

Leaders Matter: Knowledge Sharing in a Medical Discussion Forum in China

Yugang Fu

Nanjing University of Chinese Medicine
fuyugang11@foxmail.com

Yang Shen

Nanjing University of Chinese Medicine
596356450@qq.com

May 28, 2019

Abstract

Knowledge sharing is one common activity for members in a professional forum. DingXiangYuan Medical Forum, short as DXY, a famous academic medical forum in China, has gained great reputation with a decade efforts in medical and life science field. However, though some articles summarize some general operating characteristics of DXY, there hasn't yet appear research that reveals the interaction pattern in detail of DXY.

In this paper we illustrate patterns of interactions in DXY forum using social network analysis. We investigate activities in different levels to reveal what influences most the vigor of the medical forum, and find that ask-leaders, who often raise a question contribute most. In addition, we have a look at the commercial impact towards the forum. Due to DXY's regulation institution, its commercial activity hasn't influenced the academic air in the forum, but may help attract more people in this field to register.

1 Introduction

Knowledge sharing via Internet turns into a common activity for individuals. This process should be regarded

as dynamic rather than semantic because it will elicit to accumulation, transformation, and co-creation of knowledge.[10] In the early days, bulletin board system(bbs) prevails for its function of sharing and user-friendly navigation bars. In these days most forums, though lose its popularity in the public sight, still maintain active in certain professional fields where people share knowledge, experience, and stories in life. They are also known as online communities that have oriented discussion issues for particular groups of people.

In China, medical forum hasn't attracted too much attention and didn't show up in the public sight in recent years due to its professionalism. Most of them serves for people with special identities as we mentioned before, which blocked out the entrance of mass media. However, DingXiangYuan Medical Forum(short as DXY) broke this old air, with swift commercialization, the founder built up his new company, and profited via Apps in mobile terminal or other information subscription platforms like WeChat. According to Z.Lv's report[16], DXY began its way to commercial website from an academic since 2006, and first profited since 2009. On its website, we could find hyperlinks to its products facing clinical doctors and researchers in life sciences.

However, though DXY is aimed at academic communication and knowledge sharing, we have to clarify that DXY is not a typical Q&A community[4][1][3], where all posts are strictly follow the format of ask-answer. DXY has multiple forms of interaction and we will present them with concrete examples in section 2.

So far as we haven't find any paper researching on whether these promotions in business affected the pattern of the communication, neither any of former papers depict in detail what actually is the communication pattern in DXY. Though the main task of this paper is to depict the interaction patterns in the forum, we will have a look at the commercial impact in section 4.2.

We believe that the structure of communication in such professional forums is quite different from that in mass media. Unlike in mass media elites in real life may lose their influence when they are engaged in Web[14], in DXY, members who lead the discussions also have higher ranks and reputation in real life mainly because they accumulated much practical knowledge in their career. Therefore, the roles of members are more like a projection from real to the virtual, rather than rebuild a new order. We will discuss this issue in section 3.2.

Most professional forums don't care much for its enrollment strategy, and many of which are non-commercial at the beginning. Many professional forums wait for new members to come in by search engines. They seldom promote outside, for its audience limited to a small group of people. Only those who have the same interest or knowledge could get the access to, and join in the discussions. For medical forums in China, the situation hasn't changed much. These websites survived only by the support of a certain number of amateurs, or people engaging in medical related fields.

Thus this article concentrates on depicting basic interaction patterns in DXY medical forum within three analysis levels, and in the community-level analysis we additionally

take a look at the potential impact of commercial activities on the user population of the forum. In general we have endeavored on the following tasks:

1. Who are in the center of the discussion, and who are at the edge?
2. What are the roles the actors play in based on their labels and activities(ask/reply)?
3. What is the distance and reachability of actors? Are they linked to each other dense or loose?
4. The community division of the whole graph, and what might contribute to the emerge of clusters?
5. What makes the forum active, with regard to the regulation rules, or member leaders?
6. What do the motif patterns tell us about the transitivity among members?

2 Forms of Interaction

In this section, we summarize usual forms of posts and interactions in the DXY forum. In general, the forms of interaction on DXY include questions-and-answers(47.68%), series discussion(44.41%), board management(0.6%) and others(7.31%).

Questions-and-answers(Q&As) include clinical questions, academic questions, and information about hospitals. In the first scenario, practitioners or amateurs raise a question they met during work or study, and other users reply. In the second scenario, askers may introduce some newest development in some treatment, and share this information to other members, or they seek for further illustrations for some details they don't understand[8]. In the last scenario, askers intend to get the hospital information they will go for further study or training. Here's an example for scenario one.

