![DiVA logo](http://www.diva-portal.org)
This is the published version of a paper published in *Culture Unbound. Journal of Current Cultural Research*.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-91575

Crowdsourcing Knowledge
Interdiscursive Flows from *Wikipedia* into Scholarly Research

By Simon Lindgren

Abstract

Information increasingly flows from smart online knowledge systems, based on 'collective intelligence', and to the more traditional form of knowledge production that takes place within academia. Looking specifically at the case of *Wikipedia*, and at how it is employed in scholarly research, this study contributes new knowledge about the potential role of user-generated information in science and innovation. This is done using a dataset collected from the Scopus research database, which is processed with a combination of bibliometric techniques and qualitative analysis. Results show that there has been a significant increase in the use of *Wikipedia* as a reference within all areas of science and scholarship. *Wikipedia* is used to a larger extent within areas like Computer Science, Mathematics, Social Sciences and Arts and Humanities, than in Natural Sciences, Medicine and Psychology. *Wikipedia* is used as a source for a variety of knowledge and information as a replacement for traditional reference works. A thematic qualitative analysis showed that *Wikipedia* knowledge is recontextualised in different ways when it is incorporated into scholarly discourse. In general, one can identify two forms of framing where one is unmodalised, and the other is modalised. The unmodalised uses include referring to *Wikipedia* as a complement or example, as a repository, and as an unproblematic source of information. The modalised use is characterised by the invocation of various markers that emphasise – in different ways – that *Wikipedia* can not be automatically trusted. It has not yet achieved full legitimacy as a source.

Keywords: *Wikipedia*, collective intelligence, academia, encyclopaedias, citations

Introduction

This article analyses how information flows from so-called smart online knowledge systems – based on 'collective intelligence' (Lévy 1999) – and how this compares to the more traditional form of knowledge production that takes place within academia. Looking specifically at the case of *Wikipedia* and how it is employed in scholarly research, this study contributes to new knowledge about the potential role of user-generated information in science and innovation. The notion of collective intelligence is based on the idea that no single person knows everything but everyone knows something, and this collective knowledge can be harnessed through social media. People networking and sharing knowledge, experience and ideas results in a form of intelligence that, according to Lévy, is universally distributed, coordinated in real time, and constantly enhanced. This leads to an effective mobilisation of skills.

Modern digital culture makes all of us potential members of a shared virtual universe of knowledge, and the common fostering of this intelligence has the potential to make social ties the most important currency in future society. Collective intelligence, Lévy argues, can disrupt the power of government and can lead to a diversification of knowledge and creativity. Lévy argues that we will increasingly witness the development of skill zones that are 'fluid, delocalised, based on the singularities, and agitated by permanent molecular movements of association and rivalry' (Lévy 1999:5). The utopian result will be a form of real-time democracy where knowledge is no longer 'padlocked like a treasure' but instead 'pervades everything, is distributed, mediatised, spreads innovation wherever it is found' (Lévy 1999:212). This emerging knowledge system – 'the cosmopedia' – makes available 'to the collective intellect all of the pertinent knowledge available to it at a given moment, but it also serves as a site of collective discussion, negotiation, and development' (Lévy 1999: 217).

Other researchers have also pointed out similar processes of networked, nonprofit, and democratised knowledge production, including Rouse's (1991) notion of 'media circuits' as adapted by Lange (2008), Jones' (1997) concept of 'virtual settlements', Wenger's (1998) idea of 'communities of practice', and Gee's (2005) construct of 'affinity spaces'. Some of these are conceptualisations of how the 'fluid skill zones' are formed and structured, while others provide a terminology for the collaborative activities going on within them once they have been established. This conceptual redundancy – and there are certainly more terms to be found in the literature – is symptomatic of the field. A significant amount of effort has been made to name processes and patterns of online connection and engagement. It is natural for a wide array of conceptualisations to emerge in relation to new processes, and an attempt to bring the variety of overlapping theories together is needed.

Aim and Questions

While it is obvious that ICTs have the potential to dramatically transform processes of knowledge production, it is not clear how and to what extent this potential is realised. More research of real-world situations is needed. In this article, I investigate how and under what circumstances the potential that is inherent in ICT environments based on collective intelligence is, or is not, harnessed by traditional systems of knowledge production. This is done through a case study of *Wikipedia* and its relationship to the established scientific literature. Although – a user-generated online encyclopaedia that anyone can edit or contribute to – is an interesting object of study in its own right (Rosenzweig 2006; Kittur, Suh, Pendleton & Chi 2007; Fallis 2008; G. W. Lovink & Tkacz 2011; Sumi, Yasseri, Rung, Kornai & Kertész 2011), the main focus here is on the actual interaction between these types of bottom-up knowledge systems and traditional and established forms of knowledge systems. A unique dataset collected from the *Scopus* research database and processed with a combination of bibliometric techniques and qualitative analysis was used to address the following issues in quantitative terms:

1. To what extent are *Wikipedia* articles used as references in peer-reviewed academic research?
2. In which academic disciplines is *Wikipedia* most commonly used?
3. What types of *Wikipedia* articles are referenced?

After this general mapping, quantitative analysis was used to address the question of how knowledge that comes from *Wikipedia* is incorporated in academic discourses. This final and important question relates to the ways in which *Wikipedia* is used and whether or not it provides a contribution to scientific efforts that would not have been possible without a connection to collective intelligence. The overarching question has to do with whether the knowledge that can potentially flow from the social ICT platform to the academic community can be construed as merely listed information – stripped of inspiring or clarifying power – or as a powerful form of active information that can deepen and enrich the new context in which it is incorporated.

Wikipedia as a Source

One of the key characteristics of the current media landscape is the increased ability for users to create content of their own. Instead of a clear-cut division between producers and consumers, there is an increasing number of examples of 'prosumption' (Toffler 1980) and 'produsage' (Bruns 2008) as networked publics (Varnelis 2008) engage in participatory cultures (Jenkins 2006). As the cost and complexity of producing and circulating information has gone down following the development of so-called Web 2.0 technologies (Bell 2009), a much larger number of people have become involved in various forms of content creation compared to

just a few years ago. The crucial premise for much of the optimistic discourse surrounding this development (cf. Shirky 2008) is that the aggregated individual contributions will benefit the collective. As argued by Madden and Fox (2006), the new bottom-up forms can 'replace the authoritative heft of traditional institutions with the surging wisdom of crowds'.

