Temporal Interaction Patterns Based on User Attributes in Online Health Communities (Postprint)

Authors: Wu Bing, Peng Yu

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] Online health communities are developing rapidly; however, research integrating node attributes with dynamic network features of network structure is lacking, making it difficult to reveal the formation mechanisms of dynamic user interaction patterns based on node attributes. [Method / Process] This study applies a node-attribute-based temporal exponential random graph model (TERGM), taking online health communities as the research object. By incorporating user node attribute features—including sentiment orientation of user posts, length of user posts, and user community rank—we construct a research model for user temporal interaction patterns in online health communities. Data were crawled from the Diabetes Bar of Baidu Tieba between October 2016 and February 2018, comprising 2,301 valid users, 6,045 main posts, and 9,490 replies, to empirically investigate features of user temporal interaction patterns. [Results / Conclusions] User attribute features exert significant influence on the formation of reciprocal temporal patterns, k-star temporal patterns, transitive temporal patterns, and cyclic temporal patterns. Based on these findings, development suggestions are proposed for online health community construction.

Full Text

Preamble

ChinaXiv Cooperative Journal: Academic Exploration
Temporal Interaction Patterns Based on User Attributes in Online Health Communities
Wu Bing, Peng Yu
School of Economics and Management, Tongji University, Shanghai 200092

Abstract

[Purpose/Significance] Online health communities have developed rapidly, yet research integrating network node attributes with dynamic network characteristics remains scarce, making it difficult to reveal the formation mechanisms of user dynamic interaction patterns based on node attributes. **[Method/Process]** This study applies the Node Attribute-based Temporal Exponential Random Graph Model (NATERGM) to online health communities, incorporating user node attribute features—including sentiment orientation of user posts, post text length, and user community level—to construct a research model for user temporal interaction patterns. Data were collected from the Baidu Tieba Diabetes Forum, comprising 2,301 valid users, 6,045 main posts, and 9,490 replies between October 2016 and February 2018, to empirically examine the characteristics of user temporal interaction patterns. **[Results/Conclusion]** The findings indicate that user attribute features significantly influence the formation of reciprocal temporal patterns, k-star temporal patterns, transitivity temporal patterns, and cyclical temporal patterns, providing development recommendations for online health community construction.

Keywords: online health community, node attribute, user interaction pattern, NATERGM

Classification Number: C912

Citation Format: Wu B, Peng Y. Research on temporal interaction patterns based on user attributes in online health communities [J/OL]. Knowledge Management Forum, 2019, 4(3): 163-172 [citation date]. http://www.kmf.ac.cn/p/174/.

Online health communities primarily focus on medical information sharing and exchange, with increasing numbers of people seeking to address health-related issues through these platforms. Consequently, the development of online health communities is driving reforms in traditional healthcare, gradually alleviating the imbalance between medical resource supply and demand in China [1]. Based on user composition, online health communities can be categorized into three types: healthcare professional communities (e.g., Dingxiangyuan), doctor-patient communities (e.g., Haodf.com), and patient communities (e.g., Baidu Medical Tieba), which serve as bridges for mutual discovery and support among healthcare professionals, doctors and patients, and patients themselves. In recent years, the number of chronic disease patients in China has continued to grow. Taking diabetes and its complications as an example, the International Diabetes Federation (IDF) reports that the global diabetic population is projected to reach 629 million by 2045. Chronic disease groups are more inclined to use the internet to search for disease information and communicate with other patients [2], making patient-centered communities crucial platforms for information reciprocity, experience sharing, and emotional support. These communities particularly require revealing the intrinsic mechanisms between user attributes and temporal interaction pattern formation, yet relevant research remains lacking.

Therefore, this study adopts a dynamic integration perspective, applying the Node Attribute-based Temporal Exponential Random Graph Model (NATERGM) to incorporate user node attribute characteristics from online patient communities and construct a research model for user temporal interaction patterns. Using data collected from the Baidu Tieba Diabetes Forum, this study empirically investigates the formation mechanisms of dynamic interaction patterns based on user node attributes. Theoretically, this research enriches and expands the theoretical framework for studying user temporal interaction mechanisms in online health communities; practically, it provides direction and recommendations for the construction and development of online health communities.