Q1: What should you do in patients with epileptic seizures?

A1: Your failure to do anything may be the greatest help to the patient. When the patient contracts his muscles and twitches all over his body, you may hold the patient down or wrench his hand, which may cause irreparable damage or even kill him.

A2: At least you should keep head-to-side slip first.

A3: Is it okay to give muscle relaxants?

Series discussion is similar to the Q&As in content, as they both include case discussion. However, series discussions are often raised by some experienced and heart-minded members, in which he or she shares knowledge with a personal prefix in the title. In recent posts, virtual users running by the official also initiate series discussion. Here we present a title of one series discussion by a personal account.

Outpatients Clinic on Tuesday: xxx

Other forms of interaction may be daily routine sharing, personal experience in hospital, or complaints about medical environment (relationship between doctors and patients sometimes is tense). And a few posts are administrative, taking the function of keeping rules well obeyed.

In summary, knowledge sharing accounts for most of the posts, mainly includes case discussion, experience sharing, etc. Practitioners or experts will share their experience of diagnosis, prescription, surgery plan for other members to learn, and other users will give feedback on their shared cases.

We extracted the 10 most frequent words in the title of topic. You could see in the Figure 1 that most topics are around case discussion, and the scope of discussions within the neurology subject.

2.1 The datasets preparation

First, we select posts that have more than 40 threads, scraping both the interactions and attributes of actors, recorded in graph notations. As a result, we get 18294

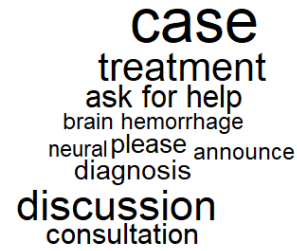


Figure 1: 10 top frequent words

nodes(actors), and 29306 edges. Second, we use this actor set as an anchor, diving into their profile page, collecting rest interactions in this sub-forum. That's to say, both sides of the interactions we catch in the second step should exist in the anchor. Therefore we get the original dataset which has 18179 nodes and 64333 edges.

We prepare two datasets to form the graph to be analyzed. The first dataset with the filename "nodes.csv", the first column of which is username, the following are node attributes; the second dataset with the filename "edges.csv", the first two columns depict the direction, "from" as the sender, "to" as the receiver, the last one is the edge attribute. Variables in the dataset show as follows.

DATASET 1: nodes.csv

- name, the user who participates in the post.
- totaltimes, speaking times of a user in all posts.
- label1, the department of the replier belonging to.
- label2, the rank of the replier, e.g. medical student, specialist, practicing physician.
- Postnum, number of posts one user initiate in the whole forum.
- Integral, integral that a user has.
- Essence, number of posts that evaluated essential by the forum administrators.
- Votenum, number of votes one user get.
- Date1, date of the earliest post.

- `Date2`, date of the latest post.

DATASET 2: `edges.csv`

- `from`, the sender in an interaction.
- `to`, the receiver in an interaction.
- `singletimes`, speaking times of a user in one post.

To make the visualization possible, we take the nodes with degrees ≥ 6 , which means in the common case the actor replies to others at least 5 times.¹ After removing the loops (including multiple loops) and reshaping the size of vertex according to the `totaltimes` variable, we get a graph in a proper size of 2663 vertexes and 30870 edges. The overview of the graph is plotted as Figure 2.

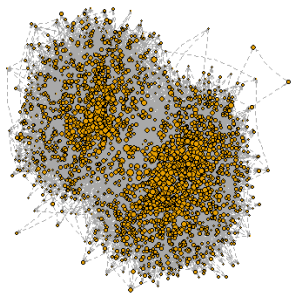


Figure 2: Overview of the dataset

3 Analysis of Individual Level

In this section, we focus on the analysis on individual level, describing their activities through seeing them as nodes, which carrying with labels about themselves.

Degrees, a symbol of the frequency that people ask or reply, depict a hierarchy of people who are active and who are not. Through degree analysis, we get a overview of the

interactions and the roles people take in, which we will discuss in detail in section 3.2. Then we measure the paths between the members, by which we could answer the question of how close the members are. Finally, we investigate whether the labels of the members affect their activities mentioned above, and how this relates to their reputation, discussed in section 3.4.

3.1 Degrees and Centrality

From the early history of social network analysis, researchers have attempted to describe basics of nodes and edges with degree. Degree centrality is one of the fundamental attributes of a node. In general speaking, there are three types of degree centrality. The term degree actually refer to the total degree, which is the sum of the in-degree and the out-degree. We call the number of edges one node send to other nodes the out-degree, and the number of the edges one node receive is called in-degree.

