Position vs. Attitude: How Topological Factors Influence Our Difference in the Attitudes on Online Interrelationships? A Case Study with Language Use

Bo Wang[1,2(✉)], Yingjun Sun[1,2], Yuexian Hou[1,2], Dawei Song[1,2], and Ruifang He[1]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
{bo_wang, sunyingjun1993, yxhou, dwsong, rfhe}@tju.edu.cn
[2] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China

Abstract. Though current researches of online collaboration often study the social relationship from an objective view, individuals' subjective attitudes on their interrelationships are more important for collaboration. Inspired by sociolinguistic theories, the latest work indicates that individuals' different attitudes on interrelationships can be measured by interactive language. However, it is still an open problem that what kind of factors influences our different attitudes on interrelationships. In this work, we investigate how individuals' position i.e., the topological factors in social network influence the differences in our bidirectional attitudes on interrelationships. Measuring the attitudes with interactive language on Enron email dataset, we analyze the correlation between attitudes and the topological factors of email network. The results indicate that individuals' differences in attitudes on interrelationships are related to some typical topological factors. These results inspire us to measure individuals' attitude in online collaboration with their topological factors in social network.

Keywords: Attitude · Online collaboration · Social network topology Interrelationship · Interactive language

1 Introduction

In online collaboration, it is an essential problem to understand the nature of social interrelationships. Most current studies [1] suppose that the properties of social relationships are independent from participants' attitudes, and topological features of social network are most widely used to understand the social relationship. However, when exploring the formation of collaboration, the role of individuals' subjective attitudes become critically important. In latest studies [2] individuals' interactive language is proved to be more capable to understand individuals' attitudes. However, the

I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 153–163, 2018.
https://doi.org/10.1007/978-3-030-00916-8_15

topological features may still be a latent cause which influences individuals' language on interrelationship, and can also be beneficial to understand and measure individuals' different attitudes on interrelationship indirectly. For example, whether two people are friends or lovers depends largely on their own attitudes, while a large number of common friends can also help to change their relationship.

1.1 Language Analysis in Signed Social Network Studies

Nowadays, researchers try to combine the language and network features to understand social relationship. Through measuring the ability of the feature sets of social behavioral and textual information, Adalı et al. [3] drew the conclusion that these two kinds of information were practically equivalent between pairs of individuals' interaction. Pang and Lee [4] extended the model of text based statistical learning approach proposed by Bramsen et al. [5]. Their improvement was inspired by an assumption of homophily, i.e., certain social relationships correlate with agreement on certain topics. Tan et al. [6] predicted attitudes about social events by utilizing Twitter follows and comments. West et al. [7] developed a model combining textual and social network information to predict the person-to-person evaluations in the signed social network.

1.2 Measurement of Bidirectional Difference in Social Interrelationship

There are also some latest studies analyzed social relationships directionally or asymmetrically. For directed relationships, Leskovec et al. [8] first considered an explicit formulation of the sign prediction problem. Their prediction methods are based on the theory of social balance and status. Bach et al. [9] and Huang et al. [10] framed sign prediction as a hinge-loss Markov random field, a type of probabilistic graphical model introduced by Broecheler et al. [11]. West et al. [7] developed a model that synthesizes textual and social network information to jointly predict the polarity of person-to-person evaluations. Wang et al. [2] investigated the subjective difference of interrelationship with interactive language features.

1.3 Our Approaches

Although current studies have proposed the measure of two individuals' different attitudes with interactive language features [2, 7], they do not explore the intrinsic causes of the difference in individuals' attitudes. If individuals' attitudes only depend on their own ideas, we can only measure their attitudes with subjective factors e.g., the interactive language. However, an individual's attitude on one of his social relationships is also affected by other relationships. Therefore, we try to examine whether social network topological factors can affect individuals' attitudes, and shall we introduce topology features into the measurement of individuals' attitudes.

2 Topological and Linguistic Features in Sociolinguistic Theories

Interactive language is a good resource to describe individuals' attitudes in social interaction. The theory of communicative action [12] reconstructs the concept of relationship with the communicative action instead of the objectivistic information. Thus we can utilize the linguistic structures to understand the social relationships. Sapir-Whorf hypothesis [13] also supposed that the semantic structure shapes the way in which a speaker formed conceptions of the world including social relationships. Therefore we can try to investigate the different attitudes with the interactive language.

