WIKIPEDIA AS A TOOL FOR INTRODUCING SOCIAL CONCERNS INTO SCIENCE EDUCATION

Tamar Groves, Carlos G. Figuerola, José Luis Alonso Berrocal, Miguel Ángel Quintanilla

The Institute of Science and Technology Studies, University of Salamanca

Abstract Wikipedia is becoming a main source for scientific information. However, we know very little of the nature of science transmitted by Wikipedia. In this paper we present the preliminary results of our attempt to characterize scientific knowledge on the Spanish Wikipedia. Our analysis consists of two stages: Identifying scientific and technological content and highlighting the relationships among its components. Comparing our results with similar procedures carried out on school textbooks, suggests that there is a significant difference between scientific culture in traditional educational aids and on the Web. Scientific content in textbooks is characterized by a traditional disciplinary division, in spite of efforts to integrate interdisciplinary approaches and significant social contexts. It seems that for its dynamic nature Wikipedia is more prone to an interdisciplinary structure of contents. This relative flexibility also opens the way to the incorporation of social concerns alongside traditional scientific disciplines. We believe thus that the popularity of Wikipedia can contribute to changing the nature of scientific content available to students.

Key words: Wikipedia, Science Education, Nature of Science

INTRODUCTION

Whether we like it or not there is mounting evidence that Wikipedia is becoming a main source for scientific information. In fact for students on all levels Wikipedia is the first step for investigating any topic (Moldwin and Miller, 2007). Its accessibility, vast content and dynamic character make it an easy and updated source of information. Even courts are using it on a regular basis for questions of geography and definitions of technical terms (Miller and Murray, 2010). More importantly the approval of Wikipedia in scientific circles is manifested by its frequent appearance in peer reviewed journal articles as a source of data (Okoli, 2009).

The novelty of Wikipedia is its collective and collaborative nature. The Wikipedia community consisting of single authors, editors, administrators and the public shape its content and structure. Although there are a few scholars that criticize its reliability, most research validates its epistemic qualities. More important it is regularly included in the 10 most visited cites demonstrating its importance as a source of knowledge (Okoli, 2009).

However, we know very little of the nature of science transmitted by Wikipedia. Most research has been limited to evaluating its reliability. A research conducted by Nature found its accuracy comparable to that of Britannica (Giles, 2005). Nevertheless other studies maintain that despite the review process, the lack of formal gatekeeping procedure ensures that the lowest common denominator will prevail (Svoboda, 2006). Still, most efforts to examine the issue confirm its position as a valuable source of

knowledge. There is even research of Wikipedia as an epistemological phenomenon examining how it affects people´s consciousness of how they know what they believe they know (Dede, 2008). Wikipedia represents thus a significant shift in how knowledge is evaluated and received.

RATIONALE

As a result we believe that there is an urgent need to understand better the character of Wikipedia's scientific contents. The Nature of Science conveyed by science textbooks is a growing field of research. The prominence of school textbooks has birthed a growing body of research aimed at characterizing the Nature Of Science (NOS) they transmit. In an early and influential attempt from 1991, Chiappetta, Fillman & Sethna determined the relative emphasis of different aspects of science. They found that "science as a body of knowledge" was the predominant theme among these texts. The second most-emphasized theme was "science as a way of investigating". The "interaction of science, technology, and society" received some coverage while "science as a way of thinking," seemed to be neglected in most of the science textbooks. Vesterinen *et al.*'s study of Nordic chemistry books found that similarly to science textbooks published in the USA, the Finnish and Swedish upper secondary school chemistry textbooks seem overly focused on the content of science and too little on the dimensions of "science as a way of knowing", and "interaction of science, technology and society".

Our research on Spanish textbooks indicates a similar pattern. We found that science books are mainly concerned with intrinsic elements of scientific culture while extrinsic elements related to the way society interacts with science tend to appear in non scientific books. We also found that Spanish curriculum as a whole is still divided along traditional disciplinary lines thus scientific information does not appear alongside its social implications or its technological applications (Groves, Quintanilla and Escobar, 2012, Quintanilla et al., 2011).

Extending the effort to understand the Nature of Science to Wikipedia is a challenging task due to its dimensions and diversity and to other specific problems. In this preliminary research we attempte to identify scientific and technological content, assessing its relative weight and highlighting the relationships among its components. This allows an initial appreciation of the potential of Wikipedia as a didactic tool as well as illuminating important characteristics of online scientific culture.

METHODS

The first task was to design an automatic procedure that would filter Wikipedia in order to detect articles with a relevant scientific content and assess their weight. The second task was to reveal the semantic structure of Wikipedia's scientific content (Chernov, Iofciu, Nejdl. & Zhou 2006), analyzing the links between different Wikipedia articles.

