

Conversation flow in Wikipedia Discussion Pages

Zahra Arjmandi ^a, Tom Stafford ^b, Christine De Kock ^c, Andreas Vlachos ^c

^a Department of Cognitive Sciences, IPM, Tehran, Iran, ^b Department of Psychology, Sheffield University, ^c Department of Computer Science and Technology, Cambridge

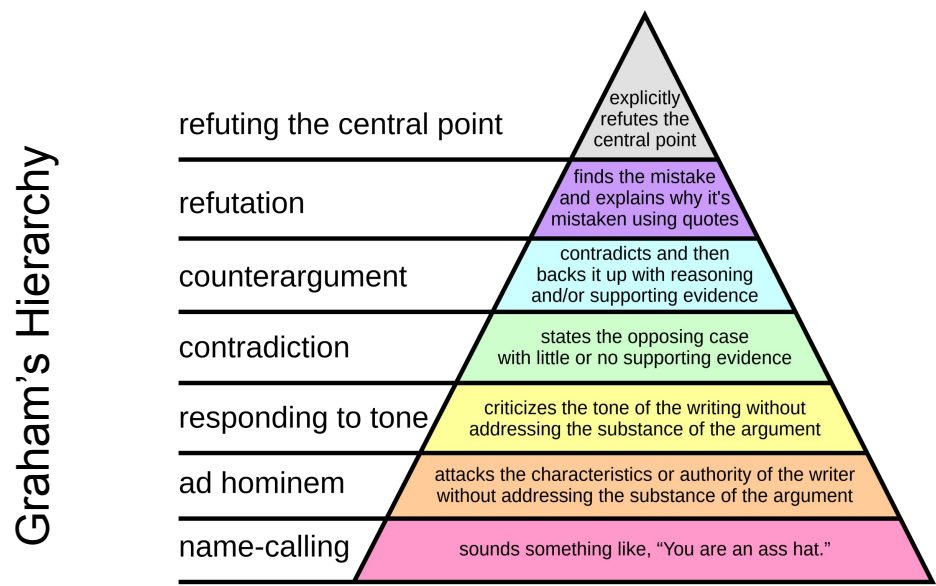
Wikipedia Discussion Pages

Wikipedia pages have a [talk/discussion pages](#) where editors discuss how to improve a page. If differences in opinions between editors cannot be resolved the talk page will be tagged as "dispute". If editors cannot resolve a dispute themselves these cases "escalated" to wikipedia mediators.

Graham Hierarchy of Disagreement

Wikipedia suggests users aspire to Paul Graham's Hierarchy of Disagreement ¹ to resolve disputes constructively (table).

Christine De Kock and Andreas Vlachos have annotated ~200 conversations (~4000 utterances) from wikipedia talk pages labelled "dispute" ². Labels reflect "rebuttal tactics", based on Graham's hierarchy, starting from name-calling (DH0) to refuting the central point (DH7). Also captured are "resolution tactics", attempts to promote understanding and consensus.



Labels used for annotations

Rebuttal	Resolution
DH7: Refuting the central point	Coordinating edits
DH6: Refutation	Contextualisation
DH5: Counterargument	Asking questions
DH4: Repeated argument/ Stating your stance	Providing clarification
DH3: Policing the discussion	Suggesting a compromise
DH2: Attempted derailings / off-topic comments	Conceding / recanting
DH1: Attacks to the credibility of the person or the argument	Bailing out
DH0: Name calling, insults and hostility	"I don't know."

