Behind The Wikipedia Medical Knowledge Factory: Understanding the Knowledge Dynamic Over Time

Reham Al Tamime
University of Southampton
Web Science Institute
Southampton SO17 1BJ
rat1g15@soton.ac.uk

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3027119

Abstract

Wikipedia has challenged the way traditional encyclopedia knowledge is built and contested by creating an open socio-technical environment that allows non-domain experts to contribute to scientific and medical knowledge. The open nature of Wikipedia has been successful in attracting readers to its medical content, but there are doubts about the quality and trustworthiness of its articles. The goal of this research is to increase transparency and trust in Wikipedia medical articles by understanding the process of medical knowledge building over time. Health-related articles in Wikipedia pass through increasing trends in editing activities. In addition, health-related articles include medical controversies that are discussed between editors. By examining the community's levels of engagement and reactions over time through the lens of Actor Network Theory and applying quantitative and qualitative analyses of actors and their relations, the contribution of this work will extend theory to offer both theoretically- and empirically-informed design principles for building and evaluating crowd-sourced knowledge environments that engender trust and maintain transparency.

The Author Keywords: Human Computer Interaction; online communities; trust; dynamic of online information; controversies.

ACM Classification Keywords:
Human Factors.

Introduction

Since its launch in 2001, Wikipedia has become the most popular general reference site on the Internet, and a prominent source of online health information compared to the other online health information providers such as MedlinePlus and NHS Direct Online [20]. Wikipedia has challenged traditional encyclopedias where only scientists or domain experts are able to claim, contest and accept scientific knowledge. There is a debate regarding the quality, accountability, and trustworthiness of articles in Wikipedia. For example, an investigation that compares Wikipedia and Britannica's coverage of science published in *Nature* found that crowd-sourced Wikipedia's content comes close to Britannica's content in terms of accuracy of its science entries [13]. However, research published in ACM Press suggests that Wikipedia is not an acceptable source for citation because there is no fact-checking mechanism in place to ensure the accuracy and the reliability of its entries [31]. The BBC warned against trusting Wikipedia's health-related entries such as in heart disease, lung cancer, depression and diabetes articles because scientists had spotted errors in entries in comparison to peer-reviewed journals [28].

These criticisms, however, are limited because they view Wikipedia entries as static, and do not show change over time. They do not examine how medical or health-related articles pass through bursts in editing activities, and how the community engages with and reacts to edits. The aim of this research, therefore, is to understand how medical entries in Wikipedia are created, contested and changed over time. This will be done by using quantitative and qualitative analyses to study Wikipedia medical articles. Looking at medical articles from both macro- and micro-level perspectives will help to understand the conditions of their construction at different periods, and the characteristics of the community that is responsible for medical articles' construction.

This research seeks to contribute theoretically, practically and methodologically. The theoretical contribution is building on Actor Network Theory to explain the dynamic of knowledge building activities. The practical contribution offers theoretically- and empirically-informed design principles for building and evaluating crowd-sourced knowledge environments that engender trust and maintain transparency. The methodological contribution is applying a mixed method approach to understand the dynamics of knowledge construction, trust, and transparency in environments such as Wikipedia.

Background and Literature Review

Current research explores Wikipedia's dynamic editing activities to achieve four different goals: identifying and describing controversy; visualizing controversies and editors' contributions; predicting and describing editors' forms of interactions and understanding the facts-building process in Wikipedia. Recent research projects focus on building models to identify and detect controversial topics in Wikipedia [31, 4, 22, 21, 23, 25, 32, 34, 10, 33]. These studies have shown that edit-history information such as the number of reverted revisions, length of discussions, editors' vote for one another in elections can be used both to automatically find conflict within articles, and for ascertaining levels of trustworthiness. Other studies have modeled, mapped and visualized controversy over time in Wikipedia [3, 4, 23, 7, 18, 5, 7, 10] or have worked on visualizing, mapping and modeling collaboration patterns and content change [11, 27, 26]. The visualization approaches that used color schemas, dashboards and representing text as lines were effective in unmasking the types of social behaviors such as negotiation and consensus that occur through the knowledge building process in Wikipedia. In addition, researchers examined the collaborative process and emergence of content creation in Wikipedia [29, 14, 6, 24, 12, 17, 8, 19]. These studies have taken approaches such as social network analysis to describe the dynamics of the Wikipedia's editing activities, and editors' agreement and disagreement. Moreover, social network analysis has been used to investigate whether the structures of breaking and nonbreaking articles are similar [16]. Such research affords insights about how peer-production communities create knowledge through affiliation networks of articles and editors [15].