Wikipedia can be defined in terms of what Rheingold wrote in 1994 about 'virtual communities'. It is based on the 'power of cooperation' and 'a merger of knowledge capital, social capital, and communion'. Such communities attract 'colonies of enthusiasts' because the digital platforms enable them 'to do things with each other in new ways, and to do altogether new kinds of things' (Rheingold 1994: xxi). A key aspect of this is the development and application of collective intelligence. Power, according to Lévy, 'is now conferred through the optimal management of knowledge whether it involves technology, science, communication, or our 'ethical' relationship with the other' (Lévy 1999: 1).

User-generated content appears in a variety of forms that range from the mere pooling or aggregation of information (e.g. collaborative filtering on sites like Amazon.com) to different broadcast models (*Twitter* or blogs) to interactive spaces (discussion forums or other types of collaborative platforms) (Flanagin & Metzger 2011). An important venue for such collectively produced information and knowledge is online encyclopaedias, of which *Wikipedia* is the prime example. Established in 2001, this openly editable encyclopaedia can 'rightfully claim to be the most successful example of online commons-based and oriented peer production' (O'Neil 2011: 309). As O'Neil claims, *Wikipedia* can be considered to be a mass project that has taken on several features of 'hacker' culture, the most prominent of which is the idea that management structures should be decentralised.

Wikipedia is, as of January 2014, the 6th most visited site online (Alexa 2014) and has increasingly become an accepted source of information that is quoted online as well as in court cases, traditional media, and popular literature (Langlois & Elmer 2009). It is also increasingly referred to in academic books and papers.

This development has led to a debate over *Wikipedia's* trustworthiness and validity. Through its model of peer-production (Benkler 2006), it aspires to produce neutral points of view (http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view). The use of open source software and content further underscores *Wikipedia's* attempt to exist as a forum for the creation and circulation of knowledge and information that is outside of the capitalist mode of production. Langlois and Elmer (2009: 775) argue that because it 'relies on a collaborative process to produce knowledge rather than the credentials of experts, the *Wikipedia* model puts into question traditional processes for legitimizing truth claims, such as relying on expert knowledge rather than the wisdom of the crowd.'

While *Wikipedia's* collaborative model for knowledge production through the use of a multitude of anonymous contributors has been praised, the same model has also been criticised and questioned. Researchers have repeatedly shown that a small core of dedicated individuals, rather than the alleged crowd of participants, has produced and controlled much of the content, especially during the first five years of *Wikipedia's* existence (Niederer & van Dijck 2010). Because various groups of users have distinctly different levels of permission to edit content, the site has never been the 'mythical egalitarian space' (ibid.: 1384) that it is often described as. Graham (2011: 271) argues that *Wikipedia* is marked by 'uneven geographies, uneven directions, and uneven politics' and states that:

The Wikipedia project has had unimaginable success in making freely provided information available to potentially anyone. However, the project is less successful in showing users where the gaps in representation lie. Part of this problem can be traced to the wording of Wikipedia's Neutral Point of View (NPOV) policy. The policy advises editors to 'assert facts, including facts about opinions – but [not to] assert the opinions themselves'. While this rule may function as an effective policy for many articles (e.g., fish anatomy, coliform bacteria, or Manchester City Football Club), it does not necessar[ily] work for articles about place. The countless ways of interpreting economic, social and political landscapes mean that articles that contribute to the palimpsests of place necessarily must only represent selective aspects of place in selective ways (ibid.: 279).

In addition to this, the credibility and reliability of *Wikipedia* has also been scrutinized. Francke and Sundin (2012), for example, have shown how on-going discussions about the credibility of participatory media are changing notions of what sources are suitable to use. Flanagin and Metzger (2011) have argued that many users are still not ready to leave traditional models of acquiring knowledge behind and that many people, especially those of older generations, still value expert-generated content more than its user-generated counterpart. Biddix et al. (2011), and studies referred to therein, have shown that college students often use *Wikipedia* as a key tool for their research process, but the site is also increasingly used as a source of reference material in academic research.

Wikipedia in Academia: General Mapping

For the purpose of this study, a dataset was created consisting of around 13 000 journal articles collected from the Scopus bibliographic database. The entire database – covering 19 500 journal titles from 5 000 different publishers – was queried for papers with the author '*Wikipedia*' cited in their reference lists. In order to exclude articles about *Wikipedia* itself from the dataset, papers with '*Wikipedia*' in their title, abstract, or keyword field were filtered out. A search was made for each year from 2003 to 2011, and key data about frequencies, research disciplines, and research areas were entered into a spreadsheet. While caution is required

when interpreting results from such small numbers, a steady increase in the use of *Wikipedia* as a reference can be seen. In all scientific areas, only one indexed paper per year included *Wikipedia* in its reference list in 2001 and 2002. As *Wikipedia* turned five years old in 2006, 1 445 articles per year referenced material from the site, and five years later in 2011 this number had increased to more than 9 000. Figure 1 shows the percentages of papers per year in *Scopus* with one or more references to *Wikipedia*.

Figure 1. Percentage of papers per year in Scopus with one or more reference to Wikipedia.
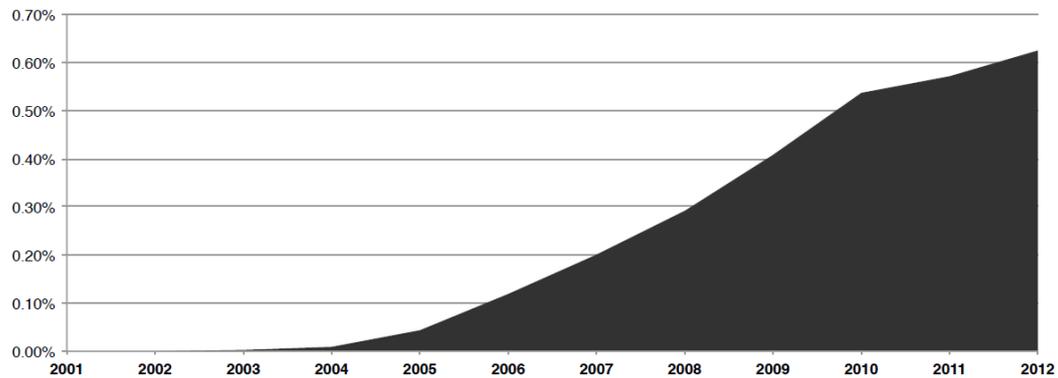

The occurrence of references to *Wikipedia* in scholarly research raises questions about how the collaborative knowledge building that takes place on this relatively open platform 'co-evolves' (Kimmerle et al. 2010) with the knowledge building that is going on within more traditional structures in academia. Langlois and Elmer (2009) have suggested that more research is needed on how the content on *Wikipedia* is circulated within, and incorporated into, other settings and how such appropriations might change the role of such content. Figure 2 shows a comparison of the annual increase (%) in the share of papers citing *Wikipedia* (grey), with the annual increase in the share of papers citing any other encyclopaedia (black). While the pattern has been levelling out in recent years, the increase in *Wikipedia* citations was quite dramatic between 2003 -2007.