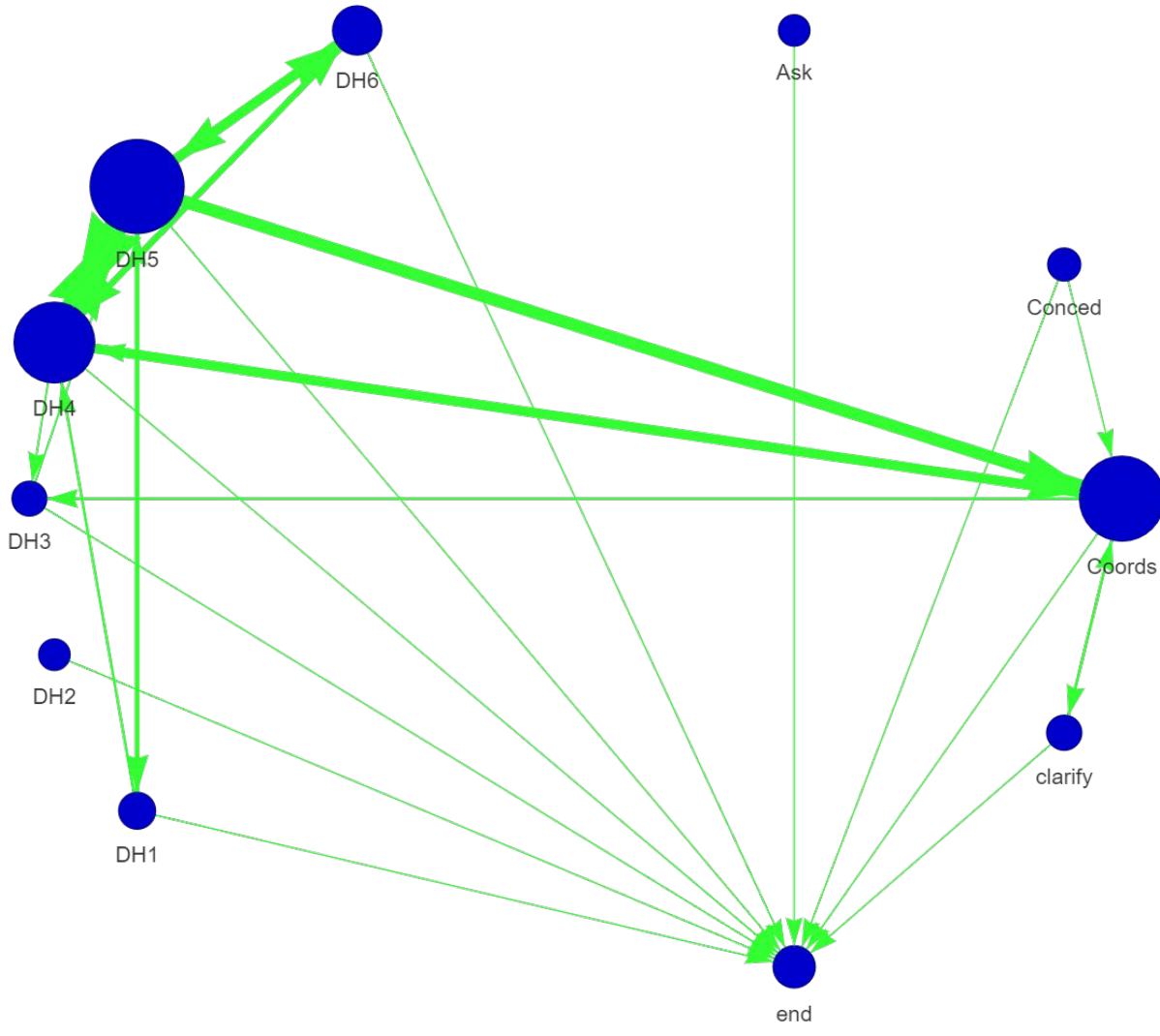
Research Questions

Can discussion flow in wikitalc disputes be effectively visualised as a network?
What differences are there between escalated and non escalated discussions?
- patterns in high vs low levels of disagreement?
- patterns in the balance of rebuttal vs resolution tactics

