

Collective memory building in Wikipedia: The case of North African uprisings

Michela Ferron
Center for Mind/Brain Sciences
University of Trento
Corso Bettini 31
Rovereto, 38068
+39 0461 314331
m.ferron@unitn.it

Paolo Massa
Fondazione Bruno Kessler
Via Sommarive 18
Trento, 38123
+39 0461 314336
massa@fbk.eu

ABSTRACT

Since December 2010, a series of protests and uprisings have shocked North African countries such as Tunisia, Egypt, Libya, Syria, Yemen and more. In this paper, focusing mainly on the Egyptian revolution, we provide evidence of the intense edit activity occurred during these uprisings on the related Wikipedia pages. Thousands of people provided their contribution on the content pages and discussed improvements and disagreements on the associated talk pages as the traumatic events unfolded. We propose to interpret this phenomenon as a process of collective memory building and argue how on Wikipedia this can be studied empirically and quantitatively in real time. We explore and suggest possible directions for future research on collective memory formation of traumatic and controversial events in Wikipedia.

Keywords

Wikipedia, Web 2.0., Collective Memory, revolutions, traumatic events, Egypt, North Africa.

1. INTRODUCTION

A series of pro-democracy rebellions and demonstrations across North Africa characterized the end of 2010 and the subsequent months. The first protests occurred in Tunisia following Mohamed Bouazizi's self-immolation in the town of Sidi Bouzid on 17 December 2010, and quickly started to spread in other countries such as Egypt, Algeria, Libya, Lebanon, Jordan, Palestine, Iran and Bahrain. Soon, the Middle East was shocked

by a wave of violent demonstrations, riots and strikes, resulting in the overthrow of two heads of state, the presidents of Tunisia Zine El Abidine Ben Ali and of Egypt Hosni Mubarak.

As these events occurred in the streets with the participation of a multitude of people, a less visible share of activity happened also on the online encyclopaedia Wikipedia: thousands of people provided their contributions to the related Wikipedia pages promptly created as the uprisings spread, trying to make sense and represent them as they unfolded over time.

Events like these can seriously traumatize the involved populations, because they shake the core of the communities identities, affecting different layers of the social and cultural tissue of these societies [1]. Cultural trauma, which is bonded to the collective identity of a community and to the formation of its emergent collective memories [2, 3], can arise as a consequence of the social change.

Theoretical perspectives agree that cultural trauma is a socially mediated process, but nevertheless it is still possible to outline a number of defining characteristics of potentially traumatic events. Neal [4] and Sztompka [1] argued that to be potentially traumatizing for a social group, an event should cause “disruption” and “radical change ... within a short period of time”, and should be wide and comprehensive, sudden and deep. In the case of North African revolutions, the protests and related happenings affected different aspects of social and personal life, involving many actors, causing radical change in the core constitution of Tunisian and Egyptian societies with the overthrow of their leaders, and rapidly spreading across other North African countries.

In order to understand how this kind of traumatic social change can affect the collectivity, it is interesting to study the progressive formation of collective memories about these events. Intending collective memory as the continuous active process of sense-making and negotiation between past and present [5], in this paper we argue that now it can be studied empirically also employing quantitative techniques, social network analysis (SNA) and natural language processing techniques, integrating the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '11, October 03 - October 05 2011, Mountain View, CA, USA
Copyright 2011 ACM 978-1-4503-0909-7/11/10...\$10.00.

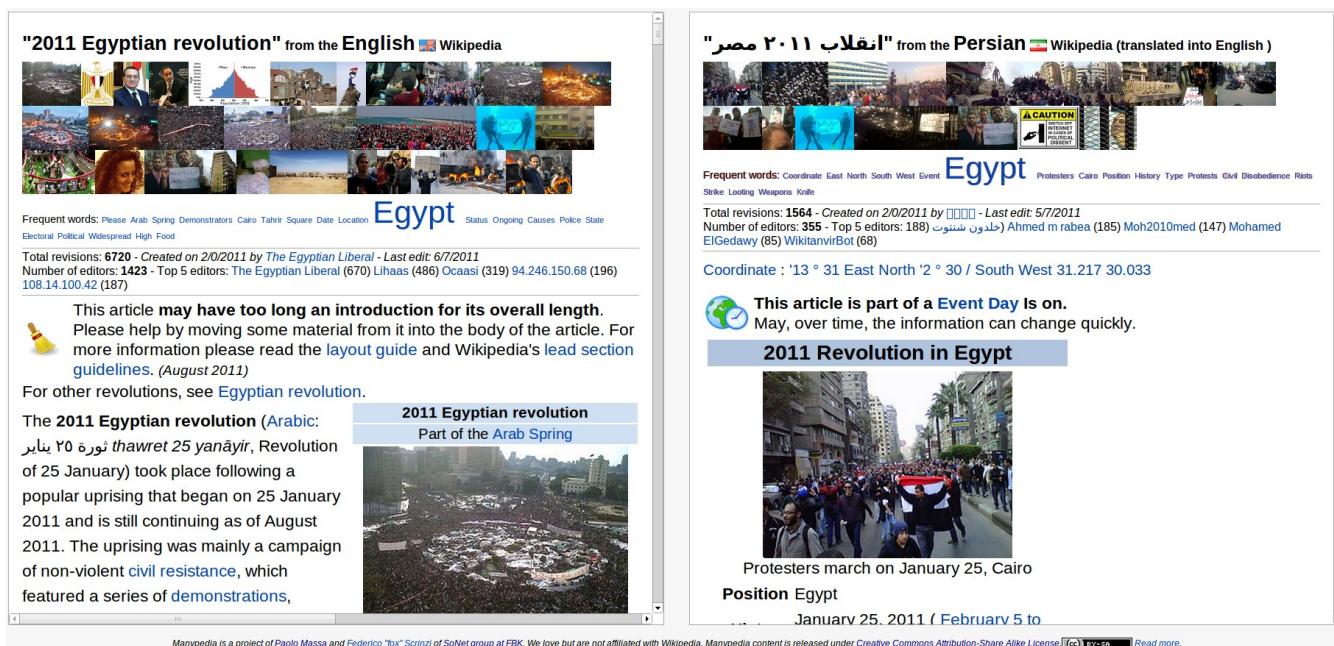


Figure 1. Pages about “2011 Egyptian revolution” from English (left) and Persian (right) Wikipedia compared using Manypedia [21]. Manypedia automatically translates pages into a language chosen among 56 different languages and adds additional descriptive information such as most frequent words, number of edits and editors, creation date, top editors and images used in the article.

perspectives provided by traditional theoretical research with a different point of view.