The whole Wikipedia is available for downloading by anyone as a database in XML format which can be imported in a MySQL database engine. For every specific language version of Wikipedia, there are snapshots of specific moments. These snapshots consist of several Gbytes of data including not only the visible part of Wikipedia, but also administrative data, historical records of editions of every article, etc.

For our research we used the snapshot of January, 29 of 2012 of the Spanish version, focusing on: articles (text), categories, and links. After cleaning the data (droping redirections, disambiguating pages and so on), we had more than 800,000 articles of all

topics. Using this data we could have built a directed graph in which the nodes are the articles and the directed edges are the links that connect them. However, this graph would have been too big to be easily processed. As a result we decided to resort to the categories assigned by the authors of the Wikipedia articles.

The categories are a kind of tags referring to the topic of the articles. After cleaning the categories list (dropping administrative categories as well as categories with very low frequencies), we had close to 64,000 categories. We analyzed all the weblinks that output from the articles of each category and point to articles of other categories. This allowed us to build a network in which categories are the nodes and the grouped links are the edges; the number of articles of every category pointing to other categories consists of the weight of the edges. The result is a directed and weighted graph of 64,000 nodes and only 2,000,000 edges. On this graph we applied techniques of Social Network Analysis in order to detect communities of Wikipedia categories (nodes). Correlated categories have more and stronger links between them, thus we can expect the emergence of communities of categories containing articles about close topics.

The technique we used is based on an algorithm known as InfoMap (Rosvall, Axelsson, & Bergstrom 2009), which takes into account not only the links between nodes, but also their direction and their intensity (weight). We chose this algorithm because it is reasonably fast in cases in which the network does not have many edges. Using Infomap, we detected 826 communities (of different sizes), which were manually revised selecting those communities that belong to Science & Technology. We found 116 communities of Science and Technology categories.

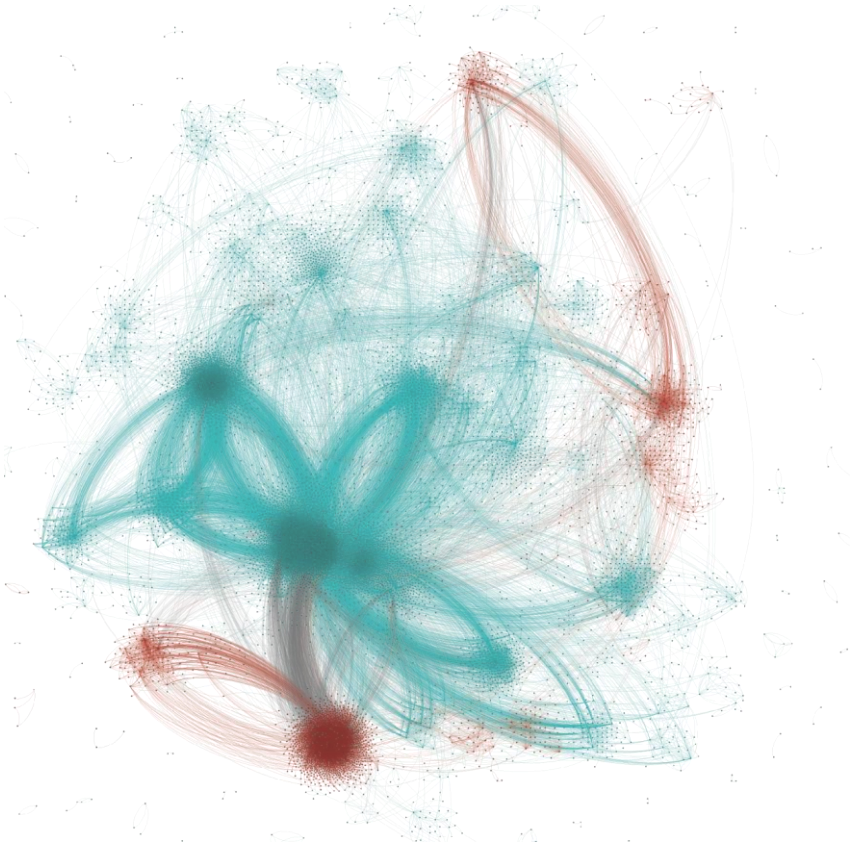


Figure 1. The network of Wikipedia categories: S & T highlighted in red

Using this procedure we detected 94,797 articles about Science and/or Technology. We checked manually a random sample and discovered that some of them were simply a title and a place in a taxonomycal tree. These kinds of articles are usually related to lists of classifications such as asteroids, zoological species, bacteria, etc. Based on the edition history and size we cleaned these empty articles and reduced the quantity to 29,639 articles on Science and/or Tecnology in the Spanish version of Wikipedia.

