A model for the analysis of online citizen deliberation: Barcelona case study

Rosa Borge, Joan Balcells and Albert Padró-Solanet

Universitat Oberta de Catalunya (UOC)

Paper prepared for the ECPR Joint Sessions of Workshops, University of Nicosia, Cyprus, 10-14 April, 2018

Abstract:
Following wins by new left-wing parties in Spain’s 2015 local elections, new participatory platforms were launched to enhance citizen participation and debate on local public policies. Platforms rolled out in Barcelona and Madrid are now being adopted by other Catalan, Spanish and European cities. The modular design of these platforms allows local governments to open up web pages for participatory deliberation with themed nesting of comments (similar to that of social media). But are these platforms really facilitating meaningful public deliberation? To answer this question, we assess the deliberative quality of one relevant online conversation developed within the local government online platform Decidim Barcelona. Our analysis focuses on the most commented citizens’ proposal discussed for the elaboration of the Barcelona’s Pla d’Actuació Municipal (strategic city planning), i.e. the granting of new licenses for tourist apartments. In order to evaluate the deliberative capacity of this conversation, we have followed the classical literature on deliberation in general and online deliberation in particular to empirically test a system of indicators that could be applied to other online media systems. We have carried out a detailed content analysis of the 336 comments from the conversation. In addition, we have examined how the different criteria of deliberative quality evolve as participants interact, by introducing the dimension of depth, inspired on the social computing research on conversation structures. The findings show that the relation between the deliberative quality and the depth of the conversation is in most cases curvilinear. The level of justification decreases as conversations go deeper, whereas the levels of reciprocity and negative empathy (disrespect) become more important over time before decreasing at a later stage. Overall, these findings show that online citizen deliberation may freely emerge, but they also reveal the difficulties of ensuring the quality of deliberation over time. Although the platform is well-designed to foster participation, citizen initiatives and open conversations, it needs a better structure to stimulate and preserve deliberation among citizens.

Keywords: Online deliberation, online conversations, deliberative criteria, online participation, participatory platforms, Barcelona, tourism.

Acknowledgements:
We acknowledge the contribution of Miguel Angel Domingo and Ricardo Serra for their help in collecting, coding and carrying out network analysis of the data. This research has been possible thanks to the grant given by the Catalan Government within the DEMOC 2017 call for projects on quality of democracy.
1.- Introduction

The aim of the paper is twofold. Firstly, following the empirical literature on online deliberation, we aim to develop a system of indicators that can serve to measure the deliberative capacity of online conversations. In this way, we intend to build an operative instrument for assessing the quality of deliberation in online conversations that can be applied to other online media systems. Secondly, we explore how the different dimensions of deliberative quality vary over time, as conversation trees progressively develop through users’ interaction. Thereby, our approach crosses between two strands of literature. On the one hand, the literature on deliberative quality, mainly focused on the comparison between deliberative practices and the ideal normative standards of good deliberation through content analysis; and on the other hand, the literature on conversation structures, which heavily relies on algorithms and computational models to describe patterns of interaction.

The paper is focused on Barcelona’s digital platform Decidim, an open-source platform especially designed to provide citizen input for governments, by allowing citizen-to-citizen and citizen-to-government interactions. This platform was created to enhance the experience of participation, and was firstly applied to the Barcelona’s Strategic City Planning, where citizens could send policy proposals to the new local government that was formed after the 2015 local elections. The paper uses as a case analysis the most commented issue in the Strategic City Planning, i.e. the granting of new licenses for tourist apartments. From a substantive point of view, this debate tackles a politically relevant and controversial issue, which has arisen enormous discussion not only in Barcelona but in many other cities affected by global mass tourism. Methodologically, the chosen debate represents a good case analysis, as it is both small enough to make content analysis through human coding available, and relatively big enough to meaningfully analyse its structure through computational techniques.

2.- The analysis of the online deliberation and its empirical criteria

Online deliberation has been analysed from two perspectives: the most classical one that focuses on how concrete online conversations fulfill the normative criteria of deliberation (Graham & Witschge, 2003; Stromer-Galley, 2007; Kies 2010; Steiner, 2012; Klinger & Russmann, 2014; Friess and Elders 2014), and the most recent one, based on Big Data, that focuses on the structure and development of the interactions along multiple conversations (Gómez et al, 2008; González-Bailón et al, 2010; Aragón et al, 2017). We think that both perspectives are useful and have theoretical and methodological strengths that can be combine in order to achieve a better understanding of online deliberation. The first perspective follows well-founded theoretical models of communication and democracy, and usually analyses in detail the content of the conversations, but in one moment in time, limited to few cases and in a non-cumulative manner (Gonzalez-Bailon et al, 2010). The second perspective derives from network analysis and computer science and examines the structural features of the conversations (such as width and length) and the interactions’ networks of
participants (such as the indegree, outdegree or centrality networks). However, this later perspective applies preconfigured models and algorithms that maybe are ill-adapted to political conversations and concentrates mainly in the skeleton of the conversation, frequently ignoring the content inside. Our research seeks to bridge the gap between these two disciplines in benefit of this very complex object of study (Vergeer & Hermans, 2008).

Following the first perspective, in order to assess the deliberative capacity of the conversation we apply widely acknowledged criteria from the classical literature on deliberation in general and online deliberation in particular. In addition, we include in our theoretical framework and analysis some of the structural measures (length and width of the conversational threads) and standpoints (evolution of the conversation, interactions’ networks) from the second perspective.

