CS 498 KGK: Critique 2

The Dynamics of Mass Interaction

In this article, the authors attempt to relate and find connections between a few very basic newsgroup demographics, several conversation strategies, and the degree of interactivity of a newsgroup. The demographics in question are the size of the newsgroup (in number of unique posters), the degree of familiarity of the users actively posting in the newsgroup (i.e. the amount of repeat-posters), and whether or not the newsgroup is a moderated newsgroup. The conversation strategies that are being studied are the amount of cross-posting in the newsgroup, the average length (in number of lines) of a single post in the newsgroup, and whether or not the newsgroup has a FAQ. Lastly, the degree of interactivity is measured (quantitatively) as the average depth of a single thread of the newsgroup.

After collecting the demographic, conversation strategy, and interaction data from 500 different Usenet newsgroups, the authors analyzed the data in relation to the common ground communication theory, hoping to show that (1) large newsgroup size implies (a) more cross-posting, (b) shorter post-length, and (c) a decreased likelihood that the newsgroup will have a FAQ; (2) high familiarity implies (a) less cross-posting, (b) longer post-length, and (c) an increased likelihood that the newsgroup will have a FAQ; (3) moderation implies (a) less cross-posting, (b) longer post-length, and (c) an increased likelihood that the newsgroup will have a FAQ; (4) large newsgroup size implies less interactivity; (5) high familiarity implies more interactivity; (6) moderation implies more interactivity; (7) high cross-posting implies less interaction; (8) long post-length implies more interactivity; and (9) the existence of a FAQ implies more interactivity. After reviewing their results, the authors found conclusive evidence to confirm 1a, 1b, 2a, 2b, 3a, 3b, 3c, and 5 and to disconfirm 7 and 8. The authors were unable, with the collected data, to either confirm or disconfirm 1c, 2c, 4, 6, and 9.

I found myself agreeing (for the most part) with the authors concerning their assessment of how demographics affected the idea of a common ground and thus the conversation strategies employed by the Usenet users (the FAQ bit seemed a bit questionable, but potentially valid). Additionally, I found many of the hypotheses concerning how newsgroup demographics and conversation strategies influenced newsgroup interactivity to be very reasonable. Indeed, it makes a good deal of sense to believe that a smaller group of friendly (i.e. familiar) people would have more ‘common ground’ on which they could stand to converse than a large group of strangers; this ‘common ground’ would almost surely yield a high degree of interactivity (assuming interactivity is something that can really be measured) within the small group of friends. After the authors moved away from the abstract and attempted to find a quantitative way to represent interactivity, however, did I find myself questioning the methods employed by the authors. Saying that a newsgroup that has a large average thread depth is necessarily exhibiting a higher degree of interactivity than a newsgroup that has a lesser average seems (to me, at least) ridiculous. In my experience, extremely deep threads tend to become so through flaming or other such unproductive posting. While posters do ‘interact’ in a flame war, I think it is wrong to say that such an exchange has a high degree of interactivity. Additionally, simply because a newsgroup has a low average thread depth does not necessarily mean that that newsgroup is not very interactive. For example, if the newsgroup is a more technically-oriented one where many of the
threads are questions (which require only a single answer and thus would ideally have a depth of 1), I would still consider that newsgroup to be a highly interactive group (provided that the response times were relatively small). Ultimately, then, I believe that it is this one thing, that the idea of ‘interactivity’ was poorly defined (and quantitatively so, which most likely was a mistake), that lead to so many problems in the conclusion of the research. Had ‘interactivity’ been defined more appropriately, I believe that the results of the research would have been much different (and more to the favor of the authors).

Becoming Wikipedian
When writing this paper, the authors, it seems to me, set out to gain an understanding of how the absurdly rich source of information known as Wikipedia grows and functions. And indeed, this seems like a noble goal, for it is quite difficult for the human mind to fathom just how such a thing, which relies on no single person but instead the internet community as a whole for upkeep, can exist and thrive as it is. In order to go about this task, the authors decided to look at how an average citizen of the internet, unfamiliar with Wikipedia, discovers Wikipedia, and from there learns and grows to the point where she becomes a significant contributor to the online, cooperatively-authored encyclopedia. To do this, the authors used the ideas behind the concept of legitimate peripheral participation (LPP), a concept which outlines the process by which an outsider integrates into a group or community. Additionally, the authors organized their research and data collection in terms of Activity Theory. That is, the authors looked at the object, subject, community, division, tools, and rules of Wikipedia when conducting their research. The bulk of the research information came from the interviews, some of which were phone and some of which were email, of nine different Wikepedians. During the interviews, the Wikipedians were asked two types of question: questions concerning their interactions with Wikipedia when they first started using and editing the online encyclopedia and questions concerning how they now, after having been involved with Wikipedia long enough to better understand how the whole system works, interact with Wikipedia. The authors were ultimately able to conclude that the LPP model aligns well with how an average internet user joins the Wikipedia community and slowly becomes a Wikipedian. That is, a to-be Wikipedian begins his transformation by starting on the periphery, doing very small yet nonetheless essential tasks, and then, after interacting with other members of the Wikipedia community and learning to use some of the more advanced tools, graduates and takes on a more caretaker-esque roll.

