Behavior Indicators for Sensemaking of Online Discussions

Livia C Gouvea[1], Ana Cristina B Garcia[1] and Adriana S Vivacqua[2]

Abstract— **Online collaborative environments, such as Reddit, have led to the rise of open and self-organizing communities. However, discussions in these places are often scattered, redundant and voluminous, and do not fully support the deliberation flow. Argument polarization, groupthink, subject bubbles, subgroup dominance and lurking are commonly identified behaviors that negatively impact the quality of online discussions. In addition, open environments allow the congregation of homogeneous groups that feed on pre-existing ideas, creating discussion bubbles and contributing to users taking extreme positions on issues. Evidence from Reddit discussions suggest that, while online collaborative environments provide powerful support for the creation and dissemination of new content, their ability to support efficient debate and idea contribution is questionable. Often, debates include hundreds or thousands of people with heterogeneous behaviors concerning post frequency, size, wording and polarity. Joining an ongoing online debate requires understanding what has happened so far. Many studies have focused on summarizing the discussion content as a way to overcome the message volume, dispersion and redundancy. However, to fully understand an online debate, it is also important to perceive the group's dynamics. The identification of dysfunctional behaviors contributes to new users' sensemaking of the debate as a whole and supports their decision whether it is worth joining the group. Additionally, understanding the group dynamics improves participants behavior and the overall discussion. We analyze the use of content summarization and behavior indicators to understand a discussion.**

I. INTRODUCTION

Online collaboration environments such as social media, forums, wikis and blogs have radically changed the way knowledge is created and managed by society. Ease of use, low cost, wide reach, and freedom of expression have led to the mobilization and aggregation of open, self-organizing communities [7]. Large-scale interaction among users enables the emergence of phenomena such as synergy of ideas, opening of space for many voices (long tails), mass verification of facts (many eyes) and mass evaluation of issues (wisdom of crowds) [10].

However, deliberation in these environments is often scattered and redundant, with little convergence of opinions to support decisions. At the same time, contributions usually present low signal-to-noise ratio, that is, there are many repetitive contributions on the same theme [11]. Another commonly encountered problem is the tendency for users to congregate in like-minded groups, of people who share the same view (also known as balkanization) [2] [16], so that there are few divergent points of view, and contributions become biased and present weak logical argumentation [11]. These very homogeneous environments also contribute to users taking extreme positions on issues (polarization) [6].

As the discussion unfolds, understanding it may become increasingly difficult for newcomers, a problem that tends to increase with group size [3]. In addition, it has been observed that technologies for sharing and collaboration are now being used to manipulate public opinion and spread fake news stories, which are successfully employed by extreme political organizations to promote their cause and recruit new supporters [7]. Comments in social networks tend to be sparse, noisy and long. Users frequently raise a number of subtopics related to the main theme under discussion. Consequently, exploring these comments to distill sub-topics and summarize comments is a hard task [17]. These elements can make it hard for a newcomer to join an ongoing conversation, preventing other voices from being heard.

While online collaboration tools and environments are powerful in supporting the creation and dissemination of new content, Joining a group discussion requires breaking the inertia of understanding what has been going on so far, which in turn requires additional work. There are few initiatives geared towards supporting the sensemaking process for an ongoing discussion. Thus, the ability to promote a fair and transparent debate in these environments is questionable.

Organizing discussions can be a way to contribute to reducing the entry barrier for new users, encouraging them to join discussions. It might also make it easier to identify problems such as balkanization, polarization and redundancy. Most effort has been on the summarization of discussion content [8] [9]. However, the group's interaction dynamics also play an important role to get a full picture of the debate. In this paper, we present an approach that combines natural language processing and information visualization techniques to produce a visual summary that contributes to create collaboration environments and online deliberations more efficient in their purpose of exchanging ideas and less scary for newcomers joining conversations.

