# Understanding co-editing mechanism of wiki-based digital humanities projects

Weihua Deng and Pei Lv
*Department of Information Management, College of Public Administration, Huazhong Agricultural University, Wuhan, China, and*
Ming Yi and Ming Liu
*School of Information Management, Central China Normal University, Wuhan, China*

Received 25 August 2019
Revised 13 October 2019
Accepted 19 October 2019

## Abstract

**Purpose** – The purpose of this paper is to reveal the co-editing mechanism aiming at content creation, and an entry of online encyclopedia is taken as a case, for the purpose of promoting and enhancing the development of wiki-based digital humanities projects (WDHPs), specifically, the projects that focus on gathering contextual information in the culture heritage domain.

**Design/methodology/approach** – An exploratory study was conducted by three steps. A representative entry's editorial records were reorganized to obtain a data set of discussion statements ($n = 608$), based on which linked-structures were built, and PageRank algorithm was used to analyze the co-editing process. Skewness statistic was applied to measure the consensus of co-editing, and consensus evolution over time was explored. Linear or curve fitting was performed to analyze the correlation between consensus evolution and its influential factors.

**Findings** – In WDHPs, co-editing activity of content creation can be considered as a large-scale group discussion, consensus can evaluate the efficiency of co-editing, which evolves with time and is influenced by the number of statements, breadth and depth of argumentation structure. Taking "Mogao Grottoes" as an example, group discussions around 15 key issues dominate the content creating process, consensus is on a rise with time, finally reaches a relatively high level, and consensus evolution is more influenced by breadth than by depth of argumentation structure, which indicates that co-editing efficiency of "Mogao Grottoes" is fine and more argumentation in a depth manner should be guided.

**Practical implications** – For researchers of WDHPs, it is beneficial to apply online encyclopedia platform combining with consensus analysis to develop WDHPs. For designers of WDHPs, the elements related to argumentation structure can be absorbed into the design to promote co-editing in an effective manner. For DH researchers, the analytic procedure can be beneficial of revealing the interest of contributors in a specific DH field.

**Originality/value** – This research is novel in comprehensively understanding co-editing mechanism of content creation in WDHPs, resulting in a three-step analytic procedure of presenting co-editing process, evaluating and improving co-editing efficiency.

**Keywords** Crowdsourcing, Digital humanities, Consensus, Co-editing, Culture heritage, IBIS

**Paper type** Research paper

## 1. Introduction

The digital revolution of the twenty-first century is empowering humanities scholars to pursue new strategies of collection, organization and evaluation of valuable resources (Travis and Holm, 2016). Wiki technology, being a prominent Web 2.0 application, has received intense interest because of its characteristics of collaboration. A certain number of digital humanities (DH) scholars have leveraged wiki tools to develop DH projects, which are called as wiki-based DH projects (WDHPs), such as "Australian Newspaper Digitization," "Wiki Loves Art," "Netherlands Institute for Sound and Vision," etc. Collaboration is defined as working together to achieve collective results (Lipnack and Stamps, 1997). In the sense,

This work has been supported by the National Social Science Foundation of China (16BTQ076).

Aslib Journal of Information Management
Vol. 72 No. 2, 2020
pp. 199-218

2050-3806
DOI 10.1108/AJIM-08-2019-0214

most WDHPs have applied Wiki tools in gathering human resources as the specific collective results. For example, the Australian Newspaper Digitization Project invites users around the world to calibrate the Optical Character Recognition (OCR)'ed newspaper text, a complex task that is difficult to deal with limited manpower. Nonetheless, collaborative editing (co-editing) as the core contribution pattern of wiki tools, being proficient in collaboratively creating web content, still gets a limited application among the current WDHPs. Existing research has confirmed that co-editing can enable users to be actively involved and work with peers, to review and share ideas and, finally, to create content in a collaborative manner (Hadjerrouit, 2014; Bradley *et al.*, 2010). Furthermore, online encyclopedia (e.g. Wikipedia) is the successful product of the co-editing pattern, and it confirms the enormous capacity of co-editing in content creation. Indeed, in the WDHPs of culture heritage domain, there exists a long tradition that gathering contextual content has been done by means of writing scientific publications, compiling magazines, etc. (Oomen and Aroyo, 2011). In this sense, gathering contextual information is similar to content creation. Therefore, it is evident that co-editing pattern offers an enhanced opportunity to engage citizen scientists and volunteers in creating scientific content, leading to potential prosperity of the WDHPs (Bonney *et al.*, 2009; Follett and Strezov, 2015; Zhao, 2017), what we lack is a clear understanding of the co-editing mechanism of the creating process of contextual content.

Despite the limited number of WDHPs orient to content creation, online encyclopedia has become widely recognized as an invaluable way to generate scientific content by co-editing, and numerous pages of online encyclopedia are devoted to humanities topics. Therefore, online encyclopedia provides an environment to explore the co-editing mechanism of WDHPs. Recently, the co-editing of online encyclopedia has received significant attention. There are a certain number of literature focusing on the exogenous factors (i.e. technology and management) about co-editing, such as the function and interaction interface of wiki system (Li *et al.*, 2004; He and Han, 2006), the organization of contributors (Miller, 2005; Nov, 2007), editing coordination (Kittur and Kraut, 2008), conflict management (Arazy *et al.*, 2013) and so on. A few studies focus on co-editing activities, such as encyclopedia contributors' behaviors and co-editing patterns (Kittur and Kraut, 2008; Kimmerle *et al.*, 2010; Zhang *et al.*, 2018). However, the scholarly literature has offered little guidance on exploring co-editing mechanism from the perspective of content creation. In fact, being the target task of co-editing, content creation has the ability to show a whole picture of how content is collaboratively created by multiple contributors. This is the gap of the existing literature.

Recently, the emerging web technologies have promoted the prosperity of informal online conversations, and the well-established theory of group discussion is used to investigate these issues, leading to a research boom in online discussion theme. Group discussion is a decision-oriented conversation in which a group weighs pros and cons of different options, articulates core values, and makes choices in a way that is respectful, egalitarian, and open (Karacapilidis and Pappis, 2000). Related literature demonstrates a strong theoretical foundation in such aspects of group discussion as the discussion process and key elements, which are applied in in-deep exploration of the online discussion phenomena in various settings, including majority opinion formation through social media (Netzer *et al.*, 2012; Lee *et al.*, 2018), e-voting activities in public deliberation domain (Chowdhury *et al.*, 2015), and collaborative learning activities in online educational environment, and so on. Co-editing in WDHPs is based on Wiki tools which are known to have some distinct characteristics such as the many-to-many conversation mechanism that greatly differs from classic collaboration (Arazy and Kopak, 2010). In terms of conversation, co-editing in WDHPs is similar to online discussion, which exhibits a broad range of discussion style. Furthermore, wiki platform offers an environment to facilitate collaborative work by providing the tools for coordinating, consensus building and

decision making. Therefore, group discussion theory offers a good way to explore the process and efficiency aspects of co-editing.

In response, the present study investigates co-editing mechanism aiming at three objectives. First, the co-editing process in WDHPs is clearly analyzed by drawing on group discussion theory (Chowdhury *et al.*, 2015; Murr, 2015). Second, it evaluates the co-editing efficiency by consensus analysis based on the method of skewness statistic (Cohen, 1996; Lee *et al.*, 2018; Munneke *et al.*, 2007; Webb *et al.*, 2016). Third, the influential factors of co-editing efficiency are investigated by incorporating the depth and breadth of argumentation structure (Cao and Protzen, 1999; Chinn and Anderson, 1998). Given that there is a successful application of co-editing in online encyclopedia, we carry out an exploratory case study of an encyclopedia entry from BaiduBaike.com (i.e. the most influential Chinese online encyclopedia), the entry formation process is considered as a contextual content creating process and the entry's editorial records are reorganized to form a data set comprising 608 statements and 90 issues. The findings reveal the co-editing mechanism of content creation, which will provide useful insights for promoting and enhancing the development of WDHPs.

