

Network Analysis of Collaboration Structure in Wikipedia

Ulrik Brandes
University of Konstanz, Germany
Ulrik.Brandes@uni-konstanz.de

Jürgen Lerner
University of Konstanz, Germany
lerner@inf.uni-konstanz.de

Patrick Kenis
TiasNimbas Business School & Tilburg
University, The Netherlands
p.kenis@tiasnimbas.edu

Denise van Raaij
Tilburg University, The Netherlands
D.P.A.M.Korssen-vanRaaij@uvt.nl

ABSTRACT

In this paper we give models and algorithms to describe and analyze the collaboration among authors of Wikipedia from a network analytical perspective. The edit network encodes who interacts how with whom when editing an article; it significantly extends previous network models that code author communities in Wikipedia. Several characteristics summarizing some aspects of the organization process and allowing the analyst to identify certain types of authors can be obtained from the edit network. Moreover, we propose several indicators characterizing the global network structure and methods to visualize edit networks. It is shown that the structural network indicators are correlated with quality labels of the associated Wikipedia articles.

Categories and Subject Descriptors

E.1 [Data Structures]: Graphs and networks; H.3.5 [Online Information Services]: Web-based services

General Terms

Algorithms, Human Factors, Measurement

Keywords

Wikipedia, social network analysis, network visualization, edit network, governance

1. INTRODUCTION

The performance of teamwork is not only dependent on the knowledge and capabilities of the individual members, but also on how they collaborate. In this paper we are interested in community forms of organizations [19]—as opposed to, say, hierarchical organizations—that rise at the Internet and that we refer to as *webbased information communities* (*WebICs*). WebICs are defined as work systems facilitated by the Internet infrastructure and composed of voluntary actors that attempt to produce a product or service such as software or encyclopedic information. (Note that we are using the term *community* in its sociological sense, denoting a group of people who share an interest, interact repeatedly, generate shared resources, develop governance policies, demonstrate reciprocity, or share cultural norms [20].) We

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

propose methods to analyze collaboration networks encoding the edit interaction among users who contribute to a Wikipedia page or a set of pages. These networks give insight into the roles of individual authors as well as into the global collaboration structure.

The *edit network* associated with a Wikipedia page has as actors the authors of the page and encodes how authors edit the page and how they respond to the edits of others. Characteristics of individual authors reveal, for instance, the most active authors, the authors that have written most of the current page content, as well as the *role* of authors (e.g., discriminate between providers of content and deleters). Information about pairs of authors can, for instance, be used to determine those authors that erase each others' edits most actively or authors who defend each others' content against deletion. Parameters characterizing the global collaboration structure include an indicator that estimates whether the author community decomposes into two poles of opinion (measure of *bipolarity*).

Contributions. The long-term goal of this project is to get insight into the social process of content-creation in Wikipedia or in similar open collaborative environments. Concrete contributions of the current paper include the following. We define a data structure, called *edit network*, that can be derived from the edit history of a Wikipedia page and that encodes the work of individual authors as well as the edit interaction between authors. We propose methods to analyze and visualize these networks, leading to characterizing quantitative parameters and illustrative images of Wikipedia communities. Finally, we establish correlation between the structural parameters of edit networks and quality labels of the associated page; for instance, pages labeled as controversial turn out to be more bipolar than so-called featured articles.

Organization of Paper. The remainder of the introduction provides arguments why network analysis may provide additional insight into Wikipedia communities and reviews related work. Section 2 defines the edit network and illustrates how various informative characteristics about individual authors and pairs of authors can be obtained from it. In Sect. 3 we propose several network indicators and a visualization technique highlighting the global collaboration structure. Section 4 shows that the structural parameters are related to distinguishing quality labels of the associated pages.

1.1 Network Structure Matters

In this section we briefly argue why we look at structural (network) features of Wikipedia communities, rather than just at their attributes. Section 1.1.1 shortly reviews general articles from the organizational science literature and Sect. 1.1.2 illustrates by means of examples that different Wikipedia pages can yield edit networks of strikingly different structure.

1.1.1 Background from Organizational Science

As a co-evolution of the network society [8], work systems are more often organized in an informal way [9], as opposed to, say, hierarchical organizations. Flows and linkages between actors characterize these network forms of organizations [16], and it is exactly these linkages that constitute the bulk of organizational activity [17]. Hence, if we want to better understand network forms of organizations, we must consider their structural features. Former research on inter organizational networks has demonstrated that a network's structure relates to outcome quality (e.g., [22, 23]), and that a network analytical perspective has the potential to complement insights derived from attribute data (e.g., knowledge).

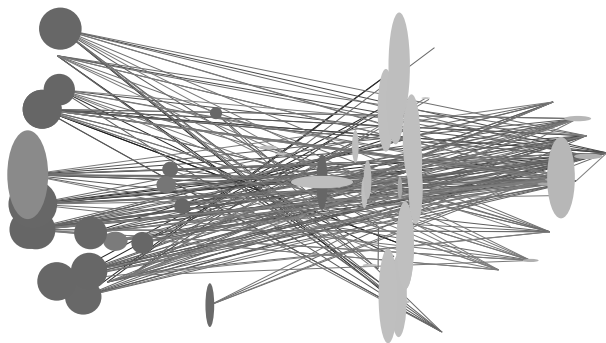


Figure 1: The edit network associated with the page on Telephone tapping. Nodes correspond to authors of the page and edges encode negative interaction describing undoing each others' edits. Size of nodes is proportional to author activity, dark authors are deleters, bright authors mostly add and restore text, high nodes mostly revise others, and broad nodes mostly get revised. These and other characteristics are defined in Sect. 2. This network has a very high bipolarity, thus decomposes into two groups that mutually undo the edits of the other group (see Sect. 3).

1.1.2 Examples of Edit Networks

We illustrate here by means of examples that the structure of edit networks can be very descriptive of how work gets done in different Wikipedia author communities. Most edges of the edit network encode *negative* relationships describing who erases the edits of whom. We emphasize that it is indeed important to know about the structure of these interactions and not just their frequency or magnitude. To illustrate this claim, we compare two edit networks computed from the page on **Telephone tapping**, which is rated by Wikipedia editors as a controversial page, and from the page on **Tropical Storm Alberto** (2006), which is rated as

a featured article. The two networks are visualized by the algorithm described in Sect. 3.2. Nodes represent authors and edges represent negative interaction where one of the connected authors undoes the edits of the other.