II. BACKGROUND

A. Argument-Centered Systems

One of the first Collaborative Computer-Supported Argument Visualization (CCSAV) approaches was the Issue-based information system (IBIS) [12]. IBIS seeks to alleviate problems in time-based deliberation tools (such as email, chat rooms or discussion forums) and also handle topic-based discussions, such as wikis and idea-sharing systems

[1]Livia C Gouvea and Ana Cristina B Garcia are with Department of Information Systems, Federal University of the State of Rio de Janeiro, Rio de Janeiro, RJ, Brazil livia.faria@uniriotec.br, cristina.bicharra@uniriotec.br

[2]Adriana S Vivacqua is with the Department of Computer Science, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil avivacqua@dcc.ufrj.br

([4], [7], [10], [13]). The IBIS approach proposes the resolution of complex problems by mapping a deliberation as a combination of 4 elements:

1) Issue to be solved (Issue);
2) Idea for resolution (Idea);
3) Argument in favor of the idea (Pro);
4) Argument against the idea (Con).

However, an analysis of current online discussion environments, shows that they have characteristics of time-based, topic-based and argument-based systems, not fitting into a single category. Examples are: the possibility of interacting directly with posts either by expressing agreement/disagreement (like/dislike) or responding to a specific comment within the topic, and the possibility of organizing the discussion according to different attributes, such as, date, relevance, type, polarity or even by the degree of controversy.

Empirical studies have shown that conversation-centered sharing tools such as wikis, blog forums and other social media are preferred by users over argument-centered tools [5]. They point out that argument-centered tools use communication formats that can be very restrictive and detrimental to user interaction.

B. Organization of Discussion Content

Content organization proposals target online deliberation environments that are based on arguments or argumentative maps. The main goal is to make the collaboration more efficient. Alves [1] proposed a mechanism to provide diagnoses of dysfunctional behaviors such as groupthink, polarization and balkanization in order to help mediators identify problems in discussions, performing a spatial representation to facilitate identification. Romeiro [15] used Rhetoric Structure Theory (RST) to perform automatic summarization of content to assist the entrance and participation of new users in deliberation environments that use argumentative maps.

However, no work was found that applies summarization to contents of online deliberation environments and combines this with visual explanations. Reddit is an online discussion platform that, according to the Alexia.com ranking, is among the 10 most visited sites in the USA. It is an ever evolving system that brings together argument-based and conversation-based features. We propose to analyze content from the Reddit database to generate statistics that may contribute to a better understanding of the content as well as the identification of dysfunctional behaviors in topics of deliberation, and generate visualizations to assist newcomers when joining the conversation.

C. Text Summarization

Zhang and Setty [17] emphasize that comments related to the topics are conversational in style, so that they touch on several aspects of the same topic and can lead to deviations of topics. Therefore, it is necessary to distill the implicit topics within the main topic and select the representative comments for each of these “subtopics”. In their solution, the authors used elastic search with the Okapi BM 25 probabilistic recovery structure to retrieve and classify the 10 most relevant discussion topics (and their respective comments) from Reddit, based on the pseudo-relevance score, such as a pseudo-result of the survey. They then pre-processed the comments to extract the interpretable threads from each query pseudo-result. Finally, they used the Sainte-Laguë (SL) diversification method to reduce redundancy and diversify search results in a topical and sentimental way, and the Comment Tree Decomposition Method (CTD), which is particular for comments in tree structure to reduce redundancy by maintaining the style of coherence and conversation of comment trees.

Ren et al. [14] emphasize that, in discussion forums, the best evaluated comments are displayed prominently, before the others. Thus, in topics with many posts, people usually read only the featured posts, ignoring the rest of the contributions. Given this, this display style reflects an incomplete view of the conversation. The authors argue that with the increase in the volume of posts in the topic, it becomes necessary to summarize the discussion to provide the user with a complete view of the discussion. Therefore, the use of document summarization techniques for the summarization of discussion threads seems convenient. However, in view of the complicated and dynamic characteristics of the forums, the traditional document summarization algorithms were not adequate to the three essential characteristics for the summarization of discussion topics:

1) topic dependencies: comments are linked to each other in the form of conversation
2) topic drifting: a topic leads to several sub-topics that may differ
3) text sparseness: most posts are composed of short messages, often filled with ellipsis. As short texts do not provide sufficient co-occurrence information, traditional text representation methods, such as ”tf-idf”, have several limitations when applied directly to mining tasks [6].

