ChangeMyView: Persuasion in Online Social Networks

Meryl Ye (may43), Ruqing Xu (rx24)

May 16, 2023

1 Introduction

Online social networks have radically transformed how people communicate and exchange information. By making information available to a larger audience, social platforms like Reddit enable the sharing of innumerable ideas. While such a platform promotes discussion and debate among diverse viewpoints, issues such as disagreement and opinion polarization can also be a by-product of such interactions. Numerous works have studied these phenomena and attributed them to be the result of a combination of factors, including psychological tendencies of homophily (i.e., selective exposure or confirmation bias) and biased assimilation [*1*]. The rise and development of recommender systems that seek to increase user engagement have also been demonstrated to influence opinion dynamics [*2*]. As these technologies expand, it becomes increasingly important to understand how users form and change opinions when they are exposed to a wide variety of allies and opponents.

In this work, we investigate factors of persuasion in online social networks via the `/r/ChangeMyView` subreddit. Since each step in the opinion updating process is essentially a case of persuasion, the factors that affect persuasion are arguably also factors that affect opinion dynamics in a broad range of networks. We study factors of persuasion from three perspectives: (1) user-level, (2) post-level, and (3) structural. As these unique perspectives address varying influences of user characteristics, interaction dynamics, and network structure, we seek to emphasize the complexity of persuasion and provide insight into what elements of an argument contribute to persuasive power.

2 Background and Related Work

Reddit is a public online discussion platform that has been used in many works as a popular and rich source from which to extract a wide variety of comments from millions of users [*3, 4*]. A distinguishing feature of Reddit is its separation of relevant discussion threads into forums called subreddits that allow users to establish their own communities with guidelines that members must follow. The `/r/ChangeMyView` subreddit claims to be "a place to post an opinion you accept may be flawed, in an effort to understand other perspectives on the issue" and advises users to "enter with a mindset for conversation, not debate"[1]. The general format of `/r/ChangeMyView` (CMV) is a thread in which an original poster (OP) posts her controversial view and invites other users (i.e. challengers) to try to change her view. The subreddit follows a Delta system: challengers who

[1] https://www.reddit.com/r/changemyview/

successfully change the view of the OP are awarded with a delta (Δ) symbol. The 'DeltaBot' acts as a moderator in verifying the awarding of deltas and recording them in the `/r/DeltaLog`. The number of past awarded deltas are maintained in user flairs.

Though CMV is a relatively popular subreddit filled with opinions and diverse viewpoints, it has not been involved in as many empirical studies as other subreddits such as `/r/Politics` [*2, 5, 6*], likely due to its unconventional guidelines and coverage of miscellaneous topics. Most studies of opinion dynamics in online social networks employ sentiment analysis to track changes in user opinion [*2, 6*]. However, this assumption implies that sentiment is analogous to opinion, which is not entirely accurate. Since users in CMV explicitly indicate a change of opinion as a result of another user, there is no need to make assumptions about whether or not an opinion change has occurred, making CMV a more robust domain for analyzing persuasion and opinion dynamics.

2.1 Interaction Dynamics and Persuasion Strategies in `/r/ChangeMyView`

Tan et al. previously studied discussions in CMV to identify interaction dynamics and language factors that contribute to successful arguments [*7*]. From the analysis, Tan et al. found that challengers with an earlier entry time were more successful. Also, while back-and-forth interactions were correlated with higher success, the likelihood of changing the OP's view seemed to taper off in longer conversations. They further found that the more challengers, the more likely the OP was to change her view. Alongside interaction dynamics, linguistic choices on the challenger's part seemed to influence the OP's view as well. Tan et al. paired similar counterarguments, one successful and one unsuccessful, to explore how language features of a comment could affect its persuasive power. Among their findings, interplay between the OP's language and the challenger's language demonstrated to be highly predictive of successful persuasion. Additionally, counterarguments using intense language showed to be less effective while intensity of the OP's language was associated with a lower likelihood of her changing her mind.

