Detecting Platform Effects
in Online Discussions

Pablo Aragón[*†]   Vicenç Gómez[*]   Andreas Kaltenbrunner[†]

[*] Universitat Pompeu Fabra   [†] Eurecat
Barcelona, Spain   Barcelona, Spain

Abstract. Online discussions are the essence of many social platforms on the Internet. These platforms are receiving increasing interest because of their potential to become deliberative spaces. Many studies have proposed approaches to measure online deliberation and to evaluate which are the best design principles for deliberative online platforms. However, little research has focused on how deliberation in online platforms is affected by the arrival of events like the emergence of new topics or the modification of platform features.

In this article we present a methodology to detect events that affect online deliberation in online discussions. Our results on Menéame, the most popular Spanish social news site, show that a change in how discussions are shown to the user, from a linear to a hierarchical conversation view, significantly enhanced deliberation. In particular we observe that this type of interface induced argumentative structures of online discussion.

1   Introduction

Nowadays, millions of citizens interchange messages in online discussion platforms. A large part of these discussions are related to political talk which is attracting an increasing number of citizens to go online to engage in political processes (Bimber, 2003; Chadwick, 2006). This might be explained by the potential of the Internet to create a new public space for political discussion (Papacharissi, 2002). Thus, information and communication technologies have been noted to facilitate the participation of citizens in democratic communication (Hacker and Van Dijk, 2000) and, ultimately, the construction of an online deliberative public sphere (Dahlberg, 2011). Public sphere, seen as "a society engaged in critical public debate" (Habermas, 1991), is the essence of deliberative processes. Although the definition and implications of deliberation are far from conclusive (Carpini et al., 2004), the many approaches to deliberation, e.g. Barber (1984); Habermas (1996); Elster (1998); Fishkin (1997); Chambers (2003), have all a common denominator: the relevance of communication in detriment of direct voting. In

this regard, public sphere and deliberation are influential concepts in the relationship of democracy and information communication technologies (Chadwick, 2008).

Despite an important fraction of research indicating the benefits of online public spheres, some other studies have adopted a more critical position regarding the potential of the Internet in facilitating deliberative processes. Early work on comparing face-to-face and online deliberation reported considerable resemblance between both types (Luskin et al., 2004; Min, 2007; Grönlund et al., 2009). However, some recent analyses have found that discussions on the Internet generate more negative emotions and, therefore, consensus is less likely to be obtained (Baek et al., 2012). The lack of consensus is commonly associated with scenarios of group polarization, which commonly occur in online discussion platforms (Van Alstyne et al., 1996; Sunstein, 2001, 2002). In this regard, uncivil attitudes in online discussions, which are contrary to deliberation by rational-critical discourse (Habermas, 1984), have been proven to play a major role in promoting polarized scenarios (Anderson et al., 2014). Given that interactions between individuals are not always civil and rational (Wilhelm, 2000), some researchers have concluded that discussions on the Internet do not necessarily lead to online deliberation (Dahlgren, 2005).

The contradictions between findings from online deliberation analyses have motivated the examination of which features in online platforms might affect their deliberative potential. A feature that has received large attention is the moderation of messages. Although online moderation can be seen as a form of censorship and a threat to freedom of speech, some studies have defended that moderation by skilled users is a relevant feature to promote deliberation (Coleman and Gotze, 2001; Edwards, 2002; Wright and Street, 2007; Wright, 2009). Another feature of interest is the anonymity of users. On the one hand, this feature is likely to improve online discussions because users feel no pressure of conventional cultural cues (Kim, 2006; Bowker and Tuffin, 2003). On the other hand, Friedman et al. (2000) indicated that this lack of pressure is precisely the reason behind the emergence of uncivil and non rational attitudes. The type of discussion, i.e. synchronous (e.g. chats) versus asynchronous (e.g. online forums), has also been examined and results often indicate that asynchronous discussions better promote deliberation (Janssen and Kies, 2005; Smith et al., 2009). Finally, online deliberation might be also conditioned by the topics under discussion. Gonzalez-Bailon et al. (2010) found that in Slashdot, an online discussion platform which has been defined as "a form of online public sphere" (Poor, 2005), political discussions were much more deliberative than discussions of other topics, e.g. online gaming.

In general, most studies of online deliberation have examined one or a few features in one single online platform and, therefore, results are limited to individual characteristics of the online community and the platform itself. This research gap has been recently addressed in Esau et al. (2016), which compares different technical features (e.g. moderation, synchrony of discussions) in a news forum, three news websites, and Facebook news pages. Their results show that

while moderation has a positive effect on online deliberation, this was not found for asynchronous discussions. Esau et al. (2016), as many others, is focused on a subset of potentially relevant features while others are not considered, e.g. anonymity. More importantly, there could be events at a specific moment in time which produce durable effects on deliberation, e.g. the deployment/change of technical features or the emergence of new topics. Given that, to our best knowledge, previous research on online deliberation has not considered the effect of events of this nature, we aim to answer the following research question:

– *Is it possible to automatically detect events which affect online deliberation in online discussions?*

To answer this question, we have chosen an online discussion platform that we suspect to have been affected by specific events. In particular, we have collected the discussion threads over five years from Menéame, the most popular Spanish social news site. Two candidate events to have affected online deliberation are:

– **E1: The 15M movement.**
  In May 2011, the 15M grassroots movement (also known as the Indignados movement) occupied the main squares of the largest cities of Spain in order to advocate for a *real* democracy. This movement has made a significant impact on Spanish politics. For instance, grassroots parties which emerged from the 15M movement (Aragón et al., 2016b), like Barcelona en Comú, Ahora Madrid and Zaragoza en Común, are currently ruling the city councils of many of these cities. Toret et al. (2015) found that the origin of the 15M movement is explained by the emergence of technopolitical practices, many of which occurred in Menéame. Some other studies stated that Menéame played an essential role in the difussion of the call for the initial demonstration (Otero and Sánchez, 2012) and, furthermore, the construction of an online space that generated many of the claims and messages adopted by the 15M movement (Mena, 2011). These effects were confirmed in Postill (2013) which proved that aggregators and link recommendation sites, especially Menéame, experienced unprecedented traffic growth during the 15M movement. Therefore, although many links in early years were related to science and technology, the eruption of the 15M movement turned Menéame into one of the most relevant online discussion platforms in Spain about social and political issues.

– **E2: Change of the conversation view.**
  Since the first version of Menéame, directly inspired by Digg, many changes have taken place. Regarding features of online discussion, we highlight the change of the *conversation view*, i.e. the way in which the discussion threads are presented. The original conversation view of Menéame displayed the comments of a thread linearly in a chronological order, regardless of reply relationships. In January 2015, this design changed and, by default, messages are now displayed hierarchically following the tree structure of the discussion thread. Figure 1 shows both interfaces: a thread from 2011 about the rise

of the 15M movement[1] presented in a *linear* conversation view (Figure 1a), and a thread from 2015 about the victory of the grassroots party Barcelona en Comú in the local elections[2] presented in a *hierarchical* conversation view (Figure 1b).

These two candidate events are motivated by different reasons. For E1, different studies confirmed that Menéame played a key role in the communication dynamics of the 15M movement (Toret et al., 2015; Otero and Sánchez, 2012; Mena, 2011; Postill, 2013). This resulted in a great increase of political talk within the platform. Given that Gonzalez-Bailon et al. (2010) found that discussion threads about politics in a similar platform exhibited higher level of deliberation, the increase of political talk might have affected deliberation in Menéame. For E2, hierarchical conversation views are the typical interfaces of asynchronous discussions, which better promote deliberation (Janssen and Kies, 2005; Smith et al., 2009). Furthermore, this type of view has been proven useful to improve different components of communication, e.g. construction of knowledge (McVerry, 2007), context of the discussion (Fuks et al., 2006; Venolia and Neustaedter, 2003), and coherence (Smith et al., 2000). Given that communication is the essence of deliberative processes, the change of conversation view from linear to hierarchical might have also affected deliberation in the platform.