## 2. Literature review and theory development

### *2.1 Wiki-based DH projects*

Wiki technology is embraced by many DH projects, leading to the development of WDHPs. Oomen and Aroyo (2011) explore a certain number of influential crowdsourcing projects in DH domain, and classify them into six categories. Two types of them belong to WDHPs, namely, correction and contextualization. Table I shows four WDHPs from Oomen and Aroyo (2011), including one correction project and three contextualization projects. In particular, the Australian newspaper digitization initiated by National Library of Australia is a typical example of correction project, which includes the huge amount of data resources, with a great need for the participation and collaboration of volunteers. The Netherlands Institute for Sound and Vision is an example of contextualization project, and it uses a wiki platform to gather contextual information on television programs, broadcasters, presenters and so on. In the four projects, three projects are based on the third-party wiki platform, such as Wikipedia, and only one builds a wiki-style platform by itself. In terms of the collective results of these WDHPs, correction project only harvests manpower, while three contextualization projects harvest contextual information, yet most contextual information harvested is in the form of the pictures or videos, and content (e.g. article) is less harvested.

Among the above WDHPs, wiki technology has been utilized as a crowdsourcing tool, yet the application of wiki technology is at an elementary level. In fact, gathering contextual information in WDHPs is a particular collaborative task, which is traditionally doing by means of writing scientific articles, compiling magazines that record the city history, investigating the family histories and so on (Oomen and Aroyo, 2011). It is not difficult to

| Project | Wiki platform | Collective results | Type of project |
|---|---|---|---|
| Australian newspaper digitization initiated by National Library of Australia | Wikipedia platform | Manpower | Correction |
| Netherlands Institute for Sound and Vision | Self-build wiki platform | Television programmers, broadcasters, presenters and so on | Contextualization |
| Wiki Loves Art initiated by a consortium of US/UK museums | Wikipedia platform | Images from museum objects on Wikimedia Commons | Contextualization |
| Wikipedian in residence initiated by British Museum | Wikipedia platform | An individual contributing to Wikipedia joining as part of the museum | Contextualization |

**Table I.** WDHPs from Oomen and Aroyo (2011)

find that gathering contextual information is similar to content creation. Although public participation is critical for content creation (Arazy and Gellatly, 2013), the simple deployment of wiki technology does not guarantee that content creation will be effective. Thus, importance has been given to understand how content is created by co-editing, which offers an opportunity for promoting and enhancing the effectiveness of WDHPs oriented to content creation.

### *2.2 Co-editing of online encyclopedia*

Wiki technology as a tool for collaborative work is designed by Ward Cunningham in 1995 (Cunningham and Leuf, 2001), and co-editing provides a mechanism of collaboratively creating web content as a specific collective result (Lipnack and Stamps, 1997). To date, co-editing has been successfully applied in the field of online encyclopedia (e.g. Wikipedia), and online encyclopedia is prominent in creating content about different topics (Pfeil *et al.*, 2006), notably for humanities topics (Spoerri, 2007).

Co-editing of online encyclopedia has been extensively explored. Earlier research focuses on the individual collaborative behavior, such as the taxonomy for categorizing contributors' actions (Pfeil *et al.*, 2006), motivations of participating in co-editing (Arazy and Gellatly, 2013), the effects of personal traits (Zhang *et al.*, 2018), etc. Some research has explored co-editing from a technical perspective, including the implementation of the co-editing system (Li *et al.*, 2004), and the interaction interface (He and Han, 2006). Other literature tries to reveal the co-editing mechanism from a managerial perspective, in which Wikipedia is described as an organization, and scholars focus on self-governance (Forte *et al.*, 2009), editing coordinate (Kittur and Kraut, 2008), conflict management (Arazy *et al.*, 2013) and so forth. A few research pays attention to the fact that co-editing about encyclopedia entry is a content-related activity, and focuses on co-editing patterns (Kittur and Kraut, 2008), co-evolution of Wikipedia entries and their authors (Kimmerle *et al.*, 2010), revision patterns of Wikipedia entries (Jones, 2008), knowledge creation or knowledge building in the entry formation process (Kittur *et al.*, 2009; Wagner, 2005), etc.

Obviously, although there are a large number of studies on co-editing of online encyclopedia, research directly investigating the co-editing mechanism which aims at content creation is very limited, and it is an obstacle in developing WDHPs.

### *2.3 The study of online discussion*

Online discussion is interpreted as a kind of group discussion phenomena in online settings. If we look for the theoretical basis of group discussion, we would be likely to turn first to the fundamental theories, including group theory and group decision-making theory. Related literature demonstrates a strong theoretical foundation in problem analysis and consensus building (Cohen, 1996; Kunz and Rittel, 1970; Loui, 2010), and online discussions in various settings have been explored comprehensively based on these theories (Chowdhury *et al.*, 2015; Netzer *et al.*, 2012). For example, Murr (2015) links e-voting to Condorcet's jury theorem, being a group decision-making theory and proposes the idea that large-scale discussions can be operational in a variety of contexts. Lee *et al.* (2018) investigate the formation of majority opinion in group interactions on Facebook from a consensus building perspective. Similarly, Langley *et al.* (2014) indicate that herding behavior in online communities is increasingly uniform with one variable of the behavior becoming increasingly dominant.

Furthermore, argumentation is a significant fraction of group discussion, which is an effective reasoning which is critical for human's dealing with conflicting information by taking into account arguments and counter-arguments relevant to certain issues (Loui, 2010). Argumentation scheme refers to a certain pattern by which a set of statements are linked with each other in an argument (Palau and Moens, 2009). Issue Based Information

System (IBIS) is a pioneering argumentation scheme proposed by Kunz and Rittel (1970). It consists of a network of essential four types of statements, namely, issue, position, argument and reference. Figure 1 shows the network of IBIS, which provides a natural and convenient form to capture related information during the discussion process. In a discussion process, a topic, as a starting point, defines the domain of relevance. Then, participants raise issues (or questions) related to the topic. They take the positions (alternatives or options) which are consistent with their opinion regarding the issues. To justify their opinions, they support (or object) the positions with arguments. To strengthen their arguments, they further refer to references.

IBIS has been applied broadly in various settings (Cao and Protzen, 1999; Aldea *et al.*, 2012), specifically, education scholars model the detailed argumentation structure of students' collaborative activities in wiki-writing processes in accordance with IBIS, and propose two dimensions, breadth and depth of the conversations, to measure the structure of argumentation (Chinn and Anderson, 1998; Jimoyiannis and Roussinos, 2017; Munneke *et al.*, 2007).

There are three main reasons why group discussion theory can be used to explore co-editing mechanism in WDHPs. First, co-editing represents a kind of online informal conversation phenomena, which exhibits a broad range of discussion style. Second, although co-editing can be basically conducted on large scale, many interactions actually occur in groups, people work around certain content creation in a group. Third, wiki platform records the content, context, time and structure of interaction in online communities, and these digital traces offer some of the rich and extensive data to explore co-editing mechanism (Lazer *et al.*, 2009).

### *2.4 Generalizing group discussion theory to co-editing of WDHPs*

Indeed, group discussion and co-editing of WDHPs share the following natures: group decision making, consensus building and a similar technical infrastructure, which makes it possible to adopt the idea and method from group discussion frameworks to explore co-editing mechanism of content creation. According to IBIS (Cao and Protzen, 1999; Munneke *et al.*, 2007), issue is the key element of a discussion, and any group discussion is basically seeking an answer to a certain issue. Thus, clarifying the key issues of content creation is the starting point of exploring co-editing mechanism. Inspired by the literature related to writing and revision of articles, the issues related to content creation include two

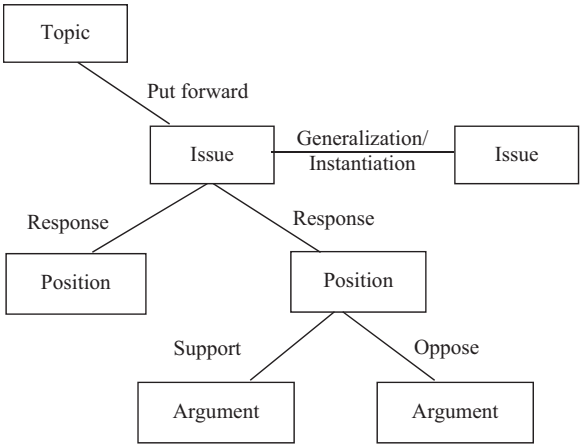



**Figure 1.**
The network of IBIS

parts: macrostructure issue (macro issue) and microstructure issue (micro issue) (Faigley and Witte, 1981). Macro issue corresponds to "major revision" that "would alter the summary of a text" (Jones, 2008), and has a close relationship with the topical structure of an article (Witte, 1983). Micro issue refers to the revision that would affect the meaning of a text, but "would not affect a summary of a text" (Jones, 2008), which mainly aims at writing a text in an appropriate manner.

