Modeling the Factors of User Success in Online Debate

Esin Durmus
Cornell University
ed459@cornell.edu

Claire Cardie
Cornell University
cardie@cs.cornell.edu

Abstract

Debate is a process that gives individuals the opportunity to express, and to be exposed to, diverging viewpoints on controversial issues; and the existence of online debating platforms makes it easier for individuals to participate in debates and obtain feedback on their debating skills. But understanding the factors that contribute to a user's success in debate is complicated: while success depends, in part, on the characteristics of the language they employ, it is also important to account for the degree to which their beliefs and personal traits are compatible with that of the audience. Friendships and previous interactions among users on the platform may further influence success.

In this work, we aim to better understand the mechanisms behind success in online debates. In particular, we study the relative effects of debaters' language, their prior beliefs and personal traits, and their social interactions with other users. We find, perhaps surprisingly, that characteristics of users' social interactions play the most important role in determining their success in debates although the best predictive performance is achieved by combining social interaction features with features that encode information on language use during the debate.

CCS Concepts

• **Information systems** → *Web applications*; *Social networks*; • **Social and professional topics** → *User characteristics.*

Keywords

online debate; social interactions; personal traits; language; persuasion

ACM Reference Format:
Esin Durmus and Claire Cardie. 2019. Modeling the Factors of User Success in Online Debate. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308558.3313676

1 Introduction

Previous work from Natural Language Processing (NLP) and Computational Social Science (CSS) that studies argumentative text and its persuasive effects has mainly focused on identifying linguistic features that are indicative of effective argumentation strategies [13, 16, 17, 33, 37, 41, 43]. For example [43] has shown that language characteristics such as conversational flow, are predictive of the success of a debater in persuading an audience member to change their stance *during a debate.* [10] has further shown that the prior beliefs of the audience with respect to the topic of discussion and the stance/beliefs of the debaters are at least as important as the linguistic characteristics of the debate in predicting the successful debater. At the same time, it has been shown that there is a strong relationship between users' social interactions and their influence on social media. For example, [35] and [7] and have shown that individuals with more activity and personal engagement are more influential on Twitter. We **hypothesize** that success in persuasion might also depend on an individual's social interactions and engagement with other users (on the debate platform) **over time**. For example, being more engaged with others over time may expose an individual to more diverse ideas and people, which in turn could foster argumentation skills that are more applicable to convincing a more diverse audience. Focusing on only individual debates and discussion threads, prior work has not investigated the relative effect of individuals' social interactions, personal traits and language use on their success in persuasion. To understand the relative effect of these factors on users' success in persuasion, we focus on online debates and study success over a user's lifetime by looking at interactions and engagement with the community over time, rather than focusing on success in individual debates.

Our study employs the DDO (debate.org) dataset [10]. Its extensive user information and multiple well-structured debates/interactions per user provides a unique opportunity to study users' success over time while accounting for the effect of individuals' social interactions, personal traits and language use. Users provide demographic information as well as their stance on controversial topics. They interact with one another in many ways: 1) debating 2) evaluating the performance of other debaters, 3) commenting on debates, 4) asking/answering opinion questions, 5) voting in polls, 6) creating polls 7) becoming friends. Ultimately, we find that the characteristics of individuals' social interactions (e.g. their friendships and voter network) are an important component in predicting their success in debates. The best predictive performance is achieved by combining social interaction features with language use features (e.g. vocabulary diversity and the extent to which their language use matches that of their opponent). **Contributions.** To the best of our knowledge, this is the first work to study computationally 1) the factors of an individual's overall success in online debating **over time** 2) by considering the combined effect of **social interactions**, personal traits and language use on success[1].

2 Dataset

Available User Information. The DDO (debate.org) dataset contains personal trait information for 45, 348 users including demographics such as gender and ethnicity as well as users' beliefs, such as their political ideology, religious ideology and stance on

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313676

[1]The dataset is publicly available at http://www.cs.cornell.edu/ esindurmus/.