—Figure 1 about here —

As we detail later in Section 2, most studies on online deliberation have examined the principles of rational-critical discourse (Habermas, 1984, 1991) with a specific coding scheme, e.g. Trénel (2004); Stromer-Galley (2007); Coleman et al. (2002); Friess and Eilders (2015). Such approaches have always relied on the human examination of linguistic features. On the one hand, these classical approaches benefited from the reliability of using human coders. On the other hand, their cost is unaffordable for large datasets as the one of Menéame. For this reason, we will measure online deliberation using the computational model in Gonzalez-Bailon et al. (2010), which is based on the structural complexity of discussion threads, i.e. the more complex, the more deliberative.

We should remark that the main objective of this study is to automatically detect events that significantly produced durable effects on online deliberation. Although we suggest two candidate events, the statistical methods of this study will examine every other possible moment in time as a possible event. Therefore, our methodology is not limited to these two events.

The organization of this paper is as follows. In Section 2, we introduce previous research on the impact of events in online platforms and the measurement of online deliberation. We then describe in Section 3 the relevance of Menéame and the details of our dataset. In Section 4, we present the methods and results of our statistical analysis of events that affected the deliberative structure of

---

[1] https://www.meneame.net/story/junta-electoral-madrid-prohibe-concentracion-convocada-acampada
[2] https://www.meneame.net/story/ada-colau-gana-barcelona

discussions. Our findings are discussed in Section 5 in relation to our research question and previous literature. We finally conclude in Section 6.

2   Previous Research

We now present previous research that motivates our methodology based on platform effects (Malik and Pfeffer, 2016) and structural indicators of online deliberation (Gonzalez-Bailon et al., 2010).

2.1   Impact of events in online platforms

Previous work has examined how the activity of online platforms is affected by the arrival of different types of events. We will distinguish between events originated by (a) the emergence of new topics under discussion, and (b) the deployment of new features in the platform.

The impact of the first type of events has been analyzed in different social media platforms. In Crane and Sornette (2008), an epidemic model was defined to prove that viewing activity on YouTube can be explained by different factors, e.g. new popular topics. A later study also found that the popularity of videos can be predicted by, among other factors, the occurrence of external events; e.g. the video being massively posted in other online social networks and blogs (Figueiredo et al., 2011). The impact of similar events, defined with data from Google Trends, was also analyzed in Wikipedia, concluding that trending topics notably affect the popularity of articles (Ratkiewicz et al., 2010). Moreover, the impact of trending topics has also received much attention in the context of microblogging services, in particular, Twitter. The factors defined in Crane and Sornette (2008) were proven effective to characterize trending topics (Ratkiewicz et al., 2010). The impact of this type of events was also found to influence the usage of mechanisms on Twitter; i.e. the average number of URLs and hashtags within the tweets (Naaman et al., 2011). Finally, Lehmann et al. (2011) revealed that peaks of activity originated by trending topics can also provide a semantic characterization of the discussions.

While the above type of events mostly corresponded to new topics under discussion, activity in online platforms can be also affected by technical changes the platform itself. Indeed, the measurement of interventions in online platforms is a well-studied issue in software development through A/B testing (Kohavi and Longbotham, 2015), which requires to control the change and deployment of technical features, a condition that cannot be assumed by external researchers. This is the motivation of a recent approach for causal inference using Bayesian structural time-series models (Brodersen et al., 2015). However, this approach requires, at least, two time series: one affected by the platform intervention, and another in which the intervention made no effect in order to construct a counterfactual. If there is no access to counterfactual information, an alternative approach is the use of experimental pretest-posttest design, which expects to infer the treatment effect of an intervention through regression discontinuity

design on a time series. This has been proposed recently to measure platform effects, i.e. "the design and technical features of a given platform which constrain, distort, and shape user behavior on that platform" (Malik and Pfeffer, 2016). Thus, this methodology can be used used to detect events (the deployment of new features in the platform and also the emergence of new topics under discussion) that might affect indicators of online deliberation.

2.2   Measurement of online deliberation

The extent to which online tools enhance the quality of discussion and decision-making has attracted increasing interest from researchers and practitioners (Davies and Gangadharan, 2009). Different studies have analysed how online deliberation occurs in online discussion platforms of diverse nature, e.g. Usenet newsgroups (Wilhelm, 1998), online forums (Coleman et al., 2002), media sites (Zhou et al., 2008; Esau et al., 2016), and online social networks (Halpern and Gibbs, 2013).

To measure deliberation in online discussion platforms many different approaches have been proposed. Dahlberg (2001b) summarized the ideal requirements to facilitate online deliberation: exchange and critique of reasoned moral-practical, validity claims, reflexivity, ideal role taking, sincerity, discursive inclusion and equality, and autonomy from state and economic power. Trénel (2004) defined a coding scheme to measure the deliberativeness of online discussions using eight dimensions: equality, rationality, respect, constructiveness, interactivity, personal experience, emotional balance, and reflexiveness. These dimensions are similar to the ones from the coding scheme in Stromer-Galley (2007): reasoned opinion expression, disagreement, equality, topic, and engagement.

As one could observe, many dimensions of these methodologies are essentially features of rational-critical discourse in consistency with the conceptualization of "public sphere" defined in Habermas (1991). This observation is even explicit in other schemes, e.g. the model in Graham and Witschge (2003), and was already observed in Coleman et al. (2002): "most researchers of online deliberation have opted to use content analysis as a means of measuring the quality of discussion, operationalizing their own conceptions of what good communication looks like". However, we should note that measuring online deliberation with content analysis has always relied on the examination of online discussions by trained human coders. Therefore, these approaches are unfeasible in large datasets.

Some recent methodologies to measure online deliberation are including features that can be automatically extracted or inferred from datasets. A deliberative analysis of Wikipedia concluded that the network structures of different groups could be useful in quantifying features like equality, influence, and group member roles (Black et al., 2011). Friess and Eilders (2015) proposed an empirical model for the analysis of online deliberation using three levels: input, throughput, and outcomes. Some of the dimensions of these levels can be automatically inferred from the texts (e.g. emotional talk with computational sentiment analysis) while some other dimensions still require the intervention of human coders (e.g. civility and constructiveness).

To the best of our knowledge, the only model to automatically measure online deliberation is the one presented by Gonzalez-Bailon et al. (2010). This approach is based on a previous Madisonian model by Ackerman and Fishkin (2004) in which deliberative processes are categorized in two dimensions: representation and argumentation. The model quantifies online deliberation without examining content features. In particular, Gonzalez-Bailon et al. (2010) uses network indicators based on the network topology of online discussions, i.e. the more complex the discussion threads, the greater the level of deliberation.

3   Menéame

The analysis of this article uses data from online discussions in Menéame, the most popular Spanish social news website (130th most visited domain in Spain according to Alexa[3]). This platform, developed in 2005, includes typical features of social news sites (e.g. Digg, Slashdot, Reddit) such as social bookmarking, blogging, and web syndication. Indeed, the developers of Menéame acknowledge Digg as an inspiration of the first version of the platform[4], and aimed to provide a similar service for the Spanish blogosphere.

The functioning of Menéame is broadly as follows. Users are able to submit links to blog posts / news (hereafter *stories*) that will appear in a queue of pending stories. Then users vote and discuss each story in a discussion thread in order to promote the most interesting stories to the front page of the platform. The selection of stories for the front page is done by an open source collaborative filtering algorithm based on multiple criteria, e.g. the voting score of the story, and the reputation index of the users who have voted the story.

