This is a preprint of Zhao, D., & Strotmann, A. (2021). Mapping Knowledge Domains on Wikipedia: An author bibliographic coupling analysis of Traditional Chinese Medicine. *Journal of Documentation.* Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JD-02-2021-0039

Mapping Knowledge Domains on Wikipedia: An author bibliographic coupling analysis of Traditional Chinese Medicine

Abstract

Purpose. Wikipedia has the lofty goal of compiling all human knowledge. The purpose of the present study is to map the structure of the Traditional Chinese Medicine (TCM) knowledge domain on Wikipedia, to identify patterns of knowledge representation on Wikipedia, and to test the applicability of author bibliographic coupling analysis (ABCA), an effective method for mapping knowledge domains represented in published scholarly documents, for Wikipedia data.

Design/methodology/approach. We adapted and followed the well-established procedures and techniques for author bibliographic coupling analysis (ABCA). Instead of bibliographic data from a citation database, we used all articles on TCM downloaded from the English version of Wikipedia as our dataset. An author bibliographic coupling network was calculated and then factor analyzed using SPSS. Factor analysis results were visualized. Factors were labeled upon manual examination of articles that authors who load primarily in each factor have significantly contributed references to. Clear factors were interpreted as topics.

Findings. Seven TCM topic areas are represented on Wikipedia, among which Acupuncture related practices, Falun Gong, and Herbal Medicine attracted the most of significant contributors to TCM. Acupuncture and Qi Gong have the most connections to the TCM knowledge domain, and also serve as bridges for other topics to connect to the domain. Herbal medicine is weakly linked to and non-herbal medicine is isolated from the rest of the TCM knowledge domain. It appears that specific topics are represented well on Wikipedia but their conceptual connections are not. ABCA is effective for mapping knowledge domains on Wikipedia but document-based bibliographic coupling analysis is not.

Originality/value. Given the prominent position of Wikipedia for both information users and for researchers on knowledge organization and information retrieval, it is important to study how well knowledge is represented and structured on Wikipedia. Such studies appear largely missing although studies from different perspectives both about Wikipedia and using Wikipedia as data are abundant. Author bibliographic coupling analysis is effective for mapping knowledge domains represented in published scholarly documents but has never been applied to mapping knowledge domains represented on Wikipedia.

Keywords

Wikipedia, mapping knowledge domains, author bibliographic coupling analysis, citation analysis, bibliometrics, Traditional Chinese Medicine

Introduction

Wikipedia has the lofty goal of compiling all human knowledge. This gigantic open encyclopaedia created and maintained collectively by volunteers world-wide has now become a one-stop shop for information on pretty much any topic. As Wikipedia articles often appear at the top of Google search result lists and are now promoted as fact check sources on Facebook, to name just two prominent examples, Wikipedia is clearly a primary information source that people see, and in many cases even the only source that people consult. Wikipedia's content (articles and categories) has also been used as knowledge base for artificial intelligence (AI)-enhanced knowledge representation and organization

Given the prominent position that Wikipedia has for both information users and for researchers on knowledge organization and information retrieval, it is important to study how well knowledge is represented and structured on Wikipedia.

Bibliometrics is effective in studying scholarly communication patterns and mapping knowledge domains represented in published scholarly documents. The present study explores what we may learn about knowledge domains represented on Wikipedia from applying bibliometric methods to Wikipedia data. In particular, we adapted and applied author bibliographic coupling analysis (Zhao and Strotmann, 2008) to examine one of the many topic areas on Wikipedia: Traditional Chinese Medicine (TCM). The objective is to map the structure of this knowledge domain on Wikipedia and to identify patterns of knowledge representation on Wikipedia.

Background and related studies

Wikipedia is a system unique in the history of civilization (Simonite, 2013). Both benefits and challenges of the Wikipedia system have been widely debated in academia, law, business, and other sectors of society. Studies from different perspectives both about Wikipedia and using Wikipedia as data are abundant (Yasseri, *et al*., 2012). These studies were possible partly because nearly every edit and discussion post are saved and available on Wikipedia.

Wikipedia started in 2001 with the lofty goal of compiling all human knowledge. It grew quickly into the largest encyclopaedia in the world (Burke and Kraut, 2008). As of January 2021, Wikipedia has over 40.7 million registered and uncounted unregistered volunteer editors and over 6.2 million articles in the English version alone (Wikipedia:Statistics). Many Wikipedia articles were found to be of a quality comparable with corresponding ones in Encyclopaedia Britannica (Giles, 2005). Wikipedia's success also made it play a symbolic role in highlighting the potential for voluntary peer-production to generate valuable collections of information. Hansen, *et al.,* (2009) even contend that Wikipedia "approximates features of the ideal speech situation articulated by Habermas" in his Theory of Communicative Actions (Habermas, 1984).