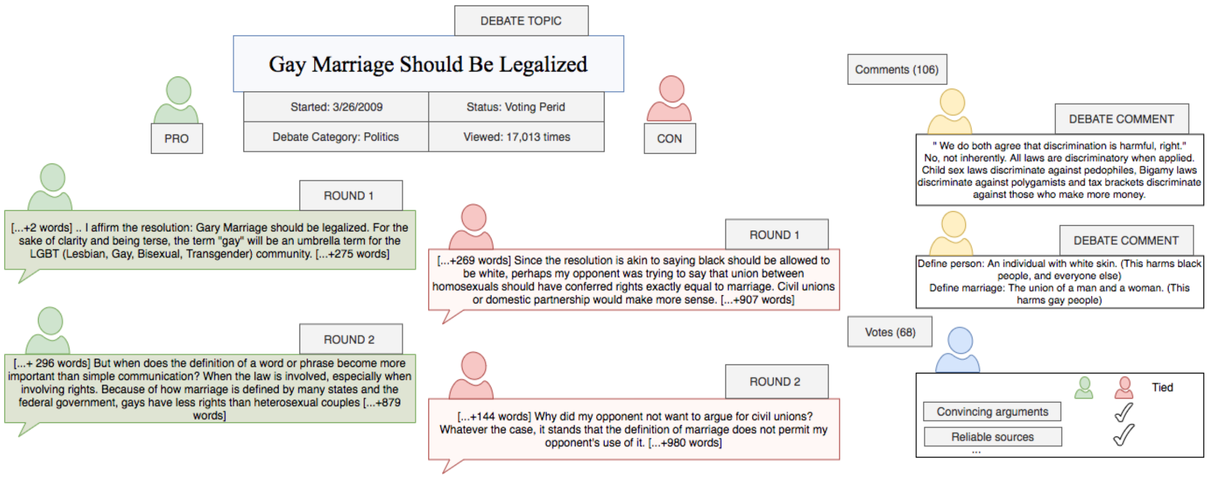

Figure 1: Example debate for the claim "*Gay Marriage Should Be Legalized*". In each round, opposing sides (PRO and CON) present their argument. Users on the platform give feedback and evaluate each debater's performance by providing comments and votes.

controversial issues. For this work, we augment DDO with social interaction data associated with each user: the debates that they participated in, voted on, and commented on; opinion questions asked, opinion arguments, poll votes and poll topics of the users[2]. Users can interact with others in more ways than simply participating in debates on DDO. For example, they can post poll and opinion questions/arguments. This allows users to exchange opinions, and to be exposed to new perspectives without engaging in a formal, structured argument. Opinion questions and arguments are more open-ended whereas polls require users to pick from a set of predefined options. An example opinion question and opinion argument are provided below:

Example Opinion Question. "Does God exist?"[3]

Example Opinion Argument. "He probably does not exist. I don't think that it's possible to say yes or no either way. We can only conclude that there is more logical evidence to say that a God probably does not exist, ..."

Debate structure. Each debate consists of multiple rounds, in which each of the two opposing debaters provides their argument. Debaters have a single opportunity to present their argument in any given round. The majority of debates have three or more rounds. There are 23 distinct debate categories. The most common debate categories are *Politics*, *Religion*, and *Society*. Figure 1 shows an example debate on GAY MARRIAGE as well as some of the comments and votes provided for this debate by other users. We see from the example that unlike monological argumentation, the debaters' arguments are shaped not only by their own opinions, but also by the way their opponents' express their opinions: debaters tend to refer to points made by their opponents throughout the debate. Voters evaluate debaters according to the quality and persuasiveness of their arguments and also provide an overall score for each debater. The winner of the debate is determined according to the total score from the voters.

[2]The dataset includes 77,655 debates, 592,390 comments, 197,231 debate votes, 262,222 poll votes, 20,300 poll topics, 10,045 opinion questions, and 61,183 opinion arguments.
[3]Full discussion on the topic can be found at https://www.debate.org/opinions/does-god-exist.

3 METHODOLOGY

In this section we describe the methods used to investigate the underlying dynamics of success in online debate.

3.1 User Success

We compute the overall *success* in debate for user u as:

$$success_u = \frac{\text{\# of debates } u \text{ won}}{\text{\# of total debates } u \text{ participated in as a debater}}$$

We treat users with $success_u \geq 70\%$ as *successful*, $success_u \leq 30\%$ as *unsuccessful* and $30\% < success_u < 70\%$ as *mediocre*.

3.2 Prediction Task

To understand the relative effect of users' personal traits, social interactions and language on their *success*, we study the following prediction task: **given a pair of debaters where one of them is *successful* and other is *unsuccessful* over the second and third stage of their lifetime, predict the *successful* one**. Note that while determining our label for success, we consider only the debates in the second and third stage of the a user's lifetime, to be able to study the relative effect of *success* in their first life stage (*success prior*) vs. other factors in a controlled way. We experiment with two settings where we control for the effect of *debate experience* and *success prior* respectively.

SETTING 1. To **control the effect of debate experience** in *success*, we create the pairs by **matching the users according to the number of debates** that they participated in (i.e. users within a pair have the same number of debates[4].

[4]There are 2154 such pairs in our dataset.