1 Literature Review

1.1 Analysis of User Interaction Patterns in Online Health Communities

Online health communities are defined as collections of users concerned with common health issues, where users can be regarded as actors and interactions between users as associations among different actors. Current research on user interaction patterns in online health communities primarily falls into three categories: (1) From a community network structure perspective, social network analysis is employed to describe the structural characteristics of social networks in specific online health communities and their influencing factors. For instance, centrality derived from social network analysis can serve as a primary feature for identifying key users in online health communities [3]. (2) From a community network individual perspective, research focuses on individual attributes and behavioral patterns in online health communities. For example, knowledge-sharing networks in online health communities exhibit small-world effects, where community members are most concerned with experience shared by others, and different user groups demonstrate significant differences in knowledge-sharing behavior and activity duration [4]; analyzing user behavior through social network analysis can explore knowledge-sharing behaviors among different groups and roles in online health community networks [5-6]. (3) From a community network formation mechanism perspective, the Exponential Random Graph Model (ERGM) is applied to study user interaction patterns in online health communities. For example, ERGM has been used to analyze how user attribute features and behavioral characteristics influence network connection patterns [7-9].

Thus, current research on user interaction patterns in online health communities faces two primary issues: (1) Network node attributes and network structure are analyzed in isolation, despite node attributes being essential components of user interaction network structures; (2) Dynamic integration research between network node attributes and user interaction networks is lacking, making it difficult to reveal the intrinsic mechanisms connecting node attributes with dynamic network formation.

1.2 NATERGM

Given an observed network, the primary task of the Exponential Random Graph Model is to statistically test which node connection patterns occur non-randomly, as shown in formula (1) [10]:

$$P(Y=y)=\frac{1}{k}\exp\left\{\sum_{A}\eta_A g_A(y)\right\} \quad (1)$$

where: (1) $P(Y=y)$ is a probability distribution, Y is a random variable matrix representing network connections, and y is the realization of Y; (2) A represents node connection patterns, $g_A(y)$ is the network statistic corresponding to A; (3) k is a scaling parameter ensuring $P(Y=y)$ is a proper probability distribution; (4) η_A is the parameter corresponding to node connection pattern A, positively correlated with the likelihood of pattern A appearing. If parameter η_A is significant, it indicates that node connection pattern A is more likely to appear in the network, suggesting that pattern A plays an important role in the network formation process.

TERGM extends ERGM to dynamic networks. Based on formula (1), the first-order Markov-dependent TERGM is shown in formula (2) [10]. TERGM's primary objective is to detect whether node connection temporal patterns A are more likely to appear across time periods $(t-1)$ and t.

$$P(Y^t=y^t|Y^{t-1}=y^{t-1})=\frac{1}{k}\exp\left\{\sum_{A}\eta_A g_A(y^t,y^{t-1})\right\} \quad (2)$$

NATERGM incorporates node attributes into TERGM. Building upon formula (2), NATERGM is calculated as shown in formula (3) [10]:

$$P(Y^t=y^t|Y^{t-1}=y^{t-1},X)=\frac{1}{k}\exp\left\{\sum_{A}\eta_A g_A(y^t,y^{t-1},X)\right\} \quad (3)$$

where: (1) A is the set of node connection temporal patterns to be validated; (2) $\eta=[\eta_a]$ is a vector parameter representing the influence of each node connection temporal pattern on network formation; (3) $g_a(y^t,y^{t-1},X)$ is the network statistic for temporal pattern a; (4) parameter estimation employs Markov Chain Monte Carlo (MCMC) methods.

Therefore, NATERGM's primary objective is to test whether node connection temporal patterns A are more likely to appear at time T for nodes with attributes X.

2 NATERGM Analysis Framework for User Temporal Interaction Patterns in Online Health Communities

NATERGM accurately tracks the temporal sequence of network node connections by examining time-stamped network node connection cross-sectional data. The basic analysis framework is shown in Figure 1 [Figure 1: see original paper].

