Wisdom of the Crowds: Decentralized Knowledge Construction in Wikipedia

Ofer Arazy, Wayne Morgan, and Raymond A. Patterson
University of Alberta, School of Business

Abstract

Recently, *Nature* published an article comparing the quality of Wikipedia articles to those of Encyclopedia *Britannica* (Giles 2005). The article, which gained much public attention, provides evidence for Wikipedia quality, but does not provide an explanation of the underlying source of that quality. Wikipedia, and wikis in general, aggregate information from a large and diverse author-base, where authors are free to modify any article. Building upon Surowiecki's (2005) *Wisdom of Crowds*, we develop a model of the factors that determine wiki content quality. In an empirical study of Wikipedia, we find strong support for our model. Our results indicate that increasing size and diversity of the author-base improves content quality. We conclude by highlighting implications for system design and suggesting avenues for future research.

1. Introduction

Wiki is a recent technology that gained prominence in the early 2000s. Wiki allows users to edit a webpage and submit new versions, immediately replacing the previous version (which is archived) (Leuf and Cunningham. 2001). One of the most prominent uses of wiki technology is Wikipedia, the online encyclopedia. In five years, Wikipedia has become one of the most popular websites (Bryant et al. 2005), and achieved quality similar to the leading print-based encyclopedia, unexpectedly perhaps for a system that allows any user to edit and overwrite web page content.

The recent interest in wikis goes well beyond the specific technology. Wikis represent an approach – referred to by Arazy et al. (2005) as "Open Content Systems" - that is poised to transform the way in which knowledge and knowledge-bases are constructed. Evans & Wolf (2005) describe the de-centralized and social approach to knowledge accumulation, which owes much to the open source movement, as a revolutionary approach to the delivery of knowledge goods.

One of the most compelling explanations for Wikipedia' success is, in short, "the wisdom of the crowds" (WoC). Surowiecki (2005) suggests that the aggregate knowledge of a large group is superior to the knowledge of one or a few experts. A simple example illustrates the WoC principle: if many people are asked to estimate a person's weight, the average of their estimates very likely will be close to the person's actual weight, and closer than (or at least as good as) the estimate an expert would give. However, not all crowds are wise. We propose that Wikipedia and wiki technology embody and promote the WoC principles, and attribute Wikipedia's success to its ability to effectively leverage the wisdom of the crowds.

In this paper we operationalize the WoC framework, and test its validity using Wikipedia data. Our findings suggest the WoC principles indeed determine Wikipedia article quality. The main contribution of this paper is providing an explanation of how open content systems successfully achieve high quality content. Wiki systems developers and administrators could utilize our findings to direct systems' design.

The paper continues as follows: in Section 2 we review related work; Section 3 presents our study's research question; Section 4 describes the empirical design; in Section 5

we report on the results of the study; finally, Section 6 concludes by discussing the implications of our findings and highlighting areas for future research.

2. Related Work

"If you put together a big enough and diverse enough group of people and ask them to 'make decisions affecting matters of general interest', that group's decision will, over time, be 'intellectually superior to the isolated individual', no matter how smart or how well-informed he is." (Surowiecki, 2005, p. *XVII)*

Surowiecki (2005) synthesizes theoretical and empirical results from various fields into what he refers to as "*The Wisdom of the Crowds*" (WoC). Surowiecki provides numerous examples, from an array of research disciplines ranging from biology to behavioral economics, which demonstrate that the aggregate knowledge of a large and diverse group is superior to that of one or a few experts. Three principles are essential, according to Surowiecki (2005) for assuring that the aggregate contributions of a crowd are high quality: (a) a large number of contributors and opinions (e.g. crowd size), (b) diversity of ideas and opinions, and (c) appropriate mechanisms for aggregating the opinions (e.g. the participants are able to express their opinions independent of influence from others, and aggregation techniques combine the independent opinions). Surowiecki (2005) asserts that wiki technology is an appropriate mechanism for content aggregation. This study focuses on: (a) the number of contributors and opinions (e.g. size), and (b) diversity of ideas and opinions. We discuss these below.