How can we describe an individual's interactive language style in order to describe his attitude on his interrelationship? In sociolinguistics, Holmes [14] proposed four important dimensions to study the language used in social interrelationship:

(1) The solidarity-social distance scale: concentrate on the solidarity of the relationships in social interrelationship.
(2) The social status scale: concentrate on the relative status of the individuals in social relationship.
(3) The formality scale: concentrate on the formality of language that individuals use in different places, topics and relationships.
(4) The referential and affective scale: concentrate on referential and affective function of the language that individuals use in social interrelationship.

Among these four dimensions, the first two concern about the topological features of social relationship, and the last two aim at the features of interactive languages.

Inspired by Holmes' theory, firstly, we use four interactive language features to indicate individuals' attitudes, including frequency, length, fluency and sentiment polarity which indicate quantity, formality and affective scale of the interactive language. Second, we then investigate the correlation between topological and linguistic factors. The calculation of linguistic features will be introduced in Sect. 4.2.

3 Topological Factors on Social Network

In this section, we introduce three most widely studied topological factors on social network which can potentially affect individuals' different attitudes on interrelationship.

3.1 Degree and Clustering Coefficient of Individuals

As for a vertex A, the degree of A which can indicate an individual's range of social relationships is defined as the number of vertexes connected with A. The clustering coefficient indicates the probability of any two friends of A are also friends in social network. It is also known as an indicator to measure A's ability to gather friends into a cluster. If the degree of A is n and the number of edges between these n vertexes is k, the clustering coefficient $C(A)$ of A can be calculated with Formula (1):

$$C(A) = \frac{k}{C_n^2} \tag{1}$$

Why degree and clustering coefficient are selected for this work? In intuition, the range of one's social relationships and the clustering degree of one's friends may affect individuals' different attitudes on social interrelationships.

3.2 Embeddedness of Interrelationships

In social network, the embeddedness of an interrelationship indicates the number of common friends of two individuals engaged in an interrelationship. Embeddedness is believed to indicate the strength of social relationship. We suppose that embeddedness is also related to the degree of attitudes' difference. This assumption is based on the similar idea of the strength measurement of social relationship: it is widely believed that more common friends make two individuals connected to each other more tightly. In this work, we will investigate whether more common friends can also make two persons' attitudes on their interrelationship more consistent.

3.3 The Balance of Triadic Closure

The Traditional Balance Theory. A triadic closure consists of any three persons and the interrelationships between them. In traditional balance theory, each relationship is singed with binary tag '+' or '−', which indicates positive or negative relationship. With binary signs, there are four different signs combinations. As shown in Fig. 1, the balance theory claims that two of them are balanced while the other two are unbalanced. The balanced triadic closures are stable while the unbalanced ones tend to become balanced.

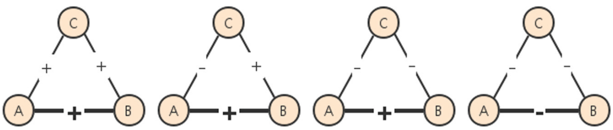


Fig. 1. All schematic diagrams of the triadic closure in the balance theory. The first and third are balanced triangles, while the other two are unbalanced.

The Extension of the Balance Theory. The traditional balance theory can be explained with the concept of homogeneity, i.e., in a triadic closure, the more consistent of the two individuals' views on the third one, the more positive their attitudes on their interrelationship are, or vice versa. The homogeneity-based explanation can help to extend the traditional balance theory from binary signs to continuous value, and from undirected relationship to directed relationship, because it describes the balance with individual's attitude instead of objective signs. This extension makes it possible to investigate the correlation between triangle balance and attitudes' difference.

