

Fort Hays State University

## FHSU Scholars Repository

---

TILT Faculty Publications

TILT - Teaching Innovation and Learning  
Technologies

---

1-1-2011

### Deconstructing Wikipedia: Collaborative Content Creation in an Open Process Platform

Andrew P. Feldstein

*Fort Hays State University*, [apfeldstein@fhsu.edu](mailto:apfeldstein@fhsu.edu)

Follow this and additional works at: [https://scholars.fhsu.edu/learning\\_tech\\_facpubs](https://scholars.fhsu.edu/learning_tech_facpubs)



Part of the [Education Commons](#)

---

#### Recommended Citation

Feldstein, A. (2011). Deconstructing Wikipedia: Collaborative Content Creation in an Open Process Platform. *Procedia - Social and Behavioral Sciences*, 26, 76–84. <https://doi.org/10.1016/j.sbspro.2011.10.564>

This Article is brought to you for free and open access by the TILT - Teaching Innovation and Learning Technologies at FHSU Scholars Repository. It has been accepted for inclusion in TILT Faculty Publications by an authorized administrator of FHSU Scholars Repository.

COINs2010: Collaborative Innovation Networks Conference

# Deconstructing Wikipedia: Collaborative Content Creation in an Open Process Platform

Andrew Feldstein\*

*Virginia State University, P.O. Box 9209, Petersburg, VA 23806 USA*

---

## Abstract

Collaboration in Wikipedia articles has widely been touted as a great leap forward and an example of how technology can be leveraged to improve collaborative processes. If we focus on the creation of individual articles, what does that creation process look like? Information was collected from the Revision History Statistics page of thirty Wikipedia featured articles to examine variables such as number of edits, number of editors and total edits by the largest contributors to a given article. This small pilot study suggests that the article creation process may more closely mirror the traditional writer/editor process than it does the “crowd as writer-editor”. It also raises questions about potential changes in how people view the content creation process.

© 2011 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).

Selection and/or peer-review under responsibility of COINs 2010 Organizing Committee

Keywords: Wikipedia; collaboration; persistent history; content creation

---

## 1. Introduction

There has been increased interest in the scope and influence of Wikipedia. In a number of popular accounts Wikipedia has been characterized as the product of “the crowd” ([1], [2], [3]) and the production of new Wikipedia articles has been described in terms of how many articles are created over increasingly short periods of time. Jaron Lanier has termed this conception of the almost mystical and powerful proliferation of Wikipedia articles as the Oracle Illusion. The oracle illusion refers to a mindset “in which knowledge of the human authorship of a text is suppressed in order to give the text superhuman validity” [4].

Recent articles have begun to dispel the oracle illusion; exploring the process of creation in Wikipedia from a number of perspectives. In a paper presented at Coins2009 Iba et al use a Wikipedia CollaboAnalyzer tool to identify influential Wikipedian editors they label “coolfarmers” [5]. Niederer and van Dijk look at the synergy between human resources and technological tools and assert “It is the intricate collaboration between large numbers of human users and sophisticated automated systems that defines Wikipedia’s ultimate success as a knowledge instrument” [6]. Geiger takes yet another approach and concentrates solely on non-human agents. His analysis of wikipedia editing processes reveals that bots are 22 of the top 30 most prolific editors and collectively make about 16% of all edits to the English language version each month’ [7].

This study looks specifically at human interactions. One of the primary features of a wiki platform is the ability to track the steps in the creation of a wiki page. This feature allows us to recreate the interactions that occur among Wikipedia editors as they collaborate to create articles. Wikis not only showcase content, but their inherent transparency allows editors to tap into the transactive group memory and gain an understanding of how an article has taken shape. By accessing the history page, the talkpage and the article itself editors can ‘re-experience’ the creation process. “Enabling re-experience constitutes the fundamental mechanism for learning and knowledge-building to occur online” [8]. It gives editors access to a form of instant replay; a way to provide context which will inform their participation.

