

# Polarization and potpourri: instructions and potential paper selections

## Schedule

Constraints I believe I satisfied:

Your name should appear (a) exactly once as a presenter and (b) exactly once as an annotator.

(c) Everybody should be available to present on their assigned day. (d) No one has to both annotate and present within a 5-day span.

	<b>Tu Apr 16</b> <i>By midnight (but actually OK if this gets pushed to midnight Wed Apr 17):</i> MM, HK post their selections to Ed	<b>Wed Apr 17.</b> <i>By noon:</i> FH annotates AB's paper. BW annotates TW's paper.	<b>Th Apr 18.</b> AB & TW upload slides and present. <i>By midnight although it would be terrific if it could be a day earlier (although it <u>must</u> be after MM and HK post):</i> DK, KL post their selections to Ed
<b>Mon Apr 22</b> <i>By noon:</i> AM annotates MM's paper; YW annotates HK's paper.	<b>Tu Apr 23</b> MM & HK upload slides and present. <i>By midnight, but days earlier much appreciated (as long as after DK and KL post):</i> EF, YW post their selections to Ed	<b>Wed Apr 24</b> <i>By noon:</i> PH annotates DK's paper TW annotates KL's paper	<b>Th Apr 25</b> DK & KL upload slides and present <i>By midnight, but days earlier much appreciated (as long as after EF and YW post):</i> PH, FH post their selections to Ed
<b>Mon Apr 29</b> <i>By noon:</i> MM annotates EF paper; HK annotates YW's paper.	<b>Tu Apr 30</b> EF YW <i>By midnight, but days earlier much appreciated (as long as after PH and FH post):</i> AM, BW post their selections to Ed	<b>Wed May 1</b> <i>By noon:</i> KL annotates PH's paper DK annotates FH's paper	<b>Th May 2</b> PH & FH upload slides and present

<b>Mon May 6</b> <i>By noon:</i> EF annotates AM's paper; AB annotates BW's paper.	<b>Tu May 7</b> AM & BW upload slides and present		
--	--	--	--

## Instructions for presenters

### 1. Paper selection Ed announcement.

My idea is that although people going earlier have less time to prepare, to compensate, they get a larger choice of papers.

So, watch Ed Discussions to see when *all* the people scheduled to present *before* you have posted. **Once they have done so, and also no later than midnight the week before you are to present, but as early as possible would be great so "downstream" people can get as much lead time as possible.** pick a paper from the papers remaining — see list below — and announce your choice on Ed Discussions.

It's nice if your Ed post copies the bibliographic info and the abstract info from what's in the paper list, and also (thanks to students for establishing the convention!) what date you'll be presenting.

Choose the "A6 - polarization/potpourri" as the Ed discussion category (and not "announcement")

**It would be a nice courtesy to begin your announcement with the name of your annotator, so they'll know.**

### 2. Conventions for the slides:

Please upload these to [this google drive folder](#) before class.

File naming convention: A6-<1st author last name and pub year>-<shorthand for title>-<your initials>

Use your initials in slides if you don't want your name public.

Please credit the sources of any images you include. (If all the screenshots are from the original paper, you can say this on the title page).

## Papers still up for grabs/presentation

Gu, Jia-Chen, Zhenhua Ling, Quan Liu, Cong Liu, and Guoping Hu. 2023. "GIFT: Graph-Induced Fine-Tuning for Multi-Party Conversation Understanding." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. <https://doi.org/10.18653/v1/2023.acl-long.651>.

