Contropedia - the analysis and visualization of controversies in Wikipedia articles

Erik Borra
University of Amsterdam
borra@uva.nl

Esther Weltevrede
University of Amsterdam
esther@digitalmethods.net

Paolo Ciuccarelli
Politecnico de Milan
paolo.ciuccarelli@polimi.it

Andreas Kaltenbrunner
Barcelona Media
kaltenbrunner@gmail.com

David Laniado
Barcelona Media
david.laniado@gmail.com

Giovanni Magni
Politecnico de Milan
gvn.magni@gmail.com

Michele Mauri
Politecnico de Milan
michele.mauri@polimi.it

Richard Rogers
University of Amsterdam
rogers@uva.nl

Tommaso Venturini
Médialab, Sciences Politiques
tommaso.venturini@sciencespo.fr

ABSTRACT

Collaborative content creation inevitably reaches situations where different points of view lead to conflict. In Wikipedia, one of the most prominent examples of collaboration online, conflict is mediated by both policy and software, and conflicts often reflect larger societal debates.

In this paper, we describe the ongoing project Contropedia which aims to build a platform for the analysis and visualization of such controversies in Wikipedia. Controversy metrics are extracted from activity streams generated by edits to, and discussions about, individual articles and groups of related articles. An article’s revision history and its corresponding discussion pages constitute two parallel streams of user interactions that, taken together, fully describe the process of the collaborative creation of an article. Our proposed platform, Contropedia, builds on state of the art techniques and extends current metrics for the analysis of both edit and discussion activity and visualizes these both as a layer on top of Wikipedia articles as well as a dashboard view presenting additional analytics. Furthermore, the combination of these two approaches allows for a deeper understanding of the substance, composition, actor alignment, trajectory and liveliness of controversies on Wikipedia.

Our research aims to provide a better understanding of socio-technical phenomena that take place on the web and to equip citizens with tools to fully deploy the complexity of controversies. Contropedia is useful for the general public as well as user groups with specific interests such as scientists, students, data journalists, decision makers and media communicators.

Contropedia is still an ongoing project and the present document has been written with the aim of asking for feedback from the Wikipedia and research community.

Categories and Subject Descriptors

D.2.2 [Design Tools and Techniques]: User interfaces; H.5.3 [Group and Organization Interfaces]: Computer-supported cooperative work

1. INTRODUCTION

The aim of Contropedia is to build a tool for the real-time analysis and visualization of discussions and controversies in systems for collaborative content generation. The tool helps to achieve a deeper multidisciplinary understanding of the web as a societal artifact and its function as a mirror of societal controversies. As experimental field we choose Wikipedia, the largest platform for such content and also the overall sixth most trafficked website\(^1\) at the moment this paper was written. We are currently developing a real-time dynamic atlas of the substance, composition, actor alignment, trajectory and liveliness of controversies on Wikipedia.

We have followed an interdisciplinary approach in designing and implementing the Contropedia platform combining experts from the fields of social science, data mining, data visualization as well as new media and digital culture studies. The tool is anchored in previous experiences of mapping public debates obtained from the ongoing EU FP7 project, EMAPS\(^2\), and its predecessor MACOSPOL\(^3\). These two projects have been developing a toolkit of conceptual and methodological instruments to explore and represent techno-scientific disputes. Drawing on the traditions


\(^3\)Mapping controversies in science and technology for politics (http://www.mappingcontroversies.net/). Last accessed on May 1, 2014.
of pragmatist political thinking [11] and of the science and technology studies [4], the method of controversy mapping has been introduced by Bruno Latour to understand and participate in public debate [20, 21]. The idea behind such method is that conflict is an essential part of collective existence and that its energies can be positively channelled (or at least the most destructive outcomes can be avoided) by equipping citizens with the tools to deploy the full complexity of controversies. Developing such tools is becoming increasingly feasible thanks to the growing traceability of collective discussions made possible by the spread of digital media [22].

