Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Emergent Influence Networks in Good-Faith Online Discussions

Henry Kudzanai Dambanemuya[1], **Daniel Romero**[2], **Emőke-Ágnes Horvát**[1]

[1] Northwestern University
[2] University of Michigan
hdambane@u.northwestern.edu, drom@umich.edu, a-horvat@northwestern.edu

Abstract

Town hall-type debates are increasingly moving online, irrevocably transforming public discourse. Yet, we know relatively little about crucial social dynamics that determine which arguments are more likely to be successful. This study investigates the impact of one's position in the discussion network created via responses to others' arguments on one's persuasiveness in unfacilitated online debates. We propose a novel framework for measuring the relationship between network position and persuasiveness using a combination of social network analysis and machine learning. Complementing existing studies that investigate the effect of linguistic aspects on persuasiveness, we show that the user's position in a discussion network is associated with their persuasiveness online. Moreover, other's recognition of one's successful persuasion is linked to one's increased dominant network position. Our findings offer important insights into the complex social dynamics of online discourse and provide practical knowledge for organizations and individuals seeking to understand the interplay between influential positions in a discussion network and persuasive strategies in digital spaces.

Introduction

Many public town hall-type debates have moved to digital spaces in recent decades. This shift has fundamentally transformed the nature of public discourse. For example, online social networks have made it easier for people to discuss with others, regardless of their physical location. Additionally, online spaces promise fewer gatekeepers and less salient signals of power and stature than in-person meetings. Thus, while the transition to online spaces could potentially lead to a diversification of voices and perspectives in public conversations, it raises questions about the role of the quality of discourse and the impact of network dynamics on persuasion (Stromer-Galley 2003). In other words, do substantive arguments or the user's network position determine which arguments will be more persuasive?

Persuasion is the process of influencing or changing one's opinions, beliefs, or behavior through argumentation (Gass 2010; Cialdini 1993). It is critical in advertising, marketing, political campaigns, and interpersonal communication (Shrum et al. 2012; Fogg 2002). Theories of persuasion, e.g., the Elaboration Likelihood Model (ELM), suggest two different paths for persuasion (Petty and Cacioppo 1986). Individuals with high levels of elaboration are more likely to scrutinize the merits or central factors of the arguments presented to them. However, people with low levels of elaboration are more likely to consider peripheral factors, such as specific argument-independent characteristics of the speaker.

Decades of communication research have focused on linguistic features of persuasion success (O'Keefe 2018, 2016, 2009, 1999). However, extensive literature in social psychology and sociology suggests that a speaker's position in a communication network is linked to how influential they are in changing others' opinions in political or health-related settings (Centola 2018, 2013; Christakis and Fowler 2007, 2008). The complex communication, persuasion, and decision-making dynamics in contemporary society have long been popular research subjects (Christakis and Fowler 2009; Fogg 2002; Tan et al. 2016). In particular, the evolving connections among individuals (or organizations) that shape the flow of information, ideas, and resources are fundamental for the critical social processes that are at play during discussions (Borgatti and Halgin 2011). These networks frequently encode influence subtly. Since they are changing rapidly in reaction to quickly evolving contexts (Cross, Cross, and Parker 2004), they are typically difficult to map and study at scale. Thus, the extent to which network features influence an individual's persuasiveness remains unknown. To address this challenge, here *we empirically investigate the role of network position in persuasiveness online*.

The rise of social media platforms, such as Facebook, X, Instagram, and Reddit, has facilitated the development of influence networks while creating the digital footprints necessary for examining them systematically (Lutu 2019; Wu and Shen 2015). Using data on who responds to whom, how and with what effect, we create discussion networks. Based on these networks, our work can move beyond existing efforts to study prominent "influencers," i.e., the few individuals or organizations with the ability to shape the opinions, attitudes, and behaviors of their followers through their curated and highly visible activities (Panagopoulos, Malliaros, and Vazirgianis 2020). Our research scrutinizes thus the conditions under which argumentation can be effective without the "social clout" of high-profile influencers (Lutu 2019) and

can be generalized to most online platform users.

Studying persuasiveness on social media platforms and fora requires a complex systems view, as these networks represent emergent systems that cannot be understood by only analyzing their components (Conte et al. 2012). Persuasion is often a collective rather than an individual effort, even in unfacilitated discussions without moderation, enforcing community governing principles. Instead, many individuals' contributions lead to the argumentation's success or failure in a self-organizing process. For this reason, we consider social interactions in online discourses that do not feature outside interventions but are still subject to the emergence of differences in how influential individual participants are (Zeng et al. 2020; Tan et al. 2016).

Within this approach, our study investigates whether, in addition to the linguistic quality of their argument, one's position in the discussion network (i.e., being influential or not) plays a role in their persuasiveness. To better understand the link between emergent influence and persuasion power, we examine successful arguments' effect on one's network position. Does successful persuasion have a positive reinforcing impact on one's influence? Or is influence a fleeting asset in these otherwise unstructured discussions? This inquiry is critical to uncover the potential endurance and implications of emergent influence networks in online discussions.

Our research focuses on a platform that hosts good-faith discussions on Reddit. This choice was motivated by extensive NLP-based work that uncovered the linguistic factors determining which arguments were persuasive in this context (Tan et al. 2016; Khazaei, Xiao, and Mercer 2017). We use this prior research as a baseline to understand the role of networks above and beyond crucial language markers. Finally, we discuss the implications of our findings in eliciting opinion change related to real-world problems, such as sustainability, migration, and global health.

Investigations into the role of one's network position in online discussions are important because they can shed light on how emergent influence networks determine the persuasiveness of information and ideas. This work can also help us understand how arguments are received and evaluated in online communities and how the structure and dynamics of discussion networks shape people's attitudes and beliefs. Furthermore, our research has implications for the design of online communication platforms and the development of strategies to enhance the quality of online discourse.

Related Work

Theories of Persuasion

To investigate the role of different factors in determining persuasiveness, we rely on the classical *Elaboration Likelihood Model (ELM)* as a theoretical base. The ELM provides a framework for understanding the basic processes underlying the effectiveness of persuasion and attitude change (Petty and Cacioppo 1986). The theory proposes two different processes by which people can be persuaded: a *central route* and a *peripheral route*. The central route focuses on a message's quality to influence opinions. For instance, the argument's quality is perceived via a central route by audiences who carefully weigh its merits (O'Keefe and Jackson 1995; Khazaei, Xiao, and Mercer 2017). The peripheral route occurs when a message receiver is unable or unwilling to decode the argument thoughtfully and instead relies on positive or negative cues associated with the message (e.g., the credibility or attractiveness of the sources of the message) (Han et al. 2018).