*Addressing the issues of who saying what to whom in multi-party conversations (MPCs) has recently attracted a lot of research attention. However, existing methods on MPC understanding typically embed interlocutors and utterances into sequential information flows, or utilize only the superficial of inherent graph structures in MPCs. To this end, we present a plug-and-play and lightweight method named graph-induced fine-tuning (GIFT) which can adapt various Transformer-based pre-trained language models (PLMs) for universal MPC understanding. In detail, the full and equivalent connections among utterances in regular Transformer ignore the sparse but distinctive dependency of an utterance on another in MPCs. To distinguish different relationships between utterances, four types of edges are designed to integrate graph-induced signals into attention mechanisms to refine PLMs originally designed for processing sequential texts. We evaluate GIFT by implementing it into three PLMs, and test the performance on three downstream tasks including addressee recognition, speaker identification and response selection. Experimental results show that GIFT can significantly improve the performance of three PLMs on three downstream tasks and two benchmarks with only 4 additional parameters per encoding layer, achieving new state-of-the-art performance on MPC understanding.*

Jichuan Zeng, Jing Li, Yulan He, Cuiyun Gao, Michael Lyu, and Irwin King. 2020. [What Changed Your Mind: The Roles of Dynamic Topics and Discourse in Argumentation Process](https://doi.org/10.1145/3366423.3380223). In Proceedings of The Web Conference 2020 (WWW '20). Association for Computing Machinery, New York, NY, USA, 1502–1513. <https://doi.org/10.1145/3366423.3380223>

*In our world with full of uncertainty, debates and argumentation contribute to the progress of science and society. Despite of the increasing attention to characterize human arguments, most progress made so far focus on the debate outcome, largely ignoring the dynamic patterns in argumentation processes. This paper presents a study that automatically analyzes the key factors in argument persuasiveness, beyond simply predicting who will persuade whom. Specifically, we propose a novel neural model that is able to dynamically track the changes of latent topics and discourse in argumentative conversations, allowing the investigation of their roles in influencing the outcomes of persuasion. Extensive experiments have been conducted on argumentative conversations on both social media and supreme court. The results show that our model outperforms state-of-the-art models in identifying persuasive arguments via explicitly exploring dynamic factors of topic and discourse. We further analyze the effects of topics and discourse on persuasiveness, and find that they are both useful — topics provide concrete evidence while superior discourse styles may bias participants, especially in social media arguments. In addition, we draw some findings from our empirical results, which will help people better engage in future persuasive conversations.*

## Already selected

Vibhor Agarwal, Sagar Prakash Joglekar, Anthony P. Young, and Nishanth R. Sastry. 2022. [GraphNLI: A Graph-based Natural Language Inference Model for Polarity Prediction in Online Debates](#). Proceedings of WWW, 2729–2737.

<https://doi.org/10.1145/3485447.3512144>

There is a follow-up journal version: Vibhor Agarwal, Anthony P. Young, Sagar Joglekar, Nishanth Sastry, [A Graph-Based Context-Aware Model to Understand Online Conversations](#), ACM Trans. Web 18, 1, Article 10 (February 2024), <https://doi.org/10.1145/3624579>

*Online forums that allow participatory engagement between users have been transformative for public discussion of important issues. However, debates on such forums can sometimes escalate into full-blown exchanges of hate or misinformation. An important tool in understanding and tackling such problems is to be able to infer the argumentative relation of whether a reply is supporting or attacking the post it is replying to. This so-called polarity prediction task is difficult because replies may be based on external context beyond a post and the reply whose polarity is being predicted. We propose GraphNLI, a novel graph-based deep learning architecture that uses graph walk techniques to capture the wider context of a discussion thread in a principled fashion. Specifically, we propose methods to perform root-seeking graph walks that start from a post and captures its surrounding context to generate additional embeddings for the post. We then use these embeddings to predict the polarity relation between a reply and the post it is replying to. We evaluate the performance of our models on a curated debate dataset from Kialo, an online debating platform. Our model outperforms relevant baselines, including S-BERT, with an overall accuracy of 83%.*

Bao, Jiajun, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. 2021. Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations. In Proceedings of the Web Conference 2021 (WWW '21). Association for Computing Machinery, New York, NY, USA, 1134–1145. <https://doi.org/10.1145/3442381.3450122>