Taken together, the object of this paper is to draw on the group discussion theories to explore the answers to the following research questions for the purpose of facilitating the contextual content creation in WDHPs:

*RQ1.* How is the content created by co-editing in WDHPs?

*RQ2.* How is the co-editing efficiency of WDHPs evaluated?

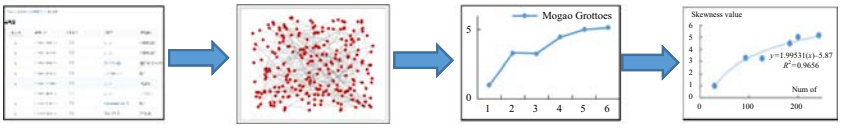
*RQ3.* How is the co-editing efficiency of WDHPs improved?

## 3. Methodology

To explore co-editing mechanism in WDHPs, this study utilized the stepwise analysis procedure consisting of four steps (see Figure 2). First, we identified the entry to be analyzed and defined the period of data collection. Second, the linked-structure was generated based on the reorganized statements, and PageRank (PR) algorithm was run based on the linked-structure to analyze the co-editing process. Third, consensus analysis was performed by conducting skewness statistic, and the consensus evolution with time was investigated. Fourth, the influential factors of consensus evolution were explored using linear or curve fitting.

### *3.1 Data collection*

We conducted an exploratory case study to explore the co-editing mechanism in WDHPs. The typical principle of Patton (1987) was referred to choose the sample. Entry "Mogao Grottoes" of BaiduBaike.com was finally selected as the sample, and there were two reasons as follows. First, countries around the world tried to develop WDHPs and achieved remarkable consequences. However, the development of WDHPs in China was still limited, which greatly constrained the promotion of DH in China. Furthermore, BaiduBaike was the largest Chinese encyclopedia platform at present (Shim and Yang, 2009), and it was very representative in the field of Chinese content creation, which provided a good way to investigate the development of WDHPs in China. Second, in order to be consistent with WDHPs oriented to content creation, selected entry must be closely related to the culture heritage field and has rich experience in content creation. Entry "Mogao Grottoes" belonged to the field of "Dun Huang," which was a famous cultural heritage domain with extensive social influence in China. Furthermore, entry "Mogao Grottoes" had a long editing history and was certified as a featured entry by BaiduBaike, which was similar to Wikipedia's




| (1) Data collection | (2) Analyzing the co-editing | (3) Consensus analysis | (4) Influential factors of consensus evolution |
|---|---|---|---|
| ✧ Identifying the entry<br>✧ Defining the period of discussion pages | ✧ Reorganizing statements<br>✧ Generating linked-structure<br>✧ PageRank algorithm | ✧ Skewness statistic<br>✧ Consensus evolution with time | ✧ Linear or curve fitting<br>✧ Comparing the degree of correlation($R^2$) |

**Figure 2.**
Overall analysis procedure

featured article. The number of contributors to this entry was 261 with 13 percent of whom had high user ratings, and the total number of editorial records was 304. As a result, the above fact indicated that entry "Mogao Grottoes" was suitable for this study.

The discussion pages of the entry served as the primary mechanism for managing co-editing activities in Wikipedia (Kittur *et al.*, 2007) and provided meta-data for this study. Similar to Wikipedia, BaiduBaike associated discussion pages with each entry's main page. Contributors commonly used the discussion pages to discuss, argue, and negotiate their views on the information that should be included in an entry's main page, and the content that should be accurate. Therefore, the discussion pages of "Mogao Grottoes" from April 20, 2006 to May 30, 2019 were completely collected, and the data items collected included: version number, entry content, contributor, update time and update reason. In the end, 304 records were obtained.

### *3.2 Linked-structure generation based on IBIS*

The linked-structure of a discussion could be generated on the basis of an entry's discussion pages, and the discussion pages contained a number of editorial records related to the entry. Each editorial record comprised a batch of editorial actions contributed by one editor at a time, involving several changes on the content of the entry, and such changes can be considered as statements in the discussion. At this point, the method of linked-structure generation was based on the reorganization of statements, and the reorganized statement information included the following items: statement number, contributor, time, issue, position, response node and response type, as shown in Table II (Aldea *et al.*, 2012).

Statement information was often implicit in the discussion pages. We reorganized them by taking three consecutive steps. First, we identified the main discussion issues (i.e. macro issue and micro issue) in the discussion process of an entry through an in-deep analysis of the discussion pages (Arazy *et al.*, 2013; Jones, 2008; Witte, 1983; Hadjerrouit, 2014). Second, editorial records related to an issue were retrieved from the editing history to form a data set, and multiple data sets corresponded to the identified issues. Editorial records of an issue were often distributed discontinuously across the discussion pages, and one editorial record always was involved in multiple issues. Third, the items of statement information were extracted based on the data set of a certain issue (Aldea *et al.*, 2012; Cao and Protzen, 1999). Besides, for the purpose of generating a linked-structure of a discussion, it was needed to identify the key discussion elements of a certain issue (i.e. issue, position and argument). Therefore, in terms of each issue, the statement with the earliest editing time was regarded as the representing statement of the issue, and statements with the response type being "Put Forward" were the representing statements of the positions. Then, the linked-structure constructed based on the IBIS model and was visualized by Python tool.

| Statement number | Contributor | Time | Issue | Position | Response node | Response type |
|---|---|---|---|---|---|---|
| 1 | Zho | April 22, 2006 | Entry topic | None | None | None |
| … | … | … | Entry topic | … | … | … |
| 3 | RobinLiang | April 23, 2006 | Entry topic | Life experience | None | Put forward |
| … | … | … | Entry topic | … | … | … |
| 11 | Yilong | September 10, 2007 | Entry topic | Life experience | 3 | Support |
| 12 | Yilong | September 10, 2007 | Entry topic | Academic philosophy | None | Put forward |
| … | … | … | Entry topic | … | … | … |
| 15 | Wu | December 12, 2007 | Entry topic | Life experience | 11 | Support |
| … | … | … | Entry topic | … | … | … |

**Table II.** Statements of issue "entry topic" of entry "Chenyinque"

Three independent coders who were undergraduate students at the Central China Normal University analyzed the discussion page of the entry, identified the discussion issues and reorganized the statements from editorial records. Our work started with a training session by studying a sample entry, namely, "DunHuang Fresco," where the three coders analyzed the entry's discussion page independently. In the training session, we established the operational procedures on various issues, developed guidelines for handling borderline cases to ensure that all the coders employed the same standard. Once the procedure and guidelines were established, coders analyzed the discussion pages of "Mogao Grottoes."

Table II showed an example of reorganized statements of entry "Chenyinque," and all the statements corresponded to issue "Entry topic." Statement 1 was the earliest one, and it was the representing statement of this issue. Its position, response node, and response type all was "None." There were two positions: "Life Experience" and "Academic Philosophy." Of them, the position of "Life Experience" was put forward by Statement 3. Statement 11 supported Statement 3 by adding text, and its response node was "3." Statement 12 proposed the position "Academic Philosophy" and its response node was "None." Then, Statement 15 added a reference to Statement 11. Its corresponding position was "Life Experience" and its response node was "11."

### *3.3 PageRank algorithm*

PageRank algorithm was used to analyze co-editing process aiming at content creation. PageRank algorithm was a link-based ranking metric, which was used to calculate the importance of a page through the relationship between pages. A fundamental idea behind PageRank was that a link from one page to another could be interpreted as a vote by a previous page for the following page. PageRank algorithm was confirmed to share the same essence of Condorcet's jury theory, namely, majority voting was a reliable indication of truth (Masterton, Olsson and Angere, 2016). PageRank value (PR value) was given by the following formula. Let PR($P_i$) be the PR value of the page $P_i$, $M_{Pi}$ be the number of outlinks of $P_i$, $L(p_j)$ be the number of inlinks of $P_i$, $n$ be the total number of all pages in the web, usually $\alpha = 0.85$ (Kleinberg, 1999; Palmer *et al.*, 2000):

$$\mathrm{PR}(P_i) = \frac{1-\alpha}{n} + \alpha \sum_{p_j \in M_{P_i}} \frac{\mathrm{PR}(p_j)}{L(p_j)}.$$

PageRank algorithm was commonly used to explore the decision making in a linked environment. For example, Palmer *et al.* (2000) explored trust on the WWW, Kleinberg (1999) searched authoritative sources in a hyperlinked setting. These studies suggested that PageRank algorithm could be used to analyze a vote-based group discussion process. Thus, it also could be utilized to analyze co-editing process. According to PageRank algorithm, one responding relationship from one statement to another (excluding disagreeing or unsupportive response, such as deletion) was a vote by a previous statement for the following. All the statements and their responding relationship would form a linked structure about an issue. Base on the generated linked-structure of statements, PageRank algorithm could be run, and PR values of all statements could be calculated. The higher the PR value of a statement was, the more important the statement was for a certain issue.