The collaborative nature of this platform has several social and political implications, as observed in previous studies. Triviño (2010) showed that many media outlets in Spain included a Menéame sharing widget, which proves the relevance of this platform in Spanish online media. Another study found that, although *El País* (the most visited media outlet in Spain) was the media source with most submitted stories to Menéame, other media outlets exhibited a greater impact within the platform (Malea et al., 2009). Thus, the social design of Menéame allows users to build a social and collaborative agenda-setting opposed to the notion of agenda-setting of traditional media, as noted in Mena (2011). Ferreras Rodríguez (2009) suggested that the true value of discussion and collaborative filtering in Menéame is the possibility to build a space of debate. Indeed, a later study found that 67.6 percent of users said that they use Menéame not only to read stories but also to participate in the discussion threads, while 31.5 percent of users were only interested in reading stories (Ferreras Rodríguez, 2011). The implicit social network of user interactions through comments has been investigated in Kaltenbrunner et al. (2011) and the heterogeneity of user behavior in Menéame was also analyzed in Marín (2015) from an ethnographic perspective. In particular, Marín (2015) described Menéame as a

---

[3] http://www.alexa.com/siteinfo/meneame.net (accessed February 6, 2017)
[4] https://www.meneame.net/faq-es

virtual community that has developed a particular cyber-culture based on social structures and their own code of practices. The development of this inner culture might be the result of the response of the online community to different events.

To generate the dataset of this study, we run a crawling process that collects all the stories in the front page of Menéame from 2011 to 2015 (both years included). We then perform a second crawling process to collect every comment from the discussion thread of each story. From both crawling processes we obtain 72,005 stories and 5,385,324 comments. For each of them, we keep associated metadata such as the id, url, user name, time-stamp, text message and received votes.

Finally, we should remark that messages in discussion threads of Menéame have to be posted as replies to either the story or another reply. For each message, the two conversation views of Menéame *always* indicate the id of the message being replied to (see Figure 1c). Therefore, to automatically generate the tree structure of each discussion thread, we also collect the parent id of every comment to comment.

4   Analysis

To better understand the activity in Menéame between 2011 and 2015, we first make a preliminary exploration of our dataset. Then, we present our statistical approach to detect events that have affected online deliberation in Menéame. Finally, we describe the results of the analysis.

4.1   Preliminary exploration

We first analyze the posting and voting activity. Figure 2 presents a scatter plot of the number of stories and the number of votes to stories for every day in the dataset. As one could expect, the plot shows a strong correlation between both variables ($R = 0.821$). Nevertheless, we detect that some days (red markers) exhibit an abnormally higher level of activity than the rest of the days, especially in the sum of votes to the stories posted on these days. The inspection of the corresponding stories reveals that these were prominent days in the Spanish 15M movement (details of these events are provided in the Appendix).

—Figure 2 about here —

We then explore discussion threads to better understand the typical tree structures in Menéame. We adapt an existing thread visualization tool (Aragón et al., 2016a) to examine differences in the structural properties between threads from 2011-2014 (i.e. when the conversation view was linear) and threads from 2015 (i.e. when the conversation view was hierarchical). We summarize our findings by illustrating two paradigmatic examples in Figure 3 (the two threads from Figure 1). In these visualizations, a discussion thread is represented as a radial tree in which nodes are messages and edges are the reply relationship between

messages. The root node is the initial message (the story) and comments are expanded radially to indicate their depth in the discussion thread. The size of each node is related to the number of replies and the color of the node is:

- **Black:** Root of the thread, i.e. the story.
- **Grey:** First level comments.
- **Random:** Comments to another comment. To identify authorship, we set the same random color to comments published by the same user.

Although both examples of discussions show some similar features, such as chains of two users that alternate messages (i.e. chains of nodes of two alternating colors), there are clear differences. In particular, the thread from 2011 (about the emergence of the 15M movement) contains much more direct comments to the original post than the thread from 2015 (about the victory of Barcelona en Comú). Furthermore, the thread from 2015 shows that comments attract often many replies and originate new sub-discussions, an effect that rarely occurs in the thread from 2011. Summing up, we observe that complex discussion structures are more likely when users discuss with the hierarchical conversation view.

—Figure 3 about here —

4.2 Statistical methods

The previous preliminary exploration showed evidence of the relevance of both events, the 15M movement (E1), and the change of the conversation view from linear to hierarchical (E2). To statistically detect events that affected online deliberation and to quantify their effect, we use a technique inspired by the methodology suggested in Malik and Pfeffer (2016), based on regression discontinuity design (RDD). RDD is a statistical quasi-experimental technique commonly applied in economics to evaluate the causal effects of interventions. Malik and Pfeffer (2016) proposed to define an intervention as a time-stamp in a time series (i.e. when an event occurred, hereafter the *cutoff*) and to observe the local average treatment effect on an outcome variable. Given a cutoff c, a (linear) regression is defined as:

$$Y_i = \omega_0 + \omega_1 \cdot x_i + \omega_2 \cdot \mathbf{1}(x_i > c) + \omega_3 \cdot x_i \cdot \mathbf{1}(x_i > c) + \epsilon_i,$$

where x_i is the time-stamp (bin size = seven days), Y_i the average value of the outcome variable, $\omega_{0...3}$ the coefficients of the regression, and ϵ_i a random error term. Thus, RDD fits data in two different linear regression functions, before and after the intervention, in order to measure the difference between both functions at the cutoff. The null hypothesis is that $\omega_2 \approx 0$ and $\omega_3 \approx 0$, i.e. the intervention generated no effect.

The purpose of our study is not to measure the effect of a given intervention but to detect from data when an invervention occurred, i.e. an event which significantly affected online deliberation. Therefore, instead of setting an arbitrary

cutoff (e.g. the rise of the 15M movement, the change of the conversation view), we apply an F-test, as suggested in Lee and Lemieux (2009), in every time-stamp of the time series. This approach allows us to find the most significant time-stamp based on the average values of the outcome variable before and after that cutoff.

To detect and measure events that affect online deliberation in discussion threads, our outcome variable is a metric suggested in Gonzalez-Bailon et al. (2010), which conjugates the two following tree network metrics:

- **width**: maximum number of comments at any reply level,
- **depth**: number of reply levels.

—Figure 4 about here —

To illustrate these two metrics, we present in Figure 4 an example thread using a radial tree. For this example, width = 14 (number of comments at the first level) and depth = 3. According to Gonzalez-Bailon et al. (2010), width and depth of discussion threads act as good proxies for representation and argumentation, respectively. This statement is based on the implicit assumption that users tend to follow a sequential posting behavior in discussion threads, i.e. replies explicitly indicate the message being replied to. Therefore, width approximates the number of different users involved in the discussion (to what extent the community is represented in the discussion), and depth indicates the number of messages of the longest chain of messages exchanged between users (how long argumentation lasts in the discussion). To illustrate this approach, we present in Figure 5 the four types of discussions defined in Gonzalez-Bailon et al. (2010), using real threads from our Menéame dataset:

- **Type I**: Wide and deep discussion. Associated with high levels of argumentation and representation and, therefore, deliberation.
- **Type II**: Deep but not wide discussion. Associated with high levels of argumentation but low levels of representation.
- **Type III**: Neither wide nor deep discussion. Associated with low levels of argumentation and representation.
- **Type IV**: Wide but not deep discussion. Associated with low levels of representation but high levels of argumentation.

—Figure 5 about here —

The combination of the width and depth of a thread is then measured through the **h-index** of the discussion thread. This metric, defined in Gómez et al. (2008), is inspired by the one proposed in Hirsch (2005) which assigns an h index to a researcher who has authored at least h papers with at least h citations each. In a discussion thread, the *h-index* corresponds to the maximum h level in which there are, at least, h comments, i.e. $h + 1$ is the first level in which there are less than h comments. In our illustrative radial tree from Figure 4, there are more than two comments at the second level but less than three comments at the third level, therefore, its *h-index* = 2.

