Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations

Palo Alto Research Center

ABSTRACT
Wikipedia is a wiki-based encyclopedia that has become one of the most popular collaborative on-line knowledge systems. As in any large collaborative system, as Wikipedia has grown, conflicts and coordination costs have increased dramatically. Visual analytic tools provide a mechanism for addressing these issues by enabling users to more quickly and effectively make sense of the status of a collaborative environment. In this paper we describe a model for identifying patterns of conflicts in Wikipedia articles. The model relies on users’ editing history and the relationships between user edits, especially revisions that void previous edits, known as “reverts”. Based on this model, we constructed Revert Graph, a tool that visualizes the overall conflict patterns between groups of users. It enables visual analysis of opinion groups and rapid interactive exploration of those relationships via detail drill-downs. We present user patterns and case studies that show the effectiveness of these techniques, and discuss how they could generalize to other systems.

CR Categories and Subject Descriptors: H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction

Additional Keywords: Wikipedia, wiki, revert, graph, collaboration, user model, visualization

1 INTRODUCTION
Human existence depends on collaborative problem solving. Nations and companies depend on collaboration to solve vexing problems such as rules of conduct, investment decisions, and resource allocation. It is therefore not surprising that visual analytics must work effectively in collaborative environments. Indeed, diverse fields that utilize visual analytics such as scientific inquiries into climate change and foreign intelligence analysis on Iraq insurgents rely on collaborative problem solving to finally reach actionable conclusions. These conclusions depend on groups of people forming a coherent picture of the problems at hand, and then developing a consensus amidst conflicting user opinions and political pressures. Recently, new collaborative knowledge systems such as Wikis are being used for collaborative problem solving and knowledge gathering [2]. As conflict and coordination costs increase in such environments, visual analytic tools may be increasingly useful for users to make sense of the status of the collaborative environment.

The largest experiment of this kind is probably Wikipedia, which has become one of the most popular knowledge repositories on the Internet. A recent study comparing it against the Encyclopedia Britannica reports that much of the Wikipedia content is of high quality in spite of issues such as vandalism and factual inconsistencies [14].

Since its inception, Wikipedia has been growing at an exponential rate [5, 6, 18, 32, 47]. Over two million articles have been collaboratively edited by more than four million users in the English Wikipedia alone. However, the high level of participation comes with corresponding costs, including conflict between user factions, vandalism, spam, and expression of political ideologies.

Keeping any social structure or organization growing often requires overhead costs, such as coordination and maintenance work [24, p.160]. Wikipedia is not an exception and our earlier work showed that the overhead cost in Wikipedia has increased significantly as Wikipedia evolves over time [18]. We showed that some of these coordination costs result from user disagreements about article content, procedures, and administrative issues.

Visual analytics may provide a mechanism for users to reduce the costs of making sense of the state of collaboration and engaging in the above activities. The result of the visual analysis could help users find major opinion groups and subjects of controversy, identify mediators and abusive editors, and understand the user conflicts within these systems.

Figure 1. Revert Graph visualizes social structures in Wikipedia articles. Revert Graph for the Charles Darwin page [37]. For a discussion, see Section 6.
In this paper, we develop a user conflict model based on users’ editing histories, specifically revisions that void previous edits, known as “reverts”. Our model extracts reverts from Wikipedia editing history and composes a node-link graph where a user is denoted as a node and a revert relationship as a link. Based on this model, we developed a tool called Revert Graph that visualizes the revert relationships between opinion groups. It utilizes a force-directed layout to cluster user groups, and provides detailed drill-down to help identify specific user opinions.

We shall show that the tool can help discover and pinpoint user patterns such as the: (a) formation of opinion groups; (b) patterns of mediation; (c) fighting of vandalism; (d) identification of major controversial users and topics. The tool can be used to identify the severity and nature of a disagreement and the number and composition of the user groups involved. Figure 1 shows some example social structures discovered and characterized using Revert Graph. We believe the tool can form the basis for conflict resolution tools in the future.

The contributions of this paper include:

- the development of a user conflict model for wiki-style online knowledge repositories;
- the creation of a visualization tool that enable analysts to understand and explore user conflicts and opinion groups;
- the presentation of user patterns and case studies that show the effectiveness of these techniques.

