That's so meta! Usability of a Hypergraph-based Discussion Model

Felix Dietze[1], André Calero Valdez[1], Johannes Karoff[2], Christoph Greven[3], Ulrik Schroeder[3], and Martina Ziefle[1]

[1]Human-Computer Interaction Center, Campus-Boulevard 57, RWTH Aachen University, Germany
{dietze,calero-valdez,ziefle}@comm.rwth-aachen.de
[2]{johannes.karoff}@.rwth-aachen.de
[3]Learning Technologies Group, Ahornstr. 55, RWTH Aachen University, Germany
{schroeder,greven}@cs.rwth-aachen.de

Abstract. Massive online communication systems such as social networks, message boards and comment sections are widely used, yet fail in conveying a diverse public opinion. Limitations of models and protocols do not allow users to precisely express their intention and to maintain a complete overview in large-scale discussions. Data-driven approaches fail as well, as they remove the nuances of human communication and use coarse representations like trends, summaries and abstract visualizations. We argue that a new discussion model and a large-scale communication protocol is needed. We evaluate the comprehensibility of a hyperedge connection in modeling arguments for online discussions. An initial mechanical turk study ($n = 200$) revealed that 30% of the subjects intuitively considered using hyperedges. This was followed by a user study of a prototype ($n = 51$), where 80% actively used hyperedges. Both findings were independent of user diversity factors (age, gender, graph theory knowledge). The prototypical implementation was evaluated positively.

Keywords: Online discussion systems; argument mapping; living document; hypergraph; user study; e-democracy

1 Failure to communicate

Mass-media, Internet, social media and higher mobility have brought the world closer together by increasing modes and quantity of communication. In particular digitally-mitigated communication has allowed real-time communication across the globe not just between individuals, but also between different and novel versions of public spaces. Traditionally only mass-media like television and newspapers had the opportunity to *broadcast* information. With public spaces such as facebook, twitter and online message boards everyone has—in theory—gained access to broadcasting media (as in twitch, facebook-live, etc.). This new form of communication, where everyone may communicate with everyone, has the potential to free access to information, publicity and opinions.

Early online mass-communication consisted of forums and chats. Both accumulate a chronologically ordered sequence of text pieces, readable by every participant of a discussion. Participants did not have to be in the same room, and therefore more people could discuss and collaborate online.

Scaling online collaboration to whole societies brings up the concept of e-democracy. In general, there is a trade-off between group size and depth of argument. Many people can collaboratively make a decision only by voting, while small groups can engage in profound discussions. E-Democracy aims at finding solutions for overcoming this trade-off [1].

One approach to deal with the increasing amount of information is to try to extract opinions and summaries via text mining. But the current state only allows for rough summaries, which in the end does not help the individual to participate. However, a structured discussion model could elevate information extraction capabilities.

2   Related Work

Quite a large body of research is relevant to this article. We try to limit the related work to what is relevant for understanding the approach in this paper.

When reading continuous text, the argument structure needs to be inferred linearly through the text. Faridani et al. [2] describe that comment lists do not scale and reinforce extreme opinions. They present a user interface called *Opinion Space* which visualizes comments based on different ratings and compare it to a list and grid interface. They confirm that users like their grid and space interface more than a list interface to navigate.

Studies have shown the benefits of working with *argument maps*, as the critical thinking ability of students increases significantly [3] and also their recall of arguments [4,5,6]. The idea of structuring argument for analysis and transparency is rather old, e.g., the model of Toulmin [7] for argument analysis or IBIS [8] for tackling wicked problems. The concept of hyperedges is also addressed by Toulmin and SIBYL [9], yet never fully investigated from a users perspective. Also, more modern implementations without hyperedges exist such as DebateGraph, which was actively used by The Independent newspaper and the White House[1]. Cosley et al. [10] find that oversight increased both the quantity and quality of contributions while reducing antisocial behavior, another benefit of argument maps.

Van Gelder argues, that software like *Rationale* is more useful for argument mapping than word processors, simply, because it was explicitly designed for that task and complements strengths and weaknesses of cognitive capabilities [11]. This strengthens the argument by Davies [5], who argues that argument mapping leads to higher information retention.