Most of the authors pinpoint that there are three levels that should be considered to evaluate the deliberation in online platforms and the conversations hosted there (Dahlgren, 2005; Wessler, 2008; Kies, 2010; Friess & Eilders, 2014): 1) the institutional, structural or technical dimension of the platform and the space for conversation; 2) the interactive or communicative traits, which basically stem from the Habermasian theory of communication (1990); and 3) the collective and individual outcomes or impact of the online deliberative process. We will focus our analysis on the quality of the communication, that is the second dimension, but taken into account that the first and third dimensions are crucial for understanding the characteristics of the conversation and how the conversation evolves.

The most common criteria and indicators for every level or dimension will be drawn from the literature, taking into account that different systems of criteria and indicators have been empirically applied (Graham & Witschge, 2003; Stromer-Galley, 2007; Kies 2010; Steiner, 2012; Klinger & Russmann, 2014; Friess and Eilders, 2014), but very often authors use different labels for the same criterion or principle. In that sense, we will do an effort trying to identify the commonalities and, therefore, we will follow the criteria systems developed by Kies (2010) and Friess and Eilders (2014) since we consider that they have sufficiently integrated the previous ones. Next, each dimension, the meaning of the related criteria and how can they be operationalized and assessed will be explained.

1) Institutional or structural dimension

This dimension refers to how online spaces should be structured and organized to foster deliberation (Friess & Eilders, 2014: 6). The building of the deliberative space or process affects the communication and interaction fostered in the online space and it is not possible to develop a deliberative space if its structure and organization are not carefully designed for deliberative purposes (Friess & Eilders 2014: 15). There are several institutional and technical characteristics that must be taken into account in order to build a deliberative communication space, such as inclusiveness or inclusion (Kies, 2010: 42-44), asynchronous communication, content visibility, moderation, identity, perceived power of the communication spaces, division of labour into smaller units, relevant information and horizontal interaction (Friess & Eilders, 2014: 6-8).
The criterion of inclusion means that all those who are affected by and/or interested in the issues under discussion should be able to participate either actively or passively (Kies, 2010: 42). Thus, inclusion should be assessed by observing the technical characteristics of the online forum: the ease of access in terms of connectivity and ICT skills, and discursive rules such as moderation, registration and identification, which should not be perceived as barriers to promoting inclusive participation (Kies, 2010: 56).

In addition, following Friess and Eilders' (2014: 6-8) explanation of the technical criteria, an asynchronous communication space is needed to allow participants to spend more time reflecting and justifying their contributions. Moreover, user content should appear immediately in order to motivate contributions and lower perceived entry barriers. Moderation is also crucial to ensure deliberation in terms of civility and rationality and for promoting inclusive participation and good organization of the discussion. Furthermore, empirical evidence shows that personal identification has positive effects on the deliberative quality of online debates. The perceived power of communication spaces refers to the building of strong discussion spaces able to influence political outcomes. This condition encourages people to participate and be more deliberative. The technical design of the online forum should enable a division of labour into smaller units focused on different issues and debate areas in order to enlarge the opportunities for and quality of deliberation. Finally, the designed structure of the online platform should enable horizontal interaction and communication with other users.

2) Communicative dimension

This refers to the deliberative attitude of participants (Kies, 2010: 42) and what the communication process should look like, mainly in relation to the reaction of participants to each other's ideas (Friess & Eilders, 2014: 8). According to most of the authors, deliberation should be rational, interactive, equal and respectful. This is the core of the normative claims of deliberation theory as defended by Habermas (1990). The most crucial feature of deliberation is rationality in communication and discourse; that is, to state positions substantiated with arguments and empirical evidence, expecting critical exchange and diversity of arguments and being willing to change one's own opinion in light of better arguments (Friess & Eilders, 2014: 8). Therefore, rationality involves criteria such as reciprocity, justification, reflectibility, empathy (including civility) and plurality. Other criteria that are also important for assessing this deliberative attitude of participants are discourse equality and sincerity (Kies, 2010: 44-54).

Among all of these criteria, we would like to test more in depth the criterion of plurality, that is the extent to which the debates host and confront all the relevant opinions on a specific topic (Kies, 2010: 53). In order to ensure plurality, a deliberative conversation should include different viewpoints but also a sufficient diversity, that means different sociodemographic profiles and a presence of the actors at stake on the issue discussed.

With relation to plurality of viewpoints, in the case of forums dealing with social or political issues this refers to disagreement and confrontation of different political ideologies and standpoints on policies. Leading theorists of the deliberative model of democracy consider that conflict acts as a trigger for deliberation (Fung, 2003) but that disagreement should go in
hand with empathy, reciprocity and justification to produce a deliberative conversation (Guttman & Thompson, 2004). Some empirical studies show that the depth of the online conversations, and also, reciprocity, is related to heterogeneous positions on the topic of the conversation (Balcells & Padró-Solanet, 2016). Other studies precisely focused on online discussions in the Barcelona’s Decidim platform demonstrate that negative stands against an initial proposal were more likely to generate discussion cascades (Aragón et al, 2017). However, these findings should be qualified since interactions’ depth and width are only structural measures of the conversation. It is necessary to examine the content of the comments or posts in order to assess how other deliberative criteria are unfolding. Our aim is finally to ascertain whether long conversations, usually triggered by disagreement, fulfil the rest of criteria needed for a conversation to be deliberative.

In the following table (see Table 1), we present the most important criteria that characterize whether communicative interaction in an online platform can be considered as deliberative. We also explain how to operationalize these criteria through content analysis of the comments or posts. In section 4 the specific application of the operationalization to the online conversation under study will be described.

Table 1: The communication process in an online forum: deliberative criteria, their meaning and operationalization, based on Kies (2010: 42, 52-57).