Here we extend these concepts to the controversies which may evolve around Wikipedia articles. While Wikipedia’s apparatus is primed to produce encyclopedic content, in a system where everyone is allowed to edit, it comes as no surprise that agreement about content is not always easily reached. To maintain its ‘encyclopedianess’ Wikipedia has mechanisms in place with which consensus is designed, such as the core content policies neutral point of view (NPoV), verifiability and no original research. By recognizing Wikipedia’s purpose to reach consensus, we suggest that by scrutinizing the disagreements which become apparent throughout an article’s edit-history and talk pages, Wikipedia can be repurposed to study not only consensus but controversies [26].

Conflicts in Wikipedia articles have previously been studied by observing article edit histories [3] - where each edit and the user who made that edit are logged, by considering reverts [1, 18] - a key mechanism in Wikipedia for repairing detrimental edits to an article, and by analyzing talk pages [6, 9] - parallel spaces for discussion about the article content. Additionally, research has been pursued to characterize and visualize conflict and coordination on Wikipedia [7, 10] and to identify which articles are controversial [24, 27]. The vast majority of these studies are primarily focused on the social dynamics between editors and only a few have taken into account the content of controversial edits [17, 23], as we do in Contropedia.

Furthermore, while most of previous literature uses articles as basic units of content, in this project we go beyond the mere identification of controversial pages by studying the controversies within an article, or a set of articles, in more detail. We combine techniques based on both mining article edit history and analysing discussion patterns in talk pages to identify and visualize what exactly is controversial within an article (which figures, concepts, references, etc.), how these controversies evolve, which users are involved, whether these users belong to opposing camps, and what issues they hold dear. While Contropedia uses up-to-date data, each controversy will be provided with a historical view to show which of the controversial issues within a (set of) articles are especially active and at which time (are they hot or not?, what is the trajectory of a contested issue?).

During the development of Contropedia we have put special focus on Wikipedia’s articles related to the debate on climate change. There are several reasons for this choice. From a practical point of view, this choice allows us to profit from the synergies of two other research project in which the partners of this consortium are participating: EMAPS and MEDEA, which are both devoted to the analysis of the discussions around global warming adaptation. From a political point of view, climate change is probably the most important current techno-scientific debate, both in terms of the amount of actors and resources involved and for the impacts that its outcomes are likely to have on collective life. Finally, climate change has always been one of the most intensely discussed topics on Wikipedia [8].

The rest of this paper is structured as follows: In Section 2 we first introduce some theoretical background on controversy mapping and explain why Wikipedia is especially well suited for controversy mapping. In Section 3 various metrics are introduced which allow one to locate the substance of controversies within an article. In Section 4 the different modules and functionalities of Contropedia’s user interface are explained and finally, in Section 5 we recapitulate the goals and uses of Contropedia and point towards future directions of research and development.

2. CONTROVERSY MAPPING ON WIKIPEDIA

Here we briefly discuss why Wikipedia is suitable for controversy mapping. We do this by connecting the defining characteristics of controversy mapping with the relevant elements of Wikipedia’s socio-technical apparatus. This discussion establishes the prerequisites with which we can explore the substance of controversies.

2.1 Controversy mapping

Controversy mapping is the practice of Actor-Network Theory (ANT) through which one aims to observe and describe the debate around techno-scientific issues [20]. Something is said to be controversial when it becomes apparent that the actors in a debate disagree and cannot ignore each other until a compromise is reached. As such, controversies constitute the best setting to observe the construction of social life, as these are the moments at which the relations between actors and their positions are articulated and, because of the fluidity of debates, constantly repositioned. Mapping out this multiplicity of viewpoints allows one to find out what is at stake in the unfolding of the controversy.

Although an elaborate treatise of ANT is beyond the scope of this paper, for the purpose of this research it is imperative to understand that in ANT an actor is understood to be anything which is acting. That means that whenever the presence or absence of something makes a difference, and whenever this difference is perceived by other actors, that something can be said to be an actor in the controversy. Actors are thus “not only human beings and human groups, but also natural and biological elements, industrial and artistic products, economic and other institutions, technical artifacts and so on and so forth” [20, pp4]. Furthermore, an actor is never isolated but always composed by, and a component of, a network. An actor exists because it inter-acts; it is always relating and being related.

Now that the very basics of controversy mapping have been introduced Wikipedia articles can be positioned as sites of controversy, which we do in the following paragraphs.