In latest work [15], the traditional balance theory is extended to a directed version based on homogeneity explanation. In this extension, a person *A*'s attitude on person *B* is presented as a continuous value $|AB|$ on a directed edge (*A, B*), and vice versa. Since our aim is to investigate the bidirectional attitudes' difference between *A* and *B*, we build an extended triadic closure with four directed edges include (*A, B*), (*B, A*), (*A, C*) and (*B, C*) as shown in Fig. 2. Instead of the binary sign, we label each directed edge with a real-value in [0, 1] to indicate the attitude of the participant on the start point, e.g., $|AB|$ indicates the value of *A*'s attitude on his relationship with *B*.

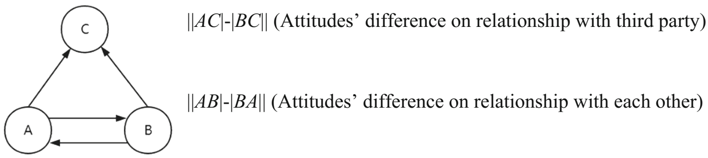


Fig. 2. Extension of the balance theory: determine the balance by comparing the difference between *A* and *B*'s attitudes on each other and the difference between *A* and *B*'s attitudes on *C*.

In this extended triadic closure, we measure the balance of the triadic closure by comparing two differences: firstly, when the difference between $|AC|$ and $|BC|$, i.e., $||AC| - |BC||$, is smaller than a threshold, we recognize the 'third party difference' between *A* and *B*'s attitudes on *C* as '−', which means *A* and *B* have similar attitudes on their interrelationships with *C*. Otherwise we recognize the 'third party difference' as '+', which means *A* and *B* have different attitudes on their interrelationships with *C*.

Secondly, when the difference between $|AB|$ and $|BA|$, i.e., $||AB| - |BA||$, is smaller than the same threshold, we recognize 'bidirectional difference' as '−' which means *A* and *B* have similar attitudes on their interrelationship. Otherwise, we recognize 'bidirectional difference' as '+' which means *A* and *B* have different attitudes. The threshold here is determined by the average of all bidirectional differences on every interrelationship. If the 'third party difference' and 'bidirectional difference' have same signs, we identify the directed triadic closure as balanced, otherwise, it is unbalanced.

4 Experiments

In the experiments, we investigate the correlation between the three topological factors and individuals' bidirectional difference in attitudes. The individuals' attitudes on relationships are characterized with four language features inspired by Holmes' theory.

4.1 Dataset

We utilized the Enron email dataset which contains 0.5M emails exchanged between 151 Enron employees. We choose this dataset because it contains both the interactive language pieces (email content) and social network topology (send and receive relationships). To make the investigation more reliable, we only selected the interrelationships where at least 15 emails were sent in each direction. The filtered dataset contains 1078 interrelationships between 647 individuals.

4.2 Attitudes (Language Features) Calculation

For each ordered pair of individuals I_i and I_j, I_i's attitude on his relationship with I_j is calculated by the value of four language features using emails sent from I_i to I_j:

(1) To calculate the feature "Frequency", we assume that the number of emails sent from I_i to I_j is N and the sending date of the first and the last email are t_1 and t_2, respectively. Then the feature "Frequency" can be calculated by Formula (2):

$$frequency\_score_{i,j} = \frac{N}{t_2 - t_1} \tag{2}$$

(2) To calculate the feature "Length", we assume that the number of emails sent from I_i to I_j is N and the total number of words in these emails is w, then the feature "Length" can be calculated by Formula (3):

$$length\_score_{i,j} = \frac{w}{N} \tag{3}$$

(3) To calculate the feature "Quality", we utilize the SRI language modeling toolkit (SRILM)[1] with Formula (4) to measure the perplexity score which has a negative correlation with the quality of a sentence. In this formula, *prob* is the generating probability of a sentence. '*words*' and '*oovs*' are the count of the words and out of vocabulary words in the sentence, respectively.

$$perplexity\_score_{i,j} = 10^{(-\log prob/(words - oovs + 1))} \tag{4}$$

(4) To calculate the feature "Sentiment", we utilize a sentiment dictionary[2] to count the sentiment words. Each positive or negative word is valued 1 or −1, respectively. Assume there are S sentences in the emails sent from I_i to I_j, and the sum of all scores of sentiment words is W, then the feature "Sentiment" is calculated by Formula (5):

$$Sentiment\_score_{i,j} = \frac{W}{S} \tag{5}$$

[1] http://www.speech.sri.com/projects/srilm/

[2] http://www.keenage.com/download/sentiment.rar.