### *3.4 Skewness statistic*

Skewness statistic was used to analyze the consensus about an issue in the discussion of content creation. Consensus referred to the distribution of votes on statements toward an issue. Positive skewness value suggested that a small number of positions gained the most votes, and the remaining positions gained a small percentage of the votes. That is, positive

skewness value meant most votes focused on certain positions. The higher the value of skewness was, the higher the consensus level was (Herrera *et al.*, 1996). Lee *et al.* (2018) examined the formation of majority opinion on Facebook by using skewness statistic to describe the distribution of likes for a post. Srinivasan *et al.* (2015) used skewness statistic to identify similar topics on blogs and Facebook. These studies indicated that skewness statistic could be used to calculate the consensus value in group discussion. In this study, skewness statistic was performed based on the PR values of all positions toward an issue, and skewness statistic with time was used to describe the consensus evolution about an issue.

### *3.5 Breadth and depth of argumentation structure*

Breadth and depth of argumentation structure were measured for the purpose of analyzing their effects on the consensus evolution of content creation. The two variables were put forward by Chinn and Anderson (1998), and they proposed that they were the quantitative measurements of the way that participants employed in doing argumentation. Following Chinn and Anderson (1998), some studies confirmed that the two variables were closely related to the discussion outcomes, including the individual knowledge building (Leitão, 2000) and the completion of collaborative writing tasks (Munneke *et al.*, 2007), etc. From the perspective of group decision making, consensus building also represented the discussion outcome of content creation, it was reasonable to infer that breadth and depth of argumentation structure had some effects on consensus evolution.

According to previous literature, breadth referred to the extensiveness of the discussion, and depth was described as the extensively people elaborate distinct individual positions (Chinn and Anderson, 1998). In this study, breadth was defined as the number of all the nodes under an issue whose response type was "Put Forward," and depth was considered as the number of all argumentation sequence under an issue. Argumentation sequence referred to the argumentative elaboration that followed a position. For example, in the discussion process of an issue, three positions were put forward. Only one position was followed by argumentative elaboration which included a supportive statement, whereas the other two positions were not. In this case, the value of breadth was three, and the value of depth was one.

## 4. Finding

### *4.1 Description analysis*

Based on the discussion pages of "Mogao Grottoes," 90 discussion issues were extracted, a total of 608 reorganized statements were obtained, and the number of statements included in one issue from 2 to 35. It was clear that there were less statements on most issues. The issues with no less than eight statements were set as the issues' selecting standard, and the reason for setting the standard is as follows. Skewness statistic with time would be applied to describe consensus evolution in the following section. According to relevant literature, the skewness statistic indicated how much the distribution of a data set was skewed. When the data set had no less than three samples, the skewness analysis of this data set could be performed (Li, Qin, and Kar, 2010). After analyzing the data set of reorganized statements, it was found that 8 is the minimum number of statements which could satisfy the requirement of skewness analysis. According to this standard, 15 issues and 285 statements were finally selected. As shown in Table III, the maximal statement of an issue was 35, the minimum number was 8, and the maximal statement corresponded to issue "Scene." The maximal number of contributors under an issue was 23, corresponding to issue "Painted Sculpture" and the minimal number of contributors was 4. Among the 15 issues, there were 12 macro issues and 3 micro issues. Generally, there was an overall macro issue that included all the 15 issues, aiming at answering the question of "What content should be included in Mogao Grottoes," and it was named the overall macro issue in the following section.

| Issue | Description | N. of statements | N. of contributors | Issue type |
|---|---|---|---|---|
| Scene | The scenic spots of Mogao Grottoes | 35 | 14 | Macro |
| Damaging | Facts of man-made destruction suffered by Mogao Grottoes | 20 | 10 | Macro |
| Style | The artistic style of Mogao Grottoes during the dynasties in Chinese history | 22 | 9 | Macro |
| Artistic | Architecture, Painting, Painted Sculpture and Other Arts in Mogao Grottoes | 31 | 14 | Macro |
| Evolution | Construction history of Mogao Grottoes | 29 | 14 | Macro |
| Protection | Protection of Mogao Grottoes by China and other countries | 17 | 9 | Macro |
| Structure | Geographical location, distribution and composition of Mogao Grottoes | 15 | 6 | Macro |
| Stolen | Historical fact that Mogao Grottoes, as a world cultural heritage, was stolen | 15 | 9 | Macro |
| Grotto | The scenic spots of Grotto | 9 | 9 | Macro |
| Mural | Murals in Mogao Grottoes | 9 | 9 | Macro |
| Art of Painted Sculpture | Artistic of painted sculptures | 9 | 9 | Macro |
| Cave Structure | The structure nature of cave | 8 | 4 | Macro |
| N. of Painted Sculptures | Number of painted sculptures | 24 | 23 | Micro |
| N. of Caves | Number of caves discovered in the excavation of Mogao Grottoes | 19 | 18 | Micro |
| Excavation Site | The location where the Mogao Grottoes was excavated | 23 | 22 | Micro |

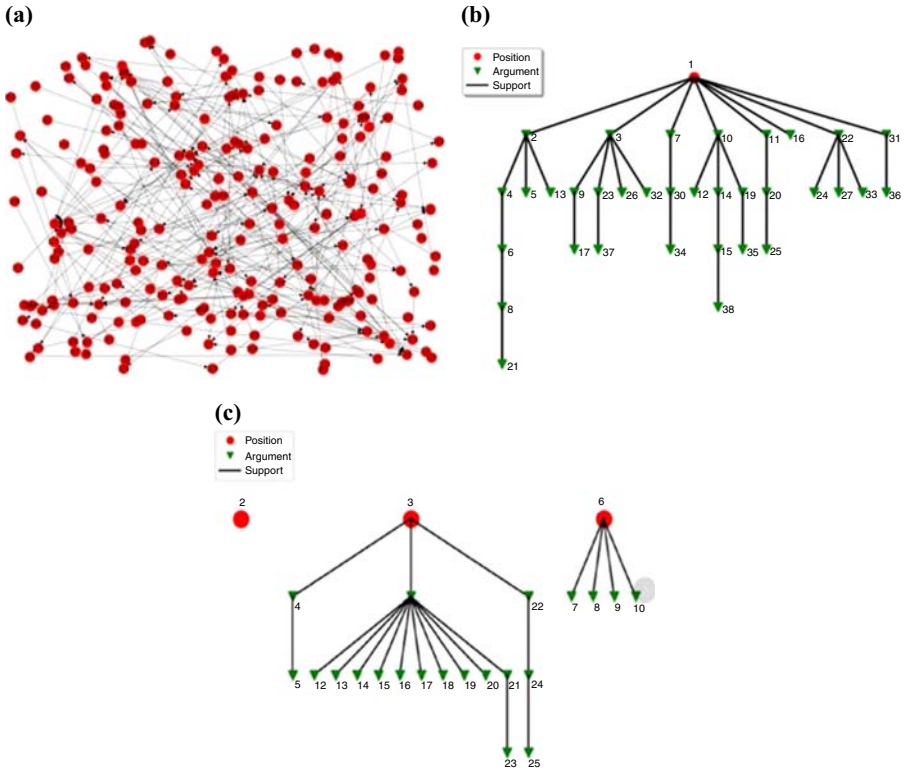
**Table III.** Description of 15 issues of "Mogao Grottoes"

*4.2 Analysis of co-editing process as group discussion*

*4.2.1 Macro issue discussion.* The discussion process of overall macro issue was analyzed first. To analyze this discussion process, we built the linked-structure with all statements of 15 issues and ran PageRank algorithm on it. Figure 3(a) showed the linked-structure of 15 issues, and the red nodes indicated the statements related to these 15 issues, and the black directed edges between the nodes represented the response relationships between the statements.