2.1 Network Construction

First, a directed network connection is constructed based on user interactions in the online health community. Y_{ij} represents the directed connection between network user nodes i and j; if node i points to node j, then $Y_{ij} = 1$, otherwise $Y_{ij} = 0$. After identifying all directed connections between user nodes, the directed network connection with N nodes can be represented as matrix $Y = [Y_{ij}], (i, j = 1, 2, \dots, N)$.

Second, based on the directed network connections of user interactions in the online health community, T_{ij} records the establishment time of the directed connection between network user nodes i and j. If node i points to node j earlier than node j points to node i, then $T_{ij} < T_{ji}$. After recording the establishment times of all directed connections between user nodes, the directed network connection times with N nodes can be represented as matrix $T = [T_{ij}], (i, j = 1, 2, \dots, N)$.

Third, based on the node attribute characteristics of users in the online health community, x_i represents the attribute value of each network node i. The directed network node attribute values with N nodes can be represented as vector $X = [x_1, x_2, \dots, x_N], (i, j = 1, 2, \dots, N)$. According to the social support function of online health communities, this study selects user post sentiment orientation, user post text length, and user community level as user attribute vectors because: (1) emotional factors play a crucial role in the formation of patient interactions in online health communities [11]; (2) users with longer average post lengths and higher writing ability help provide more information support [12]; (3) in online health communities, different user behaviors such as daily check-ins, posting, and replying are assigned varying experience values, and accumulated experience levels reflect user participation and contribution.

2.2 NATERGM Fitting for Temporal Interaction Patterns

Based on user attribute vectors in online health communities, this study focuses on dyadic relationships [13] of reciprocity and triadic relationships [13] of k-star, transitivity, and cyclicity, as shown in Tables 1 -4 . In these tables, white nodes represent general users, black nodes represent users with certain attribute features, dashed arrows represent network connections developed based on solid arrow connections; when node attribute X_i is categorical, $I(X_i) = 0$ or 1; when node attribute X_i is continuous, $I(X_i)$ is the attribute value of X_i.

2.2.1 Reciprocal Temporal Patterns Reciprocity refers to subsets formed by two mutually connected nodes in a network, described as the edge set $\{(j \rightarrow i), (i \rightarrow j)\}$. The essence of online communities is an entity based on social capital exchange and cultural reciprocity [14], and user interactions in online health communities exhibit reciprocity [7-8]. Therefore, this study examines how node attributes influence the formation of reciprocal temporal connections, with temporal hypotheses and network statistics formulas shown in Table 1:

Table 1: Temporal Reciprocity in NATERGM

Pattern	Hypothesis
Feedback	H1a: Users with positive post sentiment are more likely to receive feedback from other usersH1b: Users with longer post text are more likely to receive feedback from other usersH1c: Users with higher community level are more likely to receive feedback from other users
Response	H2a: Users with positive post sentiment are more likely to reply to other usersH2b: Users with longer post text are more likely to reply to other usersH2c: Users with higher community level are more likely to reply to other users

2.2.2 k-star Temporal Patterns The k-star pattern refers to subsets formed by a node and several other nodes connected to it. This study focuses on 2-star patterns. In directed graphs, k-stars are divided into k-in-star and k-out-star. k-in-star reflects a node's popularity or convergence in the network to some extent, while k-out-star reflects a node's divergence or social influence [15]. User popularity and social influence affect network connection formation [15], and k-star connections exist in online health community user interactions [7]. Therefore, this study examines how node attributes influence k-star temporal connections, with temporal hypotheses and network statistics formulas shown in Table 2 :

Table 2: Temporal k-star in NATERGM

Pattern	Hypothesis
k-in-star: Initial	H3a: Users with positive post sentiment are more likely to initially connect to other usersH3b: Users with longer post text are more likely to initially connect to other usersH3c: Users with higher community level are more likely to initially connect to other users
k-in-star: Laziness	H4a: Users with positive post sentiment are more likely to delay connecting to other usersH4b: Users with longer post text are more likely to delay connecting to other usersH4c: Users with higher community level are more likely to delay connecting to other users
k-out-star: Priority	H5a: Users with positive post sentiment are more likely to be prioritized for connection by other usersH5b: Users with longer post text are more likely to be prioritized for connection by other usersH5c: Users with higher community level are more likely to be prioritized for connection by other users
k-out-star: DePriority	H6a: Users with positive post sentiment are more likely to have delayed connections from other usersH6b: Users with longer post text are more likely to have delayed connections from other usersH6c: Users with higher community level are more likely to have delayed connections from other users