4.3 Attitudes' Bidirectional Difference vs. Degree and Clustering Coefficient

Degree vs. Attitudes' Difference. In the first experiment, for each individual I, we calculate $\bar{D}_{I,k}$, which is the average of the attitudes' bidirectional differences on I's all interrelationships. In formula (6), $f_k(I, I_i)$ is I's attitude to his friend I_i of language feature k. C is the set of I's all friends. Then $\bar{D}_{I,k}$ is calculated with formula (6).

$$\bar{D}_{I,k} = \frac{1}{|C|} \sum\nolimits_{I_i \in C} |f(I, I_i) - f(I_i, I)| \tag{6}$$

Then, for each language feature k, we calculated the Pearson Correlation Coefficient between individuals' degree and the attitudes' average difference on their social relationships, i.e., $\bar{D}_{I,k}$. The results are shown in the first column in Table 1.

Table 1. Pearson Correlation Coefficient between topological factors of individuals and the average bidirectional difference in the attitudes on their social relationships.

Attitudes measured by language features	Pearson Correlation Coefficient	
	Degree	Clustering coefficient
Frequency	0.085	−0.044
Length	−0.650	−0.378
Quality	−0.109	0.189
Sentiment	−0.438	0.200

Clustering Coefficient vs. Attitudes' Difference. In the second experiment, for each individual I, we calculated the clustering coefficient with Formula (1), as well as the average of attitudes' bidirectional difference $\bar{D}_{I,k}$ of I. Then, we calculated the Pearson Correlation Coefficient between individuals' clustering coefficient and the attitudes' average difference, i.e. $\bar{D}_{I,k}$. The results are shown in the second column in Table 1.

Observations. According to results in Table 1, attitudes' difference measured by 'Length' and 'Sentiment' have relatively more significant negative correlation with the degree, and the other two features have no significant correlation. This indicates that a person with more friends may be more active to cater to the partners' attitudes, i.e., have less difference in attitudes, especially reflected on the length and sentiment score.

Though 'Length' and 'Sentiment' have relatively higher correlation with clustering coefficient, the absolute values are too small, which indicates that there is no significant correlation between the clustering coefficient and the attitudes' difference in this case.

Furthermore, the correlation of frequency is the lowest. One possible explanation is that the send/replay relationships between the emails can always lead to similar bidirectional email sending frequency between two individuals, which is independent with the attitudes or social topological factors of the individuals.

In general, in this set of experiments, we find that the number of friends of the individuals may have influence on their different attitudes on their interrelationships.

4.4 Attitudes' Bidirectional Difference vs. Embeddedness

In the third experiment, we investigate the correlation between their embeddedness and the bidirectional difference in the attitudes on them, which is also measured by four language features, respectively. The results are shown in Table 2.

Table 2. Pearson Correlation Coefficient between embeddedness and the average bidirectional difference in the attitudes on social relationships

Language features	Pearson Correlation Coefficient
Frequency	−0.823
Length	−0.979
Quality	−0.678
Sentiment	−0.696

In Table 2, the embeddedness of interrelationship has significant negative correlation with the attitudes' difference. This indicates that if two individuals have more common friends, they will have more similar attitudes on their interrelationship.

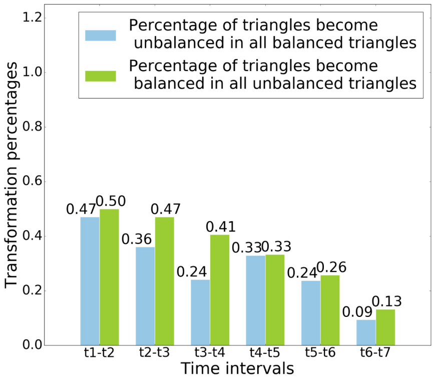
The traditional triadic closure theory states that the strength of a social interrelationship has a positive correlation with the number of common friends. Our results extend the theory to the positive correlation among the number of two individuals' common friends, the strength of their interrelationship and the consistency of their attitudes on their interrelationship.

