


| | | |
|--|---|------------------------------------|
| ITC 2/48 Journal of Information Technology and Control Vol. 48 / No. 2 / 2019 pp. 195-210 DOI 10.5755/j01.itc.48.2.20028 | Frequent Subgraph Mining Based Collaboration Pattern Analysis for Wikipedia | |
| | Received 2018/01/25 | Accepted after revision 2019/04/10 |
| |  http://dx.doi.org/10.5755/j01.itc.48.2.20028 | |

Frequent Subgraph Mining Based Collaboration Pattern Analysis for Wikipedia

Zheng Hu

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, e-mail: huzheng@bupt.edu.cn

Zhonghu Zuo

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, e-mail: zuozhonghu@bupt.edu.cn

Chunhong Zhang

Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China, e-mail: zhangch@bupt.edu.cn

Xiaosheng Tang

Laboratory of Universal Wireless Communications (Beijing University of Posts and Telecommunications), Ministry of Education, Beijing 100876, China, e-mail: txs@bupt.edu.cn

Yuqian Tang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, e-mail: yqtang@bupt.edu.cn

Corresponding author: huzheng@bupt.edu.cn

Online knowledge collaborations, where distributed members without hierarchies self-organize themselves to create valuable contents, are prevalent in many open production systems such as Wikipedia, GitHub and social networks. While many existing studies from network science have been brought to analyze the general interactive behavior patterns embedded in these systems, how the collaborations influence the achievement outcomes has not been thoroughly investigated. In this paper, we mine the collaboration patterns from a micro perspective to deeply understand the relationships between the collaboration among participants and the qualities of the Wikipedia articles. In particular, the subgraphs contained in the collaboration networks derived from the

Wikipedia revision histories are taken as the fundamental units to analyze the collaboration diversities from the subgraph properties such as size and topology. In contrast to the predefined static motifs adopted by the previous works, the collaboration subgraphs are directly found from Wikipedia dataset by a frequent subgraph mining algorithm GRAMI, which is able to capture the real dynamic collaboration patterns. Moreover, the relationships between the co-authors in the subgraphs are also discriminated to further explore the collaboration patterns. The experiments exhibit the statistical properties of the collaboration subgraphs and the efficiency of them as the metrics for the article quality assessments. We conclude that a small group of editors with relative frequent fixed collaboration patterns contribute more to the excellent article quality than the professional extents of arbitrary individuals in the collaboration group. This discovery confirms the common insight about collaboration that many heads are always better than one and concretely suggests a potential explanation for the increasing prevalence and success of the online knowledge collaborations.

KEYWORDS: collaboration, data mining, Wikipedia, article quality, social network analysis.

1. Introduction

With the widely spreading of the Internet, on-line collaboration has increasingly become a prevailing strategy by which information, contents and knowledge are created by a group of flat-organized people. The collaboration is inherently a self-organizing activity where arbitrary participant is allowed to contribute her or his specific efforts and interact with any others along the evolution process of the collaboration [3, 31]. One of the primary interesting questions arising from the web collaboration is the correlations between the variance of collaboration patterns and the expected targets of the collaboration activity. For instance, whether an expert is necessary for a successful collaboration or how the value of the aggregated knowledge produced by a crowd of heterogeneous volunteers is influenced by the collaboration interactions among them?

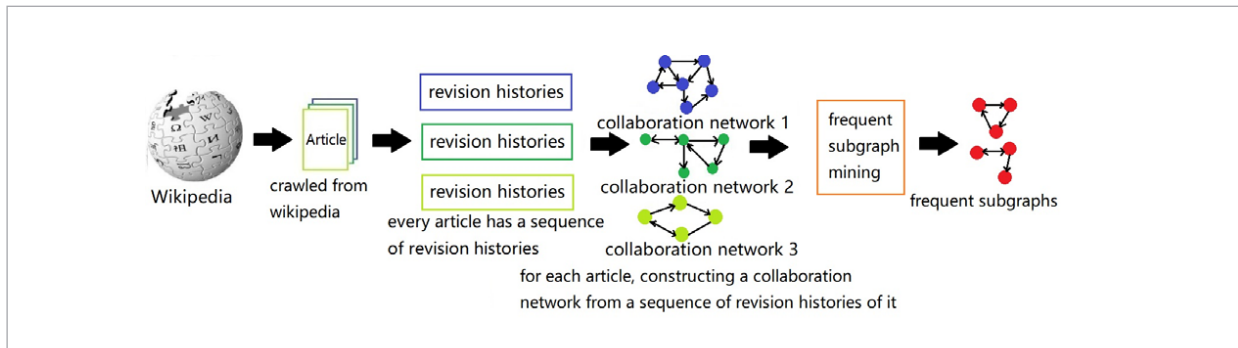
To investigate the above problems, some research efforts have been carried out [4, 19, 23, 34]. Like many of them, we also take Wikipedia as the targeting open on-line collaboration system to study these problems [8, 32]. In Wikipedia, editors who may be not familiar with each other can collaboratively edit article arbitrarily. An article could be edited by thousands of editors, with its edition duration spanning even ten years. Hence Wikipedia is deemed to be a suitable subjective for researching on-line collaboration. Furthermore, the editors' engagement and collaboration vary dramatically owing to their purpose, context and prior knowledge or other relevant attributes [7, 28, 32]. Taking the quality of user generated article in Wikipedia as the metric to evaluate the collaboration effects, researchers have found that the articles

in Wikipedia are rather diverse in terms of the quality and reliability [5, 7, 18, 21, 23, 30]. Moreover, the collaboration pattern in Wikipedia has been regarded as one of dominators leading to Wikipedia's diverse achievements and there has been some research work addressing Wikipedia collaboration patterns recently [23, 32]. Wu et al. [34] constructed edit network to represent collaboration activities of editing a Wikipedia article and then retrieved motifs from the edit network. Based on network motif counts and ratio, qualities of articles in Wikipedia could be inferred. Robertie et al. [23] utilized co-edit graph which could model collaboration among editors to measure Wikipedia article quality. Because motifs have the inherent flaw that they are predefined and co-edit graphs are unable to reflect the effect that collaboration patterns exert on a specific article, there still exists potential room to further explore collaboration patterns.

Actually, constructing collaboration networks and then mining collaborative patterns from them is also a view on researching the collaboration and this view can also make up for the flaws of motifs and the co-edit graph. The collaboration patterns are dynamically discovered from collaboration networks and these collaboration networks are representative for the reason that they frequently appear in collaboration networks. Furthermore, every article in Wikipedia has a corresponding collaboration network. Hence the influence of collaboration on any article can be easily researched through corresponding collaboration network. In this paper, each collaboration subgraph mined from collaboration networks is representative of a collaboration pattern that constantly appears in

Figure 1

The approach overview of our method



the evolution process that contents are collaboratively created. And these collaboration patterns are likely to determine the qualities of contents. However, researches on this aspect are not enough so far. Therefore, motivated by current insufficient researches on it, we conducted a series of experiments on Wikipedia to attempt to research collaboration and unveil the influence of collaboration on the qualities of collaborative outcome.

Hence, the skeleton of our methodology is as follows. For an article in Wikipedia has a sequence of revision histories [30], we can construct a collaboration network for the article from these revision histories. The collaboration network can capture structure and temporal dynamics which will emerge once one editor modifies an article which has been edited by a previous editor. In other words, the collaboration network can record various interactions among editors who have edited the same article. Because the collaboration network can capture substantial complex interactions and dependencies among editors, it can model that an article was collaboratively revised by many authors. Utilizing the collaboration network, a structural approach, we can characterize collaboration patterns among editors in an effective way and insight to the global collaboration among all editors at a macro level. Furthermore, collaboration subgraphs mined from collaboration networks can reflect the feature of collaboration networks at a micro level and are representative of various collaboration patterns that appear frequently.