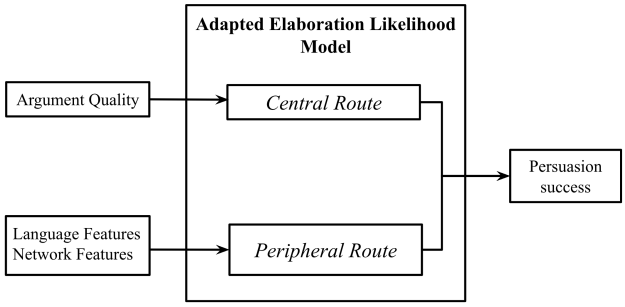

Figure 1: Proposed adaptation of the Elaboration Likelihood Model (ELM) with language and network features as part of the peripheral route.

Similarly to ELM's peripheral route, the *Pre-suasion Model* of persuasion also suggests a way to influence others by channeling attention to the persuader before exposure to the argument quality (Cialdini 2016). For example, individuals on online platforms may first seek to build trust, credibility, or popularity with their audience before attempting to persuade them. For example, high popularity can presuade audiences by shifting their attention to the popular individual's argument. Adapting these theories to the setting of online discussions, we identify from existing literature a set of factors related to the peripheral route of persuasion, as shown in Figure 1 and detailed next.

The Language of Persuasion

Extensive empirical research investigates linguistic factors associated with persuasive messages, including vocabulary and topic extraction (Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014; Tan et al. 2016), the use of semantic and syntactic rules (Hidey et al. 2017), and argument interactions and structure (Ji et al. 2018; Wei, Liu, and Li 2016; Tan et al. 2016). These language factors are essential in the peripheral route through which people process the persuasiveness of a message. The language of persuasion has been studied in text-based contexts online, including on crowdfunding platforms, advertising and marketing sites, product recommendations, and design practices that promote behavioral change (Mitra and Gilbert 2014; Shrum et al. 2012; Li and Zhan 2011; Fogg 2002; Horvát et al. 2018).

Crowdfunding studies investigate linguistic factors associated with successful fundraising for lending (Larrimore et al. 2011; Han et al. 2018), patronage (Mitra and Gilbert 2014), and charitable (Rhue and Robert 2018; Liang, Chen, and Lei 2016) campaigns. For example, Larrimore et al. (2011) rely on the Linguistic Inquiry and Word Count (LIWC) software to examine the relationship between language use and persuasiveness in peer-to-peer lending. They

find that using extended narratives, concrete descriptions and quantitative words that are likely related to one's financial situation is positively associated with receiving a loan.

Focusing on patronage sites such as KickStarter.com, Mitra and Gilbert (2014) also find persuasive language factors that predict funding success, e.g., phrases that signal lucrative offers or the attention a project has already received. In the lack of financial or material incentives, i.e., in prosocial fundraising, Rhue and Robert (2018) find that using emotions can effectively persuade others to contribute to one's campaigns. Other studies of altruistic requests in online communities also show that communicating one's needs clearly and including linguistic indications of gratitude and evidentiality are essential to convince donors to contribute to philanthropic causes (Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014).

Recently, the subreddit community `r/ChangeMyView` received significant research attention (Tan et al. 2016; Hidey et al. 2017; Wei, Liu, and Li 2016). For example, Tan et al. (2016) examined the similarity between the language of the opinion holder (the person requesting arguments that refute their opinion) and the counterarguments (the responses that challenge the initial opinion). Tan et al. (2016) find that a high linguistic similarity predicts persuasiveness. Hidey et al. (2017) annotate different types of semantic claims (e.g., interpretation, evaluation, agreement, and disagreement) to investigate whether certain claims are more likely to appear in persuasive than non-persuasive messages. The authors find that agreeing with what was previously said by others before expressing a diverging opinion, conceptual coherence (i.e., consecutive arguments of the same type), and claims backed by premises constitute persuasive rhetorical strategies.

Persuasion and Networks

As opposed to ample research on linguistic markers of persuasion, far less work focuses on other factors that may initially be processed through the peripheral routes of information processing but become critical for persuasive communication. Among these peripheral features, critical ones are related to people's influence in the discussion networks.

Social networks play a significant role in persuasion due to their influence on individuals' attitudes, beliefs, and behaviors (Christakis and Fowler 2009; Centola 2018). Extant studies suggest that social network structure determines the impact of social influence, especially when it comes to the large-scale adoption of health-related (Christakis and Fowler 2007, 2008; Centola 2013), community-building (Blair, Littman, and Paluck 2019), and sustainable behaviors (Constantino et al. 2022; White, Habib, and Hardisty 2019). The success of persuasion via social influence depends on the nature of information flows. More precisely, simple vs. complex contagion can predict the successful adoption of ideas in social networks (Centola 2018; Guilbeault, Becker, and Centola 2018). Whereas simple contagions (e.g., the spread of news much like the transmission of the flu) require a single contact activation, complex persuasive ideas and behavioral change require complex contagions, i.e., reinforcement of information that comes from multiple sources of interaction (Centola 2018; Guilbeault, Becker, and Centola 2018).

Thus, networks play a crucial role in understanding persuasiveness in online settings, as network structures that facilitate simple contagion can hinder complex contagion processes crucial for persuasion success. Additionally, studying the relationship between network position and persuasiveness can help us understand who holds more central or influential roles and how their participation in group settings shapes the conversation. Furthermore, recognizing central individuals can help target interventions or promote healthy discussions online. Therefore, studying the effect of network position in online discussions contributes to the development of social network theory and provides insights into human behavior in digital environments, advancing our understanding of communication dynamics online.