Figure 1 shows that the network associated with **Telephone tapping** can be partitioned into two groups such that authors are much more likely to undo edits from members of the other group than from members of their own group (an indicator assessing this property is called *bipolarity* and is defined in Sect. 3.1). This makes it very plausible that authors in the two groups mutually disagree with the opinion of the other group and therefore erase their edits. In addition, the two groups have very different characteristics. The group on the left-hand side consists mostly of users that delete text. In contrast, the authors on the right-hand side mostly add new text or restore previously deleted text (another difference is illustrated in Fig. 3).

In contrast, the network associated with **Tropical Storm Alberto** (2006), visualized by the same algorithm and shown in Fig. 2, does not decompose into poles of opinion. Since the page is not static and does not grow indefinitely, many of the edits delete text written by others; however, these deletions do not follow such a strict pattern as in Fig. 1. Seen from a different angle, the undo-operations in Fig. 2 do not seem to be targeted against a certain group of authors, as it seems to be the case in Fig. 1.

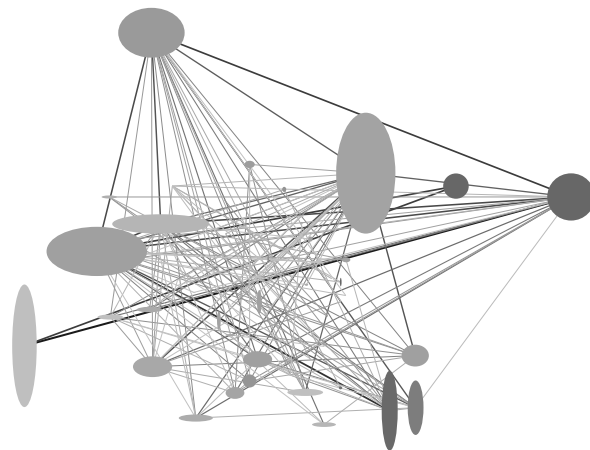


Figure 2: The edit network associated with the page Tropical Storm Alberto (2006) has a very low bipolarity. The edit interaction does not reveal distinct groups having contradicting opinions.

1.2 Related Work

1.2.1 Research on Wikipedia

The popularity of Wikipedia together with the free availability of its complete database has made Wikipedia a popular topic in academic research. Here we can only refer to a few of the papers that are most related to our work.

Early work on visualizing the history of Wikipedia articles includes the *history flow* visualization [26] that shows how sentences, labeled by the authors who have written them, persist over time, or get deleted at later revisions. Applying this approach at a later time [27] showed that the history

flow visualization reaches its limitations due to the enormous number of edits on some pages. Moreover, it does not show who deletes text written by whom.

The revision history of Wikipedia articles has been used to answer various research questions. Kittur et al. [14] found that the percentage of edits made by elite users (i.e., users with the highest number of edits) is declining over time but also that elite users change on average more words than common users and are more likely to add rather than delete content. To develop a reputation system for Wikipedia, Adler and de Alfaro [1] considered not only how many words authors add to a page, but also how long these words persist on the page. To estimate the impact of vandalism (among others), Priedhorsky et al. [21] went one step further by taking into account the number of times that a particular edit has been *viewed*. Vuong et al. [28] estimated the controversy-level of articles and authors by considering the network connecting authors to the articles they edit (in contrast to our paper which codes the author-author network associated with one or several pages).

Other research that more explicitly considered the author-author network includes the following. Kittur, Suh et al. [15, 24] constructed a network of authors, where an edge (u, v) indicates that u reverted a version submitted by v . These edges are interpreted as negative relationships between users and the networks are drawn with a force-directed layout algorithm that let heavily connected users repulse each other. A similar network has been built by coding repeated immediate revisions (not just reverts) between users [6]. In contrast to the present work, these three papers do not consider who deletes how much of whose edits or who restores whose edits deleted by whom. However, as will be demonstrated in our paper, it is exactly this information that enables us to characterize individual authors and groups of authors.

1.2.2 Network Analysis and Visualization

The bipolarity indicator (compare Sect. 3.1.1) has originally been defined to assess whether political conflict networks decompose into two opposing groups and to visualize conflict networks that have high bipolarity [4]. Later, this layout algorithm has been generalized to deal with other than just bipolar conflict structures [6, 7]. The visualization algorithm given in our paper (see Sect. 3.2) is essentially the one from the two last-mentioned papers. However, coping with the highly skewed distribution of edge weights, the pre-processing step from Sect. 3.1.2 yields significantly improved layouts.

Network analysis has been applied [25, 11] to identify *types* or *roles* of users in question-and-answer environments such as Usenet groups. Likewise, the attributes defined in Sect. 2.2 reveal user roles in environments such as Wikipedia, where users are co-authoring documents.

1.2.3 Algorithms for Text-Comparison

For constructing the edit network we repeatedly have to determine changes between two versions of a page. Detecting common sub-parts and differences between texts is a well-studied problem, see e.g., [13, 12, 18]. Here we make use of algorithms given in [13] and [18] but adapt them in a way that moving large parts of the text to a different place is not treated as massive change. Furthermore, so-called *reverts* are handled in a different manner, see Sect. 2.4.

2. THE EDIT NETWORK

In a nutshell, the edit network associated with a Wikipedia page p has as nodes the authors of p and encodes how authors contributed to p and how authors interacted with each other while editing p . This information is computed from the complete history of p , i.e., from the sequence of its revisions, by determining which part of the text has been added, has been deleted, or remained unchanged when going from one version of the page to the next. Since a description of the full details of *how* texts are transcribed is rather tedious, we first describe precisely *what* information is encoded in the edit network. We clarify in Sect. 2.4 how we handle reverts, duplicated text, and text that has been moved to a different location on the page. For space limitations we do not present an exhaustive description of the text processing algorithm.

2.1 Network Model

The *edit network* associated with a Wikipedia page p is a tuple $G = (V, E, \mathcal{A})$, whose components are defined as follows.

1. The *nodes* V of the graph (V, E) correspond to the authors that have done at least one revision on p .
2. The *directed edges* $E \subseteq V \times V$ of the graph (V, E) encode the edit interaction among authors. A particular pair of authors $(u, v) \in V \times V$ is in E , if u performed one of the following three actions with respect to v .
 - (a) u *deletes* text that has been written by v ;
 - (b) u *undeletes* text that has been deleted by v (and written by a potentially different author w);
 - (c) u *restores* text that has been written by v (and deleted by a potentially different author w).