Figure 2. Annual increase (%) in share of papers citing Wikipedia (grey) compared to annual increase in share of papers citing any other encyclopaedia (black)
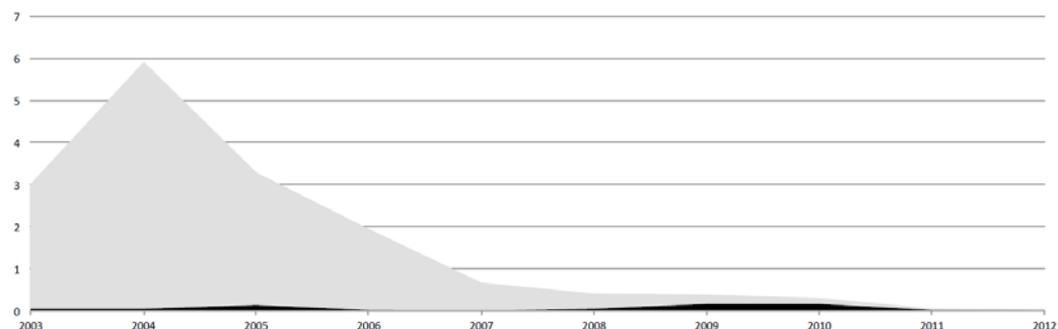
The next question to be addressed concerns in which academic disciplines *Wikipedia* is most commonly used. Figure 3 shows the percentages of papers within every subject area in the Scopus database that make one or more references to *Wikipedia*. This excludes, once again, articles that have *Wikipedia* itself as their subject matter. The general impression, which must be considered in relation to the increase in *Wikipedia* references illustrated in Figure 1, is that articles that cite *Wikipedia* are still in a clear minority ranging from around 1 to 8 out of every 1 000 articles within the respective fields of research. In Figure 3, traditionally positivist sciences (Natural Sciences, Medicine, etc.) display the lowest degree of *Wikipedia* citations while more interpretive areas like Social Science and Arts and Humanities tend to be found at the other end of the spectrum. Mathematics and its sub-field Decision Sciences rely heavily on looking up theorems and equations that are abundant and easily accessed on *Wikipedia*, thus these fields have a relatively high occurrence of *Wikipedia* citations.

Figure 3. Percentage of Scopus entries citing Wikipedia within indexed subject areas
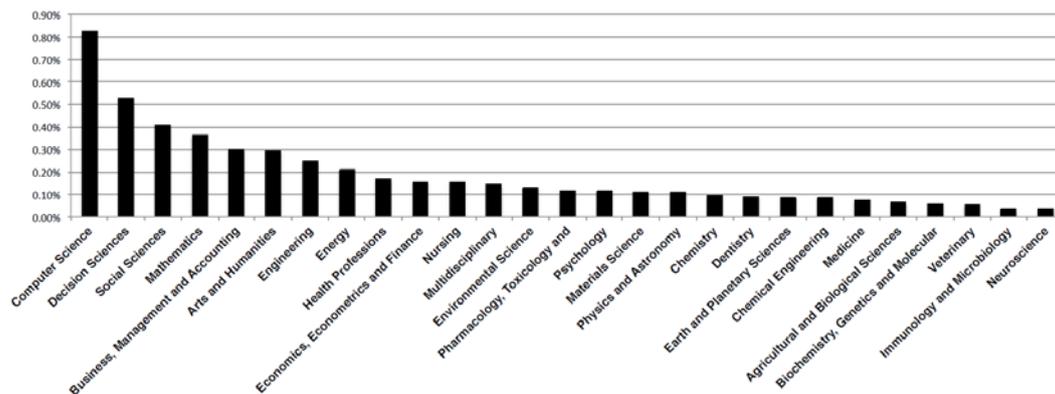
Computer Science sits at the far left of Figure 3 with 8 of every 1 000 articles citing *Wikipedia*. We can only speculate about the reason for this, but one reasonable explanation would be that this discipline, like Mathematics and Decision Sciences, builds on certain forms of knowledge – of hardware, coding languages, software, and technologies – that is sometimes better covered in *Wikipedia* than by traditional encyclopaedias. Another possible explanation could be that acceptance for looking up information on *Wikipedia* might be higher among scholars and reviewers within this inherently digital field of inquiry.

Turning to the question of what types of *Wikipedia* articles tend to be cited, the *Wikipedia* references in all articles were extracted. This was done by using regex filtering to produce a raw text list including nothing but the actual titles of cited *Wikipedia* articles. This list was then analysed using WordStat (Péladeau 2003). With this content analysis software, a list of standard English stop words were removed after which a stemming algorithm was applied to standardize the list of entries. The results of a straightforward frequency count on the resulting list, as

visualized in Figure 4, show the most often occurring words in the titles of all articles citing *Wikipedia*. The categories that come to the fore largely reflect the most represented fields (cf. Figure 3), for example, terms from Computer Science and Mathematics citations are similar to those from citations in the field of Business and Management and so on. Looking closer at the words in this context provides more information about what lies behind the different bars in the figure. The bars have been grouped and colour coded based on a rough qualitative thematisation, and this gives a somewhat more structured image of the cited articles even though the categories are not clear-cut.