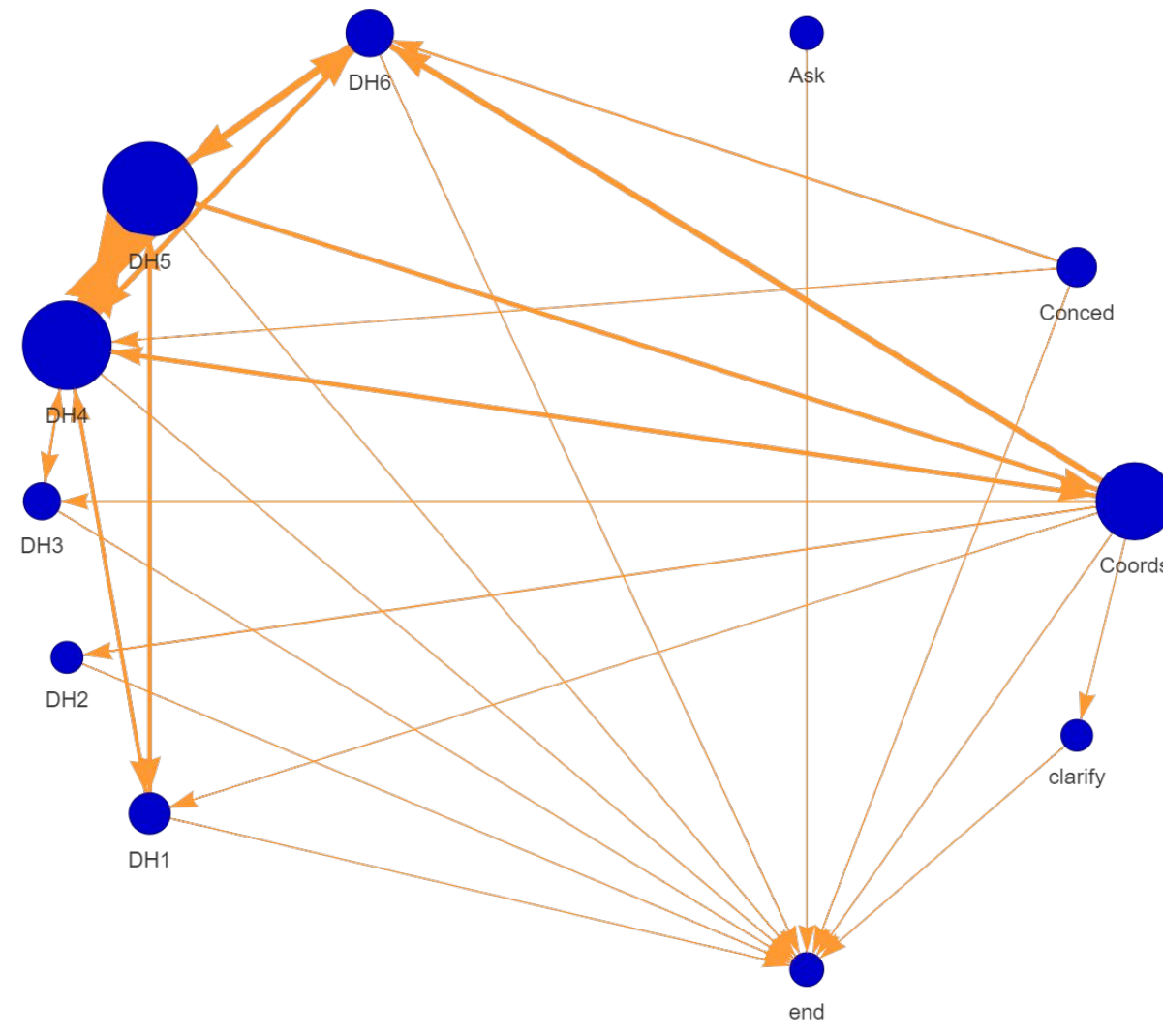
Method

- Divided the data into escalated vs. non-escalated conversations
- Each label is a node in the graph
 - size of node reflects frequency of the label
- Each transition is part of an edge
 - weight of arrow reflects number of transitions between two labels
- Used 3 methods to choose a label in case of a multi-label utterance
 - 'Max': choose the highest DH
 - 'Min': choose the lowest DH
 - 'All': include all the labels
- Used thresholds to remove low frequency nodes / edges