Since authors may as well revise text written by themselves, *loops*, i.e., edges connecting an author with herself, are allowed.

3. $\mathcal{A}$ is a set of *weighted attributes* on nodes and edges, explained in Sect. 2.2.

2.2 Attributes on Nodes and Edges

The weighted attributes on nodes and edges encode how much text users add, delete, or restore. Furthermore, in case of deletion we keep track of who has previously written the text and in case of restoration we keep track of both, the original author and the deleter of the restored text. By combining these attributes, we can get deep insight into the various roles that users play when editing, as well as into relations between users, see Sect. 2.2.2. The *amount of text* added, deleted, or restored is measured by the number of words.

We will also keep track of the timepoint when edit actions occur by indexing attributes with the revision number. In the following we assume that the history of a given page is a sequence of revisions $R = (r_1, \dots, r_N)$, ordered by increasing timestamps $1, \dots, N$.

2.2.1 Basic Attributes

Basic attributes are those that have to be computed by the network construction algorithm; others can be derived from these, see Sect. 2.2.2.

Attributes on edges. For each timepoint $i \in \{1, \dots, N\}$ and each pair of authors $(u, v) \in V \times V$,

- **delete_i(u, v)** denotes the number of words deleted by u in revision r_i and written by v at earlier timepoints j ($j < i$);
- **undelete_i(u, v)** denotes the number of words restored by u in revision r_i , deleted by v at timepoints j ($j < i$), and written by a potentially different author w at timepoints ℓ ($\ell < j < i$);
- **restore_i(u, v)** denotes the number of words restored by u in revision r_i , written by v at timepoints j ($j < i$), and deleted by a potentially different author w at timepoints ℓ ($j < \ell < i$).

Note that **delete_i(u, v)**, **undelete_i(u, v)**, and **restore_i(u, v)** are equal to zero, if u is not the author of revision r_i .

Attributes on nodes. For each timepoint $i \in \{1, \dots, N\}$ and each author $u \in V$,

- **add_i(u)** denotes the number of words that are added by u at time i ;
- **authorship_i(u)** denotes the number of words in revision r_i that have been authored by u , i.e., all words that have been added to the text by u in a revision $j \leq i$ and that are still there in r_i .

If u is not the author of r_i , then **add_i(u)** equals zero. Note however, that even in this case **authorship_i(u)** might be bigger than zero. It always holds that **add_i(u) ≤ authorship_i(u)** since, at time i , u is the author of at least those words that she added in r_i and it holds that **authorship_i(u) ≤ ∑_{j=1}ⁱ add_j(u)** since, at time i , u can only be the author of those words that she added before or at the i 'th revision.

2.2.2 Derived Attributes

Starting from the basic attributes we can define a number of other attributes that characterize pairs of authors (dyadic variables, defined on edges) or single authors (monadic variables, defined on nodes). In this section we define these attributes and describe how they are interpreted.

Attributes on edges. Summing values over all timepoints yields three weight functions for edges $(u, v) \in E$, that are given by

$$\text{delete}(u, v) = \sum_{i=1}^N \text{delete}_i(u, v),$$

and **undelete(u, v)** and **restore(u, v)** by similar formulas.

Large values of the weights **delete(u, v)** and **undelete(u, v)** imply a *negative* relationship from u to v . Indeed, if u deletes a lot of text written by v , then u apparently disagrees with v 's contributions to the article. Similarly, if u undeletes a lot of text that has been previously deleted by v , then u disagrees with v removing this text from the article. On the other hand, large values of the weight **restore(u, v)** imply a *positive* relationship from u to v , since u defends v 's contributions against deletion.

The sum over the two negative relations, denoted by

$$\text{revise}(u, v) = \text{delete}(u, v) + \text{undelete}(u, v),$$

encodes how much u undoes v 's edits. It is interpreted as a measure of how strongly u *disagrees* with v . Similarly, **restore(u, v)** is interpreted as a measure of how strongly u *agrees* with v .

Attributes on nodes. Summing the **delete** and **restore** weights of all edges starting from one author u give measures of how many words u deletes, respectively restores. Note that it is unnecessary to count the number of undeleted words, since this count is equal to the number of restored words.

$$\begin{aligned} \text{delete}_i(u) &= \sum_{v \in V} \text{delete}_i(u, v); \\ \text{restore}_i(u) &= \sum_{v \in V} \text{restore}_i(u, v) = \sum_{v \in V} \text{undelete}_i(u, v) \end{aligned}$$

Summing the attributes **add**, **delete**, and **restore** over all timepoints $i = 1, \dots, N$ defines attributes that summarize the editing-work of the respective author, given by

$$\text{add}(u) = \sum_{i=1}^N \text{add}_i(u)$$

and **delete(u)** and **restore(u)** by similar formulas. The attributes **add(u)**, **delete(u)**, and **restore(u)** characterize u 's *role* as being a provider of new content, someone who removes content, or someone who defends content from being removed.

At the last timepoint N it holds that **authorship_N(u) ≤ add(u)**, since u can only be author of those words that she added. Normally **authorship_N(u)** will be smaller than **add(u)**, since many words written by u might be deleted afterwards. An example highlighting the value of **authorship_N(u)** (i.e., the amount of text authored in the current revision) is shown in Fig. 3.

A measure of *involvement* of authors in the Wikipedia article they edit is given by the sum

$$\text{activity}(u) = \text{add}(u) + \text{delete}(u) + \text{restore}(u),$$

called the *edit activity* of author u . It is the number of words that u touched by adding, deleting, or restoring them. In the network images (e.g., Fig. 1) we visualize the activity of authors by the area size of the node, so that the most active editors correspond to the largest nodes.

A further characterization of *how* an author u contributes to a page is given by the difference

$$\text{netadded}(u) = \text{add}(u) + \text{restore}(u) - \text{delete}(u),$$

called the *net-amount of added words*. It is the number of words by which u increased the length of the text. If **netadded(u)** is positive, then u 's intention was apparently either to increase the text by adding new words or prevent it from being shortened by restoring deleted text. If **netadded(u)** is negative, then u 's intention was apparently to decrease the length of the text by deleting parts of it. The absolute value of **netadded** is always bounded by **activity**. Thus, for any author with non-zero **activity**, the ratio

$$\text{netaddedratio}(u) = \text{netadded}(u) / \text{activity}(u)$$

lies between minus one and plus one. If **netaddedratio(u) = -1**, then u dedicates all her activity to deletion of text and if **netaddedratio(u) = 1**, then u dedicates all her activity to

either adding or restoring text. In the network images (e. g., Fig. 1) the color indicates the **netaddedratio** of authors. If an author mostly deletes text, then she is shown as a dark node if an author mostly adds or restores text, then she is shown bright and if **netaddedratio** is close to zero, the node's brightness is in between.