Degree, referring to total degree, ranges from 1 to 1311, most of which are less than 1 (the median and 3rd quantile are both 1), indicating that communications happen in people accounting for small percentage. In one way, this fact means these interactions have certain purposes, aimed at answering a question technically, or expressing the opinion about former texts, or showing gratitudes toward those who give the useful information. Furthermore, there will be many triads in the whole graph that collaborating into small components, indicating that many members search for or add an answer only if they do want to know or do know about the answer.

We made a scatter plot with path through the points, in order to show the degree distribution. See Figure 3 below. We could find that large amount of nodes have zero in-degree and 1 out-degree, who only replied once, therefore degrees distribution is quite unbalanced, where a small group of people lead the discussions more than others.

¹In the sample set, the actors often reply to themselves.

This also reveals that such a real network conforms to a power-law distribution (1) and the whole graph is a scale-free network[2].

$$P(k) \sim k^{-\gamma} \quad (1)$$

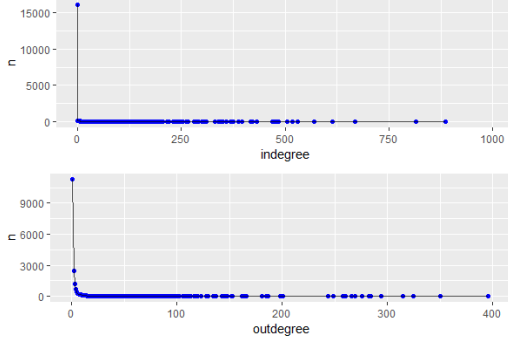


Figure 3: degree distribution

Apart from basic degree analysis, we also calculate other kinds of centrality with respect to the individual level, including closeness centrality, betweenness centrality, and eigenvector centrality. We list all the centralities in Table 1. We will come back to these properties in the next section when discussing the roles of members.

name	min	mean	max	sd
degree.cen	1.00	7.08	2796.00	49.45
close.cen	0.20	0.27	0.43	0.03
btw.cen	0.00	3976.85	8755147.38	91369.75
eigen.cen	0.00	0.00	1.00	0.01

Table 1: Table presents 4 kinds of centrality

3.2 Roles of the members

In this section we will focus on two basic activities in the forum, ask and reply, then present the characteristics of ask-leader and reply-leader. In addition, we discuss two roles about the position the members in, one is bridge, also

known as structure hole; the other is follower, whose total degree is low but also significant.

It is easy to understand that in most cases active members attract more attention, gain more reputation by replying a lot or asking a lot. We call people who ask a lot the "ask-leader", people who reply a lot the "reply-leader". It should be noticed that one replies a lot does not mean he always provides the right answer; sometimes, it triggers more discussion in detail, or shows the attitude, such as standing for the answer in upper floors.

We are interested in one question, that is do ask-leaders also reply a lot, and vice versa. In the picture below shows the correlation of in-degree and out-degree. The R value by the Pearson method is 0.61, revealing that an ask-leader is willing to be a reply-leader, but this trend is not quite strong. We then change the x-axis scale to \log_{10} (see Figure 4), which will make the graph more spreading, and thus easy to observe the trend. The R value turned down to 0.37 as many nodes that have zero in-degree changed into infinity, hence not involved in the calculation of correlation. We also draw a trend line in blue, which begins with a smooth way at first, climb up around 100 in-degree, and in the right part, the correlation seems much higher than before, and most of them are from the neurology department. Though the number of this class is quite small, they receive a lot of links from others, and reply a lot. We then discover that many of them not only lead the discussion, but take important roles in the graph, and maintain the operation of the forum in an institutional reason.

We would like to talk about bridges in the graph. Bridge, or structure hole, connects components and serves as a hub to transfer significant information. They tend to be potential sources of resources, and set up possible routes for those who haven't communicated yet. Another advantage for structure hole is that this position provides an opportunity for a kind of social "gate-keeping" — she regulates the ways of communication of both groups she belongs to.[9]