Data

We rely on data from the subreddit `r/ChangeMyView`, an active community on Reddit that started in January 2013 as a forum for debate and has grown to over 3 million users as of March 2023. `r/ChangeMyView` provides a platform for people to discuss various topics (e.g., abortion, gun control, vaccination, taxes, feminism, marriage, religious freedom, climate change, and society) and understand opposing viewpoints. On the platform, an original poster (OP) begins by posting an opinion or belief they hold to be accurate but accept that it may be flawed. They also share the reasoning behind their opinion in at least 500 characters. The OP is interested in understanding other perspectives on the issue, so challengers are invited to contest the OP's viewpoint. OPs explicitly recognize challengers' successful arguments by replying with the Δ character and explaining how and why their original viewpoint changed. The specific data sample we analyse here was released by Tan et al. (2016) and consisted of discussions from January 2013 to May 2015.

Using basic natural language processing (NLP), we follow each discussion's comment threads to identify persuasive challengers awarded Δs. Note that while multiple Δs can be awarded in the same discussion, we focus on the first Δ due to its critical role in changing the opinion of the original poster. The community also has strict rules to facilitate good-faith discussions. For example, the OP must personally hold the view they are offering for discussion, demonstrate that they are open to it changing, and be willing to have a conversation within 3 hours after posting. Challengers' direct responses to an OP must question at least one aspect of the OP's viewpoint, contribute meaningfully to the discussion and not be rude or hostile to other users.

Our data include the full text of arguments and the structure of responses (i.e., "in-reply-to" relationships between arguments). These two types of information enable us to deduce various language and network features, which could be correlated with persuasiveness. In what follows, we describe both language and network features.

Measures

Language Features

We adopt a set of measures that have been shown to be associated with persuasiveness in the r/ChangeMyView community (Tan et al. 2016; Khazaei, Xiao, and Mercer 2017). The used language features encompass the number of (in)definite articles, which are associated with an argument's specificity (Danescu-Niculescu-Mizil et al. 2012); positive and negative words suggestive of patterns of emotion (Hullett 2005; Wegener and Petty 1996); question and quotation marks prompting clarification and indicating attention to others' words (Khazaei, Xiao, and Mercer 2017); personal pronouns with self-affirmation and examples from personal experiences in argumentation (Correll, Spencer, and Zanna 2004; Cohen, Aronson, and Steele 2000); and the number of URL links citing external evidence to support an argument (Tan et al. 2016). We consider the number of words, sentences, and the readability of the arguments quantified by the Flesch-Kincaid grade level and readability score (Flesch 2007). We include lexical diversity via a word entropy measure, which quantifies the Shannon entropy of the set of words in the argument, and a token type entropy, meaning the Shannon entropy of the set of different parts of speech used in the argument. Finally, we rely on a set of features that describe an argument's (A) vocabulary overlap with the original post (O) using the following measures (Tan et al. 2016):

- *Number of common words* between an argument and original post: $|A \bigcap O|$
- *Reply fraction* of common words in the argument: $\frac{|A \bigcap O|}{|A|}$
- *OP fraction* of common words in the original post: $\frac{|A \bigcap O|}{|O|}$
- *Jaccard similarity* between the original post and the argument: $\frac{|A \bigcap O|}{|A \bigcup O|}$

Matched Sample. Following Tan et al. (2016), we compute the language features above for each persuasive challenger who received a Δ and a matched challenger who was not awarded a Δ but had the highest overlap in their arguments' vocabulary with the successful challenger. This matched sample allows us to compare properly challengers with lexically similar arguments. Hence, throughout the study, we conduct our analysis on the matched sample with an equal number of observations for arguments that won a Δ and matching arguments that were not awarded Δs.

Network Features

To explore the structure of discussions, we create directed discussion networks whose nodes are r/ChangeMyView users (i.e., the OP and their challengers) and whose edges denote "in-reply-to" relations arising from a challenger replying to another challenger or the OP. Notice that this network is weighted because, for instance, the same challenger can provide arguments to the OP multiple times during the discussion.

We calculate each node's centrality in these discussion networks to quantify the user's influence based on their network position. We choose five types of centralities, which are based purely on structural information as follows:

- **In-degree**. This centrality is defined as the sum of all incoming edge weights such that the measure quantifies the number of replies the challenger received from other challengers or the OP. In-degree has been used to quantify the popularity of social network users on platforms such as Twitter, Instagram, and Facebook (Lutu 2019). Typically, individuals with high in-degree centrality occupy an advantageous social position because they have more direct sources of information (Wasserman, Faust et al. 1994; Johnson, Everett, and Borgatti 2018).
- **Out-degree**. We sum the weights of all outgoing edges from each node to obtain a node's out-degree. This centrality quantifies how often the challenger replies to others. Proactively connecting arguments or mediating in tense discussions might indicate a critical network position.
- **Degree ratio**. As a measure of balance between incoming and outgoing edge weights, we compute the degree ratio as the ratio between out-degree and in-degree (i.e., out-degree/in-degree).
- **Authority**. This centrality is based on the Hyperlink-Induced Topic Search (HITS) algorithm developed by Kleinberg et al. (1998). It quantifies the number and quality of links pointing *to* the challenger from other high-authority users in the discussion network. It has been shown to be associated with the number of positive evaluations in crowd innovation contests that collect novel ideas from consumers (Özaygen and Balagué 2018).
- **Hubbiness**. This centrality is also based on the HITS algorithm (Kleinberg et al. 1998) but quantifies the number and quality of links pointing *from* the challenger to other high-authority nodes in the network. It has been found to be correlated with enhanced learning from various network communities (Taylor et al. 2022).
- **Betweenness**. This centrality quantifies the probability that a node lies on the shortest (directed) path between two randomly chosen nodes. It has been used to identify influential nodes or opinion leaders in online social networks (Opuszko, Gehrke, and Niemz 2019; Johnson, Everett, and Borgatti 2018).

These centralities inherently change throughout the evolution of the discussion. We compute them right before a Δ is awarded to reflect the influence accumulated by the successful challenger at that moment. For proper comparison, we compute the centralities for a matching challenger with similar arguments at the same time in the discussion. See Figure 2 for an example network illustration before and after a Δ is awarded by an OP.