To take in all the peculiarities of a discussion forum, authors developed a new topic summarization model called Posts Propagation Model (PPM). The solution used a Dirichlet tree to connect topic dependencies based on the response relationships between comments.

III. RELATED WORK

The combination of *COLLAGREE* and *Deliberatorium* projects is a proposed solution as the two main causes for large-scale deliberations. One is that the tools are not successful at extracting good ideas from major deliberations, because in these contexts, participants tend to impose their ideas rather than debating and constructing better solutions in a collaborative way. The second reason pointed out by the authors is that, although the bargaining in deliberations is able to solve simple problems, it is not able to reach consensus on complex problems[19]. The *COLLAGREE* tool utilized feelings analysis and word cloud from deliberation to support facilitators to manage large-scale discussions. ([20]). *Deliberatorim* is a tool for constructing argumentative maps where contributions are built into a tree structure consisting

of questions to be solved, possible solutions to these questions and arguments that support or refute the solutions. The map structure facilitates the organization of the content and avoids redundancy. The proposed for junction of the two systems presents an iterative cycle where participants can begin with free text discussions, then the arguments of this discussion are organized on a map of deliberation by the crowd itself. Idea filtering algorithms are then used to identify the most promising solutions ideas generated and these ideas become the starting points for consensus processes mediated by nonlinear negotiation algorithms. Alongside this, a set of alerts, metrics, and reports are generated from the deliberation data as a form of deliberative analysis to help participants gain a clearer understanding of deliberation..

The survey conducted by [21] shows how data mining can identify characteristics of behavior in social networks. The paper classified social media data in three types of entities: network, geographic information and text. The first type is the social network of the users and the network of information diffusion, constructed by the behaviors of follow and repost. The second type is the diffusion of information in space and time and the distribution of events, as well as the movement constructed from the geo-marked messages. The last type are keywords, topics and feelings derived from content in social media. The work summarized social data representation in six categories: visual monitoring, pattern extraction, event detection, anomaly detection, predictive analysis and situation awareness. The authors point out that monitoring of data provides a quick overview of monitored targets and allows the identification of discrepant patterns and values the same time as provide a representation that allows awareness of the situation.

IV. METHOD

To conduct our research, we extracted a set of discussions from Reddit, as described in Table I.

A. Data Selection

1) Selection period: To perform this work, we chose to use the most recent available data. Considering the database contains records from October 2007 to May 2015, we selected only records from the year 2015.

2) Topic selection: The discussion topics on Reddit are divided into communities such as politics, sport, funny, etc. For our research, we chose to work with the political community, based on the idea that political discussions are more likely to contain dysfunctional behaviors such polarization, balkanization, etc., which are of interest to us. We retrieved 30.017 discussion topics with specific characteristics, as described in Table II. In average, 17 people participated in the discussions which lasted 1 day posting messages with 256 characters.

B. Exploratory Analysis

A preliminary analysis of the discussion data evidenced the presence of the dysfunctional behaviors pointed out in the literature and described in the introduction of this paper.