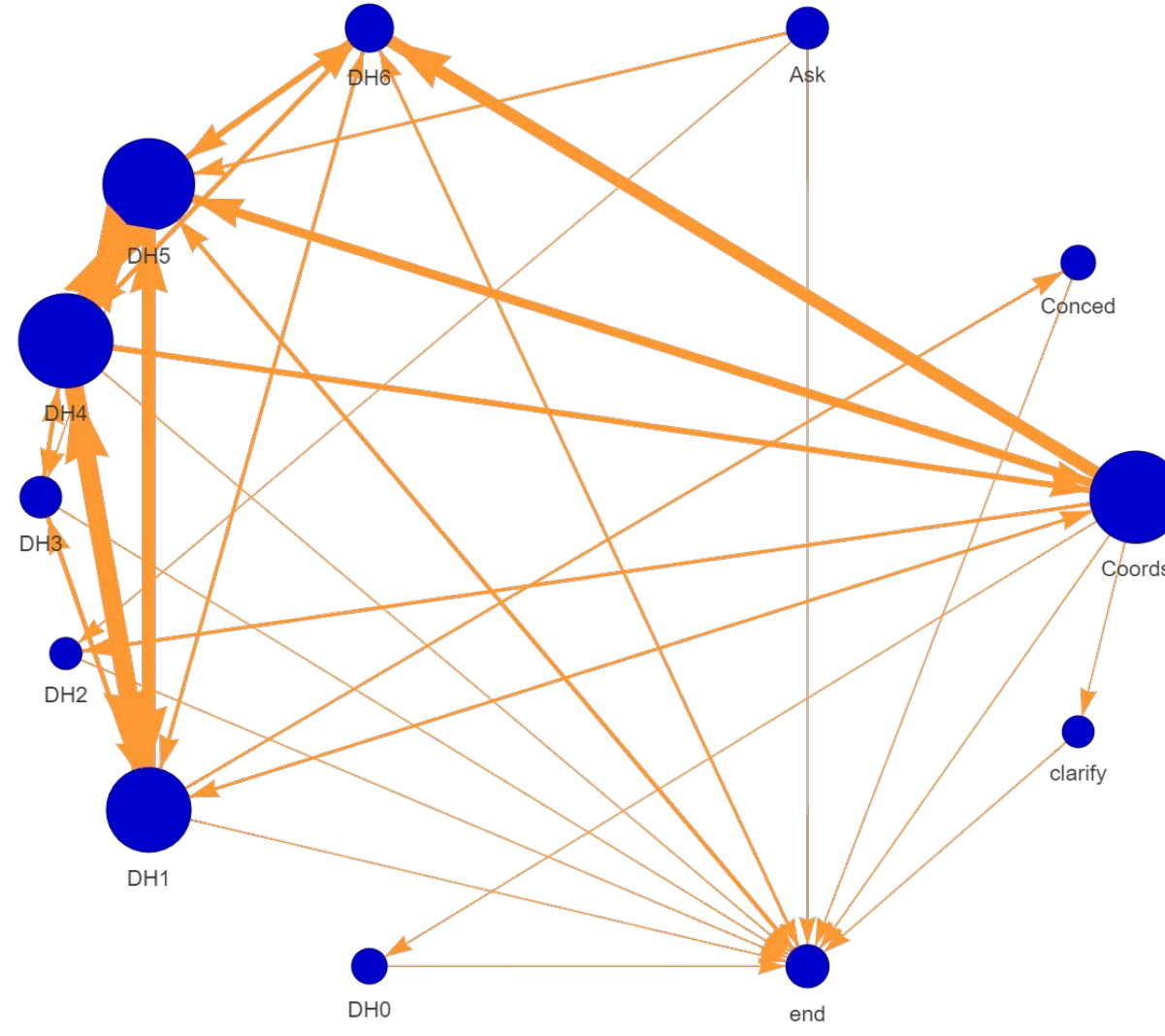