In short, the main process of our work can be sketched as follows. As can be seen in Figure 1, in the first step, articles are crawled from Wikipedia and a sequence

of corresponding revision histories are also crawled at the same time. In the second step, we utilize a sequence of revision histories of an article to construct a collaboration network. In the final step, we apply frequent subgraph mining algorithm to discover collaboration subgraphs which appear frequently from these collaboration networks. Through analyzing collaboration subgraphs, we discovered some interesting experimental phenomena. The collaboration patterns of which the topologies are stars usually produce high quality articles. The influence of a successful collaboration pattern on article quality is more remarkable than that of anyone in a collaboration pattern. Furthermore, some editors constantly cooperate with each other in a fixed collaboration pattern to produce high quality articles. Finally, we also performed some classification tasks to testify the influence of collaboration patterns on the qualities of collaborative outcome. We found that the results of the classification task were good.

Hence, the view that we construct collaboration networks and mine collaboration subgraphs from them is worthwhile. The view has been verified on Wikipedia by our work and it also has some reference value to other on-line collaboration systems.

In a word, we mainly apply the knowledge of graph theory [26] and the approach of frequent subgraphs mining to research self-organization collaboration patterns among editors. Our contributions are as follows.

- 1 Applying frequent subgraph mining to research on-line self-organization collaboration.
- 2 Making up for the flaws of motifs.

- 3 Discovering various meaningful collaboration patterns.
- 4 Proposing several metrics related to the outcome of collaboration patterns.

The remainder of this article unfolds as follows. In the next section, we will introduce some researches related with our work. In Sect.3, we will explain how the collaboration network is constructed in detail. In Sect.4, we will introduce several algorithms we utilize in experiments and some related basic concepts. In Sect.5, we will exhibit experimental results and corresponding analysis. In the final section, we will discuss the influence of collaboration on the outcome of collaboration patterns and conclude our work with a summary.

2. Related Work

The graph based approach has been long explored to represent the basic structures and properties of the collaboration behaviours in Wikipedia, which facilitates the understandings of the collaborative activities from the view point of interdependent relations among the participants. Korfiatis et al. [16] proposed a contributors network, where a vertex represents an editor and an edge represents the sequence of their interactions, to evaluate the contribution importance of the collaborative editors working on the same articles. While the definition of the vertex keeps the same with ours, the edge of our collaboration graph represents the collaborative event between the editors rather than the sequences. Similar to the contributors-network mentioned above, Laniado and Tasso [17] constructed a co-authorship network where nodes represent editors and edges represent co-authorship between editors. While inexperienced editors whose contributions are too few will not appear in the co-author network, our work takes all editors into consideration. We construct a network for every article from revision histories of the articles. Our network, therefore, can reflect editors' interactions on every specific article.

Except the form of network mentioned above, trajectory network is also widely used to research collaboration [15]. Iba et al. [12] constructed trajectory network for an article in Wikipedia, which can describe the development of the article. They focused on finding

different editor roles, namely egoboosters and coolfarmers, according to editing pattern presented in the network. In contrast, our work pays closer attention to various collaboration patterns among several editors rather than some featured editor. Furthermore, the editor-editor network proposed by Flöck and Acosta [11] is also a trajectory network, which is constructed dynamically over time. Through mining interactions from editor-editor network, interactions among editors can be visualized as time has passed. Their work gives us a transparent and intuitive understanding of the dynamic collaboration among editors. While their work only takes the disagreement interactions among the editors into consideration, our work mainly focuses on how the editors cooperating with each other can produce a good outcome.

Being different from all networks above, the networks of which nodes represent editors or articles and edges represent editing relations are also utilized by many researchers to research collaboration. For example, Li et al. [19] exploited article editor networks to assess qualities of articles in Wikipedia. They proposed three models that could be used to calculate the value of nodes in article-editor networks. Because they incorporated manual evaluation results into the third model, their experimental results may be susceptible to the number of featured articles in their datasets.

Furthermore, based on various networks, applying the analysis of motifs to study the properties of networks is a good way to research collaboration [14]. Wu and Cunningham [33] integrated three network views to predict the qualities of articles, among which are the temporal network that considers article revision histories as sequence of editor's interactions, ego network which is a static bipartite network with edges representing edits performed by editors to articles and trajectory network which is representative of revisions of an article through a sequence of editors. Then they conducted a binary classification task, based on motifs discovered in the three networks, to assess the effectiveness of their method. The three different network views capture three different aspects of interactions among authors. The performance predicting the qualities of articles, therefore, is expectedly improved by the integration of them. There is also a fly in the ointment because motifs are predefined and consequently not as flexible as frequent subgraphs. Furthermore, it remains to be seen how accurate the

results of their classification task will be when they classify articles into six classes.

There also exist non-graph-based studies. For instance, Adler and De Alfaro [1] proposed a content-driven reputation system for Wikipedia editors. In their reputation system, if the edits that one editor performed to Wikipedia articles are preserved by subsequent editors, then the editor will gain reputation otherwise reputations of the editor will be reduced. Hence, utilizing the characteristic that editors can influence and control each other, this system can apparently encourage editors to collaboratively write high-quality articles. Furthermore, one of prominent functions in the system is to predict the quality of a new article that has not yet been edited by others according to the reputation of its initial editor. In their [2] later work, after comparing seven methods to measure editor's contributions, they found that edit longevity outperforms the six methods remaining. Dalip et al. [9] extracted 69 quality indicators from three aspects, namely textual feature, review feature and network feature. Based on a machine learning algorithm, they utilized these 69 indicators to assess the qualities of articles. Lipka and Stein [20] represented an article as an n-dimensional vector of which each dimension is representative of a kind of trigram which is originally applied for writing style analysis. Then they utilized the n-dimensional vector, based on Support Vector Machine and Naive Bayes, to determine whether an article is a featured article or not. As can be seen, they all researched article quality from the perspective of article content. Indeed, the content of an article is directly related to the quality of it. Nevertheless, collaboration is also an important factor that cannot be ignored. If collaboration among editors can be also taken into consideration, their experimental results may be better.

3. Collaboration Network

In this section, we will elaborate how to construct a collaboration network from revision histories of an article. Jurgens and Lu [13] proposed a network perspective that articles can be viewed as a sequence of revisions ordered by the edits times. A sequence of revisions is fundamental elements for constructing our collaboration network. In our datasets, every revision history has been attached some edit information including not only title, revision id, parent revision id and minor but also editor name, editor id, timestamps and size. The information can be seen in the instance presented in Figure 2. Among this edit information, minor is a strange concept. In brief, a minor edit signifies that there are only superficial differences, such as typographical corrections, between the current versions and the previous versions. Conversely, a major edit should be reviewed for its acceptability to all concerned editors. Any changes affecting the meaning of an article are not a minor edit. For the sake of simplicity, we denote minor edit as 1 and major edit as 0.

Utilizing revision histories of an article mentioned above, we can construct a collaboration network for the article. For an article, firstly we establish a node representing the editor who created the first revision of the article, and label the node with a unique number. Secondly, we establish the next node representing the editor who has edited the article and then produced the second revision, and we also label this node with another unique number. Thirdly, we establish a directed edge pointing from the first node to the second one with the edge being labelled with 0 or 1 according to the edit type of this edit performed by the second editor. For the revision histories remaining, continue in this way. Then a collaboration network can be constructed completely. It is noteworthy