Figure 4. Top words in cited Wikipedia articles
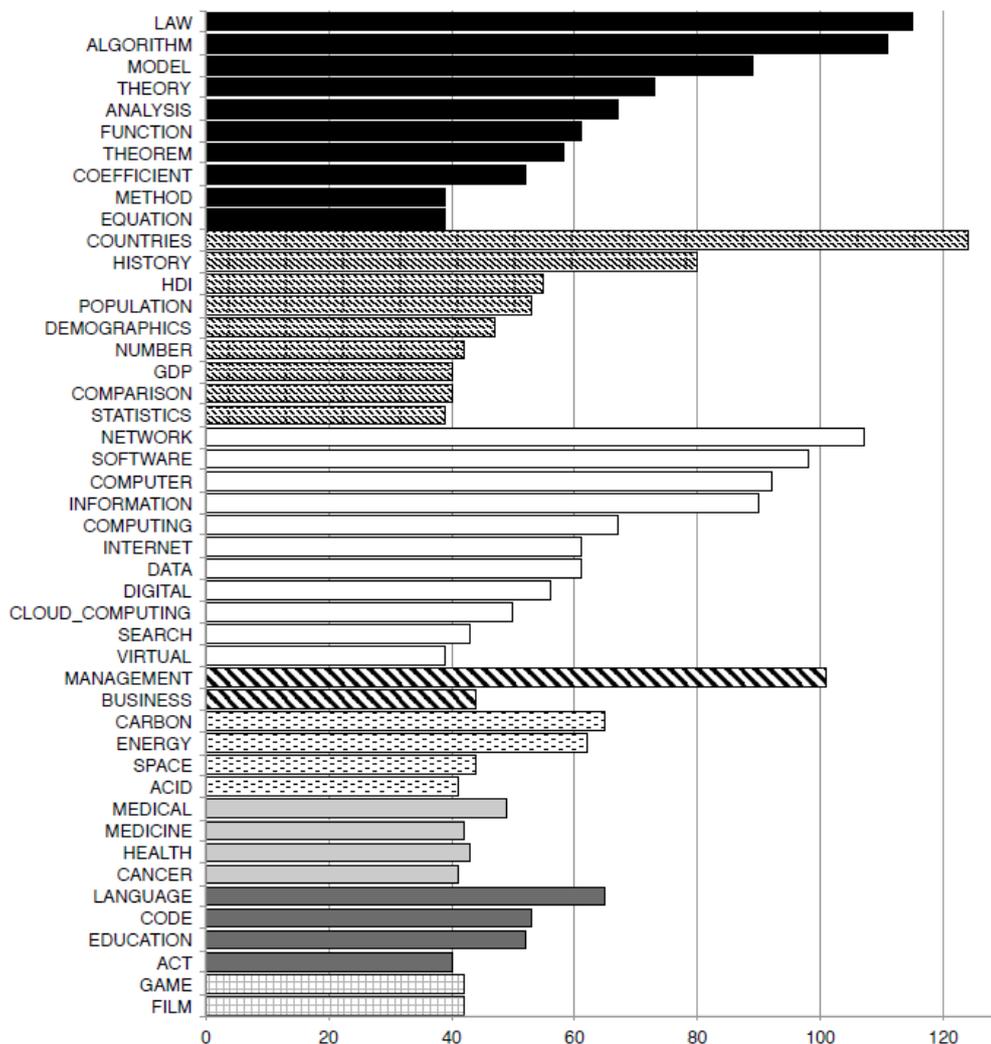
The black at the top of the graph primarily represents the relatively large number of references to articles with words like 'law', 'algorithm', 'theorem', 'coefficient', and 'equation' in their titles. Examples of frequent *Wikipedia* articles are

'Moore's law', 'Zipf's law', 'Metcalfe's law', 'Genetic algorithm', 'Greedy algorithm', 'Dijkstra's algorithm', 'Central limit theorem', 'Dominated convergence theorem', 'Spearman's rank correlation coefficient', and 'Hill equation'. This further strengthens the conclusion that *Wikipedia* tends to be used in disciplines like Mathematics and Decision Sciences for looking up and making reference to various types of principles and postulates. The third set of bars from the top, marked in white, illustrates that this type of citation behaviour extends into the field of Computer Science where top articles include 'Mobile ad hoc network', 'List of social software', and 'Cloud computing'.

The second section from the top, marked with diagonal stripes, illustrates that *Wikipedia* also seems to be employed in academic research for obtaining updated data on nations, populations, and demographics. The articles on GDP (Gross Domestic Product) and HDI (Human Development Index) are often consulted as are articles like 'List of countries by income equality', 'List of countries by military expenditures', 'List of countries by population density', and so on. In the graph as a whole, other themes that stand out are 'Management', with top articles such as 'Knowledge management' and 'Database management', as well as 'Language' with entries on 'Business execution language' and 'Swahili language' being among the most often cited. The other rough categories are Business and Management (bold diagonal stripes), Biology, Chemistry, Physics (dotted), Medicine (light grey), Media (latticed), and Miscellaneous (dark grey). All in all, this overview shows that *Wikipedia* tends to be used in academic research as a complement to, and sometimes as a replacement for, other reference works. This is especially true when it comes to current statistics because *Wikipedia* has the advantage of being constantly updated.

Wikipedia and Interdiscursivity

A key question in this article has to do with how knowledge gathered from *Wikipedia* is incorporated into academic discourse – what Latour (1987: 35) calls 'the context of citation'. This relates to the idea of interdiscursive flows, and in this case this refers to currents of discourse from a platform for user-generated content (*Wikipedia*) into a traditional context for knowledge production (scholarly research). Interdiscursive relationships are, in fact, one of the key themes in discourse studies. Assuming the social constructionist standpoint that reality can be represented in different ways entails recognising that connections between different discourses must be taken into account. Fairclough (2003: 124) writes:

> [D]ifferent discourses are one element of the relationship between different people – they may complement one another, compete with one another, one can dominate others, and so forth.

This article uses this perspective to study the points of intersection between *Wikipedia* discourse and scholarly discourse by identifying and analysing these inter-

sections in a sample of academic journal articles. What is of interest here is not the respective discourses as such, but rather the crossing points between the discourses (cf. Bjerke 2008).

The scholarly understandings that are conveyed through research papers are, in essence, a combination of elements from a number of specialised discourses that can be defined on the basis of authorship, discipline, type of source, etc. When the authors of a paper analyse their data, these discourses are brought together – they are articulated (Laclau & Mouffe 1985) – in various ways. This has to do with what Kristeva (1980) called 'intertextuality' or what Bachtin (1981) called 'dialogism'. Texts are rendered meaningful through their interdiscursive relationships with other texts. They draw on them, refer to them, contest them, assume that the reader knows them, and so on (Bachtin 1986: 69). After performing the general mapping, as outlined in research questions 1 through 3 above, the fourth research question addressed in this article revolves around analysing how *Wikipedia* cuts into scholarly discourse.