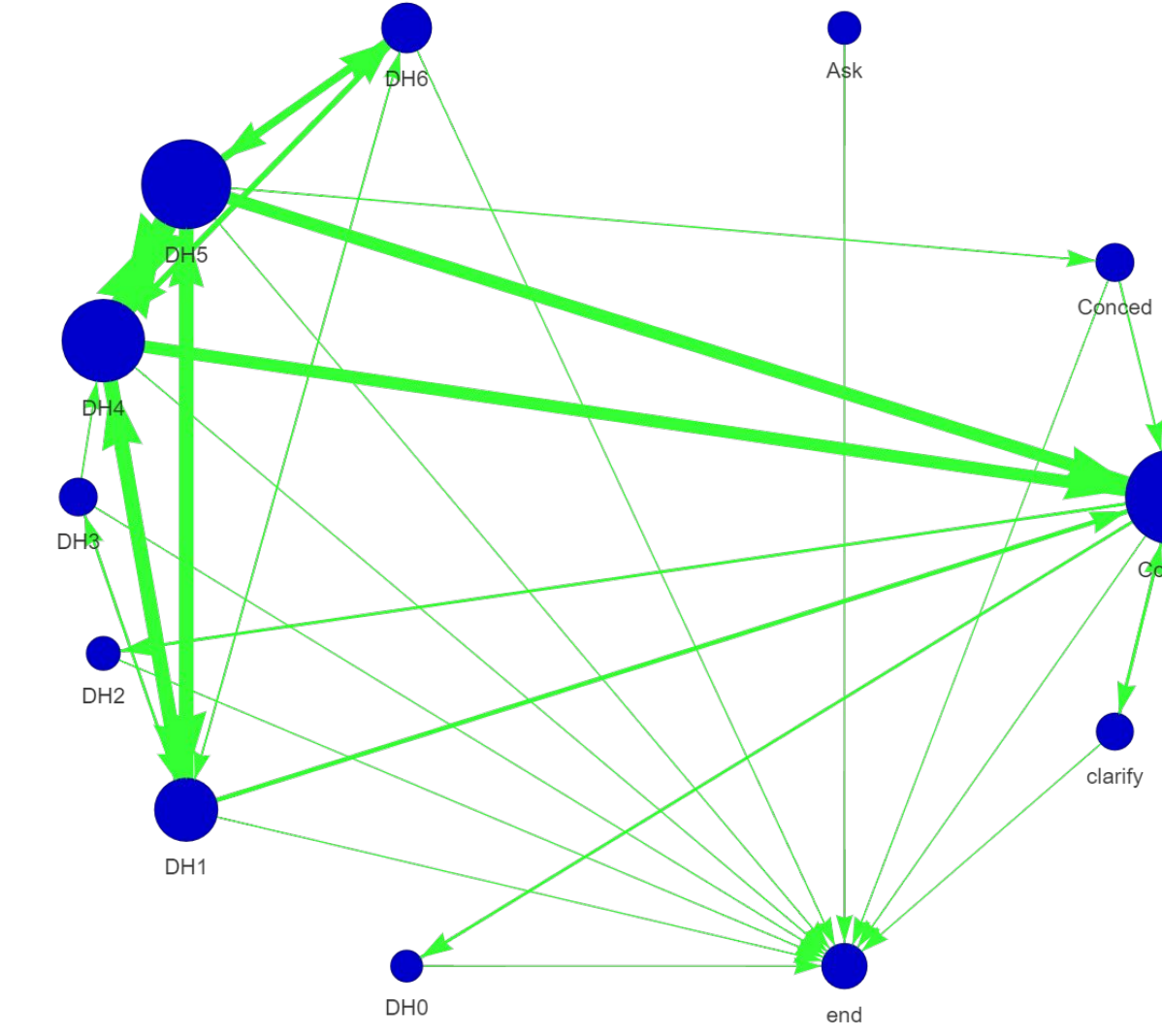
Non-escalated