Although the persistent history is primarily there to facilitate the creation process, it is also an opportunity for researchers to gain unprecedented insight into group behavior and social structure. As with Iba 2010, this paper analyzes Wikipedia "Featured Content" articles. While Iba uses these articles to identify specific roles played by individual editors, this paper will use a sample of 30 featured content articles to investigate the social structure among Wikipedians contributing to those articles. By comparing interactions among editors of award winning articles to articles identified as “needing work”, patterns emerge allowing us to explore issues of coordination and control as related to efficiency of process and the quality of Wikipedia articles. It is hoped that potential ‘best practices’ can also be identified.

Collaboration on a large scale involving a massive group of volunteer participants is not a new phenomenon. The Oxford English Dictionary is an example of an ambitious project of this nature. Simon Winchester’s account details the “thousands of volunteers (who) submitted examples of the earliest usage from public literature, which were collected, verified, and incorporated into the mammoth project” [9]. Creating the OED was a slow and painstaking process. Numerous volunteers researched the origin of words in the English lexicon and then mailed their findings to a central location where the information could be organized, verified and catalogued. The painstaking nature of the research was only one of the reasons that this was an extremely slow process. The infrastructure through which information was transmitted slowed the process even further. The mail system added days to each report as did the necessity to manually transcribe each entry [10].

The creation of articles in the Wikipedia is similar to the process of creation for the OED. Although the Oxford English Dictionary project began in 1878, the initial publication of the project wasn’t until 1928. The difference in the rate of growth can be attributed in good part to the efficiencies created by the wiki-platform and the work-flow processes built upon that platform. According to Stvilia Wikipedia’s work- flow processes “enable the use of powerful mechanisms for improving IQ (information quality) while using relatively simple collaborative technology” [9].

In particular, the wiki features of a history mechanism, notification features, and the tight integration of a discussion page with each entry, as well as other spaces for setting policy and norms (and their associated history and discussion pages) play a central role in facilitating these core factors. [9] Like Iba et al(2010) and Niederer and van Dijk (2010), the Stvilia article looks at Wikipedia as a system as it identifies and explores work flow processes that have evolved on the Wikipedia platform.

In this paper we move from the view of Wikipedia as a system and examine how Wikipedia collaboration is manifest at the level of an individual article. This exploratory study will look at the revision history of a small subset of ‘featured content’ Wikipedia articles to examine patterns manifest in the editing process. We will gather information that is readily available such as who is editing a specific article, how many edits they made, and over what period of time this occurred.

### ***1.1 Wikipedia Featured Article***

Featured content represents the best that Wikipedia has to offer. These are the articles, pictures, and other contributions that showcase the polished result of the collaborative efforts that drive Wikipedia. All featured content undergoes a thorough review process to ensure that it meets the highest standards and can serve as an example of our end goals. (Wikipedia website) A small bronze star ( ) in the top right corner of a page indicates that the content is featured. As of September 7th 2010 there were 3025 featured articles, out of a total of 3,404,587 articles on the English Wikipedia. Thirty of these featured articles were chosen for this analysis. The sample consists of the first 5 articles from the following 6 general featured content subject headings; War, Video Games, Transportation, Sports, Royalty, and Religion.

The data collected is made available by Wikipedia in the “revision history statistics” section of the article. An example of this page appears as Figure 1. Here Wikipedia provides information on each contributor including the number of edits made, whether those edits were major or minor, and the period during which each contributor made these edits. The interface also allows us to distinguish between edits made by registered members, anonymous editors, as well as automated programs known as ‘bots’. The revision history page provides an option that allows for the exclusion of edits made by bots and, since this analysis concentrates on human editors, we employed that option.

### ***1.2 Data Collection***

Two types of data were used in this analysis; direct data and extrapolated data. Direct data is information provided on the revision history statistics page that could be transcribed directly to our database. This included ‘total number of edits’, ‘major edits’, ‘minor edits’, and ‘user name’ for each individual contributor. It was decided to concentrate on ‘total number of edits’ rather than distinguish between ‘major’ and ‘minor’ edits since previous research cast doubt on the accuracy of this distinction. [9]

Extrapolated data includes information that was modified to create consistency among the articles being analyzed. The ‘first edit’ field was changed from a specific date to an absolute number whereby the actual ‘first edit’, represented on the RHS page by a specific date, was designated as ‘day one’ (1) and the most recent edit was designated as 1 plus the number of days that transpired between the first edit and the day the article was analyzed. This allowed an easy way to compare the number of days each article had been under construction. It also created an easy way to determine the duration of an individual contributor’s participation on a specific article (the absolute value of the user’s first edit was subtracted from the absolute value of the last visit).