The first step is the identification of those points where *Wikipedia* enters scholarly discourse by being called upon in peer-reviewed and published research papers. These are cases of 'direct intertextuality' (Fairclough 2003: 49; Leech & Short 2007). The analysis of these points of intersection will be taken further in a second step where the incorporation of knowledge from *Wikipedia* in the research articles is studied. This tells us how the 'imported' elements (Bjerke 2008: 7) are framed in their new context. A key concept is 'recontextualisation', which is a term from educational sociology (Bernstein 1990) that has been taken up by discourse analysis (e.g. Fairclough 2003: 33). The notion of recontextualisation highlights the fact that intertextuality always involves some sort of transformation of meanings. When a formulation, an idea, or a concept is taken out of one discursive context and put into another, certain 'adjustments' need to be made in order for the piece of content to become meaningful in the new setting:

> [I]ntertextuality is a matter of recontextualization – a movement from one context to another, entailing particular transformations consequent upon how the material that is moved, recontextualized, figures within that new context (ibid.: 51).

In this article, the recontextualisations are analysed qualitatively by focusing on how knowledge from *Wikipedia* is introduced and packaged in the scholarly texts. Particular attention is paid to what Fairclough calls 'framing'. This has to do with the choices that are made about how to frame the voice of one text as it is incorporated into another.

The Framing of *Wikipedia* Knowledge in Academic Articles

Turning to the analysis of recontextualisations, a qualitative text analysis was performed on 1,799 articles. This sample included 4% of all peer-reviewed journal articles citing *Wikipedia* within each subject area indexed in Scopus. This thresh-

old was set quite roughly in order to select a reasonably sized portion of the dataset for qualitative analysis. For areas where 4% of the articles exceeded 200, the first 200 articles (sorted by 'Relevance', as defined in Scopus) were analysed.

The analysis entailed doing batch searches in the *TextWrangler* application (www.barebones.com/products/textwrangler/) to find those places in the articles where *Wikipedia* was mentioned and then reading and thematising these passages (Braun & Clarke 2006). The text segments were coded and gradually brought together into a thematic structure that ended up including the following four types of recontextualisation: (1) as a complement or example, (2) as a repository, (3) as an uncommented incorporation, or (4) as a modalised incorporation.

The first identified way of referring to *Wikipedia* articles is in the form of complementary information or examples. In these cases, pointers to various *Wikipedia* articles are included and framed as 'extra' information that goes outside of the regular references to other types of literature. This use of *Wikipedia* is illustrated in the following excerpts from research papers:

> 1.
> Lazin, Lauren (2003): Tupac: Resurrection. Paramount; see also Wikipedia 2007, s. v. Nigger (http://en.wikipedia.org/wiki/Nigger#Nigga).

> 2.
> Tryon's several language counts – 105, 110, 113 – have subsequently been cited by many authors of published and, nowadays, online overviews of Vanuatu. Wikipedia's Vanuatu, for example, has 113 languages (Wikipedia nd); the CIA's Factbook has ''more than 100''; and Ethnologue lists 110

> 3.
> For details about this, as well as further information on TV Guide's history, see the entry 'TV Guide' from Wikipedia (2006).

> 4.
> 'Open source development' is a term that was first coined in the world of software development for software whose source code was publicly available, and thus software that anyone could modify and then contribute back to the community. For more on this topic, please see: http://en.wikipedia.org/wiki/Open source software

Excerpt number 1 is an example of a case where *Wikipedia* is used in a complementary fashion alongside another reference (in this case a movie) and framed using the wording 'see also'. Similar use is illustrated in excerpt number 2 where a reference to *Wikipedia* is packaged together with other sources covering the same issue as one example among many of a certain type of knowledge. Excerpts 3 and 4 also express a related type of framing where, in both cases, *Wikipedia* entries are suggested as sources of further background information on a particular topic. Taken together, this category consists of examples in which citations of *Wikipedia* articles are used to provide additional information or knowledge in relation to the core frame of reference of the research paper in question. Related to this recontextualisation strategy, but a bit different, is the use of *Wikipedia* by

linking to it or pointing to it as a kind of a repository. This framing invokes the site as an online archive where useful bits and pieces of information are stored, hosted, and made accessible for reference. The following excerpts are examples of this:

> 5.
> Reverend Martin Niemoller's (1946) words:
> First they came for the communists,
> and I did not speak out because I was not a communist.
> Then they came for the trade unionists,
> and I did not speak out because I was not a trade unionist.
> Then they came for the Jews,
> and I did not speak out because I was not a Jew.
> Then they came for me and there was no one left to speak out.
>
> Niemoller, M. (1946). First they came. Retrieved on January 6, 2010, from http://en.wikipedia.org=wiki=First_they_came

> 6.
> A critical mass of two-dimensional (2D) bar code users has recently emerged in Japan and it can be said that these 2D bar codes (see them illustrated in Wikipedia, 2007a) have enabled connections to be made between the mobile phone and publishing industries (see Fig. 2).

> 7.
> Another map (see Wikipedia 2006) divides the country into two regions – Jesusland and the United States of Canada.

> 8.
> Gray H. The nephron. In: Anatomy of the Human Body. Philadelphia: Lea & Febiger, 1918. http://en.wikipedia.org/wiki/File :Gray1128.png. Accessed February 2, 2009.

Excerpt number 5 refers to *Wikipedia* as a place where a famous quotation can be revisited, and excerpt number 6 points the reader to the online encyclopaedia in order to be able to see illustrations of (in this case) bar codes. Excerpt 7 refers to a map that is to be found on *Wikipedia*, and number 8 recommends *Wikipedia* as the source for looking at an image originally available in a printed book. The common denominator for this form of recontextualisation is that rather than pointing to other available – and more traditional – sources for these things, the authors have opted to make reference to *Wikipedia*. The third way of framing *Wikipedia* articles within academic publications is represented by an absence of explicit recontextualisation. In these cases, an uncommented reference is made to the encyclopaedia according to the standard conventions of scholarly writing. The following set of excerpts illustrate this:

> 9.
> Shariah covers not only religious ritual, but also many aspects of day-to-day life, politics, economics, banking, business or contract law, and social issues (*Wikipedia*, 2005).

> 10.
> The Bermuda triangle is a region in the Atlantic Ocean where some aircrafts and sur-
> face vessels have disappeared. Flight 19 is the designation of five American fighters
> which disappeared in this triangle on December 9, 1945 (Wikipedia 2008).

> 11.
> The shallowness of the focus and the density of population greatly increased the se-
> verity of the earthquake (Wikipedia 2008).