Indeed, new digital media and social networking sites allow new forms of participation on a global scale, and researchers can now access, in an unobtrusive way and almost in real time, large amounts of data describing people's thoughts and their communicative interactions [6, 7]. One of the most interesting Web 2.0 platforms for the study of collective memory processes is the online encyclopaedia Wikipedia [8], which is built every day by the voluntary work of millions of people apparently without receiving explicit credit for their contributions [9], and whose pages can be created and modified by anyone.

In the context of communication technologies we consider Wikipedia as lying between the top-down information provided by traditional news agencies and the bottom-up individual thoughts often posted by users of new media and social networking sites, such as Facebook or Twitter, whose role in the context of North African uprisings has been largely debated [10]. According to Wikipedia policies, every article should be built on the basis of reliable and published secondary sources, and not on first-hand evidence. Moreover, Wikipedia is expressively intended to be edited collaboratively and its articles should exhibit a neutral point of view [NPOV; 11]. Indeed, according to Wikipedia policies, “editing from a neutral point of view (NPOV) means representing fairly, proportionately, and as far as possible without bias, all significant views that have been published by reliable sources”.

In this article we follow Pentzold's [8] interpretation of Wikipedia as a global memory place [12], where events are discussed and their memories are continuously built through the negotiation of different narratives and points of view. For instance, the article about the “2011 Egyptian revolution” is the ongoing description of the protests and related happenings,

resulting from thousands small edits and the debate on divergent perspectives.

In the next section, focusing mainly on the recent Egyptian revolution, we argue that the current North African uprisings triggered a heavy and abrupt activity on Wikipedia, which we interpret as the beginning of collective memory building processes. In the last section we provide directions on how this vast amount of empirical data can be exploited for research purposes.

2. THE COLLECTIVE MEMORY OF THE NORTH AFRICAN UPRISINGS IN WIKIPEDIA

Since December 2010 there have been demonstrations and protests occurring in the North African countries such as Tunisia, Egypt, Libya, Syria, Yemen and more. The collective memories of each of these uprisings, later grouped in Wikipedia under the name “Arab Spring”, started to be created in a collaborative way in dedicated pages which triggered an intense participation in the online encyclopaedia since the very first days of protests. Interestingly, the title of the articles itself was an important matter of discussion as the pages were renamed by different users over time to describe the associated events as uprisings, revolutions, civil wars, major or minor protests.

In this section we focus on the Egyptian revolution in order to present a complete picture of the construction of the related Wikipedia articles in different languages, but similar considerations can be made for each of the pages related to the Arab Spring events in other countries.

On the English Wikipedia the article “2011 Egyptian revolution”, also called “2011 Egyptian uprising”, was created the same day the protests started, on 25 January 2011 at 13:26 (Greenwich

Table 1. Statistics about the articles and talk pages on the “2011 Egyptian revolution” from four different language Wikipedias, at 10 March 2011. Data are extracted using the Articleinfo tool [13].

| | English (en) | English (en) Talk | Arabic (ar) | Arabic (ar) Talk | German (de) | German (de) Talk | Egyptian Arabic (arz) | Egyptian Arabic (arz) Talk |
|---|--|--|----------------------------------|--------------------------------------|--|--------------------------------------|--------------------------------------|---|
| Creation date | 25 January 2011, 13:26 (by The Egyptian Liberal) | 25 January 2011, 13:26 (by The Egyptian Liberal) | 25 January 2011, 16:55 (by عمرو) | 25 January 2011, 20:35 (by Jo NaHaL) | 26 January 2011, 10:53 (by 195.145.91.170) | 26 January 2011, 11:24 (by Generato) | 28 January 2011, 01:52 (by Ghaly) | 05 February 2011, 18:01 (by Faris knight) |
| Edits | 6059 | 2741 | 1269 | 99 | 961 | 315 | 448 | 8 |
| Average edits per day | 134.6 | 60.9 | 28.2 | 2.2 | 21.4 | 7.0 | 10.0 | 0.2 |
| Editors | 1190 | 282 | 236 | 24 | 264 | 79 | 35 | 5 |
| Edits by the top 10% of active users | 4,406 (72.80%) | 2,193 (80.01%) | 895 (70.53%) | 42 (42.42%) | 572 (59.52%) | 166 (52.70%) | 347 (77.46%) | 2 (25.00%) |
| Most active editors | The Egyptian Liberal (559) | Lihaas (375) | Ahmed m rabea (185) | The Egyptian Liberal (18) | In dubio pro dubio (149) | In dubio pro dubio (57) | Ghaly (168) | The Egyptian Liberal (2) |
| | Lihaas (486) | The Egyptian Liberal (332) | خلدون شنتوت (174) | Osa osa 5 (14) | Goldzahn (63) | Generator (30) | Samsam2 2 (96) | Samsam22 (2) |
| | Ocaasi (275) | Ocaasi (314) | Moh2010 med (143) | Mohamed Ouda(c) (10) | WikitanvirBot (56) | Goldzahn (22) | WikitanvirBot (62) | Faris knight (2) |
| | 94.246.150.68 (196) | 94.246.150.68 (263) | WikitanvirBot (64) | 62.220.33.64 (9) | Luckas-bot (23) | BangertNo (14) | Luckas-bot (21) | Ghaly (1) |
| | 108.14.100.42 (187) | Silver seren (144) | Osa osa 5 (43) | Moh2010 med (8) | Wikifreund (23) | Tfjt (13) | EmausBot (20) | Egy Observer (1) |
| | Wipsenade (168) | Wipsenade (118) | Mohamed ElGedawy (29) | خلدون شنتوت (7) | A.Abdel-Rahim (21) | CopperBot (11) | 62.220.33.64 (14) | |
| | Ericoides (167) | Knowledge kid87 (75) | The Egyptian Liberal (27) | أسامة عباس (5) | Dinarsad (21) | Mr. Mustard (10) | Eskandara ny (8) | |
| | Aude (143) | Cs32en (69) | Luckas-bot (24) | غلام الأسمر (5) | Generator (20) | 188.174.2.207 (9) | The Egyptian Liberal (5) | |
| Anonymous edits | 1,336 (22.08%) | 426 (15.54%) | 202 (15.92%) | 11 (11.11%) | 148 (15.40%) | 73 (23.17%) | 29 (6.47%) | 0 (0.00%) |