However, Wikipedia's success was to many a surprise. Among the fundamental problems of the Wikipedia system that have been criticized are unpredictable motivations (and competences) of editors and an emphasis on consensus rather than authority (Denning *et al.,* 2005). Wikipedia evolves without supervision by certified subject experts or authorities, and its largely anonymous volunteer editors are left both to select, write about and organize the topics it includes and to define, interpret and implement its policies and resolve conflicts on their own. Wikipedia editors may be knowledge domain experts or an elementary school student, and "may be altruists, political or commercial opportunists, practical jokers, or even vandals" (Denning *et al.,* 2005, p. 152). Inaccuracy or errors may exist due to lack of supervision by certified subject experts. Bias can be introduced and maintained as long as a group of editors with that bias manage to dominate the discussion and force it to a "consensus" (Das, *et al.,* 2016; Yasseri, *et al.,* 2012). Mechanisms used in traditional systems to ensure quality and avoid abuse of power are normally based on true identity along with social expectations, norms, and status positions, and thus cannot work for Wikipedia (Arazy *et al.,* 2011; Ransbotham and Kane, 2011). The incident when Wikipedia rejected an entry about Donna Strickland, a Canadian female winner of the 2018 Nobel Prize in Physics, half year before the announcement of the prize is just one of the many examples of problems in Wikipedia's topic selection policy and practices (The Guardian, 2018).

Despite these problems, as the world's largest open encyclopaedia, the Wikipedia's content (articles and categories) "has been used extensively for tasks like entity disambiguation or semantic similarity estimation" to enhance AI-based knowledge representation and organization (Heist and Paulheim, 2019). Wikipedia's human-annotated categories (Kittur, *et al.*, 2009) have been found to be "very loosely structured" (Perez, 2021). Many studies explored how to automatically categorize Wikipedia articles (e.g., Refaei *et al,* 2018; Perez, 2021; Simone Paolo Ponzetto and Strube, 2007; Strube, and Simone Paolo Ponzetto, 2006; Gantner and Schmidt-Thieme, 2009).

A systematic review (Mesgari *et al.,* 2014) grouped studies about Wikipedia into six main areas of inquiry: general Wikipedia studies, infrastructure, content, participation, readership, and corpus. The content category is mostly focused on the quality (e.g., comprehensiveness, currency, readability, and reliability) and size of Wikipedia. For example, Sundin (2011) examined the everyday practices of Wikipedia editors (participation) and Francke and Sundin (2010) studied credibility in Wikipedia (quality). Studies on structures of knowledge domains on Wikipedia have largely been missing. Related studies are those on comprehensiveness of Wikipedia content which often use samples of articles to compare topic representation between Wikipedia and recognized traditional sources. For example, Wedemeyer *et al.* (2008) randomly picked 446 articles from *Encyclopædia Britannica*, and checked if Wikipedia articles had entries for them. Clauson, *et al.* (2008) compared medical drug information between Wikipedia and a traditionally edited database, Medscape Drug Reference.

Also related to the present study are studies on automatic categorization of Wikipedia articles. These studies focus mostly on developing computer algorithms to categorize the entire Wikipedia and have little interest in examining any knowledge domains closely. Das *et al.* (2016), for example, categorized Wikipedia articles based on a topic similarity measure that combines the degree to which two articles share references, linked terms, and Wikipedia topic categories. Kittur, *et al.,* (2009) mapped the distribution of topics on Wikipedia using its own categories.

The present study applies author bibliographic coupling analysis to examine the structure of the TCM knowledge domain closely. We chose TCM for this study because it has been reported to be one of the most controversial topic areas on Wikipedia (Koppelman, 2017; McLuhan, 2013) in which Wikipedia's problems in topic selection and treatment may be more pronounced.

Methodology

We adapted and followed the well-established procedures and techniques for Author bibliographic coupling analysis (Zhao and Strotmann, 2008a). Instead of using a citation database, we used the English version of Wikipedia as data source, and developed computer programs to collect and analyze data from Wikipedia.

Author bibliographic coupling analysis (ABCA)

Bibliographic coupling is a method that uses the number of cited references shared by two scholarly articles to measure how closely these two articles are related in terms of topics or methodological approaches (Kessler, 1963).

ABCA uses the author instead of the article as the unit of analysis, and has been found to have a number of advantages compared to article-based analysis (Zhao and Strotmann, 2008a). One of these advantages is that its wider granularity and the broader information that an author represents can smooth out the impact of outliers on the analysis. Authors represent schools of thought whereas articles represent individual pieces of evidence for or findings about concepts, theories or methods. Individuals tend to develop a set of information sources that they prefer to consult when they contribute to knowledge production (e.g., writing scholarly or Wikipedia articles) or deal with work or life problems (White, 2001). The more information sources two individuals share, the more closely their interests and beliefs may be related.

The degree to which articles share references has also been used as part of topic similarity measures between Wikipedia articles (e.g., Das *et al.,* 2016). Verifiability, one of the two fundamental principles for contributing and for resolving conflicts that Wikipedia has developed, intends to ensure that all important information and viewpoints are supported by trustworthy published resources (i.e., cited references). Studies have indeed found that information on Wikipedia was largely supported by clearly identifiable and reputable resources (Haigh, 2010; Rector, 2008) although sources cited tend to be short summary type of online resources and do not represent all voices of experts outside of the Wikipedia community (Luyt, 2012). Author-based analysis of shared references is yet to be studied.

Data collection and analysis

We identified articles on TCM from Wikipedia by starting with the articles on TCM proper and with articles under the sub-categories and sub-sub-categories of TCM. We downloaded all these articles from Wikipedia in late 2019, including the entire editing and discussion history of each article.

Following ABCA techniques, we chose the top 500 editors who have contributed the most to the downloaded articles to represent this knowledge domain. A matrix of shared reference scores was produced for all these editors. Specifically, if editors A and B have contributed substantially to article sets S1 and S2 respectively, and n different information sources were cited in both S1 and S2, n would be the shared reference score for A and B.