Overall, we describe Tan et al.'s work as following a *comment-centered* approach. For instance, one of the methods they use in the study is the construction of discussion trees from threads. Nodes in the tree represent individual comments and directed edges indicate one comment replying to another. While Tan et al. provide a thorough study of user interactions within CMV, they do not offer much insight into the potential relationships or connections between users. Furthermore, the analysis of linguistic choices is focused on comment construction and features rather than the users behind those comments. Thus, for our study, we implement a *user-centered* approach that takes into account user characteristics in conjunction with interaction dynamics as potential factors of persuasion. As later described in Section 3, we consider a measure of homogeneity between the OP and the challenger as a novel variable in influencing successful persuasion.

2.2 Theoretical Work on Opinion Dynamics

Many theoretical works of opinion dynamics incorporate one or more factors of persuasion in the models. Notably, many have assumed the importance of "node homogeneity" on one node changing the other node's opinions. Our empirical results contest this assumption.

In discrete opinion models, [*8*] generalizes the classical voter model [*9, 10*] to incorporate a preference for opinions from similar neighbors. In the model, the possible opinions (or types) define a graph in which similar opinions are adjacent. At each step, every user selects a random user whose type neighbors his own type and adopts that opinion. Similarly, the well-known threshold model [*11*] are generalized to have greater weights on neighbors whose beliefs are similar [*12*], meaning that similar neighbors have greater persuasion power.

Models with continuous opinions often embed the preference for homogeneity in the model. Probably the most famous examples are the bounded confidence models [*13*], which hypothesize that agents only interact if they are close in opinion to each other, captured by an upper confidence bound on the distance of the two opinions. Besides that, models with biased assimilation [*1*] propose that people tend to be more receptive to opinions closer to their initial opinions but critical to opinions further away from their initial opinions. Similarly, [*14*] assumes biased assimilation in a geometric model and explore the resulting opinion dynamics.

This widespread assumption that "similarity begets persuasion", however, has yet to be empirically validated in different contexts. In particular, our results provide an initial piece of evidence that this assumption may not hold in online social networks dedicated to civil discussions.

3 Data and Methods

CMV Dataset. For our study, we use the raw data from Tan et al. that contains all of the discussion in `/r/ChangeMyView` during the periods `2013/01/01-2015/05/07` and `2015/05/08-2015/09/01`, as extracted from the Reddit API [*7*]. From this data, we obtain the following information for each post: post name, original poster (OP), and comment thread. The comment threads allow us to further analyze various comment features such as length and number of comments per challenger. We construct reply networks from information provided by the comment threads, which are elaborated in Section 4.3

User Embeddings. The SNAP library provides a dataset of subreddit embeddings [*15, 16*]. Based on publicly available Reddit data from Jan 2014 to April 2017, each subreddit in the dataset is represented as a 300-dimensional numerical vector. Two subreddit embeddings are similar if the users who post in them are similar. After compiling a list of usernames from the CMV dataset, we extracted user information from the current Reddit API that revealed public subreddits in which a user has commented. From this, we generate an embedding for each user by averaging the subreddit embeddings for all the subreddits in which the user has commented outside of CMV. Therefore, each user is generated a 300-dimensional numerical vector. Two user embeddings are similar if they post in similar subreddits. We use these user embeddings to compute measures of homogeneity between each OP and challenger. For this calculation, we first find the Euclidean distance between the embeddings for each OP-challenger pair, which we define as h. We then take $H = -h$, so a greater value is correlated with greater homogeneity. Lastly, we normalize all values of H to have a mean of 0 and standard deviation of 1.

Variable Definitions and Summary Statistics. Table 1 provides summary statistics of our data. A variable described as an "indicator" is assigned 0 for False and 1 for True. We define "challenger's success" as when a challenger is awarded a delta by the OP and "OP's conversion" as when an OP awards at least one delta to a challenger. At the user-level, on average, there is one successful challenger in 100 challengers (Panel A). Looking at the post-level, however, roughly 32 percent of the posts result in one of more awarded deltas (Panel B).

Panel A also summarizes other variables at the user-level. "Homogeneity" is a measure of similarity between users that we describe in the previous paragraph. "Number of shared subreddits" are the number of public subreddits in which both the OP and the challenger have commented. "Challenger entry order" is determined by the UTC time at which a challenger makes her first comment. "Average reply length" is the average word count of a challenger's comments within a thread. "Number of past awarded deltas" are determined by the user flair included in the challenger's comments at the time of the post. "Activity level" is the number of total comments that the user has made in CMV (as either an OP or challenger).