> 12.
> (C)riminals have historically used churches and temples as a hiding place in times of
> trouble (Wikipedia, 2008).

> 13.
> Courier 1B, built by Philco, also launched in 1960, was the world's first active re-
> peater satellite (Wikipedia, n.d.)

In addition to these three more or less straightforward ways of recontextualising *Wikipedia* knowledge in academic articles, the fourth identified type of framing involves various modalisations. In critical discourse analysis, modality refers to the relationship between the author and what they write. In functional grammar, modality 'construes a region of uncertainty' (Halliday & Matthiessen 2004: 116) and it reflects the position of a speaker or writer in relation to what they say (Hodge & Kress 1988). By using certain 'markers of modalisation', an author or speaker to varying degrees commits to, or expresses affinity with, the information given. By looking at markers of modality in relation to how *Wikipedia* is referred to in scholarly papers, one can analyse with what level of assertion authors put forth these citations. In the thematic analysis, two levels of modalisation were coded with regard to the degree of modalisation. The following extracts are ex-amples of a softer form:

> 14.
> The concept of remix can refer to both material practices and ideas. Often associated
> exclusively with popular culture, as noted in Wikipedia, it is often understood as a
> 'hybridizing' practice in music

> 15.
> The Wikipedia entry for Unconferences is also a worthwhile resource as is the blog
> site on unconferences (www.unconference.net)

> 16.
> A recent Wikipedia entry reports that Christianity and Islam are the two largest reli-
> gions in the world, with 2.1 billion and 1.5 billion followers, respectively (Wikipedia
> 2008).

> 17.
> In fact, Wikipedia maintains a list of free and paid statistical software (List of Statis-
> tical Packages, n.d.).

> 18.
> Since the boom of 'Web 2.0' early this century, Social Networking Sites have been
> on the rise. As of November 2009, Wikipedia lists 167 of them.

The distinguishing feature for this type of framing, as opposed to the previously discussed type, is that it explicitly says something about *Wikipedia* in the sentence where the citation is made. Excerpt number 14 makes it clear that the information referred to is 'noted in *Wikipedia*', number 15 states that *Wikipedia* provides a resource that is 'worthwhile', and excerpt 16 notes that the presented statistics come from a 'recent *Wikipedia* entry'. Furthermore, extract 17 says that *Wikipedia* 'in fact' maintains the list used and number 18 emphasises that the number of social networking sites listed are 'as of November 2009 [on] *Wikipedia*'. While this soft modalisation constitutes no essential difference compared to the more straightforward way of recontextualising knowledge from the online encyclopaedia, there is another type of framing that represents a harder form of modalisation:

> 19.
> In contrast to the other serials described, this series was very popular, <u>at least according to a web-based source</u> (Wikipedia [nd]), the producer (Tabloid Jelita/Dv/Idh [nd]) and some of my neighbors in Semarang where I recorded this show while carrying out fieldwork.

> 20.
> <u>Those who preside over</u> the Drizzt Wikipedia page have written how 'Salvatore uses Drizzt to represent issues of racial prejudice' (Drizzt, n.d.). Drizzt has somehow rejected his evil nature but is often judged as evil.

> 21.
> <u>Wikipedia, written and edited collaboratively by volunteer authors in the general public</u>, provides a peek at the lay perception of library history. <u>The online article</u> for Public Libraries claims, 'The origins of the public library as a social institution have not been well explored or recorded. The institution may have been inspired by the libraries of European universities, which in turn attempted to imitate research libraries in antiquity.'

> 22.
> We used the 'List of Smart Card' directory in Wikipedia (2008) to identify relevant cases. <u>We believe this list to be comprehensive and accurate for two reasons</u>. First, we have followed smart card development over the past few years, and all the major initiatives that we are aware of are included. Second, <u>we used alternative search methods</u> (e.g., Google searches, and industry magazine listings) to identify <u>possible missing cases</u> and no additional cases were added.

This type of framing entails the use of different markers of modality that, in various ways, represent the above-mentioned 'region of uncertainty'. The underlined sections of excerpts 19 through 22 explicitly show the degree of affinity authors have with the statements they are making. A common pattern in the majority of cases where this framing is used for recontextualising *Wikipedia* knowledge in scholarly discourse is connected to the issues of the credibility and legitimacy of the site. Excerpt number 19, for example, modalises the reference to *Wikipedia* by stating that 'at least according to' this source the point in question can be made. Obviously, this wording presumes that other more certain or reliable sources exist, the use of which would not require this type of modalisation.

Excerpt 20 emphasises the form of agency underlying *Wikipedia*. The author(s) do not simply refer to the entry in question, but also make it clear that this knowledge comes from 'those who preside over' this page. This framing entails a modalisation that would be much less expected if the information were coming from, say, *Encyclopedia Britannica* or any other source that is more established. By making it clear that *Wikipedia* entries are 'written' by a group 'presiding over' certain areas of knowledge, the author(s) modalise their reference to the site by implying that other things might have been 'written' if other people were 'presiding' over the entry. This is, of course, also the case with a source such as *Encyclopaedia Britannica*, but this is less likely to be explicitly stated when referring to such sources. A similar recontextualisation is illustrated in excerpt 21 where a reference to a *Wikipedia* article is not only modalised as being an 'online article', but the author(s) of the research article also make it clear that the site is 'written and edited by volunteer authors' and that it can, therefore, be said to 'provide a peek at the lay perception' of the topic. While other encyclopaedias also provide 'peeks' at certain 'perceptions' of the world, the stronger legitimacy of these sources makes it less likely that references to them would be modalised in this way. Conversely, the use of these modalisations indicates that *Wikipedia* tends to be seen as a less reliable and potentially more biased source of information than many others. This impression is further strengthened by the observation that authors sometimes feel the need – as illustrated in excerpt 22 – to motivate why knowledge and information coming from *Wikipedia* can be 'believed' to be 'comprehensive and accurate'. It is possible that the inclusions of these motivations are sometimes the product of requests from peer reviewers who are sceptical about *Wikipedia* as a source of information.

Conclusion

This article has analysed how content is moving from today's much celebrated smart online knowledge systems – based on the wisdom of crowds (Surowiecki 2004) – into established processes of knowledge production. The case that has been highlighted here is the use of *Wikipedia* as a source of material in scholarly research papers. Beyond the hype of social media, actual studies such as the one presented here are needed to better understand the development of this phenomenon. Without this type of knowledge, we would be left with what Lovink (2002: 10) fittingly calls 'vapor theory'. Assessing the actual circumstances under which crowdsourced knowledge benefits scholarly research can contribute to a better understanding of the potential role of user-generated information in science and innovation.