time). During the first 45 days (from 25 January to 10 March 2011¹), it received 6059 edits by 1190 different users, with an average of 135 edits each day. In the same period, the associated

talk page received 2741 edits by 282 different users, with an average of 61 edits per day [13].

The articles about the main happenings of the 2011 Egyptian revolution were soon organized by Wikipedia users in the category “Egyptian Revolution of 2011” [14], containing 21 pages, such as “Domestic responses to the Egyptian Revolution of 2011”, “Timeline of the 2011 Egyptian revolution” or “April 6 Youth Movement”. It also comprises another subcategory, named “People of the 2011 Egyptian Revolution”, which consists of 36 pages.

¹ Wikipedia is an ongoing project and continuously changes, so the empirical data we provide in this section have limited validity in time and become obsolete very soon. Therefore, the data provided, retrieved at 10 March 2011, should be considered to get a general awareness of the richness of the phenomena described.

Table 2. The 20 most edited pages of the month on English Wikipedia, at 4 March 2011 [21]

| | |
|--|--|
| 1. 2011 Libyan uprising | 11. 2011 Egyptian protests |
| 2. 2011 Libyan protests | 12. 2010-2011 pro-democracy protests |
| 3. Deaths in 2011 | 13. 2010-2011 Pro-democracy protests |
| 4. 2011 Christchurch earthquake | 14. 2010-2011 North Africa and Western Asia protests |
| 5. 2010-2011 Arab world protests | 15. 2010-2011 Pro-democracy protests |
| 6. 2010-2011 Middle East and North Africa protests | 16. 2011 Libyan Revolt |
| 7. The Undertaker | 17. 2011 Canterbury earthquake |
| 8. Egyptian Revolution of 2011 | 18. Gary Moore |
| 9. 2011 Egyptian revolution | 19. 2011 Wisconsin protests |
| 10. Watson (artificial intelligence software) | 20. Libyan Revolution |

Moreover, the access logs of Wikipedia show that the “2011 Egyptian revolution” article was accessed at least 900,000 times during the month of February 2011 [15]. This confirms once again the popularity of Wikipedia for online consultation and information gathering. Indeed, a large share of people build their interpretation of many current controversial events also on the basis of information provided by Wikipedia articles. However, even though the “passive” act of reading a Wikipedia article has been suggested as a valuable form of participation [16], in this paper we focus on the “active” participation through edits to the articles and associated talk pages, as a sign of collective memory building activities.

In fact, the Wikipedia platform records every contribution to every article or talk page, along with the author and the timestamp of the edit, allowing to effectively keep track of the entire history of the pages. In this way, Wikipedia offers unique insights into the temporal flow of collaborative content building: the possibility to empirically access and analyze almost in real time the activity related to the formation of collective memory by Wikipedia users is an exceptional opportunity for research.

The descriptive data mentioned above on the considerable amount of edit activity to the pages related to the 2011 Egyptian revolution refer to the English Wikipedia. However, the article “2011 Egyptian revolution” links to 49 related articles on Wikipedias in 49 different languages. Table 1 reports the number of edits to the articles and talk pages about the Egyptian revolution for the most active Wikipedias, precisely the English, Arabic, German, and Egyptian Arabic Wikipedias.

Some of the data provided in this paper were extracted employing different web tools that allow to process information about the activity registered in Wikipedia [13, 15, 17, 18]. Other data, particularly those related to the evolution of edit activity on pages about traumatic events, were collected using open source scripts

which process the revision history of Wikipedia pages contained in the XML files provided by the Wikimedia Foundation [19]².

Data provided in Table 1 allow one to appreciate the timeliness with which the articles and talk pages related to the 2011 Egyptian revolution were created after the beginning of the very first protests. For instance, the articles on the English and Arabic Wikipedias were created the same day the demonstrations started to arise in the streets.

For all the reported Wikipedias, the number of edits per day highlights a considerable amount of activity. Considering the English Wikipedia, the article “2011 Egyptian revolution” registered on average 135 edits per day. To this regard, it should be noted that the average number of edits per day for a random Wikipedia article is largely smaller than 1, since in fact most pages receive no edits on most days (for more detailed statistics on the number of edits of Wikipedia pages see [19]).

Moreover, the relative differences in the amount of edits between the English Wikipedia and the others are not surprising, given that the Wikipedia in English was the first language edition to be launched in 2001, and currently has the largest community of contributors. Indeed, while the English Wikipedia has reached more than 14 millions, the Arabic one counts around 360 thousands registered users, and the Egyptian-Arabic one just around 17 thousands. Interestingly, even if the Egyptian-Arabic Wikipedia has not one of the most active communities, its article about the 2011 Egyptian revolution received a very large amount of edits, clearly indicating the high relevance of the related events for its contributors.

Table 1 also reports the number of different editors involved in the collaborative building of a shared representation of the events related to the Egyptian revolution. In particular we detected a high participation on the English, Arabic, German and Egyptian-Arabic Wikipedias. The number of edits made by the top 10% of active contributors suggests that even if some editors participate more than others in article building activities, in general the participation seems to be rather equally distributed among users. To this regard, it is relevant to note that Wikipedia articles can be edited by anyone, and Wikipedia policies expressively warn users to not consider themselves as the owners of a page.