<table>
<thead>
<tr>
<th>Deliberative criteria</th>
<th>Meaning</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse equality</td>
<td>Participants should have equal opportunity to introduce and question any assertion whatsoever and to express attitudes, desires and needs.</td>
<td>Analysis of discursive concentration by few participants and whether this concentration leads to control of the debate.</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>Participants should listen and react to the comments formulated by other participants.</td>
<td>Proportion of postings that are part of a thread versus the ones that initiate a thread but without followers, and by measuring the extent to which postings take into consideration arguments and opinions of a preceding posting.</td>
</tr>
</tbody>
</table>
### Justification
The opinions and propositions should be accompanied by reasoned, accessible, and moral justifications.

Observing whether opinions and suggestions are justified and how complex the justifications are. Analysis of whether the justification's arguments are either internal (based on personal viewpoints, values, situation) or external (based on facts, figures, comparisons, proposals or links to other information).

### Reflexivity
Participants should critically examine their values, assumptions, and interests, as well as the larger social context.

Notifying visible instances of opinion changes, conflict resolutions, moderation of the conversational tone, or solutions and conclusions based on comments from other participants.

### Empathy and civility
Participants should be sensitive to other views, opinions and situations, not only of those present during debates.

Counting the cases of disrespect, accusations, insults and negative ironies or jokes. Also collecting acknowledgements and positive remarks.

### Sincerity
Participants must make a sincere effort to make known all relevant information and their true intentions, interests, needs, and desires.

The most complex criterion to measure. Measurements based on inconsistency in speech, rhetorical forms of speech and complaints by other participants regarding the insincerity of other participants.

### Plurality and diversity (related to inclusion)
A deliberative context should be a context where a plurality of voices is heard especially if these voices are critical of the dominant opinions/ideologies. Also it should ensure users' diversity and representation of the actors at stake.

Analysing the degree to which the debates refer to different political ideologies and whether there is disagreement and conflicting standpoints. The registration system, if it is not anonymous, could give information about gender. It is also possible to see the language of the comments.

### 3) The outcome dimension
This dimension alludes to the results or impact of the deliberation, which could be individual or collective. At the individual level, participation in deliberative forums can contribute to increase tolerance, political knowledge and efficiency, public-spirited attitudes, willingness to
compromise or shift preferences (Friess & Eilders, 2014: 10; Hendricks et al., 2007). At the collective level, there are benefits related to the quality of decisions such as the generation of consensual decisions or at the very least decisions without errors, with high epistemic qualities, as they will be based on relevant reasons and evidence. As a result, the final decision will be more legitimated and supported by a wider public (Habermas, 1992; Friess & Eilders, 2014: 10).

In addition, Kies (2010: 54-55) highlights the relevance of the external impact of the deliberative process outside the context of the debate. That means that decisions resulting from online forums should have an impact on public debates and political decisions and even shape binding norms in order to contribute to the participation of citizens and guide and scrutinize official decision-making processes (Dalhberg, 2007: 49; Hendricks et al., 2007). Kies (2010: 57) argues that in evaluating the external impact it is helpful to check several indicators: signs of extension of the discussion to an external agenda, concrete political outcomes, participation of influential political personalities in the forums and participation of users in other discussion spaces.

The aforementioned deliberative criteria and operationalization is applied, through content analysis, to the assessment of the conversation on the granting of new licenses for tourist apartments. In addition to the description, we test how deliberative quality evolves through the process of interaction. Overall, our research has three interrelated objectives: 1.- To ascertain the degree of deliberation of the most commented proposal of the Barcelona’s Strategic City Planning and whether the institutional design and outcome dimension of the online participatory platform where the conversation is hosted foster deliberation; 2.- To apply and test a methodological instrument which assesses the deliberative capacity of online conversations and that take into account the conversational structure but also the content of the conversation; 3.- To analyse how the different criteria of deliberative quality evolve as participants interact in conversation. The two first objectives are more descriptive whereas the later one is more exploratory and comprises different hypotheses.

The relation between time and deliberation can be complex. According to deliberative democracy scholars (notably those that recognize the influence and legacy of Jürgen Habermas), deliberation should be ideally aimed towards a reciprocal understanding of conflicting viewpoints, to the point that its ultimate goal would be the achievement of a mutually accepted agreement. Under ideal conditions, positions should tend to converge over time towards consensus, regardless of the point of departure. Obviously, these ideal conditions are never fully achieved, but they work as a regulative ideal to drive deliberation forward. Some circumstances (such as unequal participation or unequal power, lack of publicity, predominance of interests over rationality, etc.) can make conflict persist over time, instead of mitigating it, and can have eroding effects on the quality of deliberation. Given the characteristics of the citizens’ proposals space on the platform (no moderation\(^1\), competition

\(^1\) The administrator of the platform acts as a moderator when sometimes appear in the middle or at the end of some conversations posting a link to similar proposals or announcing the merge of the proposal (not of the conversations) with others in a single action plan. There is the possibility for the participants to denounce a comment in case of disrespect, attacks or insults using a flag tag as in posts or comments in social media, in a fashion that can be considered as self-moderation.
for vote) and the issue at stake (strong positions, personal interests, political divisions), we expect high levels of confrontation between opposing views, which may temporarily increase some features of the deliberative quality such as reciprocity. But in the long run, as conversations go deeper, the lack of a common ground and moderation, and the discourse concentration by few participants is likely to negatively affect the quality of deliberation (lack of empathy, repetition of comments, no reflexivity, etc.).

We consider several possible hypotheses which relate the different indicators of deliberative quality with the development of conversations (see Figure 1). These hypotheses are not mutually exclusive, since they can simultaneously take place given the multidimensional character of deliberative quality.

