The Online Debate Networks Analysis: A Case Study of Debates at Tianya Forum

Can Wang and Xijin Tang(✉)

Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
wangcan@amss.ac.cn, xjtang@iss.ac.cn

Abstract. In this study, we examine the characteristics of online debate networks. We empirically investigate the debate networks formed by three hot threads from Tianya Forum at individual, whole-network and triad levels. At the individual level, the statistical analysis reveals that people participate different threads about one issue; authors reply to themselves; the authors of the original posts are the core of the interaction; we rank the indegree value and betweenness value of the authors, and find that they are not consistent in sequence. At the whole-network level, the structural indices reveal that the stances of the original posts affect the debate networks. At the triad level, the proportions of coded triads reveal that the common forms in debate networks are mutual dyads; the proportions of triadic closures reveal that relations between participants are different in the two camps; and the balanced triads between camps are more than those within camps.

Keywords: Debate networks · Online social network · Triads · Tianya Forum

1 Introduction

We are in an era where people can easily voice and exchange their opinions on the Internet through social media such as online forums. It is widely recognized that mining public opinion from on-line discussions is an important task, especially in the modern democratic life. Polarization phenomena often happen within the public discussions because that people have different cognitions towards one thing. There exist two streams of literature in this domain. One is automatically determining the debate participants' opinions by text mining [1–3]. The other is detecting the behavior patterns during the discussions [4]. Social network analysis can reveal the interactions patterns of the discussion participants.

The debate networks are different from other online social networks because the dyadic ties between social actors are special. An author may support or disapprove of another one. Some existing literatures focused on the debate networks [4, 5]. Most of the existing researches investigated the debate networks at the whole-network level [5]. In this paper, we analyze a classical debate topic in China at the individual, whole-network and triad levels.

J. Chen et al. (Eds.): KSS 2016, CCIS 660, pp. 140–150, 2016.
DOI: 10.1007/978-981-10-2857-1_12

2 Reply Networks and Data Sets

In this section, we define the construction of reply networks at forums and introduce the research data.

2.1 Introduction to Reply Networks

Online forums contain rich threaded discussions (threads) on all kinds of topics/issues, e.g., technology, sports, religion, and politics. In forums, a discussion starts when a user posts an initial post initiating a conversation in a particular matter. Afterwards, other users reply to the initial post or to another reply. The initial post and these replies form a thread.

In this paper, the relationships of the users' reply in a thread give rise to an reply network. The vertices (nodes) are authors and the links represent an author comments on another author's previous messages. The network is directed and there are at most two links between author i and author j (i→j, j→i). In the real world, two users may interact for many times but we do not count the weights of the links in this paper.

2.2 Debates on TCM at Tianya Forum

In China, there exist two camps of people according to their attitudes towards traditional Chinese medicine (TCM). Some people take the "abolishing TCM" stance that TCM should be abolished from the national health system. The other camp of people take the "preserving TCM" stance and insist that TCM should be preserved. In our previous research, we noticed that the discussions about TCM were always polarized and we did text analysis to mine the diverse opinions about TCM [3]. In this paper, we choose the classical controversial issue to mine people's interaction patterns among debates.

Tianya Forum is one of the most popular Chinese BBS sites. Table 1 lists three hottest threads about TCM at Tianya Forum. In this paper, we analyze online reply networks formed by the three threads. Reply network for Thread 1 (Network 1) has 4890 authors. Reply network for Thread 2 (Network 2) has 5514 authors. Reply network for Thread 3 (Network 3) has 6065 authors.

Table 1. Hot threads about TCM at Tianya Forum [6]

Thread ID	Replies	Participants	Start time	End time
1	117318	4890	2012-10-16	2013-11-29
2	36592	5514	2011-03-21	2015-01-24
3	33547	6065	2011-11-12	2015-01-24

3 Descriptive Statistics

In this section, we examine the characteristics of online reply networks at the individual level. Descriptive statistics of the participants' activities are listed.

3.1 Self Reply

Within the three threaded discussions, people can reply to posts to state their opinions. Interestingly we find that people sometimes reply to their own posts to strengthen their opinions or to make sure that their opinions are noticed. Table 2 lists the number of authors who reply to themselves. The statistics of the replies which are comments of some ones' own posts are also listed in Table 2.