One of the most prolific editors of the “2011 Egyptian revolution” article was a user nicknamed “The Egyptian Liberal” (Table 1), who also created the English version of the article on 25 January 2011. On his userpage [20] he describes himself as a native speaker of both Arabic and English, irreligious and “an advocate of democracy”. He supports the independence of Palestine and states that “there are 2000000 reasons why the 2003 invasion of Iraq was wrong”. Table 1 shows that “The Egyptian Liberal” was one of the most active editors in the English, Arabic and Egyptian-Arabic Wikipedias, especially in the talk pages associated to the articles about the 2011 Egyptian revolution. His active participation to the discussion pages in different languages suggests a strong engagement in coordinating and interacting with others about the construction of the articles and in the negotiation of different perspectives about the protests.

² The scripts are released as open source at <http://github.com/phaul/wiki-network/> and Wikipedia dumps can be downloaded at <http://dumps.wikimedia.org/>

Besides “The Egyptian Liberal” there seem not to be much overlapping among the most prolific editors in the different Wikipedias, suggesting that distinct communities might develop different shared representations of the 2011 Egyptian revolution. Interestingly, this opens the way to cross-cultural research on the construction of different collective memories about the recent controversial events [21].

While in this paper we focus mainly on the activity on Wikipedia related to the 2011 Egyptian revolution, very similar patterns can be observed also for the pages on other recent revolutions and protests in North Africa, which were promptly grouped into a category named “2010–2011 Arab world protests” [22]. At 10 March 2011, this category contained eight subcategories, including the ones related to the Algerian, Egyptian, Libyan, Tunisian and Yemeni protests.

For instance, at 4 March 2011, 16 out of the top 20 most edited pages on the English Wikipedia during the previous month were about traumatic events, most of which related to the Middle East and North Africa protests (Table 2). Moreover, it is worth to observe that two of the most edited pages are related to the 2011 Christchurch earthquake and the 2011 Canterbury earthquake, two different articles which describe the same earthquake occurred in New Zealand on 22 February 2011. This suggests that collective memory activities can possibly take place also for events, which are different from the North African revolutions, but inherently traumatic for other reasons.

This section described the richness of activities related to collective memory building on Wikipedia. Clearly, Wikipedia is an ongoing community in continuous development, and the previous empirical data have limited validity in time, so this section should be considered as a snapshot of the richness of the phenomenon. The next section will highlight possible research directions aimed at exploring these data with automated or semi-automated analysis, in order to shed light on how we, as a society, construct our collective memories in the new digital era.

3. DIRECTIONS ON EMPIRICAL RESEARCH ON COLLECTIVE MEMORY IN WIKIPEDIA

Wikipedia provides several behavioral guidelines for users. Among these, the Neutral Point of View introduced above [11] consists in presenting all the significant perspectives published by

reliable sources. An essay on Wikipedia warns against “recentism”, which consists in contributing to an article without a long-term historical view [23]. Moreover, being an encyclopaedia, Wikipedia states that, while most newspapers focus mostly on current breaking news and sometimes can disseminate the opinion of their authors, its content should be based on reliable sources and written from a neutral perspective [24]. Moreover, an official policy explains that Wikipedia should not offer first-hand news reports on breaking stories [25].

However, despite these internal norms and policies, research showed that the 77% of articles about a sample of recent traumatic events are created in at most two days after the related events [19]. This is true for the article on the 7 July 2005 London bombings, which was created only few minutes after the first bombings, at 09:18 of 7 July 2005. But this is also true for the current protests in Egypt (see Table 1), Libya, etc. In fact, in the case of events which shake the core of a community's identity, like revolutions, it is difficult to stop thousands of people from actively participating in the social sharing and in the collective representation of current happenings.

As we already mentioned, one of the most important policies of Wikipedia is the Neutral Point of View (NPOV). However, especially in the discussion pages associated to articles about controversial issues, we can often observe heated debates. These debates can sometimes end in edit wars, when users repeatedly reverse the previous user's edits trying to make their own version of the article visible [26]. For this reason, following Pentzold's analysis [8] we believe that the first period after a traumatic event represent the beginning of the collective memory building process. During these initial moments, the collective memory is still interactive, disorganized and unstable. In other words, it is still communicative [8, 27], meaning that a globally agreed perspective is still far to be reached, and different community members advance diverse points of view. Indeed, during the initial protests of the 2011 Egyptian revolution, Wikipedia was also used to advance diverging positions, as the comments on the talk page reported below suggest. However, as the events are elaborated and the strong emotional reactions to the traumatic happenings gradually soften, it seems likely that a more neutral point of view will eventually be reached. At the same time, the number of direct contributions to the articles will tend to decrease, while a substantial part of them will be minor edits or indirect work for maintenance purposes [26]. In this way, one day the community will be able to agree on a common narrative of the