Previous research has revealed that there is a relatively weak theoretical foundation to explain the knowledge-building process in Wikipedia. There is a shortage in research that explores how interactions change in crowd-sourcing environments *over time*, and the contours of trends. Therefore, the central research question is, *How to understand the dynamic of medical knowledge-building activities in Wikipedia,* and the follow-on questions are, *How to characterize the community's levels of engagement and reactions to disease-related articles over time and analyze the significance of editorial 'burstiness'.?* These questions will inform whether increasing trends in editing activities play a role in changing the knowledge flow to and from the articles. Also, answering these questions will reveal the type of editors that engage with medical articles and their level of consistency and dedication in editing these articles. This, in turn, will inform design principles that make community and editors' activities transparent to readers. Such transparency could support tools or processes to judge the trustworthiness of articles.

Research Design and Methodology

Actor Network theory (ANT), which emerged during the mid-1980s with the work of Bruno Latour, Michel Callon, and John Law, views science as a heterogeneous process in which the social, technical, conceptual, and textual are puzzled together and transformed. ANT recognizes the agency of both human actors and non-human actors (such as machines, animals, texts, and hybrids, among others). Both human and non-human actors form a heterogeneous network. *Actants* in ANT refers to any actor, collective or individual, that can associate or disassociate with other actors. Networks are processual, built activities, performed by the actants out of which they are composed [7]. Actants and their network are 'translated' as a part of their activity building. The process of translation "describes a variety of ways in which actors seek to interest others in supporting the construction of claims, enrolling actors directly or indirectly in a coalition to build a fact or a machine" [8].

Quantitative analysis will be used to ascertain and measure network activity over time and to measure correlation between burstiness in editing activities and medical articles community's levels of engagement and reactions. Qualitative analysis will be used to look at medical controversies in talk pages and how they are contested between editors in different cases of emerging and chronic diseases. Mixing methods has been recommended to gain an 'outsiders' view of the network in terms of the structure of the network and also to gain data on the perception of the network from an 'Insiders' view, including the content for those involved [9].

Medical articles and their editing history will be extracted from English Wikipedia API from the creation date of the article to 2016. The sampling frame includes all the medical articles listed under the WikiProject Medicine. Articles of emerging diseases such as Zika and chronic diseases such as Diabetes will be categorized to understand how medical controversies are compared between articles of diseases that exist for a long period and have large number of scientific research papers and articles of diseases that have less scientific research about them and confirmed knowledge about possible medication and treatments.

Phase I: Understanding the 'medical articles community', levels of engagement and reactions over time: the aim of this phase is to study different editors' level of engagement with medical articles and to look at those editors' reaction to disease outbreak articles. Also, the correlation between sudden increase in editing activities (burstiness) and the medical articles community's levels of engagement and reactions will be investigated.

Phase II: Understanding the medical articles community's reactions to controversies in certain and uncertain situations: for this phase, articles of emerging and chronic diseases will be selected for qualitative analysis. These articles' talk pages will be retrieved to look at editors' discussions around controversial medical issues. Thematic analysis will be applied using the Actor Network Theory stages of 'translation' which is effective in explaining the rationale behind the dynamic of medical knowledge building in Wikipedia.

Preliminary Findings

As a preliminary analysis, the Zika Fever article has been analyzed. The gray line summarizes the number of articles' edits over time as an indication of the public and medical articles community interest around them. The red dots depict burstiness in daily editing activities.

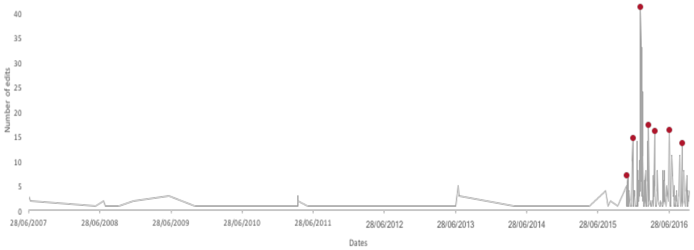

Figure 1: Zika Fever Article

These results suggest that an increasing trend could represent a change either related to the status of the medical information such as diseases outbreak (external) or a change related to the organization of co-editing activities between Wikipedia editors such as suggesting new sections and disagreement (internal). These results buttress the research questions, and call for research that understands editors' interactions at sudden increase in editing activities and whether burstiness in editing activities means new contributors to the medical knowledge in Wikipedia.