Based on the linked-structure, PageRank algorithm was used to analyze the discussion process, and the PR values of all statements were obtained. The finding showed that the statements representing 12 macro issues were at the top of all statements, and their PR values were as follows: "Scene"(0.10), "Damaging"(0.06), "Style"(0.07), "Artistic"(0.09), "Evolution"(0.08), "Protection"(0.05), "Structure"(0.05), "Stolen"(0.01), "Grotto"(0.003), "Mural"(0.005), "Painted Sculpture"(0.004) and "Cave Structure"(0.004), which indicated that these issues were agreed by all the contributors as the important content that should be included in "Mogao Grottoes." The higher the ranking was, the more important the issues were. Then, we compared these results with the latest version of "Mogao Grottoes" (BaiduBaike, 2019), and found that most of the above issues almost appeared in the entry's classification titles, and these with the high PR values are at the top of the entry content, suggesting the rationality of the above analysis.

Furthermore, the discussion process of issue "Scene" with the largest PR value was calculated. This issue addressed "What content should be included in the issue of Scene." As shown in Figure 3(b), the linked-structure with 35 statements under "Scene" issue was shown. Then, PageRank algorithm was run. The issues with relatively large PR values were shown as follows: "Grotto" (0.023), "Nine-Floor Building" (0.025), "Three-Floor Building" (0.015), "Sutra Depository Exhibition Hall" (0.024), "Mogao Grottoes Exhibition Hall" (0.011) and "Mogao Grottoes Artists" (0.008). The above results make some important sub-topics




**Notes:** (a) Linked-structure of overall macro issue; (b) linked-structure of issue "Scene"; (c) linked-structure of issue "N. of Painted Sculpture"

**Figure 3.**
Linked-structures of issues

emerge under the issue of "Scene." Similarly, this conclusion was almost consistent with the content of latest version of "Mogao Grottoes" (BaiduBaike, 2019).

*4.2.2 Micro issue discussion*. There were three micro issues in the sample, issue "Painted Sculpture" was selected as an example to exhibit the discussion process of micro issue. This issue focused on "How many painted sculptures have been discovered during the excavation of Mogao Grottoes." The reorganization of the statements revealed that there were three positions in the discussion: 2,000, 2,400 and 2,700. In total, 24 statements voted on these three positions. As shown in the Figure 3(c), the linked-structure was shown. Red nodes named "2," "3" and "6" referred to the positions of 2,000, 2,400 and 2,700, respectively. The result of PageRank showed that the maximum PR value (0.56) was observed at node "3," which was similar to the current content.

### *4.3 Consensus analysis of group discussion*

*4.3.1 Consensus evaluation*. In order to analyze the consensus level of group discussion, skewness statistic was conducted. For overall macro issue, the skewness statistic was run on the PR values of all statements. The result of skewness statistic was 5.16 which was much larger than 0, suggesting the co-editing efficiency of "Mogao Grottoes" is fine. Furthermore, skewness statistic was conducted on the 15 issues, respectively. The results showed that the consensus level of "Scene" (5.89), "Artistic" (5.53), "Evolution" (5.36),

"Damaging" (4.38), "Protection" (4.05), "Excavation Site" (4.01) and "N. of Painted Sculptures" (4.23) were higher than the average (MD = 3.89), indicating that the co-editing efficiency of these issues were better.

*4.3.2 Consensus evolution.* This study explored the consensus evolution with time by running PageRank algorithm and skewness statistic on the varying linked-structure in the co-editing process. For overall macro issue, the whole editing time was divided into six equal parts resulting in six accumulated stages of the consensus evolution. The corresponding PR values and skewness values of the six stages were obtained. Table IV exhibited that the skewness values of the six stages were 1.01, 3.31, 3.26, 4.48, 5.03 and 5.16, respectively, indicating that the consensus level took on a rise trend with a dramatic increase at early stage and a slight increase at the late stage (see Figure 4(a). Furthermore, the importance (i.e. PR value) ranking of the statements varied. For example, the ranking orders of "Scene" in six stages were the third, the fourth, the third, the first, the first and the first, respectively.

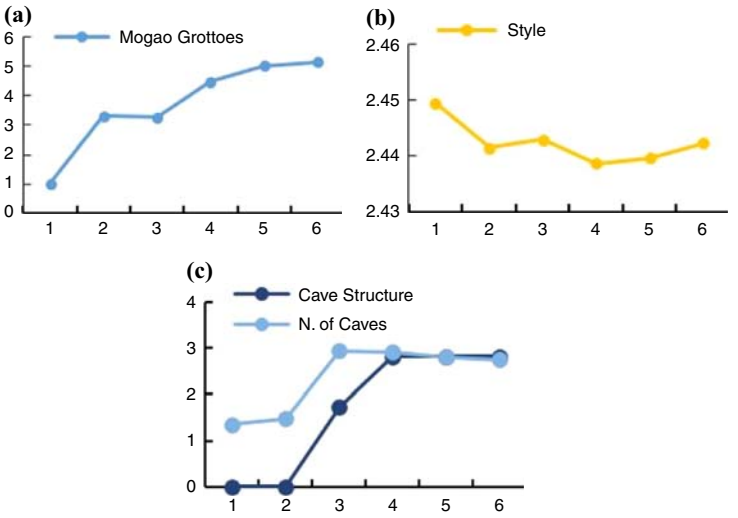
The consensus evolutions of 15 issues discussions were explored, respectively. The results showed that most consensuses of the 15 issues evolved in a same way as that of overall macro issue. Those of issue "Cave Structure" and issue "N. of Caves" only increased once (see Figure 4(c)), while the consensus of issue "Style" fluctuated significantly (see Figure 4(b)).

*4.3.3 Influencing factors of consensus evolution.* In order to clarify the differences in consensus evolution of various issues, the correlation between three influencing factors and the consensuses in six stages was explored by linear or curve fitting through the tool of Excel 2013, and the degrees of correlation ($R^2$) were calculated and compared with each other. The influencing factors included number of statements, breadth, and depth of argumentation structure.

First, the correlation between consensus evolution and number of statements in six stages was analyzed for the 15 issues. For issue "Scene," the result revealed a good logarithmic fit with $R^2 = 0.97$ (Figure 5(a). Table V showed that correlation analysis results

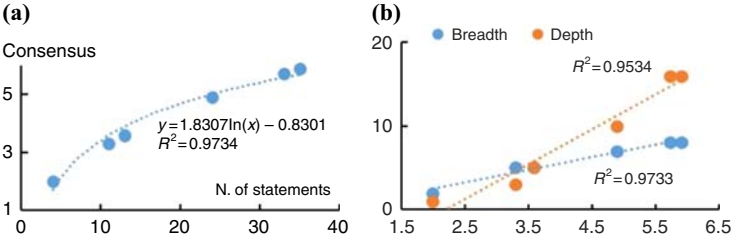
| Statement | PR | Statement | PR | Statement | PR |
|---|---|---|---|---|---|
| Stage1 | | Stage2 | | Stage3 | |
| Artist | 0.122 | Artist | 0.078 | Damaging | 0.071 |
| Structure | 0.093 | Style | 0.056 | Artist | 0.069 |
| Scene | 0.063 | Structure | 0.052 | Scene | 0.068 |
| Protection | 0.063 | Scene | 0.036 | Evolution | 0.065 |
| Evolution | 0.034 | Protection | 0.036 | Style | 0.059 |
| Style | 0.034 | Damaging | 0.036 | Structure | 0.045 |
| Damaging | 0.034 | Evolution | 0.029 | Protection | 0.033 |
| Structure | 0.005 | Excavation | 0.004 | Stolen | 0.009 |
| Construction | 0.005 | Mural | 0.004 | Grotto | 0.005 |
| Skewness value | 1.01 | Skewness value | 3.31 | Skewness value | 3.26 |
| Stage4 | | Stage5 | | Stage6 | |
| Scene | 0.108 | Scene | 0.120 | Scene | 0.104 |
| Artist | 0.100 | Artist | 0.103 | Artist | 0.092 |
| Style | 0.08 | Style | 0.080 | Evolution | 0.081 |
| Evolution | 0.076 | Evolution | 0.078 | Style | 0.066 |
| Damaging | 0.062 | Damaging | 0.057 | Damaging | 0.060 |
| Protection | 0.035 | Protection | 0.037 | Structure | 0.047 |
| Structure | 0.031 | Structure | 0.029 | Protection | 0.046 |
| Stolen | 0.008 | Stolen | 0.01 | Value | 0.024 |
| Mural | 0.005 | Mural | 0.005 | Stolen | 0.007 |
| Skewness value | 4.48 | Skewness value | 5.03 | Skewness value | 5.16 |

**Table IV.**
The result of the consensus evolution of overall macro issue




**Notes:** (a) Consensus evolution of overall macro issue; (b) consensus evolution of issue "Style"; (c) consensus evolution of issue "Cave Structure" and "N. of Caves"

**Figure 4.**
Consensus evolution of issues




**Notes:** (a) Correlations between consensus evolution and N. of statements of issue "Scene"; (b) correlations between consensus evolution and breadth or the depth of issue "Scene"

**Figure 5.**
Correlations between consensus evolution and No. of statements of issues

of the other 14 issues were same with that of issue "Scene." It was obvious that there was no significant difference among the 15 issues on the correlations of consensus evolution and number of statements.