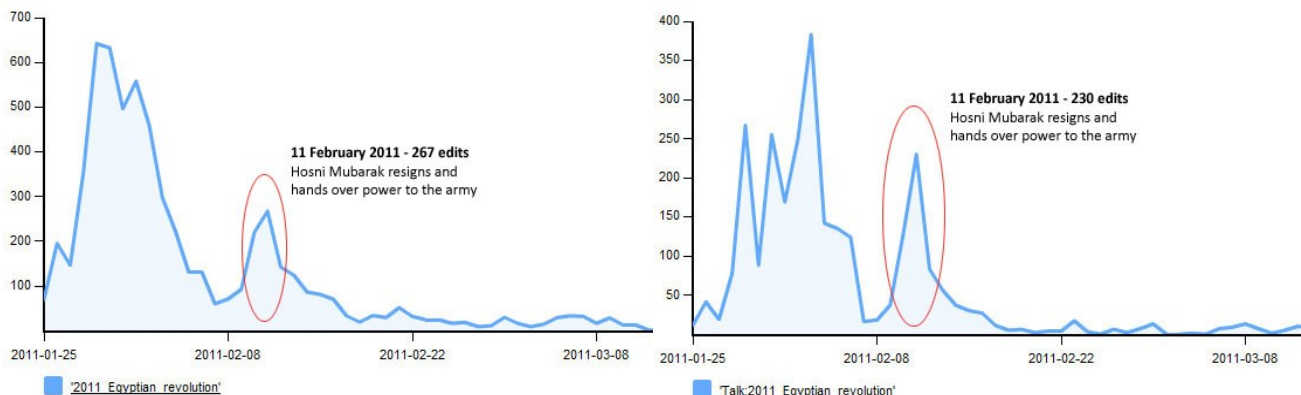


Figure 2. Number of edits per day to the “2011 Egyptian revolution” article and talk page on English Wikipedia [18].

past, allowing the collective memory to gradually reach the cultural level, becoming more stable and objectified [8, 27].

From this perspective, a first level of analysis we have pursued focuses just on the amount of edits over time [19]. For instance, Figure 2 shows the magnitude of edits in time for the “2011 Egyptian revolution” article and its talk page on the English Wikipedia. It is possible to observe that during the first days edit activity is extremely high, with an average of almost 524 edits per day from 28 January to 2 February 2011 in the article page, and about 235 edits per day from 29 January to 3 February 2011 in the talk page [18]. Furthermore, it is interesting to note that in both the article and the talk page edit activity increased on 11 February 2011, when the Egyptian President Hosni Mubarak resigned. A similar increase of activity can be observed for the pages about the Tunisian revolution on the English Wikipedia, where the number of edits to the article raised from 88 on 14 January to 192 on 15 January 2011, when Saudi Arabia officially announced to be hosting the former Tunisian president Zine El Abidine Ben Ali. In a similar way, discussions increased on the talk pages,

which went from 5 edits on 13 January to 23 and 33 respectively on 14 January and 15 January 2011 [18].

Our previous research showed that, during anniversaries, edit activity increases significantly both for articles and talk pages related to traumatic events [19]. These spikes of edit activity during anniversaries can be interpreted as a sign of commemoration processes, whose importance for memory building has been described in terms of memory retelling and interpretation of the past through the negotiation of different narratives and perspectives [28].

This first level of analysis focuses on the timing of creation of pages about traumatic events, usually within days after the first happenings, on patterns of edits over time, such as the increase in edit activity during the first weeks, at anniversaries, or corresponding to particularly meaningful facts. However, a second level of analysis can fruitfully take into account also the interactions between the contributors to the pages.