Methods

Identifying Features Associated with Persuasiveness. We train and evaluate the performance of supervised classifi-

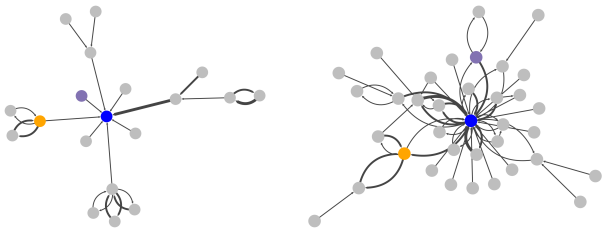

Figure 2: Example illustration of a discussion network immediately before an OP awards a Δ (left) and when the conversation concludes at a later time (right). The nodes represent the OP (blue), successful challenger (orange), matched unsuccessful challenger (purple), and other unsuccessful challengers (grey). The directed edges represent who replied to whom. The edge weights denote the number of replies and are signaled by line thickness.

cation models to predict which challengers will successfully persuade an OP to change their view. Thus, the outcome of a classification is whether or not a challenger's arguments are persuasive. If the challenger is awarded a Δ, the outcome is 1; otherwise, it is 0. We classify based on the challenger's language and network features computed from observations immediately before the OP awards a Δ. To ensure generalizability, we use various models, including Decision Trees, Random Forests (Breiman 2001), Adaptive Boosting (Freund and Schapire 1997), Logistic Regression, and Gaussian Naive Bayes. We report the area under the receiver operating characteristic curve (AUC) scores computed on hold-out test sets during 5-fold cross-validation to evaluate classification model performance. We rely on Random Forest permutation importance scores to evaluate the relative importance of network features and select the network feature with the most explained variance in predicting a challenger's persuasiveness.

Difference in Difference (DID) Estimation in the Change of Network Centrality over Time. To estimate the effect of receiving a Δ on users' network position, we rely on an econometric approach for inferring the causal effects of treatment. In other words, we compare longitudinal data between a treatment group (users awarded Δs) and a control group of matching users not awarded Δs. Specifically, we compare the magnitude of the gap between the treatment and control groups *before* an OP awards a Δ to a challenger and *after* an OP awards a Δ to a challenger, i.e., when the discussion concludes. We estimate the DID effect using the following specification: $y = \beta_0 + \beta_1 T + \beta_2 G + \beta_3 (T \cdot G) + \epsilon$, where y is a type of network centrality, B_0 is the regression intercept, G is a dummy variable for group membership (0 = control, 1 = treated), T is a dummy variable for the period (0 = before the Δ, 1 = after the Δ), $T \cdot G$ is the interaction term between time and treatment group, and ϵ is the error term. The coefficient β_3 of the interaction of T and G is the DID effect.

Results

We analyzed 283,751 comments made by 34,907 challengers in 3,051 posts. On average, each post contains 93 comments and 38 unique challengers. On average, posts receive 64.08 (standard deviation = 102.89) replies before a challenger receives a Δ and 28.84 (standard deviation = 72.84) replies after a Δ is awarded. Thus, conversations in the `r/ChangeMyView` persist despite OPs acknowledging when they change their original viewpoint.

Language Features. Consistent with prior findings, we observe that successful arguments are wordier, use more URL links as supporting evidence, utilize more examples (i.e., phrases such as "for example", "for instance", or "e.g.") and punctuation, have more lexical diversity (e.g., as measured by different words or different parts of speech), and have more words in common with the original post compared to unsuccessful arguments, c.f., (Tan et al. 2016; Khazaei, Xiao, and Mercer 2017). We also find that successful arguments contain more positive and negative words, supporting two hypotheses in the existing literature. On the one hand, existing literature suggests that using positive words may lead to persuasion success by energizing and directing one's behavior to react positively to an argument or request (Liang, Chen, and Lei 2016). On the other hand, the "empathy helping hypothesis" suggests that using negative words can lead to persuasion success by making people more empathetic towards one's plight (Fisher, Vandenbosch, and Antia 2008). Determining which types of arguments that either use positive or negative words will result in success in persuasion is beyond the scope of our work. Table 1 compares persuasive and non-persuasive arguments along various linguistic features.

Network Features. We observe that having higher initial out-degree, hubbinness, and betweeness centralities is positively associated with successful argumentation (Table 1). Accordingly, the more arguments one provides (hence higher out-degree), the more likely they will succeed. The higher the number of links pointing from the challenger to other high-authority challengers (hence greater hubbiness), the more likely a challenger will succeed. Additionally, the observation that higher betweenness centrality is associated with persuasion success implies that challengers who play a crucial role in transmitting information between different parts of the communication network have better chances of persuasion success.

However, having a higher initial in-degree and authority centrality is negatively associated with successful argumentation (Table 1). Thus, before getting recognition from an OP, we observe that successful challengers reach out to other challengers more than they receive replies. Hence, they also have a higher degree ratio than unsuccessful challengers. This finding suggests an "exploration $\rightarrow$ exploitation" argumentation strategy whereby successful challengers might first explore multiple viewpoints that enable them to combine different viewpoints and then provide substantive arguments (c.f., Figure 2).

We further observe that most network centralities are highly correlated. Using Random Forest permutation impor-

Features	Non-Persuasive Challengers	Persuasive Challengers
Language features		
# words	305.318 (297.102)	440.757 (432.011)
# sentences	17.052 (17.730)	23.476 (22.895)
# positive words	8.612 (9.254)	12.420 (13.793)
# negative words	6.859 (9.260)	9.876 (12.411)
# examples	1.295 (2.461)	1.892 (3.102)
# hedges	5.740 (6.498)	8.017 (8.898)
# definite articles	6.068 (7.211)	8.883 (9.928)
# indefinite articles	0.091 (0.327)	0.116 (0.350)
# 1st person pronouns	5.891 (9.301)	8.003 (11.227)
# 1st person plural pronouns	1.692 (3.417)	2.614 (5.119)
# question marks	1.860 (3.309)	2.304 (3.649)
# quotation marks	11.274 (13.611)	16.007 (19.065)
# url	0.076 (0.535)	0.123 (0.653)
word entropy	6.672 (0.792)	7.018 (0.700)
token type entropy	3.534 (0.175)	3.565 (0.099)
# common words	44.298 (28.431)	50.422 (30.736)
argument reply fraction	0.285 (0.110)	0.245 (0.103)
OP fraction	0.241 (0.125)	0.271 (0.128)
Jaccard similarity	0.132 (0.051)	0.129 (0.045)
Network features		
in-degree	1.515 (3.171)	1.374 (2.909)
out-degree	2.015 (2.771)	2.140 (2.524)
degree ratio	1.508 (0.576)	1.664 (0.548)
betweenness	20.971 (235.408)	27.275 (222.365)
authority score	0.039 (0.140)	0.036 (0.135)
hub score	0.492 (0.320)	0.523 (0.328)

Table 1: Mean (standard deviation) of language and network features for all matched user pairs computed immediately before an OP awards a Δ.