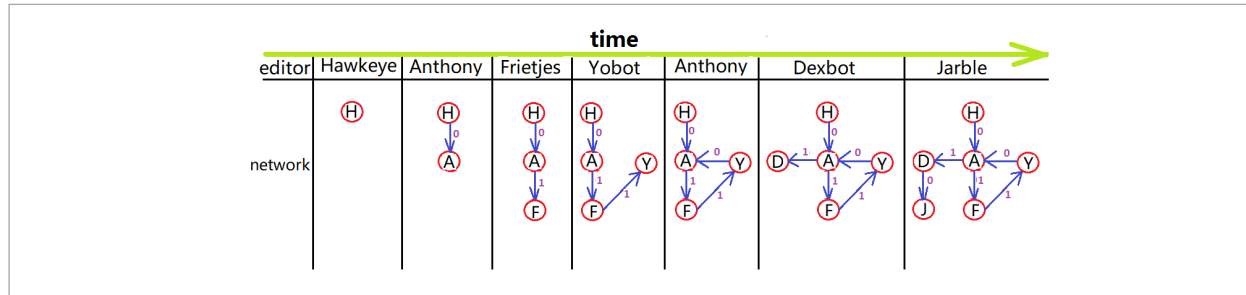
Figure 2

Edit information of seven revision histories of an article

| title | revision id | parent revision id | minor | editor name | editor id | timestamp | size |
|-----------------------|-------------|--------------------|-------|-------------|-----------|----------------------|--------|
| J. Robert Oppenheimer | 687560830 | 687558683 | 0 | Jarble | 302228 | 2015-10-26T10:15:25Z | 111661 |
| J. Robert Oppenheimer | 687558683 | 687558682 | 1 | Dexbot | 422129 | 2015-10-25T08:20:23Z | 111640 |
| J. Robert Oppenheimer | 687558682 | 687558653 | 0 | Anthony | 653217 | 2015-10-24T11:18:30Z | 111642 |
| J. Robert Oppenheimer | 687558653 | 687558666 | 1 | Yobot | 302121 | 2015-10-24T12:50:25Z | 111670 |
| J. Robert Oppenheimer | 687558666 | 687558654 | 1 | Frietjes | 422129 | 2015-10-23T10:25:45Z | 111651 |
| J. Robert Oppenheimer | 687558654 | 687558672 | 0 | Anthony | 712309 | 2015-10-21T14:46:40Z | 111631 |
| J. Robert Oppenheimer | 687558672 | 687557634 | 0 | Hawkeye | 302229 | 2015-10-18T17:55:42Z | 111672 |

Figure 3

A simple example which demonstrates the process of constructing a collaboration network from several revision histories of an article. Each node label is the initial of corresponding editor's name



that when constructing a collaboration network for an article, self-loop edges are neglected to avoid the case that an article was continuously edited multiple times by the same editors. Figure 3 demonstrates the process of how the collaboration network is constructed from the information in Figure 2.

As is shown in Figure 3, Hawkeye created a new article. This is the first revision of the article. Then Anthony edited this article. The second revision consequently comes into being. For these two revisions, we establish two nodes for them which are just the node H and node A in Figure 3. Owing to the edit performed by Anthony is a major edit, the directed edge pointing from node H to node A is labelled with 0. In the next time, Frietjes edits the revision of Anthony and then the directed edge pointing from node A to node F is labelled with 1 for the edit performed by Frietjes to the revision of Anthony is a minor edit. Following Frietjes's editing to the revision of Anthony, Yobot edited the revision of Frietjes. The label of the edge is 1 for the same reason. Then Anthony returned to edit the revision of Yobot with the edge being labelled with 0. The process for Dexbot and Jarble is similar to previous editors and we will not give unnecessary details here.

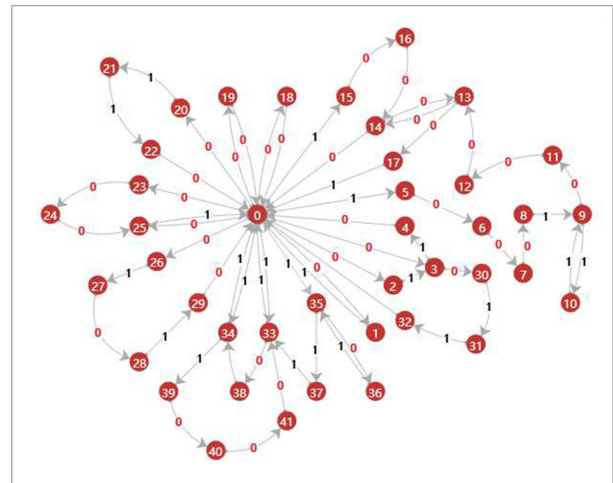
It is noteworthy that the collaboration network does not contain complete time information. Although a sequence of revision histories is ordered chronologically, our work mainly focuses on whether an editor interacts with another editor or not. We do not care the specific chronological sequence. In other words, it is unnecessary to know exactly which editor edited the article before which one.

For every article in our datasets, a collaboration network can be constructed according to the process

mentioned above. Finally, we select a collaboration network from 3916 collaboration networks as an example, which is presented in Figure 4.

Figure 4

A complete collaboration network of a "FA" article (Each label of a node is the id of the corresponding editor)



4. Algorithms

In this section, we will introduce two algorithms we utilize to accomplish our work smoothly and some basic concepts about frequent subgraphs mining. Assume that we have n articles denoted as a_1, a_2, \dots, a_n . for $1 \leq i \leq n$, let $e_{i1}, e_{i2}, \dots, e_{im}$ be the editors of corresponding revision histories.

For each article, due to a sequence of revision histories are arranged chronologically in our datasets, the

algorithm `constructCollaborationNetwork` naively iterates over a sequence of revision histories in order to establish nodes different from each other and edges pointing from one node constructed from the previous revision history to a node constructed from the current one. Through the iteration process above, collaboration networks for every article can be constructed completely. It is noteworthy that in this algorithm there probably exist multiple edges appearing between two different nodes. Multiple edges are also important, which may represent disputes between two authors or closed collaboration.

Finally, we will give a brief introduction to some basic concepts about frequent subgraphs mining and the algorithm GRAMI by taking the specific problems in our work for example.

Algorithm 1: `constructCollaborationNetwork`

Input: revision histories of articles

Output: a single large collaboration network

```

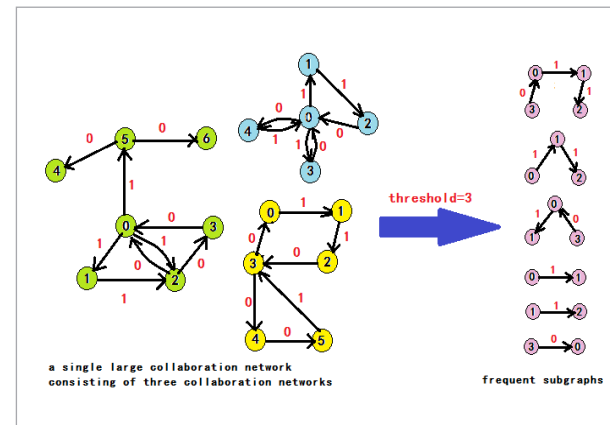
1 editors  $\leftarrow \phi$ 
2 singleLargeGraph  $\leftarrow \phi$ 
3 for each  $a_i$  in articles do
4   previousNode  $\leftarrow$  null
5   currentNode  $\leftarrow$  null
6   nodes  $\leftarrow \phi$ 
7   edges  $\leftarrow \phi$ 
8   for each  $v_{ij}$  in revision histories of  $a_i$  do
9     if  $e_{ij}$  not in editors then
10      editors  $\leftarrow e_{ij} \cup$  editors
11    end
12    currentNode  $\leftarrow$  construct a new Node
13    if currentNode not in nodes then
14      nodes  $\leftarrow$  currentNode  $\cup$  nodes
15    end
16    else
17      currentNode  $\leftarrow$  get currentNode in nodes
18    end
19    if previousNode  $\neq$  null and currentNode  $\neq$  previousNode then
20      edges  $\leftarrow$  construct a new Edge  $\cup$  edges
21    end
22  end
23  singleLargeGraph  $\leftarrow$  singleLargeGraph  $\cup$  construct new graph
24 end
25 return singleLargeGraph

```

In short, frequent subgraph mining is the process of discovering in graph database or in a single large graph all subgraphs whose frequencies of occurrences are no less than a specific threshold previously set by users. The frequency of occurrences is called support and denoted as *supp*. As is shown in Figure 5, there are three collaboration networks constructed

Figure 5

A simple example of frequent subgraph mining



from revision histories of three articles. Six frequent subgraphs can be mined when the threshold equals 3.