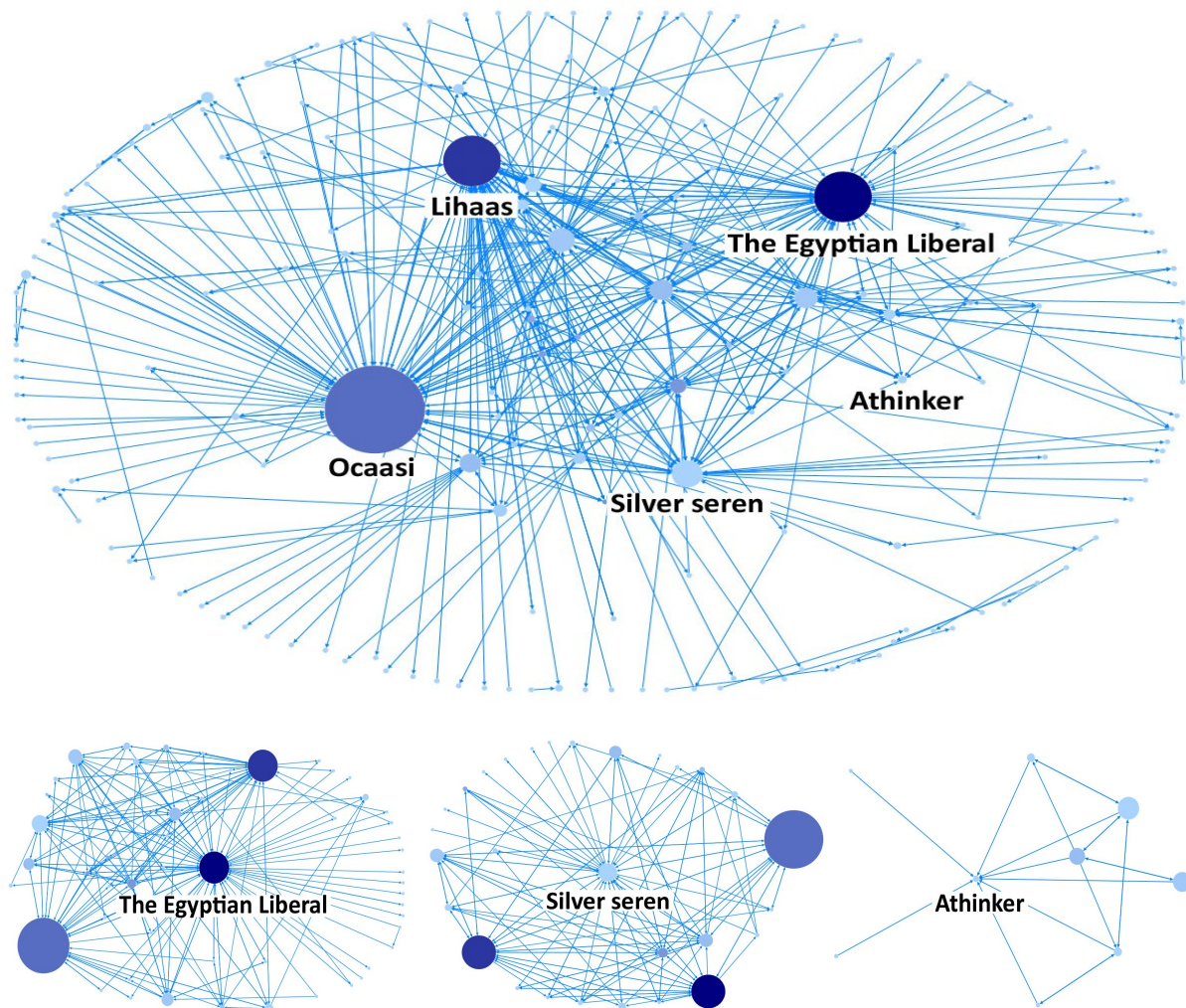


Figure 3. Discussion network of the talk page associated to the "2011 Egyptian revolution" article. In the network graph, which is directed and weighted, nodes are users and edges represent direct replies between users in the talk page. Nodes attributes are size, which represents the value of actor's betweenness centrality, and color, representing the number of direct edits to the associated article. So, a large and dark node is very central in the network and made many contributions to the article page. Below, the egonetworks of three particular users are reported, obtained by including the ego, his/her adjacent nodes and the edges between these nodes. The graphical representation of the network has been obtained using NodeXL [29].

As already mentioned, the process of collective memory building is the result of a complex ecology of discussions on different perspectives. The excerpt reported below gives an idea of such negotiations of different interpretations between users. The discussion took place during the second day of protests on the talk page of “2011 Egyptian revolution” in the English Wikipedia. In the example, the first user challenges the neutrality of the current article and an anonymous user replies accusing him to be a “defender of the regime”.

This article is not taking a neutral stance. I believe it is forged by anti-regime contributors who are taking the side of the protesters by any means available. I'm neither contributing nor will I contribute to the article, but for those who are doing so, have some Wikipedian ethics. Thank you. [[User:MagedMahfouz|Maged Mahfouz]] [[User talk:MagedMahfouz|talk]] 16:33, 26 January 2011 (UTC)

Well like it or not Maged, when a people rise up and overthrow a despotic regime, the facts of the matter are what they are. One is not anti-fascist by reporting on the Third Reich or the Italian New Order. Nor is there any “forgery”. Your input as a defender of the regime will definitely help the article be a better report of the objective facts, but on the basis of the evidence and assuming no one is ready to provide such support, I suggest the NPOV tag be removed. [[Special:Contributions/72.228.177.92|72.228.177.92]] [[User talk:72.228.177.92|talk]] 19:06, 26 January 2011 (UTC)

To this regard, Wikipedia is an excellent source of data for Social Network Analysis (SNA), providing large datasets in different languages. Offering many different types of pages and diverse ways for users to interact, Wikipedia also provides a social context which is more complex compared to many other communication platforms, such as for example chat, email or even Twitter. The richness and the diversity of possible interactions in Wikipedia allows researchers to represent these relations in a variety of different network graphs.

For instance, conversations occurred in the talk page of “2011 Egyptian revolution” can be analyzed from a SNA perspective, interpreting the communicative interactions between users as ties between nodes in a network.

Figure 3 shows the network of discussions happened on the talk page of “2011 Egyptian revolution”. In the network graph, nodes are Wikipedia users and a directed edge is present from user A to B if A has written at least one comment indented under an entry by user B [30]. The size of nodes in Figure 3 depends on their betweenness centrality, a typical SNA measure representing the relative importance of each user in the network, and based on the number of shortest paths from all vertices to all others passing through a particular node [31]. Betweenness centrality can be interpreted as the potential of a node for control over the information flow passing between two other nodes [32]. The color of the nodes represents the number of edits the users have performed on the related article page and is an indicator of participation in the direct creation of the collective representation of the event. In other words, a large and dark node represents a user who is both very central in the network and contributed heavily to the editing of the related article.

It is possible to observe that the three largest nodes, which are the most central ones in the network according to their betweenness centrality, are also the ones who performed most of the edits (see also Table 1). An initial observation of the network seems to suggest that heavily involved users who deeply care about the topic tend to be very active both on the article, with direct edits, but also on the associated talk page, participating in discussions in order to reach consensus (for example, “The Egyptian Liberal”, “Lihaas”, and “Ocaasi” in Figure 3). However, an interesting exception can be easily identified just looking at the visual representation of the network: the user “Silver seren” is the forth based on betweenness centrality score but is very light in color, meaning that he did not edit the article very much. Indeed, data show that this user edited the content page just 5 times, but instead contributed to the associated talk page 144 times (see second column of Table 1). This apparently unusual behavior is explained by the user himself, who describes his social role [33] in some of the comments left on the talk page:

I don't think i'm up for article writing right now. Especially not on such an extensive one such as this. But I am definitely up for reference finding. Just let me know what statements you need to have referenced or the types of references that you need and i'll find them for you. Silverseren 08:47, 31 January 2011 (UTC)

Cool. Add it to the article -- The Egyptian Liberal (talk) 20:07, 2 February 2011 (UTC)

No editing from me, that's up to you guys. I'm just the reference guy. :P Silverseren 20:11, 2 February 2011 (UTC)

In Figure 3, we highlighted another user nicknamed “Athinker” because s/he plays another interesting social role [33]. This user started 9 different threads on the talk page with opinionated statements but never took part in the discussions which were originated. In fact, “Athinker” is characterized by a marked difference between his/her indegree of 9, which can be interpreted as the number of direct replies s/he received, and his/her outdegree of 0, meaning that s/he did not replied to anyone in the network (indeed, an edge from A to B is added when A replies to a discussion started by B). This also explains his/her small betweenness centrality. Moreover, “Athinker” never edited the article page. Figure 3 also includes the ego networks of “The Egyptian Liberal”, “Silver seren”, and “Athinker”, obtained by including the ego, his/her adjacent nodes and the edges between these nodes. Ego networks have been used in past research for detecting structural signatures to identify social roles [33, 34].