- **H₀** assumes no relation between the development of the conversation and the criteria of deliberative quality. This null hypothesis would consider the quality of deliberation to be a stable phenomenon with no variation over time.

- **H₁** assumes indicators of deliberative quality to increase as conversations further develop. This hypothesis would imply that certain quality elements are especially activated in the heat of debate. It could also indicate a kind of virtuous circle; the existence of deliberative quality would progressively lead to better quality. This would be the expected relation of criteria such as reciprocity, as continuous interaction is likely to encourage a more personal and direct way of communication.

- On the contrary, **H₂** assumes deliberative quality indicators to decrease as conversation further develop. This hypothesis would show instead a reduction of quality that could be explained by a plurality of different factors, such as a certain fatigue due to persistent conflict and strong disagreement. For example, this could be the case of criteria such as justification if arguments are presented in a more complete form at the beginning of the conversation and then progressively fade away.

- **H₃** is a combination of the two previous hypotheses, that could probably fit better with the complexity of conversations as it assumes no linearity. H₃.₁ shows that quality grows until a certain point of fatigue or exhaustion, from which it starts to decrease. H₃.₂ instead shows the reversal pattern, quality decreases over time but is reactivated at the end of the conversation.
3.- The new participatory platforms *Decidim* and the Barcelona case study

Barcelona constitutes a good laboratory for citizen participation for several reasons. Firstly, the city and its surroundings encompass a vibrant and dynamic urban area with strong civic networks and a rich associative life. Secondly, it has a long record of participatory experiences and has led important citizen mobilizations, most recently connected with the cycle of protest and discontent triggered by the economic crisis. Thirdly, since the recovery of democracy in the last quarter of the 20th century, participation has been a paramount concern for the successive local governments in Barcelona, which have invested a big deal of resources in participatory policy-making, though not always successfully. This has been intensified after 2015 local elections, with the arrival to power of a new leftist political party (*Barcelona en Comú*) mostly composed by civil society activists with a strong commitment towards citizen participation. The mayor herself -Ada Colau- was a prominent leader of the anti-eviction movement. Once in office, the not always easy combination of activism and institutional
politics has generated some interesting discussions about the role of political institutions, and the difficulties of bridging the logic of representative and direct democracies.

Barcelona’s Pla d’Actuació Municipal 2016-2019 (Strategic City Planning) was an attempt to bridge these two worlds by giving citizens the possibility of helping to define from below the policy priorities for the local government. It was also designed to facilitate and enhance participation by opening a digital space for discussion and vote. According to a recent public opinion poll carried out in the city of Barcelona, 16% of citizens would be very much predisposed to participate in a face-to-face meeting to discuss a local issue or policy; but this percentage rises up to 38% when asked for giving their opinion online, and up to 51% for voting in referenda or citizen consultations (Ajuntament de Barcelona 2017). The Strategic City Planning was actually based on a hybrid method, combining both face-to-face meetings with online participation through a new digital and open-sourced platform (Barcelona Decidim). Citizens were able to make policy proposals through the digital platform, and each proposal could be openly discussed in a forum and voted by other citizens who had previously registered in. The Plan had in total 10,860 citizen proposals, 18,192 comments and 25,435 online participants. The debates on policy proposals were not conducted by moderators, but relied on spontaneous participation. Even though registration was necessary, citizens could use nicknames to make communications anonymous.

Somehow paradoxically, the most discussed issue did not overcome the veto of the city council. The most commented policy proposal was granting new licenses for tourist apartments, which received a total of 336 comments. The proposal was finally rejected regardless of the supports, because it was considered to go against the political line of the local government, one of whose main priorities has been to regulate the tourist boom and keep it under control, especially after the expansion of new platforms for renting apartments. The debate on tourism is at this moment highly controversial in Barcelona, as it is in other cities affected by mass global tourism. Though an important source of income, the negative externalities of mass tourism -such as gentrification- have become a troublesome problem for locals. The discussion on the model of tourism has pervaded all spheres of local politics, and has become a matter of intense political debate. The opening of the digital space for conversation was probably seen as an opportunity to bypass or influence decision-makers before a public audience, and put the issue of new licenses into the political agenda. The online debate that took place in the platform reflects the different interests at stake, namely the confrontation between property owners and local neighbours, and their reasons for favouring or blocking the proposal.

4.- Operationalization, data and methods

Regarding the first objective, we have assessed the deliberative quality of the most commented citizen proposal and how the online participatory platform hosting it contributes to make a deliberative space. Following the criteria explained in previous sections, first we ascertained how the online platform were structured and technically organised in order to foster deliberation. Secondly, regarding the communicative dimension, we carried out content and network analysis of the conversation raised around the proposal to grant new licenses for
tourist apartments. Thirdly, for the outcome dimension, we determined whether the results of the debates analysed were accepted by the municipal government.

As for the communicative dimension, we analysed the most commented proposal in the Strategic City Planning that stands for new licences for tourist apartments, against the restrictions decided by the local government. This proposal was registered on February 28, 2016 and the conversation ended up on April 4, 2016. It received 336 comments, 196 votes and 72 citizens took part in the conversation. Among the 20 most commented proposals, 3 focused on tourist lodging and one on the Strategic Tourism Planning. However, these proposals are not among the ones with the majority of votes showing that neither regulation or deregulation did get so much unanimity. It is evident that the issue of tourism arouses interest and controversy and in that sense it is an ideal case for studying dialogical practices.

Following the deliberative criteria compiled in section 2, we present in Table 2 the operationalization of the indicators for each criterion. As recommended for any empirical analysis of deliberation, the operationalization should adapt to different contexts and types of conversations and spaces (Kies, 2010: 55). Consequently, we have designed a specific set of indicators to measure the deliberative quality of an online conversation on a socio-political issue.
### Table 2: Operationalization of deliberative criteria.