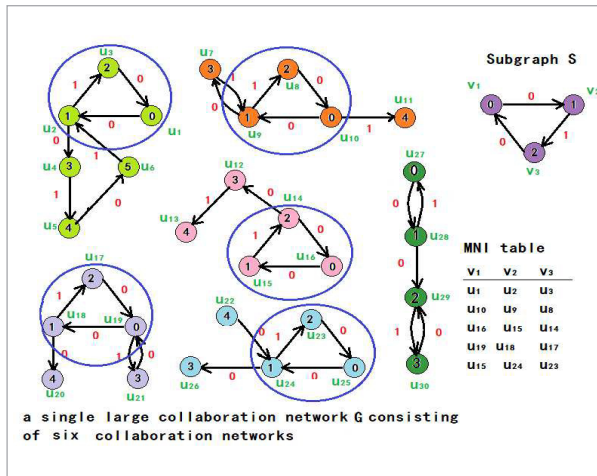
Currently, frequent subgraph mining technologies are rather mature and there are many algorithms for frequent subgraph mining such as CloseGraph, FSG, gSpan, GRAMI [10, 22]. Different frequent subgraph mining algorithms are applicable for different types of graphs and also different in performance. Taking above factors into consideration synthetically, we finally selected the last one. Furthermore, the algorithm GRAMI also possesses some merits mentioned below.

The algorithm GRAMI only finds the minimum set of instances of a candidate subgraph. That is to say, the algorithm will terminate the progress of discovering instances of a candidate subgraph as soon as the number of instances of the candidate subgraph is no less than the support threshold. Hence, this feature of the algorithm can improve the performance of the algorithm. To handle single large graphs, the algorithm adopts two optimizations which improve the performance of the algorithm significantly. The first optimization is to evaluate expensive nodes after evaluating lightweight ones. The second optimization is to prune the search space by utilizing the previous subgraph evaluations and the features of graph structure.

How to judge whether a subgraph is frequent is a basic problem for all frequent subgraph mining algorithms. The algorithm GRAMI adopts the support metric MNI which is an antimotone support metric [6]. For example, in Figure 6 there are five valid solu-

Figure 6

A single graph G and its subgraph S



highlighted with circles and a corresponding MNI table calculated according to these solutions. A valid solution is a mapping from S to a subgraph of G according to labels of nodes and edges. For all columns in the MNI table, the minimum size of all the columns is 5. So the support of S is 5. If the threshold is no more than 5, then S is a frequent subgraph of G.

5. Experiment

5.1. Datasets

According to Wikipedia article quality scale, articles in Wikipedia can be classified to seven classes as follows:

FA: these articles are the best and they are outstanding, thorough and professional.

A: these articles are pretty complete but to be promoted “FA”, some problems still need solving.

GA: these articles have no obvious problems and almost every user thinks that these articles are useful.

B: these articles, in terms of content, are complete and have no major problems.

C: these articles are substantial while they miss the content of importance yet or contain much irrelevant material.

Start: these articles are incomplete and need to be perfected further.

Stub: these articles just describe some basic content of a topic.

Hence the ranking of these seven classes of articles, in terms of quality, is as follows:

$$FA \succ A \succ GA \succ B \succ C \succ Start \succ Stub.$$

The datasets we select for our experiments consist of 3916 articles which were collected on February 20, 2016. These articles have totally 2364520 revision histories. Table 1 shows the article distribution in our datasets. By the way, the reason why “A” class articles are not included in our datasets is that these articles are too few in Wikipedia and we grouped “A” class articles with “GA” class articles. From Table 1, we can catch sight of that the quantity of high quality articles, generally speaking, is less than the quantity of less excellent articles. In fact, it is similar to the articles distribution in Wikipedia as well. Anyone can get Wikipedia articles by aid of MediaWiki API help and API sandbox.

Table 1

The article distribution in our datasets

| Class | FA | GA | B | C | Start | Stub |
|-----------|-----|-----|-----|-----|-------|------|
| #articles | 316 | 437 | 611 | 715 | 1156 | 681 |

For the sake of simplicity, we divided the original datasets into six parts according to Table 1 and then constructed collaboration networks respectively. Finally, we mined collaboration subgraphs from these six types of collaboration networks separately. The sizes of these six types of collaboration networks are displayed in Table 2. It is noteworthy that thresholds set to mine collaboration subgraphs from six types of collaboration networks are in proportion to the number of corresponding class articles in our datasets. Therefore, according to Table 1, the thresholds denoted as th for six types of collaboration networks satisfy the equation below:

$$\frac{th_{FA}}{316} = \frac{th_{GA}}{437} = \frac{th_B}{611} = \frac{th_C}{715} = \frac{th_{Start}}{1156} = \frac{th_{Stub}}{681}.$$

For example, if the threshold of “FA” collaboration networks is 4, then the thresholds for the other five types of collaboration networks are 6, 8, 10, 15 and 9, respectively. For the sake of simplicity, without spe-

cific explanations, we only give the threshold of “FA” collaboration networks later and the thresholds for the other five types of collaboration networks can be easily got by this equation.

Table 2

The sizes of six types of collaboration networks

| Class | FA | GA | B | C | Start | Stub |
|--------|--------|--------|--------|--------|--------|-------|
| #nodes | 173217 | 164317 | 299280 | 243191 | 97102 | 17876 |
| #edges | 310486 | 290131 | 494794 | 368404 | 128772 | 20829 |

5.2. Experimental Results and Analysis

5.2.1. Analysing Collaboration Patterns among Editors

As can be observed from Figure 7, collaboration subgraphs of “FA” collaboration networks, in terms of size, are more diverse than those of the other five types of collaboration networks. Furthermore, the sizes 2, 3, 4, 5 or 6 are common. To be more specific, the collaboration patterns that 2 to 6 people coauthor articles in Wikipedia are very common. Moreover, compared to the other five classes of articles, frequent collaboration patterns in “FA” articles are more diverse and the scale of these frequent collaboration patterns is larger.

Figure 7

The distribution of collaboration subgraphs size. The threshold is 4. The quantity of nodes in a collaboration subgraph is the size of it here

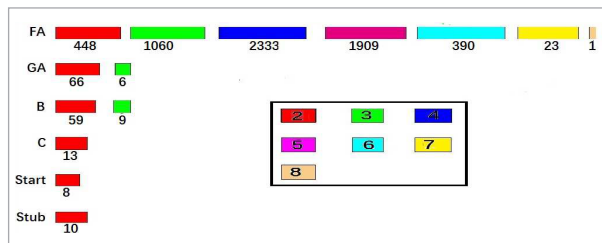
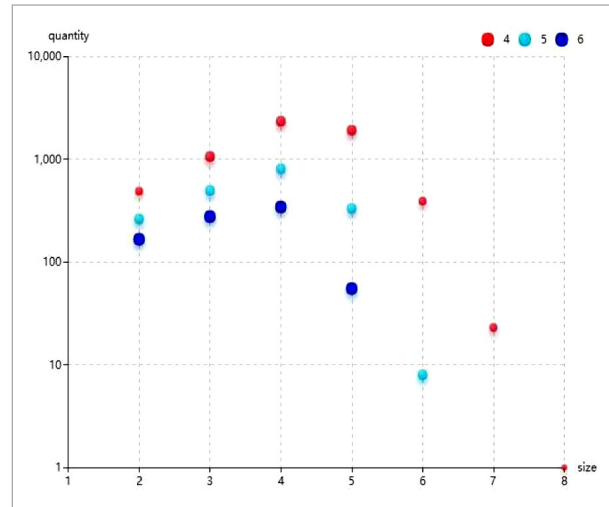


Figure 8 depicts the quantities of collaboration subgraphs of different sizes under three different thresholds, which only belong to “FA” collaboration networks. Owing to the quantities of different sizes, collaboration subgraphs belonging to the other five types of collaboration networks are too small, as can be seen from Figure 7, we did not present them in this figure. As can be seen from Figure 8, as the threshold

Figure 8

The distribution of collaboration subgraphs size at the thresholds 4, 5 and 6, respectively. The quantity of nodes in a collaboration subgraph is the size of it here

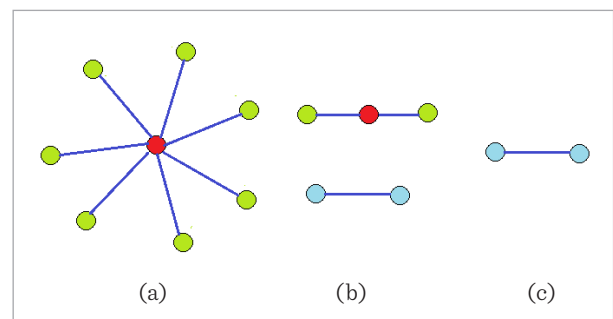


goes higher, the quantity of collaboration subgraphs of the same size goes down. In other words, when the scales of collaboration patterns are the same, types of frequent collaboration patterns are less than those of less frequent collaboration patterns.