We deleted those editors whose vectors contain only zeros, i.e., those who do not share cited references with any other editors in the set, which resulted in a 380x380 matrix. We left the diagonal cells empty in this matrix; they were treated as missing values and replaced with the mean in the Factor Analysis routine in SPSS that we used to explore the underlying structure of the interrelationships between these editors (McCain, 1990; White & McCain, 1998; Zhao & Strotmann, 2008a; 2008b).

Factors were extracted by Principal Component Analysis. The number of factors extracted was determined based on an examination of the Scree plot, total variance explained, and correlation residuals – the differences between observed correlations and correlations implied by the factor model (Hair, et al., 1998). This resulted in an 18-factor model that explains 57.5% of the total variance, and the differences between observed and implied correlations are smaller than 0.05 for the most part (92%).

An oblique rotation was applied resulting in a pattern matrix and a structure matrix. We use the highest loading of a factor in the pattern matrix to indicate its distinctiveness. The size or prominence of a factor is indicated by the number of editors who load primarily on this factor in the pattern matrix. A Component Correlation Matrix showing how closely factors are related to each other was also produced by the Factor Analysis routine.

Factors are labeled upon manual examination of articles that authors who load primarily in each factor have significantly contributed references to. A factor is labeled as Undefined if all loadings in this factor are lower than 0.7, although an attempt may still be made at interpreting it.

Visualization of factor structures

Both pattern matrix and structure matrix provide important information about the structure of the knowledge domain: the pattern matrix shows the memberships of editors in topic areas while the structure matrix indicates the interrelationships between these topics. They are visualized in the present study in a single two-dimensional map, which combines the informative features of both matrices (Zhao and Strotmann, 2014; 2015). The central idea of this technique is to directly visualize the factor analysis results as a bipartite network of authors and factors (specialties) linked to each other according to the loadings of authors on the factors.

For a combined visual representation of the network, we use the sparse pattern matrix to draw lines connecting editor and factor nodes in order to clearly show editors' memberships in topic areas. We use the densely connected structure matrix to automatically position the factor and editor nodes in relation to each other in order to obtain a relatively stable (and therefore potentially meaningful) layout of the visual map. The width of a line that connects an editor with a factor is proportional to the loading of this editor on this factor in the pattern matrix, as is its gray-scale value, with wider and darker lines signifying higher loadings (thus stronger memberships). Only significant loadings (i.e., 0.1 or higher) are shown on the map.

Factors and their primary members are color-coded on the map, which shows the size of each factor visually. Since we are not interested in individual editors in the present study, we did not differentiate editor nodes by size nor were we concerned about the readability of their labels on the map.