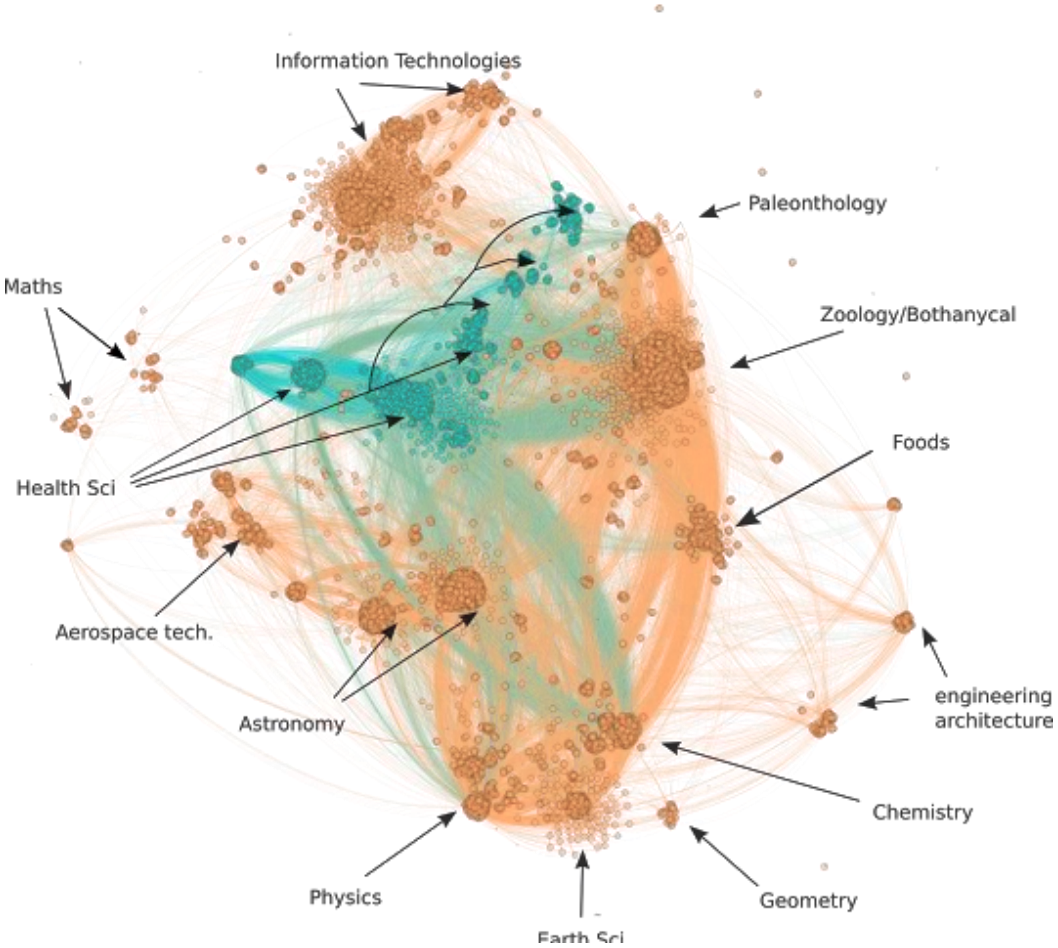
The second stage of our analysis was concerned with the web links connecting the different articles. We created a new graph in which the articles are the nodes and the weblinks are the edges. This process highlighted the existence of a different kind of communities, those of groups of articles which are highly connected among themselves. This analysis was naturally limited to articles already classified as science and technology. As we wanted to see how science and technology articles are connected to other contents we looked at links to articles not classified as science and technology. For the purposes of this paper we only checked the situation with regard to the links of Health Science articles to contents not classified as science and technology.

RESULTS

As can be seen in graph 2 the communities of articles based on the analysis indicates a relative traditional division of scientific contents.

Figure 2. The communities of Wikipedia articles based on the analysis of weblinks

However when we look at the net of links of a specific field, such as Health Science, it seems to reflect a more interdisciplinary arrangement as there are strong connections to physics and chemistry (Graph 3). An analysis of the weblinks of in Health Science of non scientific article also reflects a more flexible structure of contents as among the most connected contents we find articles about Philosophy, Education, Religion, Society, Culture, Ethics and Civil Rights (Table 1). We also find links to the TV series Grey's Anatomy a phenomenon that reflects how popular culture is intertwined with scientific knowledge on the Web. There is a need for further analysis to evaluate the significance of this tendency.