2.2 Wikipedia as a controversy defusing device

While much research has been devoted to the identification of controversial articles, see e.g. [24, 27], to our knowledge little research has yet revealed which content within an article is controversial. Our project seeks to go beyond the
mere identification of controversial articles and seeks to discover within a controversial article what exactly is disputed, when it is, and to what extent.

Wikipedia’s goal is the production of encyclopedic content through an open and collaborative environment. It will come as no surprise that agreement about content is not always easily reached. Therefore, Wikipedia has a socio-technical apparatus [13] in place to support consensus reaching – consensus is even said to be “the primary way decisions are made”[14]. It does not imply unanimity nor is it the result of a vote, it is a form of decision-making which involves “an effort to incorporate all editors’ legitimate concerns, while respecting Wikipedia’s norms”[15]. Phrased differently, consensus is the guiding principle for knowledge production and conflict resolution on Wikipedia. Consensus is achieved most notably through the core content policies neutral point of view (NPoV), verifiability[6] and no original research[7], although Wikipedia also offers a plethora of other tools and techniques to defuse controversies: to avoid edit wars there is a three-revert rule[8], controversial matters should be further discussed on talk pages, etc.

The built-in edit history and the talk pages of a Wikipedia article are rich sources for the controversy mapper as they meticulously document the work involved in reaching consensus. They provide a detailed record of present and past changes to the content of articles and the unfolding of discussions on the talk page; they reveal the fabric of collective knowledge production as editors of Wikipedia engage in tying and untying relations, and argue about categories and terminology. Because of this, we propose that the edit history of a controversial article can be controversial within that article [26].

Returning to ANT’s notion of an actor, we posit that each linked element in a controversial Wikipedia article can be considered an actor in the controversy. According to Wikipedia’s guidelines links should generally be created to “any linked element in a controversial Wikipedia article can be controversial within that article [26].”

Before our metrics are run, our system will retrieve the full edit history for a controversial article, including the wiki text of each revision and the meta data conveying at what time it was edited, who the editor was, as well as the editor’s comment. As we are specifically interested in the substance of the controversy, or more specifically, those edits where substantive contributions are made, our measures do not take into account vandalism edits or vandalism reverts[13]. Additionally, we have filtered out punctuation and maintenance edits[14].

3. THE SUBSTANCE OF CONTROVERSIES

Now that we have established links within an article, or more generically wiki objects[12], as the actors and focal points within a controversial article, it is possible to find out which of these actors have been most controversial within the article. In this section we first discuss metrics based on the edit-history of a controversial article, then we discuss a metric based on the talk page of an article, and lastly we discuss how we combine these two metrics.

3.1 Assigning controversy scores based on the edit history

We have implemented two metrics based on the edit-history to indicate which of the actors have been the locus of most edit activity or debate. By looking at edits to the sentences in which these wiki objects appear, they both use the activity around a specific actor as an indication of controversiality.

Let \( \{R_1, \ldots, R_{r-1}, R_r, R_{r+1}, \ldots\} \) be the set of revisions of a Wikipedia article. As we are specifically interested in the substance of the edits we consider the edit activity on a sentence level by comparing every revision \( R_{r-1} \) with its successor \( R_r \). We first split each revision into sections [16] and then make a pairwise comparison of the sections via a Levenshtein distance smaller than five.

We do this both because [10] and [14] have shown that controversy is located in specific sections of an article and because it makes comparison easier.

11. We consider an edit as a punctuation or maintenance edit if the textual difference between two revisions has a Levenshtein distance smaller than five.
12. Whenever a user makes multiple consecutive edits, we compare the last version made by the user with the version which she started revising.
13. We do this both because [10] and [14] have shown that controversy is located in specific sections of an article and because it makes comparison easier.

diff algorithm. Whenever a difference between two sections is detected, we identify the individual sentences which are edited and the wiki objects contained in them.

For each wiki object thus obtained we store in a database the article’s revision id, the user who made the edit, the comment added by the user, the sentence which was changed, the diff of the edit, the wiki object, its canonical form, and the type of edit (inserted, deleted, changed, or change in surrounding sentence).

### 3.1.1 Edit-based measure

Although we are interested in finding out how controversial a wiki object $O_k$ is, we are comparing the edit activity of sentences in which $O_k$ appears. Intuitively, the more wiki objects appear in an edited sentence, the less focus there is on one particular wiki object, i.e., an actor. For every edited sentence $S_j$ we thus divide the weight attributed to a wiki object by the total number of wiki objects $o(S_j)$ which appear in that sentence.