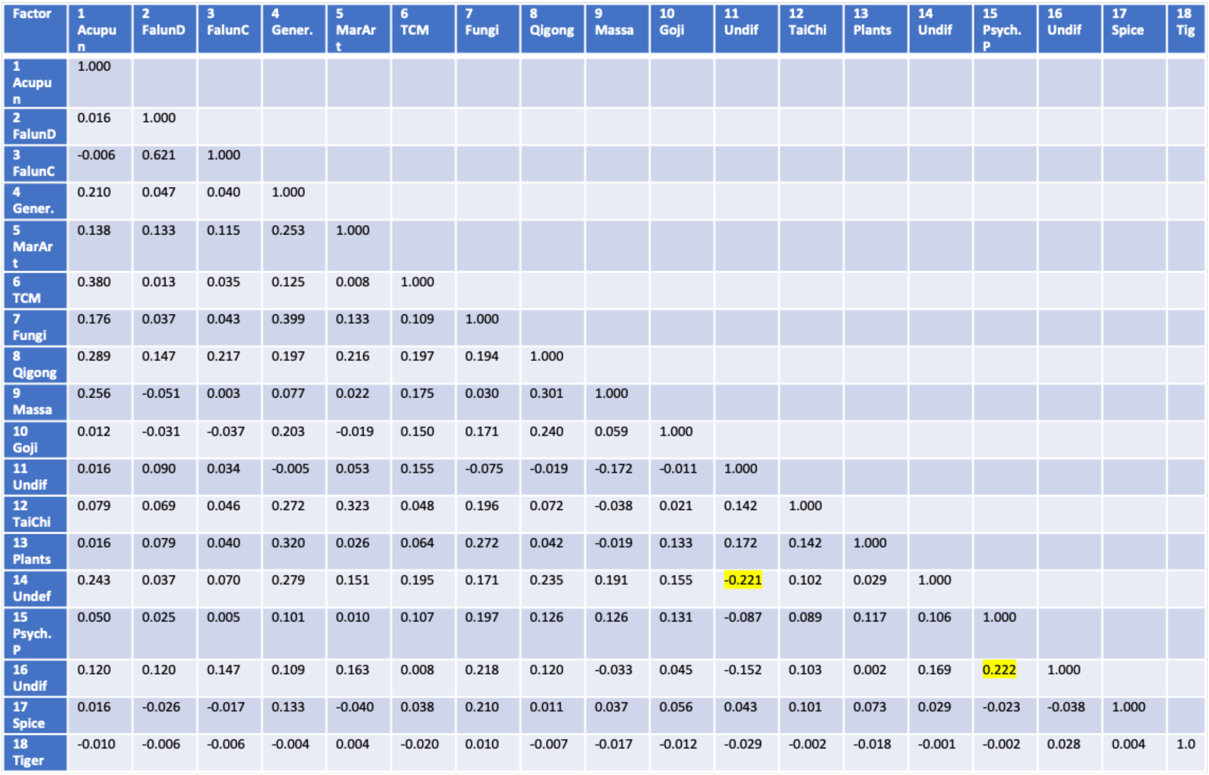
Results and discussion

Table 1 shows the topics identified and their distinctiveness and prominence indicated by the highest loading and the number of editors loading primarily and significantly on a factor respectively. Table 2 is the Component Correlation Matrix showing how closely these topics are related to each other. Correlations that are higher than 0.2 are highlighted and are considered as indicating a substantially close connection in the discussions below. Figure 1 is a visualization of the factor analysis results showing visually the size of and interrelationships between identified topic areas in the TCM knowledge domain.

Table I. Topics and their prominence (Size) and distinctiveness (Highest loading)

Factor	*Topical Label*	*Size*	*Highest_loading*
1	Acupuncture	84	0.998
2	Falun Gong - destructive	35	0.998
3	Falun Gong - constructive	35	1.011
4	Generalists	20	0.945
5	Chinese martial arts	16	0.952
6	TCM and its Common concepts	19	0.989
7	Fungi used in TCM	19	0.894
8	Qi Gong	14	0.982
9	Chinese Massages	17	0.805
10	Goji	21	0.861
11	Undefined	7	0.661
12	Tai Chi	19	0.852
13	Medicinal plants	20	0.892
14	Undefined	11	0.681
15	Psychoactive plants	11	0.832
16	Undefined (Flowering plants)	14	0.674
17	Plants used as both spice and medicine	12	0.802
18	Tiger bone wine	6	0.944

Table II. Component Correlation Matrix



Factor	1 Acupun	2 FalunD	3 FalunC	4 Gener.	5 MarArt	6 TCM	7 Fungi	8 Qigong	9 Massa	10 Goji	11 Undif	12 TaiChi	13 Plants	14 Undif	15 Psych. P	16 Undif	17 Spice	18 Tig
1 Acupun	1.000																	
2 FalunD	0.016	1.000																
3 FalunC	-0.006	0.621	1.000															
4 Gener.	0.210	0.047	0.040	1.000														
5 MarArt	0.138	0.133	0.115	0.253	1.000													
6 TCM	0.380	0.013	0.035	0.125	0.008	1.000												
7 Fungi	0.176	0.037	0.043	0.399	0.133	0.109	1.000											
8 Qigong	0.289	0.147	0.217	0.197	0.216	0.197	0.194	1.000										
9 Massa	0.256	-0.051	0.003	0.077	0.022	0.175	0.030	0.301	1.000									
10 Goji	0.012	-0.031	-0.037	0.203	-0.019	0.150	0.171	0.240	0.059	1.000								
11 Undif	0.016	0.090	0.034	-0.005	0.053	0.155	-0.075	-0.019	-0.172	-0.011	1.000							
12 TaiChi	0.079	0.069	0.046	0.272	0.323	0.048	0.196	0.072	-0.038	0.021	0.142	1.000						
13 Plants	0.016	0.079	0.040	0.320	0.026	0.064	0.272	0.042	-0.019	0.133	0.172	0.142	1.000					
14 Undef	0.243	0.037	0.070	0.279	0.151	0.195	0.171	0.235	0.191	0.155	-0.221	0.102	0.029	1.000				
15 Psych. P	0.050	0.025	0.005	0.101	0.010	0.107	0.197	0.126	0.126	0.131	-0.087	0.089	0.117	0.106	1.000			
16 Undif	0.120	0.120	0.147	0.109	0.163	0.008	0.218	0.120	-0.033	0.045	-0.152	0.103	0.002	0.169	0.222	1.000		
17 Spice	0.016	-0.026	-0.017	0.133	-0.040	0.038	0.210	0.011	0.037	0.056	0.043	0.101	0.073	0.029	-0.023	-0.038	1.000	
18 Tiger	-0.010	-0.006	-0.006	-0.004	0.004	-0.020	0.010	-0.007	-0.017	-0.012	-0.029	-0.002	-0.018	-0.001	-0.002	0.028	0.004	1.0