*Online conversations can go in many directions: some turn out poorly due to antisocial behavior, while others turn out positively to the benefit of all. Research on improving online spaces has focused primarily on detecting and reducing antisocial behavior. Yet we know little about positive outcomes in online conversations and how to increase them—is a prosocial outcome simply the lack of antisocial behavior or something more? Here, we examine how conversational features lead to prosocial outcomes within online discussions. We introduce a series of new theory-inspired metrics to define prosocial outcomes such as mentoring and esteem enhancement. Using a corpus of 26M Reddit conversations, we show that these outcomes can be forecasted from the initial comment of an online conversation, with the best model providing a relative 24% improvement over human forecasting performance at ranking conversations for predicted outcome. Our results indicate that platforms can use these early cues in their algorithmic ranking of early conversations to prioritize better outcomes.*

Federico Bianchi, Marco Marelli, Paolo Nicoli, and Matteo Palmonari. 2021. [SWEAT: Scoring Polarization of Topics across Different Corpora](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10065–10072, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

*Understanding differences of viewpoints across corpora is a fundamental task for computational social sciences. In this paper, we propose the Sliced Word Embedding Association Test (SWEAT), a novel statistical measure to compute the relative polarization of a topical wordset across two distributional representations. To this end, SWEAT uses*

~~two additional wordsets, deemed to have opposite valence, to represent two different poles. We validate our approach and illustrate a case study to show the usefulness of the introduced measure.~~

~~Dorottya Demszky, Jing Liu, Zid Mancenido, Julie Cohen, Heather Hill, Dan Jurafsky, and Tatsunori Hashimoto. 2021. [Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions](#). In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1638–1653, Online. Association for Computational Linguistics.~~

~~In conversation, uptake happens when a speaker builds on the contribution of their interlocutor by, for example, acknowledging, repeating or reformulating what they have said. In education, teachers' uptake of student contributions has been linked to higher student achievement. Yet measuring and improving teachers' uptake at scale is challenging, as existing methods require expensive annotation by experts. We propose a framework for computationally measuring uptake, by (1) releasing a dataset of student teacher exchanges extracted from US math classroom transcripts annotated for uptake by experts; (2) formalizing uptake as pointwise Jensen-Shannon Divergence (pJSD), estimated via next utterance classification; (3) conducting a linguistically-motivated comparison of different unsupervised measures and (4) correlating these measures with educational outcomes. We find that although repetition captures a significant part of uptake, pJSD outperforms repetition-based baselines, as it is capable of identifying a wider range of uptake phenomena like question answering and reformulation. We apply our uptake measure to three different educational datasets with outcome indicators. Unlike baseline measures, pJSD correlates significantly with instruction quality in all three, providing evidence for its generalizability and for its potential to serve as an automated professional development tool for teachers.~~

Ding, Xiaohan, Horning, Michael, & Rho, Eugenia H. (2023). Same Words, Different Meanings: Semantic Polarization in Broadcast Media Language Forecasts Polarity in Online Public Discourse. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1), 161-172.  
<https://doi.org/10.1609/icwsm.v17i1.22135>

*With the growth of online news over the past decade, empirical studies on political discourse and news consumption have focused on the phenomenon of filter bubbles and echo chambers. Yet recently, scholars have revealed limited evidence around the impact of such phenomenon, leading some to argue that partisan segregation across news audiences can not be fully explained by online news consumption alone and that the role of traditional legacy media may be as salient in polarizing public discourse around current events. In this work, we expand the scope of analysis to include both online and more traditional media by investigating the relationship between broadcast news media language and social media discourse. By analyzing a decade's worth of closed captions (2.1 million speaker turns) from CNN and Fox News along with topically corresponding discourse from Twitter, we provide a novel framework for measuring semantic polarization between America's two major broadcast networks to demonstrate how semantic polarization between these outlets has evolved (Study 1), peaked (Study 2) and influenced partisan discussions on Twitter (Study 3) across the last decade. Our results demonstrate a sharp increase in polarization in how topically important keywords are discussed between the two channels, especially after 2016, with overall highest peaks occurring in 2020. The two stations discuss identical topics in drastically distinct contexts in 2020, to the extent that there is barely any linguistic overlap in how identical keywords are contextually discussed. Further, we demonstrate at scale, how such partisan division in broadcast media language significantly shapes semantic polarity trends on Twitter (and vice versa), empirically linking for the first time, how online discussions are influenced by televised media. We show how the language characterizing opposing media narratives about similar news events on TV can increase levels of partisan discourse online. To this end, our work has implications for how media polarization on TV plays a significant role in impeding rather than supporting online democratic discourse.*