For a given revision $R_r$ and each wiki object $O_k$ in edited sentences $S_j$ we assign a controversy score $c(O_k)$

$$c(O_k) = \sum_{i=1}^{r} \sum_{S_j \in R_i} \frac{1}{o(S_j)}$$  \hspace{1cm} (1)

In other words: over all revisions up to $R_r$ and over all edited sentences where $O_k$ appears in a certain revision we sum the inverse of the number of wiki objects in these sentences.

We can also make reverts count more:

$$c(O_k) = \sum_{i=1}^{r} g_i \cdot \sum_{S_j \in R_i} \frac{1}{o(S_j)}$$  \hspace{1cm} (2)

where $g_i = 2$ when the edit of revision $R_i$ is a revert and $g_i = 1$ if it is a normal edit.

We obtain the normalized controversy score for a wiki object in a particular article by normalizing all the scores by the wiki object with the maximum controversy score. To find out which wiki object is most controversial, i.e. around which actors most negotiations took place, we simply rank the wiki objects of the article in descending order.

### 3.1.2 User-based measure

While the ‘edit-based’ measures of Eqs. (1) and (2) already provide wiki objects with an intuitively intelligible controversy rank, we have also experimented with a ‘user-based’ measure, inspired by the mutual revert measure of [19].

Again, let $\{R_1, \ldots, R_{q-1}, R_q, R_{r+1}, \ldots\}$ be the set of revisions of a Wikipedia article. We denote by $n^u_r$ the number of times a user $u$’s author of revision $R_r$ has edited the article. We characterize edits by tuples $(n^i_r, n^u_r)$, where $u$

odes the index of the user who makes edits to the article versions edited by user $U_u$. Note that $r = q + 1$.

As we want to focus on wiki objects instead of revisions, we denote by $n^w_r(O_k)$ the total number of times (until revision number $r$) in which wiki object $O_k$ was edited by user $U_u$.

Let us denote in analogy to Eq. (1) $\tilde{n}_r(O_k) = \sum_{S_j \in R_r} \frac{1}{o(S_j)}$ as a weighted version of this counter so that we include the notion that there should be less weight attributed to object $O_k$ if multiple wiki objects are present in the edited sentences $S_j$. Weighted edit pairs for a specific wiki object $O_k$ are then defined as $\tilde{n}_r^w(O_k) = \frac{1}{o(S_j)}$, where

$$\tilde{n}_r^w(O_k) = \sum_{R_r \in U_u \text{ changes } O_k} \frac{1}{o(S_j)}$$  \hspace{1cm} (3)

In other words: we sum the weighted edit count of all revisions of user $U_u$ where she changes Object $O_k$.

When ranking editor contributions, two main edit types can be distinguished: when one or both of the editors have made few edits to the article, these are typically editors who are not “invested” in the controversy, and when both editors are heavily invested. In order to express this distinction numerically, we use the lesser of the re-weighted edit counter $\tilde{n}_r^w(O_k)$, so that the total count includes edits by less invested pairs of users as well, but with a much smaller weight.

Finally, we multiply by the number of editors $|E_i|$ (the larger the armies, the larger the war).

$$c(O_k) = |E_i| \sum_{i=1}^{r} \min \left( \tilde{n}_r^w(O_k), \tilde{n}_r^w(O_k) \right)$$  \hspace{1cm} (4)

Additionally, we can censor the editor pairs with greatest weighted edit counts to eliminate cases with conflicts between two persons only.

$$c(O_k) = |E_i| \left( \sum_{i=1}^{r} \min \left( \tilde{n}_r^w(O_k), \tilde{n}_r^w(O_k) \right) - \max_{i=1} \left( \min \left( \tilde{n}_r^w(O_k), \tilde{n}_r^w(O_k) \right) \right) \right)$$  \hspace{1cm} (5)

The results are similar to the ‘edit-based’ measures, with small differences in emphasis.20

Section 4 reports about the uses of these metrics. First, however, we introduce how we measured controversy scores for actors based on the activity in the talk pages.