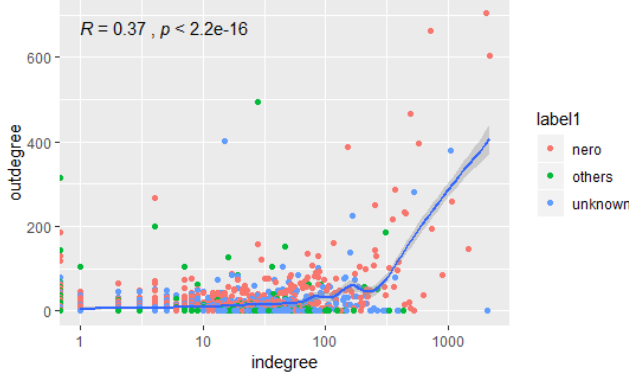


Figure 4: Correlation of in-degree and out-degree

In this analysis, we find that active members take in the role of bridge, and most of them become the administrators of the forum in volunteer. Respectively, they still show high ranks even we choose different methods to measure the centrality because they usually function as both hubs and authorities, and own large number of neighbors.

Like real world organization, online communities also need administration to maintain such a forum with many boards. This position needs competence and role identity, and increases the member's reputation and authority.[5] People volunteer to take over it not only because the power to delete/highlight some posts, but also for satisfaction reinforced in years of work.[11][13] Thus, to promote active members in the forum to serve as administrators is an operating strategy of DXY medical forum.

Last, we reckon followers who post a few may fall into several conditions. First, their question was quickly solved and thus they leave. Second, previous research figured that the first experience is significant for the following participation[15]. If one felt his question is ignored, they might leave the forum. Third, they are more active in other sub-forums rather than this one since we observe half of the 1-outdegree follower has high Postnum.

3.3 Distance and reachability

In this section, we present some basic characteristics of the graph, such as diameter, density, keeping the accuracy of 2 decimals place. This part put the emphasis on the attributes of the whole graph, rather than the attributes of the nodes.

The average density measures how the edges are allocated in the graph, telling us whether the graph is loose or dense. In this analysis, the average edge density is low as 1.94×10^{-4} , most attributed to those small components. Generally it's a loose graph where members do not interact frequently, which also accords to the degree centrality before.

The mean distance measures the average shortest path across any two nodes. In this directed graph it ends up with 3.75, which tells us each member is likely connected through three people. If we remove the direction, this number will go up to 3.77. Since there exists the possibility that communications might happen in the future in the opposite direction that hasn't yet occurred, though these two numbers are quite close, we reckon the number 3.77 is a more appropriate one indicating the average shortest path between any two members.

The diameter, marked with red nodes and orange edges, in Figure 5, implies the longest path of any two members. It depicts the longest path ever emerged in the graph, which often pass by the valuable clusters, thus give us priorities to some nodes that may take with more information. In this directed graph, the diameter is 11, size of nodes are reshaped based on the variable `totaltimes`. The nodes in the diameter appears among large components, some of them shadowed beneath the Grey nodes.

3.4 Do labels matter?

We collect some labels of the users before, examining whether their ranks, integrals, departments affect their

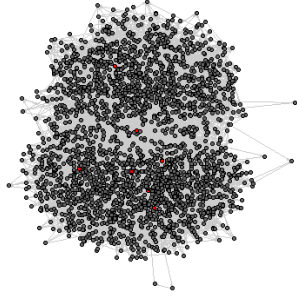


Figure 5: Diameter

behavior in the forum. First, we show the possible combinations of different labels in the Table 2 below.

	label1	label2	n
1	nero	identified doctor	323
2	nero	practicing doctor	3069
3	nero	specialist	381
4	others	identified doctor	235
5	others	practicing doctor	4227
6	others	specialist	245
7	unknown	identified doctor	105
8	unknown	medical student	1729
9	unknown	practicing doctor	345
10	unknown	unknown	7520

Table 2: 10 possible combinations of labels

About half of the nodes have department information, 20% of them are belonging to neurology, 26% belonging to others. The percentage of ranks are 48% for doctors, 10% students, and 42% unknowns. Then how will these labels affect the appearance of the graph? Based on labels and total speaking times, we make two plots concerning department and rank (Figure 6). We only select hundreds of nodes to make the image clear enough to show its characteristic. Therefore it's more like a diagram, rather than the accurate graph.

If we suppose influential discussions involve amounts of people to engage in, and an influential person speaks a lot

as to lead the talk, we could see that most influential discussions are lead by neurology doctors, second by unidentified members, other departments seldom lead the discussions. From the right picture, we find that both doctors and unknowns speak a lot, but they seem to separate with each other. We have to distinguish the two "unknowns" that have different meanings, one for department, one for rank. If someone with medical background want to join in the discussion when he someday occasionally find the website, he will quickly get himself enrolled without filling in his complete information. However, if he does fill in the form next day when he wants to see this post again, he's most likely to write down both department and rank. That's why we see combination 7 and 9 are a few hundred, while the unknown-unknown combination rush to about 7500.