Fu et al. compare the *usability of indented tree and graph visualizations* of ontologies. They find that tree visualization is more approachable and familiar

---

[1] https://www.whitehouse.gov/blog/2009/06/05/open-government-brainstorm-collaboration-action

for novice users. Other subjects reported the graph visualization to be more tractable and intuitive, because of less visual redundancy, especially for ontologies with multiple inheritance [12]. Additionally, Fu et al. study the usability with eye-tracking and find that indented lists are more efficient at supporting information searches while graphs are more efficient at supporting information processing [13].

Google Wave was an approach to address the problems that arose with email communication [14]. It models conversations as *living documents*, where users reply inline and can change their written content at any time, similar to the ideas proposed by Sumner & Shum [15].

3   New Requirements

We think that a tool to actually scale online discussions in the number of participants is needed for project teams and democracy. Our idea to create such a tool is twofold.

1) Create a data model which is able to model human communication in a manner that is as useful as possible. At the same time, this model should reflect the mental model of participants. Users should be able to intuitively express themselves regarding other people's contributions. Content should consist of atomic pieces of information to allow precise referencing.

2) Create a protocol for participants to develop and improve the current state of discussion as a living document [16]. This includes removing outdated and unnecessary content collaboratively. This is the opposite of traditional discussion protocols, where contributions can only be appended to existing, immutable content.

Our conjecture is that the combination of an expressive data model with a collaborative moderation system allows to break out of the classic model of online communication and therefore scale better in the number of participants. In such a system a new kind of interaction could emerge, where participants collaboratively develop the current state of discussion instead of just lining up pieces of text. This current state could be easily determined by readers as well as new participants to enable immediate contribution.

3.1   Our contribution

In this work we address the first question of finding a suitable data structure which approximates the expressiveness of human communication as closely as possible, while still being usable for its participants.

We propose an unconstrained hypergraph-based discussion model and a user interface to modify and interact with the discussion. Our proposed model is not completely new, but cherrypicks concepts of both argument mapping models and internet forums (e.g. Toulmin, IBIS, reddit, etc.). In an initial mechanical turk study, we asked participants where they would connect an argument to an existing discussion (see Fig. 2). To verify the results and investigate the impact

of our interface, we replicated the study in the lab. Prior user studies were used to fix major usability issues, allowing us to improve our system and focus the evaluation on our model. From the questionnaire based Mechanical turk (mturk) study, we can measure the intuitiveness of the hypergraph model itself. The lab study allows to reason about the acceptance of hyperedges while actively using the prototype implementation. However, this paper does not evaluate scalability, it merely looks into comprehension of a new connection type.

4   Generalizing Discussion Topologies

When a discussion participant cannot explicitly express his intention within a discussion model, the semantics and relation to other contributions can only be described in the unstructured text field. If more text creates higher cognitive load, the barrier to read and contribute will thereby be raised.
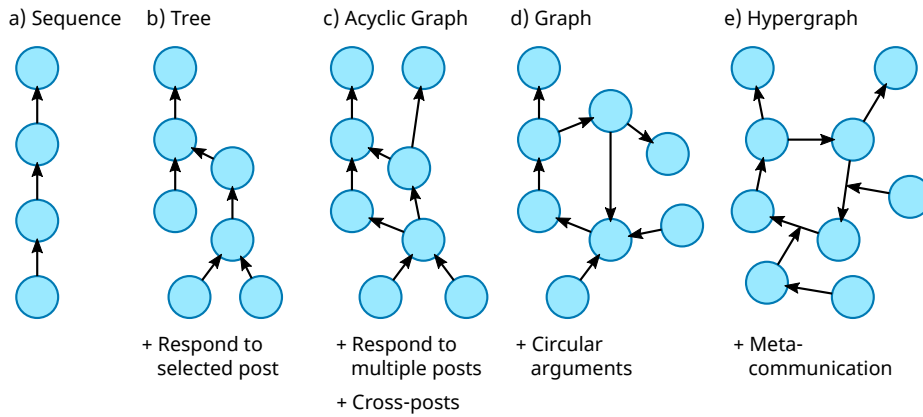

Fig. 1. A sequence (a) of posts corresponds to a protocol of spoken language and has no semantic structure. A tree (b) models a *responds-to*-relation to one parent post, a directed acyclic graph (c) to multiple parents and posts of other threads. A graph with cycles (d) allows to model circular arguments, while a hypergraph (e) allows to model meta-communication.