### 3.2 Assigning controversy scores based on the talk pages

Besides looking at controversies in article edit histories we analyze talk pages, which are special wiki pages associated to each article and devoted to discussion about how to improve article quality. Here is where controversies are explicitly discussed by editors of an article. In cases in which an article

\[\text{\textcopyright 2023 American Meteorological Society. All rights reserved.} \]
is protected from editing\textsuperscript{21}, talk pages are the only place where controversy takes place.

First, we identify all talk pages associated to an article and parse them to detect thread titles, and signature and date of each comment. Additionally we look at comment indentation to extract the thread structure of messages and replies, i.e. we reconstruct the discussion tree. For this step we follow the methodology of [6] and [9], but instead of quantifying controversiality at the level of articles, we do it at the level of discussion threads.

We represent each thread as a tree of comments and replies, and we characterize controversiality of each thread based on two metrics, representing respectively depth and width of the discussion. The first metric, derived from the structure of the discussion tree, is the maximum depth of the thread, i.e. the maximum level of indentation of comments. When comments are directly written under the thread title without indentation, i.e. no comment is a reply to another comment, maximum depth of the thread is \( 1 \). If some comment is a reply to another comment, maximum depth is \( 2 \), while if there is a reply to a reply maximum depth is \( 3 \), and so on. High values of this metrics indicate high presence of argumentation among users and is therefore a straightforward proxy for controversiality. However, a deep discussion thread can be created just by a few users, typically just by two users arguing with each other. Thus, as a complementary metric we take the number of users participating in a thread, representing the width of the discussion.

### 3.3 Combining controversy scores from the edit history and the talk pages

Discussions in talk pages are not explicitly associated to the specific sections of the article to which they are associated, so we establish this connection relying on a set of heuristics to associate each discussion thread to one or more sections.

The first method consists of detecting when a user mentions a discussion at the moment the article is edited. When submitting an edit, a user has to fill in a comment field describing the content of the edit. If in this comment we find a link to a discussion thread, we deduce that the editor is executing what has been deliberated in the corresponding thread, or is anyway referring to that discussion as related to the edit. We therefore interpret this as straightforward evidence of a connection between the article section affected by the edit and the discussion thread mentioned in the edit comment.

As a second method, we look for each section’s title in discussion thread titles, or in the text of the discussion threads. If the section’s title is mentioned in a discussion thread, we establish a link between the two. Analogously, we also look for the word “abstract” in the talk pages to detect connections between a discussion thread and the article’s abstract.

To be able to match a higher number of discussion threads, and have a more robust matching, we are currently developing more sophisticated methods, based on common elements between a discussion thread and an article section. Namely, we look for the co-occurrence of common actors mentioned, and of common users active in the same time-period. In other words, if we find that the same actors are mentioned in a discussion thread and in a section of the article, we hypothesize a relationship between the two. Similarly, if some users are editing the article and discussing in the associated talk page at the same time, we interpret this as a hint for associating the discussion thread in which these users are discussing to the article section they are editing.

After associating discussion threads to article sections, we are able to assign to each actor and each section the controversiality indexes coming from the analysis of both sources.

### 4. THE COMPONENTS OF THE CONTRO-PEDIA PLATFORM

Contropedia will be an online platform\textsuperscript{22} that seeks to provide insights into the disagreement about the substance of controversial Wikipedia articles. As addressed in the previous section, the platform makes use of the rich data that is available on the history and talk pages of Wikipedia articles to map out the disputed aspects of controversial articles. The aim is to provide insights into the debate itself and gain an understanding of the matters of concern, i.e. the disputed actors within an article. In this section we describe and present the design of the main components of the Contropedia platform. The questions we have currently sought to answer are: What is controversial and to what extent? Who is involved in the controversy? When are controversies controversial?

These answers are addressed in a number of components, which consist of a layer on top of the controversial Wikipedia article as well as a dashboard view of controversial actors. These visualizations depict three main sources of data taken from the Wikipedia platform: controversial elements within an article, discussion threads from talk pages, and interaction networks of Wikipedians editing each other’s revisions.