We call the value

$$\text{revisor}(u) = \sum_{v \in V} \text{revise}(u, v)$$

(i. e., u 's weighted out-degree with respect to the **revise** relation) u 's *degree as a revisor*. It is the number of words that u deletes after they have been added, or restores after they have been deleted, i. e., it is a measure that characterizes the *undo-activity* of u . Similarly,

$$\text{revised}(u) = \sum_{v \in V} \text{revise}(v, u)$$

(i. e., u 's weighted in-degree with respect to the **revise** relation) is called u 's *degree as being revised*. It is a measure of how much u 's edits are undone later. It holds that both, **revisor**(u) and **revised**(u), are bounded from above by **activity**(u). Actors with **revisor** $\approx$ **activity** show a *reactive* behavior, since they dedicate most of their activity to undo changes made by others. In contrast, actors with **revised** $\approx$ **activity** are those that do not succeed in making their edits permanent, as these are mostly undone afterwards. In the network images (e. g., Fig. 1) the ratio **revisor** divided by **revised** is proportional to the height/width ratio of the ellipses, so that authors that mostly revise others and are not much revised themselves are high and narrow. The computation of the coordinates of nodes (i. e., the layout of the network) is detailed in Sect. 3.2.

2.3 Edit Networks of Several Pages

A straightforward way to compute the edit network of several pages—possible having related topics—is to process the individual pages separately, take the union of the author sets, and add up the attribute functions. However, in some cases a different strategy is preferable.

As a matter of fact, some Wikipedia pages grow very large and some of their content, describing certain subtopics, is turned into a page of its own. When processing pages separately we would treat this as a massive deletion of text in the former page and addition of newly created text in the latter page which obviously ignores the fact that the text has just been moved. A better way to handle a set of pages that emerged out of a common page is to consider the concatenation of text, i. e., do the analysis as if the pages had never been split. By the conventions we make in Sect. 2.4 the order of the text does not matter and, therefore, splitting a page into two would not be treated as a massive change.

In general, sets of related pages could be identified by looking for pages strongly referencing each other, pages that have a strong author-overlap (see [5]), or pages that are members of the same categories. The edit network associated with such a set of related pages would then represent the collaboration structure of a *topic* in Wikipedia.

2.4 Text-Processing Conventions

In this section we describe the conventions that we adopt when processing the text, especially how to treat duplicated

text, text that is cut and pasted to a different location, and edits that are reverts.

Table 1: Example of four revisions on a page. Greek letters stand for words.

author	text	authorship of words
Alice	$\alpha \beta.$	$A(\alpha \beta)$
Bob	$\alpha \beta. \alpha \gamma \delta.$	$A(\alpha \beta), B(\alpha \gamma \delta)$
Charlie	$\alpha \delta. \gamma \beta.$	$B(\alpha \delta), C(\gamma), A(\beta)$
Alice	$\alpha \beta. \alpha \gamma \delta.$	$A(\alpha \beta), B(\alpha \gamma \delta)$

We consider the granularity of authorship on the word level, i. e., each word has exactly one author and different words may have different authors.

The main point to clarify is whether the ordering of words is taken into account and how to treat duplicated text. Considering the ordering of words has an obvious disadvantage in the context of Wikipedia pages: if an author restructures the page by cutting and pasting large parts of the text to different places, then we would count this as massive deletion of text and addition of newly created text. On the other hand, if we modeled the whole text as an unordered set of words, it would be impossible to determine authorship of duplicated words. (Accounting all instances of a word to the first author who has written it would also be a misinterpretation.) To overcome this dilemma, we propose to make use of the fact that words are assembled to sentences and we make here the assumption that one sentence represents one statement, gives one fact, or makes one claim. More precisely, we will model the whole text as an unordered *set* of sentences, which in turn are modeled as ordered lists of words. In particular, moving a complete sentence to another position, duplicating a complete sentence, or deleting a duplicated sentence is not considered as a change. Note however, that two words within the same sentence might have different authors. For instance, if an author changes a sentence *partially* by adding some words to it, she becomes only the author of the newly added words.

Punctuation and capitalization is used only to determine the boundaries of sentences. Once the text has been split into sentences, punctuation and capitalization is ignored.

The last point we have to consider is that many edits on Wikipedia pages are so-called *reverts*, i. e., edits that set back the page to an earlier version. For instance, if a user u deletes the whole content of a page in revision i and another user v restores it in revision $i+1$ to the version $i-1$, it would not be reasonable to credit v as the author of the whole text (actually, she has not written it and might be completely ignorant of the topic). Rather, we set back the authorship of all words as it has been assigned in revision $i-1$ and just keep track of the fact that v undeletes a lot of text deleted by u .

Table 1 gives an example of four revisions and the resulting authorship of words as determined by the conventions listed above. In this example Greek letters stand for words and periods delimit sentences. Note that the third revision is interpreted in the way that **Charlie** interchanged the first and second sentence, deleted γ in sentence $\alpha \gamma \delta$, and changed the word α in sentence $\alpha \beta$ to γ . The interchange of the two sentences is established by the fact that sentence $\alpha \gamma \delta$ and sentence $\alpha \delta$ have a common subsequence of length two and are, therefore, the most similar pair of sentences. After the third revision, **Charlie** is the author of γ , **Alice** is the au-

thor of β , and Bob is the author of α and δ . Furthermore, it is accounted that Charlie deleted one word of Alice (namely the α from $\alpha\beta$), deleted one word of Bob (namely the γ from $\alpha\gamma\delta$), and added the γ in sentence $\gamma\beta$. The forth revision is a revert that makes Charlie's changes undone. Therefore, Alice deleted Charlie's γ , restored her own α , and restored Bob's γ , setting $\text{undelete}(\text{Alice}, \text{Charlie}) = 2$.

3. ANALYSIS AND VISUALIZATION

Naturally, the edit network can be analyzed with any method from social network analysis (see e.g., [3] and [29] for an overview). In this section we focus on network analysis and visualization methods that are especially useful for networks having many *negative* edges and whose basics have been developed in [4].