Aspect	Features
Personal traits	1) **match of the personal traits (e.g. gender, political ideology, religious ideology and ethnicity) with friends and voters**. 2) **opinion similarity with friends and voters**.
Social Interactions	1) **participation features** : # of comments, # of votes, # of friends, # of opinion questions and arguments, # of voted debates, # of poll votes and topics. 2) **friendship network features** : degree, degree centrality, page rank scores. 3) **voter network features**: in-degree, out-degree, in-degree centrality, out-degree centrality, page rank, hub and authority scores.
Language	1) **features of debaters' own language** : # of words, # of definite articles, # of indefinite articles, # of person pronouns, # of positive words, # of negative words, # of hedges, # of swear words, # of punctuation, # of links, average sentiment, type-token ratio, # of quotes, distribution of POS tags, distribution of named entities, BOW. 2) **features to encode the interplay** : exact content word match, exact stop word match, content word match with synonyms.

Table 1: Features

SETTING 2. Given that we're interested in understanding the factors that correlate with *success*, we **control for the *success prior*** in a very specific way – we only consider users that were *unsuccessful* in their initial life stage (**success prior**$\leq$ 30%[5]). This allows us to directly study the factors that are correlated with users that were initially *unsuccessful*, but later went on to become *successful* debaters.

3.2.1 Personal Traits It has been shown that characteristics of the audience's personality [22], and the degree to which the debaters' beliefs match with that of the audience [10], are important to consider in persuasion studies. We further investigate this effect in debaters' *success* over their lifetime. We also extend this study by considering additional personal traits, such as the degree to which debaters' demographics (e.g. gender and ethnicity) matches with those of their friends and the voters participating in the debates. We extract features to encode the similarity for a user's opinion, political ideology, religious ideology, gender, and ethnicity with that of her friends and voters. To compute opinion similarity, we used the information about users' opinions on 48 different controversial topics which they share on their user profiles and we measure the match of the opinions for these controversial topics[6].

3.2.2 Social Interactions The users interact with each other on the platform in following ways: 1) debating 2) evaluating the performance of other debaters, 3) commenting on debates, 4) asking/answering opinion questions, 5) voting in polls, 6) creating polls, 7) becoming friends.

We hypothesize that modeling these interactions is important to understand the differences between how *successful* and *unsuccessful* users interact on this platform, and whether or not these are important factors for success. The ability to interact with others in a myriad of different ways, provides users with ample opportunity to learn interesting new strategies and improve their skills over time, as they are exposed to a diverse set of perspectives.

[5]There are 957 such pairs in our dataset.
[6]We consider issues where users identified their side as either PRO or CON and measure the similarity of their opinion for these issues with their friends and voters.

Friendship network. We represent the friendship network as an **undirected** graph $G = (V, E)$ where V represents the set of users, and E represents the set of edges where $(x, y) \in E$ if $x \in V$ and $y \in V$ are friends.

Voter network. We represent the voter network as a weighted **directed** graph $G = (V, E)$ where V represents the set of users, and E represents the set of edges where $(x, y) \in E$ if $x \in V$ voted in a debate in which $y \in V$ participated as a debater. The weight of the graph represents how many times x voted in debates y was a debater. Note having (x, y) edge in the graph **does not** imply that x voted for y in a debate.

Using HITS algorithm [19], we compute hub and authority scores for each node (user) in the voter network graph. We find that *successful* users have, on average, a significantly higher hub score than *unsuccessful* users ($p < 0.001$).

3.2.3 Language To capture the linguistic style of the debaters' language and its relationship to their *success*, we use textual features that encode 1) users' own language and 2) the interplay between users' and their opponents' language.

Modeling users' own language. We extract features from the text of users' debates, opinion questions, opinion arguments, poll votes and poll topics. These features includes # of words, word category features (e.g. # of personal pronouns, # of positive and negative words), structural features (e.g. distribution of POS tags and named entities) and features to encode the characteristics of the entire language (e.g. type-token ratio)

Modeling the interplay between the debaters' and opponents'. We measure the interplay between debaters and their opponents by measuring how similar a debater's language is to the previous statement made by her opponent. To measure the similarity of a debater's language (D) to that of the opponent's (O) in a round, we look at # of content words that are in both D and O, # of stop words that are in both D and O and # of content words that are in D and have synonyms in O.

The *content word match with synonyms* feature aims to capture the cases where the opponent refers to similar concepts but doesn't necessarily use the exact same words as the debater.