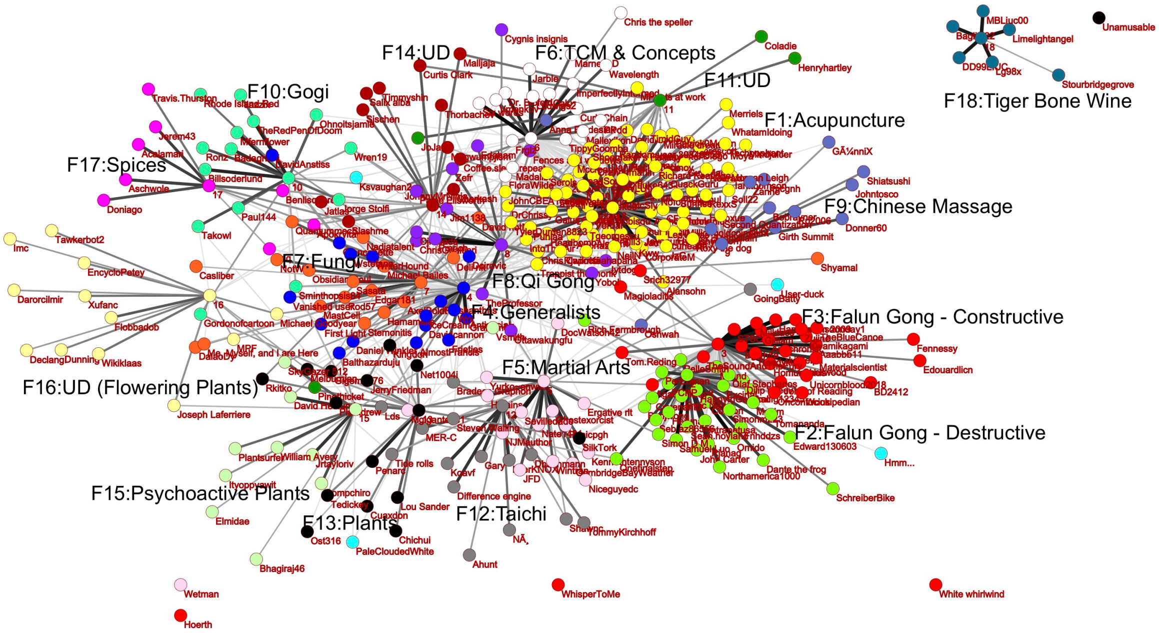


Figure 1. Structure of TCM knowledge domain

Identified topics of the TCM knowledge domain on Wikipedia (Table 1) can be grouped into seven areas: Acupuncture and related practices (F1, F9), Falun Gong (F2, F3), Herbal Medicine (F7, F10, F13, F15, F16, F17), TCM and its common concepts (F6), Qi Gong (F8), Chinese Martial Arts (F5, F12), and Non-herbal medicine (F18). The first three are the most prominent topic areas as indicated by the large numbers of Wikipedia editors associated with them primarily.

Acupuncture, Qi Gong, and related practices

As seen from Table 1 and Figure 1, the most prominent topic in the TCM knowledge domain on Wikipedia is Acupuncture (F1), which reflects the fact that among TCM theories, techniques and practices, acupuncture is the most recognized, accepted and practiced in the Western world. In many western countries, acupuncture therapists are trained and certified, and acupuncture treatments are covered by health insurance plans while other TCM practices such as herbal medicine are not. Qi Gong (F8) and Chinese massages (F9; cupping therapy, Guasha, etc.) are practices closely related to acupuncture (Table 2), but are much less recognized as seen from their much smaller sizes (Table 1).

Qi Gong, "an amalgam of traditional medical and self-cultivation practices", was often attacked by some "for their 'superstitious' nature as well as for their links to religion and spirituality" even in China where it has been primarily practised (Ownby, 2016).

Among all the identified topics, Acupuncture and Qi Gong have the most connections to the TCM knowledge domain, with 5 and 6 highlighted correlations in Table 2 respectively. They also serve as bridges for other topics to connect to the knowledge domain: Acupuncture for the topic "TCM and its common concepts" and Qi Gong for the topics Falun Gong and Chinese Martial Arts (to be discussed below). The strongest connections (i.e., correlation above 0.3) across the seven topic areas listed above (excluding connections with generalists) are between Acupuncture and TCM concepts, and between Qi Gong and Chinese massages.

Falun Gong

Falun Gong, an offshoot of Qi Gong, is the second most prominent topic with a size only slightly smaller than Acupuncture when considering both constructive (F3) and destructive (F2) contributions to the topic.

Falun Gong has been a highly controversial and political topic, and attracted attention from many. Introduced in China in 1992, Falun Gong is both a form of meditation with potential health benefits and "a process of moral self-improvement that is meant to lead to spiritual enlightenment" (Roose, 2020). Its teaching, however, led to many obsessed followers who "were driven to insanity, committed suicide and killed their loved ones", and it was banned by the Chinese government as a "heretical cult" (The Washington Post, 1999). "Today, the group is known for the demonstrations it holds around the world to 'clarify the truth' about the Chinese Communist Party, which it accuses of torturing Falun Gong practitioners and harvesting the organs of those executed" (Roose, 2020).

Editors who load high on the constructive Falun Gong factor added or modified significantly more references than they removed on Falun Gong on Wikipedia, and tend to have broad interest in TCM beyond Falun Gong as shown from them having contributed more broadly across the range of the TCM knowledge domain. In contrast, editors who load high on the destructive Falun Gong factor removed significantly more references than they added or modified, and tend to be highly focused on Falun Gong – presumably on the politics involved.

It is interesting to see the identification of separate constructive and destructive groups of editors who contribute references to articles in the same topic area on Wikipedia. This has never been observed in citation analysis studies of intellectual structures of knowledge domains represented in research articles, but is made possible by Wikipedia data that record how (e.g., adding, deleting, modifying) each editor contributes references to each article.

Compared to document-based citation analysis, author-based citation analysis has long been recognized as having the potential for studying the social structure of the community of authors implied in the intellectual structure that citation networks represent (White and Griffith, 1981; White and McCain, 1998; Zhao and Strotmann, 2008b). Actual studies have yet to realize this potential, however, and Wikipedia data that record details of editor contributions may turn out to be very useful for doing so eventually.