2.2.3 Transitive Temporal Patterns For three nodes i, j, k in a network, transitivity can be described as the edge set $\{(i \rightarrow j), (i \rightarrow k), (j \rightarrow k)\}$. This study examines transitive bridging, transitive follow-up, and transitive reference. In social networks, people tend to establish relationships with friends of friends [16], and transitive connection relationships exist in online health community user interactions [7]. Therefore, this study examines how node attributes influence transitive temporal connections, with temporal hypotheses and network statistics formulas shown in Table 3 :

Table 3: Temporal Transitivity in NATERGM

Pattern	Hypothesis
Bridge	H7a: Users with positive post sentiment are more likely to bridge other usersH7b: Users with longer post text are more likely to bridge other usersH7c: Users with higher community level are more likely to bridge other users
Follow-up	H8a: Users with positive post sentiment are more likely to transitively follow other usersH8b: Users with longer post text are more likely to transitively follow other usersH8c: Users with higher community level are more likely to transitively follow other users
Reference	H9a: Users with positive post sentiment are more likely to be transitively referenced by other usersH9b: Users with longer post text are more likely to be transitively referenced by other usersH9c: Users with higher community level are more likely to be transitively referenced by other users

2.2.4 Cyclical Temporal Patterns For three nodes i, j, k in a network, cyclicity can be described as the edge set $\{(i \to j), (j \to k), (k \to i)\}$. Corresponding to transitivity research, this study examines reversed bridging, reversed follow-up, and reversed reference. Research on baboon social networks has shown that when food is scarce, individuals exhibit obvious cyclical interactions [17], and cyclical connection relationships exist in online health community user interactions [7]. Therefore, this study examines how node attributes influence cyclical temporal connections, with temporal hypotheses and network statistics formulas shown in Table 4 :

Table 4: Temporal Cyclicity in NATERGM

Pattern	Hypothesis
Reversed Bridge	H10a: Users with positive post sentiment are more likely to reverse-bridge other usersH10b: Users with longer post text are more likely to reverse-bridge other usersH10c: Users with higher community level are more likely to reverse-bridge other users

Pattern	Hypothesis
Reversed Follow-up	H11a: Users with positive post sentiment are more likely to cyclically follow other usersH11b: Users with longer post text are more likely to cyclically follow other usersH11c: Users with higher community level are more likely to cyclically follow other users
Reversed Reference	H12a: Users with positive post sentiment are more likely to be cyclically referenced by other usersH12b: Users with longer post text are more likely to be cyclically referenced by other usersH12c: Users with higher community level are more likely to be cyclically referenced by other users

2.3 NATERGM Fitting Effect Comparison

The comparison procedure is as follows: First, based on the actual network observed at time point $t-1$, estimate NATERGM parameters η^{t-1}. Next, based on NATERGM, simulate and generate K networks at time point t, denoted as $y_k^t(k = 1, 2, \ldots, K)$. Third, compare the absolute differences between the K simulated networks and the actual network, as shown in formula (4):

$$AD_a = \frac{1}{K}\sum_{k=1}^{K} |g_a(y_k^t) - g_a(y_0^t)| \quad (4)$$

where y_0^t is the actual network at time point t. Fourth, since NATERGM adds node attributes to the TERGM temporal patterns, this study uses TERGM as the baseline model to compare NATERGM fitting effects. The Wilcoxon signed-rank test is employed to verify whether NATERGM fitting errors are significantly lower than TERGM fitting errors, as shown in formula (5) [10]:

$$W = \sum_{i=1}^{T} I(AD_i^{\text{baseline}} > AD_i^{\text{NATERGM}})R_i \quad (5)$$

where: (1) T is the comparison period, $I()$ is the indicator function for whether the baseline model TERGM fitting error is greater than NATERGM's, and R_i is the rank of the ith comparison pair; (2) $\mu_W = n(n+1)/4$; (3) $\sigma_W^2 = n(n+1)(2n+1)/24$. A higher W value indicates that NATERGM fitting errors are significantly lower than TERGM fitting errors.