Second, the correlation between breadth or depth of argumentation and the consensus evolution were explored. Figure 5(b) showed the result of issue "Scene." The orange line showed the correlation between consensus evolution and depth, and the blue line showed the correlation between consensus evolution and breadth. The results revealed good linear fittings with $R^2 = 0.95$ and $R^2 = 0.97$, respectively. However, the effect of breadth was slightly different from that of depth ($0.97 > 0.95$). The results of 15 issues exhibited differences (Table V). For issues "Scene," "Artistic," "Evolution," "Damaging" and "Stolen," both breadth and depth correlations with consensus evolution reached the significant level of $R^2 > 0.9$. For issue "Cave Structure," neither correlations were significant ($R^2 < 0.8$). For issues "Protection," "Structure," "Grotto," "Mural," "Painted Sculpture," "Excavation Site," "N. of Painted Sculptures" and "N. of Caves," only one correlation was significant.

**Table V.**
$R^2$ of correlations between influential factors and consensus of 15 issues

| Issue | Consensus | $R^2$ (Consensus vs N. of statements) | $R^2$ (Consensus vs depth) | $R^2$ (Consensus vs breadth) |
|---|---|---|---|---|
| Scene | 5.89 | 0.98 | 0.95 | 0.97 |
| Artistic | 5.53 | 0.97 | 0.99 | 0.99 |
| Evolution | 5.36 | 0.97 | 0.96 | 1.00 |
| Damage | 4.38 | 0.98 | 0.85 | 0.88 |
| Protection | 4.06 | 0.99 | 0.83 | 0.56 |
| Stolen | 3.86 | 0.96 | 0.99 | 0.89 |
| Structure | 3.83 | 0.94 | 0.78 | 0.85 |
| Grotto | 3.13 | 0.97 | 0.64 | 0.80 |
| Mural | 2.99 | 0.99 | 0.22 | 0.93 |
| Painted Sculpture | 2.98 | 0.995 | 0.83 | 0.69 |
| Cave Structure | 2.82 | 0.55 | 0.11 | 0.38 |
| Style | 2.44 | 0.35 | 0.26 | 0.82 |
| N. of Painted Sculpture | 4.23 | 0.95 | 0.98 | 0.50 |
| Excavation Site | 4.10 | 0.95 | 0.96 | 0.59 |
| N. of Caves | 2.76 | 0.85 | 0.89 | 0.49 |

## 5. Discussion

From the perspective of group discussion, this study has deeply explored co-editing mechanism of WDHPs oriented to context creation. The main conclusions can be revealed including the following three points, which are invaluable for promoting and enhancing WDHPs.

### *5.1 Co-editing process and vote-based group discussion*

Corresponding to the question of "How the contextual content of WDHPs is created by co-editing?," we consider the co-editing process as a large-scale group discussion activity based on the discussion element of issue. Drawing upon the idea of Murr (2015), PageRank algorithm is used to analyze this process, and related findings are helpful in enhancing the co-editing efficient in WDHPs.

First, according to the group discussion theory, issue is the starting point to understand co-editing process aiming at content creation, including two types, macro issue and micro issue. Macro issue is responsible for the construction of hierarchical topical structure, whereas micro issue takes charge of the appropriateness of the mini content unit, such as spelling, picture, linking, etc. In the context of "Mogao Grottoes," after reorganizing statement information, we have found that there are more macro issues in the formation process of the entry than micro issues ($12 > 3$), which are consistent with the findings of Jones (2008). This result indicates that contributors are more involved in discussions on macro issues than in discussions on micro issues during the co-editing process. Twice analysis of co-editing process of macro issues in our study have confirmed the results of previous studies that macro issue has a close relationship with topical structure of an entry and the topical structure is hierarchical (Witte, 1983; Jones, 2008). Furthermore, micro issue is considered as wicked problem concerned by IBIS, and focuses on the question of "How to describe specific knowledge elements appropriately?" (Aldea *et al.*, 2012). In this paper, "How many painted sculptures have been excavated in Mogao Grottoes" is an example of micro issue. Its final conclusion is "about 2,400," which is a more appropriate description than "2,400."

Second, in the context of an issue, the vote-based group discussion proceeds in two steps, namely, putting forward position and position argumentation. Among them, the step of putting forward position is responsible for proposing the alternatives/options of the issue, involving such editorial actions as adding link, and adding information, etc. The step of

position argumentation is a response to previous position, providing proof, supplement or elaboration, and involving such editorial actions as clarifying information, fixing link, correcting misspelling, etc. (Hadjerrouit, 2014). These actions of position argumentation (excluding deleting action) show a same attitude of supporting the corresponding position, so as to be regarded as affirmative votes for the position. Analysis of co-editing process by applying PageRank algorithm results in some rational results, which extend the opinion of Murr (2015) by confirming a large-scale discussion existing in the co-editing process of WDHPs oriented to content creation.

These above findings will be valuable for developing WDHPs by applying co-editing in an appropriate way. For example, the results regarding the discussion element of issue provide a possible way to enhance study design and data collection guideline of a WDHP. Like the common public participated science programs, WDHPs vary greatly with regard to primary objectives, which indicates that creating contextual content must following the specific objectives. Some studies from citizen science domain indicate that collaboration between professional scientists and public contributors could lead to data collection schemes focusing on specific research questions (Crall *et al.*, 2010; Stohlgren and Schnase, 2006). Therefore, the appropriateness of content creation can be ensured by encouraging professional scientists to set the proper issues and engaging many contributors in co-editing them.

### *5.2 Co-editing efficiency and consensus*

Traditionally, consensus means a strict and unanimous agreement of all the participants regarding all possible alternatives during a discussion process (Ben-Arieh and Chen, 2006; Meng and Chen, 2015). In the context of content creation, consensus can be defined as the outcome of co-editing activities. In the words of Pfeil *et al.* (2006), online encyclopedia aims to build consensus among contributors about different topics. Furthermore, the value of consensus can be used to measure the efficiency of co-editing, and a high consensus value corresponds to fine collaborative efficiency, which indicates that the generated content is a final decision that majority contributors can support (Cabrerizo *et al.*, 2010). For the purpose of ensuring the well-work of WDHPs, professional scientists should monitor the progress of projects (Bonter and Cooper, 2012; Cooper *et al.*, 2007). In this sense, the method of consensus analysis offers a way to obtain the information of co-editing efficiency.

Taking "Mogao Grottoes" as an example, the consensus value of overall macro issue is 5.16, and the consensus evolution holds a growth trend in the process of co-editing. Compared with the result of the previous study of the majority opinion formation on Facebook, the consensus value of content creation about "Mogao Grottoes" is larger, suggesting its co-editing efficiency is finer (Lee *et al.*, 2018). Furthermore, the findings of consensus analysis of 15 issues show that the development trends of their consensus are almost similar, but their co-editing efficiencies vary greatly, and the 15 issues can fall into two categories. The first category includes "Scene," "Artist," "Evolution," "Protection," "Structure," "Damaging" and their consensus values are higher. The second category includes "Structure," "Grotto" and "Mural," and their consensus values are lower. The result indicates that the co-editing efficiency of the first category is desirable, while the co-editing efficiency of the second category is less satisfactory. Therefore, to improve the content creation of "Mogao Grottoes," more effective co-editing on the second category remains to be further explored.

### *5.3 Influencing factors of co-editing efficiency and argumentation structure*

Starting from the requirement of improving the co-editing efficiency of WDHPs, the influencing factors are explored. The finding indicates that the correlation between statement number and consensus evolution forms a satisfactory logarithmic fitting

($R^2 > 0.9$), which is consistent with the result of previous study (Lee *et al.*, 2018). However, it is not evident that there exists significant difference between 15 issues on the influence of statement number. Based on the theory of argumentation, this paper examines two influential factors, breadth and depth of argumentation structure, leading to some valuable results (see Table IV). First, among the 12 macro issues, most consensus values benefit more from breadth than from depth with the exception of "Painted Sculpture," "Protection" and "Stolen." Second, for the issues with higher consensus values (i.e. consensus value > 0.5), such as "Scene," "Artistic," "Evolution," the combination of breadth and depth is more effective.