Figure 9 shows that the topological structures of collaboration subgraphs belonging to “FA” collaboration networks are often star topologies and the topological structures of collaboration subgraphs that belong to the other five types of collaboration networks are fre-

Figure 9

The typical topological structures of collaboration subgraphs corresponding to different types of collaboration networks. Figure 9.a belongs to “FA” collaboration networks. Figure 9.b belongs to the collaboration networks of “GA” and “B”. Figure 9.c belongs to the other three types of collaboration networks



quently linear topologies. Figure 9.a stands for a collaboration pattern that several editors coauthor on an article and one of them plays leading role in the process of collaboratively producing the article. Figure 9.b and Figure 9.c are representative of two collaboration patterns, respectively. The collaboration pattern consisting of three editors also has one editor who plays leading role. The collaboration pattern which is composed of two editors is the simplest case and the importance of these two editors is equal in essence. Figure 9 indicates that the collaboration patterns that Figure 9.a stands for constantly produce articles of high quality. In fact, a successful collaboration team often has relatively more complex relationship among members of the team and there constantly exist some people in the team shouldering important responsibility. The person should lead the team and coordinate members of the team in order to lead the outcome of the team to develop towards a good direction.

Table 3 shows that how many edits editor “Ealdgyth” has performed respectively to six classes of articles in our datasets. The editor often edited articles of “FA” class, but the editor edited articles of other classes not so often. From the results of collaboration subgraphs mined from collaboration networks, we discovered that the node that editor “Ealdgyth” stands for appears in the center of many collaboration subgraphs of “FA” collaboration networks and also appears in the collaboration subgraphs of “Stub” collaboration networks. To be more specific, the editor plays an important role in the collaboration patterns that constantly appear in “FA” articles. However, the role of the editor in collaboration patterns that frequently appear in “Stub” articles is not so important. These collaboration patterns that respectively appear in these two classes of articles are actually different. Although the editors who play an important role in the evolution of these two classes of articles are same, the qualities of them differ greatly.

Table 4 shows that how many edits editor “Addbot” has performed, respectively, to six classes of articles

Table 3

Number of edits performed by editor “Ealdgyth” to per class articles

| Class | FA | GA | B | C | Start | Stub |
|--------|------|------|-----|-----|-------|------|
| #edits | 7250 | 2606 | 100 | 228 | 339 | 190 |

Table 4

Number of edits performed by editor “Addbot” to per class articles

| Class | FA | GA | B | C | Start | Stub |
|--------|-----|-----|-----|-----|-------|------|
| #edits | 285 | 311 | 506 | 587 | 718 | 362 |

in our datasets. The editor constantly edited articles of “Start” class and performed many edits to other classes of articles as well. According to results of collaboration subgraph mined from collaboration networks, we found that the node the editor is representative of is a node in most frequent subgraphs of “Start” collaboration networks. The editor who constantly appears in bad collaboration patterns, however, has also collaborated with other editors to produce many “FA” articles and “GA” articles.

Therefore, Tables 3 and 4 indicate that collaboration patterns among editors can exert more remarkable influence on article quality than that of the ability of any editor in these collaboration patterns. Without a successful collaboration pattern among editors, they may also produce articles of bad quality though some of them may be rather capable. While many editors who may not be very capable, they can also produce articles of high quality with a successful collaboration among them.

Furthermore, we also found that there exist some collaboration subgraphs that only appear in “FA” or “GA” collaboration networks. Some collaboration subgraphs can be only discovered in “Start” or “Stub” collaboration networks. The former is larger in size than the latter and the quantity of the latter is much less than that of the former as well. This experimental phenomenon infers that there are some editors constantly cooperating with each other in a fixed collaboration pattern to produce articles of high quality.

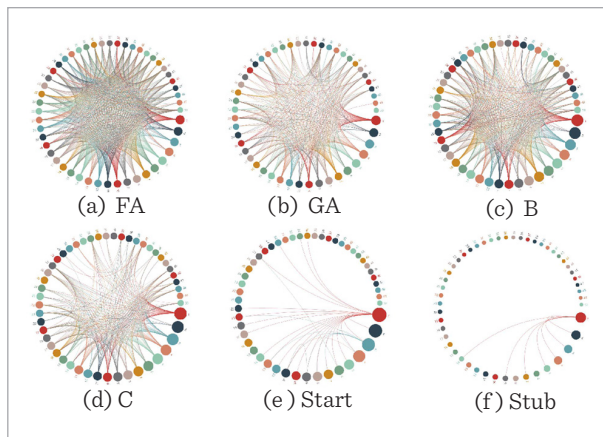
In general, as can be noted from Figure 10, from Figure 10.a to Figure 10.f, curves in each figure go denser. Only the case that curves in Figure 10.b are sparser than curves in Figure 10.c is an exception. Denser curves in Figure 10 mean closer collaboration among editors. Consequently, Figure 10 implies that closer collaboration among editors will produce more excellent articles more often than not.

In conclusion, without appropriate collaboration among members of a team, the sum of one plus one may be less than one. One plus one may be greater than

two if all the members cooperate with each other appropriately. Hence discovering and analysing various collaboration patterns is of great importance. Based on analysing various collaboration patterns which have been discovered already, evaluation of current collaboration and some suggestions may be given.

Figure 10

Figure 10a stands for the frequent collaborations in “FA” class articles among top 50 editors who edit most “FA” articles. If the number of “FA” articles that two editors have coauthored surpasses the threshold 25, there will be a curve connecting one node with another node. Each node stands for an editor. The bigger the size of a node is, the more “FA” articles the editor, the node is representative of, has edited. It is the same case to the five figures remaining



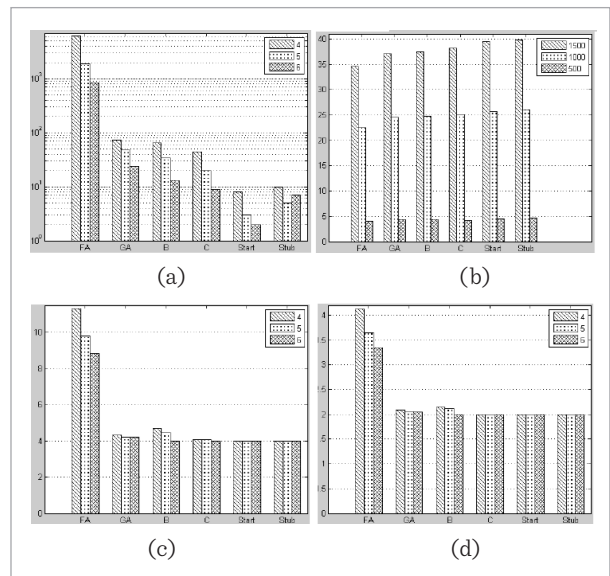
5.2.2. Analysing the Influences of Collaboration Patterns on Collaborative Outcome

As is shown in Figure 11.a, the lower the quality of articles is, the less the quantity of collaboration subgraphs is. Only the case that quantity of collaboration subgraphs of “Start” collaboration networks is less than that of collaboration subgraphs of “Stub” collaboration networks is an exception.