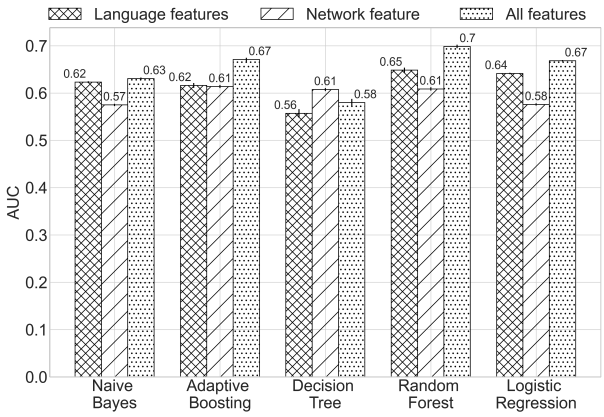

Figure 3: Area under the receiver operating characteristic curve (AUC) for language features, the network feature (degree ratio), and all features combined in each supervised learning model. The reported AUC scores are based on a balanced data set with matched samples.

tance score, however, we find that degree ratio has the highest importance, explaining 32.9% of the variance among network features. Given the importance of this feature in relation to other network features, we will only use degree ratio in our models of persuasion.

Network Features Enhance the Prediction of Persuasion

We train and evaluate the performance of different supervised learning models based on language and/or the most significant network feature (degree ratio) and report the AUC for each feature group in Figure 3. Accordingly, language features are core to predicting persuasiveness. Still, we observe significant boosts in AUC from adding our network feature to models based on language features. Across the different classifiers, Random Forest attains the highest performance improvement (7.95%) when adding network to language features. These findings suggest that network position plays a significant role in persuasion success. Across the different classification models, Random Forest performs best on all features (AUC=0.701). Figure 4 shows the corresponding ROC curves obtained with Random Forest.

Persuasive Arguments Re-enforce the Challengers' Influential Network Position

When comparing the network position of challengers who received a Δ to their matches who do not receive Δs, we observe that receiving a Δ has non-trivial effects on successful challengers' network position (Table 2).

In-degree. We find that receiving a Δ increases the number of comments that a successful challenger receives, thus shifting other challengers' attention to the successful challenger ($\beta_3 = 2.219, p < 0.001$).

Out-degree. Receiving a Δ also increased a successful challenger's outgoing replies to others ($\beta_3 = 0.889, p <$

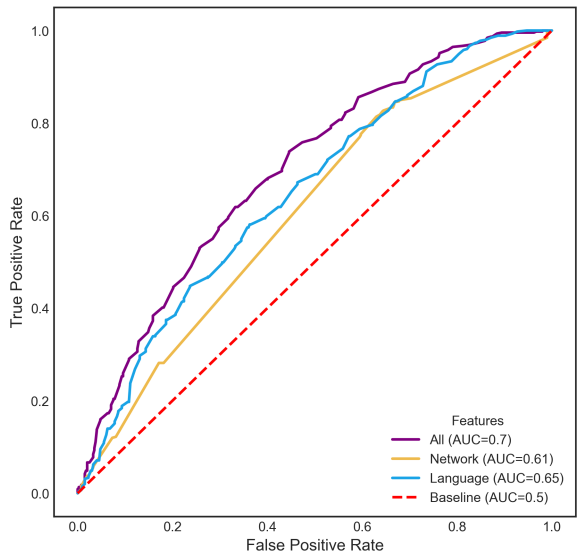

Figure 4: Receiver Operating Characteristic (ROC) curves of the best-performing model, Random Forest, for each feature group. The model evaluation is based on a balanced data set with matched samples.

Network Centrality	Main analysis	Robustness check
In-degree	2.219***	2.165***
	(0.203)	0.204)
Out-degree	0.889***	0.588***
	(0.157)	(0.156)
Authority	0.035***	0.040***
	(0.006)	(0.007)
Hub	0.065***	0.006
	(0.014)	(0.014)
Betweenness	112.061***	106.034***
	(27.118)	(26.890)
Observations	7940	7786

Table 2: Difference in Difference (DID) results on weighted communication networks. The main analysis includes replies to the winning argument and all subsequent arguments. The robustness check excludes replies to the winning argument.

Features	Accuracy	Precision	Recall	F1	AUC
Network	0.599	0.570	0.810	0.669	0.614
Language	0.600	0.601	0.594	0.597	0.648
All	0.641	0.640	0.644	0.642	0.701

Table 3: Random forest classification results on balanced data set of matched samples

0.001). However, it is worth noting that receiving a Δ has a much higher effect on successful challengers' incoming than outgoing replies and is thus negatively associated with the degree ratio, i.e., out-degree/in-degree ($\beta_3 = -0.618, p < 0.001$).

Authority. We further observe that receiving a Δ increases the number of replies to a successful challenger from other challengers with high authority ($\beta_3 = 0.035, p < 0.001$). Thus, challengers who receive $Deltas$ from OPs increasingly become important sources of information that are highly visible and influential to other challengers involved in the conversation.

Hubbiness. Complementary to a challenger's importance as a source of information, we observe that receiving a Δ increases a successful challenger's hubbiness, which represents their importance as gateways or connectors to other important challengers in the network, thereby helping other challengers to navigate different arguments and viewpoints in the conversations ($\beta_3 = 0.065, p < 0.001$).

Betweenness centrality. Additionally, we observe that receiving a Δ is associated with successful challengers occupying unique network positions and serving as bridges or intermediaries between different parts of the network ($\beta_3 = 112.061, p < 0.001$). Thus, successful challengers can better control or mediate the flow of ideas and viewpoints to other challengers during the discussions.