		SETTING 1			SETTING 2		
	Feature	Precision(%)	Recall(%)	F1(%)	Precision(%)	Recall(%)	F1(%)
	(1) Majority	$26.47_{\pm 1.11}$	$51.44_{\pm 1.08}$	$34.95_{\pm 1.22}$	$26.67_{\pm 1.61}$	$51.62_{\pm 1.56}$	$35.16_{\pm 1.76}$
	(2) Debating experience	$52.70_{\pm 2.91}$	$52.04_{\pm 1.77}$	$41.76_{\pm 2.06}$	$46.00_{\pm 0.89}$	$50.16_{\pm 1.02}$	$38.98_{\pm 3.92}$
	(3) Success prior	$65.20_{\pm 0.77}$	$64.39_{\pm 0.65}$	$63.63_{\pm 0.50}$	$55.60_{\pm 0.93}$	$55.07_{\pm 0.39}$	$52.10_{\pm 0.47}$
Personal Traits	(4) Overall similarity with voters	$61.93_{\pm 1.60}$	$60.86_{\pm 1.70}$	$59.44_{\pm 1.67}$	$56.55_{\pm 2.43}$	$55.69_{\pm 1.31}$	$52.68_{\pm 1.47}$
	(5) Overall similarity with friends	$62.70_{\pm 0.86}$	$59.98_{\pm 1.05}$	$56.94_{\pm 1.14}$	$55.87_{\pm 3.43}$	$54.23_{\pm 2.35}$	$47.52_{\pm 3.18}$
Social Interactions	(6) Participation features	$67.78_{\pm 1.66}$	$66.02_{\pm 2.33}$	$64.82_{\pm 2.70}$	$59.39_{\pm 4.09}$	$57.68_{\pm 2.34}$	$55.08_{\pm 3.16}$
	(7) Friendship network features	$64.23_{\pm 1.40}$	$63.60_{\pm 1.40}$	$62.92_{\pm 1.35}$	$57.94_{\pm 1.87}$	$57.16_{\pm 1.50}$	$55.41_{\pm 1.67}$
	(8) voter network features	$72.39_{\pm 0.19}$	$70.75_{\pm 0.34}$	$70.20_{\pm 0.70}$	$70.54_{\pm 1.78}$	$69.91_{\pm 1.79}$	$69.65_{\pm 1.76}$
	(6) + (7) + (8)	$72.67_{\pm 0.73}$	$72.29_{\pm 0.93}$	$72.12_{\pm 1.03}$	$71.66_{\pm 0.71}$	$71.47_{\pm 0.51}$	$71.38_{\pm 0.51}$
Language	(9) # of words	$70.37_{\pm 1.41}$	$70.15_{\pm 1.55}$	$69.97_{\pm 1.59}$	$65.78_{\pm 0.85}$	$64.99_{\pm 1.03}$	$64.41_{\pm 1.16}$
	(10) Features of debaters' interplay	$62.11_{\pm 1.09}$	$62.07_{\pm 1.03}$	$61.92_{\pm 1.01}$	$57.47_{\pm 1.42}$	$57.16_{\pm 1.31}$	$56.41_{\pm 1.29}$
	(11) Features of debaters' own language	$72.65_{\pm 2.45}$	$72.66_{\pm 2.45}$	$72.64_{\pm 2.44}$	$64.48_{\pm 0.74}$	$64.37_{\pm 0.90}$	$64.24_{\pm 0.97}$
Combinations	(6) + (7) + (8) + (11)	$78.49_{\pm 1.29}$	$78.46_{\pm 1.32}$	$78.45_{\pm 1.32}$	$75.44_{\pm 0.90}$	$75.44_{\pm 0.90}$	$75.43_{\pm 0.89}$
	(6) + (7) + (8) + (10) + (11)	$\mathbf{81.63_{\pm 1.63}}$	$\mathbf{81.62_{\pm 1.65}}$	$\mathbf{81.61_{\pm 1.65}}$	$\mathbf{78.06_{\pm 0.88}}$	$\mathbf{78.05_{\pm 0.89}}$	$\mathbf{78.05_{\pm 0.88}}$

Table 2: Prediction Task Results for SETTING 1 and SETTING 2. For both SETTING 1 and SETTING 2, voter network features are the most predictive social interaction features. Combining interaction and language features achieves the best predictive performance.

The full list of features modeling the aspects of personal traits, social interactions and language features is shown in Table 1.

3.3 Prediction Results

We use weighted logistic regression and choose the amount and type of regularization ($\ell 1$ or $\ell 2$) by grid search over 5 cross-validation folds. We compute **weighted** precision, recall and F1 scores.

In SETTING 1, we create user pairs (u_1,u_2) where:

- u_1 and u_2 have an equal number of debates they participated in as debaters.
- One of u_1 or u_2 is *successful* and the other one is *unsuccessful* over the second and third stage of their lifetime [7].