Acknowledgement

I would like to thank my supervisors Professor Wendy Hall and Dr. Richard Giordano for their guidance and advice.

References

1. Berry, D. 2012. *Understanding digital humanities.* Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
2. Borra, R., Weltevrede,R., Ciuccarelli, P., Kaltenbrunner, A., Laniado, D., Magni, G., Mauri, M., Rogers, R. and Venturini, T. 2015. Societal Controversies in Wikipedia Articles. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 193-196. DOI=http://dx.doi.org/10.1145/2702123.2702436
3. Borra, R., Weltevrede,R., Ciuccarelli, P., Kaltenbrunner, A., Laniado, D., Magni, G., Mauri, M., Rogers, R. and Venturini, T. 2014. Contropedia - the analysis and visualization of controversies in Wikipedia articles. In *Proceedings of The International Symposium on Open Collaboration* (OpenSym '14). ACM, New York, NY, USA, Pages 34 , 1 pages. DOI=http://dx.doi.org/10.1145/2641580.2641622
4. Brandes U, Lerner J. 2007. Revision and co-revision in Wikipedia. *In Proceedings of the International Workshop on Bridging the Gap Between Semantic Web and Web 2.0 at the 4th European Semantic Web Conference (ESWC'07*), Innsbruck, Austria*,* 85-96
5. Brandes, U and Lerner, J. 2007. Visual Analysis of Controversy in User-generated Encyclopedias. *2007 IEEE Symposium on Visual Analytics Science and Technology* (2007). DOI:http://dx.doi.org/10.1109/vast.2007.4389012
6. Burke, M and Kraut, R. 2008. Taking up the mop: identifying future wikipedia administrators. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '08). ACM, New York, NY, USA, 3441-3446. DOI=http://dx.doi.org/10.1145/1358628.1358871.
7. Crawford, C. (2004). 'Actor network theory'. In: Ritzer, G. *Encyclopedia of social theory*. Thousand Oaks: Sage Publications Inc.
8. Cunha, et al. (200*9). Handbook of Research on Social Dimensions of Semantic Technologies and Web Services*. London: IGI Global.
9. Edwards, G. (2010) *Mixed-method approaches to social network analysis*. Discussion Paper. NCRM.
10. Ekstrand, M and Riedl, J. 2009. rv you're dumb: identifying discarded work in Wiki article history. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (WikiSym '09). ACM, New York, NY, USA, Article 4, 10 pages. DOI=http://dx.doi.org/10.1145/1641309.1641317
11. Fernanda, B.,Viégas, M. and Kushal, Dave. 2004. Studying cooperation and conflict between authors with *history flow* visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04). ACM, New York, NY, USA, 575-582. DOI=http://dx.doi.org/10.1145/985692.985765
12. Ferron, M and Massa, P. 2011. Collective memory building in Wikipedia: the case of North African uprisings. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (WikiSym '11). ACM, New York, NY, USA, 114-123. DOI=http://dx.doi.org/10.1145/2038558.2038578
13. Giles, J. 2005. Internet encyclopaedias go head to head. *Nature news@nature* 438, 7070 (2005), 900–901. DOI:http://dx.doi.org/10.1038/438900a
14. Kaltenbrunner , A and Laniado, D. 2012. There is no deadline: time evolution of Wikipedia discussions. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (WikiSym '12). ACM, New York, NY, USA, Article 6 , 10 pages. DOI=http://dx.doi.org/10.1145/2462932.2462941.
15. Kane, G and Ransbotham, S. (2014) Research Note—Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities*. Journal of Information Systems Research*, 27(2), 424 – 439. http://dx.doi.org/10.1287/isre.2016.0622
16. Keegan, B, et al. (2013) Hot Off the Wiki: Structures and ynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist.* 7, 595 – 622.
17. Keegan, B., Gergle, D and Contractor, N. 2011. Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (WikiSym '11). ACM, New York, NY, USA, 105-113. DOI=http://dx.doi.org/10.1145/2038558.2038577.