4.3   Results

We first analyze whether the *h-index*, our measure for online deliberation, is affected by any event detected by our proposed method. Figure 6 (left) shows the longitudinal F-test statistic as a function of time. The best cutoff appears on 10/01/2015 and corresponds to the exact moment when the original linear conversation view was replaced with a hierarchical one[5] (E2). The regression discontinuity analysis corresponding to that cutoff is shown in Figure 6 (right). The discontinuity shows a notable increase in both the *h-index* itself and in the slope of the regression, indicating an acceleration after the intervention. In particular, the break at the cutoff is 0.28 ($\omega_2 = -2.550$; $\omega_3 = 0.0134$).

—Figure 6 about here —

Since the *h-index* is a non-trivial combination of the width and depth of a discussion thread, we also examine these two metrics separately. The width as a function of time is shown in Figure 7. In this case, we observe a strong coupling of the width with a seasonal pattern, possibly reflecting the drop of activity during winter holidays. This prevents direct application of RDD using a linear model. Alternatively, we present a symmetric moving average of 24 weeks, to indicate cyclic activity, and 52 weeks, to completely detrend the time series. This reveals a progressive decay trend in 2013, not related to a sudden change.

Figure 8 shows the results corresponding to our analysis of the discussion depth. Unlike the width, this metric does not exhibit a seasonal pattern and it is amenable for RDD analysis using a linear model. As before, in Figure 8 (left) we show the F-test value as a function of time. In this case, the global maximum coincides with a local maximum of the *h-index*, four months after the hierarchical view was introduced. Interestingly, by looking for possible explanations of such a change, we found that the hierarchical conversation view was modified at that time. In particular, the maximum depth of the visualized discussion was increased from four to size levels[6]. The RDD results for the thread depth are presented in Figure 8 (right). The break at the cutoff is 1.614 ($\omega_2 = -0.488$; $\omega_3 = 0.009$) and confirms the discontinuity, while the null hypothesis does not capture such effect.

—Figure 7 about here —

—Figure 8 about here —

Finally, we look at the relation between depth and width, also over time. In Figure 9 we show scatter plots, with horizontal and vertical axes corresponding to the width and depth, respectively (bin size = seven days). The color gradient in Figure 9 (left) goes from the oldest threads (blue) to the most recent ones (red). We observe that the first discussion threads are characterized by wide

---

[5] https://www.meneame.net/notame/2002188
[6] https://github.com/gallir/Meneame/commit/b35a6b2

but not deep structures, as in our example thread of 2011 presented in Figure 3 (left). Threads progressively acquire more depth and reduce width. This trend changes abruptly in January 2015, when the hierarchical view replaced the original linear view. Subsequently, the width remains stable while the depth grows much faster, especially with the second version of the hierarchical view in which the maximum visual depth is increased. This may explain why the slope of the *h-index* increased in Figure 6 (right): the second version of the hierarchical conversation view induced much deeper conversations. Figure 9 (right) makes explicit this segmentation using different colors for each period: blue for the linear conversation view, yellow for the first version of the hierarchical view, and red for the second version hierarchical conversation view (increased maximum visual depth).

From these results we can conclude that our methodology detects the change of the conversation view (E2) as the most significant event in Menéame in terms of promoting deliberation, since the intensity of argumentation in the discussion threads is increased, an effect which is accentuated with the second version of the hierarchical view.

—Figure 9 about here —

5   Discussion

The analysis of Menéame in this study has been designed to answer our research question which deals with the automatic detection of events that affect online deliberation.

Our first candidate event was the rise of the 15M movement (E1). This was motivated by two observations: (a) the relevance of Menéame in the communication of this social movement (Toret et al., 2015; Otero and Sánchez, 2012; Mena, 2011; Postill, 2013) which led to an increase of political talk in the platform, and (b) political subforums in similar social news sites have exhibited greater levels of deliberation than other subforums (Gonzalez-Bailon et al., 2010). The preliminary exploration of the dataset revealed outstanding levels of activity when actions from the 15M movement occurred. However, our statistical analysis of online deliberation did not find any significant effect induced by E1. Although politics might get relevance by the emergence of this movement, this did not affect the deliberative structure of online discussions.

Our second candidate event was the change of the conversation view from linear to hierarchical (E2). Previous studies indicated that hierarchical views helped to improve communication (Smith et al., 2000; Venolia and Neustaedter, 2003; Fuks et al., 2006; McVerry, 2007). Indeed, this is the typical interface of asynchronous and deliberative discussions (Janssen and Kies, 2005; Smith et al., 2009). Our statistical methodology allowed us to detect the change from linear to hierarchical in January 2015. When discussion threads started to be displayed hierarchically, the indicator of deliberation (*h-index*) suddenly increased, i.e. discussion structures became much more complex. Therefore, E2 was a significant

event. Given that this change occurred in isolation, i.e. no other features took place at that specific time-stamp, this confirms that the adoption of hierarchical conversation views has a positive effect in online public spheres.

As we have indicated in the introduction of the paper, our methodology was not limited to these two events, any time-stamp was susceptible to be the most significant cutoff in the regression discontinuity design. In particular, we found an additional significant event when the visual depth of online discussions was increased. Given that we were unaware of this event in the design of the experiment, this finding illustrates the flexibility of our statistical approach.

In general, the increase of the depth of discussion threads (associated with higher argumentation) is induced by the long chains of reciprocal interaction between users, as shown in Figure 3. Since reciprocity, sometimes referred to as interactivity or mutuality, is one of the most common features when measuring online deliberation (Dahlberg, 2011; Trénel, 2004; Black et al., 2011; Friess and Eilders, 2015), future work will focus on whether reciprocity is also affected by these technical changes of the conversation view. In this context our results may prove useful to understand how design of online platforms — in terms of what social information they present — may shape our decision-making environment (Margetts, 2017).

Despite the significance of the results, we should reflect on both the benefits and limitations of detecting events with structural indicators of online deliberation from Gonzalez-Bailon et al. (2010). The decision of measuring online deliberation using the complexity of discussion threads, while disregarding content, will allow academics to easily replicate this methodology on large datasets from online discussions platforms of very diverse nature. In addition, our characterization of the structure of reply structures is aligned, in part, with Dahlberg (2001a) which suggested to focus on the contestation rather than emphasizing communicative rationality. Nevertheless, we are aware that our approach represents a confrontation with the Habermasian conceptualization of public spheres (Habermas, 1991) and the existing language-based coding schemes for online deliberation, e.g. Trénel (2004); Stromer-Galley (2007); Coleman et al. (2002); Friess and Eilders (2015). Language-independent approaches to online deliberation, as applied here, examine the strength of exchanges rather than content. Therefore, these cannot characterize whether users back their comments in a respectful manner or simply fight with a flaming or trolling behavior. Given that these coding schemes are too expensive for large datasets and inspired by Black et al. (2011), which showed that some features can be automatically inferred, future work might address these limitation with natural language processing techniques to also compute linguistic features.

6 Conclusion

This study has proposed a language-independent methodology to detect events which affect deliberation in large datasets from online discussion platforms. Results on the discussions from the social news site Menéame show the influence of

online discussion interfaces on the emergence of deliberative network structures. In particular, the change of the conversation view from linear to hierarchical induced deeper discussion threads which are associated with higher argumentation. This was accentuated when the maximal visual depth was increased. This event was not considered when designing the experiment and, therefore, illustrates the flexibility of our methodology. In conclusion, with this study we aim to provide a better understanding of the interplay between the design of online discussion platforms and the effective development of online public spheres.