Efstratiou, Alexandros. 2024. [Deliberate Exposure to Opposing Views and its Association with Behavior and Rewards on Political Communities](#). To appear at the ACM Web Conference.

*Engaging with diverse political views is important for reaching better collective decisions, however, users online tend to remain confined within ideologically homogeneous spaces. In this work, we study users who are members of these spaces but who also show a willingness to engage with diverse views, as they have the potential to introduce more informational diversity into their communities. Across four Reddit communities (r/Conservative, r/The\_Donald, r/ChapoTrapHouse, r/SandersForPresident), we find that these users tend to use less hostile and more advanced and personable language, but receive fewer social rewards from their peers compared to others. We also find that social sanctions on the discussion community r/changemyview are insufficient to drive them out in the short term, though they may play a role over the longer term.*

Mia Mohammad Imran, Preetha Chatterjee, and Kostadin Damevski. 2024. [Uncovering the Causes of Emotions in Software Developer Communication Using Zero-shot LLMs](#). In Proceedings of the IEEE/ACM 46th International Conference on Software Engineering (ICSE '24). Association for Computing Machinery, New York, NY, USA, Article 182, 1–13.  
<https://doi.org/10.1145/3597503.3639223>

*Understanding and identifying the causes behind developers' emotions (e.g., Frustration caused by 'delays in merging pull requests') can be crucial towards finding solutions to problems and fostering collaboration in open-source communities. Effectively identifying such information in the high volume of communications across the different project channels, such as chats, emails, and issue comments, requires automated recognition of emotions and their causes. To enable this automation, large-scale software engineering-specific datasets that can be used to train accurate machine learning models are required. However, such datasets are expensive to create with the variety and informal nature of software projects' communication channels.*

~~In this paper, we explore zero-shot LLMs that are pre-trained on massive datasets but without being fine-tuned specifically for the task of detecting emotion causes in software engineering: ChatGPT, GPT-4, and flan-alpaca. Our evaluation indicates that these recently available models can identify emotion categories when given detailed emotions, although they perform worse than the top-rated models. For emotion cause identification, our results indicate that zero-shot LLMs are effective at recognizing the correct emotion cause with a BLEU-2 score of 0.598. To highlight the potential use of these techniques, we conduct a case study of the causes of Frustration in the last year of development of a popular open-source project, revealing several interesting insights.~~

~~Jo, Yohan, Shivani Poddar, Byungsoo Jeon, Qinlan Shen, Carolyn Rose, and Graham Neubig. 2018. "Attentive Interaction Model: Modeling Changes in View in Argumentation." In *Proceedings of NAACL*, 103–16. New Orleans, Louisiana: Association for Computational Linguistics.~~  
~~<http://www.aclweb.org/anthology/N18-1010>.~~

~~We present a neural architecture for modeling argumentative dialogue that explicitly models the interplay between an Opinion Holder's (OH's) reasoning and a challenger's argument, with the goal of predicting if the argument successfully changes the OH's view. The model has two components: (1) vulnerable region detection, an attention model that identifies parts of the OH's reasoning that are amenable to change, and (2) interaction encoding, which identifies the relationship between the content of the OH's reasoning and that of the challenger's argument. Based on evaluation on discussions from the Change My View forum on Reddit, the two components work together to predict an OH's change in view, outperforming several baselines. A posthoc analysis suggests that sentences picked out by the attention model are addressed more frequently by successful arguments than by unsuccessful ones.~~

Khare, Prashant, Ravi Shekhar, Mladen Karan, Stephen McQuistin, Colin Perkins, Ignacio Castro, Gareth Tyson, Patrick Healey, and Matthew Purver. 2023. [Tracing Linguistic Markers of Influence in a Large Online Organisation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 82–90, Toronto, Canada. Association for Computational Linguistics.