Typical online conversations are modeled as sequences of posts sorted by creation time (chats, threaded forums, see Fig. 1a). Such a protocol has no semantic structure. Referring to a specific post can only be achieved by quoting, thus inducing redundancy.

Tree based models, such as reddit, make use of a *responds-to*-relation between posts (Fig. 1b), which eliminates the need to repeat content. Still, the tree model forces users to post the argument twice if it applies to two different positions, which creates redundancy.

The tree topology can be generalized as a directed acyclic graph (DAG), allowing redundancy-free posts responding to multiple posts within and across

separate discussions (Fig. 1c). E.g., the idea of driving by bike might be an answer to two different questions. Directed graphs with cycles can additionally model *circular arguments* or feedback-loops (Fig. 1d).

Hypergraphs[2] can model a relation between an arbitrary number of posts. This allows to model *meta-communication* by responding to a connection between two posts, which models the act of communication (Fig. 1e). Technically, meta-communication does not require hypergraphs, but using our type of model, which links meta-communication to its referent, simplifies deixis, and thus reduces redundancy from quoting, which is typically used in meta-communication.

4.1 Proposed Discussion Model

To allow users to precisely express their intention and to avoid redundancy in discussions, we propose a hypergraph-based discussion model. Here, posts are the vertices of the graph, which consist of a mandatory title and an optional (more detailed) description. The title is used to visualize many posts in a limited amount of space. This should also motivate participants to split their contribution into separate units with distinct meaning, which increases interactivity [17]. Posts can be connected with directed edges in a *responds-to* semantic. We use the properties of hypergraphs to model cross-posts, circular arguments, and meta-communication.

Depending on context, the correct entry point to a discussion-graph may be ambiguous. Therefore, we use tags to label entry-points. A tag defines a topic and accumulates relevant conversations introducing the concept of abstraction to deal with the complexity of big discussions.

5   Method

In order to understand whether users would use a protocol and model proposed by us, we decided to conduct a two-part study. We first start with a mechanical turk study investigating how users would connect a meta-communication argument to a graph-based visualization ($n = 200$). We then let users use our prototypical implementation and ask the same question about where to connect a meta-communication argument in a graph-based visualization ($n = 51$).

5.1 Mechanical Turk Study

The mturk study was designed to capture the opinion of non-informed users. The survey was designed to be as short as possible. We asked for the users' age, gender, graph theory knowledge (GTK) and hypergraph theory knowledge

---

[2] A hypergraph consists of a set of vertices and a set of hyperedges. A hyperedge, in contrast to a normal edge, can connect an arbitrary number of vertices. Hypergraphs can be generalized by additionally allowing edges to point at other edges—instead of only nodes. We refer to hypergraphs, even though—in a strict mathematical sense—we are talking about generalized hypergraphs.
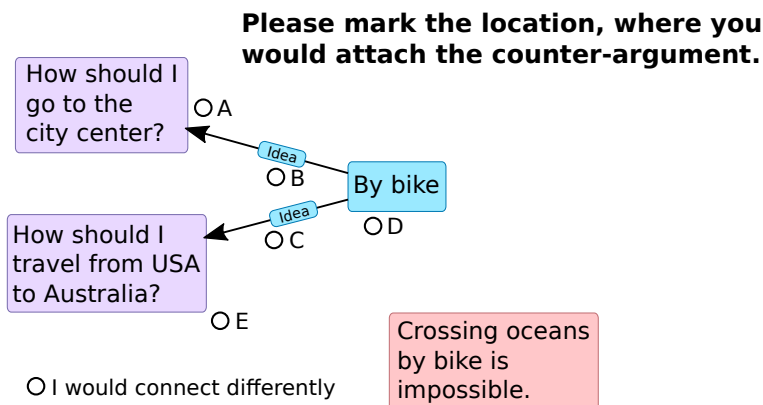
Please mark the location, where you would attach the counter-argument.