Figuere 3. The net of weblinks between Health Science and other scientific content (highlighted in blue)

Table 1

Non scientific articles most connected to Health Science articles

Article	No. of weblinks in Health Science	Article	No. Of weblinks in Health Science
ISBN	247	Arte	14
Filosofía	35	Política	14
Anatomía_de_Gray	31	Derechos_humanos	13
Alimento	25	Cristianismo	12
Escritor	21	Alimentación	12
Tabaco	18	Dios	11
Educación	18	Raza	11
Aprendizaje	18	Meditación	11
Población	18	Internet_Archive	10
Religión	17	Budismo	10
Familia	17	Símbolo	10
Persona	16	Sociología	10
Sociedad	16	Ganado	10
Cultura	15	Felicidad	10
Ética	14	País	10

DISCUSION AND CONCLUSIONS

Comparing these results with similar procedures carried out on school textbooks, suggests that there might be a difference between scientific culture in traditional educational aids and on the Web. Scientific content in school textbooks is characterized by a traditional disciplinary division, in spite of efforts to integrate interdisciplinary approaches and significant social contexts. It seems that in Wikipedia there is more flexibility than in traditional academic contexts.

Our first preliminary results with regard to the Wikipedia give the impression that the distribution of scientific contents does not depend solely on traditional academic disciplines, but tends towards a more interdisciplinary structure, and maybe even allows the introduction of current social concerns. It is important to remember that the links are set up by the authors of the articles and that they reflect a collective perception of the relations among different contents. It is difficult to predict how the educational process will be affected by tools such as Wikipedia. In order to confirm our hypothesis there is a need to extend and elaborate more precise methodologies in order to analyze the links between scientific and technological articles and other contents on the Wikipedia.

REFERENCES

Abd-El-Khalick, F., Waters, M., & Le, A. P. (2008). Representations of Nature of Science in High School Chemistry Textbooks over the Past Four Decades. *Journal of Research in Science Teaching*, 45 (7), 835–855.

Boldwin Mark B. and Miller T. (2007). Wikipedia's Role in Science Education and Outreach. *Eos, Transactions American Geophysical Union*, Chernov, S.,Iofciu, T., Nejdl, W. & Zhou, X. (2006). Extracting semantic relationships between wikipedia categories. *Proc. of Workshop on Semantic Wikis (SemWiki).* estudian los enlaces entre categorías de la wikipedia.

Chiappetta, E. L., Fillman, D. A., &. Sethna, G. H. (1991). A Method to Quantify Major Themes of Scientific Literacy in Science Textbooks. *Journal of Research in Science Teaching*, 28, 713-725.

Dede, C. (2008). A Seismic Shift in Epistemology. *EDUCAUSE Review*, vol. 43, no. 3, 80–81.

Figuerola, C. G. and Mateos Sánchez, M. (2009). Aplicación de técnicas de clustering en la recuperación de información web, Gijon: Trea.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature* 438, 900-901.

Groves, T., Quintanilla M. Á., Escobar M. (2012). Scientific and Technological Culture in Secondary Education Textbooks in Spain. In: *Os Manuais escolares e os jóvenes: tédio ou curiosidade pelos saberes?.* Lisbon: Universidade Lusófona de Humanidades e Tecnologias.

Okoli, C. (2009). A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals. *Proceedings of the 2009 Third International Conference on Digital Society*, 155-160.

Quintanilla, M.Á. (2012). Cultura, Tecnología e innovación. In E. Aibar & M. A. Quintanilla (Eds.), *Ciencia, tecnología y sociedad. Enciclopedia Iberoamericana de Filosofía*. Madrid: Trotta.

Quintanilla, M. Á., Escobar M. Groves T, Montero Becerra, A. Palacios Sánchez R. Montañés Perales O. and McBride O., *Scientific and Technological Culture in ESO Textbooks*. Salamanca: El Instituto Universitario de Estudios de la Ciencia y la Tecnología, 2011.

Rosvall, M., Axelsson, D., & Bergstrom, C.T. (2009) The map equation, *European Physical Journal Special Topics*, 178, 13 -23

Svoboda, E. (2006). One Click Content, No Guarantees . *IEEE Spectrum* vol. 43, Issue 5, 64-65.

Vesterinen, V.M., Aksela, M., & Lavonen, J. (2011). Quantitative Analysis of Representations of Nature of Science in Nordic Upper Secondary School Textbooks Using Framework of Analysis Based on Philosophy of Chemistry. *Science & Education, Online First*. http://www.springerlink.com/content/r71913nk28771142/?MUD=MP