3 Empirical Study of Baidu Diabetes Tieba

Since its establishment in 2003, although Baidu Tieba is not a professional online health community, it has attracted a large number of patients with numerous advantages. Currently, it has over 1 billion registered users, more than 300 million monthly active users, and 19 million themed forums. This study selected the Baidu Diabetes Tieba as the research object primarily because diabetes is a relatively common disease with a broad patient population, and patients are more inclined to seek emotional comfort and information support.

3.1 Data Collection and Preprocessing

3.1.1 Network Connections Since Baidu Tieba regularly cleans redundant data, the data collection period spanned from October 2016 to February 2018, totaling 66 weeks. Using Python's Scrapy framework, the Baidu Diabetes Forum contained 23,777 users, 77,970 main posts, and 80,729 replies during the collection period. Specifically: (1) Main posts are broadcast directly by users in the community and appear on the community homepage in chronological order, serving as invitations to establish user connections; (2) Replies are made in the discussion area under main posts, establishing user connections through free responses.

After data collection, the following cleaning and preprocessing steps were applied: (1) Delete main posts without any replies; (2) Delete users who did not participate in any interactions; (3) Delete posts containing only videos/emoticons/images; (4) Delete duplicate posts with similar timestamps. After preprocessing, 2,301 valid users, 6,045 main posts, and 9,490 replies were obtained.

3.1.2 Node Attributes **(1) User Post Sentiment Orientation.** Based on the Chinese sentiment polarity dictionary NTUSD, the sentiment score $\mathrm{senti}(p_{ij})$ for the jth post by user i is calculated as shown in formula (6) [18]:

$$\mathrm{senti}(p_{ij}) = \frac{1}{|p_{ij}|} \sum_{s \in p_{ij}} \delta(s) \frac{\sum_{w \in s} \mathrm{ntusd}(w, \mathrm{pol}(w))}{\sum_{w \in s} |\mathrm{pol}(w)|} \quad (6)$$

where: (1) $|p_{ij}|$ is the number of sentences in the jth post by user i; (2) s is a sentence in post p_{ij}; (3) If sentence s is emotionally non-neutral, $\delta(s) = 1$; if neutral, $\delta(s) = 0$; (4) w is an emotional expression word in sentence s; (5) $\mathrm{pol}(w)$ is the polarity of w in sentence s; (6) $\mathrm{ntusd}(w, \mathrm{pol}(w))$ is the NTUSD score of w based on $\mathrm{pol}(w)$.

Thus, the post sentiment orientation $A_{\mathrm{senti},i}$ for user i is the average sentiment of all n posts during the data collection period, as shown in formula (7). After standardizing $A_{\mathrm{senti},i}$ to $[0, 1]$, it serves as user i's sentiment value.

$$A_{\text{senti},i} = \frac{1}{n}\sum_{j=1}^{n} \text{senti}(p_{ij}) \quad (7)$$

(2) User Post Text Length. The text length $A_{\text{len},i}$ for user i is the average text length of all n posts during the data collection period, as shown in formula (8):

$$A_{\text{len},i} = \frac{1}{n}\sum_{j=1}^{n} \text{Length}(p_{ij}) \quad (8)$$

where $\text{Length}(p_{ij})$ is the text length of the jth post p_{ij} by user i. After standardizing $A_{\text{len},i}$ to $[0,1]$, it serves as user i's post text length value.

(3) User Community Level. Baidu Tieba's level system, implemented on September 20, 2011, awards community levels based on accumulated online duration, community discussion participation, and daily check-ins. In Tieba, user community levels range from 1-9, reflecting user participation. The community level classification value for user i is calculated as shown in formula (9):

$$\text{Level}_i = \begin{cases} 0 & \text{if level } 1-5 \\ 1 & \text{if level } 6-9 \end{cases} \quad (9)$$

where: (1) When user community level is in the 1-5 range, since required experience points are minimal and new registered users can reach these levels within days, they cannot accurately reflect user seniority, thus user i's community level is assigned 0; (2) When user community level is in the 6-9 range, requiring accumulated experience over time, user i's community level is assigned 1.