The empirical analysis presented in this article has shown that there has been an increase in the use of *Wikipedia* as a reference within all areas of science and scholarship. This development is clearly illustrated with the data from the *Scopus*

database showing that 14 papers cited *Wikipedia* in 2003, around 1,500 cited *Wikipedia* in 2006, and over 9,000 cited *Wikipedia* in 2011. It was further shown that *Wikipedia* is used to a larger extent within subject areas like Computer Science, Mathematics, Social Sciences, and Arts and Humanities than in the Natural Sciences, Medicine, and Psychology. *Wikipedia* is used as a source for a variety of knowledge and information and as a replacement for traditional reference works.

The thematic and qualitative analysis presented here showed that *Wikipedia* knowledge is recontextualised in different ways when it is incorporated into scholarly discourse. In general, one can identify both unmodalised and modalised forms of framing *Wikipedia* citations. The unmodalised uses include referring to *Wikipedia* as a complement, as an example, as a repository, and as an unproblematic source of information. The modalised use is characterised by the use of various markers that emphasise in different ways that *Wikipedia* cannot be automatically trusted. It is said to be 'web-based', 'online', and founded on a type of authorship that differs from the traditional form. Authors using a modalised framing appear to feel obligated to motivate why they have chosen to cite *Wikipedia*. This illustrates the following key conclusion of this study: *Wikipedia* is increasingly used as a reference in scholarly research, but it has not yet achieved full legitimacy as a primary source. Traditionally positivist sciences use it less than interpretive disciplines, and those citing it sometimes feel the need – or might have been urged – to explain why they have chosen *Wikipedia* rather than other sources. Looking at the modalisations used, it seems that the biggest issues with the site are the fact that it is 'online' and that its collective and volunteer authorship might lead it express 'lay' rather than 'professional' perceptions and might make it prone to bias when only some groups 'preside' over certain pages.

The increased use of crowdsourced knowledge for academic references is not limited to *Wikipedia*. Figure 5 provides an overview of the occurrence of the micro-blogging platform *Twitter* (launched in 2006), the social network site *Facebook* (launched in 2004), the social video site *YouTube* (launched in 2005), and the blogging platform *WordPress* (launched in 2003) in *Scopus* reference lists since 2006. This figure excludes articles that discuss or analyse these services in particular or social media in general. Even though the absolute numbers are still small, the increase is obvious.

Figure 5. "Collective intelligence" site URLs in Scopus reference lists
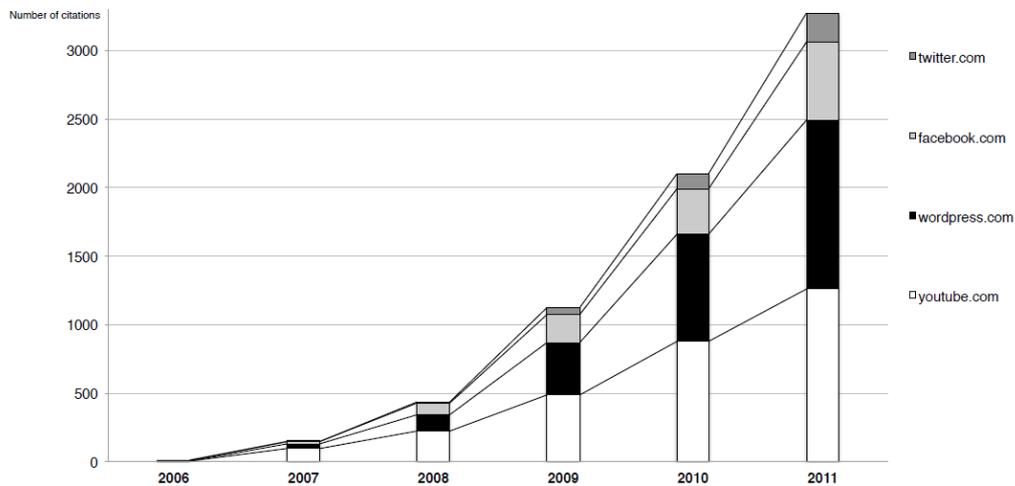
These services, when used as sources of information and knowledge, can – like *Wikipedia* – be seen as platforms for crowdsourced knowledge. But in the cases of Twitter, Facebook, *WordPress*, and YouTube we are also dealing with potentially less structured and more diverse forms of content. As academia gradually embraces the wisdom of crowds – as enabled by collective intelligence through social content platforms – the ways in which this wisdom is used will have to be negotiated within the scholarly community. As this study of *Wikipedia* – maybe the most popular collaborative online platform – shows, the use of collective intelligence sources has not changed scholarly citation practices to any significant degree. The use of these sources is still marginal, and the ways in which they are used suggest that they are only incorporated in ways that sit well with established traditions for scholarly citations. The future will present two challenges. First, scholars will have to find ways to maintain rigour in the face of increasingly diverse sources of knowledge. Second, the academic community will have to find ways to benefit from the wisdom of crowds without being discouraged by the open and vernacular nature of such wisdom.

Simon Lindgren is Professor of Sociology at Umeå University, Sweden. He researches digital culture with a focus on social connections, social organisation and social movements. He works with developing theoretical as well as methodological tools for analysing discursive and social network aspects of the evolving new media landscape. His publications cover themes like hacktivism, digital piracy, citizen journalism, subcultural creativity and learning, popular culture and visual politics. Simon is the author of *New Noise: A Cultural Sociology of Digital Disruption* (2013) and the editor of *Hybrid Media Culture: Sensing Place in a World of Flows* (2013). More information can be found at www.simonlindgren.com. E-mail: simon.lindgren@umu.se

References

Alexa (2014): *ALEXA Top 500 global Sites*, http://www.alexa.com/topsites, (accessed 13 January 2014).

Bachtin, Michail (1981): *The Dialogic Imagination: Four Essays*, Austin: University of Texas Press.

Bachtin, Michail (1986): *Speech Genres and Other Late Essays,* Austin: University of Texas Press.

Bell, Ann (2009): *Exploring Web 2.0: Second Generation Internet Tools: Blogs, Podcasts, Wikis, Networking, Virtual Worlds, and More*, Georgetown: Katy Crossing Press.