As mentioned above, different kinds of network graphs can be extracted from Wikipedia, and one such example is represented by networks of User talk pages, which are Wikipedia pages devoted to direct communication with a specific user [30, 33, 34, 35]. In this directed and weighted networks, nodes are users, and an edge is drawn from A to B when A leaves a comment on the User talk page of B. Following we report two messages written by “The Egyptian Liberal” on the User talk page of “Lihaas”. Differently from discussions occurring on article talk pages, which are in general related to the associated content article, these direct conversations are more personal [35].

Hopefully, Mubarak will leave before the article get bigger (for us) and more ppl die (for the Egyptians) -- The Egyptian Liberal (talk) 13:48, 3 February 2011 (UTC)

OK. Now I am pissed. I saw Omar and Mubarak's Speech. They want war, they got one. Tomorrow is going to be the biggest protest the world have seen in this MILLENNIUM. I got the word right now that people are coming out with numbers that has not been seen on TV. They are also moving to the palace. Bloody fucking idiots, if he stepped down today, noone will be killed tomorrow. -- The Egyptian Liberal (talk) 21:46, 10 February 2011 (UTC)

By complementing networks of discussions occurring on article talk pages with one-to-one conversations occurring on User talk pages, it is possible to get a more complete picture of the dynamics involved in the creation of the collective memories of events such as the North African uprisings.

So far we have extracted networks where edges represent a communicative interaction between two users, but a deeper understanding would be reached by considering what is exchanged in these edges and by assigning at least a binary classification into positive or negative interaction. In fact, a recent paper [36] proposed algorithms for extracting a signed network from Wikipedia edit activity considering edits over commonly-authored articles, other activities such as votes for adminship, the restoring of an article to a previous version, or the assignment of barnstars (a prize acknowledging valuable contributions). Indeed, an edit on an article talk page or User talk page can represent alignment and positive feedback as well as disagreement and negative feedback. Zhu et al. [37] exploited machine learning models to automatically identify four types of leadership behaviors (positive, negative, directive and social) from 4 million messages sent between Wikipedia users, finding that editors in peripheral roles produce a large proportion of leadership behaviors. Welsch et al. [34] instead performed a manual coding to classified users as confrontational, deliberative and neutral.

This short survey shows that a further level of analysis can provide an even more complete picture of the collective memory building processes occurring on Wikipedia. Not only occurrences of edits in time and users' interactions are taken into account, but also the actual textual content of users' communicative exchanges. As introduced briefly in the previous paragraph, different techniques are possible and have been recently proposed, ranging from manual coding [34, 8] to automatic natural language processing technologies augmented with machine learning [37].

With this regard, we are currently implementing automated content analysis tools to explore the patterns of language used in Wikipedia pages related to traumatic events. Cohn and colleagues [7] used the text-analysis program Linguistic Inquiry and Word Count [LIWC; 38] to analyze blog posts for two months prior to and after the 11 September 2001 attacks. They found signs of pronounced psychological changes in the language used by bloggers, mainly related to an increase of words associated to negative emotions, cognitive processing, social engagement and psychological distancing during the first days following the attacks.

Our aim is to adapt this methodology and extend it to a larger sample of Wikipedia pages about traumatic events, whose collective memory is built over time by thousands of editors. To

this end we have developed PYWC, a software tool similar to LIWC, and released it as open source³. We are currently analyzing if there are statistically significant differences in the use of language associated to various psychological categories, such as words related to positive and negative emotions, between pages about traumatic events and other Wikipedia pages. We are also interested in the longitudinal study of these indexes in pages related to traumatic events, our hypothesis being that this is another way to detect the gradual transformation from communicative to cultural memory.

The last research direction we briefly highlight in this contribution is related to cross-cultural studies. Even though there are Wikipedias written in more than 270 different languages, and many of them are built by very active communities, only few studies compared them [for example, 39]. For example it would be interesting to analyze how two or more different language communities form their memories about the same or similar events. To this regard, we developed Manypedia [21], a web tool whose goal is to allow to explore this empirical question by comparing two different versions of the same page, as it appears in two different Wikipedias. Currently Manypedia automatically translates both pages into one among 56 possible languages and adds some descriptive statistics on top of the pages, such as those presented in Table 1 concerning the “2011 Egyptian revolution” article. Our future work will also be aimed at integrating Manypedia with additional information extracted with PYWC, such as, for example, the levels of different emotions (anger, anxiety, happiness, ...) expressed in the text of the pages under comparison.

As mentioned above, automated text analysis can be implemented in the study of the shared representations that different language communities form about similar events. For instance, it would be interesting to compare the articles about the 2005 London and 2006 Madrid bombings in the English and Spanish Wikipedias. LIWC provides dictionaries in seven languages, allowing to compare the articles and talk pages about the same events on different Wikipedias, without having to rely on automatic translation techniques which often reduce the richness of the text.

In this paper we presented evidence showing that the collective memory of the recent North African revolutions and uprisings is indeed built through the collaboration, and sometimes conflicting interaction, of users in Wikipedia. Our goal in this contribution was to suggest and provide evidence that Wikipedia is a global memory place in which the processes of negotiation of different narratives can be studied empirically. We believe Wikipedia offers an unprecedented opportunity for researchers for studying how we, as a society, build our cultural representations of the past. Through the analysis of Wikipedia edits and discussions, also employing SNA and natural language processing techniques, it is possible to study this crucial phenomenon in real-time and on a large scale.

4. ACKNOWLEDGMENTS

We would like to thank David Laniado and Riccardo Tasso for providing the discussion network of the talk page related to the 2011 Egyptian revolution. This work has been realized also thanks to the support from the Provincia autonoma di Trento and

³ The software can be downloaded at <https://github.com/phaully/wiki-network/blob/master/pywc.py>

the Fondazione Cassa di Risparmio di Trento e Rovereto, in the frame of the project “Active Digital Memories of Collective Life”.