Based on formulas (7)-(9), descriptive statistics for node attributes are shown in Table 5 :

Table 5: Descriptive Statistics of Node Attributes (N=2301)

Attribute	Mean	Std. Dev.	Min	Max
User post sentiment orientation	0.45	0.12	0.00	1.00
User post text length	0.38	0.15	0.00	1.00
User community level	0.41	0.49	0.00	1.00

3.2 Empirical Results

3.2.1 Influence of User Post Sentiment Orientation The NATERGM fitting results for user post sentiment orientation in the Baidu Diabetes Tieba are shown in Table 6 . Among the 12 hypotheses for user post sentiment orientation, 4 are not supported. This is because users with positive post sentiment are more likely to initially connect to other users, and consequently more likely to be prioritized for connection, transitively referenced, and cyclically referenced by other users.

Table 6: NATERGM Fitting Results for User Post Sentiment Orientation

Parameter	Estimate	Std. Error	Hypothesis Test
Feedback	1.683***	0.412	H1a supported
Response	2.017***	0.389	H2a supported
k-in-star: Initial	0.195***	0.034	H3a supported
k-in-star: Laziness	-0.071	0.090	H4a not supported
k-out-star: Priority	0.139***	0.041	H5a supported
k-out-star: DePriority	0.090	0.112	H6a not supported
Bridge	-0.034	0.078	H7a not supported
Follow-up	0.473***	0.095	H8a supported
Reference	0.334***	0.087	H9a supported
Reversed Bridge	0.268	0.156	H10a not supported
Reversed Follow-up	0.339***	0.098	H11a supported
Reversed Reference	0.285***	0.089	H12a supported

Note: ** p<0.001; ** p<0.01; * p<0.05*

3.2.2 Influence of User Post Text Length The NATERGM fitting results for user post text length in the Baidu Diabetes Tieba are shown in Table 7 . Among the 12 hypotheses for user post text length, 4 are not supported. This is because users with longer post text are more likely to delay connections, bridge, and reverse-bridge other users, and consequently more likely to be transitively referenced and cyclically referenced by other users.

Table 7: NATERGM Fitting Results for User Post Text Length

Parameter	Estimate	Std. Error	Hypothesis Test
Feedback	1.412***	0.398	H1b supported
Response	1.863***	0.412	H2b supported
k-in-star: Initial	-0.071	0.087	H3b not supported
k-in-star: Laziness	0.195***	0.041	H4b supported
k-out-star: Priority	-0.034	0.095	H5b not supported
k-out-star: DePriority	0.231***	0.067	H6b supported
Bridge	0.813***	0.134	H7b supported
Follow-up	-0.034	0.102	H8b not supported
Reference	0.391***	0.089	H9b supported
Reversed Bridge	1.079***	0.187	H10b supported

Parameter	Estimate	Std. Error	Hypothesis Test
Reversed Follow-up	0.071	0.134	H11b not supported
Reversed Reference	0.192***	0.067	H12b supported

Note: ** p<0.001; ** p<0.01; * p<0.05*

3.2.3 Influence of User Community Level The NATERGM fitting results for user community level in the Baidu Diabetes Tieba are shown in Table 8 . Among the 12 hypotheses for user community level, 5 are not supported. This is because users with higher community levels are more likely to initially connect, bridge, and transitively follow other users, and consequently more likely to reverse-bridge and cyclically follow other users.

Table 8: NATERGM Fitting Results for User Community Level

Parameter	Estimate	Std. Error	Hypothesis Test
Feedback	3.390***	0.512	H1c supported
Response	1.770***	0.423	H2c supported
k-in-star: Initial	1.490***	0.187	H3c supported
k-in-star: Laziness	0.110***	0.034	H4c not supported
k-out-star: Priority	0.562***	0.134	H5c not supported
k-out-star: DePriority	0.318**	0.112	H6c not supported
Bridge	0.341***	0.098	H7c supported
Follow-up	0.235***	0.087	H8c supported
Reference	-0.034	0.102	H9c not supported
Reversed Bridge	0.813***	0.156	H10c supported
Reversed Follow-up	0.473***	0.134	H11c supported
Reversed Reference	0.090	0.134	H12c not supported