TABLE I

VARIABLES EXTRACTED FROM THE REDDIT DATABASE

Variable Name	Description
retrieved_on	Variable that identifies the date that was extracted from the site archived. The site allows users to archive their topics. Archived topics can not be answered or voted on, only queried. This variable identifies whether the topic is archived or not.
controversiality	Comments that contain many positive and negative votes are defined as controversial. This variable indicates whether the comment is controversial or not.
edited	Variable that records whether the comment was edited.
parent_id	Identifies who the comment responds to. The comment may be responding to the initial topic or to another comment.
author_flair_text	Flair is a type of tag that can be added to posts. They are often used to help readers filter out a particular type of post. This variable displays the comment flair.
score_hidden	Moderators can hide the score of new posts and comments for a certain period of time, up to 24 hours maximum. This feature was introduced to discourage the effect it commands, when a comment receives negative or positive feedback initially it usually continue to receive votes of the same type. This variable identifies whether the comment's score is hidden.
gilded	Sets whether a comment received a gold. Gold is a paid feature from Reddit that offers some perks to users. A user can offer these perks to another by assigning gold to his comment.
created_utc	Records the date the comment was created
link_id	Identifies which discussion thread this comment is associated with.
subreddit	Reddit topics are grouped by communities like politics, cinema, etc. This variable holds the record of which community the comment belongs to.
author_flair_css_class	Another type of flair is the user's, a text or visual tag associated with a username. Some more serious topics use to provide additional context about the author's expertise comment but most topics use it to allow users to select something to represent them.
score	The comments can be positively evaluated "up vote" and negatively "down vote". This variable displays the total votes.
downs	Variable that records the total of "Down votes".
ups	Variable that records the total of "Up votes".
id	Comment identifier.
body	Comment text.
distinguished	Reddit moderators may assign the "distinguished" symbol to their comments as a way to show that comment is from an official Reddit moderator. This variable records whether the comment was assigned as "distinguished."
subreddit_id	All discussion topics belong to a subreddit that is like a community. This variable registers the comment's subreddit id.
name	Name of comment that is the id preceded by "t1_"

TABLE II
DATASET DESCRIPTION

Variable	Average	Minimum	Maximum
Number of Participants	17	1	4.041
post size (characters)	256	3	10.797
duration (days)	1	1	133

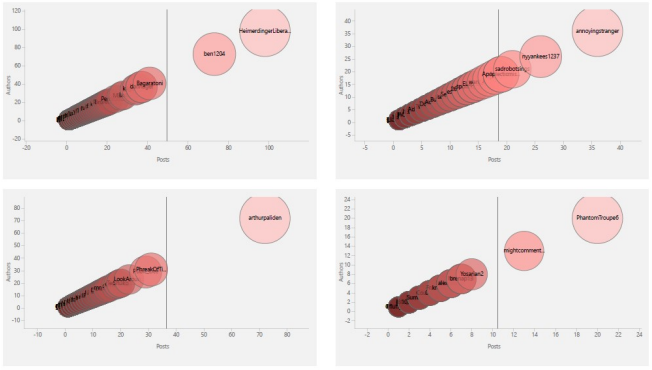

Fig. 1. Number of posts per author in 4 discussion topics

Subject bubbles, dominance of speakers and disengagement of users were identified. Before generating visualizations, two changes were made in the data:

1) For the generation of the graph in Fig.1, which represents the number of posts by authors, the posts of whose authors had deleted their accounts or have been banned by moderators were removed because they were being counted as one author, causing an imbalance in the graph.
2) For the generation of the graph in Fig.2, of proportion of replies to comments, the initial posts in discussions were removed, because these posts were the most replied to. This was impacting in the visualization of the proportion of responses for other posts of the discussion.

From this selection of data, it was possible to generate some visualizations for exploratory analysis of the data as well as reports that show the existence of the dysfunctional behaviors mentioned above. The bubble charts in Fig.1 show the volume of posts per author on four randomly selected discussion topics. It can be observed that one or two authors speak significantly more than the rest of the participants, what could be an evidence of dominance of these authors in the discussion. Also, Fig.2 shows in bubble charts the volume of responses to the posts in four randomly selected discussion topics. From these graphs, it can be seen that a few posts are the focus of the discussion, what could be a representation of concentration of the discussion in bubbles.

