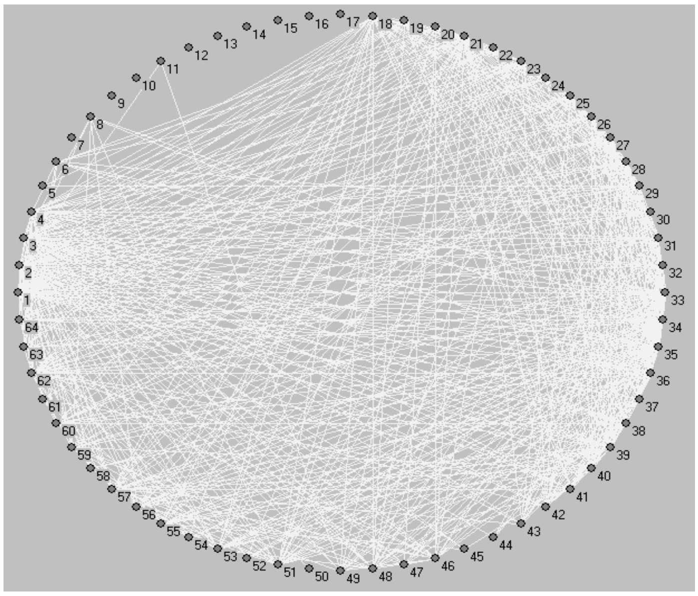
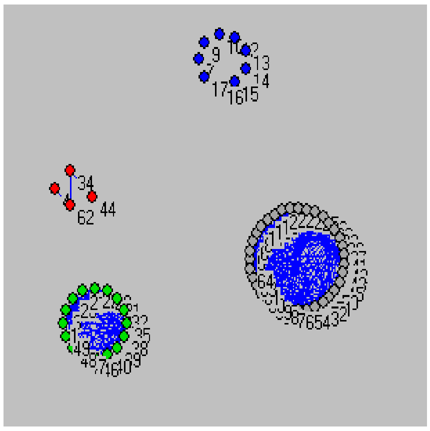
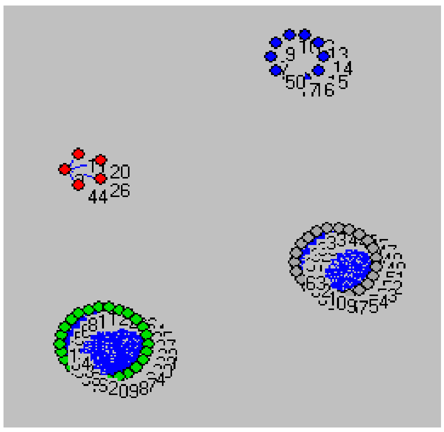


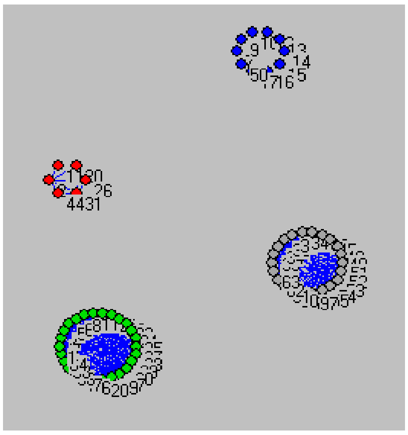
Fig 6. Social relationships on BB8100



2006



2007



2008

Fig 7. Sentiment communities at different time points

In Figure 7, by comparing the sentiment communities of 2006 and 2007, we can find that the positive sentiment community grows (Growth); the neutral sentiment community contracts (Contraction); while the negative sentiment community is stable. But for the communities of 2007 and 2008, they are relatively stable.

Discussion & Limitations:

The initial experiment evaluation indicates the proposed method can discover the sentiment communities with different polarities. And also some evolvement patterns can be found, such as Growth and Contraction. But some other evolvement patterns do not appear, such as merging and splitting. Maybe this is related with the dataset, because most of users will not change their sentiment polarities in online shopping. This is different with politics. In politics, people always change their sentiment polarities with time changing, which will lead to the merging and splitting of sentiment communities. In the future, we will explore the dataset on politics.