Panel B summarizes the variables describing the aggregate properties of a particular post. In particular, the average, variance, maximum, and minimum of challenger homogeneity describe the distribution of the homogeneity measure of challengers with the OP under a particular post. The rest of the list are the post-level version of the Panel A variables.

Panel C summarizes the variables at the network structural level in a larger sample of posts, since there is no issue with deleted users. The graph distance represents a measure of reply distance between challengers and OP, a notion we will make formal in section 4.3.

Table 1: Sample summary statistics

	(A) Mean	(B) SD	(C) Min	(D) Max
Panel A. User-level variables				
Indicator for challenger's success	0.01	0.10	0.00	1.00
Homogeneity with OP	0.00	1.00	-6.63	2.43
Number of shared subreddits with OP	0.24	0.52	0.00	6.00
Challenger entry order	30.33	46.80	1.00	762.00
Average reply length (in 100 words)	1.01	1.14	0.00	18.79
Number of past awarded deltas	5.77	16.42	0.00	132.00
Number of replies to the same post	2.03	2.27	1.00	73.00
Activity level (in 100 comments)	1.70	3.48	0.01	19.57
Indicator that challenger has been OP	0.25	0.43	0.00	1.00
N (# OP-challenger pairs): 197,643				
Panel B. Post-level variables				
Indicator for OP's conversion	0.32	0.47	0.00	1.00
Average challenger homogeneity with OP	-0.04	0.70	-5.10	1.24
Variance of challenger homogeneity with OP	0.58	0.46	0.00	4.66
Maximum challenger homogeneity with OP	0.85	0.75	-4.89	2.43
Minimum challenger homogeneity with OP	-1.82	1.12	-6.63	1.03
Number of challengers	22.77	28.99	2.00	762.00
Total past awarded deltas of all challengers	131.46	103.49	0.00	850.00
Total replies made by all challengers	46.29	66.77	2.00	1189.00
Maximum number of replies made by one challenger	6.63	5.67	1.00	73.00
Average activity level of challengers (in 100 comments)	2.17	1.30	0.06	10.94
Indicator that OP has been challenger	0.64	0.48	0.00	1.00
N (# posts): 8,680				
Panel C. Network structural level variables				
Indicator for OP's conversion	0.31	0.46	0.00	1.00
Average graph distance between challengers and OP	1.30	0.30	0.91	2.68
Number of challengers	32.64	36.58	11.00	969.00
N (# posts): 12,272				

User Roles. We also look at user roles–whether a user is exclusively an OP or a challenger, or participates in CMV as both. As seen in Figure 1, most users participate exclusively as challengers, with only about 8 percent of challengers having also been OPs. On the other hand, more than half of OPs have engaged in CMV dicussions as challengers.

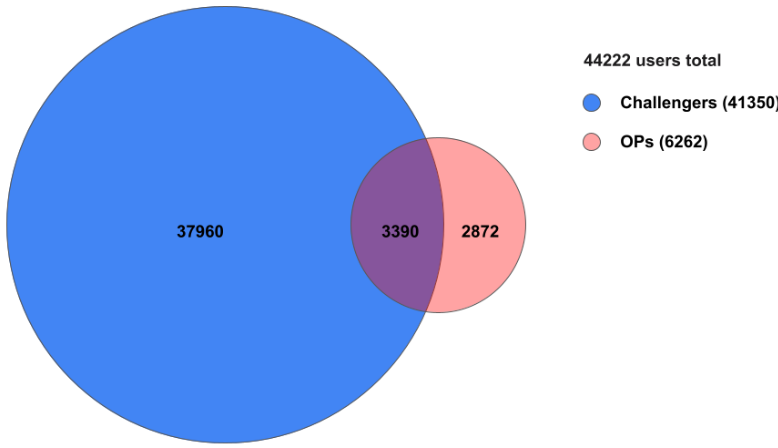

Figure 1: Venn diagram of OPs and Challengers

4 Factors of Persuasion

Our primary objective is to identify factors that affect the likelihood of persuasion in CMV discussions. We study this objective from three different perspectives. First, the user-level perspective focuses on traits of each challenger that influence their success of persuasion (i.e. the challenger is awarded a delta by the OP). Second, the post-level perspective collectively looks at traits of all challengers within a post that may indicate the OP's conversion (the OP awards a delta to any challenger). Lastly, the structural perspective addresses properties of the user network structure associated with the OP's conversion.