Negative relationships between two authors are expressed by the **delete** and **undelete** relations, since these encode disagreements or disputes. In Sect. 3.1 we present an indicator (called *bipolarity*) that estimates to what extend the set of authors can be partitioned into two subsets such that disagreements are more frequent between members of different clusters than between members of the same cluster. If this is the case, the clusters are likely to represent groups of authors that have contradicting opinions. (Note that this approach to detect communities based on negative relations is fundamentally different from the more usual computation of densely-connected groups in networks.) A visualization algorithm that highlights controversy structure in the Wikipedia collaboration network is given in Sect. 3.2 while Sect. 3.3 proposes several indicators measuring whether there is balance between the contributions of opinion groups—if there are opinion groups in the network.

3.1 Bipolarity

A straightforward formalization of bipolarity would be based on the MAXCUT-problem which is sketched in the following. A *cut* is a partition of the actorset V of a weighted graph $G = (V, E, \omega: E \rightarrow \mathbb{R})$ into two disjoint subsets $V = V_1 \cup V_2$. In our case, the edge weights encoding disagreement are given by $\omega(u, v) = \text{revise}(u, v)$. The *weight* of a cut (V_1, V_2) is simply the sum of the weights of edges crossing the cut, i.e.,

$$\omega(V_1, V_2) = \sum_{u \in V_1, v \in V_2} \omega(u, v) + \omega(v, u) . \quad (1)$$

If (V_1, V_2) is a cut with maximal weight for a given graph G , then the value $\omega(V_1, V_2)$ is denoted by $\text{MAXCUT}(G)$.

Clearly, $\text{MAXCUT}(G)$ is bounded by the sum over all edge weights $W = \sum_{(u,v) \in E} \omega(u, v)$. Thus, the ratio $\text{MAXCUT}(G)/W$ could be taken as a measure for the bipolarity of G . If $\text{MAXCUT}(G)/W = 1$ (or very close to 1), then the partition is perfect in the sense that (almost) all disagreement edges are between actors from different clusters—thereby revealing a very clear opposition of the two groups.

However, the MAXCUT problem has some properties that make this formulation inappropriate for our task. First of all, computing $\text{MAXCUT}(G)$ is computationally intractable and partitions that yield a cut of maximal weight are highly sensitive to small changes in the edge-weights (e.g., caused by noise in the input data). Furthermore, discrete partitions put actors completely into one group or the other and do therefore not distinguish between authors that are central

members of one of the clusters and authors that are rather between the two clusters.

To overcome these problems, a method has been developed in [4] that does not put authors completely into one group or the other, but that determines real-valued degrees of membership. We present the essentials of this method in Sect. 3.1.1.

3.1.1 Continuous Projection into Controversy Space

Given a weighted graph $G = (V, E, \omega: E \rightarrow \mathbb{R})$ with $V = \{1, \dots, n\}$ and a positive integer $k \leq n$, a *projection* (of dimension k) is a real $k \times n$ matrix P that has orthogonal rows of unit length, i.e.,

$$\forall i, j \in \{1, \dots, k\} \text{ it is } \sum_{u \in V} P_{iu} \cdot P_{ju} = \delta_{ij} ,$$

where $\delta_{ij} = 1$, if $i = j$ and zero otherwise. A projection is interpreted as a continuous version of a discrete partition into k classes. The *degree of membership* of node u in class i is the real number P_{iu} . The normalization constraint is necessary to avoid arbitrary scaling of membership values. The name *projection* is motivated by the fact that these matrices correspond to orthogonal projections from $\mathbb{R}^n$ to $\mathbb{R}^k$.

In the case of a discrete partition on a weighted graph, (1) defines a value describing how much two classes V_1 and V_2 are connected. This can be generalized to a continuous projection P , as explained in the following. For $i, j \in \{1, \dots, k\}$ the weight connecting class i to class j is defined by

$$\omega(i, j) = \sum_{u, v \in V} (\omega(u, v) + \omega(v, u)) \cdot P_{iu} \cdot P_{jv} .$$

Edge weights are here multiplied with the membership values of actors to classes.

Translating the formulation of the MAXCUT problem to continuous projections gives the following problem: given a weighted graph, determine a two-dimensional projection, such that $\omega(1, 2)$, i.e., the weight of edges crossing the cut, is maximized. In contrast to the discrete MAXCUT problem, the problem of determining a projection that maximizes $\omega(1, 2)$ is efficiently solvable since it is given by the orthogonal projection to the space spanned by the two eigenvectors associated with the largest and the smallest eigenvalues [4]. In addition, the optimal projection is, under reasonable assumptions which are typically satisfied in real world data, robust to noise in the input data [4].

Let $w = \omega(1, 2)$ be the maximal weight of edges between the two groups and

$$c = \sum_{u, v \in V} \omega(u, v) \cdot P_{1u} \cdot P_{1v} + \sum_{u, v \in V} \omega(u, v) \cdot P_{2u} \cdot P_{2v}$$

be the weight of edges connection authors in the same group (using the same projection P that maximizes w). We define the *bipolarity* of a network G to be

$$\text{bipolarity}(G) = \frac{w - c}{w + c} .$$

The bipolarity lies between minus one and plus one. It equals plus one if the graph is bipartite, i.e., edges connect only members from different groups and, therefore, the division into opposing groups is perfect. For instance, the bipolarity of the network on **Telephone tapping** shown in Fig. 1 has a very high bipolarity of 0.85.

The bipolarity equals zero if the graph is complete (i. e., all pairs of actors are connected) and all edges, including loops, have the same weight. In this case, the partition into two groups is completely arbitrary and does not represent opposing groups. (Actually, a random projection on a random graph would have expected bipolarity zero.) For instance, the bipolarity of the edit network on *Tropical Storm Alberto* (2006) shown in Fig. 2 has a low bipolarity of 0.26.

It can be shown that the bipolarity can be computed by

$$\text{bipolarity}(G) = -\frac{\lambda_{\min}}{\lambda_{\max}},$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of G , respectively [4]. This implies an efficient computation by iterative numerical algorithms.

3.1.2 Logarithmic scaling of the edge-weights

Computing the bipolarity as it has been outlined above is, without further pre-processing, only appropriate for graphs whose edge weights do not differ dramatically. Since most edit networks have a distribution of edge weights which is extremely skewed (more precisely, most edges have very small weights and few edges have very large weights), we need to normalize the edge weights before computing the optimal projection and bipolarity. In fact, without prior normalization the bipolarity indicator would effectively estimate only the bipolarity of the small subnetwork that is induced by the few edges with the highest weights. All authors that are not incident to the strongest edges (i. e., the vast majority of authors) would get membership values close to zero and, therefore, the structure of the largest part of the network would be ignored.