## en.wikipedia.org, by Edits (reverse), with Page = Agatha\_Christie:\_And\_Then\_There\_Were\_None

show [100](#) / [250](#) / [500](#) / [1000](#) | [next 100](#) >>

Edits ↑	User	first edit	last edit
216 (216/0)	<a href="#">Jtomlin1uk</a>	2007-08-19 12:44	2010-05-17 14:33
68 (64/4)	<a href="#">Grover.cleveland</a>	2006-09-10 03:46	2008-12-07 18:41
61 (57/4)	<a href="#">ImperatorExercitus</a>	2009-02-20 22:40	2009-04-09 14:23
44 (24/20)	<a href="#">Accounting4Taste</a>	2007-05-27 04:57	2010-05-21 19:41
26 (23/3)	<a href="#">Ellsworth</a>	2004-05-10 19:35	2009-02-27 03:53
24 (9/15)	<a href="#">Jason.Palpatine</a>	2006-06-11 17:07	2007-11-05 07:29
22 (22/0)	<a href="#">Marktreut</a>	2007-05-06 23:34	2009-02-14 19:12
21 (21/0)	<a href="#">Ricardiana</a>	2009-03-31 15:51	2009-04-17 00:54
20 (6/14)	<a href="#">Haemo</a>	2007-06-01 08:11	2008-02-18 23:03
19 (12/7)	<a href="#">DCEdwards1966</a>	2008-10-09 15:14	2010-06-29 17:39
17 (17/0)	<a href="#">loverreading12</a>	2009-03-14 04:07	2009-03-23 09:48
15 (15/0)	<a href="#">207.200.116.14</a> (anon)	2006-01-25 22:26	2006-05-23 08:46
14 (5/9)	<a href="#">ClueBot</a> (bot)	2007-10-14 21:20	2009-11-15 20:34
13 (13/0)	<a href="#">134.53.145.103</a> (anon)	2006-08-24 13:47	2006-08-24 14:51
13 (0/13)	<a href="#">Goochelaar</a>	2009-06-05 16:15	2010-02-23 17:24
13 (13/0)	<a href="#">64.131.231.66</a> (anon)	2006-04-15 02:47	2006-04-22 05:47
13 (13/0)	<a href="#">24.8.11.15</a> (anon)	2006-08-04 07:49	2006-11-24 08:30
13 (13/0)	<a href="#">Heyguys12</a>	2009-04-07 03:31	2009-04-08 06:50
12 (0/12)	<a href="#">VoABot II</a>	2007-09-24 03:13	2008-05-02 07:52
11 (11/0)	<a href="#">Girolamo Savonarola</a>	2006-04-22 00:33	2006-08-17 20:41
11 (11/0)	<a href="#">Barek</a>	2009-07-17 16:09	2010-08-08 20:44
10 (8/2)	<a href="#">Shannermann</a>	2006-09-04 07:34	2006-10-01 06:23
10 (10/0)	<a href="#">6afraidof7</a>	2008-04-20 17:42	2008-09-16 16:49
10 (10/0)	<a href="#">Marieblasdell</a>	2007-07-24 18:00	2009-03-24 06:36
10 (0/10)	<a href="#">DuncanHill</a>	2008-09-03 11:20	2009-01-06 00:55
9 (9/0)	<a href="#">Hesloopian</a>	2009-02-10 23:00	2009-08-21 18:14



Fig 1. Wikipedia Revision History Statistics Page

## 2. Results

Tables 1 and 2 provide a summary of the variables extrapolated from the Revision History Statistics pages of the 30 featured content articles. Table 1 displays the subject area of the featured article, its title, as well as the “Total Edits” to date, the “Total Number of Editors” to date, and how long the article has been under construction. The last two columns describe the duration of the article creation process in days and years, respectively.