*Social science and psycholinguistic research have shown that power and status affect how people use language in a range of domains. Here, we investigate a similar question in a large, distributed, consensus-driven community with little traditional power hierarchy—the Internet Engineering Task Force (IETF), a collaborative organisation that designs internet standards. Our analysis based on lexical categories (LIWC) and BERT, shows that participants’ levels of influence can be predicted from their email text, and identify key linguistic differences (e.g., certain LIWC categories, such as “WE” are positively correlated with high influence). We also identify the differences in language use for the same person before and after becoming influential.*

Christine De Kock, Tom Stafford, and Andreas Vlachos. 2022. [How to disagree well: Investigating the dispute tactics used on Wikipedia](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3824–3837, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

*Disagreements are frequently studied from the perspective of either detecting toxicity or analysing argument structure. We propose a framework of dispute tactics which unifies these two perspectives, as well as other dialogue acts which play a role in resolving disputes, such as asking questions and providing clarification. This framework includes a preferential ordering among rebuttal-type tactics, ranging from ad hominem attacks to refuting the central argument. Using this framework, we annotate 213 disagreements*

~~(3,865 utterances) from Wikipedia Talk pages. This allows us to investigate research questions around the tactics used in disagreements; for instance, we provide empirical validation of the approach to disagreement recommended by Wikipedia. We develop models for multilabel prediction of dispute tactics in an utterance, achieving the best performance with a transformer-based label powerset model. Adding an auxiliary task to incorporate the ordering of rebuttal tactics further yields a statistically significant increase. Finally, we show that these annotations can be used to provide useful additional signals to improve performance on the task of predicting escalation.~~

~~Ian Stewart and Rada Mihalcea. 2022. [How Well Do You Know Your Audience? Toward Socially-aware Question Generation](#). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 255–269, Edinburgh, UK. Association for Computational Linguistics.~~

~~When writing, a person may need to anticipate questions from their audience, but different social groups may ask very different types of questions. If someone is writing about a problem they want to resolve, what kind of follow up question will a domain expert ask, and could the writer better address the expert's information needs by rewriting their original post? In this paper, we explore the task of socially-aware question generation. We collect a data set of questions and posts from social media, including background information about the question askers' social groups. We find that different social groups, such as experts and novices, consistently ask different types of questions. We train several text generation models that incorporate social information, and we find that a discrete social representation model outperforms the text-only model when different social groups ask highly different questions from one another. Our work provides a framework for developing text generation models that can help writers anticipate the information expectations of highly different social groups.~~

Tierney, Graham and Alexander Volfovsky. 2021. [Sensitivity Analysis for Causal Mediation through Text: an Application to Political Polarization](#). In *Proceedings of the First Workshop on Causal Inference and NLP*, pages 61–73, Punta Cana, Dominican Republic. Association for Computational Linguistics.

*We introduce a procedure to examine a text as mediator problem from a novel randomized experiment that studied the effect of conversations on political polarization. In this randomized experiment, Americans from the Democratic and Republican parties were either randomly paired with one another to have an anonymous conversation about politics or alternatively not assigned to a conversation — change in political polarization over time was measured for all participants. This paper analyzes the text of the conversations to identify potential mediators of depolarization and is faced with a unique challenge, necessitated by the primary research hypothesis, that individuals in the control condition do not have conversations and so lack observed text data. We highlight the importance of using domain knowledge to perform dimension reduction on the text data, and describe a procedure to characterize indirect effects via text when the text is only observed in one arm of the experiment.*