4.5 Attitudes' Bidirectional Difference vs. Balance Theory

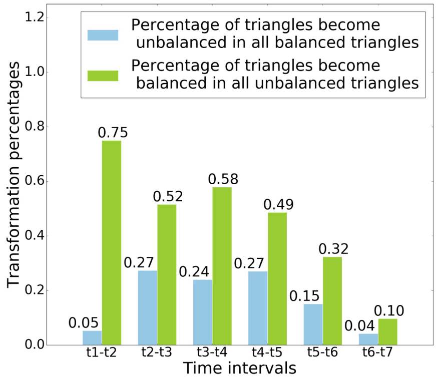
In this experiment, based on extended balance theory and bidirectional attitudes measuring, we investigate whether unbalanced triangles tend to become balanced in social network evolution.

In this experiment, we divided emails into seven time intervals. From Jan, 1999 to Jun, 2002, every six months is divided as an interval. We firstly calculated the number of balanced and unbalanced triangles on each time interval, according to traditional and extended balance theory. Secondly, for each two adjacent time intervals pair, we calculated two transformation percentages: (1) The percentage of the balanced triangles in previous interval which become unbalanced in next interval; (2) The percentage of the unbalanced triangles in previous interval which become balanced in next interval. The two percentages on all six adjacent time intervals pairs are shown in Fig. 3. In Fig. 4, we calculated the average transformation percentages between balanced and unbalanced triangles on all six adjacent time intervals pairs.

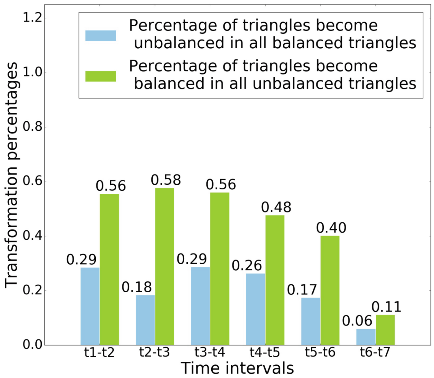
As we can see from Fig. 3, the percentage of unbalanced triangles changed to balanced triangles (green bars) is larger than balanced triangles changed to unbalanced (blue bars) in all cases in. In Fig. 4, we also find that in average the triangles tend to become balanced on both traditional and extended balance theory.



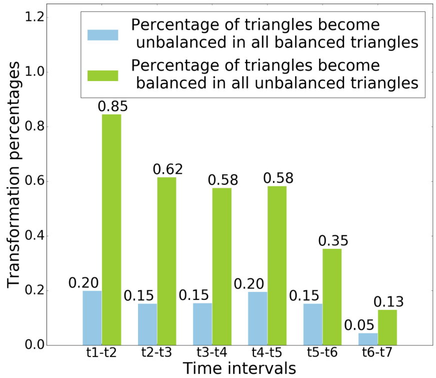
(a) Frequency



(b) Length

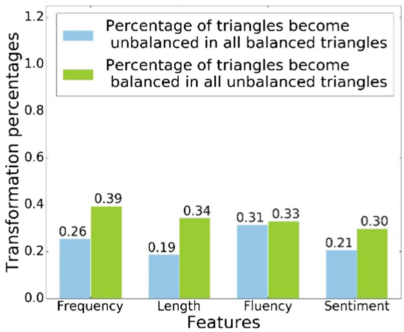


(c) Quality

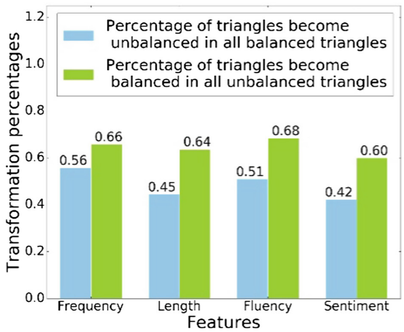


(d) Sentiment

Fig. 3. Transformation percentages between extended balanced and unbalanced triangles measured by four language features on adjacent time intervals pairs.