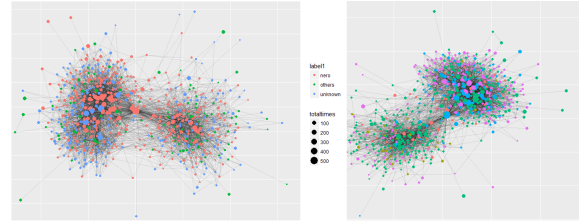


Figure 6: Left: Labeled with department Right: Labeled with rank.

Last, we check if the incentive rules affect members activity in posting. We test it with a OLS multiple regression, purposed at whether people tend to speak more if they know these virtual rewards. Members get integrals when they post, and their rank get promoted as the integrals increase. To balance, people with higher rank receive less integrals than new comers. Essence is a recognition of valuable posts, and Votenum accumulate when your answer/question is appraised by others. These information serve as special labels presented in the profile when people surfing the answers. But we don't find strong correla-

tion between them(the adjusted R squared is 0.13). This might because these labels also count posts in other sub-forums, and not strongly related to the neurology posts.

4 Analysis of Community Level

4.1 Community detection

Community detection should always be the core task of a social network analysis. We tag the dataset with two labels to divide them into several groups, as to see how they interact within and outside the group. First, we make the distinction due to the departments they belong to, the neurology department with label "nero", the other departments with label ""others", the unidentified with label "unknown". In addition, according to the professional ranks that recorded by the forum, we set five labels, mapping to different members, which are, "practicing doctor", "identified doctor", "specialist", "medical student", the others are labeled as "unidentified".

Using cluster fast greedy algorithm(FGA)[6] with two strategies, one with no weight, one with weight based on `singletimes` variable, that is, times one user speak per edge. We transform the graph into undirected, removing all the self-loops and merging the multiple edges. Then we divide the graph into different communities, showed in Table 3.

We select one image of the community division. The Figure 7 presents the condition whose members' degree is larger than 5, under which scenario about 2900 nodes are left. In this figure we could see that the whole graph is divided into two parts, and the communities overlap over other. This means they could be generally seen as counterparts. Also this result accords to the overview plot, see Figure 2 in section 2. In the next section, we interpret this result and present potential mechanisms that make the division happen.

degree \geq	no_weight	weight	remaining
1	24	22	18179
2	11	10	6757
3	15	5	4474
4	9	3	3472
5	6	2	2905
6	9	3	2538
7	6	3	2250
8	3	3	2045
9	7	2	1849
10	6	2	1693

Table 3: First column shows the minimum degree of the nodes we divide; column 2 and 3 show number of communities we get, weighted or not by the times one user speaks in a post. The last column records the members left after we do the cut-off in the first column.

4.2 Lifespan Analysis and Commercial impact

From the findings of community division, we could see that, for nodes that have at least 5 degrees, they obviously could be divided into two parts with weight. In this section, we clarify why does this happen, and how could it relate to the vigor of the forum.

We collect all the time tags that members in the anchor send messages. The total number of the time tags is 182646, time ranging from 2002-07-06 to 2019-04-25. Notice that since we care about what the members sent, so we do not restrict both the senders and receivers into anchor set. Figure 8 shows the number of posts along the time line. Actually, the path of the line plot is quite tuberculous, sharply moving up and down. The blue smooth line is the trend line, and the red line is the mean value of posts.

We take a closer look at the year scope, and see that around year 2005 and 2016 clustering many posts, more clear shown in the figure right side .

We select a group of people in the top-level to see whether this pattern still holds. The benchmark of "top-

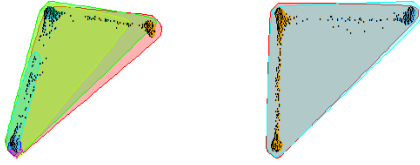


Figure 7: Left: Six communities when the parameter `Weight` is not used in FGA. Right: Two communities when `Weight` is based on edge attribute `singletimes`. We could see the two communities are counterparts for each other.

level” is people with degree more than 25, which accounts for about 4.5%. From Figure 9, we could find that these people follow almost the same path, comparing to the trend line.

We suppose that these two clusters explains why the graph could be divided into two communities. Lots of people posted in the forum around 2005 and 2016, and they enrich contents of discussion. Moreover, top-level members accounts for one-third of the posts, and their posting path is quite similar to the general trend line. Next, we give some assumptions of why this two-hump shape appears.