To deal with the heavily skewed distribution of edge weights, we use edge-weights

$$\log(1 + \omega(u, v))$$

instead of $\omega(u, v) = \text{revise}(u, v)$. In the whole article, we always applied this logarithmic scaling before computing the optimal projection and bipolarity, even if not stated explicitly. The observed effect of this normalization is that the bipolarity decreases on most networks.

3.2 Visualizing Edit Networks

The degrees of membership of authors to two opposing groups, as defined in Sect. 3.1.1, could directly be taken as coordinates to draw the edit network. Authors that are strong members of the first group are then drawn far on the (say) left-hand side, authors that are strong members of the second groups are drawn far on the right-hand side, and authors that are members of both groups are drawn in the middle. This yields exactly the visualization method developed by [4]. However, the coordinates given by the membership values yield a good drawing only if the network has high bipolarity (see [7]). A slightly different method for computing the coordinates that can deal with more general (i. e., not only with bipolar) conflict structures has been proposed in [6, 7] and will be sketched below.

Graph layouts that are based on eigenvectors are frequently proposed in various application scenarios and—dependent on the chosen objective criterion—use different matrices associated with graphs, among others the adjacency matrix, Laplacian, or modularity matrix. Taking the eigenvectors of the adjacency matrix that have the lowest eigenvalues is

a well established choice if strongly connected nodes should be separated (as in our case). An outstanding paper for this approach (although it deals with graph partitioning, rather than graphdrawing) is the work of Alon and Kahale [2].

The coordinates that are determined by the following algorithm can be shown to interpolate smoothly between, e. g., bipolar and three-polar conflicts, see [6]. Let A be the real matrix whose rows and columns are indexed by the authors and whose entry in row u and column v equals $\log(1 + \omega(u, v)) + \log(1 + \omega(v, u))$.

1. Compute the two minimal (negative) eigenvalues $\lambda_{\min}$ and $\lambda'_{\min}$ of A together with (orthogonal and normalized) eigenvectors x and y .
2. The horizontal (x -) coordinates of authors are directly given by the vector x . The vertical (y -) coordinates are given by $y \cdot \frac{\lambda'_{\min}}{\lambda_{\min}}$ (i. e., y scaled with the ratio of the next-to-minimal eigenvalue divided by the minimal eigenvalue).

The so-defined coordinates of authors are centered around the origin in two-dimensional space. Since the images are more crowded in the center than on the periphery, we rescale the distance from the origin $r_u = \sqrt{x_u^2 + y_u^2}$ of author u by assigning coordinates

$$(x_u, y_u) \cdot \frac{\sqrt{r_u}}{r_u},$$

instead of (x_u, y_u) . The choice of the square-root is clearly motivated by the fact that the area of a two-dimensional circle grows quadratically with its radius.

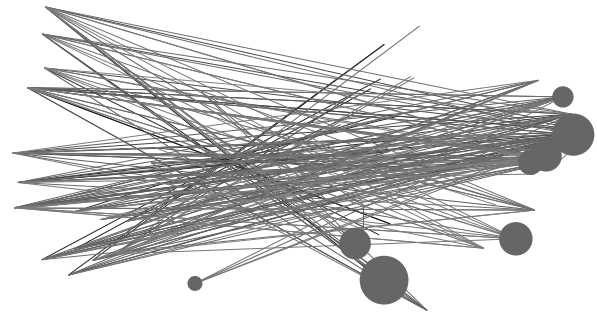


Figure 3: edit network associated with the page on Telephone tapping with actor size proportional to the authorship attribute. This network is extremely imbalanced in this respect.

An author u is drawn at the coordinates defined above as an ellipse whose area size is proportional to $\text{activity}(u)$. The ratio $\text{revisor}(u)$ divided by $\text{revised}(u)$ is proportional to the height/width-ratio of the ellipse, so that actors that frequently revise others but are not revised themselves are shown as high and narrow. If $\text{netadde dratio}(u)$ is close to one (i. e., if u mostly add or restores text), then u is drawn in bright color, if it is close to minus one (i. e., if u mostly deletes text), then u is drawn dark, and if u equally adds and deletes (netadde dratio close to zero), then u 's color is in between. The edges are drawn as grey lines with darkness and width proportional to their value of the revise attribute. Normally, we draw only the edges with highest

weights. Example-images of edit networks, visualizing the various characteristics of authors, are given in Fig. 1 and Fig. 2.

Naturally, other characteristics can be highlighted in the network images. For instance, Fig. 3 shows the same network layout as Fig. 1 but where the actors are drawn as circles whose area is proportional to the number of words they have written in the current version of the page on **Telephone tapping**. As it can be seen, the text has been written by only a few authors and, in addition, these do almost all belong to the group on the right-hand side. According to the authorship balance indicator, defined in Sect. 3.3.2, this network is very imbalanced. Actually, its authorship balance is only 0.1 indicating that one group has authored ten times more text than the other.

3.3 Further Structural Indicators

3.3.1 Partitioning Authors into Opposing Groups

The continuous assignment of authors to opposing groups can be turned into a discrete partition, as it is explained in this section. A discrete partition is necessary, for instance, if we want to determine whether one of the two opposing groups has authored more of the text than the other, or whether one group consists mostly of deleters and the other of authors that mostly add or restore text. The local optimization algorithm that we describe in the following is quite established, for instance, in *generalized blockmodeling* [10].

We determine an initial partition (V_1, V_2) of the authorset V by making use of the graph's eigenvector x associated with the smallest (most negative) eigenvalue. More precisely, we put an author u into V_1 if the u 'th coordinate of x is bigger than zero, otherwise we put u into V_2 . (Of course, there is no meaning associated with what is the first and what is the second group.) The value induced by this initial cut (V_1, V_2) is given by (1).

This initial partition is improved in a series of local optimization steps. The *neighborhood* of a partition (V_1, V_2) is defined to consist of all partitions that can be obtained from (V_1, V_2) by one of the following two operations

- move one author from its class to the other;
- swap two authors that are in different classes.

We compute the cut-value of all partitions in the neighborhood of the current partition (V_1, V_2) and, if an improvement is possible, we change (V_1, V_2) to the partition that yields the largest improvement. This step is repeated until no further improvement is possible.