<table>
<thead>
<tr>
<th>Deliberative criteria</th>
<th>Indicators</th>
</tr>
</thead>
</table>
| **Discourse equality** | ● Number of comments per participant, mean and sd  
   ● Number of participants that concentrate more comments  
   ● Gini Index |
| **Reciprocity** | REC Index:  
   1. Literal mentioning of the name of the participant to whom the comment is referring (1)  
   2. Literal citation in quotes of previous comments or parts of the comments (1)  
   3. Questions that refer to previous comments or are rhetorical or an interpellation (1)  
   Other:  
   ● Number of single comments (first level without comments)  
   ● Number of comments ≥ 2 level (all levels with comments) |
| **Justification** | JUST Index:  
   1. External justification based on data or on the content of laws and regulations.  
   2. External justification based on facts or statements as if they were facts  
   3. External justification based on comparisons or similes (among cities, countries, cases, social classes, collectivities, kinds of phenomena..)  
   4. External justification referring to the common good or interest, justice or rights or to a collective or social class: 4.1) Regarding the city in general, a country, citizenry, etc.. 4.2) Regarding an specific collectivity, social class, socio-demographic group, etc..4.3) Regarding a collective good: the economy, wealth, peaceful coexistence, civism, culture, cosmopolitanism, employment, health, etc..  
   5. External justification based on links to webs, videos, or different information sources.  
   6. External justification based on solutions or proposals.  
   7. Internal justification referring to the personal experience or situation.  
   8. Internal justification referring to own values, rights, ideologies, viewpoints.  
   Other:  
   ● Very short comments without justification (e.g. comments of support or rejection without argumentation or over-simplified statements)  
   ● Number of characters of the comments. |
| **Reflexivity** | REF Index:  
   1. Opinion changes.  
   2. Conflicts resolution.  
   3. Moderation of the conversational tone, conciliation.  
   4. Conclusions based on previous comments.  
   5. Solutions based on proposals from other participants. |
Empathy  
Positive empathy: Thanks, acknowledgement, admiration, enthusiastic agreement.

Negative empathy or disrespect index:
1. Accusations.
2. Ironies.
4. Insults,
5. Ridiculizations

Other:
• Repeated comments.
• Number of trolls o participants that hinder or boycott the conversation.

Sincerity  
Indirect indicators of sincerity (SINC Index):
• Inconsistency in speech.
• Rhetorical forms of speech, demagogic expressions.
• Perception by other participants of the insincerity of other participants.

Plurality and Diversity  
Plurality of viewpoints:
• Approval or disapproval of the initiating proposal.
• Conflictual comment that is an opposing comment to the previous one regarding the initiating proposal.
• In favour of local government position or regulations on the issue at stake or against local government position or regulations.

Diversity:
• Proportion of women and men.
• Proportion of different languages: catalan or spanish.

Among the deliberative criteria examined, the criterion of sincerity was not studied, as its measurement is somehow complex (Kies, 2010: 52, 57). Data on gender could only be gathered when participants wrote their names.

The coding process developed in the following phases. First, four coders (the three coauthors and one research assistant) carried out a previous codification of 50 comments in order to test and agree on the coding dictionary. Secondly, the total coding of the 336 comments was carried out by three coders (two of the coauthors and one research assistant). After the coding was completed, inter-coder reliability through Krippendorff’s alpha was calculated. The results were not satisfactory for 20 indicators (out of the total 28). So, thirdly, we decided to clarify the description of the indicators, improve the instructions for codification, add more information to the codebook and finally recodify the values for the 20 indicators. Eventually, 6

---

2 It has to be measured and assessed at the user level, rather than at the comment level, as an overall assessment of the consistency or the evolution of the expressed or the deduced positions of the users on the issue discussed.
of them were disregarded because more of the 95% of their values were 0. That was the case of the majority of the indicators of reflexivity and two indicators of incivility. In political discussions it is difficult to achieve a high level of reflexivity that refers to changes of the initial positions and efforts to arrive to a common agreement (Hendricks et al., 2007; Wojcieszak, 2011). In addition, the indicator of external justification regarding the common or collective good, interests or rights was discharged since was very difficult for the coders to agree on the codification. Therefore 13 indicators were recodified and their reliability increased significantly. In total we have 22 indicators or variables, most of them aggregated into single indexes. The final Krippendorff’s alphas are listed in Table 3 in the Appendix. All of them are between 0.713 and 0.922 so the coding can reasonably be seen as reliable.

To test the hypotheses, we have run several regression analysis between the different indicators of deliberative quality and the evolution of conversations over time. To capture the temporal evolution of conversations, we have firstly identified the different conversation threads that compose the whole discussion (by conversation thread we understand any initial comment that directly responds to the proposal and starts a line of discussion, together with all the subsequent comments that are derived from it). And secondly, we have identified the position of each comment in the thread, by specifying its level of depth (that is to say, the layer they belong to). Depth can be regarded as a good indicator of development because each level of depth represents a step further in the chain of replies from previous comments; thereby, it captures the natural give and take of conversations. To make it comparable across different conversation threads (since some can be very short, but some other very long), the variable ‘depth’ has been rescaled to range between 0 and 1; where 0 means that the comment is the beginning of a conversation thread, 1 means that it is the ending, and all the other values in-between show intermediate developments of the conversation thread.