Also we can find that, the method mainly find three communities: positive, negative, and neutral ones. It can not find more fine- granularity communities. This phenomena maybe comes from two reasons. One is the relationships of users in this dataset are dense, and most of nodes with same sentiment polarity can link with each other with + label, so they are allocated into one cluster. In the future, other datasets can be explored. Another reason is the principle of clustering only considers the agreement of sentiment polarity, without considering the link density in cluster. Later, the principle of clustering will be revised with considering the link density in community.

Conclusion & Future Work

Identifying users' sentiment communities in social network is very useful for doing customer segmentation and target marketing. Exploring the evolvement of sentiment communities can help enterprises trace users' sentiments and learn their propagation. In this study, a novel method is proposed for discovering users' sentiment communities and exploring their evolvement, and an initial experiment evaluation shows the effectiveness of this method.

In the future, other datasets will be used for the experiment evaluation. These datasets will come from forums on politics and IT topics, and people change their sentiment polarities frequently, so some other evolvement patterns maybe will be found. New principles of discovering sentiment communities will be design for considering the link density in sentiment community. Also, the competition analysis can be done by comparing the sentiment communities of competitive products.

References

[1] Abbasi A., Chen H., Salem A., Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems 26(3) (2008).

[2] Chau M., Xu J., Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human-Computer Studies, 65(1) (2007).

[3] Chen H., Intelligence and security informatics: information systems perspective. Decision Support Systems 41(3) (2006)

[4] Guha R., Kumar R., Raghavan P., A. Tomkins. Propagation of Trust and Distrust. Proceedings of the 13th international conference on World Wide Web. 2004.

[5] Kim Y.A., Le M.T., Lauw H.W., Lim E., Liu H., and Srivastava J., Building a Web of Trust without Explicit Trust Ratings. ICDE 2008 Workshop on Data Engineering for Blogs, Social Media, and Web 2.0. 2008.

[6] Leskovec J., Huttenlocher D., Kleinberg J., Predicting Positive and Negative Links in Online Social Networks. Proceedings of the international conference on World Wide Web. 2010.

[7] Liu B., Web Data Mining- Exploring Hyperlinks, Contents and Usage Data. (Springer, 2006).

[8] Liu H., Lim E., Lauw H.W., Le M.T., Sun A., Srivastava J., Kim Y.A., Predicting Trusts among Users of Online Communities – an

Epinions Case Study. Proceedings of the 9th ACM conference on Electronic commerce. 2008.

[9] Matsuo Y., Yamamoto H., Community gravity: measuring bidirectional effects by trust and rating on online social networks. Proceedings of the 18th international conference on World wide web. 2009.

[10] Pang B., Lee L., Opinion mining and sentiment analysis. in: J. Callan, F. Sebastiani (Ed.) Foundations and Trends in Information Retrieval 2(1-2), (Now Publishers; 2008).

[11] Pang B., Lee L., Vaithyanathan S., Thumbs up? Sentiment Classification using Machine Learning Techniques. in: D. Yarowsky, D. Radev, Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA, USA (2002) 79-86.

[12] Raghu T.S., Chen H., Cyberinfrastructure for homeland security: Advances in information sharing, data mining, and collaboration systems, Decision Support Systems 43 (4) (2007)

[13] Wiebe J.M., Bruce R.F., O'Hara T.P., Development and use of a gold-standard data set for subjectivity classifications. in: R. Dale, K. Church (Ed.), Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, College Park, Maryland, USA (1999) 246 – 253

[14] Yu H., Hatzivassiloglou V., Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences, in: M. Collins, M. Steedman (Ed.), Proceedings of the 2003 conference on Empirical methods in natural language processing, Sapporo, Japan (2003) 129-136

[15] Zhang Y., Wu Z., Chen H., Sheng H., Ma J., Mining Target Marketing Groups from Users' Web of Trust on Epinions. The 2008 AAAI Spring Symposium. 2008.

[16] Correlation Clustering.

[17] Correlation Clustering: Maximizing Agreements via Semidefinite Programming

[18] 18. http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

[19] 19. Empirical Comparison of Algorithms for Network Community Detection

[20]