5. REFERENCES

- [1] Sztompka, P. 2000. The trauma of social change. In *Cultural Trauma and Collective Identity*, A. J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, and P. Sztompka, Eds. University of California Press, Berkeley.
- [2] Alexander, J. C. 2004. Toward a theory of cultural traumas. In *Cultural Trauma and Collective Identity*, A. J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, and P. Sztompka, Eds. University of California Press, Berkeley.
- [3] Eyerman, R. 2004. Cultural Trauma: Slavery and the Formation of African American Identity. In *Cultural Trauma and Collective Identity*, A. J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, and P. Sztompka, Eds. University of California Press, Berkeley.
- [4] Neal, A. 1998. *National Trauma and Collective Memory: Major Events in the American Century*. M.E. Sharpe, Armonk, New York.
- [5] Olick, J. K., and Levy, D. 1997. Collective memory and cultural constraint: Holocaust myth and rationality in German politics. *American Sociological Review*. 62, 6, 921-936.
- [6] Castells, M. 2001. *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press, Inc., New York, NY.
- [7] Cohn, M. A., Mehl, M. R., and Pennebaker, J. W. 2004. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*. 15, 10, 687-693.
- [8] Pentzold, C. 2009. Fixing the floating gap: the online encyclopaedia Wikipedia as a global memory place. *Memory Studies*. 2, 2, 255-272.
- [9] Forte, A., and Bruckman, A. 2005. Why do people write for Wikipedia? Incentives to contribute to open-content publishing. In *GROUP 05 workshop: Sustaining community: The role and design of incentive mechanisms in online systems*. Sanibel Island, FL.
- [10] Beaumont, P. 2011. The truth about Twitter, Facebook and the uprisings in the Arab world. *The Guardian*. Friday 25 February 2011
- [11] Wikipedia.org, Wikipedia:Neutral point of view. <http://en.wikipedia.org/wiki/Wikipedia:NPOV>
- [12] Nora, P. 1996. From lieux de mémoire to realms of memory. In *Realms of memory. Rethinking the French Past, Vol. 1: Conflicts and divisions*, P. Nora, Ed. Columbia University Press, New York.
- [13] Toolserver.org, Article Info. <http://toolserver.org/~soxred93/articleinfo/>
- [14] Wikipedia.org, Category:Egyptian Revolution of 2011. http://en.wikipedia.org/wiki/Category:Egyptian_Revolution_of_2011
- [15] Stats.grok.se, Wikipedia Article Traffic Statistics. <http://stats.grok.se/>
- [16] Antin, J., and Chesire, C. 2010. Readers are not free-riders: reading as a form of participation on Wikipedia. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*. ACM, New York, NY, 127-130.
- [17] Wikirage.com, <http://www.wikirage.com/>
- [18] Nunes, S., Ribeiro, C., and David, G. WikiChanges: exposing Wikipedia revision activity. In *Proceedings of Int. Sym. Wikis*. 2008.
- [19] Ferron, M., and Massa, P. 2011. Studying collective memories in Wikipedia. *3rd Digital Memories Conference* (Prague, Czech Republic, March 14 - 16, 2011). Available at <http://www.inter-disciplinary.net/critical-issues/cyber/digital-memories/conference-programme-abstracts-and-papers/>
- [20] Wikipedia.org, User:The Egyptian Liberal. http://en.wikipedia.org/wiki/User:The_Egyptian_Liberal
- [21] Massa, P., and Scrinzi, F. 2011. Exploring Linguistic Points of View of Wikipedia. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. ACM, New York, NY.
- [22] Wikipedia.org, 2010-2011 Arab world protests. http://en.wikipedia.org/wiki/2010-2011_Arab_world_protests
- [23] Wikipedia.org, Wikipedia:Recentism. <http://en.wikipedia.org/wiki/Wikipedia:Recentism>
- [24] Wikipedia.org, Wikipedia:Wikipedia is not a newspaper. http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_not_a_newspaper
- [25] Wikipedia.org, Wikipedia:What Wikipedia is not. <http://en.wikipedia.org/wiki/Wikipedia:NOT#JOURNALISM>
- [26] Kittur, A., Suh, B., Pendleton, B. A., and Chi., E. 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the ACM Conference on Human-factors in Computing Systems*. CHI '07. ACM, New York, NY.
- [27] Assmann, J. 1995. Collective memory and cultural identity. *New German Critique*. 65, 125-33.
- [28] Zerubavel, Y. 1995. *Recovered roots: collective memory and the making of Israeli national tradition*. University of Chicago press, Chicago.
- [29] Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., and Dunne, C., 2010. *NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010*, <http://nodexl.codeplex.com/> from the Social Media Research Foundation, <http://www.smrfoundation.org>
- [30] Laniado, D., Tasso, R., Volkovich, Y., and Kaltenbrunner, A. 2011. When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages. *ICWSM 2011 - 5th International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.
- [31] Wasserman, S., and Faust, K. 1994. *Social network analysis. Methods and applications*. Cambridge University Press, Cambridge, MA.
- [32] Freeman, L. C. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry*. 40, 1, 35-41.

- [33] Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. 2011. Finding social roles in Wikipedia. In *Proceedings of the ACM iConference* (Seattle, WA). ACM, New York, NY.
- [34] Welser, H. T., Underwood, P., Cosley, D., Hansen, D., and Black, L. 2010. Wiki-networks: Connections of creativity and collaboration. In *Analyzing Social Media Networks with NodeXL*, D. Hansen, B. Shneiderman, and M. Smith, Eds. Morgan Kaufmann, New York.
- [35] Massa, P. 2011. Social networks of Wikipedia. Under review at *ACM Hypertext 2011 Conference*.
- [36] Maniu, S., Cautis, B., and Abdessalem, T. 2011. Building a signed network from interactions in Wikipedia. In *Databases and Social Networks (DBSocial '11)*. ACM, New York, NY.
- [37] Zhu, H., Kraut, R. E., Wang, Y., and Kittur, A. 2011. Identifying shared leadership in Wikipedia. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*. ACM, New York, NY.
- [38] Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. *Linguistic inquiry and word count (LIWC): LIWC 2001*. Erlbaum: Mahwah, NJ.
- [39] Hara, N., Shachaf, P. and Hew, K. F. 2010. Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology*. 61, 2097–2108.