5.- Analysis and discussion

5.1. Institutional design and impact of the deliberation.

The institutional design refers to the main structural and technological conditions that help to foster deliberation in a deliberative space. The examination of the different functionalities of the platform Decidim that hosted the conversations on citizens’ proposals confirms that most of the structural requirements for deliberation are satisfied. The platform allows the asynchronous participation of the users. Conversation is open so users can contribute with their post at any moment (they can spend more time reflecting and justifying their contributions), and the user contents appear immediately, allowing horizontal interaction (between users) by either commenting on other user’s threads and comments or voting on their contributions. There is a non intrusive moderation team, although they participate occasionally to facilitate conversations by posting links to similar proposals or by grouping the proposals that bring about action plans endorsed by the local government. User registration only requires a user name (or nickname), a password and an email address. With this simple registration it is possible to post a comment but in order to vote the system requires the postal address and ID card.
As we will show in the next sections, we have detected some problems affecting the deliberative quality of the conversation. Maybe these problems could have been solved with a better moderation and a more organised structure around the citizens’ proposals. There were dozen of proposals that were related or belonged to the same issue but were not properly merged into a single conversation or a general proposal. So in that sense the participants were not debating and working productively in smaller talks that can be aggregated into larger conversations. In addition, the relevant information for the discussion was not normally present because it depended on the will of the participants and the official administrators did not appear to supply it. There was also a lack of representation of all the interests at stake in the discussed issue: its seems that only small owners of flats and local residents were debating, which resulted in discourse concentration by some few and intransigent participants.

Regarding the political impact of these conversations, the local government committed itself publicly to accept the proposals with more votes. Eventually, proposals with many comments were also accepted in spite of having few votes. In fact 70% of the proposals presented for the Strategic City Planning were accepted. However, the local government reserved to itself the right to filter and reject proposals that were not in line with their political standpoints or priorities. That is what happened with the proposal under study, which was rejected with a very short feedback “The local government is not working on this line” appearing at the platform. Evidently, the feedback is not sufficient and can cause a lack of motivation among the participants and defendants of the proposal. It seems that the local government was not able to send a proper feedback to the 9.560 citizen proposals. We have to take into account that the online debates about the Strategic City Planning took place from January, 31 to April, 4, 2016, only 6 months after the electoral victory. Currently, the feedback functionality at the platform is being improved and more civil servants have been derived to serve in this important task (Decidim Team, 2017).

5.2 The deliberative quality

The success of the online debate on new licences for tourist apartments is due to its controversial nature and the involvement of a few users –with personal interests on the issue - that have triggered a noteworthy cascade of comments. Comparatively, that is indeed an exceptional case, since the majority of proposals in the platform have not generated any debate at all. If stakeholders have decided to take part in that particular debate, it is because they have considered the issue as relevant and have perceived that their participation could make the difference between the success and the failure of the proposal.

The debate has been mostly dominated by few users who have posted the majority of comments (see Figure 2 in the appendix). The level of discourse equality is indeed very low, as can be seen in all indicators that measure discourse equality; e.g. only 4 individuals, who are slightly above 5% of the total number of users, concentrate 65% of the posts, with a Gini index

---

3 “El govern municipal no està treballant en aquesta línia”
of .71; the average number of posts written by users is 5 with a standard deviation of 14 and a median of 1 (which clearly shows the distribution asymmetry).

Despite the uneven participation, as it is shown in Table 4, the levels of plurality show that the debate has captured the attention of stakeholders with antagonistic interests (66% of the comments are in favour of the proposal and 31% are against), and that the confrontation between defenders and opponents of the proposal – though asymmetrical – has been pretty balanced (69% of comments imply the interaction between individuals who hold opposite opinions). Both the conflictual disagreement and the balance between both sides have transformed the debate into a lively and vibrant discussion. The structure of the conversation tree shows that most threads are actually a chain of replies that successively alternate pro and against positions (see Figure 3 in the appendix). However, there is a participation bias in favour of those who support the proposal. As mentioned in a previous section, that might be so because participation in the platform could be seen as an opportunity for property owners who had no other means of influence to protest against the local government’s policy of limiting tourist licences. Thereby, participation in the debate was a way to make visible their discontent by publicly exposing their views and arguments as a sign of protest. Actually, there are in the debate some criticisms personally directed against the mayor and the government. That obviously has triggered the reaction of local neighbours who are affected by the negative externalities of mass tourism and strongly oppose any measure that, from their perspective, could make things worst.

A more careful analysis on the different criteria of deliberative quality provides some insights to better describe the quality of the debate. The lively interaction and attention between users is captured by the reciprocity dimension, with direct mentions between users, the use of direct questions and literal quotes of other users’ expressions or statements. Only 20.2% of the comments to the initial proposal do not trigger any comment whereas 79.8% produce following comments at different levels of depth. As it is shown in the distribution of comments by level of interaction (Figure 4 in the appendix), only 19.3% of the comments belong to the first level of interaction, i.e. they are a reply to a comment on the initial proposal. Remarkably, 43% of comments are at the third level or above in the chain of replies and the maximum depth in the conversation is 8 levels, which indicates that reciprocity between participants is relatively high.

Interestingly enough, the debate is characterized by a real exchange of arguments, which combine a repertoire of different forms of justification, such as the description of facts (62.4%), comparisons (23.3%), proposals (22.4%), references to personal experience (19.1%), and to a lesser extent, own ideological values (8.7%) and links (4.2%). These relatively high percentages show that there has been a serious attempt to provide evidence for supporting one’s point of view or attacking the weaknesses of opposing arguments. It is also remarkable that 1 out of 5 comments suggest solutions or proposals with regard to the issue under discussion, which show also the propositive nature of the debate with the discussion of different alternatives.
The levels of reflexivity (in terms of moderation of the tone of the debate) and positive empathy (in terms of recognition) are comparatively more modest, with percentages below 10%. The high levels of interaction and argumentation are not incompatible with elements that we have labelled as ‘negative empathy’ or ‘disrespect’ such as accusations (34.3%), irony (10.8%) or even insults (3.9%). Even though these elements cannot be seen as a sign of ‘quality’, it is not obvious that they necessarily play a negative role in deliberation. For instance, when combined with the presentation and development of arguments, they can increase the degree of engagement or involvement in conversation, by inviting users to reciprocate and respond. However, they may become problematic if they come at the price of replacing the exchange of arguments tout court.