The Falun Gong topic area is largely separated from the rest of the TCM knowledge domain on Wikipedia as shown on Figure 1 and indicated by the mostly very low correlations with all the other topics (Table 2). The only relatively substantial connection that the two Falun Gong factors have is with Qi Gong, which reflects the belief that Falun Gong is an offshoot of Qi Gong. We can speculate on reasons for this separation. Although there are controversies over several TCM topics, such as acupuncture and Chinese herbal medicine, those are mostly about whether they are scientific and effective. Controversies around Falun Gong, however, are highly political as discussed above. Essentially, editors who contributed to Falun Gong related Wikipedia articles may not be interested in TCM *per se* but rather in the politics around Falun Gong whereas editors who contributed to other TCM topics were drawn to the medicine aspects of the topics.

TCM and its common concepts

According to resources on the University of Minnesota website (University of Minnesota, 2021), “two concepts that are unique and fundamental to Chinese medicine are Qi (usually translated as "vital energy") and yin and yang (the harmony of all the opposite elements and forces that make up existence). These two concepts form what we might call the "roots" of Chinese medicine.”

It is somewhat surprising to see that TCM concepts (F6) are perceived on Wikipedia as only having substantially close connections to a single topic: Acupuncture, which suggests that TCM concepts have not been applied to the discussions of the other conceptually related TCM topics such as Qi Gong and Tai Chi. “Yin and yang are two aspects of one unity or holism. This unity is usually expressed in a diagram known as the Tai Chi symbol. Tai Chi is usually translated as the cosmos” (Wong, 2002).

Herbal Medicine

Herbal medicine has many different types such as those that are also used as spices (e.g., Cinnamomum cassia) and those used as psychoactive drugs (e.g., Cannabis). This topic area is linked to the TCM knowledge domain through the generalists who made low contributions to all TCM topics except for Falun Gong and Tiger Bone Wine.

It is interesting to see that sub-areas of herbal medicine are not interrelated to each other but form a weak partial linear chain of links: Fungi used in TCM (F7) is related to Plants used as both spice and medicine (F17) and to the topic that appears to be flowering plants such as rose and magnolia (F16) which is then related to psychoactive plants (F15). Goji (F10) and the generic topic area on medicinal plants (F13) are not related to any other sub-areas of herbal medicine. Similar to the earlier observation that TCM concepts are perceived on Wikipedia as only closely related to a single topic area, it is somewhat surprising to see that the general topic on medicinal plants is not perceived as closely related to any of the specific types of medical plants.

Several sub-areas of herbal medicine, including the generic topic area on medicinal plants (F13), Plants used as both spices and medicine (F17), and Psychoactive plants (F15), only have a substantial connection with one of the other TCM topics. The topic “Plants used as both spice and medicine” (F17) is largely separated from the rest of the TCM knowledge domain, as indicated by the mostly very low correlations with all the other topics. Contributors to this topic who are only interested in the plants as spice may have separated this topic out from the rest of TCM.

Non-herbal medicine

In contrast to herbal medicine, only one non-herbal medicine topic (Tiger bone wine) stands out as a separate factor, although “TCM ingredients include a wide range of plants, herbs, minerals, and parts from over 1,500 animals” (Guynup, 2014). Tiger bone wine is “a tonic made by steeping a tiger carcass in rice wine to produce an extremely expensive elixir. It’s thought to impart the animal’s great strength, a status symbol product bought or gifted by the elite” (Guynup, 2014). This very small topic area is largely isolated from the rest of the TCM knowledge domain as seen visually on Figure 1 and indicated by the very low correlations with all other factors in Table 2.

Chinese Martial Arts

Tai Chi (F12) is correctly recognized on Wikipedia as a form of Chinese martial arts (F5) as indicated by the substantially close correlation between these two factors. Often known as "moving meditation," Tai Chi is a series of slow, gentle motions that are patterned after movements in nature, and its teaching almost always includes the concepts and theories, and usually movements of Qi Gong as its foundation (Piedmont Healthcare, 2021). One would thus expect a close relationship between Tai Chi and Qi Gong on Wikipedia, and it is surprising to see that this is not the case as indicated by the very low correlation (0.07) between the two. Instead, Qi Gong (F8) is perceived on Wikipedia as more closely related to Chinese massages (0.3) than to Chinese martial arts (0.2).

Conclusions

The present study applied ABCA, one of the effective bibliometric methods for mapping knowledge domains represented in published scholarly documents, to the examination of the TCM knowledge domain on Wikipedia. It is interesting to explore what we may learn from bibliometric studies of knowledge domains represented on Wikipedia, a gigantic open encyclopaedia created and maintained collectively by volunteers world-wide.

We found that seven TCM topic areas are represented on Wikipedia: Acupuncture and related practices, Falun Gong, Herbal Medicine, TCM and its common concepts, Qi Gong, Chinese Martial Arts, and Non-herbal medicine. The first three areas attracted most of the Wikipedia editors who contributed significantly to the TCM knowledge domain. Acupuncture and Qi Gong have the most connections to the TCM knowledge domain on Wikipedia, and also serve as bridges for other topics to connect to the TCM knowledge domain: Acupuncture for the topic "TCM and its common concepts" and Qi Gong for the topics Falun Gong and Chinese Martial Arts. Herbal medicine is only linked to the TCM knowledge domain through the generalists who made low contributions to many TCM topics. Non-herbal medicine is focused on a single topic (i.e., Tiger Bone Wine) which is a very small topic area isolated from the rest of the TCM knowledge domain. Falun Gong and Plants used as both spice and medicine are largely separated from the rest of the TCM knowledge domain, due probably to editors whose primary interests were in the non-medical aspects of the topic, i.e., politics around Falun Gong, and cooking, respectively.