The scatter plot in Fig.3 shows that the number of posts in a discussion increases proportionally with the number of users. Another scatter plot shown in Fig.4 shows that the controversy variable maintains a relation of proportion to the size of the discussion, that is, the larger the discussion,

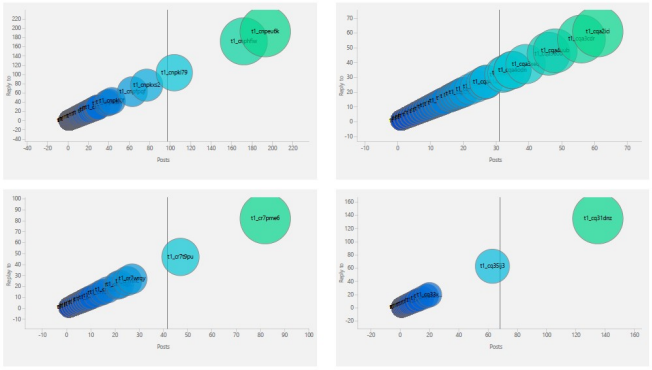

Fig. 2. Volume of replies to a comment in thread topics

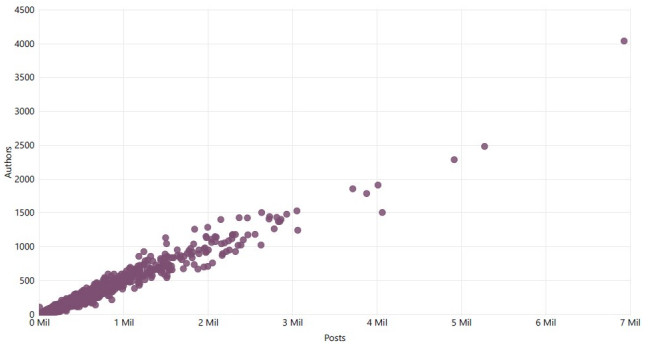

Fig. 3. Number of authors by size of the discussion

the more controversial it becomes.

Fig.5 shows in bar graphs the behavior over time for four randomly selected discussions. From this graph, it can be observed that the discussions start heated and with passing time, they lose intensity.

The correlation matrix in Fig.6 of the variables shows that the post size and punctuation variables have a strong correlation. From this view, one can also observe a strong relationship between the score of the post and the amount of responses to it. Another relationship focuses between the size of the post and its controversy.

C. Resulting Design recommendations

Based on our analysis, we identified important information concerning the group interaction that should be available to a newcomer to make sense of an ongoing discussion. We identified a set of descriptive and diagnostic variables. We believe this information concerning the group dynamics is important and should be shared with a newcomer.

We propose to present to the new user a description of the debate containing basic information, such as:

- number of participants
- starting date
- number of posts so far
- average posts size (number of characters)
- average frequency of posts per day
- average posts per participant

