

Medical Science in Wikipedia: The Construction of Scientific Knowledge in Open Science Projects

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

PhD Supervisors:
Professor Wendy Hall & Dr. Richard Giordano

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, but there are concerns about the quality and trustworthiness of its articles. The goal of my research is to build a theoretical framework to explain the dynamic of knowledge building in crowd-sourcing based environments like Wikipedia and judge the trustworthiness of the medical articles based on the dynamic network data. By applying Actor Network Theory and Social Network Analysis, the contribution of my research is theoretical and practical as to build a theory on the dynamics of knowledge building in Wikipedia across times and to offer insights for developing citizen science crowd-sourcing platforms by better understanding how editors interact to build health science content.

Keywords

Open collaboration; trust; wikipedia; facts' dynamics; citizen science.

1. 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 [17]. Wikipedia has challenged the traditional encyclopedia where only scientists are responsible. By relying on the wisdom of crowds, anyone can create, claim, contest and approve scientific facts in Wikipedia's socio-technical environment. Despite Wikipedia's popular success, this open nature of knowledge production is viewed with skepticism by its consumers. The quality, accountability, and trustworthiness of the articles in Wikipedia have been debated heavily [10, 11, 26, 28] especially with respect to trusting critical health and medical information.

According to Bruno Latour, the knowledge production process starts with a *statement* that goes through several operations in laboratories such as "adding modalities, citing, enhancing, diminishing, borrowing and proposing new combinations" [15, p.86-87]. Statements could remain claims such as a result of no

interest or operations in them even after publication. Claims are statements that could not convince other people of their truthfulness. In contrast, "some statements could change their status rapidly, following a kind of alternate dance, as they are proven, disproven and proven again." [15, p.87]. Such statements remain in flux as a consequence of disputes about them, until some reach a point of stabilization. This is when the statement becomes a *fact* that forms textbook knowledge. Facts are "taken-for-granted features" [15, p.87] that are neither questioned or modified, and slowly disappear from the concerns of daily scientific activity.

Latour's work was limited to understand scientific knowledge construction inside labs. The aim of my research is to understand scientific knowledge construction in open environment like Wikipedia. Latour asserts that "statements lie along a continuum according to the extent to which they refer to the conditions of their construction" [15, p.176]. Therefore, understanding the dynamic of knowledge creation will help us to ascertain the veracity of statements in Wikipedia. Additionally, my research illuminates how citizens are participating in creating scientific knowledge.

2. Goals

The research question is *How do scientific knowledge's social network structure properties differ or stay the same at different phases and trends happening in Wikipedia articles?* Recent research projects have focused on building models to identify and detect controversial topics in Wikipedia [27, 2, 19, 18, 20, 22, 29, 31, 7, 30]. 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 to automatically find conflict within articles for administrative and trust in articles purposes. Other studies have modeled, mapped and visualized controversy over time in Wikipedia [1, 2, 18, 5, 14, 3, 5, 7] or have worked on visualizing, mapping and modeling collaboration patterns and content change [8, 24, 23]. These visualization approaches that utilized color schemas, dashboards and representing text as lines were effective in term of unmasking the types of social behaviors such as negotiation and consensus that occur through the facts' building process in Wikipedia. In addition, researchers have tried to understand the collaborative process and emergence of content creation in Wikipedia [25, 12, 4, 21, 9, 13, 6, 16]. These studies have used techniques such as social network analysis and bounded confidence (BC) model to describe dynamic of the Wikipedia's editing activities, editors' agreement and disagreement. However, little is communicated about the rational behind facts' transformation over time in Wikipedia. While these researches have touched on the issues that I am going to study in my research, there are still some gray areas that are worth highlighting.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

OpenSym '16 Companion, August 17-19, 2016, Berlin, Germany

ACM 978-1-4503-4481-4/16/08.

<http://dx.doi.org/10.1145/2962132.2962141>

The previous research revealed that there is still lack in developing theoretical foundations to explain the facts' building process in Wikipedia. Knowledge is not static, but is a process of acquisition and reflection that depends on the involvement and interactions of several actors. Knowledge in Wikipedia is also dynamic in nature and this calls for longitudinal analysis of Wikipedia editing activities at different phases and trends.

3. Methods

Medical articles in Wikipedia will be studied exclusively because they represent trusted online scientific content, which raises questions about the process of their creation and veracity of their statements after or during controversy. Therefore, medical articles and their history will be extracted from English Wikipedia API¹ from the creation date of the article until 2016. The English language Wikipedia will be used because of its global nature and large number of articles, editors, edits and active users. Actor Network Theory will be applied to uncover the contributions of human and non-human actors in this socio-technical community.

Social network analysis (SNA) will be used to identify patterns of editors' interactions when opening and closing facts and to uncover relationships' changes as a result of addressing controversies and accepting facts. Specifically, SNA helps to address if there is a relationship between the dynamic of network structure and the dynamic of medical statements. Branching in the network and small group's discussions at several stages of facts' production will be revealed.

As a preliminary analysis, four medical related articles from Wikipedia have been sampled. The first two articles can be described as articles about diseases that exist for a long period of time and attract large number of scientific research and attention from patients and non-patients. These two articles are the Cancer and Diabetes Wikipedia articles. The other two articles can be described as articles about emerging diseases appeared recently and have less scientific research about them and confirmed knowledge about possible medication and treatments. These two articles are the Ebola and Zika Wikipedia articles. Revert edit analysis on these Wikipedia medical articles have been conducted. Revert edits means reversing a prior edit, which typically leads to an article being restored to a previous version. Revert edits reflect the presence of controversies. So here revert edit analysis will serve as a way to quantify the *intensity* of the controversies between editors.