How should I go to the city center? ○A

Idea

○B By bike

Idea ○D

How should I travel from USA to Australia? ○C

○E

○ I would connect differently

Crossing oceans by bike is impossible.

Fig. 2. Task description and six possible answers. By clicking an option the inserted edge was visualized.

(HGTK). The compensation for the worker was set to 0.06$. The compensation was chosen to ensure an hourly rate of approx. 8.50$. GTK and HGTK were measured by asking the familiarity of graph theory concepts on a six-point Likert Scale (1=very unfamiliar, 6=very familiar).

The main task in the mturk study was for users to attach the (meta-communication) argument "Crossing oceans by bike is impossible" to the argument graph shown in Figure 2. According to our protocol the correct choice would be option C. Thus the experiment aims to measure how users intuitively attach an argument that does not addresses an idea directly (i.e. a bike is a valuable method of transportation), but its relation to a specific question (i.e. a bike is not a valuable method for crossing an ocean, as suggested in the graph).

5.2 Lab Study

To evaluate the discussion model and the corresponding user interface, we built an interactive website for our prototypical discussion platform. The prototype is based on the concepts described in the previous section. It supports multi-user realtime collaborative editing of discussions in a graph-based visualization.

The prototype was built using Scala [18], the graph database neo4j with renesca [19], AngularJS and D3 [20]. The implementation was iteratively improved in two iterations with nine users to ensure that usability was no major hindrance in the actual experiment.

The goal of the lab study was to see whether using a graph-based discussion system would affect how users would attach a meta-communication argument in a later task.

We recruited 51 users from the authors' social networks and invited them to a lab study. Users were asked the same demographic questions as in the mturk study (age, gender, GTK, HGTK). After completing some tasks in the

graph-based discussion system, we asked the users the same meta-communication question: "Where would you attach the following argument?". Furthermore, we assessed usability of the prototype using the System Usability Scale (SUS).
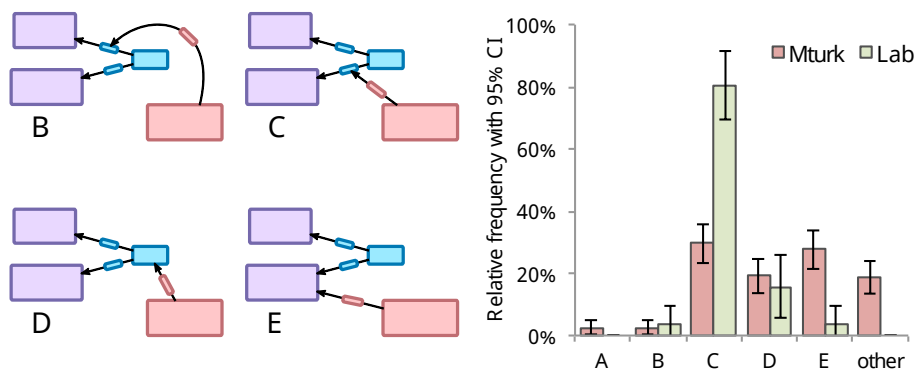
6   Results

Fig. 3. Relative frequencies of participants that selected a specific connection for the mturk and lab study. Option A was omitted, see Fig. 2.

We report data as descriptive statistics and 95% confidence intervals when comparing between subjects. We use χ^2-tests to measure effects of categorial variables.

6.1   Mechanical Turk study

From the mechanical turk study we see that the largest part of the sample wants to map the argument as a hyperedge (C, $n = 59$). The second largest group ($n = 55$) attaches the argument to the question (E). Attaching the argument to the answer and other options were chosen similarly often (see Fig. 3)

When looking at the cats eyes plots of the measured demographic factors (see Fig. 4), we see that no differences in the demographics are evident between any of the chosen connections. Gender showed an effect on choice ($\chi^2(5) = 11.492, p < .05$). Men chose the hyperedge more frequently than women (44% and 20% respectively).

The relative high ratings of GTK and HGTK for the "other" option might be caused by non-serious "click-through" users. We tried removing nonsensical data (e.g. response times too short), but not all could be removed.