The number of people (i.e the number of ideas and opinions) in a crowd directly impacts the crowd's aggregate knowledge. The accuracy of judgments of a statistical group is best explained by reference to the Jury Theorem (Condorcet, 1976). In a group of *n* people, making a binary decision, where each person has the same probability *p* of choosing the right answer, and people's choices are independent, the probability that the group will reach the correct answer is $P_n = \sum_{h=(n+1)/2}^{n} \left[n!/h!(n-h)! \right] p^h (1-p)^{n-h}$. When probability of an individual's correct answer is higher than 50%, P_n converges with *n*, group size, to 1 (the larger the group, the quicker the convergence). This effect has been demonstrated by numerous studies (Surowiecki 2005; Sunstein 2006).

Diversity also impacts the crowd's wisdom and is expected in large groups. As Surowiecki (2005) notes, diversity adds perspectives that would otherwise be absent. Group diversity also reduces some more destructive aspects of group decision making, such as groupthink and conformity. Following organizational theorist J.G. March, diversity enables groups to continue to learn as new perspectives and information are introduced (March 1991). Mannix and Neale (2005), reviewing fifty years of diversity research, point out how differences in group members' skills and knowledge are associated with increased group performance.

If wiki systems, and specifically Wikipedia, embody the WoC principles, then we would expect Wikipedia articles to exhibit high quality. Viegas et al. (2004) demonstrate that vandalism is quickly corrected. Emigh & Herring (2005) study the format and writing-quality of Wikipedia articles, and report that Wikipedia content maintains high standards, suggesting this as evidence for content quality. Stvilia et al. (2005) demonstrate that Wikipedia's

decentralized quality control mechanisms indeed identify and correct many quality threats. Finally, a recent news article in *Nature* (Giles 2005) directly tests Wikipedia quality, based on an expert peer review of a sample of articles, and finds Wikipedia achieves a level of accuracy comparable to Encyclopedia *Britannica*[1].

While these recent empirical findings regarding Wikipedia strengthen our confidence in the quality of Wikipedia content, they provide little insight on the factors driving that quality. O'Reilly (2005), in a column on his website, suggests that the WoC phenomenon may be responsible for wikis' success. Although this, to date, has not been demonstrated in research, there are initial indications that WoC principles (size, diversity, and aggregation mechanisms) may be associated with Wikipedia's content quality. Some (e.g. Lih 2004, Stvilia et al. 2005) argue that the number of contributors and edits in an article is associated with an article's quality. A large author-base ensures diversity of opinions (Surowiecki 2005), and Wikipedia contributors exhibit diversity in terms of background and interests (Bryant et al. 2005). Diversity of opinion is evident in an article's discussion page (where conflicting opinions are deliberated), and high-quality pages (i.e., those selected as Featured Articles) have larger discussion pages (Stvilia et al. 2005).

3. Research Question

To date, there is little empirical evidence that WoC principles indeed determine Wikipedia article quality. The objective of this study is to asses the extent to which the WoC principles explain Wikipedia article quality. The aggregation mechanism is an inherent feature of wiki technology, and thus we do not empirically examine its existence in Wikipedia. Crowd size and diversity, on the other hand, may vary greatly across wiki articles and wiki implementation, and thus we pose the following research question: "to what extent do crowd size and diversity impact Wikipedia article quality?"

4. Research Method

To investigate the determining factors of Wikipedia's quality and validate the proposed framework, we conducted an empirical study, utilizing the 42 Wikipedia articles included in the *Nature* study (Giles 2005). We choose this sample (using the full-text of these articles) because it provides us with independent quality assessment of articles, obtained through an objective process, and published in a highly-respected journal.

The notion of information quality is multi-faceted, and the most referenced quality dimensions are: accuracy, completeness, relevance, and timeliness (Kim et al 2005). Giles (2005) reports the number of errors per article, and these errors include inaccuracies and missing data. Since we can assume relevance (the articles were judged by experts on the topic) and timeliness (a recent version of the articles was used), Giles' measure captures an

[1] It is worth mentioning that the article was followed by a debate between Britannica and Nature, where Britannica challenged the validity of the study (see http://www.nature.com/nature/britannica/eb_advert_response_final.pdf), and Nature defended their methodology and findings (see www.nature.com/nature/britannica/index.html).

article's (lack of) quality[2]. To transform the error counts into a quality measure we use the inverse function, thus $Quality = 1/Number\ of\ Errors$.