We investigate the first date and last date people spoke in the forum, and we believe that this two dates are the start and end of a user’s lifespan. The scatter plot (Figure 10) presents that many people registered in around 2005 and 2016, then quickly became quiet in the forum(dark region in the southwestern and northeastern edge). However, lifespan is a tricky word, though the area of 2018,2019 is dense, too, we cannot claim that the last post is the end of their lifespan because they may also be active in future, and the current observation cannot prove that’s the termination of their activities.

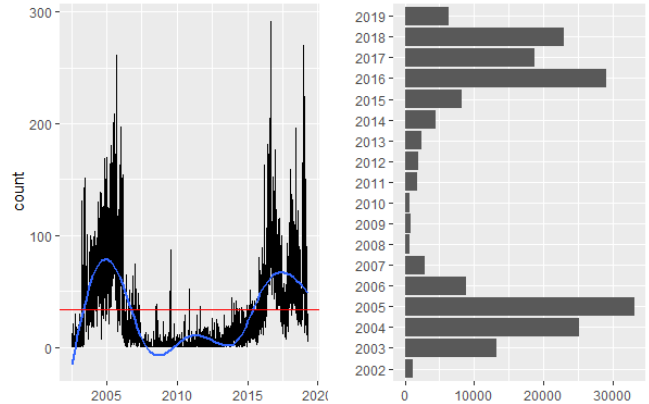


Figure 8: Posts from year 2002 to 2019

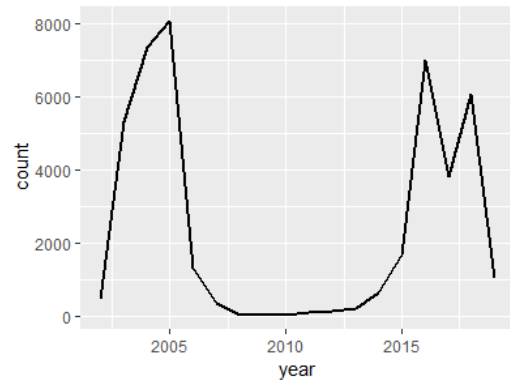


Figure 9: Posts of the top-level people

We then track lifespan of the top-level members mentioned before, and it concludes that these people show much adherence to the all-user pattern. Most of them keep active more than one year, and some of them keep active for decade, since the initiation of the forum. Another part of top-level members joined in the forum later than 2015, and they still keep active now. See figure 11. This fact indicates that these members participate in the discussions for a long term, rather than contribute with short-term curiosity and enthusiasm.

Finally, we could deduce from these two facts that key

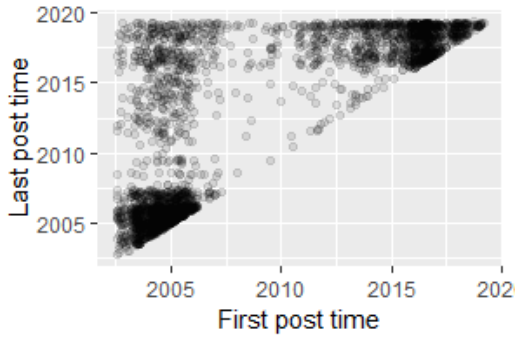


Figure 10: The lifespan of members

members with long lifespan lead the discussions, contributing to the healthy environment of the forum. When they are active, the forum turns to be prosperous; when they keep silent, the forum fall into the trough.

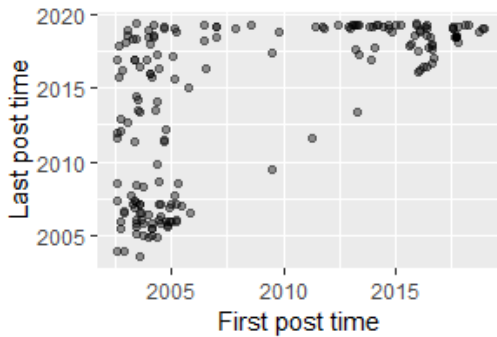


Figure 11: The lifespan of top-level members

In the end of this section, we discuss what factors may cause the phenomenon that many people registered around 2005 and 2016.

It is easy to make some implications about year 2005. After we surfed some posts retrospecting the "Golden Era" of 2005-2006, we found that, in 2005, DXY medical forum was introduced to many doctors because its focus on academic after three years of effort, and its founder as a doctor, received much recognition by his colleges all over the country.