In SETTING 2, in addition to the requirements of SETTING 1, we also require u_1 and u_2 to both have *success prior* ≤ 0.3.

Task. For both SETTING 1 and SETTING 2, we aim to predict whether u_1 or u_2 is *successful* over the second and third stage of her lifetime.

In SETTING 2, by only studying user pairs with low *success priors*, we aim to understand the factors that are important for a user to improve as a debater over time.

3.3.1 Results for SETTING 1 Table 2 shows the results for SETTING 1 and SETTING 2. We compare our model with three simple baselines – majority, debating experience, and success prior. For the majority baseline, for each example, we predict the most common label in the training data. For debating experience baseline, we use # of debates as the only feature to predict the *successful* debater. For success prior baseline, we pick the user with the higher *success prior* as successful.

In SETTING 1, since we do not control for the *success* in the first life stage, we see that the *success prior* information alone can achieve 63.63% F1 score. This implies that there is a correlation between users' *success* in their early life stage and later life stages. This factor may be related to users' prior debating skills. We observe that the features that encode debaters' overall similarity with voters and friends achieves significantly better F1 score than majority and debating experience baselines. However, these features do not have as high a predictive power as the *success prior*. We perform an ablation study for participation features, friendship network features and voter network features. We find that voter network features are significantly more predictive than the baselines, personal trait features and other social interaction features. We also perform an ablation study for the language features and find that # of words is a very predictive feature of *success*. When we combine the language features with the interaction features, we get the best predictive performance (81.61% F1 score) for this task which is significantly better than the baselines. This indicates that it is important to account for both social interaction and language factors to determine the *successful* debater since these two components encode different kinds of information about the users.

3.3.2 Results for SETTING 2 In this task, by controlling for *prior success*, we aim to understand the factors that are correlated with *success* by reducing the effect of prior debating skills of the users. As shown in Table 2, the F1 score for the *success prior* baseline is not as quite as high as in SETTING 1, since we control for this aspect by ensuring both users in the pair are *unsuccessful* in their initial life stage. However, this does not necessarily mean that the two paired users will have the exact same success prior, which explains why success prior still performs better than the other baselines. We do not observe any significant difference between the performance of the features encoding personal traits, participation, and the baseline. However, consistent with the SETTING 1, we see that features of the voter network are significantly better (69.65%) in predicting *success*. Although language features achieve significantly better F1 score than the baseline, they perform significantly worse than

[7] We consider success only over the second and third stage of users' lifetime in our prediction task, in order to study the effect of *success prior* vs. the other aspects. We use the success in the first life stage as *success prior*.

the voter network features. Similar to SETTING 1, combining these language features with the social interaction features improves the performance significantly (78.05% F1 score).

3.3.3 Feature Analysis

To understand the important social interaction and language features, we 1) compute the correlation coefficients for the feature values and the labels, 2) analyze the coefficients of the logistic regression classifier and 3) apply recursive feature elimination method [15] to rank the features according to their importance. In this section, we present the features that are consistently important with respect to each of these methods.

Analysis of Social Interaction Features. We find that the **most important** social interaction features for SETTING 1 are authority score, hub score, in and out-degree centrality and the page rank of the voter network. Note that all these important features are **positively correlated** with *success*. Although participation and friendship network features (e.g. # of voted debates, degree of the user node in friendship network) are also positively correlated with *success*, the correlation values for these are not as high as the ones of the voter network features. We also find high correlation between some of the user activities. For example, users with more # of comments are more active in making friends, voting, providing poll votes and they have higher centrality value in the friendship network. Perhaps surprisingly, we do not observe any correlation between # of voted debates and hub/authority scores in voter network. However, we see a highly positive correlation between hub scores, authority scores, in-degree centrality, out-degree centrality and page rank values of voter network. This implies that *success* is not only about the quantity of voted debates but also about the characteristics of the debaters that are involved in these debates, because the hub score of a user is influenced by the authority scores of the debaters they vote for. Similarly, the authority score of a user is influenced by the hub scores of the voters that participate in her debates. Therefore, besides the frequency of interactions, type of interactions and characteristics of users involved in the interactions are important to take into account. Consistent with SETTING 1, in SETTING 2, the most important features (positively correlated with *success*) are authority score, hub score, in and out-degree centrality and the page rank of the voter network. We observe the same patterns of user activities and authority and hub scores as in SETTING 1.