4.1 User-Level Perspective

In our study of factors of successful persuasion from the user-level perspective, we estimate a regression of the form:

$$\Delta_{ip} = \alpha + \beta H_{ip} + \gamma X_{ip} + \varepsilon_{ip} \tag{1}$$

where Δ_{ip} is an indicator variable that equals 1 if a challenger i is assigned a "Δ" in post p. The main variable of interest is H_{ip}, which is the measure of homogeneity between the challenger and the OP, as described in section 3. X_{ip} includes all the control variables listed in Panel A of Table 1. Each entry of the dataset is an OP-challenger pair, so this regression lets us compare successful and unsuccessful pairs and explore the features that enable a challenger to successfully persuade an OP. Specifically, we hope to identify how homogeneity between users can affect the likelihood of one changing the other's opinion.

Table 2: User-level regression results

	(A)	(B)	(C)	(D)
	All posts		Successful posts	
Variables	Univariate regression	With controls	Univariate regression	With controls
Homogeneity with OP	-0.0011^{***}	-0.0005^{**}	-0.0069^{***}	-0.0027^{**}
	(0.0002)	(0.0002)	(0.0011)	(0.0011)
Challenger entry order		-0.0001^{***}		-0.0006^{***}
		(0.0000)		(0.0000)
Average reply length (in 100 words)		0.0063^{***}		0.0308^{***}
		(0.0002)		(0.0009)
Number of past awarded deltas		0.0005^{***}		0.0022^{***}
		(0.0000)		(0.0001)
Number of replies to the same post		0.0025^{***}		0.0133^{***}
		(0.0001)		(0.0005)
Activity level (in 100 comments)		-0.0010^{***}		-0.0052^{***}
		(0.0001)		(0.0006)
Indicator that challenger has been OP		0.0022^{***}		0.0084^{***}
		(0.0006)		(0.0026)
N (# OP−challenger pairs)	197,643	197,643	40,837	40,837

Notes: Parentheses contain standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 displays the estimates from the user-level regression. Column (A) presents a univariate regression of Δ_{ip} on H_{ip} (regression (1) without variable X_{ip}). The coefficient of homogeneity is negative and significant – one standard deviation increase of homogeneity with the OP leads to 0.1 percent lower chance of persuading the OP as a challenger. The impact of homogeneity stays negative and significant after adding in control variables (column B), while its magnitude reduces by half.

The estimates on control variables are also worth discussing. The entry order of a challenger to the post appears to be negatively correlated with the success of the challenger, which confirms the finding in [7] that the success rate decreases roughly linearly when sorted by entry orders. Reply length of the challenger has a positive impact on persuasion – a 100 word increase in reply improves the success rate by an average of 0.6 percent. Number of repeated replies also have a positive impact, which captures the effect of "back-and-forth exchange" studied in [7]. As expected, the challenger's past success, as captured by the number of past awarded deltas, also has a positive predictive power on persuading the OP. Interestingly, the activity level of the challenger in CMV is negatively correlated with persuasion, which might indicate a lack of contemplation and effort for frequent challengers. Finally, we find that if a challenger has been an OP of another CMV post, she is more likely to be a successful challenger. This suggests that being able to "stand in other people's shoes" indeed helps persuasion, as shown in the CMV context.