4. Results

The four articles have been extracted from Wikipedia API containing revision id, timestamp, username, and comment. Revisions that have reverted as a result of vandalism have been excluded from analysis. The results summarize the number of articles' edits over time as indication of the public interests around them. The following is a plot to show the results as aggregated monthly and presented in a scale spaced in a six-month period.

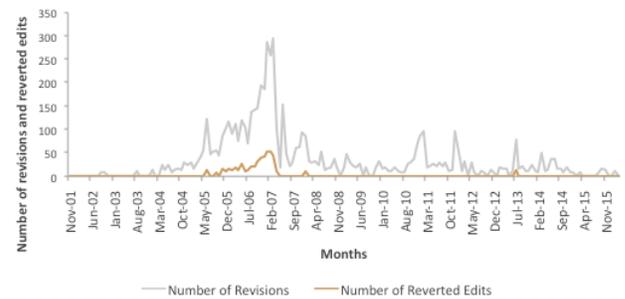


Figure 1: Cancer Wikipedia Article

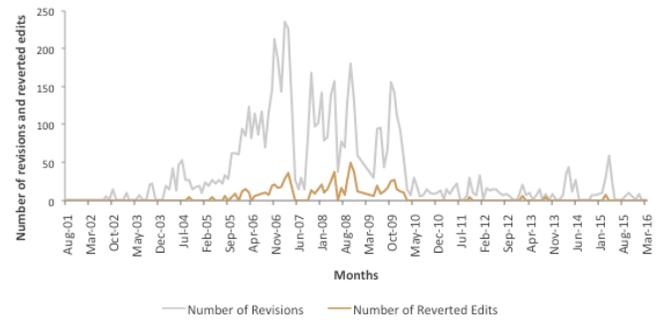


Figure 2: Diabetes Wikipedia Article

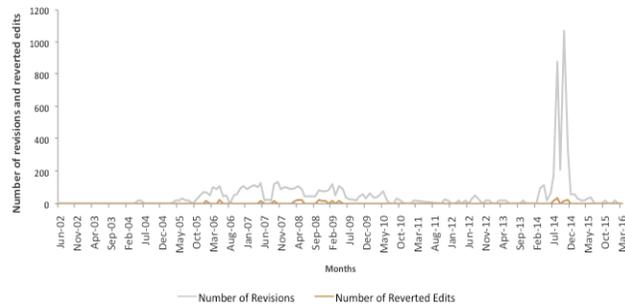


Figure 3: Ebola Wikipedia Article

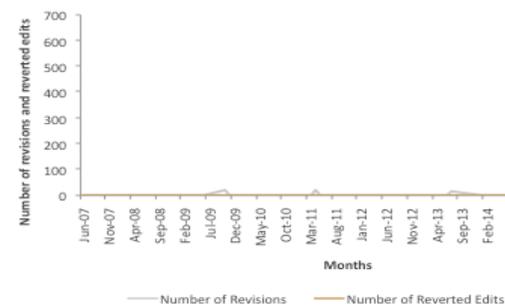


Figure 4: Zika Wikipedia Article

As we can notice from the plots, there are certain period of times that bring burst in edit and revert activities. These times will be probably associated with news or scientific updates about the diseases or with the start of the epidemics spread. It is also interesting to see in the Cancer and the Diabetes articles that increase in edits also bring increase in reverting of edits. These reverts dissolve as time moves on which makes us wonder: what happened behind the scene at different phases and trends.

¹ https://www.mediawiki.org/wiki/API:Main_page

5. Discussions

The shown results of dynamics of increase and decrease of revert activities are consistent with Latour's studies of scientific statements' change from being controversial to being accepted over time. Also, the results are consistent with other studies [9, 13] that showed trends in Wikipedia editing activities as a result of external factors such as political movement and natural disasters. These preliminary findings endorse the validity of the research question and call for a framework that understands the iterative process that Wikipedia articles go through over time.

6. Questions

The OpenSym's Wikipedia and Wikimedia Research Track will touch on issues relating to how and why modes of organization and collective behavior support more inclusive and collaborative participation in Wikimedia projects. I am very interested to learn about how and why modes of organization and collective behavior support reaching consensus or facilitating controversial discussions in crowd-sourcing projects such as Wikipedia. In addition, I am looking forward to get feedback about my research methodology from the Doctoral Symposium Committee at the OpenSym.

7. 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] 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>
- [8] 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>
- [9] 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>
- [10] 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>
- [11] Hafner, K. "Seeing Corporate Fingerprints in Wikipedia", *The New York Times*, Aug 17th, 2007.
- [12] 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>.
- [13] 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>.
- [14] 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>
- [15] Latour, B. and Woolgar, S. 1979. *Laboratory Life: the Social Construction of Scientific Facts*. Los Angeles: Sage.
- [16] 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>
- [17] 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>
- [18] 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>.
- [19] 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>.

- [20] 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>.
- [21] 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>.
- [22] 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.
- [23] 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>.
- [24] 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>.
- [25] 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>.
- [26] Vargas, J.A. "On Wikipedia. Debating 2008 Hopefuls' Every Facet", The Washington Post, Sept 17th 2007.
- [27] 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>
- [28] Waters, N.L. "Why You Can't Cite Wikipedia in My Class", *Comm. ACM*, 50(9), ACM Press (2007), 15-17.
- [29] 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>.
- [30] 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.
- [31] 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. 1-3.