It appears that specific topics are represented well on Wikipedia but their conceptual connections are not, especially those between a general topic (e.g., TCM concepts or Medical plants) and its sub-topics (e.g., Qi Gong and Tai Chi or Psychoactive plants).

The present study shows that ABCA is effective for mapping knowledge domains represented on Wikipedia. We also attempted using document-based bibliographic coupling analysis (DBCA), which is often used effectively to cluster research articles, to directly categorize Wikipedia articles on TCM, but found that the result did not make sense. Many Wikipedia articles are far from fully developed and therefore were often placed by DBCA into the same group based on a single shared reference, resulting in many small groups each essentially representing a single cited reference instead of a broader topic as with ABCA. This kind of impact of Wikipedia data on the analysis appears to have been smoothed out by the wider granularity and broader information that an author represents than an article does, which is one of the advantages of ABCA over DBCA discussed in previous studies (Zhao and Strotmann, 2008a; 2015).

In addition to mapping knowledge domains, the ABCA of Wikipedia articles sometimes identifies separate groups of editors who contribute references to the same topic area in different ways (e.g., constructive and destructive groups in the Falun Gong topic area). Compared to ABCA of research articles, this additional benefit for studying social structures is made possible by Wikipedia data that records details of editor contributions, and warrants closer scrutiny in follow-up research.

Acknowledgments

This study was supported by an Insight Development Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC). We would like to thank our research assistant Xinru Zhang for writing some of the data collection and analysis programs for this study.

References

Arazy, O., Nov, O., Patterson, R., and Yeo, L. (2011), "Community-Based Collaboration in Wikipedia: The Effects of Group Composition and Task Conflict on Information Quality". *Journal of Management Information Systems,* vol 21 issue 4, pp 71–98.

Burke, M., and Kraut, R. (2008). "Mopping up: modeling Wikipedia promotion decisions". *Proceedings of the 2008 ACM conference on Computer supported cooperative work,* pp. 27-36.

Clauson, K.A., Polen, H.H., Boulos, M.N.K., and Dzenowagis, J.H. (2008). "Scope, completeness, and accuracy of drug information in Wikipedia". *The Annals of Pharmacotherapy, vol 42 issue* 12, pp 1814–1821.

Das, S., Lavoie, A., and Magdon-Ismail, M. (2016). "Manipulation among the arbiters of collective intelligence: How Wikipedia administrators mold public opinion". *ACM Transactions on the Web*, Vol 10 issue 4.

Denning, P., Horning, J., Parnas, D. and Weinstein, L. (2005). "Wikipedia risks". *Communications of the ACM*, vol 48 issue 12, p 152.

H Francke, H., and Sundin, O. (2010). "An inside view: Credibility in Wikipedia from the perspective of editors". *Information Research*, Vol 15 No 3. Available at http://informationr.net/ir/15-3/colis7/colis702 (Accessed 26 Mar 2021).

Gantner, Z., and Schmidt-Thieme, L. (2009). "Automatic Content-based Categorization of Wikipedia Articles". *People's Web '09: Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources*, pp 32–37. Available at: https://www.aclweb.org/anthology/W09-3305.pdf (accessed 19 Jan 2021)

Giles, G. (2005). "Internet Encyclopedias Go Head to Head". *Nature*, Vol 438 No 7070, pp 900-901.

Guynup, S. (2014). Tigers in Traditional Chinese Medicine: A Universal Apothecary. *National Geographic*. Available at https://blog.nationalgeographic.org/2014/04/29/tigers-in-traditional-chinese-medicine-a-universal-apothecary/ (Accessed 28 Mar 2021).

Habermas, J. (1984). *The theory of communicative action: Reason and the rationalization of society.*: Beacon Press, Boston.

Haigh, C.A. (2010). "Wikipedia as an evidence source for nursing and healthcare students". *Nurse Education Today*, Vol 31 No 2, pp 135–139.

Hair, J.F. Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate data analysis* (5th edition). Prentice Hall, Upper Saddle River, NJ.

Hansen, S., Berente, N., & Lyytinen, K. (2009). "Wikipedia, Critical Social Theory, and the Possibility of Rational Discourse". *The Information Society – An International Journal*, Vol 25 No 1, pp 38-59.

Heist N., Paulheim H. (2019), "Uncovering the Semantics of Wikipedia Categories". In: Ghidini C. et al. (eds) *The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science*, 11778. Springer, Cham. https://doi.org/10.1007/978-3-030-30793-6_13.

Kessler, M. M. (1963), Bibliographic coupling between scientific papers. *American Documentation*, vol 14, pp 10–25.

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* (pp. 1509–1512). New York: ACM.

Koppelman, M.H. (2017). "WikiTweaks: The Encyclopaedia that Anyone (Who is a Skeptic) Can Edit". *Journal of Chinese Medicine*. Feb2017, Issue 113, pp 35-40.

McCain, KW. (1990). "Mapping authors in intellectual space: A technical overview". *Journal of the American society for information science*. Vol 41 No 6, pp 433-443.

McLuhan, R. (2013). *Guerrilla Skeptics*. Available at: https://monkeywah.typepad.com/paranormalia/2013/03/guerrilla-skeptics.html. (Accessed 21 Feb 2021).