While column (A) and (B) present the regression results in all posts, column (C) and (D) only looks at the successful posts (the post where the OP ultimately award a delta). When comparing successful and unsuccessful challengers in the successful posts, the signs of the coefficients do not change but the magnitudes become much more pronounced. In successful posts, an one standard

Table 3: User-level regression results using the discrete measure of homogeneity

	(A)	(B)	(C)	(D)
	All posts		Successful posts	
Variables	Univariate regression	With controls	Univariate regression	With controls
Number of shared subreddits with OP	-0.0018^{***}	-0.0009^{*}	-0.0079^{***}	-0.0027
	(0.0005)	(0.0005)	(0.0022)	(0.0021)
Challenger entry order		-0.0001^{***}		-0.0006^{***}
		(0.0000)		(0.0000)
Average reply length (in 100 words)		0.0063^{***}		0.0308^{***}
		(0.0002)		(0.0009)
Number of past awarded deltas		0.0005^{***}		0.0022^{***}
		(0.0000)		(0.0001)
Number of replies to the same post		0.0025^{***}		0.0133^{***}
		(0.0001)		(0.0005)
Activity level (in 100 comments)		-0.0010^{***}		-0.0052^{***}
		(0.0001)		(0.0006)
Indicator that challenger has been OP		0.0023^{***}		0.0087^{***}
		(0.0006)		(0.0026)
N (# OP−challenger pairs)	197,643	197,643	40,837	40,837

Notes: Parentheses contain standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

deviation increase in homogeneity reduces the challenger's chance of persuasion by almost 0.7 percent. We discuss the interpretation and potential mechanisms of this finding in section 5.

Our results remain robust when we adopt a discrete measure of homogeneity, the number of subreddits other than CMV in which the challenger and the OP have both posted. Note that this is a stronger definition of homogeneity than the embedding measure, since two embeddings can still be similar if the users posted on similar but not identical subreddits. Table 3 displays the estimates from regression (1) with H_{ip} replaced by the number of shared subreddits between the challenger and the OP. The signs and magnitudes of the coefficients remain very consistent with the main specification. An increase in shared subreddits with the OP by one decreases the challenger's probability of persuasion by 0.2 percent in all posts and 0.8 percent in successful posts.

4.2 Post-Level Perspective

From the post-level perspective that concentrates on the OP's conversion, we estimate a regression of the form:

$$\Delta_p = \alpha + \beta H_p + \gamma X_p + \varepsilon_p \tag{2}$$

where Δ_p equals 1 if a delta is assigned to *any* challenger in the post (signifying the OP's conversion). H_p describes the properties of the distribution of challenger homogeneity with the OP, for example, the average homogeneity, the maximum and minimum homogeneity, and the variance of the homogeneity within a single post. The control variables X_p now measure other properties of the posts, which are listed in the Panel B of Table 1. We are specifically interested in what properties of the mixture of challenger homogeneity under the same post best predict the OP's conversion.

Table 4: Post-level regression results

	(A)	(B)
	All posts	
Variables	Single contorl	All controls
Average challenger homogeneity with OP	0.0304	0.0293
	(0.0199)	(0.0199)
Variance of challenger homogeneity with OP	0.0316	0.0373*
	(0.0193)	(0.0202)
Maximum challenger homogeneity with OP	−0.0384**	−0.0415**
	(0.0185)	(0.0187)
Minimum challenger homogeneity with OP	0.0155*	0.0185*
	(0.0091)	(0.0097)
Number of challengers	0.0011***	0.0033***
	(0.0002)	(0.0007)
Total past awarded deltas of all challengers		0.0001*
		(0.0001)
Total replies made by all challengers		−0.0011***
		(0.0003)
Maximum number of replies made by one challenger		−0.0015
		(0.0016)
Average activity level of challengers (in 100 comments)		−0.0070
		(0.0052)
Indicator that OP has been challenger		0.0346***
		(0.0105)
N (# posts)	8,680	8,680

Notes: Parentheses contain standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 displays the regression results at the post level. In column (A), we estimate the regression on H_p only controlling for the total number of challengers under the post. This control is necessary since the the number of challengers are arithmetically correlated with the maximum, minimum, and variance of the distribution and thus may confound the analysis. We find that, controlling for the total number of challengers, a higher maximum homogeneity with the OP harms the likelihood of persuasion, while a higher minimum homogeneity helps. Taking together, these results are suggestive of a "sweet spot" of the challenger mixture – having someone too similar to the OP does not help (it may strengthen OP's original stance), while increase the similarity of the most dissimilar challenger facilitates persuasion (OP feels less attacked).