The rest of this paper is structured as follows: First, we present related work on Wikipedia social dynamics, collaborative problem solving, conflict resolution, and visualization of social environments. Second, we introduce Wikipedia and its particular structure and statistics. Next, we introduce the user conflict model as well as the particular visualization method we used. We then demonstrate the tool by presenting interesting user patterns and case studies we have found in Wikipedia. We conclude by discussing how the tool can generalize to other systems, future directions, and final remarks.

2 RELATED WORK

Collaborative knowledge spaces and collaborative problem solving are big research areas with many researchers working on different aspects of the problem. In order for these collaboration spaces to be effective, we need to understand how conflicts arise and how conflicts are resolved in these systems. Here we summarize some past works in social dynamics and conflict resolution.

There appears to be a number of intuitions about why collaboration might improve analytics [24, Chapter 8]. First, an individual receiving information cues from a group of other analysts is more likely to be more efficient in exploring a domain area. Second, since the work covering a large domain area can be divided up into the light work of many, one can ensure important pieces of evidence are not missed. Third, knowledge specialization ensures that years of skilled training in one particular area might enable one to build up better ability to spot patterns. Fourth, the diversity of viewpoints can potentially overcome cognitive biases toward one single interpretation of the data.

Researchers have studied this area in the field of Computer Supported Cooperative Work (CSCW) and tried to understand distributed problem solving and cognition. Documents and spreadsheets often are focuses of collaboration. For example, various researchers have studied collaborative writing and commenting [21, 22]. Nardi and Miller found that spreadsheets are used as collaborative tools for problem solving in organizations [20]. Flor and Hutchins described how distributed cognition occurs in a team during software development [12]. Changes in social computing spaces have caused psychologists such as McGrath to examine theories of how groups collaborate over time, and how they resolve conflicts [19].

Indeed, communities of practice have changed greatly since the introduction of the Web and online communities [3, 7, 9]. Arguello et al. studied how individuals must act in order to receive a reply in a discussion forum [3]. Cosley et al. studied how intelligent task routing can help with individuals wishing to collaborate and create useful value to social groups [7]. Dibbell described how conflicts, vandalism, and anti-social behavior manifest itself in virtual worlds such as MUDs and MOOs [9]. Dourish and Bellotti studied how awareness is required for coordination in shared workspaces [10].

Viewed from this CSCW perspective, the rise of conflict and the costs of coordination are unavoidable in a distributed collaboration system [13] such as Wikipedia, and manifest in scenarios such as conflicts between users, communication costs between users, and the development of procedures and rules for coordination and resolution. Researchers have seen similar costs in other computer mediated communication (CMC) systems [8, 9]. Even though researchers have documented the growth of Wikipedia [5, 32], the impact of coordination costs for adding content and users has not been well studied, though Burrow et al. [5] discussed maintenance cost such as reverts and vandalism.

Perhaps the most relevant study to coordination and conflict cost is our previous study [18]. In the study, we discovered that the number of article edits is decreasing while the overhead (i.e., the number of edit intended for communication and policy making) is increasing. The study also introduced a model for the degree of conflict based on simple metrics. According to the study, the increased number of revisions made on article discussion pages is the most significant predictor for the degree of conflict in an article. This finding strongly hints that the degree of conflict and disagreement between users has a close relationship with the amount of total overhead. Motivated by earlier results, here we focus on developing a user conflict model for Wikipedia.

Visual analytics [29] offers a potential solution for understanding conflicts and coordination costs in collaborative social spaces. For example, researchers have visualized conflicts between political blogs [1] and entities in political news [4]. Researchers have also used visualization to understand social spaces by visualizing relationships amongst people in chat rooms [30], friends [17], and other social networks [23].

Visualization also offers a potential way to understand the dynamics of content creation between users of a collaborative space. Viegas et al. showed how visualization can be used to understand the editing history of a single article in Wikipedia [31]. The tool was able to reveal some social patterns such as vandalism, negotiations, edit wars, and anonymous edits.

Visual analytic tools not only can help us characterize the collaboration space such as Wikipedia, but they also potentially could offer benefits to end-users who need to understand how others in their social groups are performing, what they are paying attention to, and what conflict patterns might exist during collaboration. In other words, visual analytics applied to these social spaces will enable a kind of situation awareness [11] of the analytical topic area. These tools reveal social structures and user behavior that shed light on how we can characterize conflicts and coordination costs.