It is striking to note the differences in these variables for the 30 articles examined here. The total number of edits for these articles ranged from as few as 99 edits to as many as 2858 edits. The average number among the 30 articles was 740 total edits. The total duration these articles, from first to most recent edit, ranged from 105 days to 3389 days; that is a little over 3 months to a slightly more than 9 years. The average duration was 2275 days or just over 6 years.

The extrapolated data from the revision history statistics page is examined in more detail in Table 2. The first column provides the number of edits contributed by the editor who is identified as initiating the article. An initial supposition was that the editor who initiated an article would have a high level of involvement in the article’s creation. That is, if someone was motivated to begin an article on a specific topic, they might have some special knowledge or interest in the topic and contribute a large number of edits. This wasn’t supported since only five of the 30 editors initiating an article were also the largest contributors.

Table 1. Data from Revision History Statistics Page

Subject Area	Article Title	Total Edits	Total # editors	duration in days	duration in years
War1	13th Airborne Division	260	63	2432	6.66
War2	1968 Thule Air Base B-52 crash	763	116	1180	3.23
War3	1982 British Army Gazelle Friendly Fire Incident	314	33	758	2.08
War4	1984 Rajneeshee bioterror attack	671	125	1037	2.84
War5	1994 Black Hawk shootdown incident	527	85	1309	3.59
VideoG1	1080° Snowboarding	475	169	2127	5.83
VideoG2	4X	1167	286	2294	6.28
VideoG3	Agatha Christie: And Then There Were None	2858	1435	3389	9.28
VideoG4	Agatha Christie: Murder on the Orient Express	758	445	2818	7.72
VideoG5	Age of Empires	1144	595	1941	5.32
Transport1	American Airlines Flight 11	1506	611	3292	9.02
Transport2	American Airlines Flight 77	1967	713	3293	9.02
Transport3	Baker Street and Waterloo Railway	99	18	731	2.00
Transport4	Baltimore Steam Packet Company	178	76	859	2.35
Transport5	Albert Bridge, London	203	82	2124	5.82
Sports1	1896 Summer Olympics	1307	550	3084	8.45
Sports2	1910 London to Manchester Air Race	188	17	105	0.29
Sports3	1923 FA Cup Final	357	73	1655	4.53
Sports4	1926 World Series	516	90	2070	5.67
Sports5	1930 FIFA World Cup	1397	513	3003	8.23
Royalty1	Áedán mac Gabráin	208	61	2724	7.46
Royalty2	Ælle of Sussex	351	162	3235	8.86
Royalty3	Æthelbald of Mercia	305	75	2902	7.95
Royalty4	Æthelberht of Kent	324	143	3249	8.90
Royalty5	Æthelred of Mercia	255	64	2871	7.87
Religion1	Ælfheah of Canterbury	524	172	3299	9.04
Religion2	Adi Shankara	1504	615	2770	7.59
Religion3	The Age of Reason	937	285	3201	8.77
Religion4	Anekantavada	849	148	1727	4.73
Religion5	Asser	310	115	2786	7.63
Avg.		740.73	264.5	2275.5	6.23

Table 2 also reports the number of edits made by the largest contributor. The largest editor was, far and away, the most prolific editor in almost every article. While most editors contributed only one or two edits to an article, the edits made by the largest contributor ranged from 44 to 506; with an average of 201 edits by the largest contributor for each article. These same editors showed their commitment, not just in the number of edits they made, but in the time over which they made them. The shortest period was 97 days and the longest was 1522 days ( just over 4 years).

‘Total number of edits’ in each article was divided by the total number of edits by the largest contributor to that article to determine the percentage of edits by the largest contributor. In our sample the “% of edits by the largest contributor” varied from as little as 8% to as much as 82%. The average number of contributions by the largest contributor in this group is 39%. This percentage declines when articles have a relatively large number of editors.

