The Dynamics of Mass Interaction

The paper analyzes some qualities of Usenet such as interactivity, familiarity, and demographics. First, quantitative information was automatically culled from a stratified selection of Usenet groups. Then, based on a partially connected causal model of the variables of interest, relationships were automatically fitted to the statistical data to find an approximate causal model. This model was compared to predictions made about Usenet interaction based on the “common ground” model, which basically says that interaction is linked to the amount of shared knowledge and interest of the members. Based on the comparison, the authors found two discrepancies: one, cross-posting was actually positively linked to interactivity (measured by thread depth), and short messages also had this effect.

The analysis of demographics was surprising. I wished the authors had given a GINI coefficient; it might have been more informative than the graph and table. It might also be wise to apply some basic anti-spam filtering before determining what constitutes a “single post.” I liked that the paper used a more-or-less scientific method: the hypothesis was chosen based on a well-founded theory (common ground), predictions were made, statistical measurements taken, and conclusions made. It was very clean and well presented. However, I have some arguments with the qualitative aspects of the paper. Some of the points I made note of while reading the paper were explicitly admitted in the conclusions section, but I’ll repeat them.

The primary fault I found was the use of thread depth as an indicator of interactivity. Without any kind of content analysis, it’s not clear to me at all that a deep thread indicates an “active” topic that has grabbed the attention of the group. In my experience, the most common cause of really deep threads is a flamewar, usually between just two or three posters (one or more of whom is a troll). The authors could very easily have compared email addresses to determine which threads were dyadic or had a small set of contributors, and which threads were both deep and involved a large subset of the community. This simple addition would have made the numbers pretty convincing. Even multiplying thread depth by average number of lines would help; flamewars often have flurries of pedantic, one or two line posts.

The method used to automatically analyse the causal relationships was very clever. The only recommendation I have is to use a full Bayesian network to do the same analysis, as it can provide information on more specific probabilistic queries. But again, even if the numbers in the chart are correct, we have to accept the meaning the authors attach to their variables. For instance, the chart shows that cross-posting had a positive contribution to interactivity (or thread depth). Also, if the subject of the thread really was interesting to members, then counting all the messages of the threads from the group the message was posted to might be considered multiple-counting, inflating the effect.
2 Becoming Wikipedian

This article analyzes the ways that users participate in the Wikipedia project. In particular, it looks at how users go from consumers to editors. The analysis uses two tools to do this analysis: Legitimate Peripheral Participation (LPP) and Activity Theory (AT). LPP describes how persons in the role of an apprentice or newcomer are able (or unable) to gradually take on the role of master. Several attributes of LPP are given, such as access to tools, participation with more advanced users, and hands-on practice in low-risk areas. Activity Theory describes six dimensions of the complex interaction of people attempting to achieve some goal. The paper describes how these dimensions are altered as users progress from consumer to editor.

The article's greatest strength was in identifying the differences in the growth process of Wikipedia contributors with the process that was previously identified in offline communities. The article notes, for instance, that novice users of Wikipedia are more likely to identify the other contributors as just a “collection of users” rather than a community with division of labor. I thought this was an interesting point that might affect design choices in anonymous editing and communication. Since the “induction process” is inherently different online, with no immediate physical sense of community, the attitudes of the members and how they develop will be different.

Largely though, it seemed like much of what the article had to say was obvious: novices have a limited understanding of tools, make “safe” edits in low-risk areas, are corrected by experienced users but given some leeway, etc. There may be some value in fitting these ideas into a standard model from psychology, but in itself these observations are not very revealing. It was interesting that the paper made specific mention of wiki editing features and identified how they enabled Wikipedia to be a successful community effort. Perhaps more attention could have been paid to the editing interface itself.

One way the article could be improved would be to include a comparative study of users who do not end up as such “good citizens.” It could try to find the reasons for vandalism or intentional misinformation and suggest why, how, and where it is likely to occur. It seems that this would be another area that makes Wikipedia different from physical LPP scenarios.
3 Studying Cooperation and Conflict between Authors with history flow Visualizations

The paper presents a visualisation of Wikipedia article edits named “history flow.” This visualisation presents in a graphical form the history of individual articles in all the changes made by authors over time. It employs a simple difference algorithm to show which edits are moves, insertions, or deletions. The lines of the graph are color-coded to indicate the author of the edit. By using a visualisation to attempt to grasp the entirety of an article’s history at a glance, patterns emerge which might not otherwise have been obvious. The authors noticed that they could quickly notice certain phenomena: Vandalism and repair, Negotiation, Authorship, and Temporal patterns. They found, among other things, that many kinds of vandalism are corrected very quickly, often within a few minutes; that there are frequent “edit wars” where the page is reverted between two version repeatedly before settling; that the initial text of an article tends to set the tone of the article and remains largely intact; and that pages tend to grow over time instead of stabilizing in size.

The paper, but more importantly the software it presents, is an amazing tool in its own right and a good example of the power of visualisations to bring hidden patterns to light. The authors concluded that without the tool to analyse the information graphically, some of the patterns they mention would not have been apparent. Even without having done any statistical analysis and not being intimately familiar with the Wikipedia edit process, I was able to grasp the meaning of the images and could identify in them the trends which were mentioned. Even though the visualisation algorithm is not particularly complicated, and the information it utilises is not heavily processed and is freely available, the value added to the data by being able to see it graphically is enormous.

One thing I would have liked to have seen included in the paper is the response of a veteran Wikipedia editor using the history flow software in addition to the usual watch list functionality. The history flow tool might be useful for editors to spot vandalism even in articles with which they are not immediately familiar and don’t usually watch. A scrolling version of the flow on a portion of their screen real estate could run continuously. With a small amount of training, even if the editor was busy doing something else, the “feel” of the visualisation might subconsciously trigger the urge to edit a page they may never have seen before. Or, perhaps editors might make edits that produced interesting patterns in the flow itself, which would have an unknown but perhaps aesthetically significant effect on the article.