Note: ** p<0.001; ** p<0.01; * p<0.05*

3.3 NATERGM Goodness-of-Fit Test

The goodness-of-fit test employs degree distribution parameters [19], including standard indegree, standard outdegree, skewness indegree, and skewness outdegree, as degree distributions best reflect how users connect and participate in interactions in social media networks. First, for the data during the collection period, weekly fitting tests were conducted according to formula (4). For each interaction pattern a, absolute difference vectors AD_a^{TERGM} and AD_a^{NATERGM} were obtained based on TERGM and NATERGM, respectively. Then, using TERGM as the reference, the Wilcoxon signed-rank test was performed, with results shown in Table 9 .

Table 9: Wilcoxon Signed-Rank Test Results

Interaction Pattern	W Statistic	Significance
Reciprocity	5.98***	p<0.001
k-star	5.67***	p<0.001
Transitivity	4.45***	p<0.001
Cyclicity	4.89***	p<0.001

Note: ** p<0.001; ** p<0.01; * p<0.05*

The results indicate that NATERGM fitting errors are significantly lower than TERGM fitting errors, demonstrating better model fit.

4 Research Conclusions and Community Development Recommendations

Based on NATERGM, this study constructed an analytical model for user temporal interaction patterns in online health communities, using the Baidu Diabetes Tieba as an empirical research object. The study validated the effectiveness of applying NATERGM to analyze user temporal interaction patterns and revealed how user attributes influence the formation of these patterns [19].

4.1 Formation of Reciprocal Temporal Patterns

All reciprocal temporal hypotheses H1a, H1b, H1c, H2a, H2b, and H2c are supported, meaning that users with positive post sentiment, longer post text, and higher community levels tend toward reciprocal temporal feedback and response. Therefore, it is recommended that online health communities provide instant messaging-like mechanisms to facilitate direct user communication and offer recommendation lists for post authors to invite users with positive sentiment, longer posts, and higher community levels to participate in discussions, thereby promoting active user interactions.

4.2 Formation of k-star Temporal Patterns

For k-in-star temporal initial hypotheses, H3a and H3c are supported while H3b is not; for k-in-star temporal laziness hypotheses, H4b is supported while H4a and H4c are not. This means that when users receive answers from many forum members, answers from users with positive sentiment and higher community levels are likely to appear earlier—these users tend to reply promptly; users with longer post text tend to reply later but more comprehensively.

For k-out-star temporal priority hypotheses, H5a is supported while H5b and H5c are not; for k-out-star temporal lag hypotheses, H6b is supported while H6a and H6c are not. This means that when many users seek answers, users with positive sentiment are more likely to be prioritized for response, but there is no prioritization for users with higher community levels; users with longer post text are more likely to receive delayed responses because interpreting long texts requires time. Online health community users focus more on emotional and information support.

Emotional support helps build an active community atmosphere, while information support directly reflects users' expression level and engagement. Therefore, it is recommended that online health communities provide reply rewards based on post content richness to improve user reply quality.

4.3 Formation of Transitive and Cyclical Temporal Patterns

For transitive temporal bridging hypotheses, H7b and H7c are supported while H7a is not; for transitive temporal follow-up hypotheses, H8a and H8c are supported while H8b is not; for transitive reference hypotheses, H9a and H9b are supported while H9c is not. For cyclical temporal bridging hypotheses, H10b and H10c are supported while H10a is not; for cyclical temporal follow-up hypotheses, H11a and H11c are supported while H11b is not; for cyclical temporal reference hypotheses, H12a and H12b are supported while H12c is not.

This indicates that users with positive sentiment are more likely to actively follow other users and consequently be referenced by others, but are not yet sufficient to serve as bridging intermediaries; users with longer post text are more likely to serve as bridges and be referenced by other users because their posts contain substantial information that provides important information support during user interactions; users with higher community levels are more likely to bridge and follow other users due to their rich experience and ability to seek effective information sources.

Information support and experience sharing largely function as intermediaries in interaction networks. Therefore, it is recommended that online health communities provide recommendation lists that leverage the influence of users with longer posts and the self-experience of users with higher community levels to provide interaction opportunities for other users.