Acknowledgments

Appendix: Events related to the 15M movements

In this appendix we describe the days with an abnormally higher level of posting/voting activity (see Figure 2), which were precisely prominent days of the 15M movement:

– **17-19/05/2011** The rise of the 15M movement
  On May 15, 2011, the first demonstration took place in the largest cities of Spain. At the end of the demonstration in Madrid, a group of 40 protesters decided to camp in Puerta del Sol Square (*Acampada Sol*). The next day, although police forces attempted to evict the camp, more protesters joined Acampada Sol and around 200 people also decided to camp in Catalunya Square in Barcelona (*Acampada BCN*). This trend continued in the following days and the main squares of cities in Spain were occupied for weeks under the motto '15M movement'.
– **27/05/2011** Violent police eviction of Acampada BCN
  The City Council of Barcelona sent 350 police officers to dismantle the protesters in Acampada BCN early in the morning. This action resulted in a violent clash between police and citizens. New calls to protest emerged in all the squares yet occupied in Spain and, in the evening, protesters rebuilt Acampada BCN.
– **21/02/2012** 15M Outbreak in Valencia (*Primavera Valenciana*)
  Inspired by the actions of the 15M movement, schoolchildren and university students in Valencia started a rally of daily protests against the Spanish Government because of corruption scandals and the austerity measures proposed for debt control.
– **11/07/2012** Asturian miners' march
  Coal miners from Asturias organized a march in Madrid in order to protest against the plans of the Government to reduce subsidies for 40 mines. Asturian miners arrived to Puerta del Sol Square and received the support of thousands of citizens.
– **25-27/09/2012** Encirclement of the Parliament (*25S Rodea el Congreso*)
  On September 25, 2012, protesters from the 15M movement decided to surround the Spanish Parliament to claim against austerity measures, the tax system and the overall Spanish political system. Protests resulted into riots between police forces and citizens and, two days later, new surrounding actions were made by protesters.
– **31/01/2013** Podemos' anti-austerity march (*Marcha del Cambio*)
  Podemos, emerging political party founded in the aftermath of the 15M movement, organized an anti-austerity march in Madrid. Tens of thousands of citizens attended the event, hosted in Puerta del Sol Square.
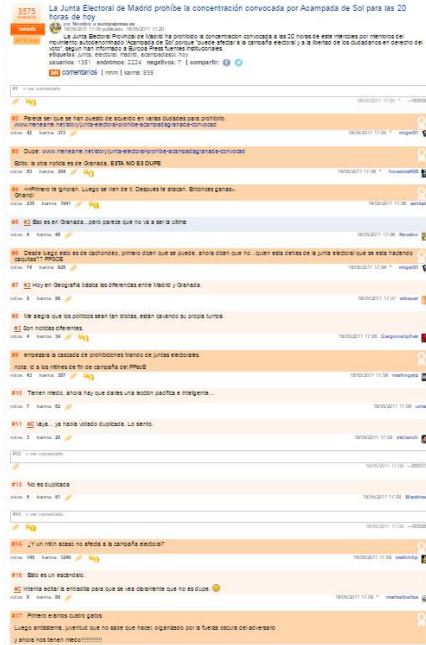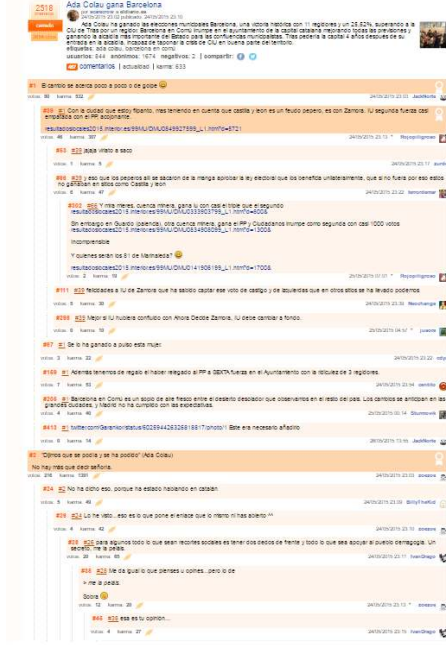
(a) **Linear** conversation view.   (b) **Hierarchical** conversation view.
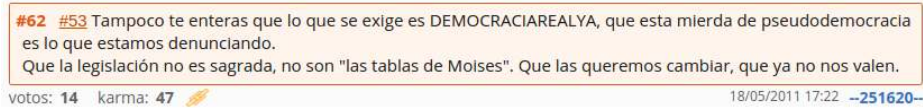
(c) Example of how a comment (#62) replies a previous comment (#53).

Fig. 1: Comparison between (a) the linear conversation view of a discussion thread from 2011 about the 15M grassroots movement occupation of Puerta del Sol Square in Madrid, and (b) the hierarchical conversation view of a discussion thread from 2015 about the victory of the grassroots party Barcelona en Comú in the local elections. In (c) we show a comment to another comment (identical in both views). Thus, every comment starts with the symbol # followed by the id. If the comment is replying to another comment, not to the story, it automatically shows the symbol # followed by the id of the comment being replied to. Comments posted by the story's author are blue, comments scored negatively are white (text is hidden unless clicked upon), and the rest of the comments are orange using intensity to indicate the voting score (i.e. the better, the darker).
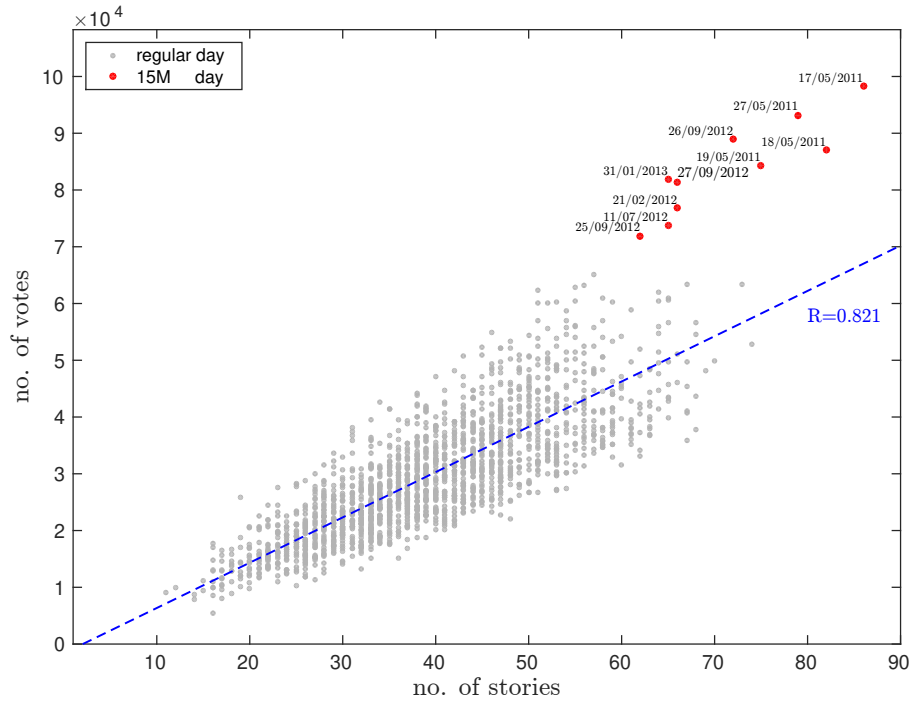
Fig. 2: Scatter plot of days in the dataset of Menéame (2011-2015). Each day is represented by a dot with coordinates consisting of the number of stories in the front page (X-axis) and the sum of the votes to these stories (Y-axis). Although both dimensions are highly correlated, prominent days in the Spanish 15M movement (red markers) exhibit an abnormally high level of activity.
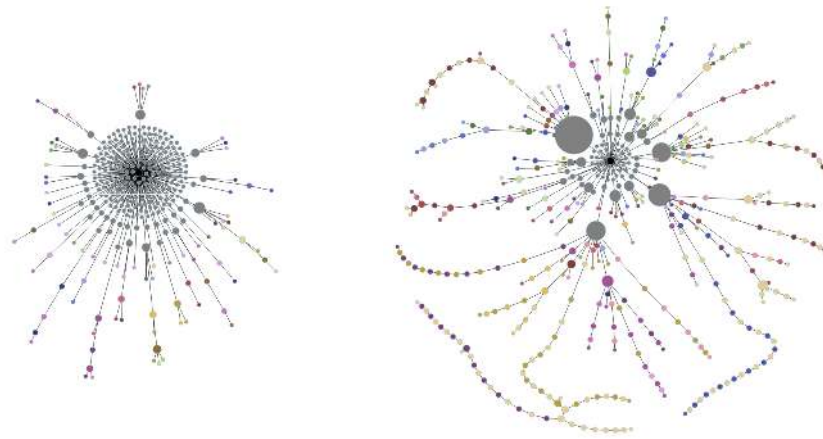
(a) Thread from 2011 discussed with the **linear** conversation view