To reveal the correlation between quality of articles and the quantity of collaboration subgraphs further, we respectively sampled 500, 1000 and 1500 articles at random from our datasets according to the article class distribution in Table 1. Then we gave each article a ranking according to the quantity of collaboration subgraphs belonging to it. Articles will be given the same ranking if quantities of collaboration subgraphs belonging to them are equal. Finally, we calculated average ranking of six classes of articles, respectively. As

Figure 11

Figure 11.a shows the quantities of collaboration subgraphs respectively mined from six types of collaboration networks under three different thresholds. Figure 11.b is average rankings of six classes of articles. Figure 11.c is average size of collaboration subgraphs respectively mined from six types of collaboration networks under three different thresholds and the number of nodes in a collaboration subgraph is the size of it. Figure 11.d is average complexity of collaboration subgraphs respectively mined from six types of collaboration networks under three different thresholds. Corresponding standard deviations have been attached in appendices



can be observed in Figure 11.b, the higher the quality of articles is, the higher the average ranking of them is. Figure 11.c indicates that the average sizes of frequent subgraphs mined from “FA” collaboration networks is the biggest and from the other five types of collaboration networks remaining is almost the same.

We can intuitively imagine that if two graphs have the same number of nodes but the number of edges in one graph is more than that in another, then the former will be deemed more complex than the latter. We can also imagine that if the number of edges in a graph divided by the number of nodes in it equals that of another, the graph which has more nodes will be deemed more complex. Hence the complexity of a graph is simply defined by us as follows:

$$Complexity = \frac{N_{edge}}{N_{node}} + N_{node}$$

where N_{edge} is the number of edges and N_{node} is the number of nodes. According to this definition, we calculated average complexity of collaboration subgraphs mined from six types of collaboration networks, respectively. Then Figure 11.d indicates that the average complexity of collaboration subgraphs mined from “FA” collaboration networks is the most complex and from the five types of collaboration networks remaining is almost the same.

Table 5

Result of frequent itemset mining. For there are 316 “FA” articles in our datasets, there are 316 corresponding transactions. Each transaction is composed of all editors of corresponding article. It is the same case to “GA”, “B”, “C”, “Start” and “Stub” articles. The threshold for frequent itemset mining is 50

| Class | FA | GA | B | C | Start | Stub |
|----------|-------|-----|------|-----|-------|------|
| 1-itemse | 120 | 52 | 70 | 48 | 16 | 5 |
| 2-itemse | 1326 | 204 | 442 | 184 | 12 | 2 |
| 3-itemse | 4828 | 186 | 911 | 223 | 2 | 0 |
| 4-itemse | 8587 | 44 | 800 | 118 | 0 | 0 |
| 5-itemse | 7250 | 0 | 277 | 23 | 0 | 0 |
| 6-itemse | 2772 | 0 | 29 | 1 | 0 | 0 |
| 7-itemse | 414 | 0 | 0 | 0 | 0 | 0 |
| 8-itemse | 12 | 0 | 0 | 0 | 0 | 0 |
| sum | 25309 | 486 | 2529 | 597 | 30 | 7 |

The experimental results presented in Figure 11.c and Figure 11.d indicate that the metric average size of frequent subgraphs and the metric average complexity of frequent subgraphs can distinguish “FA” class articles from other classes of articles very well but the two metrics cannot work very well in distinguishing among the other five classes of articles. Figure 12.a can also illustrate this experimental phenomenon.

Furthermore the results of frequent subgraph mining above, we also utilize the algorithm FPGrowth to mine frequent itemsets from our datasets [24, 25, 27]. This experiment is complementary to the experiment of frequent subgraph mining. As is shown in Table 5, for the same column in the table, the quantity of frequent itemsets goes lower from “FA” to “Stub” except “GA”. Hence, generally speaking, the experimental results indicate that there are strong and positive correlations between the quantity of frequent itemsets

and the qualities of articles. The results of frequent itemset mining are consistent with the results of frequent subgraph mining.

5.2.3. Testifying the Influence of Collaboration Networks on Collaborative Outcome

To further testify the correlations between collaboration and the qualities of collaborative outcome, we performed a six-classification task. We sampled 1506 articles from our datasets, in which the quantities of “FA”, “GA”, “B”, “C”, “Start” and “Stub” articles are all 251. We selected 100 articles from the 1506 articles at random as test set data, in which the numbers of “FA” and “GA” articles are both 16 and the numbers of the remaining four classes of articles are all 17. The 1406 articles remaining were selected as training set data. We then mined collaboration subgraphs from collaboration networks constructed from these 1506 articles at the threshold of 6. Here, 1506 collaboration networks are constructed from the 1506 articles and collaboration subgraphs are mined from the 1506 collaboration networks together at the threshold of 6. In the next step, we extracted features from those collaboration subgraphs and each collaboration network. In Figure 12.a, the first three features are mentioned above. Furthermore, P, C and L are three statistical properties of complex networks, which are defined in [29]. In this paper, P, C and L are the properties of frequent subgraphs which can also be regarded as complex networks. It is noteworthy that C is defined in [29] as follows:

$$C = \frac{1}{n} \sum_{i=1}^n \frac{2k_i}{k_i(k_i - 1)},$$

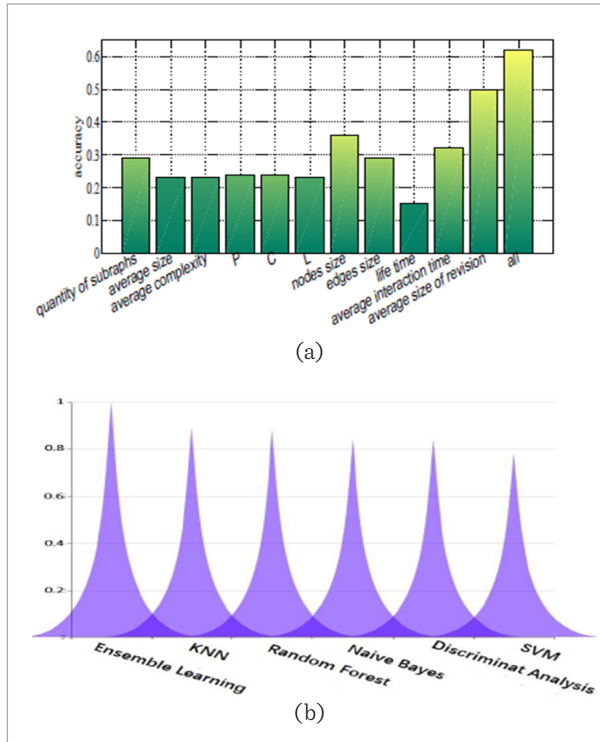
where k_i is the degree of a node in a network and n is the number of nodes in the network. However, if $k_i = 1$, then C is infinite. In our collaboration subgraphs the degree of a node may equal 1. To avoid this case, we revised it into the following:

$$C = \frac{1}{n} \sum_{i=1}^n \frac{4k_i}{k_i(2k_i - 1)}.$$

The nodes size and edges size in Figure 12.a are the quantities of nodes and edges in a collaboration network, respectively. The feature of lifetime is the time of existence of an article. The feature of average interaction time is the mean of time that how long ed-

Figure 12

Figure 12.a shows the accuracy of classification by separately applying 11 features and the combination of these features, based on Random Forest (using 5 trees), to classify 100 articles into six classes. Figure 12.b is the accuracy of classification by applying six classification algorithms respectively to classify articles into two classes, among which “FA” articles are treated as one class and the remaining articles are treated as another class. The detail parameterizations of classifiers in Figure 12.b have been attached in appendices



itors interact with each other one times. This feature reflects the frequency of interactions among editors. The last feature is the average size of revision histories of an article which is the mean of all revision histories’ size. As can be seen from Figure 12.a, all the features do have something to do with the quality of an article. The accuracy of classification by combining all 11 features is as high as 62%. Without learning about anything about the content of an article, we can just utilize these features to identify the class of an article with an accuracy of 62%. Therefore, 62% is a relatively high accuracy in fact.

Furthermore, we also calculated recall per class according to the results of classification mentioned above. Table 6 also indicates that the results of the

classification are good. The recall per class in Table 6 is not inferior to that in [23].