In order to ensure that the actual visual representation in the main task did not influence the answer (e.g. shortest mouse-paths, etc.) we switched option A and E (and B,C respectively) for 50% of the participants. No significant differences (χ^2-Test) between answers in both groups were found ($p > .05$).
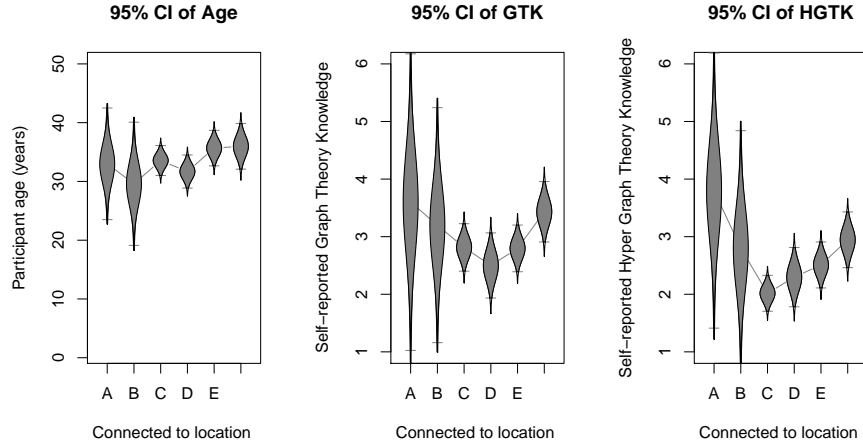
Fig. 4. Demographics for the different connections. 95% CI of means for age, graph theory knowledge and hypergraph theory knowledge. The sixth plot refers to the "other" location.

6.2   Lab Study

Looking at different answer types, we see basically six different representations. Most users attached the response only as a hyperedge (C, $n = 26$), as intended. Some included the idea (C & D, $n = 7$), some the question (C & E, $n = 5$), while two users connected all three positions (C, D, E). Then again, eight who only marked the idea (D) obviously did not use something similar to a hyperedge. Two users marked the wrong hyperedge (B, see also Fig. 3).

From this we can argue that two stances exist. Forty-one users correctly want to address the hyperedge, while eight want to address the node. When comparing user diversity of these two stances, we could not find differences for age ($CI[-8.1; 9.841]$), gender ($p = .181$), system usability ($CI[-8.1; 10.18]$) or graph-theory knowledge (edges $CI[-2.27; 0.48]$ or hyperedges $CI[-1.68; 0.38]$).

The usability of our prototype was rated as above average [21] ($SUS = 76$, $SD = 12$), indicating good usability. Gender was equally distributed in both studies and no gender effect on the SUS scale was found (unequal variances $F = 2.179$, $t(20.294) = .778$, $p = .446$, CI of differences $[-5.96; 13.1]^3$). Since we have no further data on gender and other variables, as well as the absence of this effect in the lab study, we assume the effect to be a methodological artifact, for which our data provides no satisfactory explanation. Further research is required.

---

[3] Confidence intervals of differences that contain 0, can be seen as non-significant differences, but provide more information then simply reporting significance testing results.

7   Discussion

Our results show that a large part of users are able to conceptualize and understand meta-communication modeled by hyperedges. Furthermore, when using an argument-mapping system the proportion of people intuitively using a hyperedge increases to 80%.

The main difference between Mturk and the lab study was the prior exposure to our software-prototype. Mturk participants should not use our system, to establish a large sample baseline. The lab study participants could use our system. The difference in percentages is interpreted as caused by the hypergraph-based interface of our system. The Mturk study merely serves as a baseline-measure for using hyperedges without the software-prototype context. No user-diversity factors influenced understanding or the usability evaluation significantly in the lab study.

We conclude from these findings, that using hyperedges in an argument mapping system may indeed be used, without confusing a majority of users.

7.1   Future Work

Large discussions often require a higher level of *abstraction* to express complex arguments besides using tags. This may happen, e.g. when a sub-discussion should be separate but contained in another post. Here, we propose using nested hypergraphs as a possible solution and want to investigate their comprehensibility. A concrete solution could be to merge the concepts of posts and tags to construct overlapping abstraction hierarchies.