In this paper, we employ a novel method for constructing the conflict model between users in Wikipedia, and apply visual analytic techniques to understand the social relationships between these users. We show how the tool is able to find interesting patterns such as the formation of opinion groups and mediation patterns, and controversial editors.
3 INTRODUCING WIKIPEDIA

Before we explain how we have identified conflict between users in Wikipedia, we first introduce relevant Wikipedia structures and policies, and describe the data that we used in the analysis.

3.1 WIKIPEDIA DATA ANALYZED

In this paper, we used a complete history dump of the English Wikipedia that was generated on July 2, 2006. The dump included over 58 million revisions, of which 2.4 million are article entries in the encyclopedia, totaling approximately 800 gigabytes of data. To process this data, we imported the raw text into the Hadoop [15] distributing computing environment running on a cluster of commodity machines. The Hadoop infrastructure allowed us to explore new full-scale content analysis techniques quickly. We also cloned Wikipedia’s own metadata databases for direct analysis. This database allowed us to inspect Wikipedia statistics in their native format.

3.2 WIKIPEDIA ORGANIZATION

Here we briefly introduce the internal organization of Wikipedia:

**Pages:** To manage the work on the encyclopedia while still presenting a usable encyclopedia interface to non-participants, Wikipedia organizes its editable pages into several categories. The encyclopedia content itself is kept on article pages, while discussion of ongoing changes and other disputes to a specific article are kept on an associated discussion, or “talk” page. Similarly, individual contributors have a unique user page, and an associated user “talk” page to enable discussion with or about a user. There is also a revision comment field, which is a free-form text field supplied by each user when submitting an edit, intended to describe what kind of changes were made.

**Policies:** Most activities on Wikipedia are governed by various policies. Users are encouraged to follow a general policy to Be Bold [36], i.e., take action themselves when there is an apparent reason. Another policy is to refrain from altering content so that it presents a non-neutral point of view [41]. A further policy asserts that material should be attributed to a reliable published source [35]. Together, these and other policies are also subject to the same editing and review process that governs the encyclopedia content, but play an intrinsic character in determining the social norms that the communities is expected to adopt.

**Users:** Users of Wikipedia fall into distinct categories. Since it is not required for a user to register to edit most pages, many users are anonymous. Users may establish a user account, which allows them to maintain a consistent identity across different sessions and network addresses. Additional permissions can be granted to such users. The most significant of which is the role of Administrator, who has the ability to change many settings, such as locking a page for editing, banning abusive users, and the ability to access content which has been otherwise deleted from the encyclopedia.

3.3 COORDINATION AND MAINTENANCE COST IN WIKIPEDIA

Users interacting with each other create conflicts and incur coordination costs. These interactions take place on user and discussion pages, which constitute a category of work we label “overhead”. Overhead work represents the coordination costs of working in such collaboration spaces. One form of overhead consists of edits made to non-article pages. These overhead activities can be roughly divided into three categories: (1) maintenance work such as formatting and fighting vandalism, (2) developing administrative procedures and policies for Wikipedia, and (3) communication cost between users such as edits on article talk, user talk pages.

Our recent study [18] found that work going into editing article pages has been steadily decreasing from 95% to 65% (see Figure 2). Meanwhile, work going into communication purposes has been increasing to about 15–20% of total edits. Vandalism has also been increasing but still accounts for less than 1% of edits.

Figure 2. Changing number of edits over time made on various types of pages in Wikipedia

Recent research shows that Wikipedians spend a significant amount of their effort on coordination and conflict resolution [18]. This strongly indicates that there exist underlying social processes that govern coordination. In this paper, we focus on exploring the social dynamics, which is crucial to understanding the coordination and conflict resolution process in Wikipedia.

4 CONFLICT USER MODEL

To understand disagreement between users, we need to build a model of how users engage in disputes. One way to discover disagreements is to look in the revision log and analyze the substance of each user’s edits. However, we found this approach rather impractical for large data sets such as Wikipedia, because it requires accurate identification of the intention of each edit, which is not viable for machine-only processing. Many edits may have non-obvious intentions based on the context in which it was made.

Instead, we use users’ editing history to approximate disagreements between users. Specifically, we use revert revisions as a way of identifying disagreement and conflicts between users. The user survey from the previous study enabled us to collect insight that Wikipedians often treat being reverted as an indication that another user strongly disagrees with an edit [18]. We employ revert as a proxy for dispute and develop a user model based on revert activity, described in the next section.