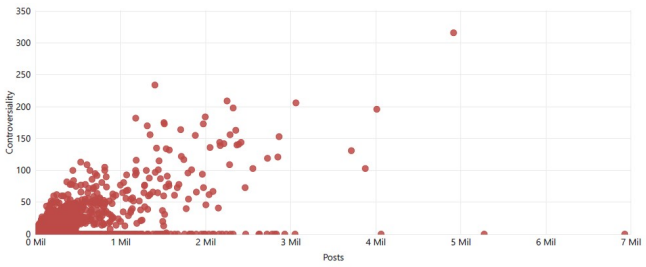

Fig. 4. Controversy by size of the discussion

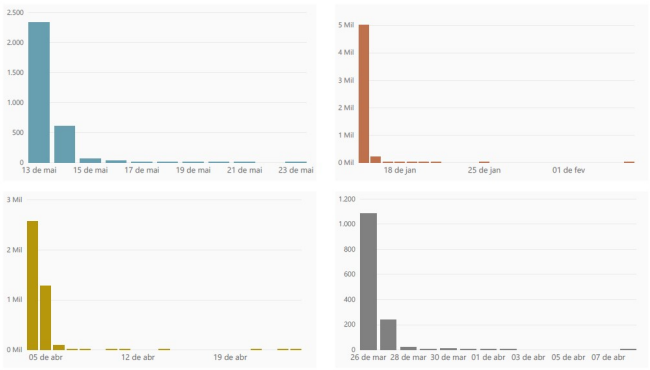

Fig. 5. Evolution of discussions over time

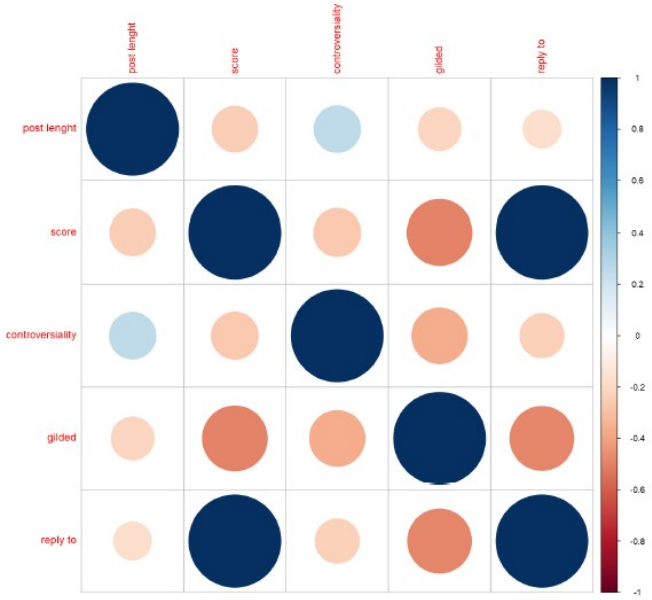

Fig. 6. Variable correlation matrix

In addition, it is important to provide information concerning the interaction dynamics of participants including information such as:

- Participant dominance: based on the average number of posts each participant inserted, identify if there are a few participants that dominated the discussion. Dominance is specified as number of posts greater than the average number of posts plus 10 times the standard deviation.

$$\frac{\textit{ParticipantDominance} : \exists\, \textit{Partic posts too much}}{\begin{array}{l}\forall\, i\ P_{max} = max(Posts(participant_i)) \\ \textbf{if } P_{max} \geq \mu_{posts} + 10 \times \sigma_{posts} \\ \qquad \textit{ThenParticipant}(P_{max})\ \textit{dominates the debate}\end{array}}$$

- Post Dominance: similar to participant dominance, this variable identifies posts that attracted most replies.

$$\frac{\textit{PostDominance} : \exists\, \textit{Post too many replies}}{\begin{array}{l}\forall\, i\ P_{max} = max(Replies(post_i)) \\ \textbf{if } P_{max} \geq \mu_{posts} + 10 \times \sigma_{posts} \\ \qquad \textit{ThenPost}(P_{max})\ \textit{dominates the debate}\end{array}}$$

- Degree of Controversy: this indicates how hot the debate is. It is calculated as the average of the posts' controversy variable (available in the Reddit data).

$$\frac{\textit{DegreeOfControv} :\ \textit{Most controversial post}}{\forall\, i\ \textit{DegreeOfControv} = max(Controv(Post_i))}$$

- Groupthink: this is an inference indicator. The value is a probably yes OR probably No. We say the debate tends to groupthink if most posts and supports goes forward one single idea. Posts that reply to posts with the most Likes are an indication of support.

$$\frac{\textit{Groupthink} : \forall\, \textit{Participants} \simeq \textit{Idea}}{\begin{array}{l}\exists\, Post_j, \forall\, Post_i : Post_i\ HangsOn\ Post_j \wedge \\ \qquad Sentiment(Post_i, Post_j) = + \in S\end{array}}$$

- Polarization: indicates the debate has two subgroups with opposite polarity.

$$\frac{\begin{array}{l}\textit{Polarization} : \forall\, \textit{Participants} \\ \qquad\qquad Support(Idea_1) \oplus support(Idea_2)\end{array}}{\begin{array}{l}K = Cluster(Posts) \\ \textbf{if } K = 2 \\ \textit{Then} \\ \qquad Cluster_1 = Posts(Group_1) \\ \qquad Cluster_2 = Posts(Group_2) \\ \qquad \textbf{if } Polarity(Cluster_1, Idea_2) = Negative \wedge \\ \qquad\qquad Polarity(Cluster_2, Idea_1) = Negative \\ \qquad\qquad \textit{Then Polarization is True}\end{array}}$$

- Balkanization: It is calculated similarly to Polarization, but k is greater than 5.
- Post frequency decay: It indicates the speed the debate is dying.