(b) Thread from 2015 discussed with the **hierarchical** conversation view

Fig. 3: Visualization of a discussion thread from 2011 about the 15M grassroots movement occupation of Puerta del Sol Square in Madrid (left), and a discussion thread from 2015 about the victory of the grassroots party Barcelona en Comú in the local elections (right). Nodes (i.e. comments) are sized based on the number of replies. To identify authorship, we set the same color to comments published by the same user except for the root node (black) and the comments at the first level (gray). The visualizations indicate that the left thread has a star-like structure (width=256, depth=5, h-index=4) while the right thread has a much more complex structure (width=110, depth=37, h-index=8).
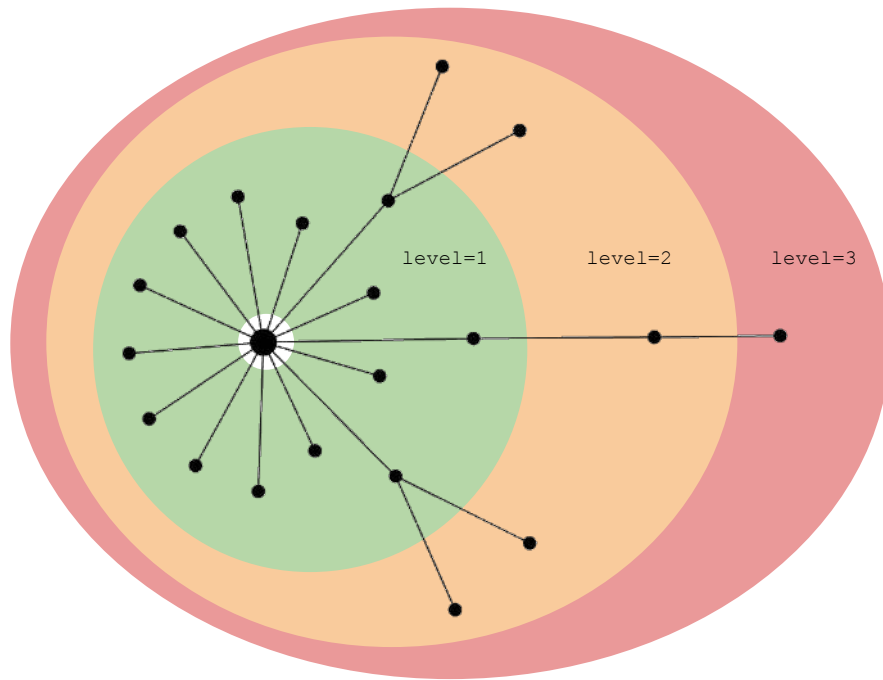
level=1   level=2   level=3

Fig. 4: Example of a discussion thread presented as a radial tree. Width=*14*, because is the maximum number of comments at any reply level. Depth=*3* because is the number of levels. *h-index=2* because is the maximum h level in which there are, at least, h comments; i.e. there are more than two comments at the second level but less than three comments at the third level.
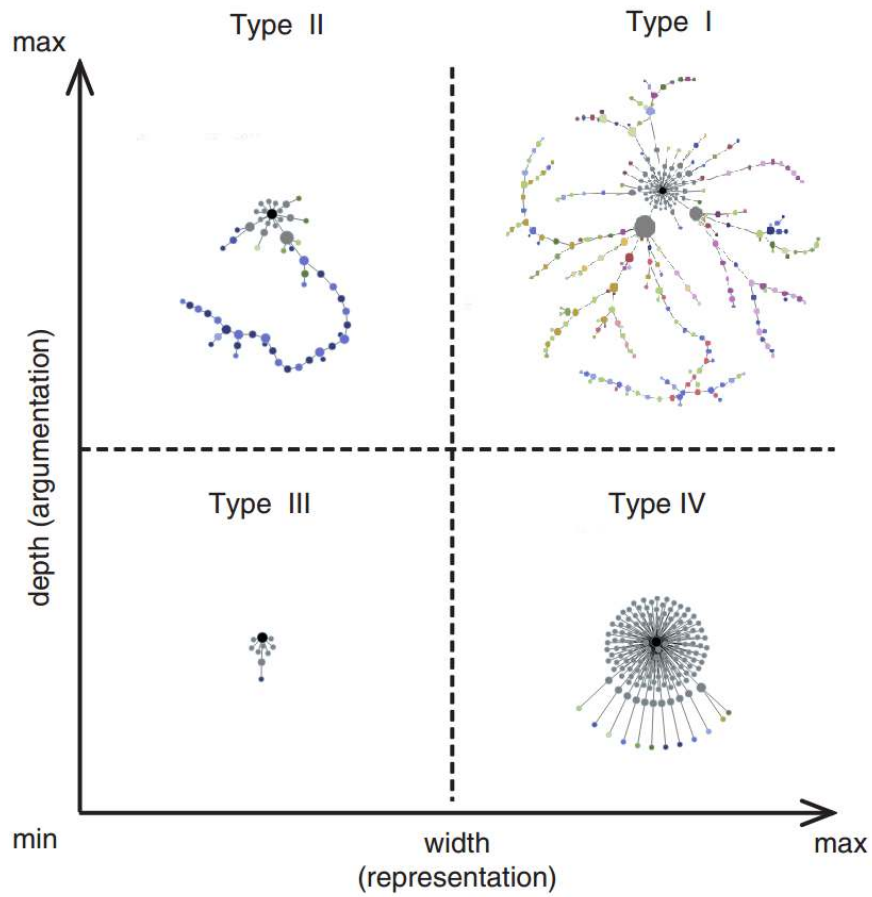
Fig. 5: The four types of discussions defined in Gonzalez-Bailon et al. (2010) using real threads from our Menéame dataset. Type I presents the best conditions for online deliberation: representation (width of the thread) and argumentation (depth of the thread).
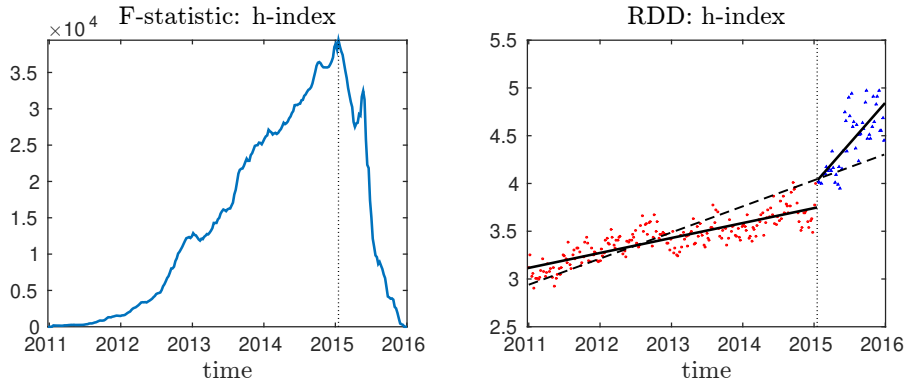
Fig. 6: Regression discontinuity design (RDD) applied to the *h-index* (bin size = seven days). The vertical line denotes the optimal cutoff obtained through an F-test. Red circles correspond to discussion threads before the optimal cutoff and blue triangles correspond to threads after the optimal cutoff. The solid line is the result of the discontinuous linear regression and the dashed line corresponds to the linear regression of the null model. The *h-index* shows an abrupt increase after the cutoff. The slope also increases.
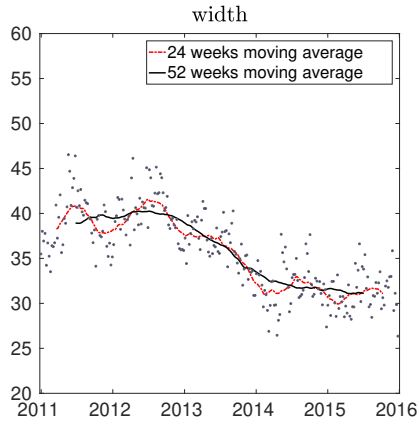