Table 2. Authors reply to themselves and the corresponding replies

Tread ID	Authors		Replies	
	Size	Percentage	Size	Percentage
1	288	5.88 %	20760	17.70 %
2	191	3.46 %	1602	4.38 %
3	164	2.70 %	1428	4.26 %

3.2 Participate in Different Threads

Thread 1, Thread 2 and Thread 3 are all about the issue of TCM. By comparing the participants we find that some authors reply to different threads about one issue. As shown in Fig. 1, 364 authors participate both in Thread 1 and Thread 3, 174 authors participate both in Thread 1 and Thread 2, 261 authors participate both in Thread 2 and Thread 3, and 92 authors participate in the three threads.

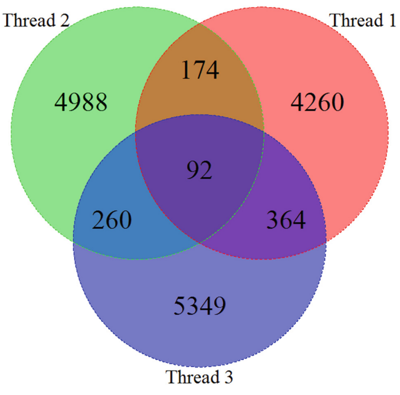

Fig. 1. The venn diagram of the authors in the three threads

3.3 Consistent Relationships Across Networks

In the three networks, we find same reply relationships among authors which means that some of the linking relationships among these authors are consistent across networks. For example, there are 14 pairs of direct links exist in Network 1 and Network 2, 80 pairs of direct links exist in Network 1 and Network 3, 23 pairs of direct links exist in

Table 3. Examples of consistent relationships

Source (author ID)	Target (author ID)	Thread 1	Thread 2	Thread 3
50425395	15956478	17	2	48
11440453	50425395	8	45	21
50425395	13010103	2	1	5

Network 2 and Network 3. Table 3 lists some examples of consistent relationships and the appearing times of these links in the threads.

3.4 Key Players

Key players in networks are determined by their topology attributes and structural attributes. Indicators of centrality identify the most important vertices within a graph. In this paper, we use indegree centrality and betweenness centrality to identify the key players. Table 4 lists the top 10 authors by indegree value. Authors of the original posts ("61681904", "3865013" and "60219641") gain the highest indegree. In other words, the authors of the original posts are the core of the participants.

Table 4. Top 10 authors in decreasing order of their indegree values

Thread 1		Thread 2		Thread 3	
Author ID	Indegree	Author ID	Indegree	Author ID	Indegree
61681904	854	3865013	779	60219641	600
60233507	198	34658621	300	50425395	217
68378554	125	50425395	275	60480958	157
73117788	108	26546902	168	42004025	90
61908805	99	35721022	164	61908805	83
56462278	92	**37971276**	**159**	60244454	72
50425395	91	**5695593**	**154**	15956478	68
72458058	88	53167987	141	2558039	52
47548691	**84**	**14734994**	**138**	**13249554**	**46**
64925795	**64**	**34888324**	**120**	**48969130**	**43**

Betweenness centrality is an indicator of a node's centrality in a network. The betweenness value of a node equals to the number of shortest paths from all vertices to all others that pass through that node[1]. Table 5 lists the top 10 authors by their betweenness value. The bold characters in Table 4 are these authors who do not appear in Table 5. The bold characters in Table 5 are these authors who do not appear in Table 4. We can infer that the indegree value and betweenness value of the authors are not consistent in sequence.

[1] https://en.wikipedia.org/wiki/Betweenness_centrality.

Table 5. Top 10 authors in decreasing order of their betweenness values

Thread 1		Thread 2		Thread 3	
Author ID	Betweenness	Author ID	Betweenness	Author ID	Betweenness
61681904	2074041	3865013	713402	50425395	2211015
60233507	317674	26546902	201356	**50191233**	**1965446**
61908805	166961	50425395	146851	60219641	1896792
50425395	147006	34658621	143833	60480958	1329253
73117788	133075	34306001	86521	42004025	265711
72458058	113751	39086442	60060	61908805	259360
74186058	**106932**	53167987	55408	15956478	227940
39086442	**92645**	**13561653**	**53886**	60244454	201634
68378554	88941	35721022	49680	**28721827**	**155379**
56462278	88827	**5608751**	**48757**	2558039	132718

4 Structural Characteristics of the Reply Networks

Structural indices such as connected components, average degrees, clustering coefficients, etc. are given in this part to reveal the structural characteristics of the reply networks at the whole-network level.