5 Research Limitations and Future Directions

This study selected user post sentiment orientation, post text length, and community level as user node attributes. Future research needs to improve node attribute completeness, such as adding user post professionalism. Regarding the influence of node attributes, this study examined the impact of single node attributes on network dynamic formation; future research should extend to investigate network dynamic formation mechanisms under the interaction of multiple node attributes based on existing NATERGM temporal patterns. In terms of empirical object selection, this study used the Baidu Diabetes Tieba as the research community; future research should conduct empirical studies in more online health communities to explore differences in user interaction temporal patterns across different communities.

References

[1] Wu J, Zhou L. Cross-regional flow of medical information resources: research on the role of online medical communities in optimizing medical resource allocation [J]. Information Science, 2017, 4(1): 58-65.

[2] Long T. Research on continuous usage behavior in online medical communities and its impact on doctor-patient relationships [D]. Hefei: Hefei University of Technology, 2017.

[3] Zhao K, Greer G E, Yen J, et al. Leader identification in an online health community for cancer survivors: a social network-based classification approach [J]. Information systems and e-Business management, 2015, 13(4): 629-645.

[4] Wu J, Zhou L. Research on knowledge sharing networks and knowledge interaction behavior in online medical communities [J]. Journal of Information Resources Management, 2017, 35(3): 144-151.

[5] Zhai Y, Zhang X, Wang F. User participation behavior in online health communities—taking "Baidu Quit Smoking Bar" as an example [J]. Library and Information Service, 2017(7): 75-82.

[6] Stewart S A, Abidi S S R. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum [J]. Journal of medical Internet research, 2012, 14(6): e170-e175.

[7] Liu X, Wang L, Li J, et al. Research on the influencing mechanism of user reply behavior in online health communities [J]. Management Science, 2017, 1(30): 62-72.

[8] Wu B, Jiang S, Chen H C. Effects of individual gratifications in online health forums [J]. Social behavior and personality, 2016, 44(2): 299–312.

[9] Zhang X, Xia H, Chen X, et al. Research on influencing factors of information credibility in online health communities [J]. Library and Information Service, 2015(22): 88-96.

[10] Jiang S, Chen H C. NATERGM: A model for examining the role of nodal attributes in dynamic social media networks [J]. IEEE transactions on knowledge and data engineering, 2016, 28(3): 729-740.

[11] Luo P, Chen K, Wu C H, et al. Exploring the social influence of multichannel access in an online health community [J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 98-109.

[12] Larsen-Freeman D. Adjusting expectations: the study of complexity, accuracy, and fluency in second language acquisition [J]. Applied linguistics, 2009, 30(4): 579-589.

[13] Zhang M, Ma Z, Zhang Y. The formation path of users' subjective knowledge hiding behavior in online health communities [J]. Information Studies: Theory & Application, 2018(10): 111-117.

[14] Fan X, Ai S. An empirical study on the impact of participation behaviors from both sides in online medical communities on knowledge exchange effectiveness [J]. Journal of Intelligence, 2016(7): 173-179.

[15] Jiang S, Gao Q, Chen H C. The roles of sharing, transfer, and public funding in nanotechnology knowledge diffusion networks [J]. Journal of the Association for Information Science and Technology, 2014, 66(1): 150-160.

[16] Goodreau S M, Morris M. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks [J]. Demography, 2009, 46(1): 103-125.

[17] Zhang M, Liu X, Zhang Y. Online health community users' diagnostic information seeking behavior—external factors, individual motivations, and formation paths [J]. Modern Information, 2018(11): 18-24.

[18] Zimbra D, Chen H C. A stakeholder approach to stock prediction using finance social media [J]. IEEE intelligent systems, 2011, 26(6): 88-92.

[19] Hanneke S, Fu W, Xing E P. Discrete temporal models of social networks [J]. Electronic journal of statistics, 2010, 4(1): 585-605.

[20] Goh J M, Gao G D, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities [J]. MIS quarterly, 2016, 40(1): 247-263.

Author Contributions

Wu Bing: Research design and manuscript writing;
Peng Yu: Data collection and model fitting.

Note: Figure translations are in progress. See original paper for figures.