The results remain qualitatively and quantitatively similar when we add in more controls (column B). Notably, the variance of the challenger homogeneity appear to be positive and significant, indicating the OP's preference for a diverse opinion set. Interestingly, the number of total replies by all challengers appears to be negatively correlated with persuasion, implying that merely "bombarding" the OP helps little with persuasion. Finally, we see a significant "reciprocal" behavior of OP – the OP is more likely to award a delta if she has been a challenger herself before. Combined with the mirroring result from the challenger side, this could be indicative of a healthy interaction dynamics of CMV.

4.3 Structural Perspective

From the structural perspective, we seek to emphasize our view of threads in CMV as social networks. We construct user networks for each thread. Nodes represent users and undirected edges are created between any user pair for which one has replied to the other. We define "reply distance" as the minimum path length (number of edges) between two nodes. For instance, a challenger has a reply distance of 1 from the OP if she has directly replied to the OP.

Examples of these user reply networks are shown below. Blue nodes are OPs, red nodes are unsuccessful challengers, and green nodes are successful challengers.

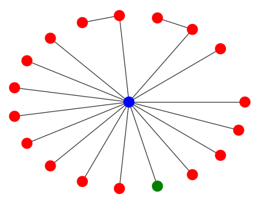

Figure 2

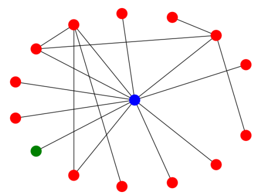

Figure 3

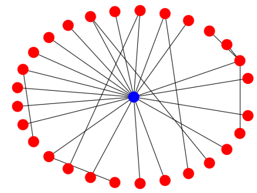

Figure 4

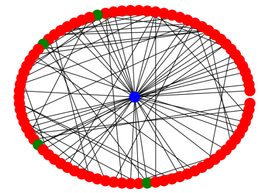

Figure 5

Figure 1 illustrates a mostly star-like network for which there is one successful challenger with a reply distance of 1. In a star-like network, almost all challengers have a reply distance of 1 from the OP. In this example, only two users have a reply distance of 2, and they were unsuccessful. Figure 2 illustrates a network with more web-like features, but still an underlying star-like structure. Unsurprisingly, the one successful challenger has a reply distance of 1 from the OP. Figure 3 illustrates another network that has a mixture of star-like and web-like features. Despite having more challengers, however, there are no successful challengers. Figure 4 illustrates a network with many challengers and therefore most closely resembles a web. There are four successful challengers, all of whom have a reply distance of 1 from the OP.

Table 5: Network structural level regression results

	(A)	(B)
	All posts	
Variables	Univariate regression	With controls
Average graph distance between challengers and OP	-0.0783^{***}	-0.1955^{***}
	(0.0142)	(0.0175)
Number of challengers		0.0016^{***}
		(0.0001)
N (# posts)	12,272	12,272

Notes: Parentheses contain standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In our study of factors of successful persuasion from the structural perspective, we estimate a regression of the form:

$$\Delta_p = \alpha + \beta D_p + \gamma X_p + \varepsilon_p \tag{3}$$

where Δ_p equals 1 if a delta is assigned to *any* challenger in the post (signifying the OP's conversion). D_p describes the average reply distance that an OP has from her challengers. The control variable X_p measures the number of challengers for an OP.

Table 5 displays the regression results at the structural level. In column (A), we estimate the regression on D_p with no controls. We find that a lower average reply distance between the OP and her challengers increases the likelihood of persuasion. The impact of average reply distance stays negative and significant after adding in the control variable (column B), with its magnitude more than doubled. These results suggest that an OP is more likely to change her view when she has directly interacted with her challengers.

5 Discussion

Negative impact of user homogeneity. One of the most unexpected results from our analysis is the negative impact of homogeneity on the likelihood of persuasion at the user-level. We provide several potential mechanisms of this effect: (1) the challenger with a more distinct background and viewpoint from the OP may have a stronger incentive to persuade the OP and therefore may exert more effort in terms of word choice and argument construction. (2) the OP in the CMV community commits to keeping an open mind and therefore appreciates a fresh perspective coming from a very different challenger. Future work could look more into the correlation between homogeneity and linguistic choices of the challengers as well as explore other possible mechanisms of such an effect.