Robustness Checks

To ensure the robustness of our findings, we conducted several additional analyses. First, we varied the evaluation metrics for our language and network feature-based classification models. We obtained similar results across different measures, including accuracy, precision, recall and F1 score (Table 3). Then, we used different selections of network features (e.g., in/out-degree, betweenness, hub, and authority) to complement the language-based Random Forest classifier and still observed significant improvements in prediction performance in terms of the evaluation measures reported in Table 3. However, we observe the highest performance improvement from the degree ratio since it explains the highest variance. Additionally, we examined the impact of Δs on network position using both weighted and unweighted networks (i.e., considering the frequency as opposed to the mere presence of replies) and found consistent patterns of change in both cases. Furthermore, we conducted an analysis in which we ignored responses to the argument that won a Δ and observed similar changes in the network structure, indicating that successful challengers not only receive more interactions in response to their winning arguments but also attract greater attention from other participants in the conversation (Table 2). Future research could extend these analyses to consider additional robustness checks, such as variations in network sampling methods or alternative model specifications, to assess further the conditions under which winning arguments benefit from the challengers' network position and the network position is further elevated after the recognition of a successful argument.

Discussion

It is widely accepted that we live in a connected world where people are often influenced by the attitudes, beliefs, and behaviors of others around them (Easley and Kleinberg

2010; Christakis and Fowler 2009). Such social influences are prevalent in online social networks, where people come together to share ideas, exchange information, and sometimes engage in debates. Our work focuses on the latter, where individuals engage in discussions in an attempt to persuade someone to change their views on specific topics. Through our analysis, we uncover the dynamics of persuasion in this particular social context.

First, *we investigate what makes some individuals successful persuaders*. Consistent with prior findings, we find that language features play an essential role in persuasion (Tan et al. 2016; Khazaei, Xiao, and Mercer 2017). The language features include using external URLs to provide evidence in support of one's claims and the lexical diversity of one's arguments. These language features are important because they reflect the true merits of an argument. Our study emphasizes the crucial role of language in successful persuasion as a critical peripheral route in the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986).

Second, *we deduce emergent influence networks of who replies to whose arguments*. We compute relevant network centrality measures and quantify their role in persuasion. Adding network-based features to language-based supervised learning models significantly enhances the model's predictive performance in classifying (un)successful challengers. Although the emergent network position is not directly reflected in the arguments, it represents another important peripheral cue. Our finding suggests that people deduce heuristics about who to interact with based on their position in the influence network. Combined, our first two findings provide empirical support for the peripheral route of information processing in the ELM persuasion theory (Petty and Cacioppo 1986).

Third, *we investigate the effect of persuasiveness on network position* and find that successful persuasion leads to elevated centrality in the discussion network. Our difference-in-difference estimation shows that network centrality changes over time, such that persuasiveness consolidates influential nodes' positions in the network. For example, we observe that successful challengers benefit from more challengers interacting with them. Thus, compared to unsuccessful challengers, successful challengers incur indegree and authority score benefits, further enhancing their influence in the network. This result is essential in light of the pre-suasion model, which suggests that challengers can influence others by capturing and channeling attention to themselves. Persuasive challengers will be more likely to accrue higher recognition (Cialdini 2016).

Finally, *we find that successful arguments have spillover effects.* We observe that successful challengers interact more with others, even when we ignore replies to the message that received a Δ in both weighted and unweighted networks. This means that successful challengers do not simply have more incoming responses to their successful arguments. Instead, their other arguments garner more attention as well. This observation also supports the pre-suasion model of persuasion in the sense that having a winning argument directs attention towards other arguments that one may have made, but were not publicly acknowledged as successful. Additionally, we observe that the degree distributions are skewed, which could arise through a generative mechanism of preferential attachment (Barabási and Albert 1999). Combined, these findings demonstrate the reinforcing effect of successful arguments on influential positions in the network.

Practical Implications. Given its importance for democratic societies, it is crucial that we improve the quality of online deliberation. Our findings about the role of emergent influence networks in deliberation point to a couple of practical ideas that could improve outcomes in online deliberative spaces. First, by understanding different network positions' role in unmoderated conversations, platform maintainers can either highlight or, on the contrary, de-emphasize influential individuals based on whether they encourage constructive discussions, promote critical thinking, or, conversely, contribute to the spread of incivility and misinformation. Second, studying network positions can reveal how existing algorithms interact with network dynamics, potentially exacerbating the visibility of harmful comments by vocal influential users. Our study demonstrates that persuasion is a collective effort. Hence, instead of universally incentivizing user engagement, platforms could weigh engagement metrics (such as likes, shares, and comments) based on network position to also highlight helpful perspectives of non-central users. Such incentives to redistribute recognition, could contribute to a plurality of voices and retain users in online deliberation (Becker, Almaatouq, and Horvát 2020). Beyond the context of Reddit, similar efforts could also improve content ranking and recommender systems.

Broader Impact. We apply network science methods to provide a better understanding of persuasiveness in online discussions. We offer new insights on how to create effective communication strategies not simply based on the contents of the message, but also on the messenger's position in the influence network. Our findings have broad implications for understanding the effectiveness of different messages and people's network positions in eliciting opinion change to solve pressing problems around, e.g., sustainability, migration, and global health. Thus, our findings could be applicable to digital behavior interventions that aim to foster and support positive change (Valente 2012). For example, understanding what makes persuasion efforts successful can help overcome problems of low uptake or high attrition rates that are associated with behavior interventions in both online and offline settings.

Our findings also have implications for democracy and political campaigns where emergent influence networks play a significant role in shaping political opinions and mobilizing voters. By understanding how network position relates to persuasiveness, researchers can gain a better understanding of the strategies used by political actors to persuade and mobilize their supporters, as well as the potential risks and opportunities associated with online political communication.

Limitations and Future Work. Several factors beyond language and network features affect persuasiveness in on-

line debates, e.g., characteristics of the communicator (such as one's credibility, liking, and similarity with the original poster), the message's properties (e.g., using different kinds of arguments, narratives, fear appeals, and so on), and characteristics of the recipient (such as one's moods, defensive reactions, and personality traits) (O'Keefe 2016). While such personal characteristics may not be feasible to accurately measure in public online forums, they may confound the effect of the language and network measures examined in this study. Additionally, since we do not conduct randomized controlled experiments, we cannot establish causal claims from the above findings. Furthermore, our control group comprise a matched sample developed by (Tan et al. 2016) based on Jaccard similarity in vocabulary overlap between successful and unsuccessful arguments. Using this dataset we maintain consistency in terms of both data and methods which allows for comparisons with previous studies. Yet, we acknowledge that better state-of-the-art methods in Semantic Textual Similarity (STS) now exist, e.g., BERT (Devlin et al. 2018) and RoBERTa (Liu et al. 2019). Future work could also consider matching approaches that take into account other language features such as the length of the post. More importantly, real-world applications typically contain imbalanced data. As noted, in the original data set, only 1.09% of the arguments were persuasive, i.e., were awarded Δs. By performing our analysis on a balanced data set of matched samples, we potentially lose real-world representations that could lead to models that perform well on balanced data sets but poorly on imbalanced data sets. Finally, our work focuses on persuasion in good-faith online discussions in a single Reddit community. We recognize that good-faith discussions are not overly representative of online spaces. Hence we do not know whether and how our findings may generalize to other forms of interaction that do not require mutual trust, honesty, and a willingness to engage in civil discourse, even when there are disagreements or differences in opinion.