### 4.1 The overall platform design

Contropedia provides visual access to the our analyses, in order to assist the user in understanding how controversies are deployed. Diagrammatic tools are particularly suited for this purpose, as controversies need tools that do not divide or analyze the elements separately but present them in an interconnected and indivisible manner [15]. In Contropedia, information visualization is used to convey the page history, its genesis and the social dynamics which generated it. Providing visual access to the page history provides an overview and allows the user to study the dynamics behind controversial topics. It also allows to compare controversial pages, to identify patterns, similarities and differences. Compared to other tools that use visualization on Wikipedia, see e.g. [7, 25], Contropedia proposes a shift in visualization goals. The visualization does not just represent an abstract view of Wikipedia data but overlays our measures of controversiality onto Wikipedia’s contents, similar to annotations. We have applied methods from information visualization to identify and design visual models in order to communicate the data and to provide useful interactions to explore controversies [2]. The aim of these visualizations is not to convey data in a more efficient way, but rather to use them to shape new knowledge about the topic [12].


\textsuperscript{22}A demo of Contropedia can be found at \url{http://www.contropedia.net/demo}. We provide it here for review purposes only. Please do not distribute the link or make it public, as the system is still in an experimental phase.
Contropedia seeks to annotate and rework the original Wikipedia article layout to show which actors are controversial and in which part of the article they are. The whole page is represented through a minified version of the article (see Figure 1)\(^2\) that simplifies the original article and highlights the controversial elements while respecting the original article’s layout. To achieve this, a visual contrast has been created among controversial and non-controversial elements. Controversial elements are represented through five colour shades, from the most controversial (red) to the least one (pale blue). To enhance the visual contrast, colors and graphical elements were removed from the page. Text is rendered using a block font, enhancing each section proportion and controversial elements position. Images are replaced with a standard placeholder. We used gray colours to represent all non-controversial elements.

Contropedia’s visual navigation is based on the semantic zooming pattern: “Zooming need not follow a strict geometric metaphor: semantic zooming methods can modify both the amount of information shown and how it is displayed as analysts move among levels of detail.”\(^5\) In Contropedia, each page is subdivided according to article sections. Empirically, article sections are a good solution to sample the article in coherent sub-elements,\(^2\) as sections can be seen as minimum thematic elements. In our visual representation, each section will share the same graphical structure. The use of a compact, minified, representation allows one to get an overall view of the article and to identify the most relevant, controversial, sections at a glance. The user is then able to expand part of the article and see the article’s text in its original layout. By hovering a wiki object, a further level of zoom is provided showing the list of edits related to the wiki object, i.e. the debate about the actor.

4.2 Where is the controversy located, and what is controversial and to what extent?

Contropedia shows how much a given part of an article is edited sequentially by disagreeing article editors (to appreciate the lack of consensus on it). Additionally, if the section of an article is discussed in the talk pages, the correspond-

\(^2\) Figures 1, 2 and 3 can be seen as a functional demo on http://bit.ly/i1E0GcQ. Last accessed May 1, 2014.


4.2.1 Controversial edits

When the application is loaded, each section is represented with its title, a lateral bar showing the overall controversy level, and two controversy navigation buttons: discussions and content (see Figure 1). When clicking discussions the main threads related to the section are represented through dots. When clicking content the minified version of that section can be extended so that a zoom overlay of the original article’s section is shown (see Figure 2).

The content button reveals the original Wikipedia structure of the article with additional highlights on the most controversial wiki objects. Hovering those actors allows the users to go into deeper detail and shows them the section as if they were on the original page (see Figure 2). All the original elements are set in gray scale mode, as links are highlighted with a colored background according to their controversiality index as the images are overlaid by a shape, this way we can show controversial parts directly on the Wikipedia format in a less abstract way.

On mouse-over the highlighted controversial wiki objects within an article can be expanded to show the edits involving, or debate about, that wiki object (see Figure 3). This function allows the user to zoom in and further look into the edit history of a particular wiki object as the edits are divided and presented by wiki objects, as opposed to the overall article history over time as offered by the original Wikipedia interface. The edit view allows the user to further scrutinize the construction of controversial elements in an article and to gain an understanding of what precisely is disputed about that particular subsection of the controversy.

4.2.2 Controversial discussion threads

Underneath the discussion button at the top of the section, dots represent the discussion threads in the talk pages that are about that section. The size indicates the number of Wikipedia editors involved in the discussion thread and the redness indicates the controversiality of the thread.