Benkler, Yochai (2006): *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven: Yale University Press.

Bernstein, Basil (1990): *Class, Codes and Control. Vol. 4, The Structuring of Pedagogic Discourse*, London: Routledge.

Biddix, J. Patrick, Chung Joo Chung & Han Woo Park (2011): 'Convenience or Credibility? A Study of College Student Online Research Behaviors', *The Internet and Higher Education,* 14:3, 175-182.

Bjerke, Flemming (2008): 'Interdiscursivity and Ideology', unpublished manuscript, Roskilde University.

Braun, Virginia & Victoria Clarke (2006): 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology*, 3:1, 77-101.

Bruns, Axel (2008): *Blogs, Wikipedia, Second Life and Beyond: From Production to Produsage*, New York: Peter Lang.

Fairclough, Norman (2003): *Analysing Discourse: Textual Analysis for Social Research*, New York: Routledge.

Fallis, Don (2008): 'Toward an Epistemology of Wikipedia', *Journal of the American Society for Information Science and Technology*, 59:10, 1662-1674.

Flanagin, Andrew J. & Miriam J. Metzger (2011): 'From Encyclopaedia Britannica to Wikipedia', *Information, Communication & Society*, 14:3, 355-374.

Francke, Helena, & Sundin, Olof (2012): 'Negotiating the Role of Sources: Educators' Conceptions of Credibility in Participatory Media', *Library & Information Science Research*, 34:3, 169-175.

Gee, James Paul (2005): 'Semiotic Social Spaces and Affinity Spaces: From the Age of Mythology to Today's Schools', David Barton & Karin Tusting (eds): *Beyond Communities of Practice: Language, Power and Social Context,* New York: Cambridge University Press, 214-232.

Graham, Mark (2011) 'Wiki Space: Palimpsests and the Politics of Exclusion', Geert Lovink & Nathaniel Tkacz (eds): *Critical Point of View: A Wikipedia Reader*, Amsterdam: Institute of Network Cultures, 269-282.

Halliday, Michael. A., & Christian M. Matthiessen (2004): *An Introduction to Functional Grammar*, London: Arnold.

Hodge, Bob & Gunther R. Kress (1988): *Social Semiotics*, Ithaca, N.Y.: Cornell University Press.

Jenkins, Henry (2006): *Convergence Culture: Where Old and New Media Collide*, New York: New York University Press.

Jones, Quentin (1997): 'Virtual Communities, Virtual Settlements & Cyber-Archaeology: A Theoretical Outline', *Journal of Computer Mediated Communication*, 3:3, 35-49.

Kimmerle, Joachim, Johannes Moskaliuk, Andreas Harrer, & Ulrike Cress (2010): 'Visualizing Co-Evolution of Individual and Collective Knowledge', *Information, Communication & Society,* 13:8, 1099-1121.

Kittur, Anniket, Bongwon Suh, Bryan A. Pendleton, & Ed H. Chi (2007): *He Says, She Says: Conflict and Coordination in Wikipedia*: http://www-users.cs.umn.edu/~echi/papers/2007-CHI/2007-Wikipedia-coordination-PARC-CHI2007.pdf, (accessed 13 January 2014).

Kristeva, Julia (1980): *Desire in Language: A Semiotic Approach to Literature and Art*, New York: Columbia University Press.

Laclau, Ernesto & Chantal Mouffe (1985): *Hegemony and Socialist Strategy*, London: Verso.

Lange, Patricia (2008): 'Publicly Private and Privately Public: Social Networking on YouTube', *Journal of Computer-Mediated Communication,* 13:1, 361-380.

Langlois, Ganaele, & Elmer, Greg (2009): 'Wikipedia Leeches? The Promotion of Traffic through a Collaborative Web Format', *New Media & Society,* 11:5, 773-794.

Latour, Bruno (1987): *Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge, MA.: Harvard Univ. Press.

Leech, Geoffrey N. & Mick Short (2007): *Style in Fiction: A Linguistic Introduction to English Fictional Prose,* Harlow: Longman.

Lévy, Pierre (1999): *Collective Intelligence: Mankind's Emerging World in Cyberspace,* Cambridge, MA: Perseus Books.

Lovink, Geert & Nathaniel Tkacz (2011): *Critical Point of View: A Wikipedia Reader*, Amsterdam: Institute of Network Cultures.

Lovink, Geert (2002): *Dark Fiber: Tracking Critical Internet Culture*, Cambridge, MA: MIT Press.

Madden, Mary, & Fox, Susannah (2006): 'Riding the Waves of "Web 2.0."', *Pew Internet and American Life Project* website: http://www.pewinternet.org/~/media/Files/Reports/2006/PIP_Web_2.0.pdf.pdf, (accessed 13 January 2014).

Niederer, Sabine & José van Dijck (2010): "Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System", *New Media & Society,* 12:8, 1368-1387.

O'Neil, Mathieu (2011): 'Wikipedia and Authority', Geert Lovink & Nathaniel Tkacz (eds): *Critical Point of View: A Wikipedia Reader*, Amsterdam: Institute of Network Cultures, 309-324.

Péladeau, Normand (2003): *WordStat: Content Analysis Module for SIMSTAT,* Montreal, Canada: Provalis Research.

Rheingold, Howard (1994): *The Virtual Community: Homesteading on the Electronic Frontier*, London: Secker & Warburg.

Rosenzweig, Roy (2006): 'Can History be Open Source? Wikipedia and the Future of the Past', *The Journal of American History*, 93:1, 117-146.

Rouse, Roger (1991): 'Mexican Migration and the Social Space of Postmodernism', *Diaspora,* 1:1, 8-24.

Shirky, Clay (2008): *Here Comes Everybody: The Power of Organizing without Organizations*, New York: Penguin Press.

Sumi, Róbert, Taha Yasseri, András Rung, András Kornai, & János Kertész (2011): 'Edit Wars in Wikipedia', *2011 IEEE International Conference on Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust (Socialcom '11)*, Los Alamitos, CA., Washington, DC: IEEE Computer Society, 724–727.

Surowiecki, James (2004): *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Society and Nations*, London: Little, Brown & Co.

Toffler, Alvin (1980): *The Third Wave*, London: Collins.

Varnelis, Kazys (2008): *Networked Publics*, Cambridge, MA: MIT Press.

Wenger, Etienne (1998): *Communities of Practice: Learning, Meaning, and Identity*, Cambridge: Cambridge University Press.