According to citizen science theory, adhering closely to the scientific method/procedure in managing the project is crucial for WDHPs (Loss *et al.*, 2015), and professional scientists should maintain two-way communications with WDHPs and regularly provide useful feedbacks (Bonter and Cooper, 2012; Chandler *et al.*, 2012; Cooper *et al.*, 2007). Therefore, the approach of exploring the influential factors of co-editing efficiency enables the managers of WDHPs to obtain useful feedbacks. In terms of "Mogao Grottoes," its content creation is more affected by breadth than by depth of argumentation structure, which leads to a feedback that editorial actions in a depth manner should be encouraged and guided, especially for the issues of "Structure," "Grotto" and "Mural."

## 6. Conclusion

This study is among the first attempts to reveal co-editing mechanism of WDHPs from the perspective of group discussion and to investigate the efficiency of co-editing by analyzing the consensus and its influential factors. Related findings can be valuable for the development of WDHPs and other similar crowdsourcing projects.

Co-editing in WDHPs is an example of crowdsourcing activity, which has become a powerful mechanism to harvest collective wisdom from thousands of social participants (Howe, 2008). Booming crowdsourcing trend encouraged a stream of research focused on the collaborative mechanism. However, most of the existing literature focus on the collaborative mechanism at an individual level, such as behaviors and motivations (Arazy and Gellatly, 2013). This paper is concerned with collaborative mechanism in WDHPs at a group level, and highlights the process and consensus of content creation. At this point, the findings presented in this paper extend related research.

This research contributes a three-step analytic procedure which illustrates how content is created by co-editing, how co-editing efficiency is evaluated and how to enhance co-editing efficiency by adjusting its influential factors. In particular, the basic approach of constructing linked-structure of discussion statements in this procedure is distinct and efficient, which is the first attempt to apply IBIS theory in a large-scale online discussion. The application of PageRank algorithm and skewness statistics in the consensus analysis is rational, which not only has a strong theoretical basis, but also is confirmed by relative results. Taken together, the three-step analytic procedure is a highly effective approach of presenting, monitoring and guiding the co-editing process, which can work in multiple contexts of WDHPs and on different wiki platforms.

This study is of important implications for DH researchers. For managers of WDHPs, the application of online encyclopedia platform combined with consensus analysis is beneficial. Managers can start a WDHP by launching an entry on the encyclopedia platform, and obtain co-editing efficiency through consensus analysis to ensure the project's good work. For designers of WDHPs, the elements related to argumentation structure can be utilized to optimize the design to guide contributors to participate in co-editing of content creation in an effective manner. Besides, our analytic procedure can be helpful in finding the interest of contributors in a certain DH field, which will be valuable for the researchers who are interested in this field. For example, the possible difference between the interest of

contributors and the interest of experts can be explored to promote some potential research topics in specific DH domains.

The limitation of our study is focusing on a unique entry (i.e. Mogao Grottoes), which provides a useful example for the preliminary validation. Further research is needed to test whether our findings is valid for large number of entries, and to reveal whether the difference exists between various types of entries, such as information entry and knowledge entry. In addition, other studies are suggested to explore other forms of wiki-based collaborative activities in DH domain based on the analytic procedure of this study. However, the method of reorganizing statement information is restricted by the nature of our sample (i.e. BaiduBaike.com), therefore, how to optimize the information reorganization method according to research object is a further research issue.

## References

Aldea, A., Bañares-Alcántara, R. and Skrzypczak, S. (2012), "Managing information to support the decision making process", *Journal of Information & Knowledge Management*, Vol. 11 No. 3, pp. 1250016-1-1250016-13.

Arazy, O. and Gellatly, I.R. (2013), "Corporate wikis: the effects of owners' motivation and behavior on group members' engagement", *Journal of Management Information Systems*, Vol. 29 No. 3, pp. 87-116.

Arazy, O. and Kopak, R. (2010), "On the measurability of information quality", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 1, pp. 89-99.

Arazy, O., Yeo, L. and Nov, O. (2013), "Stay on the Wikipedia task: when task-related disagreements slip into personal and procedural conflicts", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 8, pp. 1634-1648.

BaiduBaike (2019), "The latest version of Mogao Grottoes", available at: https://baike.baidu.com/item/莫高窟/303038 (accessed July 20, 2019).

Ben-Arieh, D. and Chen, Z. (2006), "Linguistic group decision-making: opinion aggregation and measures of consensus", *Fuzzy Optimization and Decision Making*, Vol. 5 No. 4, pp. 371-386.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. and Shirk, J. (2009), "Citizen science: a developing tool for expanding science knowledge and scientific literacy", *BioScience*, Vol. 59 No. 11, pp. 977-984.

Bonter, D.N. and Cooper, C.B. (2012), "Data validation in citizen science: a case study from Project Feeder Watch", *Frontiers in Ecology and the Environment*, Vol. 10 No. 6, pp. 305-307.

Bradley, L., Lindstrom, B., Rystedt, H. and Vigmo, S. (2010), "Language learning in a wiki: student contributions in a web based learning environment", *Themes in Science and Technology Education*, Vol. 3 Nos 1-2, pp. 63-80.

Cabrerizo, F.J., Moreno, J.M., Pérez, I.J. and Herrera-Viedma, E. (2010), "Analyzing consensus approaches in fuzzy group decision making: advantages and drawbacks", *Soft Computing*, Vol. 14 No. 5, pp. 451-463.

Cao, Q. and Protzen, J.-P. (1999), "Managing design information: issue-based information systems and fuzzy reasoning system", *Design Studies*, Vol. 20 No. 4, pp. 343-362.

Chandler, M., Bebber, D.P., Castro, S., Lowman, M.D., Muoria, P., Oguge, N. and Rubenstein, D.I. (2012), "International citizen science: making the local global", *Frontiers in Ecology and the Environment*, Vol. 10, pp. 328-331.

Chinn, C.A. and Anderson, R.C. (1998), "The structure of discussions that promote reasoning", *Teachers College Record*, Vol. 100 No. 2, pp. 315-368.

Chowdhury, W., Burt, C., Akkaoui, A. and Davies, J. (2015), "Quanty: an online game for eliciting the wisdom of the crowd", *Computers in Human Behavior*, Vol. 49, pp. 213-219.

Cohen, J. (1996), "Procedure and substance in deliberative democracy", in Benhabib, S. (Ed.), *Democracy and Difference: Contesting the Boundaries of the Political*, Princeton University Press, Princeton, NJ, pp. 95-119.

Cooper, C.B., Dickinson, J., Phillips, T. and Bonney, R. (2007), "Citizen science as a tool for conservation in residential ecosystems", *Ecology and Society*, Vol. 12 No. 2, available at: www.ecologyandsociety.org/vol12/iss2/art11

Crall, A.W., Newman, G.J., Jarnevich, C.S., Stohlgren, T.J., Waller, D.M. and Graham, J. (2010), "Improving and integrating data on invasive species collected by citizen scientists", *Biological Invasions*, Vol. 12 No. 10, pp. 3419-3428.

Cunningham, W. and Leuf, B. (2001), *The Wiki Way – Quick Collaboration on the Web*, Addison-Wesley Longman Publishing Co., Inc., Boston.

Faigley, L. and Witte, S. (1981), "Analyzing revision", *College Composition and Communication*, Vol. 32 No. 4, pp. 400-414.

Follett, R. and Strezov, V. (2015), "An analysis of citizen science based research: usage and publication patterns", *PLOS One*, available at: https://doi.org/10.1371/journal.pone.0143687

Forte, A., Larco, V. and Bruckman, A. (2009), "Decentralization in Wikipedia governance", *Journal of Management Information Systems*, Vol. 26 No. 1, pp. 49-72.

Hadjerrouit, S. (2014), "Wiki as a collaborative writing tool in teacher education: evaluation and suggestions for effective use", *Computers in Human Behavior*, Vol. 2014 No. 32, pp. 301-312.

He, F. and Han, S. (2006), "A method and tool for human–human interaction and instant collaboration in CSCW-based CAD", *Computers in Industry*, Vol. 57 Nos 8-9, pp. 740-751.

Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1996), "A model of consensus in group decision making under linguistic assessments", *Fuzzy Sets and Systems*, Vol. 78 No. 1, pp. 73-87.