\(^2\) At the time of writing this paper, this specific link has not been implemented in the demo yet.
4.2.3 Controversial wiki objects

The layer view provides a powerful way to present what the controversial elements are within the context of the article. In addition to the layer view the Contropedia platform presents a dashboard view to quickly identify the most controversial elements (see Figure 5)\(^\text{26}\). This view is designed to address a more analytical view on the data and also includes highly controversial elements which might have been deleted in the shown revision. This dashboard consists of a table listing all the controversial elements, from the most controversial to the least one. Each cell contains a different variable associated to it, represented through a suitable visualization. The table adopts the small multiples idea, allowing a vertical comparison of the same variable among controversial elements. Through its color represents the bar on the left how controversial and actor is.

A timeline shows the amount of edits through time, allowing the user to identify historical periods where several edits affected the same element. The timeline is visualized using a horizon chart; a visualization which is particularly suitable for items with high variability that should be represented in a compact way. The type of actor is represented through an icon. All the other variables are represented through bars. Deletes, inserts and changes are normalized on the maximum value of the three. In this way it is possible to confront the same variable among different elements but also to compare values of the same item. The remaining variables are normalized independently of their maximum value. This is due the fact that are different properties of the page, and it would not make sense to compare them. The timeline visualization is also meant to provide access to the full list of single revisions. Following the folded interaction which was used in the minified layout, when the timeline is clicked, all edits to that wiki object are shown (see Figure 3). This revisions table adopts a classical “diff visualization” showing the added and removed parts with color-coding. It also shows several other relevant variables like the name of the editor, the timestamp of the edit, the comment, and the name of the section in which the edit was made.

The interactive layer view on top of the article, as well as the additional dashboard view, show various analytics and provide the user of Contropedia with various possibilities to quickly gain an understanding of what is controversial, where something is controversial within the article, and to what extent. A feature which we have not yet implemented will be placed at the top of the layer view and will allow the user of Contropedia to set the time frame to visualize the selected article. This will allow the user to explore the trajectory of which sections and actors were controversial at which time, i.e. at what point an actor was still hot or whether its controversiality has cooled down over time.

4.3 Who is involved in the discussion?

The Contropedia platform helps in discovering who are the main actors of the controversy. In addition to extracting the actors in the debate, Contropedia also extracts the actor of the debate, that is to say, the Wikipedians which have been most active in rousing the controversy. In Contropedia the editors of an article are used as proxies to identify

---

\(^{26}\)The figure is accessible in the Contropedia Demo under http://bit.ly/1s4bks3. Last accessed May 1, 2014.
4.3.1 Interaction graphs

An interaction in a reply chain means that in a specific discussion thread on the article talk page the users have mutually replied to each other consecutively. The most simple case we consider is that user A writes a comment, user B replies to this comment and user A replies back. Much longer chains can occur.

Examples of these graphs can be observed in Figure 6 (reply network) and Figure 7 (revert network). In both cases anonymous users have been removed from the networks. The edge weights indicate the number of reply chains between each pair of users in the reply network and, in the case of the reverts network, the number of reverts between the two users (more exactly the sum of the reverts between the two users in both directions divided by two and rounded up to the nearest integer).
4.3.2 “My enemy’s enemy is my friend” network

All the interactions described in the previous sub-section can be considered interactions between antagonists in the controversies. Thus, standard community detecting methods which aim at maximizing the number of connections within (and minimize those outside) a community are not suitable here.

We have therefore developed a Gephi plugin\(^{27}\) aimed at finding communities when edges are connecting antagonists. It is based on the idea that “the enemy of my enemy is my friend”. The plugin first computes a “friendship” score between each pair of nodes. The more common antagonists two nodes have, the higher their friendship score will be. The more connected two nodes are, the lower their friendship score will become. Then all original edges of the network are removed (for community detection) and each pair of nodes that has a score higher than a given threshold will become connected. Finally, Gephi’s modularity algorithm is used to detect communities in this new graph. Later the original edges of the network are restored and the provisional edges are removed again.