We operationalize the Size and Diversity WoC principles as follows. Size is estimated through two proxies (1) the number of unique contributors to an article (*number of authors*), and (2) the total number of contributions (*number of edits*). Diversity would ideally be measured through the variance of background, networks, education, and skills for an article's authors (Mannix & Neale 2005); however, such data is not available in Wikipedia. Instead, we employ two proxies: (1) the number of words in an article's discussion page, and (2) the number of edit wars (where Wikipedia's defines an edit war as three edits by a particular user within 24 hours, with edits from other users in between). The greater the variance in opinions is, the more likely the authors are to engage in edit wars and argue their opinions on the discussion page. In order to rule out the explanation that the number of errors was affected by an article's length, as proposed by (Stvilia et al. 2005), we test for this effect.

The complete *Nature* sample includes 42 articles[3], 4 of which are reported to have no inaccuracies. Since our measure of quality is *1/Number of Errors*, we removed articles without errors from our sample. To calculate the measures of Size and Diversity, we harvested Wikipedia articles, relying heavily on articles' history logs. We conducted a thorough data cleansing process, and removed all non-human authors (i.e. software bots) and edits made by these bots. In addition, since Wikipedia articles change content continually, it was important that the data collected corresponds to the date of quality assessments. We obtained the exact cutoff dates (mid-October 2005) of the articles from the Nature article author through personal correspondence.

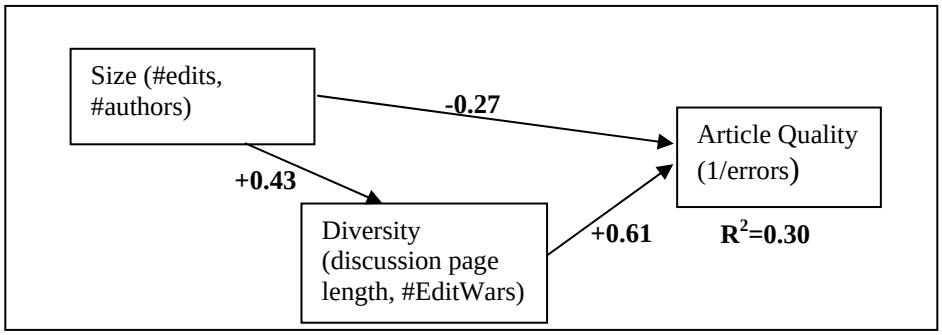
5. Results

We analyzed our data using partial least squares (PLS), since this statistical method tests for causal relations and allows collapsing of several measures into one latent construct. The results support our model. First, the items grouped into a latent construct do represent one concept (the number of authors and the number of edits are highly correlated, 0.94; similarly, discussion page length and the number of edit wars are highly correlated, 0.78; both correlations are significant at 0.01). Second, a **causal** relationship between *Diversity* and the dependent variable (*Quality*) is established, and is statistically significant. Finally, *size* has a significant effect on *Diversity*. Figure 1 below graphically presents the PLS results. *Size* has a positive effect on *Diversity*, and *Diversity* has a positive effect on *Quality* (path coefficients are 0.43 and 0.61 respectfully, both significant at 0.05). Overall, the proposed model explains 30% of the variance in Article Quality ($R^2 = 0.302$).

---

[2] Although not all errors represent the same threat to quality and some errors are more serious than others, judging error seriousness is highly subjective and requires domain expertise in a wide array of fields. We, thus, assigned equal weight to all errors.

[3] For a complete list of the articles assessed by *Nature* please refer to (Giles 2005a).



It is interesting to note that the coefficient for the path linking *Size* to *Quality* has a negative sign. This result could be explained by the conflicting effects of diversity. *Diversity* is a multi-dimensional construct, where some diversity dimensions (e.g. skills and knowledge) have a positive effect on group decision-making, while other dimensions (e.g. age, gender, status) have negative effects on group performance (Mannix & Neale, 2005). Group size increases both types of diversity, and the correlations between *Size* and *Quality* is very close to zero. In the PLS model, the positive effect of group size on quality is represented through the link to opinion diversity, and the remaining negative effect is evident in the direct relationship between size and quality.