Howe, J. (2008), *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*, Random House, New York, NY, p. 311.

Jimoyiannis, A. and Roussinos, D. (2017), "Students' collaborative patterns in a wiki-authoring project", *Journal of Applied Research in Higher Education*, Vol. 9 No. 1, pp. 24-39.

Jones, J. (2008), "Patterns of revision in online writing", *Written Communication*, Vol. 25 No. 2, pp. 262-289.

Karacapilidis, N. and Pappis, C. (2000), "Computer-supported collaborative argumentation and fuzzy similarity measures in multiple criteria decision making", *Computers & Operations Research*, Vol. 27 Nos 7-8, pp. 653-671.

Kimmerle, J., Moskaliuk, J., Harrer, A. and Cress, U. (2010), "Visualizing co-evolution of individual and collective knowledge", *Information, Communication & Society*, Vol. 13 No. 8, pp. 1099-1121.

Kittur, A. and Kraut, R.E. (2008), "Harnessing the wisdom of crowds in Wikipedia: quality through coordination", *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW 2008, ACM, San Diego, CA, November 8–12.*

Kittur, A., Chi, E.H. and Suh, B. (2009), "What's in Wikipedia: mapping topics and conflict using socially annotated category structure", *Proceedings of the 27th International Conference on Human Factors in Computing Systems – CHI '09, ACM, Boston, MA, April 4-9.*

Kittur, A., Suh, B., Pendleton, B.A. and Chi, E.H. (2007), "He says, she says: Conflict and Coordination in Wikipedia", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '07, San Jose, CA, April 28-May 3.*

Kleinberg, J.M. (1999), "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, Vol. 46 No. 5, pp. 604-632.

Kunz, W. and Rittel, H.W.J. (1970), "Issues as elements of information systems", Working Paper No. 131, Center for Planning and Development Research, University of California, Berkeley.

Langley, D.J., Hoeve, M.C., Ortt, J.R., Pals, N. and van der Vecht, B. (2014), "Patterns of herding and their occurrence in an online setting", *Journal of Interactive Marketing*, Vol. 28 No. 1, pp. 16-25.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Albert-László, B., Brewer, D. *et al.* (2009), "Computational social science", *Science*, Vol. 323 No. 1, pp. 721-723.

Lee, S., Ha, T., Lee, D. and Kim, J.H. (2018), "Understanding the majority opinion formation process in online environments: an exploratory approach to Facebook", *Information Processing & Management*, Vol. 54 No. 6, pp. 1115-1128.

Leitão, S. (2000), "The potential of argument in knowledge building", *Human Development*, Vol. 43 No. 6, pp. 332-360.

Li, W.D., Ong, S.K., Fuh, J.Y.H., Wong, Y.S., Lu, Y.Q. and Nee, A.Y.C. (2004), "Feature-based design in a distributed and collaborative environment", *Computer-Aided Design*, Vol. 36 No. 9, pp. 775-797.

Li, X., Qin, Z. and Kar, S. (2010), "Mean-variance-skewness model for portfolio selection with fuzzy returns", *European Journal of Operational Research*, Vol. 202 No. 1, pp. 239-247.

Lipnack, J. and Stamps, J. (1997), *Virtual Teams: Reaching Across Space, Time, and Organizations with Technology*, John Wiley & Sons, New York, NY.

Loss, S.R., Loss, S.S., Will, T. and Marra, P.P. (2015), "Linking place-based citizen science with large-scale conservation research: a case study of bird-building collisions and the role of professional scientists", *Biological Conservation*, Vol. 2015 No. 184, pp. 439-445.

Loui, R.P. (2010), "Process and policy: resource-bounded nondemonstrative reasoning", *Computational Intelligence*, Vol. 14 No. 1, pp. 1-38.

Masterton, G., Olsson, E.J. and Angere, S. (2016), "Linking as voting: how the Condorcet jury theorem in political science is relevant to webometrics", *Scientometrics*, Vol. 106 No. 3, pp. 945-966.

Meng, F. and Chen, X. (2015), "A new method for group decision making with incomplete fuzzy preference relations", *Knowledge-Based Systems*, Vol. 73, pp. 111-123.

Miller, N. (2005), "Wikipedia and the disappearing 'author'", *ETC: A Review of General Semantics*, Vol. 62 No. 1, pp. 37-40.

Munneke, L., Andriessen, J., Kanselaar, G. and Kirschner, P. (2007), "Supporting interactive argumentation: influence of representational tools on discussing a wicked problem", *Computers in Human Behavior*, Vol. 23 No. 3, pp. 1072-1088.

Murr, A.E. (2015), "The wisdom of crowds: applying condorcet's jury theorem to forecasting US presidential elections", *International Journal of Forecasting*, Vol. 31 No. 3, pp. 916-929.

Netzer, O., Feldman, R., Goldenberg, J. and Fresko, M. (2012), "Mine your own business: market-structure surveillance through text mining", *Marketing Science*, Vol. 31 No. 3, pp. 521-543.

Nov, O. (2007), "What motivates Wikipedians?", *Communications of the ACM*, Vol. 50 No. 11, pp. 60-64.

Oomen, J. and Aroyo, L. (2011), "Crowdsourcing in the cultural heritage domain: opportunities and challenges", *Proceedings of the Fifth International Conference on Communities and Technologies, C&T'11, ACM, Brisbane, June 29-July 2.*

Palau, R.M. and Moens, M.F. (2009), "Argumentation mining: the detection, classification and structuring of arguments in text", *Proceedings of the 12th International Conference on Artificial Intelligence & Law, ACM, Barcelona, June 8-12.*

Palmer, J.W., Bailey, J.P. and Faraj, S. (2000), "The role of intermediaries in the development of trust on the www: the use and prominence of trusted third parties and privacy statements", *Journal of Computer-Mediated Communication*, Vol. 5 No. 3, available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2000.tb00342.x/full

Patton, M.Q. (1987), *How to Use Qualitative Methods in Evaluation*, Sage, London.

Pfeil, U., Zaphiris, P. and Ang, C.S. (2006), "Cultural differences in collaborative authoring of Wikipedia", *Journal of Computer-Mediated Communication*, Vol. 12 No. 1, pp. 88-113.

Shim, J.P. and Yang, J. (2009), "Why is wikipedia not more widely accepted in korea and china? Factors affecting knowledge-sharing adoption", *Decision Line*, Vol. 40 No. 2, pp. 12-15.

Spoerri, A. (2007), "What is popular on Wikipedia and why?", *First Monday*, Vol. 12 No. 4, available at: https://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1765/1645

Srinivasan, B.V., Natarajan, A., Sinha, R. and Sinha, M. (2015), "Community reaction: from blogs to Facebook", *Proceedings of the Second ACM IKDD Conference on Data Sciences – CoDS '15*, *Bangalore*, pp. 69-74.

Stohlgren, T.J. and Schnase, J.L. (2006), "Risk analysis for biological hazards: what we need to know about invasive species", *Risk Analysis*, Vol. 26 No. 1, pp. 163-173.

Travis, C. and Holm, P. (2016), "The digital environmental humanities – what is it and why do we need it? The NorFish project and SmartCity Lifeworlds", in Travis, C. and von Lünen, A. (Eds), *The Digital Arts and Humanities*, Springer International Publishing, pp. 187-204.

Wagner, C. (2005), "Breaking the knowledge acquisition bottleneck through conversational knowledge management", *Information Resources Management Journal*, Vol. 19 No. 1, pp. 70-83.

Webb, P., Whitlow, J.W. and Venter, D. (2016), "From exploratory talk to abstract reasoning: a case for far transfer?", *Educational Psychology Review*, Vol. 29 No. 3, pp. 565-581.

Witte, S.P. (1983), "Topical structure and revision: an exploratory study", *College Composition and Communication*, Vol. 34 No. 3, pp. 313-341.

Zhang, X., Fang, Y., He, W., Zhang, Y. and Liu, X. (2018), "Epistemic motivation, task reflexivity, and knowledge contribution behavior on team wikis: a cross-level moderation model", *Journal of the Association for Information Science and Technology*, Vol. 35 No. 1, pp. 1-14.

Zhao, Y. (2017), "A preliminary exploration on citizen science projects based on scientific crowdsourcing perspectives: conceptualization, pattern design and research opportunities", *Journal of Library Science in China*, Vol. 43 No. 5, pp. 42-56.

**Corresponding author**
Ming Yi can be contacted at: yiming0415@mail.ccnu.edu.cn