Due to many other researchers [5, 33, 34], applying various classification algorithms, classified articles into two classes, we also utilized the combination of 11 features presented in Figure 12.b, based on six types of classification algorithms, to classify articles into two classes, respectively. From Figure 12.b, we can see that the classification accuracy of Ensemble Learning (ensembles for Boosting, Bagging and Random Subspace) is the highest, which is nearly 100%. The classification accuracy of SVM (Support Vector Machine) is the lowest with accuracy of 78%. It was noteworthy that the high classification accuracy can be expected for the reason that “FA” articles can be apparently distinguished from other articles by these features, which can be clearly seen in Figure 11. Hence the accuracy of classification in our work is very high and exceeds, to the best of our knowledge, the accuracy of classification in all related work performed by other researchers.

Table 6

Recall per class

| Class | FA | GA | B | C | Start | Stub |
|--------|------|--------|--------|---------|--------|--------|
| Recall | 1.00 | 0.4375 | 0.5882 | 0.41.18 | 0.4706 | 0.8224 |

Table 7

The results of the quantity of nodes and quantity of edges in Table 2 divided by the quantity of articles in Table 1

| Class | FA | GA | B | C | Start | Stub |
|------------------|-----|-----|-----|-----|-------|------|
| #nodes/#articles | 548 | 376 | 489 | 340 | 84 | 26 |
| #edges/#articles | 983 | 664 | 810 | 515 | 111 | 31 |

Table 8

Result of quantity of edges divided by quantity of nodes in Table 2

| Class | FA | GA | B | C | Start | Stub |
|---------------|-------|-------|-------|-------|-------|-------|
| #edges/#nodes | 1.792 | 1.766 | 1.653 | 1.515 | 1.326 | 1.165 |

6. Discussion and Conclusion

Researching collaboration among editors in online collaboration systems is a significant challenge. In this paper, we have utilized collaborative subgraphs to deeply

anatomize the influences of collaboration patterns on the collaborative outcome. Some corresponding classification tasks are performed to testify the influence on the collaborative outcome. There, in fact, exist some explanations for the correlations between collaboration characteristics among editors and the collaborative outcome. Compared with a less excellent article, articles of high quality, generally speaking, will be edited by more editors and many of them may edit these articles repeatedly. That is to say, usually an article of high quality will be produced by the closer collaboration among more editors. The collaboration networks of excellent articles are accordingly larger and more complex. From Table 7, it can be verified that collaboration networks of high quality articles are, on average, larger than networks of less excellent articles. From Table 8, we can roughly see that collaboration networks of high quality articles are, on average, more complex than collaboration networks of less excellent articles. Naturally, the quantity of collaboration subgraphs mined from the former collaboration networks is more than that of collaboration subgraphs mined from the latter collaboration networks and collaboration subgraphs mined from the former are more complex than those mined from the latter as well. Figure 13 can intuitively demonstrate the explanations stated above.

In future work, we will further explore the view proposed in this paper and the features of collaboration subgraphs. We will apply this view to other online collaboration systems to further verify this view. Furthermore, because the process of producing contents

by collaboration among editors is dynamic, the quality of contents created collaboratively in these systems may gain improvement materially assuming authors can learn about the status of current collaboration and adopt corresponding suggestions for future collaboration. These suggestions may be provided based on our future work. Therefore, in the future, we will try to research how to concretely guide editors according to collaboration patterns discovered in this paper to collaboratively create contents of high quality.

Appendices

Table 9

The standard deviation for Figure 11.b

| #articles | FA | GA | B | C | Start | Stub |
|-----------|--------|--------|--------|--------|--------|--------|
| 500 | 1.1412 | 0.9279 | 0.7743 | 0.9233 | 0.6723 | 0.5430 |
| 1000 | 6.0686 | 3.8332 | 3.0273 | 1.9459 | 1.6689 | 1.4675 |
| 1500 | 8.2465 | 5.1357 | 4.6544 | 4.2470 | 1.8949 | 1.7258 |

Table 10

The standard deviation for Figure 11.c

| Thresholds | FA | GA | B | C | Start | Stub |
|------------|--------|--------|--------|---|-------|------|
| 4 | 1.0329 | 0.2764 | 0.3454 | 0 | 0 | 0 |
| 5 | 0.9368 | 0.2421 | 0.3222 | 0 | 0 | 0 |
| 6 | 0.8669 | 0.1998 | 0 | 0 | 0 | 0 |

Figure 13

A simple example of two collaboration networks, among which the left belongs to a less excellent article and the right belongs to an article of high quality. Edge labels are ignored to simplify the visualization

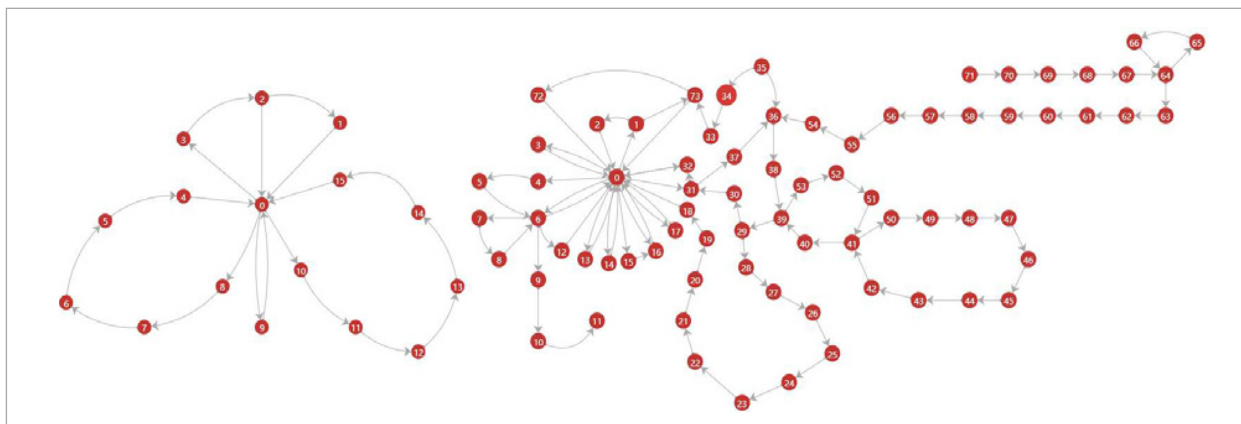


Table 11

The standard deviation for Figure 11.d

| Thresholds | FA | GA | B | C | Start | Stub |
|------------|--------|--------|--------|--------|-------|------|
| 4 | 1.1337 | 0.3426 | 0.4890 | 0.1041 | 0 | 0 |
| 5 | 1.0589 | 0.2887 | 0.4092 | 0.1090 | 0 | 0 |
| 6 | 1.0050 | 0.2496 | 0 | 0 | 0 | 0 |

Acknowledgement

This work is partially supported by Natural Science Foundation of China (No. 61302077), and Funds for Creative Research Groups of China (No. 61421061)