Similar to the bipolarity, the computed cut yields an index encoding to what extent the two groups of authors disagree. More precisely, let $w = \omega(V_1, V_2)$ denote the cut-value for the computed partition (V_1, V_2) , as it has been defined in (1) and let

$$c = \sum_{u,v \in V_1} \omega(u, v) + \sum_{u,v \in V_2} \omega(u, v)$$

denote the aggregated weight of edges within the classes. The ratio

$$\text{CUT}(G) = \frac{w - c}{w + c}$$

operationalizes the same concept as **bipolarity**(G) and is normalized to the same interval $[-1, 1]$. It equals one if

the network is bipartite, i.e., all disagreement edges connect authors from different groups. It should be noted that the algorithm described above is only a heuristic and not guaranteed to yield the partition with the highest cut value.

3.3.2 Measures of Balance

Having computed the partition (V_1, V_2) , we can define several measures that characterize whether the network is balanced (i.e., both groups have similar characteristics with respect to the attributes defined in Sect. 2.2) or imbalanced.

Let (V_1, V_2) be the partition of authors into opposing groups and define group-authorship values by

$$a_i = \sum_{u \in V_i} \text{authorship}(u), \quad (2)$$

for $i = 1, 2$. The *authorship balance* of the edit network is defined to be $\text{autbalance}(G) = \min(a_1, a_2) / \max(a_1, a_2)$. This value equals one if both groups have authored equally many words and it equals zero if one group has written the whole text of the current page.

Figure 3 visualizes the authorship attribute for the edit network associated with the page on **Telephone tapping**. As it can be seen, almost all dominant authors belong to the group on the right-hand side. The authorship balance of this network is 0.1 indicating that one group has authored ten times more text than the other.

Similar to the authorship balance, it can be computed whether the network is balanced with respect to any other characteristic of the authors. For instance, it can be determined whether one of the two groups does all the deletion-edits and the other all the adding and restoring of text. The node colors in Fig. 1 show that the page on **Telephone tapping** is also very imbalanced in this respect.

3.3.3 Consistency of Positive Edges

In contrast to the negative disagreement edges that are given by the attribute **revise**(u, v), high values of the attribute **restore**(u, v) indicate a positive relationship from u to v . If the two opposing groups (V_1, V_2) computed in Sect. 3.3.1 really represent contradicting opinions, we expect that positive edges are mostly between members of the same group. The following indicator estimates to what extent this property holds.

Let

$$c = \sum_{u,v \in V_1} \text{restore}(u, v) + \sum_{u,v \in V_2} \text{restore}(u, v)$$

be the aggregated weight of positive edges within the groups and

$$w = \sum_{u \in V_1, v \in V_2} \text{restore}(u, v) + \text{restore}(v, u)$$

be the aggregated weight of positive edges between the groups. The ratio

$$\text{grouprestore}(G) = \frac{c - w}{w + c}$$

is normalized to the interval $[-1, 1]$. It equals 1 if all **restore**-edges are within the groups, i.e., if no author restores text from an author of the other group, indicating contradicting opinions. It equals 0 if the **restore**-edges are independent on the group membership, indicating that the two groups do not have contradicting opinions.

4. STRUCTURE AND OUTCOME

In this section we want to explore whether articles that are labeled differently typically exhibit a different collaboration structure. Note that it is not our intention here to design a method that predicts featured or controversial article status. Rather, the analysis here provides preliminary results that structural parameters of the edit network are correlated with article labels assigned by Wikipedians. This suggests that it matters how the author community collaborates.

4.1 Case Selection and Results

To test whether high values of the bipolarity indicator point to controversy in authors' opinions, we computed the bipolarity of articles linked from the page `Wikipedia:List of controversial issues`, with our hypothesis being that bipolarity is high on those controversial articles and lower on non-controversial ones. However, the selection of non-controversial articles is more involved due to the extreme skewedness of various distributions over the Wikipedia database.

As a matter of fact, the vast majority of articles is not labeled as controversial, is very short, and has very small numbers of authors and revisions. Thus, one could argue that the vast majority of articles is not labeled as controversial just for the reason that these pages have not yet received enough attention for any controversy to emerge. To compare controversial articles with non-controversial ones that did receive enough attention, we have chosen so-called *featured articles* which are listed on the page `Wikipedia:Featured articles`.

From both sets, controversial and featured, we have chosen a random sample of 60 articles and downloaded their complete history in March 2008. We computed for each of these 120 articles the associated edit network and the network parameters **bipolarity**, **grouprestore**, and **autbalance** defined in Sect. 3. These parameters are averaged over the sample of featured articles and over the sample of controversial articles; the results are shown in Table 2 (estimated standard errors in brackets).

Table 2: Estimated mean and standard errors of structural network parameters for featured (F) and controversial (C) articles; the last row gives the significance level of the mean values being different.

	bipolarity	grouprestore	autbalance
F	0.60 (0.022)	0.71 (0.045)	0.20 (0.032)
C	0.72 (0.022)	0.79 (0.026)	0.29 (0.033)
<i>p</i>	< 0.01	—	< 0.1

4.2 Discussion

The bipolarity index of controversial articles is significantly higher than the bipolarity of featured articles. Thus, the controversy of topics is indeed reflected in the edit behavior on the associated Wikipedia article. The **grouprestore** indicator is also higher for controversial articles (although not significantly), indicating that authors are more likely to restore text written by authors in their own group. This suggests that for controversial articles the two computed groups indeed represent contradicting opinions, while the opposition is less clear for featured articles.

The **autbalance** indicator is higher for controversial than for featured articles, i. e., the amounts of text written by the

two groups differ less for controversial articles. This could be explained by the reasoning that the authors of controversial articles indeed try to include both points of view. Seen from a different angle, no group succeeds entirely in erasing all edits of the other group. On the other hand, in high-quality articles most readers can accept the facts stated on the page and, therefore, fewer authors try to delete the text of the dominant authors and fewer authors feel that they have to add their own point of view for balancing opinions. It has to be validated in future research whether this reasoning can also be empirically supported.

It should be noted that the **bipolarity** and **grouprestore** index—although varying in the expected direction—are quite high for both controversial and featured articles. We hypothesize that this is due to the fact that controversial articles but also featured articles are often the target of vandalism. Vandalism, such as mass-deletions or insertion of vulgarities, is normally reverted afterwards which induces a bipartite structure between vandals and vandal-fighters. It will be addressed in future research whether removing vandals from the edit network gives even better insight into *opinion-triggered* controversies.