(a) Traditional balance theory



(b) Extended balance theory

Fig. 4. Average transformation percentages between balanced and unbalanced triangles measured by four language features on all six adjacent time intervals pairs.

These results illustrate that in social network evolution, social relationship triangles tend to become balanced in general, especially measured by the bidirectional different attitudes, i.e., measured by our extended balance theory.

5 Conclusions

We investigate whether the topology of social network are latent factors of individuals' difference in attitudes on their interrelationships. We characterize individuals' bidirectional attitudes with interactive language features. As a case study, on Enron email dataset, we analyzed the correlations between three most popular topological factors of users' relationship and their attitudes' difference. The investigated topological factors include the degree and clustering coefficient of individuals, the embeddedness of the interrelationship and the balance of the interrelationship triangle. Especially, to analyze the interrelationship triangle, we extend the traditional balance theory to redefine the balance by measuring bidirectional attitudes on interrelationship.

The experimental results reveal evidences that topological factors can influence the difference in individuals' bidirectional attitudes on interrelationship. First, a person with more friends tends to have more similar attitudes with his partners. Second, two individuals sharing more common friends tend to have more similar attitudes on their interrelationship. Third, the two individuals having more similar attitudes on their common friends also tend to become having more similar attitudes on each other, which can be an extension of the traditional social balance theory.

These results reveal that individuals' attitudes on their interrelationship are not only related to their own idea but also related to their topological context in social network. This study contributes to the social network research by showing the potential interaction between topology and language, and encourages the study of the online collaboration on social relationship synthesizing the objective and subjective features.

Acknowledgments. This work is supported by National High-tech R&D Program of China (2015AA015403), National Natural Science Foundation of China (Key Program, U1636203), Major Project of National Social Science Fund of China (14ZDB153) and Tianjin Natural Science Foundation (14JCQNJC00400).

References

1. Aggarwal, C.C.: Social Network Data Analytics. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-8462-3
2. Wang, B., Yu, Y.S., Zhang, P.: Investigation of the subjective asymmetry of social interrelationship with interactive language. In: Proceedings of the World Wide Web (2016)
3. Adalı, S., Sisenda, F., Magdon-Ismail, M.: Actions speak as loud as words: predicting relationships from social behavior data. In: Proceedings of World Wide Web (2012)
4. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of Association for Computational Linguistics (2004)

5. Bramsen, P., Escobar-Molano, M., Patel, A., Alonso, R.: Extracting social power relationships from natural language. In: Proceedings of Association for Computational Linguistics (2011)
6. Tan, C.H., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P.: User-level sentiment analysis incorporating social networks. In: Proceedings of ACM Special Interest Group on Knowledge Discovery and Data Mining (2011)
7. West, R., Paskov, H.S., Leskovec, J., Potts, C.: Exploiting social network structure for person-to-person sentiment analysis. Trans. Assoc. Comput. Linguist. (2014)
8. Leskovec, J., Huttenlocher, D., Kleinberg, J.: Predicting positive and negative links in online social networks. In: Proceedings of World Wide Web (2010)
9. Bach, S., Huang, B., London, B., Getoor, L.: Hinge-loss Markov random fields: convex inference for structured prediction. In: Proceedings of Uncertainty in Artificial Intelligence (2013)
10. Huang, B., Kimmig, A., Getoor, L., Golbeck, J.: A flexible framework for probabilistic models of social trust. In: Proceedings of International Social Computing, Behavioral-Cultural Modeling and Prediction Conference (2013)
11. Broecheler, M., Mihalkova, L., Getoor, L.: Probabilistic similarity logic. In: Proceedings of Uncertainty in Artificial Intelligence (2010)
12. Habermas, J.: The Theory of Communicative Action. Beacon Press, Boston (1981)
13. Sapir, E.: The status of linguistics as a science. Linguistic Society of America (1929)
14. Holmes, J.: An Introduction to Sociolinguistics. Routledge, London (2013)
15. Wang, B., Sun, Y.J., Han, B., Hou, Y.X., Song, D.W.: Extending the balance theory by measuring bidirectional opinions with interactive language. In: Proceedings of the World Wide Web (2017)