Ethics Statement

As researchers, we recognize the potential ethical concerns that arise when using online user-generated content in research and have taken steps to address these concerns. Our study is based solely on a publicly available data set that has already been used in previous research, such as (Tan et al. 2016; Khazaei, Xiao, and Mercer 2017; Hidey et al. 2017; Ji et al. 2018). The raw data do not contain or reveal any sensitive information about users. To further protect the privacy of users, we carefully consider the context in which the data are presented. We only show aggregated trends and do not reveal individual comments. We also acknowledge that our research findings may have social implications and therefore, to avoid any potential misapplication of our results, we take a neutral stance regarding the quality of arguments analyzed. Our focus is solely on the dynamics of persuasion and not on identifying who is right or wrong. Finally, we acknowledge the potential for manipulation and misinformation. The complexity of emergent influence networks can also create opportunities for manipulation and the spread of misinformation. By understanding the mechanisms of persuasion within these networks, we hope that researchers can develop strategies for countering the negative effects of misinformation and promoting accurate information online. While minimizing potential risks, we believe that the expected benefits of our contributions are substantial and outweigh unlikely and unintended harms.

Conclusion

In this study, we set out to investigate the impact of emergent influence networks on persuasiveness in unfacilitated online discussions. Through the use of a novel combination of social network analysis and machine learning, we were able to measure the influence of network position on persuasiveness, and demonstrate the impact of persuasiveness on successful users' network centralities over time. Our findings provide empirical support for the Elaboration Likelihood Model (ELM) of persuasion, and offer important insights into the complex social dynamics of online discourse. Looking ahead, our framework and methods could be applied in a variety of real-world settings to help organizations and individuals better understand how network position and persuasion strategies interact in digital spaces.

Acknowledgments

This work was partly supported by the U.S. National Science Foundation under Grant No. IIS-1755873.

References

Althoff, T.; Danescu-Niculescu-Mizil, C.; and Jurafsky, D. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Barabási, A.-L.; and Albert, R. 1999. Emergence of scaling in random networks. *Science*, 286(5439): 509–512.

Becker, J.; Almaatouq, A.; and Horvát, E.-Á. 2020. Network structures of collective intelligence: The contingent benefits of group discussion. *arXiv preprint arXiv:2009.07202*.

Blair, G.; Littman, R.; and Paluck, E. L. 2019. Motivating the adoption of new community-minded behaviors: An empirical test in Nigeria. *Science Advances*, 5(3): eaau5175.

Borgatti, S. P.; and Halgin, D. S. 2011. On network theory. *Organization Science*, 22(5): 1168–1181.

Breiman, L. 2001. Random forests. *Machine Learning*, 45(1): 5–32.

Centola, D. 2013. Social media as a tool in medicine. *Circulation*, 127: 2135–2144.

Centola, D. 2018. The truth about behavioral change. *MIT Sloan Management Review*.

Christakis, N. A.; and Fowler, J. H. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4): 370–379.

Christakis, N. A.; and Fowler, J. H. 2008. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21): 2249–2258.

Christakis, N. A.; and Fowler, J. H. 2009. *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown Spark.

Cialdini, R. 2016. *Pre-suasion: A revolutionary way to influence and persuade*. Simon and Schuster.

Cialdini, R. B. 1993. Influence: The psychology of persuasion (Rev. ed.). *New York: Morrow*.

Cohen, G. L.; Aronson, J.; and Steele, C. M. 2000. When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, 26(9): 1151–1164.

Constantino, S. M.; Sparkman, G.; Kraft-Todd, G. T.; Bicchieri, C.; Centola, D.; Shell-Duncan, B.; Vogt, S.; and Weber, E. U. 2022. Scaling up change: A critical review and practical guide to harnessing social norms for climate action. *Psychological science in the public interest*, 23(2): 50–97.

Conte, R.; Gilbert, N.; Bonelli, G.; Cioffi-Revilla, C.; Deffuant, G.; Kertesz, J.; Loreto, V.; Moat, S.; Nadal, J. P.; Sanchez, A.; et al. 2012. Manifesto of computational social science. *The European Physical Journal Special Topics*, 214: 325–346.

Correll, J.; Spencer, S. J.; and Zanna, M. P. 2004. An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, 40(3): 350–356.

Cross, R. L.; Cross, R. L.; and Parker, A. 2004. *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business Press.

Danescu-Niculescu-Mizil, C.; Cheng, J.; Kleinberg, J.; and Lee, L. 2012. You had me at hello: How phrasing affects memorability. *arXiv preprint arXiv:1203.6360*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Easley, D.; and Kleinberg, J. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

Fisher, R. J.; Vandenbosch, M.; and Antia, K. D. 2008. An empathy-helping perspective on consumers' responses to fund-raising appeals. *Journal of Consumer Research*, 35(3): 519–531.

Flesch, R. 2007. Flesch-Kincaid readability test. *Retrieved October*, 26(3): 2007.

Fogg, B. J. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December): 2.

Freund, Y.; and Schapire, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1): 119–139.

Gass, R. H. 2010. *Persuasion, Social Influence, and Compliance Gaining*. Allyn & Bacon.

Guilbeault, D.; Becker, J.; and Centola, D. 2018. Complex contagions: A decade in review. *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks*, 3–25.

Han, J.-T.; Chen, Q.; Liu, J.-G.; Luo, X.-L.; and Fan, W. 2018. The persuasion of borrowers' voluntary information in peer to peer lending: An empirical study based on elaboration likelihood model. *Computers in Human Behavior*, 78: 200–214.