5. CONCLUSION

In this paper we have defined *edit networks* to encode the collaborative work of open content production communities, such as Wikipedia. An edit network summarizes the work of individual authors as well as positive and negative interaction among authors. Information about individual authors reveal the *roles* that users play during article writing; for instance, it shows which users are the providers of content and which users fulfill a control-function by watching the page and reverting edits that they do not approve. Coding the edit interaction among users enables us to spot pairs of users who erase each others' edits and pair of users who prevent each others' edits from being deleted.

A convenient approach for qualitative exploration of edit networks is enabled by adequate visualization techniques. The generated images give a good overview over the distribution of editor roles in the community, for instance, they show which are the most active authors and whether there are editors particularly focused on deletion or restoration of content. The network images also show that the global collaboration structure of some pages indeed differs dramatically (compare Fig. 1 and Fig. 2) and that some networks are very imbalanced with respect to which group provides the content (Fig. 3).

Quantitative indicators for the network structure that are proposed in this paper assess whether the author community partitions into poles of opinion and, if so, whether the contributions of these groups are balanced. We have shown that articles of different quality differ on average in their structural parameters. Naturally, analysis of the edit network is not bounded to the computation of opposing groups, but many more methods from social network analysis can be fruitfully applied.

In future work we plan to analyze how edit networks evolve over time. For instance, it would be interesting to know about the collaboration structure of an article before it gets featured (being the time when the high-quality edits are performed). Exploratory dynamic analysis could be done by animated visualization showing when users become active and how controversies evolve and fade out over time.

Another promising avenue to follow is to augment the edit network by taking into account other types of interaction among authors. Wikipedia provides talk-pages on which users can discuss articles, user-pages where authors write about themselves, and associated usertalk-pages where users can make supportive comments or express their criticism to other users. A more complete network among authors would incorporate these positive or negative discussion along with the edit interaction. However, it should be noted that it is probably difficult to design a purely automatic method for determining whether statements on talkpages are meant in a supportive way or whether they express criticism.

Acknowledgments. We gratefully acknowledge financial support by the *Netherlands Organization for Scientific Research* (program *Networks of Networks*).

6. REFERENCES

- [1] B. T. Adler and L. de Alfaro. A content-driven reputation system for the Wikipedia. In *Proc. 16th Intl. Conf. WWW*, pages 261–270, 2007.
- [2] N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing*, 26:1733–1748, 1997.
- [3] U. Brandes and T. Erlebach, editors. *Network Analysis*. Springer Verlag, 2005.
- [4] U. Brandes, D. Fleischer, and J. Lerner. Summarizing dynamic bipolar conflict structures. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1486–1499, 2006.
- [5] U. Brandes and J. Lerner. Revision and co-revision in Wikipedia. In *Proc. Intl. Workshop Bridging the Gap Between Semantic Web and Web 2.0, 4th Europ. Semantic Web Conf. (ESWC'07)*, pages 85–96, 2007.
- [6] U. Brandes and J. Lerner. Visual analysis of controversy in user-generated encyclopedias. *Information Visualization*, 7:34–48, 2008.
- [7] U. Brandes and J. Lerner. Visualization of conflict networks. In M. Kauffmann, editor, *Building and Using Datasets on Armed Conflicts*, volume 36 of *NATO Science for Peace and Security Series E: Human and Societal Dynamics*. IOS Press, 2008.
- [8] M. Castells. *The rise of the network society (Vol. 1. The information Age: Economy, society and culture)*. Oxford University Press, 1996.
- [9] G. DeSanctis and P. R. Monge. Introduction to the special issue: communication processes for virtual organizations. *Organization Science*, 10:693–703, 1999.
- [10] P. Doreian, V. Batagelj, and A. Ferligoj. *Generalized Blockmodeling*. Cambridge University Press, 2004.
- [11] D. Fisher, M. Smith, and H. T. Welser. You are who you talk to: Detecting roles in usenet newsgroups. *Hawaii Intl. Conf. System Sciences*, 3:59b, 2006.
- [12] P. Heckel. A technique for isolating differences between files. *Communications of the ACM*, 21(4):264–268, 1978.
- [13] D. S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM*, 24(4):664–675, 1977.
- [14] A. Kittur, E. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 2007.
- [15] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in Wikipedia. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pages 453–462, 2007.
- [16] P. R. Monge and N. S. Contractor. *Theories of Communication Networks*. Oxford University Press, 2003.
- [17] P. R. Monge and J. Fulk. Communication technology for global network organizations. In G. DeSanctis and J. Fulk, editors, *Shaping organizational form: Communication, connection and community*, pages 71–100. Sage, 1999.
- [18] E. W. Myers. An $O(nd)$ difference algorithm and its variations. *Algorithmica*, 1(1):251–266, 1986.
- [19] S. O'Mahony and F. Ferraro. The emergence of governance in an open source community. *The Academy of Management Journal*, 50(5):1079–1106, 2007.
- [20] J. Preece. *Online Communities: Designing Usability, Supporting Sociability*. John Wiley & Sons, 2000.
- [21] R. Priedhorsky, J. Chen, S. T. K. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in Wikipedia. In *Proc. Intl. ACM Conf. Supporting Group Work*, pages 259–268, 2007.
- [22] K. G. Provan and H. B. Milward. A preliminary theory of interorganizational network effectiveness: A comparative study of four community mental health systems. *Administrative Science Quarterly*, 40(1):1–33, 1995.
- [23] K. G. Provan and J. G. Sebastian. Networks within networks: service link overlap, organizational cliques, and network effectiveness. *Academy of Management Journal*, 41(4):453–462, 1998.
- [24] B. Suh, E. H. Chi, B. A. Pendleton, and A. Kittur. Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *Proc. IEEE VAST*, pages 163–170, 2007.
- [25] T. C. Turner, M. A. Smith, D. Fisher, and H. T. Welser. Picturing usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10(4), 2005.
- [26] F. B. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pages 575–582, 2004.
- [27] F. B. Viégas, M. Wattenberg, J. Kriss, and F. van Ham. Talk before you type: Coordination in Wikipedia. In *Proceedings HICSS*, 2007.
- [28] B.-Q. Vuong, E.-P. Lim, A. Sun, M.-T. Le, and H. W. Lauw. On ranking controversies in Wikipedia: Models and evaluation. In *Proc. Intl. Conf. Web Search and Web Data Mining*, pages 171–182, 2008.
- [29] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.