6. Discussion and Conclusions

The success of Wikipedia has attracted much public attention recently, since it seems counter-intuitive that a large and diverse author-base could generate high-quality content in a seemingly unorganized and uncontrolled manner. Several alternative explanations for this phenomenon are possible. Some see the key to Wikipedia's success in the cohesiveness of the Wikipedia community and authors' motivations, while others focus on the specific decentralized control mechanisms (e.g. WatchLists) that Wikipedia employs. In this paper, we proposed an alternative explanation that is based on the Wisdom of the Crowds framework, and demonstrated that two factors – crowd size and diversity – explain Wikipedia content quality. Our findings suggest that a simple model can account for a relatively high percentage of the variance in article quality. The main contributions of this work are in (1) proposing a model of the WoC phenomena, based on Surowiecki's (2005) framework, and (2) in demonstrating empirically that this model indeed explains Wikipedia's article quality.

Our findings have direct implications for wiki system designers and administrators. In order to achieve high quality content, it is essential that many users participate in authoring wiki pages. Also, it is important that participation levels be high (i.e. each page has many edits). To entice participation, organizations using wikis should strive to eliminate barriers (e.g. allow users to post anonymously) and provide incentives for contributions. It is important that the author-base be diverse, thus organizations should encourage cross-departmental collaboration, and perhaps inter-organizational collaboration, in wiki authoring. Lastly, it is essential that wiki systems incorporate aggregation mechanisms that will ensure the independence of users' opinions.

Although our empirical analysis was performed on a wiki system, we believe that our results generalize to other open content system (e.g. advanced discussion forums, such as

Slashdot) and collaborative software development. In future research we plan to expand on the current study and address some limitations. We employed indirect measures of crowd size and diversity, and thus it is suggested that future research explores alternative measures. Our study used a small sample of Wikipedia articles, and future research should try to generalize these results to alternative samples of Wikipedia articles and to other wiki systems. Specifically, we plan to explore these ideas in organizational settings. Lastly, it would be interesting to apply our model to other types of open content systems, such as append-based systems (e.g. discussion groups), open source systems, and other types of decentralized knowledge management systems where information quality is paramount.

References

Arazy, O. & R. Patterson. 2005. An analysis of open content systems. *Proceedings of the Fifteenth Annual Workshop on Information Technologies and Systems.* Las Vegas, Nevada. December 10-11, 2005.

Bryant, S., A. Forte & A. Bruckman. 2005. Becoming wikipedian. *Group05.*

Condorcet, 1976. Selected Writings, editor K.M. Baker.

Emigh W. & Herring S., 2005, Collaborative authoring on the web: A genre analysis of online encyclopedias. 38th International Conference on System Sciences.

Evans, P. & B. Wolf. 2005. Collaboration rules. *Harvard Business Review.* 83 (7/8): 96-104.

Giles, J. 2005. Internet encyclopedias go head to head. *Nature.* 438 (15): 900-901.

Giles, J. 2005a. Supplementary information to "Internet encyclopedias go head to head". Online at *Nature* www.nature.com/news/2005/051212/full/438900a.htm

Kim, Y., R. Kishore, & G. L. Sanders. 2005. From DQ to EQ: understanding data quality in the context of e-business systems. *Communications of the ACM.* 48 (10): 75-81.

Leuf B. & Cunningham W. 2001, The Wiki Way: quick collaboration on the web. Addison-Wesley.

Lih, A. 2004. Wikipedia as participatory journalism: reliable sources? Fifth International Symposium on Online Journalism, April 16-17, 2004. University of Texas at Austin.

March, J. 1991. Exploration and exploitation in organizational learning. *Organization Science.* 2: 71-87.

Mannix, E. & M. Neale. 2005. What differences make a difference? *Psychological Science in the Public Interest.* 6 (2): 31-55.

O'Reilly, T. 2005. What is web 2.0. Design patters and business models for the next generation of software. www.oreillynet.com, September 30, 2005.

Stvilia, B., M. Twidale, L. Gasser, & L. Smith. 2005. Information quality discussions in wikipedia. Working paper. University of Illinois at Urbana-Champaign.

Sunstein C., 2006. Infotopia, Oxford University Press.

Surowiecki, J. 2005. *The wisdom of crowds.* New York: Random House.

Viegas, F., M. Wattenberg & K. Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations, *CHI 2004.*