Fig. 7: The width of discussion threads as a function of time (bin size = seven days). Results show that the width is affected by a seasonal pattern. The dashed lines are symmetric moving average using 24 weeks (red) and 52 weeks (black). A decrease is observed during 2013.
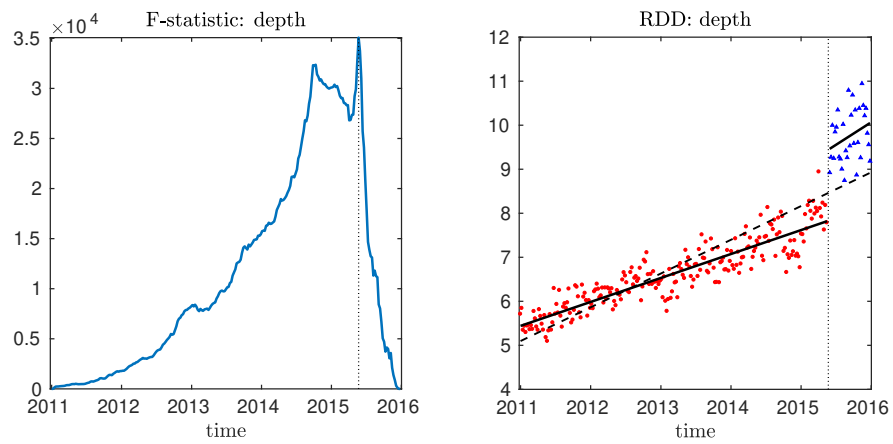
Fig. 8: Regression discontinuity design (RDD) applied to the thread depth (bin size = seven days). The vertical line denotes the optimal cutoff obtained through an F-test. Red circles correspond to discussion threads before the optimal cutoff and blue triangles correspond to threads after the optimal cutoff. The solid line is the result of the discontinuous linear regression and the dashed line corresponds to the linear regression of the null model. The thread depth shows an abrupt increase after the cutoff. The slope also increases.
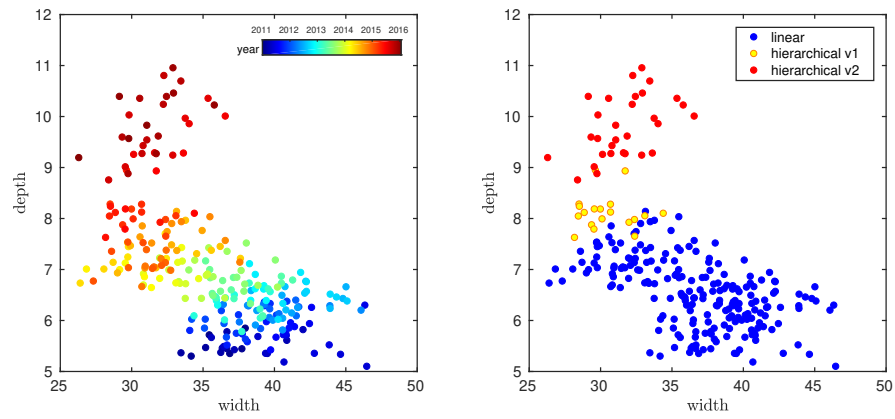
Fig. 9: Scatter plot of width versus depth of the discussion threads (bin size = seven days). Left: dots are colored in a scale which indicates time. Right: blue dots are threads presented in a linear conversation view, yellow dots are threads presented with the first version of the hiearchical view, and red dots are threads presented in the second version hierarchical conversation view (increased maximum depth). From Jan 2011 to Jan 2015, depth increases while width decreases. Then, the linear conversation view is replaced by the a hierarchical one and the depth grows much faster while the width remains stable.

Bibliography

Ackerman, B. A. and Fishkin, J. S. (2004). *Deliberation day*. Yale University Press.

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., and Ladwig, P. (2014). The "nasty effect:" online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, 19(3):373–387.

Aragón, P., Gómez, V., and Kaltenbrunner, A. (2016a). Visualization Tool for Collective Awareness in a Platform of Citizen Proposals. In *ICWSM-16 - 10th International AAAI Conference on Web and Social Media*. The AAAI Press.

Aragón, P., Volkovich, Y., Laniado, D., and Kaltenbrunner, A. (2016b). When a movement becomes a party: Computational assessment of new forms of political organization in social media. In *Tenth International AAAI Conference on Web and Social Media*.

Baek, Y. M., Wojcieszak, M., and Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? why? what? with what effects? *New media & society*, 14(3):363–383.

Barber, B. R. (1984). *Strong democracy: Participatory politics for a new age*. Univ of California Press.

Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge University Press.

Black, L. W., Welser, H. T., Cosley, D., and DeGroot, J. M. (2011). Self-governance through group discussion in wikipedia: Measuring deliberation in online groups. *Small Group Research*, 42(5):595–634.

Bowker, N. and Tuffin, K. (2003). Dicing with deception: People with disabilities' strategies for managing safety and identity online. *Journal of Computer-Mediated Communication*, 8(2):0–0.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274.

Carpini, M. X. D., Cook, F. L., and Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annu. Rev. Polit. Sci.*, 7:315–344.

Chadwick, A. (2006). *Internet politics: States, citizens, and new communication technologies*. Oxford University Press, USA.

Chadwick, A. (2008). Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance. *Isjlp*, 5:9.

Chambers, S. (2003). Deliberative democratic theory. *Annual review of political science*, 6(1):307–326.

Coleman, S. and Gotze, J. (2001). *Bowling together: Online public engagement in policy deliberation*. Hansard Society London.

Coleman, S., Hall, N., and Howell, M. (2002). *Hearing voices: the experience of online public consultations and discussions in UK governance*. Hansard Society.

Crane, R. and Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653.

Dahlberg, L. (2001a). Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication*, 7(1):0–0.

Dahlberg, L. (2001b). The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4):615–633.

Dahlberg, L. (2011). Re-constructing digital democracy: An outline of four 'positions'. *new media & society*, page 1461444810389569.

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. *Political communication*, 22(2):147–162.

Davies, T. and Gangadharan, S. P. (2009). *Online deliberation: Design, research, and practice.* University of Chicago Press. CSLI Lecture Notes.

Edwards, A. R. (2002). The moderator as an emerging democratic intermediary: The role of the moderator in internet discussions about public issues. *Information Polity*, 7(1):3–20.

Elster, J. (1998). *Deliberative democracy*, volume 1. Cambridge University Press.

Esau, K., Friess, D., and Eilders, C. (2016). Design matters! an empirical analysis of online deliberation on different news platforms. In *Internet, Policy & Politics Conferences: The Platform Society.*

Ferreras Rodríguez, E. M. (2009). Nuevas herramientas en el ecosistema digital. la promoción social de noticias en castellano. análisis de menéame. *Comunicación y pluralismo*, 8:155–167.