The year 2016 attracts us more. We have once guessed whether it is relevant to the commercial activities of DXY forum, but after querying the financing date of the company, it cannot accord to the finding we get. DXY forum started its financing since 2010, and then 2012, 2014 and 2019. While we see that posts from 2010 to 2014 are all low-posts-years, though there may be a time lag, and people started to enter after the financing, it couldn't make much sense.

Then we look over the posting rules and ads institution of DXY. We find that DXY has strictly rules on ads, administrators actually block out the ads in the boards, and there's no chance for sellers to come in to promote their drugs, equipments, etc. The rules were made by members themselves, not by an official authority, hence each one could suggest on them. This gives the forum large rights of autonomy[12], and the commercial impacts little on the academic air of DXY.

On the other hand, on the index page we could find hyper-links to other products of DXY in the margin of the navigation bar, and other products like DX-JobSeeking, DX-DrugHelper also have links to DXY medical forum. We assume that this bring more people engaging in the discussions.

To summarize, DXY as a vigorous forum have a group of people who join in the discussions in a long cycle, and many of them lead the discussions for their expertise in the field. Moreover, the academic air was not influenced by the commercial activities, due to the institution that blocks out ads and outwards links. Last, other product of DXY company may bring more people in the discussions.

5 Motif Analysis

Motif analysis allows one to discover small local patterns of interaction that are indicative of particular social dynamics. It connects the analysis between individual level

and community level, aimed at telling some truth from the middle layer of the graph.

Triad census was defined by David and Leinhardt[7], focus on all sixteen possible directed stages among three connected users within a forum. All possible sixteen triad patterns are gathered in figure 12.

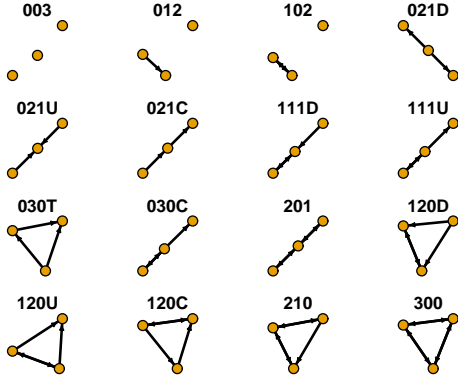


Figure 12: 16 stages of triads

To make the narrative simple, we call the patterns 1-16, from left to right, top to down. We first remove the multiple edges to reduce amount of calculation, then standardize the number of motifs detected, using the formula (2), and find many of the triads have high z-scores, which means subgraph N_i is a network triad of graph. Actually, social network analysis cares about triads that reveals transitivity, which indicates that weak ties may emerge in the future if a transitivity pattern is confirmed. The typical pattern of transitivity is pattern 6, where $A \rightarrow B \rightarrow C$, see in Figure 12.

Pattern 5 is over-presented, the number reaching to 847. This pattern is named as "in-star", where B,C points to A, indicating leaders receive many links from others and earn high in-degree, the opposite pattern of which is pattern 4, negative in the plot.

$$Z_i = (N_i^{real} - \overline{N_i^{random}}) / std(N_i^{random}) \quad (2)$$

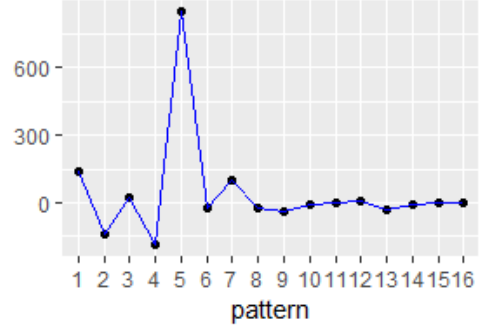


Figure 13: Z-scores fluctuate around zero, pattern 5 seems over-presented

Another interesting fact is that triads that all three nodes interact with others is low-represented, almost the same as the random conditions. But in the community visualization, we could find that top-level group interact strongly with others, it covered by the standardization because of the average.

We find 1204 pattern sixteen triads, which represents full-complete triad. This tells us nodes A, B, and C have established very strong connections, every edge between them is a strong tie. Still, some patterns that are also present infinity, like pattern 13-15, are weaker versions of pattern sixteen.

We suppose the following scenarios may help us understand why it is possible to have these triads appear with high probability comparing to random graphs.

We think small groups of discussion are likely to tentatively built up especially when the administrator also have quite good knowledge of the field. When dealing with Q&A questions, people priorly answer those easy ones, and leave the hards ones for skilled experts. As we mentioned, skilled members volunteer in administration of the forum, when he supervises the posts, he would detect those not answered and then it's highly possible for him to join as a "replier", rather than the role of administrator. Then others may be attracted to follow in this post, which promotes

the triads to emerge, in both weak and strong way.