Mesgari, M., Okoli, C., Mehdi, M., Finn Årup Nielsen, F., and Lanamäki, A. (2014). "The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia". *Journal of the Association for Information Science and Technology,* Vol 66 No 2, pp 219-245.

Ownby, D. (2016). Falun Gong: Chinese spiritual movement. *Britannica*. Available at https://www.britannica.com/topic/Falun-Gong (Accessed 28 Mar 2021).

Perez, B., West, A.G., Feo, C., and Lee, I. (2021). *WikiCat: A graph-based algorithm for categorizing Wikipedia articles*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.641.5224&rep=rep1&type=pdf. (Accessed 19 Jan 2021)

Piedmont Healthcare (2021). The difference between tai chi and qi gong. Available at https://www.piedmont.org/living-better/the-difference-between-tai-chi-and-qi-gong (Accessed 28 Mar 2021).

Ransbotham, S., and Kane, G.C. (2011). "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia". *MIS Quarterly,* Vol 35 No 3, pp 613–627.

Rector, L.H. (2008). "Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles". *Reference Services Review*, Vol 36 No 1, pp 7–22.

Refaei1, N., Hemayed, E.E., and Mansour, R. (2018). "WikiAutoCat: Information Retrieval System for Automatic Categorization of Wikipedia Articles". *Arabian Journal for Science and Engineering,* Vol 43, pp 8095–8109. https://doi.org/10.1007/s13369-018-3244-9.

Roose, K. (2020). How The Epoch Times Created a Giant Influence Machine. Th*e New York Times.* Available at https://www.nytimes.com/2020/10/24/technology/epoch-times-influence-falun-gong.html (Accessed 28 Mar 2021).

Simone Paolo Ponzetto and Strube, M. (2007). "Deriving a large scale taxonomy from Wikipedia". *Proceedings of the 22nd national conference on Artificial intelligence*, Vol 2, pp 1440–1445.

Simonite, T. (2013). "The Decline of Wikipedia". *MIT Technology Review*. Available at: https://www.technologyreview.com/s/520446/the-decline-of-wikipedia/ (Accessed 21 Feb 2021)

Strube, M., and Simone Paolo Ponzetto. (2006). "Wikirelate! computing semantic relatedness using Wikipedia". *Proceedings of the 21st national conference on Artificial intelligence,* Vol 2, pp 1419–1424.

The Guardian (2018). *Female Nobel prize winner deemed not important enough for Wikipedia entry.* Available at: https://www.theguardian.com/science/2018/oct/03/donna-strickland-nobel-physics-prize-wikipedia-denied (Accessed 19 Jan 2021)

University of Minnesota (2021). *What Is Qi? (and Other Concepts)*. Available at https://www.takingcharge.csh.umn.edu/explore-healing-practices/traditional-chinese-medicine/what-qi-and-other-concepts (Accessed 28 Mar 2021).

Wedemeyer, B., Yakubova, N., Kallenbach, J., Ekdahl, A., Lesko, L., Reed, E., and Schwartz, K. (2008). "Quality of the science articles on the English Wikipedia: Preliminary results". *Wikimania 2008*. Available at: http://www.youtube.com/watch?v=B7bCZbHHeZI (accessed 21 Feb 2021)

White, H. D. (2001). “Authors as citers over time”. *Journal of the American Society for Information Science and Technology,* Vol 52, pp 87-108.

White, H. D. & Griffith, B.C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, Vol 32, pp 163-171.

White, H. D., and McCain, K.W. (1998). “Visualizing a discipline: An author co-citation analysis of information science, 1972-1995”. *Journal of the American Society for Information Science.* Vol 49 No 4, pp 327-355.

Wong, K.K. (2002). *The Complete Book of Tai Chi Chuan*. Tuttle Publishing. Extracts available at https://taijiquan.org/general/the_philosophy_of_yin-yang.html (Accessed 28 Mar 2021).

Yasseri T, Sumi R, Rung A, Kornai A, and Kertész, J. (2012), “Dynamics of Conflicts in Wikipedia”. *PLoS ONE,* Vol 7 No 6, e38869. https://doi.org/10.1371/journal.pone.0038869.

Zhao, D., & Strotmann, A. (2008a). “Evolution of research activities and intellectual influences in Information Science 1996-2005: Introducing author bibliographic coupling analysis”. *Journal of The American Society for Information Science and Technology,* Vol 59 No 13, pp 2070-2086.

Zhao, D., & Strotmann, A. (2008b). “Information science during the first decade of the web: An enriched author cocitation analysis”. *Journal of The American Society for Information Science and Technology,* Vol 59 No 6, pp 916-937.

Zhao, D., & Strotmann, A. (2014). “The knowledge base and research front of Information science 2006-2010: An author co-citation and bibliographic coupling analysis”. *Journal of the Association for Information Science and Technology*, Vol 65 No 5, pp 996-1006.

Zhao, D., & Strotmann, A. (2015). *Analysis and Visualization of Citation Networks*. Morgan & Claypool Publishers. (doi:10.2200/S00624ED1V01Y201501ICR039)

The Washington Post (1999). The Falun Gong Controversy. *The Washington Post.* Available at https://www.washingtonpost.com/archive/opinions/1999/09/06/the-falun-gong-controversy/97fa4b0f-96a5-4a42-864b-51dcc7037975/. (Accessed 28 Mar 2021).