As it is hard to investigate the effect of a graph-based argument mapping system on communication without conducting actual arguments, real world tests will need to be carried out next. We want to compare the effect of using our prototype in discussions in the e-learning system of seminars. Two similar seminars will use two different systems (graph-based argument mapping vs. regular message board) and report on usability and expressiveness in their evaluation. This allows to investigate differences between the discussions resulting from the two different protocols.

Before scalability can be evaluated within our approach, challenges are twofold: new methods for visualizing and navigating large graphs must be developed and large discussions must be investigated within our model.

8   Acknowledgments

References

1. Hilbert, M.: The maturing concept of e-democracy: From e-voting and online consultations to democratic value out of jumbled online chatter. Journal of Information Technology & Politics **6**(2) (2009) 87–110
2. Faridani, S., Bitton, E., Ryokai, K., Goldberg, K.: Opinion space: a scalable tool for browsing online comments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2010) 1175–1184
3. Kunsch, D.W., Schnarr, K., van Tyle, R.: The use of argument mapping to enhance critical thinking skills in business education. Journal of Education for Business **89**(8) (2014) 403–410
4. Dwyer, C.P., Hogan, M.J., Stewart, I.: The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. Thinking Skills and Creativity **5**(1) (2010) 16–22
5. Davies, M.: Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? Higher education **62**(3) (2011) 279–301
6. Shum, S.B., De Liddo, A., Klein, M.: Dcla meet cida: Collective intelligence deliberation analytics. In: 2nd International Workshop on Discourse-Centric Learning Analytics, LAK14: 4th International Conference on Learning Analytics & Knowledge. (2014)
7. Toulmin, S.E.: The Uses of Argument. Cambridge University Press (1958)
8. Kunz, W., Rittel, H.W.: Issues as Elements of Information Systems. Volume 131. Institute of Urban and Regional Development, University of California Berkeley, California (1970)
9. Lee, J.: Sibyl: a tool for managing group design rationale. In: Proceedings of the 1990 ACM conference on Computer-supported cooperative work, ACM (1990) 79–92
10. Cosley, D., Frankowski, D., Terveen, L., Riedl, J.: Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM (2006) 1037–1046
11. Van Gelder, T.: The rationale for rationale. Law, probability and risk **6**(1-4) (2007) 23–42
12. Fu, B., Noy, N.F., Storey, M.A.: Indented tree or graph? a usability study of ontology visualization techniques in the context of class mapping evaluation. In: International Semantic Web Conference, Springer (2013) 117–134
13. Fu, B., Noy, N.F., Storey, M.A.: Eye tracking the user experience–an evaluation of ontology visualization techniques. Semantic Web (Preprint) (2015) 1–19
14. Trapani, G., Pash, A.: The complete guide to Google Wave. 3ones Inc (2010)
15. Sumner, T., Shum, S.B.: From documents to discourse: Shifting conceptions of scholarly publishing. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co. (1998) 95–102
16. Garcia-Castro, A., Labarga, A., Garcia, L., Giraldo, O., Montana, C., Bateman, J.A.: Semantic web and social web heading towards living documents in the life sciences. Web Semantics: Science, Services and Agents on the World Wide Web **8**(2) (2010) 155–162
17. Whittaker, S., Terveen, L., Hill, W., Cherny, L.: The dynamics of mass interaction. In: From Usenet to CoWebs. Springer London (2003) 79–91

18. Odersky, M., Altherr, P., Cremet, V., Emir, B., Maneth, S., Micheloud, S., Mihaylov, N., Schinz, M., Stenman, E., Zenger, M.: An overview of the scala programming language. Technical report (2004)
19. Dietze, F., Karoff, J., Calero Valdez, A., Ziefle, M., Greven, C., Schroeder, U. In: An Open-Source Object-Graph-Mapping Framework for Neo4j and Scala: Renesca. Springer International Publishing, Cham (2016) 204–218
20. Bostock, M., Ogievetsky, V., Heer, J.: D^3 data-driven documents. Visualization and Computer Graphics, IEEE Transactions on **17**(12) (2011) 2301–2309
21. Sauro, J.: Sustisfied? little-known system usability scale facts. UX Magazine **10**(3) (2011) 2011–3