References

- Adler, B. T., Alfaro, L. d. A Content-Driven Re-putation System for the Wikipedia. Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada, May 08 - 12, 2007, 261-270. <https://doi.org/10.1145/1242572.1242608>
- Adler, B. T., Alfaro, L. d., Pye, I., Raman, V. Measuring Author Contributions to the Wikipedia. Proceedings of the 4th International Symposium on Wikis, Porto, Portugal, September 08 - 10, 2008, 1-10. <https://doi.org/10.1145/1822258.1822279>
- Bernot, G., Gaudel, M. C., Marre, B. Software Testing Based on Formal Specifications: A Theory and a Tool. Software Engineering Journal, 1991, 6(6), 387-405. <https://doi.org/10.1049/sej.1991.0040>
- Betancourt, G. G., Segnine, A., Trabuco, C., Rez-gui, A., Jullien, N. Mining Team Characteristics to Predict Wikipedia Article Quality. Proceedings of the 12th International Symposium on Open Collaboration, Berlin, Germany, August 17 - 19, 2016, 1-9. <https://doi.org/10.1145/2957792.2971802>
- Blumenstock, J. E. Size Matters: Word Count as a Measure of Quality on Wikipedia. Proceedings of the 17th International Conference on World Wide Web, Beijing, China, April 21 - 25, 2008, 1095-1096. <https://doi.org/10.1145/1367497.1367673>
- Bringmann, B., Nijssen, S. What Is Frequent in a Single Graph? Pacific-Asia Conference on Knowledge Discovery and Data Mining, Berlin, Heidelberg, 2008, 858-863. https://doi.org/10.1007/978-3-540-68125-0_84
- Calzada, G. D. I., Dekhtyar, A. On Measuring the Quality of Wikipedia Articles. Proceedings of the 4th Workshop on Information Credibility, Raleigh, North Carolina, USA, April 27-27, 2010, 11-18. <https://doi.org/10.1145/1772938.1772943>
- Chevalier, F., Huot, S., Fekete, J.-D. Wikipediaviz: Conveying Article Quality for Casual Wikipedia Readers. 2010 IEEE Pacific Visualization Symposium (PacificVis), Taipei, Taiwan, 2-5 March, 2010, 49-56. <https://doi.org/10.1109/PACIFICVIS.2010.5429611>
- Dalip, D. H., Gonçalves, M. A., Cristo, M., Calado, P. Automatic Quality Assessment of Content Created Collaboratively by Web Communities: A Case Study of Wikipedia. Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, Austin, TX, USA, June 15-19, 2009, 295-304. <https://doi.org/10.1145/1555400.1555449>
- Elseidy, M., Abdelhamid, E., Skiadopoulou, S., Kalnis, P. Grami: Frequent Subgraph and Pattern Mining in a Single Large Graph. Proceedings of the Vldb Endowment, 2014, 7(7), 517-528. <https://doi.org/10.14778/2732286.2732289>
- Flöck, F., Acosta, M. Whovis: Visualizing Editor Interactions and Dynamics in Collaborative Writing over Time. Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, May 18 - 22, 2015, 191-194. <https://doi.org/10.1145/2740908.2742846>
- Iba, T., Nemoto, K., Peters, B., Gloor, P. A. Analyzing the Creative Editing Behavior of Wikipedia Editors: Through Dynamic Social Network Analysis. Procedia - Social and Behavioral Sciences, 2010, 2(4), 6441-6456. <https://doi.org/10.1016/j.sbspro.2010.04.054>
- Jurgens, D., Lu, T.-C. Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia. Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012.

Table 12

The parameterizations of classifiers in MATLAB for Figure 12.b

| Classifiers | parameterizations of classifiers |
|-------------------|---|
| Ensemble Learning | <i>fitensemble(data, labels, 'AdaBoostM1', 100, 'tree', 'type', 'classification')</i> |
| Random Reforest | <i>TreeBagger(5, trainData, trainLabel)</i> |
| KNN | <i>ClassificationKNN.fit(data, labels, 'NumNeighbors', 5)</i> |

14. Juszczyszyn, K., Kazienko, P., Musiał, K. Local Topology of Social Network Based on Motif Analysis. *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Berlin, Heidelberg, 2008, 97-105. https://doi.org/10.1007/978-3-540-85565-1_13
15. Keegan, B., Gergle, D., Contractor, N. Staying in the Loop: Structure and Dynamics of Wikipedia's Breaking News Collaborations. *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, Linz, Austria, August 27-29, 2012, 1-10. <https://doi.org/10.1145/2462932.2462934>
16. Korfiatis, N. T., Poulos, M., Bokos, G. Evaluating Authoritative Sources Using Social Networks: An Insight from Wikipedia. *Online Information Review*, 2006, 30(3), 252-262. <https://doi.org/10.1108/14684520610675780>
17. Laniado, D., Tasso, R. Co-Authorship 2.0: Patterns of Collaboration in Wikipedia. *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia*, Eindhoven, The Netherlands, June 06 - 09, 2011, 201-210. <https://doi.org/10.1145/1995966.1995994>
18. Lih, A. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource. *Nature*, 2004, 3(1), 16-17.
19. LiJintao, X., TangTing, WangZhunchen, Rijke, L. d. Automatically Assessing Wikipedia Article Quality by Exploiting Article-Editor Networks. *European Conference on Information Retrieval*, Cham, 2015, 574-580. https://doi.org/10.1007/978-3-319-16354-3_64
20. Lipka, N., Stein, B. Identifying Featured Articles in Wikipedia: Writing Style Matters. *Proceedings of the 19th International Conference on World Wide Web*, Raleigh, North Carolina, USA, April 26-30, 2010, 1147-1148. <https://doi.org/10.1145/1772690.1772847>
21. Liu, J., Ram, S. Who Does What: Collaboration Patterns in the Wikipedia and Their Impact on Article Quality. *ACM Transactions on Management Information Systems*, 2011, 2(2), Article No. 11. <https://doi.org/10.1145/1985347.1985352>
22. Ramraj, T., Prabhakar, R. Frequent Subgraph Mining Algorithms - a Survey. *Procedia Computer Science*, 2015, 47, 197-204. <https://doi.org/10.1016/j.procs.2015.03.198>
23. Robertie, B. d. L., Pitarch, Y., Teste, O. Measuring Article Quality in Wikipedia Using the Collaboration Network. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, Paris, France, August 25-28, 2015, 464-471. <https://doi.org/10.1145/2808797.2808895>
24. Savulioniene, L., Sakalauskas, L. A Stochastic Algorithm of Frequent Set Search for Mining Association Rules. *Information Technology and Control*, 2014, 43(2), 121-132. <https://doi.org/10.5755/j01.itc.43.2.3135>
25. Shekhar, S., Xiong, H. Frequent Itemset Mining. *Encyclopedia of GIS*, 2008, 323-323. https://doi.org/10.1007/978-0-387-35973-1_433
26. Siu, M.-K. Introduction to Graph Theory. *The Mathematical Gazette*, 1998, 82(494), 343-344. <https://doi.org/10.2307/3620453>
27. Suresh, J., Rushyanth, P., Trinath, C. Generating Associations Rule Mining Using Apriori and Fpgrowth Algorithms. *International Journal of Computer Trends and Technology*, 2013, 4(4), 887-892.
28. Suzuki, Y., Yoshikawa, M. Assessing Quality Score of Wikipedia Article Using Mutual Evaluation of Editors and Texts. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, San Francisco, California, USA, October 27 - November 01, 2013, 1727-1732. <https://doi.org/10.1145/2505515.2505610>
29. Wang, R., Gao, J., Gao, Z., Gao, X., Jiang, H. Complex Network Theory-Based Condition Recognition of Electromechanical System in Process Industry. *Science China Technological Sciences*, 2016, 59(4), 604-617. <https://doi.org/10.1007/s11431-016-6025-2>
30. Wang, S., Iwaihara, M. Quality Evaluation of Wikipedia Articles through Edit History and Editor Groups. *Asia-Pacific Web Conference*, Berlin, Heidelberg, 2011, 188-199. https://doi.org/10.1007/978-3-642-20291-9_20
31. Wilkinson, D. M., Huberman, B. A. Cooperation and Quality in Wikipedia. *Proceedings of the 2007 International Symposium on Wikis*, Montreal, Quebec, Canada, October 21-25, 2007, 157-164. <https://doi.org/10.1145/1296951.1296968>
32. Wöhner, T., Peters, R. Assessing the Quality of Wikipedia Articles with Lifecycle Based Metrics. *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, Orlando, Florida, October 25-27, 2009, 1-10. <https://doi.org/10.1145/1641309.1641333>
33. Wu, G., Cunningham, P. Integration of Multiple Network Views in Wikipedia. *Knowledge and Information Systems*, 2015, 45(2), 473-490. <https://doi.org/10.1007/s10115-014-0802-7>
34. Wu, G., Harrigan, M., Cunningham, P. Classifying Wikipedia Articles Using Network Motif Counts and Ratios. *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, Linz, Austria, August 27-29, 2012, 1-10. <https://doi.org/10.1145/2462932.2462948>