Figure 7 presents the prototype interface with the debate information to assist newcomers to make sense of an ongoing debate. We believe this debate summary will also help improve discussion quality, avoiding fallacies and helping participants reflect on their behavior as a group.

V. CONCLUSIONS

It is often hard for newcomers to join ongoing discussions, especially when they have been going on for a while.

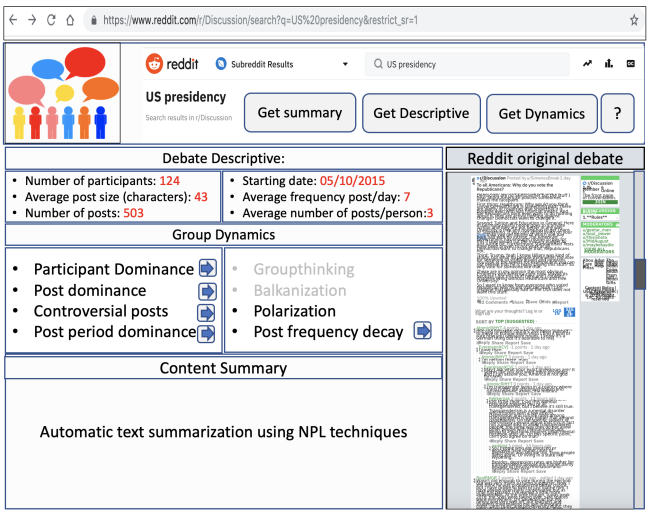

Fig. 7. Debate Explanation Interface: in this example, the debate is polarized and neither groupthink nor balkanization behavior were identified. Arrows are buttons for specific information concerning the feature

Discussion length, large number of posts and distinct arguments can make it hard for someone entering the discussion halfway through to make an effective point. In addition, group dynamics can make it even harder to be heard and understood, or to put one's opinion forward. Knowing that a discussion is dominated by a single individual, leans heavily towards a certain point of view or is heavily polarized, with opinions in different extremes and unlikely to change, is important for a newcomer when considering the best way to express his/her ideas, and deciding whether or not to join the discussion at all.

In a heavily polarized discussion, for instance, a newcomer (or even a current participant) might go through the discussion and consider whether a new line of argumentation can be made. In a discussion dominated by an individual, it might be useful to counter that individual's arguments, challenging his/her leadership. Different strategies can be adopted when entering a discussion, depending on what the participant's goals are. Visualization techniques have proven useful in multiple situations to help individuals interpret and understand large quantities of data. Thus, we adopt visual representations of relevant variables to inform a user's decisions. We expect the visual overview, coupled with text processing techniques will be a useful tool for sensemaking of discussions, both for newcomers and current participants.

ACKNOWLEDGMENT

This work was supported by CAPES, FAPERJ and CNPq.

REFERENCES

[1] Alves, B. (2016). Uma Abordagem Para Identificação De Problemas Em Discussões Estruturadas Por Mapas Argumentativos.

[2] Dishon, G., & Ben-Porath, S. (2018). Don'@t me: rethinking digital civility online and in school. Learning, Media and Technology, 43(4), 434-450.

[3] Garcia, A. C. Klein, M. (2015). Making Sense of Large-Group Discussions using Rhetorically Structured Text. Sotics, (c), 26–35.