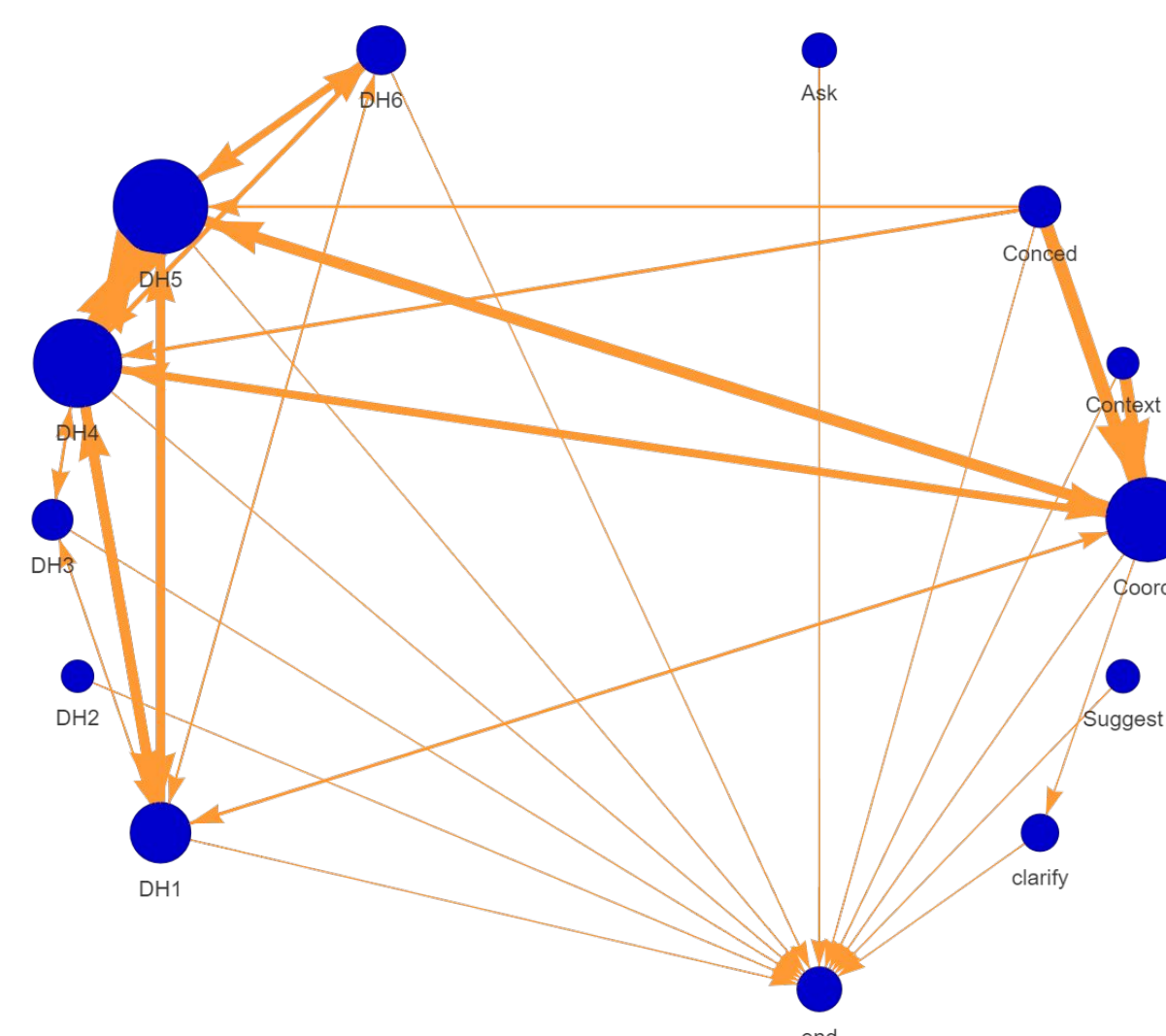
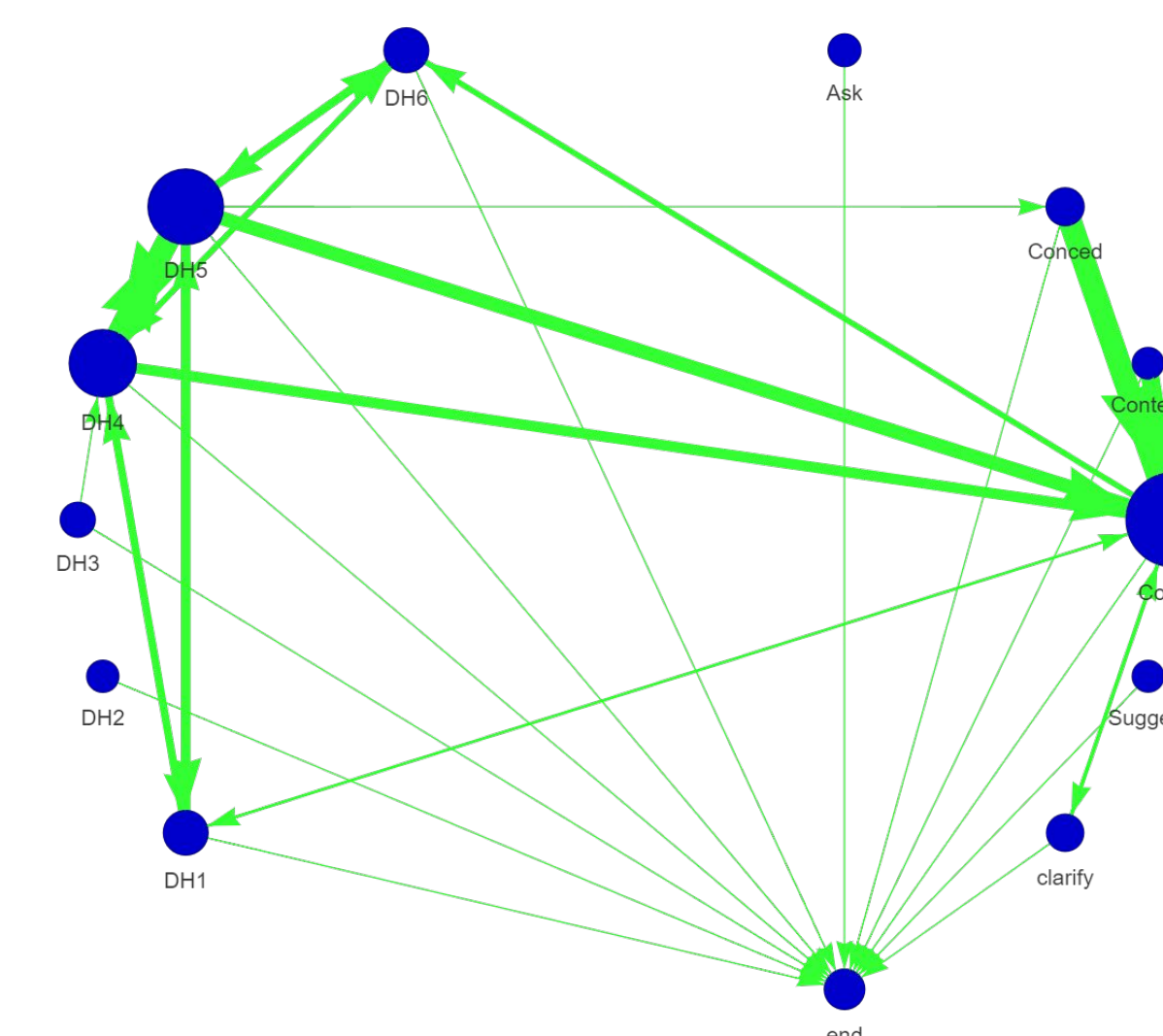
Escalated



"MAX" method: Both "escalated" vs "non-escalated" conversations use many high-order arguments. Non-escalated discussion associated with transitions between arguments and coordinating moves



"MIN" method: reveals more lower-order arguments in "escalated" conversations, suggesting conjoint use of higher and lower order arguments (especially DH1) could be a reason why less coordination was seen.



"ALL" method: fewer rebuttal-resolution (left-right) transitions in the "escalated" graph. Escalated conversations used the same level of argument, but failed in the coordination moves such "Asking questions", "providing a clarification", or "suggesting a compromise".

References

- Paul Graham. 2008. *How to disagree*. <http://www.paulgraham.com/disagree.html>
- Christine De Kock, Tom Stafford, Andreas Vlachos. 2022. *How to disagree well: Investigating the dispute tactics used on Wikipedia*. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.

Acknowledgement

This work has been supported by EPSRC grant EP/T024666/1: "Opening Up Minds" and the University of Sheffield and the University of Cambridge

Further Information

Check out the report in length here: https://github.com/zahra-arjm/wikipedia_discussions



z.arjmandi@ipm.ir
<https://github.com/zahra-arjm>