Ferreras Rodríguez, E. M. (2011). New proposals for the creation of an informative agenda: The case of menéame. *Innovation: Economic, Social, and Cultural Aspects*, page 157.

Figueiredo, F., Benevenuto, F., and Almeida, J. M. (2011). The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 745–754. ACM.

Fishkin, J. S. (1997). *The voice of the people: Public opinion and democracy.* Yale University Press.

Friedman, B., Khan Jr, P. H., and Howe, D. C. (2000). Trust online. *Communications of the ACM*, 43(12):34–40.

Friess, D. and Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3):319–339.

Fuks, H., Pimentel, M., and De Lucena, C. J. P. (2006). RU-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. *International Journal of Computer-Supported Collaborative Learning*, 1(1):117–142.

Gómez, V., Kaltenbrunner, A., and López, V. (2008). Statistical analysis of the social network and discussion threads in slashdot. In *Proceedings of the 17th international conference on World Wide Web*, pages 645–654. ACM.

Gonzalez-Bailon, S., Kaltenbrunner, A., and Banchs, R. E. (2010). The structure of political discussion networks: a model for the analysis of online deliberation. *Journal of Information Technology*, 25(2):230–243.

Graham, T. and Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications-Sankt Augustin then Berlin*, 28(2):173–204.

Grönlund, K., Himmelroos, S., et al. (2009). The challenge of deliberative democracy online–a comparison of face-to-face and virtual experiments in citizen deliberation. *Information Polity*, 14(3):187–201.

Habermas, J. (1984). The theory of communicative action, volume i. *Boston: Beacon*.

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society.*

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy.*

Hacker, K. L. and Van Dijk, J. (2000). What is digital democracy. *Digital democracy: Issues of theory and practice*, pages 1–9.

Halpern, D. and Gibbs, J. (2013). Social media as a catalyst for online deliberation? exploring the affordances of facebook and youtube for political expression. *Computers in Human Behavior*, 29(3):1159–1168.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 102(46):16569–16572.

Janssen, D. and Kies, R. (2005). Online forums and deliberative democracy. *Acta política*, 40(3):317–335.

Kaltenbrunner, A., Gonzalez, G., Ruiz De Querol, R., and Volkovich, Y. (2011). Comparative analysis of articulated and behavioural social networks in a social news sharing website. *New Review of Hypermedia and Multimedia*, 17(3):243–266.

Kim, J.-Y. (2006). The impact of internet use patterns on political engagement: A focus on online deliberation and virtual social capital. *Information Polity*, 11(1):35–49.

Kohavi, R. and Longbotham, R. (2015). Online controlled experiments and a/b tests. *Encyclopedia of machine learning and data mining*, pages 1–11.

Lee, D. S. and Lemieux, T. (2009). Regression discontinuity designs in economics. Technical report, National Bureau of Economic Research.

Lehmann, J., Gonçalves, B., Ramasco, J. J., and Cattuto, C. (2011). Dynamical classes of collective attention in twitter. *WWW 2012*.

Luskin, R. C., Fishkin, J. S., and Iyengar, S. (2004). Considered opinions on us foreign policy: Face-to-face versus online deliberative polling. *International Communication Association, New Orleans, LA*, pages 1369–1387.

Malea, E. O., Ontalba, J. A., et al. (2009). Propuesta de indicadores métricos para gestores sociales de noticias: análisis de la prensa digital española en menéame. *Information research*, 14(3):2.

Malik, M. M. and Pfeffer, J. (2016). Identifying Platform Effects in Social Media Data. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media, ICWSM-16*.

Margetts, H. (2017). Political behaviour and the acoustics of social media. *Nature Human Behaviour*, 1:0086.

Marín, A. (2015). El usuario prosumidor en la promoción social de las noticias: estudio etnográfico en menéame. In *Desafíos éticos de la comunicación en la era digital: III Congreso Internacional de Ética de la Comunicación, libro de actas. Facultad de Comunicación de la Universidad de Sevilla, días 24, 25 y 26 de marzo de 2015. Luis Alfonso Guadarrama Rico, Juan Carlos Suárez Villegas y María del Mar González Jiménez (eds.).(pp. 500-513).* Sevilla: Universidad de Sevilla.

McVerry, J. G. (2007). Forums and functions of threaded discussions. *New England Reading Association Journal*, 43(1):79.

Mena, J. C. (2011). La dimensión híbrida del movimiento 15m: entre lo físico y lo virtual. In *Actas del V Congreso Online del Observatorio para la Cibersociedad "Hybrid Days". Recuperado el*, volume 16.

Min, S.-J. (2007). Online vs. face-to-face deliberation: Effects on civic engagement. *Journal of Computer-Mediated Communication*, 12(4):1369–1387.

Naaman, M., Becker, H., and Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on twitter. *Journal of the American Society for Information Science and Technology*, 62(5):902–918.

Otero, T. P. and Sánchez, C. C. (2012). Ciberactivismo y redes sociales. el uso de facebook por uno de los colectivos impulsores de la 'spanish revolution', democracia real ya (dry). *Observatorio (OBS*)*, 6(3).

Papacharissi, Z. (2002). The virtual sphere the internet as a public sphere. *New media & society*, 4(1):9–27.

Poor, N. (2005). Mechanisms of an online public sphere: The website slashdot. *Journal of Computer-Mediated Communication*, 10(2):00–00.

Postill, J. (2013). Democracy in an age of viral reality: A media epidemiography of spain's indignados movement. *Ethnography*, page 1466138113502513.

Ratkiewicz, J., Flammini, A., and Menczer, F. (2010). Traffic in social media i: paths through information networks. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 452–458. IEEE.

Smith, G., John, P., Sturgis, P., and Nomura, H. (2009). Deliberation and internet engagement: initial findings from a randomised controlled trial evaluating the impact of facilitated internet forums. In *ECPR General Conference*.

Smith, M., Cadiz, J. J., and Burkhalter, B. (2000). Conversation trees and threaded chats. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 97–105. ACM.

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of public deliberation*, 3(1).

Sunstein, C. R. (2001). *Republic. com 2.0*. Princeton University Press.

Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2):175–195.

Toret, J., Calleja-López, A., Marín, O., Aragón, P., Aguilera, M., Barandiaran, X., and Monterde, A. (2015). Tecnopolítica y 15m: La potencia de las multitudes conectadas.

Trénel, M. (2004). Measuring the quality of online deliberation. coding scheme 2.4. *Unpublished paper*, 18:2004.

Triviño, A. I. B. (2010). Presencia de las redes sociales en los cibermedios españoles. *Textual & Visual Media: revista de la Sociedad Española de Periodística*, 3:25–42.

Van Alstyne, M., Brynjolfsson, E., et al. (1996). Electronic communities: global village or cyberbalkans? In *Proceedings of the 17th International Conference on Information Systems. New York: Wiley.*

Venolia, G. D. and Neustaedter, C. (2003). Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 361–368. ACM.

Wilhelm, A. G. (1998). Virtual sounding boards: How deliberative is on-line political discussion? *Information Communication & Society*, 1(3):313–338.

Wilhelm, A. G. (2000). *Democracy in the digital age: Challenges to political life in cyberspace.* Psychology Press.

Wright, S. (2009). The role of the moderator: problems and possibilities for government-run online discussion forums. *Online deliberation: Design, research, and practice*, pages 233–242.

Wright, S. and Street, J. (2007). Democracy, deliberation and design: the case of online discussion forums. *New media & society*, 9(5):849–869.

Zhou, X., Chan, Y.-Y., and Peng, Z.-M. (2008). Deliberativeness of online political discussion: A content analysis of the guangzhou daily website. *Journalism Studies*, 9(5):759–770.