[4] Gürkan, A., Iandoli, L., Klein, M., Zollo, G. (2010). Mediating debate through on-line large-scale argumentation: Evidence from the field. Information Sciences, 180(19), 3686–3702. https://doi.org/10.1016/j.ins.2010.06.011

[5] Iandoli, L., Quinto, I., De Liddo, A., Buckingham Shum, S. (2016). On Online Collaboration and Construction of Shared Knowledge: Assessing Mediation Capability in Computer Supported Argument Visualization Tools Luca. Journal of the Association for Information Science and Technology, 3(2), 80–90. https://doi.org/10.1002/asi

[6] Iandoli, L., Klein, M., Zollo, G. (2009). Enabling On-Line Deliberation and Collective Decision-Making through Large-Scale Argumentation: A New Approach to the Design of an Internet-Based Mass Collaboration Platform. International Journal of Decision Support System Technology, 1(1), 69–92. https://doi.org/10.4018/jdsst.2009010105

[7] Iandoli, L., Quinto, I., Spada, P., Klein, M., Calabretta, R. (2018). Supporting argumentation in online political debate: Evidence from an experiment of collective deliberation. New Media and Society, 20(4), 1320–1341. https://doi.org/10.1177/1461444817691509

[8] Kano, R., Miura, Y., Taniguchi, M., Chen, Y. Y., Chen, F., Ohkuma, T. (2018). Harnessing Popularity in Social Media for Extractive Summarization of Online Conversations. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 1139-1145).

[9] Kim, B., Kim, H., Kim, G. (2018). Abstractive Summarization of Reddit Posts with Multi-level Memory Networks. arXiv preprint arXiv:1811.00783.

[10] Klein, M. (2012). Enabling large-scale deliberation using attention-mediation metrics. Computer Supported Cooperative Work, 21(4–5), 449–473. https://doi.org/10.1007/s10606-012-9156-4.

[11] Klein, M., Iandoli, L. (2008). Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium. Directions and Implications of Advanced Computing; Conference on Online Deliberation. University of California, Berkeley.

[12] Kunz, W., Rittel, H. W. J. (1970). Issues as elements of information systems. Book, 131(131), 10. https://doi.org/10.4324/9780203851586

[13] Quinto, I. (n.d.). A Debate Dashboard to Enhance On-Line Knowledge Sharing Simon Buckingham Shum Anna De Liddo Luca Iandoli, 1–15.

[14] Ren, Z., Ma, J., Wang, S., Liu, Y. (2011). Summarizing web forum threads based on a latent topic propagation process. Proceedings of the 20th ACM International Conference on Information and Knowledge Management - CIKM '11, 879–884. https://doi.org/10.1145/2063576.2063703

[15] Romeiro, A. K. Q. (2007). Um Estudo Sobre O Uso Da Teoria Da Sumarizar A Sabedoria Da Coletividade.

[16] Zhang, J., Danescu-Niculescu-Mizil, C., Sauper, C., & Taylor, S. J. (2018). Characterizing Online Public Discussions through Patterns of Participant Interactions. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 198.

[17] Zhang, H., Setty, V. (2016). Finding diverse needles in a haystack of comments. In Proceedings of the 8th ACM Conference on Web Science - WebSci '16. https://doi.org/10.1145/2908131.2908168

[18] Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., Li, X. (2011). Comparing Twitter and Traditional Media Using Topic Models, 338–349. https://doi.org/10.1016/0952-8180(89)90027-5

[19] Ito, T., Imi,Y., Ito,T., Hideshima,E. (2014) COLLAGREE: Facilitator-mediated Large-scale Consensus Support System. In Collective Intelligence 2014, Poster Sessions. June 10-12, 2014. MIT, Cambridge, USA.

[20] Fujita, K., Ito, T., Klein, M. (2017) Enabling Large Scale Deliberation using Ideation and Negotiation-Support Agents. In 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW) (360-363).

[21] Chen, S., Lin, L., Yuan, X. (2017) Social media visual analytics. In Computer Graphics Forum (Vol. 36, No. 3, pp. 563-587).