Analysis of Language Features. We find that # of words is positively correlated with *success*. It may be the case that longer text may convey more information and explain the points more explicitly [29, 30]. Bag of words features are not as predictive as the # of words feature. For both SETTING 1 and SETTING 2, we observe that the value of average sentiment is negatively correlated with *success*. The reason for this may be that negative information is more attention grabbing than positive information [9, 18, 34] since people are more used to seeing arguments that are phrased in a more positive way [25]. We also find that *type-token ratio* (diversity of language) is negatively correlated with *success* for both settings. It may be the case that people who talk about a smaller set of topics gain expertise on these topics over time; therefore, they may be more *successful*. We observe that other textual features are positively correlated with *success* for both of these settings. However, the degree of correlation is not as high as it is for type-token ratio and sentiment.

4 UNDERSTANDING THE LOSS OF *SUCCESS*

In the previous section, we show that social interaction and language features are important to predict *successful* debaters. Our findings are consistent for the case when 1) we only control for users' debating experience and 2) we also control for users' *success prior*. Users' participation, the types of interactions they have on the platform, and the characteristics of the users they interact with are predictive of their *success*, regardless of a their prior expertise in debating (encoded by the *success prior*).

In SETTING 1, since we did not control for the *success prior*, we studied the factors that are important for a user to become *successful* in their second and third life stages, regardless of their *success* in the beginning. In SETTING 2, we studied the factors that are important for an *unsuccessful* users to improve their performance and become *successful* over time. As a natural follow-up, we would also like to understand what factors are correlated with users who are initially *successful*, but later become *unsuccessful* in their lifetime. To do that, in SETTING 3, in addition to the requirements of SETTING 1, we have an additional criterion for all user pairs (u_1,u_2):

- u_1 and u_2 both have *success prior* ≥ 0.7 [8].

4.1 Results

As shown in Table 3, features of personality traits, social interactions and language perform significantly better than the baselines. For this task, *success prior* baseline performs relatively worse than in the previous two settings. Upon closer examination, we observed that the variance of success priors for this task is an order of magnitude smaller than in SETTING 2, and therefore, as a possible explanation, the *success priors* may not be as predictive for this task.

In social interaction features, similarity with friends is the most predictive feature. However, participation features perform significantly better than the features of personal traits. For this task, contrary to SETTING 1 and SETTING 2, we see that participation features are the most predictive in the set of social interaction features. This implies users' participation is important for them to remain *successful*. Lower participation could be a contributing factor for these users to eventually become unsuccessful. Although friendship and voter network features are still significantly more predictive than the baselines, they are not as highly predictive as the participation features. For users with high *success priors*, continued participation may be the most important aspect of their social interaction. We observe that language features alone achieves a similar performance as the social interaction features. Consistent with the SETTING 1 and SETTING 2, combining social interaction and language features gives the best predictive performance (73.43% F1 score).

Analysis of Social Interaction Features. The **most important** social interaction features include # of voted debates, degree of the user node in the friendship network, in-degree, out-degree, in-degree centrality and out of the user node in the voter network. All these features are indicative of a higher participation on the platform and they are positively correlated with staying *successful*.

[8] We have 700 user pairs with these criteria.

	Feature	Precision(%)	Recall(%)	F1(%)
	(1) Majority	$26.97_{\pm 2.69}$	$51.86_{\pm 2.62}$	$35.46_{\pm 2.95}$
	(2) Debating experience	$53.77_{\pm 2.95}$	$52.43_{\pm 2.91}$	$43.02_{\pm 6.19}$
	(3) Success prior	$39.94_{\pm 7.63}$	$51.00_{\pm 2.23}$	$36.04_{\pm 2.35}$
Personal Traits	(4) Overall similarity with voters	$55.17_{\pm 1.58}$	$55.00_{\pm 2.36}$	$53.94_{\pm 2.99}$
	(5) Overall similarity with friends	$66.38_{\pm 4.11}$	$63.43_{\pm 2.77}$	$60.87_{\pm 3.33}$
Social Interactions	(6) Participation features	$68.88_{\pm 3.57}$	$68.00_{\pm 2.86}$	$67.88_{\pm 2.96}$
	(7) Friendship network features	$65.60_{\pm 4.83}$	$64.00_{\pm 3.81}$	$62.81_{\pm 3.73}$
	(8) voter network features	$64.36_{\pm 1.57}$	$62.72_{\pm 2.37}$	$61.44_{\pm 2.87}$
	(6) + (7) + (8)	$67.80_{\pm 1.86}$	$67.14_{\pm 1.43}$	$66.97_{\pm 1.42}$
Language	(9) # of words	$67.63_{\pm 3.90}$	$66.57_{\pm 2.70}$	$66.29_{\pm 2.39}$
	(10) Features of debaters' interplay	$58.76_{\pm 2.03}$	$57.43_{\pm 0.86}$	$56.60_{\pm 0.93}$
	(11) Features of debaters' own language	$68.47_{\pm 0.21}$	$68.14_{\pm 0.14}$	$68.10_{\pm 0.17}$
Combinations	(6) + (7) + (8) + (11)	$69.32_{\pm 2.48}$	$69.00_{\pm 2.42}$	$69.00_{\pm 2.41}$
	(6) + (7) + (8) + (10) + (11)	$\mathbf{73.60_{\pm 0.80}}$	$\mathbf{73.43_{\pm 0.70}}$	$\mathbf{73.43_{\pm 0.72}}$