4.1 Connected Components

Table 6 lists the sizes of the authors of Thread 1, Thread 2 and Thread 3. In Table 6, authors interact with original post authors are whose who only reply to the original posts. Authors interact with others refer to authors having links with others because that they reply to others' replies. C is the number of connected components in Table 6, GC is the sub-graph with most vertices, and the number of nodes in GC is |GC|.

Table 6. Connected components and GC

Thread ID	Size of authors	Authors interact with original post authors	Authors interact with others	C	|GC|
1	4890	2132	2758	15	2737
2	5514	3348	2166	39	2100
3	6065	1858	4207	9	4193

2758 authors in Network 1 interact with other authors and they make up 56.40 % of the participants of Thread 1. In Network 2, the authors who interact with other authors make up 39.28 % of the participants. In Network 3, the authors who interact with other authors make up 69.37 % of the participants.

4.2 Average Degree

Degree of a vertex of a graph is the number of edges incident to the vertex[2]. Average degree in a network with N nodes is $\bar{d}_i = \frac{\sum_{i=1}^{N} d_i}{N}$ where d_i represents the degree of a vertex v_i Table 7 lists the average degree of the three reply networks.

Table 7. Average degrees of the reply networks

Thread ID	Connected nodes		The whole network	
	Average degree	Size of authors	Average degree	Size of authors
1	2.818	2758	1.586	4890
2	2.445	2166	0.959	5514
3	1.958	4207	1.358	6065

4.3 Clustering Coefficient

Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together[3]. The local clustering coefficient C_i for a vertex iis given by the proportion of links between the vertices within its neighborhoods divided by the number of links that could possibly exist between them [7]. The local clustering coefficient for directed graphs is given as $C_i = \frac{|\{e_{jk}: v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$, the edge e_{jk} connects vertex v_j and v_k. k_i is the number of neighbors of a vertex v_i. If $k_i = 0$ or $k_i = 1$, $C_i = 0$. The overall level of clustering in a network is measured as the average of the local clustering coefficients of all the vertices.

Table 8 lists the clustering coefficients of the three reply networks. The clustering coefficient of Network 2 is lower than others, which means the authors interact less in Network 2.

Table 8. Clustering coefficients of the reply networks

Thread ID	Authors interact with others		Reply networks	
	Clustering coefficient	Size of authors	Clustering coefficient	Size of authors
1	0.323	2758	0.182	4890
2	0.145	2166	0.057	5514
3	0.254	4207	0.176	6065

From the viewpoints of percentage of connected components, average indegree and clustering coefficient, the participants of Network 2 interact less frequently than participants of the Network 1 and Network 3. This is because that the original posts'

[2] https://en.wikipedia.org/wiki/Degree_(graph_theory).

[3] https://en.wikipedia.org/wiki/Clustering_coefficient.

stances about TCM are different. The original posts of Thread 1 and Thread 3 hold the same stance of "abolishing TCM". The expressions of the two original posts provoke heated arguments about TCM. The original post of Thread 2 holds the "preserving TCM" stance. The author of the original post claims that she is a daughter of a TCM practitioner, so some participants join in to ask for advices about therapies. It is inferred that these authors in Network 2 do not interact with others as frequently as authors in Network 1 and Network 3 do.

5 Triads Analysis

This section examines the characteristics of online reply networks at the triad level.

5.1 The Coded Triads

Davis and Leinhardt proposed that social relations could be tested on directed rather than undirected triad census, and forwarded 16 different types of directed triads as shown in Fig. 2 [8]. Their classification scheme describes each triad by a string of four elements: the number of mutual dyads within the triad; the number of asymmetric dyads within the triad; the number of null dyads within the triad; a configuration coding (U for up, D for down, C for cyclical and T for transitive) the triads which are not uniquely distinguished by the first three distinctions. For example, 120D refers the triad includes 1 mutual dyad, 2 asymmetric dyads, 0 null dyad and the down orientation.

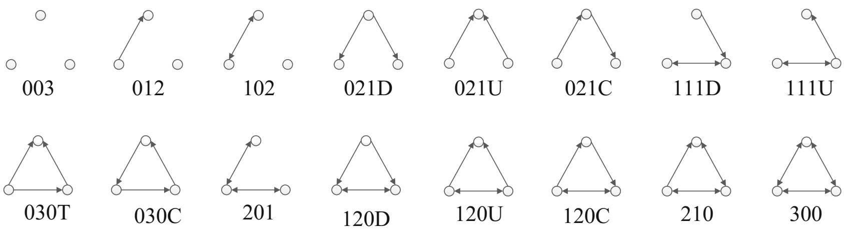

Fig. 2. 16 different types of directed triads [9]

We find the most active participants who post more than ten times in the three threaded discussions. Then we get three sub-graphs. Sub-graph 1 has 254 nodes and 1473 edges. Sub-graph 2 has 210 nodes and 1054 edges. Sub-graph 3 has 208 nodes and 854 edges. We use the "SNA" package[4] in R to calculate the types of the coded triads of these sub-graphs.