Examples of the outcome of this algorithms can be found in Figure 6 (reply network) and Figure 7 (revert network). Nodes from the same communities (those “fighting” with the same users) have the same color. We observe that the large majority of interaction happens in between the users of the different communities. The algorithm is still in an experimental phase and extensive user evaluation are currently being performed to verify their outcome. Nonetheless an initial verification of the outcome of the algorithm indicates that the obtained communities indeed correspond to persons with similar positions in the controversies.

While currently these networks are generated on the basis of an article’s full edit history, thus involving all wiki objects within the article, we will soon also implement these networks around specific wiki objects. This will allow us to indicate per actor whether the debate around it is polarized or involving multiple groups, whether the controversy is artificially kept alive by a specific set of users (which might point towards slant or purposeful bias), etc.

5. CONCLUSIONS AND OUTLOOK

The DMI proverb to “learn from medium method”\(^ {16}\) entails that any online (search) engine or platform has its own specific methods and objects with which digital researchers can work. This medium-specific approach seeks to identify the core objects of the medium, to study how these objects are handled by the medium and to repurpose medium method productively. Closely looking at the inner workings of Wikipedia and thereby recognizing that Wikipedia works as a controversy defusing device, and that each link within an article can be seen as an actor in a debate, has allowed us to repurpose controversial Wikipedia articles as ideal sites to map controversies. We have shown in this paper how actors in a Wikipedia article may be assigned controversy scores by using edit histories and talk pages. We then introduced the design of Contropedia which is intended to get a quick overview of what is controversial in an article, where and to what extent. The design, and accompanying demo, furthermore allow the users of Contropedia to zoom in to the specific discussion and changes around actors in the unfolding debate. The last part of our paper dealt with detecting which Wikipedians which have been most active in rousing the controversy, identifying the multiplicity of viewpoints around actors in a controversy and showing which issues they hold dear. We thus believe that Contropedia, as an implementation of controversy mapping, provides an elaborate toolkit of social life as it unfolds on Wikipedia. We thus foresee many different types of users:

- Wikipedians can gain insight into the substance and build-up of controversies, allowing informed decisions about the management of “edit wars” and disagreements about the articles’ content.
- For scientists it will provide a tool to visualize in real-time the dynamics of techno-scientific debates, stimulating the framing and phrasing of scientific issues, helping to clarify and defuse the conflict.
- For teachers the platform could be employed to teach the complexity of techno-scientific controversies. By showing the full complexity of a scientific controversy as well as its entanglement of relations and networks, Contropedia can provide an innovative viewpoint on ‘science in action’.
- For decision makers, both public and private, it will allow to more readily understand the essence and the extent of open controversies and therefore the assessment of their potential consequences.
- For communicators, on professional media or open platforms (e.g. blogs), access to the background of the controversies will facilitate a better informed description of facts and issues to the public.
- For citizens wishing to be better informed, and for society in general, access to an integrated platform mapping the nature and the extent of Wikipedia controversies will help to build a greater trust in the online encyclopedia as well as to make its limits explicit.

Although we are still in early phases of the Contropedia platform this paper is intended to share our work as soon as possible with the Wikipedia community and Wikipedia researchers, so that we can get early feedback about the usefulness of our system. As such we have already identified at various points in our paper where there is still room for improvement or experimentation. A few other areas could also use further exploration. First, as a link should generally “appear only once in an article, but if helpful for readers, [they] may be repeated in infoboxes, tables, image captions, footnotes, and at the first occurrence after the lead”\(^ {28}\), to increase the resolution of our metrics it could be fruitful to also detect where in the text a link could have been. Second, as in our current metrics each article will always have a most controversial actor, we currently focus only on articles which are already known to be controversial. Third, while we currently focus on a subset of the controversial articles in the English Wikipedia we would like to expand these

\(^{27}\) The Antagonist Based Community Detection plugin for Gephi can be found at https://github.com/jacomyal/ABCD-Gephi-Plugin.

techniques to other controversial articles and other language Wikipedias. And last, apart from just looking at the controversy of actors within one article, we also intend to allow Contropedians to explore related sets of articles. While currently we provide a zoom in to actors within an article, we could thus also allow one to zoom out and see which articles are more controversial in the network of related articles.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the EU FP7 EINS under grant agreement No 288021. We also wish to thank Alexis Jacomy and Benjamin Ooghe for their help in the development of this project.

7. REFERENCES