Table 4: Summary of variables by dimension of deliberative quality (percentage of appearances). The last column includes a summary index of each dimension (excluding plurality).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Variables</th>
<th>Percentage</th>
<th>Index*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity (0-3)</td>
<td>REC1 Mentions</td>
<td>32.8</td>
<td>0.68 (0.81)</td>
</tr>
<tr>
<td></td>
<td>REC2 Quotes</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>REC3 Questions</td>
<td>27.8</td>
<td></td>
</tr>
<tr>
<td>Justification (0-7)</td>
<td>JUS1 EXT Data</td>
<td>13.4</td>
<td>1.53 (1.06)</td>
</tr>
<tr>
<td></td>
<td>JUS2 EXT Facts</td>
<td>62.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JUS3 EXT Comparisons</td>
<td>23.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JUS4 EXT Links</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JUS5 EXT Proposals</td>
<td>22.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JUS6 INT Personal experience</td>
<td>19.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JUS7 INT Values</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>Reflexivity (0-1)</td>
<td>REF1 Moderation</td>
<td>3.3</td>
<td>0.03 (0.18)</td>
</tr>
<tr>
<td>Positive empathy (0-1)</td>
<td>EMP1 Recognition</td>
<td>9.0</td>
<td>0.09 (0.28)</td>
</tr>
<tr>
<td>Negative empathy (0-3)</td>
<td>EMP2 Accusation</td>
<td>34.3</td>
<td>0.49 (0.61)</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>EMP3 Irony</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EMP4 Insults</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>Plurality</td>
<td>PLU1 Approval / disapproval</td>
<td>66 (approval)</td>
<td>31 (disapproval)</td>
</tr>
<tr>
<td></td>
<td>PLU2 Conflict**</td>
<td>69.2</td>
<td></td>
</tr>
</tbody>
</table>

*Mean value and standard deviation (between parentheses).
** Conflict means a reply to a person that holds the opposite opinion on the issue.
5.3. The evolution of deliberative quality

As stated in the hypotheses, the quality of deliberation varies as conversation threads unfold, but indicators evolve differently. The regression analyses (see Table 5 and Figure 5) show that the level of justification decreases as conversations go deeper, whereas the levels of reciprocity and negative empathy both become more important over time. In both cases, the pattern is more similar to what was stated in H3, since the relation is not completely linear (as the coefficient of the quadratic transformation of the variable ‘depth’ shows). Reflexivity and positive empathy have not statistically significant coefficients, and thus they might be read carefully.

Table 5: Regression Models (OLS) between different indicators of deliberative quality and the development of conversations in terms of depth. The model includes the quadratic transformation of the variable depth to test for a non-linear relation. It also contains additional relevant variables such as the number of characters of the posts, whether the post agrees or disagrees with the initial proposal (PLU1) and the existence of conflict (PLU2).

<table>
<thead>
<tr>
<th></th>
<th>Reciprocity</th>
<th>Justification</th>
<th>Reflexivity</th>
<th>Positive Empathy</th>
<th>Negative Empathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>1.486</td>
<td>-1.250</td>
<td>0.125</td>
<td>0.045</td>
<td>1.380</td>
</tr>
<tr>
<td>Depth$^2$</td>
<td>-1.025</td>
<td>0.936</td>
<td>-0.093</td>
<td>-0.086</td>
<td>-1.041</td>
</tr>
<tr>
<td>Number of Characters</td>
<td>0.001</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PLU1 Approval (landlords)</td>
<td>0.014</td>
<td>0.201</td>
<td>-0.044</td>
<td>0.006</td>
<td>0.020</td>
</tr>
<tr>
<td>PLU2 Conflict</td>
<td>0.182</td>
<td>0.315</td>
<td>0.006</td>
<td>-0.156</td>
<td>0.107</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.071</td>
<td>0.701</td>
<td>0.018</td>
<td>0.212</td>
<td>0.035</td>
</tr>
</tbody>
</table>