4.1 REVERTS IN WIKIPEDIA

A revert is defined as undoing the actions of another editor in whole or in part [44]. Reverts are often used to fight vandalism to bring articles back to their original state. However, users also use reverts to block other users’ contributions. Edit wars [40] are a typical example, where disagreeing users repeatedly revert each others’ edits. Revert is a very reliable sign that one user clearly opposes other’s edit, and, therefore, is a reasonable proxy for underlying conflict and disagreement.

We identify reverts in Wikipedia by two different methods. In the data driven method, we computed a unique identifier of every revision made to every article using the MD5 hashing scheme [25], which is commonly used to check that data objects are identical. We use the hashing function to generate a small fingerprint of each revision, which is suitable for rapidly comparing all revisions of an article. Using MD5 values for all revisions of an article, we can identify when a later revision
exactly matched the hash of a previous article, indicating a revert. The advantage of this method is that it does not depend on users to label reverts, which is not always done consistently.

However, the disadvantage of this method is that it does not pick up partial reverts, in which only some of the text in an article is reverted. To capture partial reverts we used a user-labeled metric, counting revisions whose revision comments included the text “revert” or “rv” (a commonly used abbreviation of revert). The combination of both the data-driven and user-labeled methods provides converging evidence on the true change in reverts over time.

Table 1 shows that the statistics for reverts calculated by the two methods have slightly different characteristics. The MD5 identity revert discovery technique captures more revisions than user-labeled (comment) reverts (3.7M vs. 2.4M), suggesting that a substantial number of reverts are not labeled by users as such. The union of the two methods provides the most accurate view of reverts, resulting in 3,917,008 reverts marked by either comments or MD5 hashes. In other words, approximately 6.7% of all changes in Wikipedia goes to restoring articles to previous versions.

Vandalism in Table 1 is calculated by a similar method to the comment method for reverts. We looked through the revision comments of each article for any form of the word “vandal” or “comment method for reverts. We looked through the revision versions. Changes in Wikipedia goes to restoring articles to previous versions.

5 REVERT GRAPH – VISUALIZING USER CONFLICT

Revert Graph is a tool that enables an analyst to quickly understand the relationships and patterns of activity that embody the conflict between users. Our layout algorithm simulates the social dynamics that result from the user conflict model. We accomplish this by visually gathering users with similar or compatible opinions together, while separating disagreeing users. We implemented this approach using a force-directed graph layout algorithm [16] that assigns forces such that the edges (representing revert relationships) act as springs, while the individual users are represented as particles with gravitational fields (as shown in Figure 3).

Table 1. User, Revision, Revert and Vandalism Statistics

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users Total</td>
<td>3,769,347</td>
</tr>
<tr>
<td>Users who made at least one revert</td>
<td>402,454</td>
</tr>
<tr>
<td>Revisions Total</td>
<td>58,545,791</td>
</tr>
<tr>
<td>Reverts (MD5 hash method)</td>
<td>3,711,638</td>
</tr>
<tr>
<td>Self-reverts</td>
<td>582,373</td>
</tr>
<tr>
<td>Pages with at least one revert</td>
<td>721,866</td>
</tr>
<tr>
<td>Pages with 50 reverts or more</td>
<td>9,973</td>
</tr>
<tr>
<td>Reverts (Comment method)</td>
<td>2,422,482</td>
</tr>
<tr>
<td>Vandalism (Comment with vandal, rvv, etc)</td>
<td>577,643</td>
</tr>
<tr>
<td>Reverts (Union of both methods)</td>
<td>3,917,008</td>
</tr>
</tbody>
</table>

Figure 3. Force-directed layout structure employed in Revert Graph. Users (represented as nodes) attract each other unless they have a revert relationship. A revert relationship is represented as an edge, thus pushing such users apart. Left figure: Nodes are evenly distributed as an initial layout. Right figure: When forces are deployed, nodes are rearranged in two user groups.

Initially, the tool loads a group of users participating in editing an article as a uniformly distributed node-link graph. As the simulation runs, forces in the graph stabilize, and social structures between users begin to emerge, as shown in Figure 4.

Node size is proportional to the log of the number of reverts or revisions. Nodes are color-coded based on users’ registration status. An administrator is drawn as a green node, a normal registered user as a grey node, and an unregistered anonymous user as a white node, respectively.