We add the proportions of 16 different types of directed triads respectively in the three sub-graphs. Table 10 lists the coded triads descending by their proportions.

[4] http://cran.r-project.org/web/packages/SNA/.

Table 9. Coded triads and their distributions

Coded triads	Sub-graph 1		Sub-graph 2		Sub-graph 3	
	Triads	‱	Triads	‱	Triads	‱
003	2323374	8608.26	1279216	8407.49	1295411	8763.10
012	158591	587.59	120101	789.35	91527	619.16
021C	5231	19.38	3219	21.16	2260	15.29
021D	4629	17.15	4810	31.61	2835	19.18
021U	4951	18.34	2281	14.99	3043	20.59
030C	27	0.10	24	0.16	23	0.16
030T	365	1.35	512	3.37	242	1.64
102	134097	496.84	83743	550.39	61637	416.96
111D	13079	48.46	7243	47.60	3407	23.05
111U	20503	75.97	7775	51.10	8836	59.77
120C	503	1.86	345	2.27	208	1.41
120D	367	1.36	698	4.59	142	0.96
120U	871	3.23	339	2.23	349	2.36
201	27173	100.68	8728	57.36	7174	48.53
210	2867	10.62	1505	9.89	743	5.03
300	2376	8.80	981	6.45	419	2.83

Table 10. Coded triads descending by their distributions

No.	Coded triads	‱	No.	Coded triads	‱
1	003	25778.85	9	021U	53.92
2	012	1996.09	10	210	25.54
3	**102**	**1464.19**	11	300	18.09
4	**201**	**206.57**	12	120U	7.82
5	**111U**	**186.84**	13	120D	6.91
6	**111D**	**119.11**	14	030T	6.35
7	021D	67.94	15	120C	5.54
8	021C	55.83	16	003C	0.42

Excluding the normal triads coded “003”, “012”, the triads coded “102”, “201”, “111U” or “111D” with mutual dyads outnumber other triads. The most common forms of the relationships between celebrities and audience or fans are one way relationships while the most common forms of the relationships between close friends are two way relationships [10]. In this paper, we can infer that the common forms in debate networks are mutual dyads.

5.2 Triadic Closure

Triadic closure process is one of the fundamental processes of structure formation. There is an increased chance that a friendship will form between two persons if they already have a friend in common [11].

To disclose the behavior patterns of participants with different opinions, we divide the participants in the debate into different groups by their stances: an "abolishing TCM" camp and a "preserving TCM" camp. We manually label the stances of authors by reading all their replies. We study the characteristics of the two main opposite groups in this sub-network. We find that nobody change their opinions during the debate. We compare the coded triadic closures in the two opposite camps. We choose the coded triads in triadic closure patterns: "030C", "030T", "120C", "120D", "210" and "300" in Fig. 2. Table 11 lists the size of the coded triadic closures and the ratio of these triads. From Table 11, we can see that a obviously higher proportion of directed triadic closures is in the "abolishing TCM" camp than that in the "preserving TCM" camp.

Table 11. Coded triadic closures in different camps

Thread ID	Preserving TCM		Abolishing TCM	
	Coded triadic closures	‰o	Coded triadic closures	‰o
1	141	5.66	67	183.28
2	123	9.48	69	97.91
3	100	5.00	58	70.33

5.3 Structural Balance

According to the balance theory, some social relations are more stable than others. For the triads, a friend of my friend is possibly more of my friend than my enemy. Balance is achieved when there are three positive links or two negatives with one positive. Two positive links and one negative creates imbalance as shown in Fig. 3.