While reading this article, I found myself mostly agreeing with everything the authors said (mostly because most of what the authors had to say was not terribly profound). Indeed, it is quite natural to think that one would begin ones initiation into Wikipediadom by first making small changes to articles you know much about or that you have a great interest in. After these initial edits, if a novice user is still interested in becoming a fully-fledged Wikipedian, she may begin creating new articles of her own; during this creation, she most likely will learn a few more rules and how to use a few of the more advanced tools. Additionally, if she at this time (during the creation of her first article) fails to use the proper formatting, she will be contacted by a more experience user who will explain how she should format her article. Also, she may begin to use the discussion pages and thus will start interacting with other members of the Wikipedia community. She may then decide to, because she likes her article so
much (and perhaps several others), start using a watch list. At this point the once-novice will have advanced to a point where she can call herself a Wikipedian, albeit a somewhat inexperienced one. This all seems like a very natural process (and indeed one quite similar to the way one would become a part of almost any other community).

Some things I did find quite interesting about the article were the discussion on motivation and also the discussion concerning why Wikipedia has been so successful. Concerning the former, I would say that Wikipedians contribute freely to Wikipedia at least partially because of altruistic motivations. However, I believe it is the sense of belonging (along with the instant gratification and feeling of efficacy) to a community that is the primary motivator. Also, I believe very much that it is the ease of editing that has let Wikipedia grow so large and be so successful (it is difficult to make people give some personal information if they are not familiar with the website).

### History Flow Visualization

For this article, it seems that, rather than having some pre-defined set of questions that they sought to answer or beliefs that they sought to verify, the authors simply wanted to create a visualization tool that would, perhaps, allow them to more easily see trends in and thus more easily examine Wikipedia. The visualization created by the authors of the article is something that they dubbed History Flow, and it manages to take textual information (revision histories and page edits of Wikipedia) and convert it into a fairly readable and intuitive graphical display. The primary component of History Flow’s display is a graph whose horizontal axis represents time and whose vertical axis represents the size of the article. Also, in addition to showing the size of the article, the vertical dimension shows, using color, shading, and other such things, which authors wrote which parts of the article (and also show which segments were kept over time and which were removed). After taking and using History Flow on several different (and, because of controversy or long revision history, almost certainly interesting) Wikipedia articles, the authors were able to show how article vandalism (which includes mass deletion, offensive copy, phony copy, phony redirection, and idiosyncratic copy) was often quickly removed, the back-and-forth behavior of an edit war (and how these edit wars are not restricted to only pages with controversial topics), how the information first posted for a given article has great longevity (i.e. it is not often deleted), how articles (even very large ones) tend to continue to grow and change, and how authorship can be easily traced by the visualization. In addition to using History Flow for their research, the authors performed a large-scale statistical analysis of the Wikipedia archives. This statistical analysis helped the authors to show that vandalism was often removed very quickly.

I thought that, for the most part, the visualization worked well. It was, however, not quite perfect. I found myself needing to re-read the section of the article explaining how the visualization worked several times before I completely understood how to ‘read’ the visualization. Perhaps the visualization would be easier to understand if the article length grew in the more natural direction of up rather than slightly confusing direction of down. Additionally, I found that the visualization often appeared to be ‘smooshed’ horizontally, making it somewhat difficult to read. I imagine a simple thing such as viewing the data on a wider screen would help this problem tremendously (a problem, then, which should not necessarily be attributed to the visualization itself). However, after I gained an understanding of how to ‘read’ the visualization (which took less than 10 minutes), I found that the graphical representations of
the revision histories to indeed be quite revealing. Additionally, after understanding History Flow, I was able to see how the visualization could be effectively be applied to situations of mass software development. Rather than simply displaying the differences between your current version of the software and the most up-to-date (and committed) version of the software, History Flow would provide a means by which the entire history of the software development (and who contributed what) could be easily viewed. Aside from the discussion of the application of History Flow to various Wikipedia articles, I found most of the other conclusions made by the authors to be good ones. If errors and vandalism of postings were not quickly fixed or removed, Wikipedia would not be read by as many people as it is. Additionally, I did not find terribly surprising the conclusion that edit wars could take place even when the topic was not controversial. The internet, if it has done nothing else, has provided a means for people to quickly and easily engage in argument with one another. So long as a topic has passionate followers (which most, like chocolate, do), bickering of that topic will occur. I did not, however, expect that Wikipedia postings, even the largest ones which contain the most information, would never cease expanding and growing. I would expect that an article would eventually become saturated with information and would no longer need constant updating. However, because most things tend to be ever-changing and in flux, the discovery that the articles associated with these things are also ever-changing seems less shocking. Lastly, although I am not myself a ‘Wikipedian’, I could see how the visualization could be very useful for those who use a watch list. A new color appearing on the History Flow visualization would let the watcher know when a closer inspection of the article (to insure that the mystery user altered the post in a positive way) was required.