In this context 8% still represents a solid commitment to an article. “Agatha Christie: And Then They Were None”, which is about the Microsoft video game not the book or the movie, was edited by

1425 editors; the article in our sample with the largest number of editors. With just over 2800 total edits, the average contributor would have made just two edits, however ‘jtomlinluk’ made 216 edits over 1000 days. His commitment is not surprising when you reference his Wikipedia user page and learn that his main hobby is book collecting and his specialty is crime fiction. A quick look at the RHS page for “Agatha Christie: Murder on the Orient Express”, which is also on our list of 30 articles, reveals that ‘jtomlinluk’ is the most prolific editor on the article. His 84 edits represent 11% of the total edits and they were contributed over a period of 849 days.

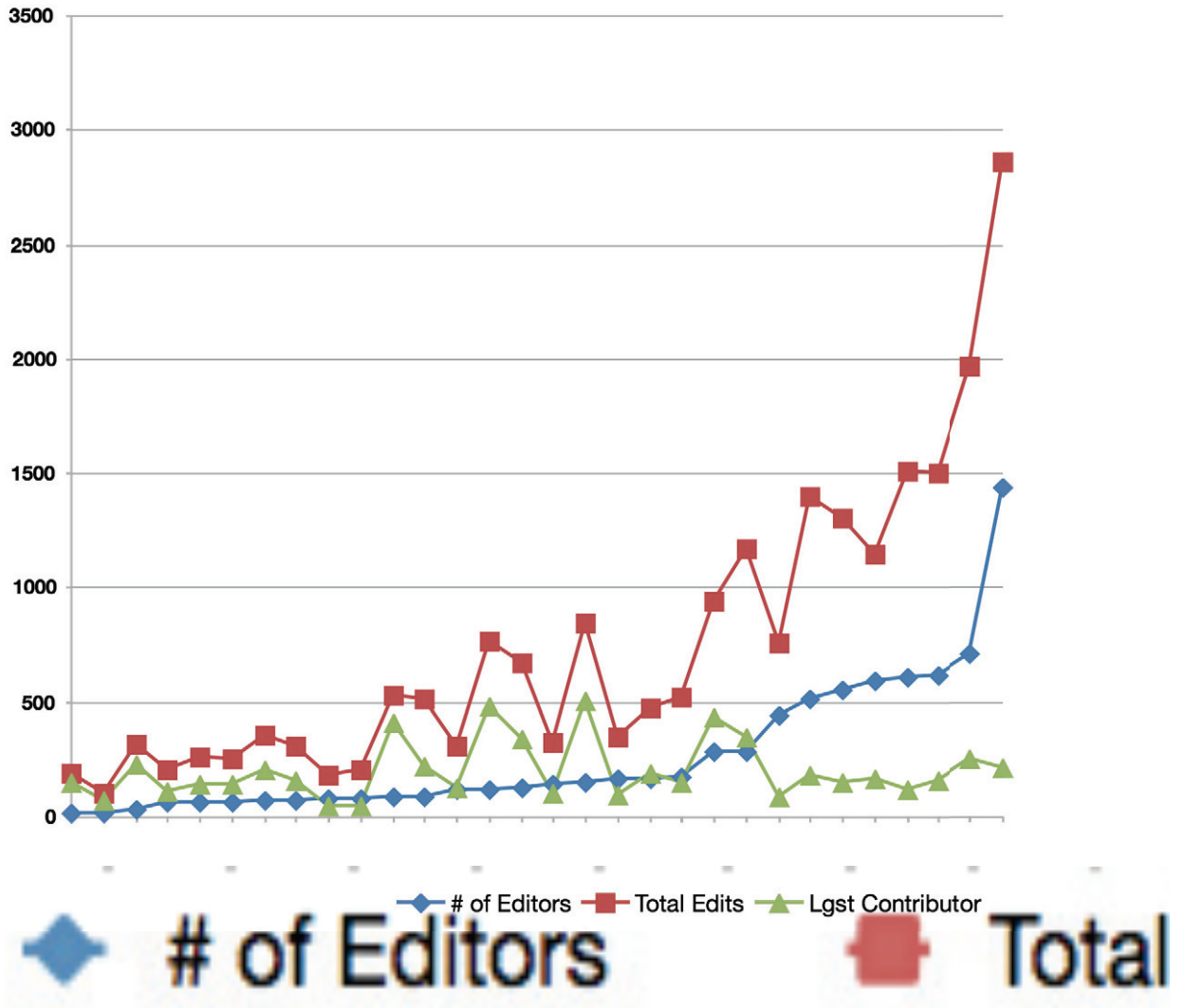
Table 2. Data Extrapolated from Revision History Statistics Page