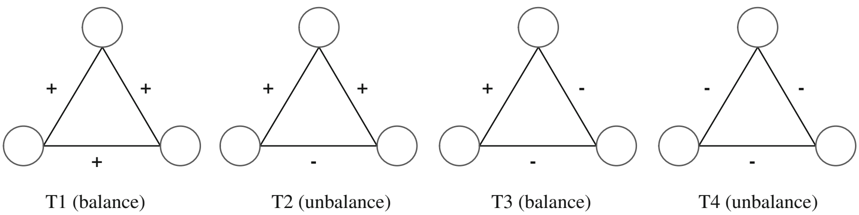

Fig. 3. Balanced and unbalanced triadic relationship [12]

We classify the edges into two types, the inter-camp edges whose end-points belong to different camps, and the inner-camp edges whose end-points come from the same camp. Then triads can be classified into two types "inner-camp" triads and "inter-camp" triads. We calculate the triads coded "300" (in Fig. 2) in the three

Table 12. Balanced "300" triads

Sub-graph	"preserving TCM" camp	"abolishing TCM" camp	inner-camp	inter-camp
1	48	312	360	2016
2	43	113	156	825
3	10	42	52	367

sub-graphs including the active participants in Table 9. Table 12 lists the sizes of the balanced "300" triads in the two camps and the size of the "inner-camp" balanced "300" triads (T1 in Fig. 3) and "inter-camp" balanced "300" triads (T3 in Fig. 3). Obviously, the balanced triads between camps are more than those within camps.

An interesting characteristic of many newsgroups is that people are more likely to respond to a message which they disagree than which they agree. This behavior is in sharp contrast to the WWW link graph, where linkage is an indicator of agreement or common interest [5]. This paper verifies the founding of Agrawal et al. from the perspective of structural balance.

6 Conclusions

This study explores the characteristics of online debate networks. Taking the debate on TCM at Tianya Forum as instance, we mine the interaction patterns at three levels.

At the individual level, some users reply to different threads related to the same topic. Authors of the original posts are the key players in reply networks. Some relationships among these authors are consistent across networks.

At the whole network level, the stance of an author who starts a thread affects the structural indices of the reply network. In our corpus, authors of original posts of Thread 1 and Thread 3 hold the "abolishing TCM" stance and the author of original post of Thread 2 holds the "preserving TCM" stance. That may explain why the percentage of connected nodes, average indegree, clustering coefficient of Network 2 are the lowest.

At the triad level, we find that mutual dyads are common form relationship in debate networks; balanced triads between the two camps are more than those within camps; there are more triadic closures within the "abolishing TCM" camp, which means the participants holding this opinion are more active to interact with each other.

In the future, we will do more studies to identify the behavior patterns of participants within debates. We will also combine the reply network analysis and text analysis to explore how opposing perspectives and arguments are put forward.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Nos. 61473284 and 71371107).

References

1. Somasundaran, S., Wiebe, J.: Recognizing stances in ideological on-line debates. In: NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, pp. 116–124 (2010)
2. Liu, B.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
3. Wang, C., Tang, X.J.: Stance analysis for debates on traditional Chinese medicine at Tianya Forum. In: 5th International Conference on Computational Social Networks, Ho Chi Minh, pp. 321–332 (2016)
4. Agrawal, R., Rajagopalan, S., Srikant, R., et al.: Mining newsgroups using networks arising from social behavior. In: Proceedings of the 12th International Conference on World Wide Web, pp. 529–535. ACM, Budapest (2003)
5. Yang, Y., Chen, Q., Liu, W.: The structural evolution of an online discussion network. Phys. A Stat. Mech. Appl. **389**(24), 5871–5877 (2010)
6. Zhao, Y.L., Tang, X.J.: In-depth analysis of online hot discussion about TCM. In: 15th International Symposium on Knowledge and Systems Science, pp. 275–283. JAIST Press, Sapporo (2014)
7. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. Nature **393** (6684), 440–442 (1998)
8. Davis, J.A., Leinhardt, S.: The structure of positive interpersonal relations in small groups. Sociological Theories in Progress **2**, 218–251 (1972)
9. Li, Z.P., Tang, X.J.: Triadic and M-cycles balance in social network. In: 15th International Symposium on Knowledge and Systems Science, pp. 275–283. JAIST Press, Sapporo (2014)
10. Lou, T.C., Tang, J., Hopcroft, J., et al.: Learning to predict reciprocity and triadic closure in social networks. ACM Trans. Knowl. Discov. Data **7**(2), 1825–1835 (2013)
11. Granovetter, M.S.: The strength of weak ties. Soc. Sci. Electron. Publishing **78**(2), 1360–1380 (2015)
12. Li, Z.P., Tang, X.J.: Group polarization: connecting, influence and balance, a simulation study based on Hopfield modeling. In: 12th Pacific Rim International Conference on Artificial Intelligence, Kuching, pp. 710–721 (2012)