Table 3: Prediction Task Results for loss of *success*. Participation features are the most important social interaction features. Combining the social interaction features with the language features gives the best prediction performance.

Although the other social interaction features, such as authority and hub scores of the voter network are also positively correlated with *success*, the value of correlation for these is not as high as the previously mentioned features. For users who are initially *unsuccessful*, participation alone may not be enough for them to become *successful* debaters – the type of interactions and the characteristics of people with whom they interact are crucially important for their *success*. On the other hand, users who are initially *successful* may already be experienced debaters, and staying active and participating may be sufficient for them to remain *successful*.

Analysis of Language Features. As in SETTING 1 and SETTING 2, # of words is positively correlated with staying *successful*. We find that the # of first person pronouns is the language feature with the highest positive correlation with staying *successful*. We observe that users who refer to their personal experiences and opinions use first person pronouns more often. It may be the case that debaters may try to appeal to logos by citing personal experience [8]. Consistent with SETTING 1 and SETTING 2, the value of average sentiment is negatively correlated with staying *successful*.

5 Related Work

5.1 Modeling Social Interactions

There has been a tremendous amount of research on understanding user interactions and behaviour on social media [2, 4, 5, 14, 20, 21, 23, 24, 27, 42]. For example, [42] analyze the interaction graphs of Facebook user traces and show that interaction activity on Facebook is significantly skewed towards a small portion of each user's social links. [21] investigates how people interact in multiple online social networks. Although there is a lot of work on understanding user behavior on social media sites such Facebook and Twitter, the work on understanding user behaviour on online debate platforms has been limited. [35] is the most similar to our work, in that the authors study the effect of interaction dynamics, such as participant entry order and degree of back-and-forth exchange in the discussion, on success in changing an opinion holder's stance in a thread. Note that unlike our study, this work does not consider the effect of social interaction features (such as friendship network or voter network) on users' *success*. Moreover, in our work, we study the overall *success* of users over their lifetime, rather than a single debate or discussion thread.

5.2 Language in Persuasion

Previous work in psychology and linguistics has studied the effect of language on people's perception of the arguments and persuasion [3, 12, 25, 39]. These studies explore the effects of *framing*, – whether presenting different formulations of the same problem (e.g. presenting something as a gain or loss) would result in different effects on people. Besides the effects of framing, researchers investigate the effects of structural and lexical characteristics of language on success. For example, [36] found that using concrete words yields more positive outcomes than using abstract words, in terms of convincing customers to have a positive attitude towards a product.

NLP (Natural Language Processing) studies in argumentation mining focus on 1) identifying the structure of the persuasive aspects of text and relationships between them [1, 26, 28, 31, 32, 38, 40] and 2) the characteristics of persuasive text [6, 13, 16, 17, 33, 37, 41, 43]. There is very limited work in NLP that considers the factors of the individuals who are involved in the argumentation [11, 22]. [22] analyzes the effect of people's personality traits on their perceptions of the monologic arguments. However, in their work, they do not have any information about the source of the arguments. Moreover, in their setting, there is no notion of social interactions. [10] considers the effect of prior beliefs of the audience members and debaters on success in an individual debate. However, unlike this study, it does not consider the effect of social interactions on a user's success in online debate over time.

References

[1] Khalid Al Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. 2016. A News Editorial Corpus for Mining Argumentation Strategies.

In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, 3433–3443. http://aclweb.org/anthology/C16-1324

[2] Lars Backstrom, Eytan Bakshy, Jon M Kleinberg, Thomas M Lento, and Itamar Rosenn. 2011. Center of attention: How facebook users allocate attention across friends. (2011).

[3] Sara M Banks, Peter Salovey, Susan Greener, Alexander J Rothman, Anne Moyer, John Beauvais, and Elissa Epel. 1995. The effects of message framing on mammography utilization. *Health psychology* 14, 2 (1995), 178.