Subject Area	# Edits by 1st Editor	Edits by Lgst Contributor	Day of 1st Edit by Lgst	Day of Last Edit by Lgst	Length of Involvement	% Edits by Lgst	% Time Lgst
War1	1	145	1568	2106	539	56%	22%
War2	5	478	507	1104	598	63%	51%
War3	231	231	1	737	737	74%	97%
War4	338	338	1	1037	1037	50%	100%
War5	6	410	10	1151	1142	78%	87%
VideoGame1	2	191	791	1224	434	40%	20%
VideoGame2	3	349	1302	1788	487	30%	21%
VideoGame3	1	216	2262	3264	1003	8%	30%
VideoGame4	1	84	1694	2543	850	11%	30%
VideoGame5	3	168	684	1284	601	15%	31%
Transport1	3	119	1755	3276	1522	8%	46%
Transport2	2	251	1616	2940	1325	13%	40%
Transport3	72	72	1	731	731	73%	100%
Transport4	50	50	1	854	854	28%	99%
Transport5	1	44	1674	2047	374	22%	18%
Sports1	54	149	2262	3077	816	11%	26%
Sports2	154	154	1	105	105	82%	99%
Sports3	7	208	321	1579	1259	58%	76%
Sports4	1	225	590	1277	688	44%	33%
Sports5	3	179	1705	2990	1286	13%	43%
Royalty1	1	109	738	2211	1474	52%	54%
Royalty2	2	95	2033	3192	1160	27%	36%
Royalty3	1	155	1678	2293	616	51%	21%
Royalty4	1	102	2034	3218	1185	31%	36%
Royalty5	1	141	1940	2747	808	55%	28%
Religion1	1	151	2155	3292	1138	29%	34%
Religion2	1	157	1206	1302	97	10%	3%
Religion3	2	432	1999	3181	1183	46%	37%
Religion4	2	506	876	1712	837	60%	48%
Religion5	6	126	1646	2658	1013	41%	36%
Avg.	31.87	201.17			863.3	39%	47%

In contrast, the 231 edits ‘Ryan4314’ made to the “1982 British Army Gazelle Friendly Fire Incident” article over a 746 day time period represents 74% of the total edits made to that article. Unlike the Agatha Christie article, which had a total of 1435 different editors, the 1982 British Army incident had only 33. Unlike ‘jtomlinluk’ whose participation began on Day 2262 of an article spanning 3389 days, ‘Ryan4314’ initiated the British Army article and stayed with it for the entire process.



Figure 2 illustrates how the number of editors contributing to an individual article can influence the behavior of the largest contributor. The graph shows that, when the number of editors reaches approximately 300, the role of the largest contributor seems to be diminished. In the first twenty-two articles the number of editors ranges from 17 to 286. Up to this point the trend line for the number of total contributions and the number of edits by the largest contributor seem to mirror that of total editors. That is, the edits made by the largest contributor seems to keep pace with the number of total edits. At article twenty-three the contributions by the largest editor seem to tail off dramatically. Whether this is because the number of edits has become overwhelming or because there are now too many editors involved is a topic for further discussion.





### 3. Discussion

The view of the collaborative process made available through the revision statistics history page provided some real insight into the Wikipedia article creation process. Here is a brief summary of the insights gleaned from this analysis. First, just because someone initiates an article, doesn't mean he or she has a lot to say about the subject. Out of the 30 articles 11 of the editors contributing the first post, did only that. Another 13 initiating editors contributed 7 edits or less.

This indicates that there may be multiple reasons someone initiates a wikipedia article. In cases where the initiating editor makes only a very few contributions they might be interested in a subject of which they have little knowledge. Their contributions are meant to serve as a way to guide the article creation process. On the other hand, when the editor who initiates the article, make the majority of the contributions, they might be more interested in sharing their knowledge than in gaining knowledge. This view can be supported by the fact that the five initiating editors who were also the largest contributors, were involved for the entirety of the article editing process.