Reciprocal behavior across time. Both the user-level and post-level analysis reveal strong reciprocal behaviors of the OPs and the challengers. Having been a challenger increases the OP's overall likelihood of assigning any delta, while having been an OP increases the likelihood of the challenger's success in persuasion. This over-time relationship is indicative of the importance of past interactions on altering current opinion dynamics. Most of the existing literature focus on a "memoryless" procedure of opinion updating, which falls short of capturing over-time interactions between the users. We encourage the empirical validation and theoretical modeling of the impact of over-time interactions on opinion dynamics.

Missing user data. In our extraction of user data from the current Reddit API, we were unable to generate embeddings for approximately 34 percent of users present in the CMV dataset. This shortcoming is because either (1) the user has been deleted since the posting of her comment or (2) the user has made their Reddit activity private.

Predictions. Tan et al.'s original dataset is separated into a training period and a heldout period. From their analysis of the training period data, they made predictions about factors of persuasion and subsequently tested them in the heldout period data. In our work, we combine the data from the training and heldout periods. Our results are drawn from this combined dataset. For a future study, we consider drawing predictions from this work and testing them on new CMV data.

Alternative factors of persuasion. There remain factors of persuasion that we still have yet to examine. For example, the degree of nodes from the structural perspective may reveal new findings about interactions among challengers. Changing the OP's mind may sometimes be a collective effort rather than an individual effort. Other alternative factors include user's total Reddit activity, interactions between users outside of CMV, and user karma.

6 Conclusion

In this work, we seek to identify factors contributing to successful persuasion in the online discussion forum `ChangeMyView`. In addition to analyzing how homogeneity between users may influence persuasive power, we report results from user-level, post-level, and structural-level perspectives that portray the complexity of changing a person's opinion. A successful argument is not solely dependent on the challenger, but also influenced by contextual interactions and other individuals involved in the discussion. While having less homogeneity with the OP decreases a challenger's change of success at changing her mind, a greater probability of the OP's conversion requires a mixture of similar and dissimilar challengers. Furthermore, the more direct interactions an OP has with her challengers, the more likely she is to change her view.

Overall, our findings offer novel insights into factors of persuasion in online social networks, taking into account user characteristics as well as interaction dynamics and structure. As the micro-foundation for how opinions propagate in online social networks, persuasion is an important process to understand, for its implications impact all kinds of social networks.

References

Pranav Dandekar, Ashish Goel, and David T Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.

Uthsav Chitra and Christopher Musco. Analyzing the impact of filter bubbles on social network polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 115–123, 2020.

Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2):20563051211019004, 2021.

Alexey Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. *The Anatomy of Reddit: An Overview of Academic Research*, pages 183–204. 05 2019.

Nikita Bhalla, Adam Lechowicz, and Cameron Musco. Local edge dynamics and opinion polarization. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, WSDM '23, page 6–14, New York, NY, USA, 2023. Association for Computing Machinery.

Abir De, Sourangshu Bhattacharya, Parantapa Bhattacharya, Niloy Ganguly, and Soumen Chakrabarti. Learning a linear influence model from transient opinion dynamics. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, page 401–410, New York, NY, USA, 2014. Association for Computing Machinery.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, page 613–624, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee.

David Kempe, Jon Kleinberg, Sigal Oren, and Aleksandrs Slivkins. Selection and influence in cultural dynamics. *Network Science*, 4(1):1–27, 2016.

Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

Richard A Holley and Thomas M Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

Scott Cox, Kathy J Horadam, and Asha Rao. The spread of ideas in a weighted threshold network. In *Complex Networks & Their Applications V: Proceedings of the 5th International Workshop on Complex Networks and their Applications (COMPLEX NETWORKS 2016)*, pages 437–447. Springer, 2017.

Hegselmann Rainer and Ulrich Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. 2002.

Jan Hazła, Yan Jin, Elchanan Mossel, and Govind Ramnarayan. A geometric model of opinion polarization. *Mathematics of Operations Research*, 2023.

Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 933–943. International World Wide Web Conferences Steering Committee, 2018.

Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1269–1278. ACM, 2019.