In the second scenario, series discussion would gather many members who are attracted by the contents or the popularity of this series. For followers who constantly attend in the discussion, they're apt to know and communicate with each other because they regard this activity as a daily routine, thus trust among these members shall be larger than the average.

6 Conclusion

This study represents one of first efforts to explore the interaction pattern in medical online communities of China, achieving some interesting findings about the relationship between the vigor of the forum and roles members taking in three levels of analysis. These three parts correlate with each other, sharing some facts about the activities and their reasons, together describing the knowledge sharing behavior in the forum.

In the individual level, degree distribution presents highly unbalanced, in line with the power law. By analyzing ask-reply activities and labels, we conclude that in the early days people aren't aware of completing their label info; neurology doctors are the main participants in popular posts. Virtual awarding like Integrals don't affect the times people speak.

In the community level, we discuss the reasons of two biggest components emerge, and linked it to lifespan analysis where we stretch out to factors that may influence on enrollment of new members. We find that the leaders matter in discussion as they take over two roles: both participants and administrators, thus lead the vigor of the whole forum. Last we make some suppositions about the boom of registrations in two special years, and find that academic air in the forum hasn't changed by commercial activities.

In the motif analysis, we confirm the previous findings from a middle-layer level, and build two scenarios to show

how could the transitivity happen.

This research has its limitations as we only study on one sub-forum. Looking forward, we plan to concentrate on the registration strategies of DXY, and seek for more reasonable mechanisms that may contribute to the boom of enrollment in certain years.

References

- [1] Lada A Adamic et al. "Knowledge sharing and yahoo answers: everyone knows something". In: *Proceedings of the 17th international conference on World Wide Web*. ACM. 2008, pp. 665–674.
- [2] Albert-László Barabási and Réka Albert. "Emergence of scaling in random networks". In: *science* 286.5439 (1999), pp. 509–512.
- [3] Mohamed Bouguessa, Benoît Dumoulin, and Shengrui Wang. "Identifying authoritative actors in question-answering forums: the case of yahoo! answers". In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2008, pp. 866–874.
- [4] Cody Buntain and Jennifer Golbeck. "Identifying social roles in reddit using network structure". In: *Proceedings of the 23rd international conference on world wide web*. ACM. 2014, pp. 615–620.
- [5] Brian Butler et al. "Community effort in online groups: Who does the work and why". In: *Leadership at a distance: Research in technologically supported work* 1 (2002), pp. 171–194.
- [6] Aaron Clauset, Mark EJ Newman, and Christopher Moore. "Finding community structure in very large networks". In: *Physical review E* 70.6 (2004), p. 066111.

- [7] James A Davis and Samuel Leinhardt. "The structure of positive interpersonal relations in small groups." In: (1967).
- [8] Crystal Dieleman and Edward AS Duncan. "Investigating the purpose of an online discussion group for health professionals: a case example from forensic occupational therapy". In: *BMC health services research* 13.1 (2013), p. 253.
- [9] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets*. Vol. 8. Cambridge university press Cambridge, 2010.
- [10] Samer Faraj, Sirkka L Jarvenpaa, and Ann Majchrzak. "Knowledge collaboration in online communities". In: *Organization science* 22.5 (2011), pp. 1224–1239.
- [11] Xiao-Ling Jin, Matthew KO Lee, and Christy MK Cheung. "Predicting continuance in online communities: model development and empirical test". In: *Behaviour & Information Technology* 29.4 (2010), pp. 383–394.
- [12] Diane Maloney-Krichmar and Jenny Preece. "A multilevel analysis of sociability, usability, and community dynamics in an online health community". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 12.2 (2005), pp. 201–232.
- [13] Yuanyuan Shi, Xitong Yue, and Jin He. "Understanding social network sites (SNSs) preferences: personality, motivation, and happiness matters". In: *International Conference on Online Communities and Social Computing*. Springer. 2013, pp. 94–103.
- [14] Alan Tidwell. "The virtual agora: Online ethical dialogues and Professional communities". In: *First Monday* 4.7 (1999).
- [15] Zhijun Yan et al. "Knowledge sharing in online health communities: A social exchange theory perspective". In: *Information & Management* 53.5 (2016), pp. 643–653.
- [16] Z.Lv. "DXY:From academic forum to commercial website(trans)". In: *Entrepreneur* 009 (2014), pp. 24–24.