Second, in each of the featured content articles examined here, there is a single Wikipedian whose contributions far exceed all others. The edits made by this single, largest contributor range from 8% to 82% of all edits in a given article. The majority of editors make very few edits to an individual article and most make only one contribution. This raises a very interesting question regarding the characterization of the Wikipedia as something created by 'the crowd'. If one member of the crowd makes a significantly greater contribution than other crowd members, what kind of crowd do we have?

It is possible that when we get down to the level of the individual article the nature of the crowd dynamic changes. It was evident that the control of an individual editor seemed to be reduced as more editors joined the process and maybe there is a critical mass at some point where the process is too cumbersome for individual control to be efficient or effective.

A significantly larger number of articles will need to be analyzed to determine whether we can make the assertion that, "Too many cooks spoil the stew", or whether articles with more editors make for an improved article. It is possible that there is something inherently different about topics that have a large number of editors and those with fewer editors. It would be interesting to look at the articles themselves to see if there is pattern among different topics or categories. As it stands it does seem likely that, after a certain cutoff point, the largest contributor becomes less influential.

Third, Wikipedia articles, at least the ones examined here, don't just appear in finished form. The creation process is continuous and can go on for a very long time. One might conjecture that articles that describe historic events or long defunct computer games might have a shorter time span than articles dealing with more current events but articles dealing with "Ælle of Sussex", the first king of the South Saxons circa 477, and the "American Airlines Flight 77, which deals with much more recent events have both been in process for over nine years.

This raises the question; when, if ever, can an article be considered complete? The thirty articles examined for this article have been measured in time for a duration relative to the date the article was begun until the time we imposed an artificial end date in September 2010. Who is to say how we will

compare the duration of each of these articles in another 10 years. A possible avenue of investigation might be to use Wikipedia activity on certain topics to gauge interest in certain historical events or cultural phenomena.

#### 4. Conclusion

Common wisdom has it that the Wikipedia has been created by “the crowd”. This characterization does not hold up at the level of article creation; at least not in the sense that a large swarm of Wikipedia editors descends upon a blank topic page and, when the dust settles, a fully formed Wikipedia article appears. This small pilot study suggests that the article creation process, at least, seems to more closely mirror the traditional writer/editor process than it does the “crowd as writer-editor”. It also should get us thinking about how the wiki platform, itself, is influencing the creation process. How is the process changing the way people create content and even how the process may be changing how people feel about the necessity to complete something are only a couple of issues raised here that will make for interesting areas to study.

#### References

- [1] Tapscott, D. and A. Williams (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York, Portfolio.
- [2] Shirky, C. (2009). *Here comes everybody : the power of organizing without organizations* (p. 352). London, UK: Penguin Books.
- [3] Surowiecki, J. (2006). The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economics, Societies and Nations. *Personnel Psychology*, 59(4), 982-985. Doubleday.
- [4] Lanier, J. (2010). *You Are Not a Gadget: A Manifesto*. Knopf.
- [5] Iba, T., K. Nemoto, et al. (2010). "Analyzing the Creative Editing Behavior of Wikipedia Editors Through Dynamic Social Network Analysis." Elsevier *Procedia: 2009 Proceedings in the Social and Behavioral Sciences* 2(5).
- [6] Niederer, S., & Van Dijck, J. (2010). Wisdom of the Crowd or Technicity of Content? Wikipedia as a socio-technical system. *New Media & Society*, 12(8), 1368-1387. SAGE Publications.
- [7] Geiger, R. Stuart Geiger, “The Lives of Bots.” in *Wikipedia: A Critical Point of View*. 2011
- [8] Hemetsberger, A. (2006). Learning and Knowledge-building in Open-source Communities: A Social-experiential Approach. *Management Learning*, 37(2), 187-214.
- [9] Stvilia, B., B. Twidale, et al. (2008). "Information Quality Work Organization in Wikipedia " *Journal of the American Society for Information Science and Technology* 59(6): 983-1001.
- [10] Winchester, S. (1998). *The Professor and the Madman: A Tale of Murder, Insanity and the Making of the Oxford English Dictionary*. New York, Harper Collins.