18. Kittur, A., Suh, B., Pendleton, B. and Chi., E 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 453-462. DOI=http://dx.doi.org/10.1145/1240624.1240698
19. Roth, C., Taraborelli, D., and Gilbert, N. 2008. Measuring wiki viability: an empirical assessment of the social dynamics of a large sample of wikis. In *Proceedings of the 4th International Symposium on Wikis* (WikiSym '08). ACM, New York, NY, USA, , Article 27 , 5 pages. DOI=http://dx.doi.org/10.1145/1822258.1822294
20. Laurentm, M.R., Vickers, T. R. 2009. Seeking Health Information Online: Does Wikipedia Matter? *Journal of the American Medical Informatics Association* 16, 4 (January 2009), 471–479. DOI:http://dx.doi.org/10.1197/jamia.m3059
21. Sepehri-Rad, H., and Barbosa, 2012. Identifying controversial articles in Wikipedia: a comparative study. *In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (WikiSym '12).* ACM, New York, NY, USA, Article 7, 10 pages. DOI=http://dx.doi.org/10.1145/2462932.2462942.
22. Sepehri-Rad, H., and Barbosa, D. 2015. Identifying Controversial Wikipedia Articles Using Editor Collaboration Networks. *ACM Trans. Intell. Syst. Technol.* 6, 1, Article 5 (March 2015), 24 pages. DOI=http://dx.doi.org/10.1145/2630075.
23. Sepehri-Rad, H., and Barbosa. 2011. Towards identifying arguments in Wikipedia pages. In *Proceedings of the 20th international conference companion on World wide web* (WWW '11). ACM, New York, NY, USA, 117-118. DOI=http://dx.doi.org/10.1145/1963192.1963252.
24. Slattery S. 2009. "edit this page": the socio-technological infrastructure of a wikipedia article. In *Proceedings of the 27th ACM international conference on Design of communication* (SIGDOC '09). ACM, New York, NY, USA, 289-296. DOI=http://dx.doi.org/10.1145/1621995.1622052.
25. Suh, B. ; Chi, E. H. ; Pendleton, B. A. ; Kittur, A. 2007 . Us vs. them: understanding social dynamics in Wikipedia with revert graph visualizations. *IEEE Symposium on Visual Analytics Science and Technology* (VAST '07). Sacramento, CA. Piscataway, NJ: 163-170.
26. Suh, B., Chi, E., Kittur, A . 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 1509-1512. DOI=http://dx.doi.org/10.1145/1518701.1518930.
27. Suh, B., Chi, E., Kittur, A and Pendleton, B. 2008. Lifting the veil: improving accountability and social transparency in Wikipedia with wikidashboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 1037-1040. DOI=http://dx.doi.org/10.1145/1357054.1357214.
28. Stephens, p. (2014) 'Trust your doctor, not Wikipedia, say scientists.' *BBC* . 28 May. Available at: http://www.bbc.co.uk/news/health-27586356.
29. Swarts. J. 2009. The collaborative construction of "fact" on Wikipedia. In *Proceedings of the 27th ACM international conference on Design of communication* (SIGDOC '09). ACM, New York, NY, USA, 281-288. DOI=http://dx.doi.org/10.1145/1621995.1622051.
30. Vuong, B., Lim,B., Sun, A.,Le, M., Lauw, H., and Chang, K. 2008. On ranking controversies in wikipedia: models and evaluation. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (WSDM '08). ACM, New York, NY, USA, 171-182. DOI=http://dx.doi.org/10.1145/1341531.1341556
31. Waters, N.L. "Why You Can't Cite Wikipedia in My Class", *Comm. ACM,* 50(9), ACM Press (2007), 15-17.
32. Yasseri, T., Sumi, R., Rung, A., Kornai, A, and Kertész, J. 2012. Dynamics of Conflicts in Wikipedia. *PLoS ONE* 7, 6 (2012). DOI:http://dx.doi.org/10.1371/journal.pone.0038869 .
33. Yasseri, T., Sumi, R., Rung, A., Kornai, A, and Kertész,J. 2011. Edit Wars in Wikipedia. In *2011 IEEE Third International Conference on Social Computing (SocialCom),* Boston, MA, USA. 724-727.
34. Yasseri, T., Sumi, R., Rung, A., Kornai, A, and Kertész,J. 2011. Characterization and prediction of wikipedia edit wars. In *Proceedings of the ACM WebSci'11*, Koblenz, Germany.