Hidey, C.; Musi, E.; Hwang, A.; Muresan, S.; and McKeown, K. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, 11–21.

Horvát, E.; Wachs, J.; Wang, R.; and Hannák, A. 2018. The Role of Novelty in Securing Investors for Equity Crowdfunding Campaigns. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 6(1): 50–59.

Hullett, C. R. 2005. The impact of mood on persuasion: A meta-analysis. *Communication Research*, 32(4): 423–442.

Ji, L.; Wei, Z.; Hu, X.; Liu, Y.; Zhang, Q.; and Huang, X.-J. 2018. Incorporating argument-level interactions for persuasion comments evaluation using co-attention model. In *Proceedings of the 27th International Conference on Computational Linguistics*, 3703–3714.

Johnson, J. C.; Everett, M. G.; and Borgatti, S. P. 2018. Analyzing social networks. *Analyzing Social Networks*, 1–384.

Khazaei, T.; Xiao, L.; and Mercer, R. 2017. Writing to persuade: Analysis and detection of persuasive discourse. *IConference 2017 Proceedings*.

Kleinberg, J. M.; et al. 1998. Authoritative sources in a hyperlinked environment. In *SODA*, volume 98, 668–677.

Larrimore, L.; Jiang, L.; Larrimore, J.; Markowitz, D.; and Gorski, S. 2011. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1): 19–37.

Li, J.; and Zhan, L. 2011. Online persuasion: How the written word drives WOM: Evidence from consumer-generated product reviews. *Journal of Advertising Research*, 51(1): 239–257.

Liang, J.; Chen, Z.; and Lei, J. 2016. Inspire me to donate: The use of strength emotion in donation appeals. *Journal of Consumer Psychology*, 26(2): 283–288.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Lutu, P. E. N. 2019. Using Twitter mentions and a graph database to analyse social network centrality. In *2019 6th International Conference on Soft Computing & Machine Intelligence (iscmi)*, 155–159. IEEE.

Mitra, T.; and Gilbert, E. 2014. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 49–61.

O'Keefe, D. J. 1999. How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. *Annals of the International Communication Association*, 22(1): 209–249.

O'Keefe, D. J. 2009. Theories of persuasion. *The SAGE Handbook of Media Processes and Effects*, 269–282.

O'Keefe, D. J. 2016. Persuasion and social influence. *The International Encyclopedia of Communication Theory and Philosophy*, 1–19.

O'Keefe, D. J. 2018. Persuasion. In *The Handbook of Communication Skills*, 319–335. Routledge.

Opuszko, M.; Gehrke, S.; and Niemz, S. 2019. Peer Influence and Centrality in Online Social Networks. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1377–1382. IEEE.

Özaygen, A.; and Balagué, C. 2018. Idea evaluation in innovation contest platforms: A network perspective. *Decision Support Systems*, 112: 15–22.

O'Keefe, D. J.; and Jackson, S. 1995. Argument quality and persuasive effects: A review of current approaches. In *Argumentation and Values: Proceedings of the Ninth Alta Conference on Argumentation*, 88–92.

Panagopoulos, G.; Malliaros, F. D.; and Vazirgianis, M. 2020. Influence maximization using influence and susceptibility embeddings. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 511–521.

Petty, R. E.; and Cacioppo, J. T. 1986. The elaboration likelihood model of persuasion. In *Communication and Persuasion*, 1–24. Springer.

Rhue, L.; and Robert, L. P. 2018. Emotional delivery in prosocial crowdfunding success. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, 1–6. Montreal QC, Canada: Association for Computing Machinery. ISBN 978-1-4503-5621-3.

Shrum, L.; Liu, M.; Nespoli, M.; and Lowrey, T. M. 2012. *Persuasion in the Marketplace*. Sage.

Stromer-Galley, J. 2003. Diversity of political conversation on the Internet: Users' perspectives. *Journal of Computer-Mediated Communication*, 8(3): JCMC836.

Tan, C.; Niculae, V.; Danescu-Niculescu-Mizil, C.; and Lee, L. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, 613–624.

Taylor, K. L.; Kenny, N. A.; Perrault, E.; and Mueller, R. A. 2022. Building integrated networks to develop teaching and learning: the critical role of hubs. *International Journal for Academic Development*, 27(3): 279–291.

Valente, T. W. 2012. Network interventions. *Science*, 337(6090): 49–53.

Wasserman, S.; Faust, K.; et al. 1994. Social network analysis: Methods and applications.

Wegener, D. T.; and Petty, R. E. 1996. Effects of mood on persuasion processes: enhancing, reducing, and biasing scrutiny of attitude-relevant information.

Wei, Z.; Liu, Y.; and Li, Y. 2016. Is this post persuasive? Ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 195–200.

White, K.; Habib, R.; and Hardisty, D. J. 2019. How to SHIFT consumer behaviors to be more sustainable: A literature review and guiding framework. *Journal of Marketing*, 83(3): 22–49.

Wu, B.; and Shen, H. 2015. Analyzing and predicting news popularity on Twitter. *International Journal of Information Management*, 35(6): 702–711.

Zeng, J.; Li, J.; He, Y.; Gao, C.; Lyu, M.; and King, I. 2020. What changed your mind: The roles of dynamic topics and discourse in argumentation process. In *Proceedings of The Web Conference 2020*, 1502–1513.

Paper Checklist

1. Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, see *Ethics Statement*. We use publicly available data that do not contain sensitive information about users. We present aggregated trends that do not reveal individual comments or identities tied to societies or cultures.
2. Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, our work applies statistical modeling to observational data. We describe our evidence as correlational, not causal.
3. Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see *Methods* section.
4. Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see *Data* and *Ethics Statement*.
5. Did you describe the limitations of your work? Yes, see *Limitations and Future Work* in the *Discussion*.
6. Did you discuss any potential negative societal impacts of your work? Yes, see *Ethics Statement*. The expected benefits our our work are substantial and outweigh unlikely and unintended harms.
7. Did you discuss any potential misuse of your work? Yes, see *Ethics Statement*.
8. Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see *Ethics Statement*. The data are anonymized, publicly available and have been used in prior studies. Our code is available online at https://github.com/LINK-NU/ICWSM24-rCMV.
9. Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we can confirm that our paper conforms to the ethics review guidelines.