$R^2$ | 0.089 | 0.364 | 0.016 | 0.073 | 0.054 |

N= 291
(Comments on the initial proposal that have not started a line of discussion have been removed from the models, that is the reason why the final N is slightly below the total number of comments.)
These findings indicate that conversation threads in this debate follow different phases. Initially, users try to present convincing arguments to justify their positions. At this initial moment, most of the efforts are focussed on developing persuasive arguments and providing enough evidence to support them. There are attempts to put oneself in other people’s shoes and make reasonable arguments that could be acceptable by sceptics and even opponents. However, there is a certain general effect of exhaustion through interaction. This can be partly explained because a common pattern for conversations is to start as a statement with a more or less complete justification of the argument, and as the discussion develops, they become more focused on specific details. In addition, the incompatibility of viewpoints and the lack of a common ground can also diminish the emphasis on justification, by transforming the conversation into a more personal and less rational exchange of views in the form of irony, personal attacks and accusations reciprocally responded by each side. This phase of degradation can lead to an insolation of the conversation, making it less appealing for other users to take part in the debate and, as a consequence, reducing the visibility and the necessity to present widely acceptable arguments. The fact that conversation threads are mostly dominated by a few highly active users does not help to keep personal confrontations out of the debate. Reiterative interactions between the same users lead sometimes to a repetition of arguments, even a literal copy and paste of previous posts, transforming the debate into a sort of dialogue of the deaf.
Overall, these findings show that spontaneous deliberation is possible, but they also reveal the difficulties of ensuring and preserving the quality of deliberation over time. The discussion mostly tended towards exposing the reasons for or against the proposal and, despite real attempts to exchange arguments, there were at the end few concessions to either side. Two elements have probably helped to produce this result. On the one hand, the only effect of the debate was the potential impact on the vote; thus, incentives were placed for campaigning to add or subtract supports for the proposal, but not for abandoning maximalist positions and compromising. On the other hand, stakeholders have probably perceived through successive interactions that interests were so antagonistically opposed that no possible common ground was achievable, and that it was not worth to struggle for a mutually agreed solution. Thus, continuous interaction has probably served to reinforce initial positions instead of moderating conflict and reconciling the different viewpoints. At the very least, conversations may have helped to voice the different interests at stake and make visible the diversity of stances and claims.

6. Conclusions

Deliberative theory presupposes that for deliberation to emerge a set of ideal conditions are required. When understood as a continuum, it is possible to ascertain the degree of deliberativeness of a given conversation, by analysing the number of deliberative criteria that are fulfilled. In this paper, we have provided some measurements of the deliberative quality applied to a specific debate, and we have also showed that indicators of quality are not necessarily stable but may evolve differently depending on how conversations develop over time. In that way, we try to bridge the gap between the literature on deliberative quality and the analysis of conversation structures, by focusing on how deliberative quality indicators vary as conversation threads move forward. More or less genuine citizen deliberation can emerge spontaneously in online spaces habilitated for debate - as it was the case of the platform Decidim, but it can be easily eroded too, if there are no favourable conditions. For instance, we find that the justification of arguments becomes less important as conversation threads develop, while certain elements such as negative reciprocity gain further relevance. The identification of these trends may offer useful information for institutions to understand the dynamics of conversations, and provide the right conditions for keeping online debates within good standards of deliberation.

The deliberative erosion found in this conversation can be partly explained because of some problems identified in the institutional design. Even though the platform under analysis seems to work pretty well in enabling citizens to voice their claims, certain additional features are probably needed if conversations are expected to more specifically aim towards deliberation. Social-media like structure, as the one used in the Decidim platform, can facilitate interactions between users in a very intuitive way, and thus might provide a good basis for communication. However, certain requirements - such as active moderation, previous information, some basic regulatory norms and representation of all stakeholders - cannot be simply neglected if deliberation is to be less vulnerable to the randomness of spontaneity, which may or may not lead towards a promising exchange of views and arguments in terms of deliberative quality.
7.- References


### 8.- Appendix

Table 3: Indexes and reliability of the indicators of deliberative criteria.

<table>
<thead>
<tr>
<th>Final index</th>
<th>Indicators</th>
<th>Krippendorff's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity (0-3)</td>
<td>REC1 Mentions</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>REC2 Quotes</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>REC3 Questions</td>
<td>0.791</td>
</tr>
<tr>
<td>Justification (0-7)</td>
<td>JUS1 EXT Data</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>JUS2 EXT Facts</td>
<td>0.715</td>
</tr>
<tr>
<td></td>
<td>JUS3 EXT Comparisons</td>
<td>0.713</td>
</tr>
<tr>
<td></td>
<td>JUS4 EXT Links</td>
<td>0.776</td>
</tr>
<tr>
<td></td>
<td>JUS5 EXT Proposals</td>
<td>0.803</td>
</tr>
<tr>
<td></td>
<td>JUS6 INT Personal experience</td>
<td>0.732</td>
</tr>
<tr>
<td></td>
<td>JUS7 INT Values</td>
<td>0.861</td>
</tr>
<tr>
<td></td>
<td>Short comments (no just.)</td>
<td>0.903</td>
</tr>
<tr>
<td>Reflexivity (0-1)</td>
<td>REF1 Moderation</td>
<td>0.963</td>
</tr>
<tr>
<td>Positive empathy (0-1)</td>
<td>EMP1 Recognition</td>
<td>0.708</td>
</tr>
<tr>
<td>Negative empathy (0-3)</td>
<td>EMP2 Accusation</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>EMP3 Irony</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>EMP4 Insults</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>Repeated comments</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td><strong>Plurality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLU1 Approval / disapproval</td>
<td>0.848</td>
<td></td>
</tr>
<tr>
<td>PLU2 Conflict</td>
<td>Based on PLU 1</td>
<td></td>
</tr>
<tr>
<td>PLU3 Local government</td>
<td>0.733</td>
<td></td>
</tr>
<tr>
<td><strong>Diversity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLU4 Gender</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>PLU5 Language</td>
<td>0.758</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Network of participants in the debate (for simplicity reasons, only comments above level one of depth are represented). The nodes’ size is dimensioned according to the level of degree. Green nodes are participants who favour the proposal, whereas red ones are against it (yellow means undefined positions). Hedges are scaled to reflect the intensity of exchanges; black lines mean interaction between opponents, while red/green lines mean interaction between like-minded users (blue lines imply interactions that involve undefined users).
Figure 3: Radial tree of the debate on the proposal for new licenses for tourist apartments. Green nodes are comments in favour of the proposal and red nodes are comments against the proposal (yellow ones are undefined cases). The central node corresponds to the initial post that defends the proposal. Nodes (N=336) are scaled according to the number of characters of each post. Hedges indicate the line of replies.
Figure 4: Distribution of comments by level of interaction or depth.