[4] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. 2009. Characterizing User Behavior in Online Social Networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC '09)*. ACM, New York, NY, USA, 49–62. https://doi.org/10.1145/1644893.1644900

[5] Moira Burke, Cameron Marlow, and Thomas Lento. 2009. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 945–954.

[6] Amparo Elizabeth Cano-Basave and Yulan He. 2016. A Study of the Impact of Persuasive Argumentation in Political Debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 1405–1413. https://doi.org/10.18653/v1/N16-1166

[7] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. (2010).

[8] M. Cooper and W.L. Nothstine. 1992. *Power Persuasion: Moving an Ancient Art Into the Media Age*. Educational Video Group. https://books.google.com/books?id=mkEqAQAAMAAJ

[9] Peter Ditto and David F. Lopez. 1992. Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions. *Journal of Personality and Social Psychology* 63 (10 1992), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

[10] Esin Durmus and Claire Cardie. 2018. Exploring the Role of Prior Beliefs for Argument Persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, 1035–1045. https://doi.org/10.18653/v1/N18-1094

[11] Esin Durmus and Claire Cardie. 2018. Understanding the Effect of Gender and Stance in Opinion Expression in Debates on "Abortion". In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. Association for Computational Linguistics, New Orleans, Louisiana, USA, 69–75. https://doi.org/10.18653/v1/W18-1110

[12] Adrian Edwards, Glyn Elwyn, Judith Covey, Elaine Matthews, and Rolsin Pill. 2001. Presenting risk information a review of the effects of framing and other manipulations on patient outcomes. *Journal of health communication* 6, 1 (2001), 61–82.

[13] Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 987–996.

[14] Scott A Golder, Dennis M Wilkinson, and Bernardo A Huberman. 2007. Rhythms of social interaction: Messaging within a massive online network. In *Communities and technologies 2007*. Springer, 41–66.

[15] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46, 1-3 (2002), 389–422.

[16] Ivan Habernal and Iryna Gurevych. 2016. What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation.. In *EMNLP*. 1214–1223.

[17] Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum. In *Proceedings of the 4th Workshop on Argument Mining*. Association for Computational Linguistics, Copenhagen, Denmark, 11–21. http://www.aclweb.org/anthology/W17-5102

[18] Pamela M. Homer and Sun-Gil Yoon. 1992. Message Framing and the Interrelationships among Ad-Based Feelings, Affect, and Cognition. *Journal of Advertising* 21, 1 (1992), 19–33. https://doi.org/10.1080/00913367.1992.10673357 arXiv:https://doi.org/10.1080/00913367.1992.10673357

[19] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.

[20] Shamanth Kumar, Reza Zafarani, and Huan Liu. 2011. Understanding User Migration Patterns in Social Media.

[21] Bang Hui Lim, Dongyuan Lu, Tao Chen, and Min-Yen Kan. 2015. # mytweet via instagram: Exploring user behaviour across multiple social networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 113–120.

[22] Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument Strength is in the Eye of the Beholder: Audience Effects in Persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, 742–753. http://aclweb.org/anthology/E17-1070

[23] Sofus A Macskassy and Matthew Michelson. 2011. Why do people retweet? anti-homophily wins the day!

[24] Marcelo Maia, Jussara Almeida, and Virgílio Almeida. 2008. Identifying User Behavior in Online Social Networks. In *Proceedings of the 1st Workshop on Social Network Systems (SocialNets '08)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/1435497.1435498

[25] Beth E Meyerowitz and Shelly Chaiken. 1987. The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of personality and social psychology* 52, 3 (1987), 500.

[26] Raquel Mochales and Marie-Francine Moens. 2011. Argumentation Mining. *Artif. Intell. Law* 19, 1 (March 2011), 1–22. https://doi.org/10.1007/s10506-010-9104-x

[27] Meenakshi Nagarajan, Hemant Purohit, and Amit P Sheth. 2010. A Qualitative Examination of Topical Tweet and Retweet Practices. (2010).

[28] Vlad Niculae, Joonsuk Park, and Claire Cardie. 2017. Argument Mining with Structured SVMs and RNNs. In *Proceedings of ACL*.

[29] Daniel J. O'Keefe. 1997. Standpoint Explicitness and Persuasive Effect: A Meta-Analytic Review of the Effects of Varying Conclusion Articulation in Persuasive Messages. *Argumentation and Advocacy* 34, 1 (1997), 1–12. https://doi.org/10.1080/00028533.1997.11978023 arXiv:https://doi.org/10.1080/00028533.1997.11978023

[30] Daniel J. O'Keefe. 1998. Justification Explicitness and Persuasive Effect: A Meta-Analytic Review of the Effects of Varying Support Articulation in Persuasive Messages. *Argumentation and Advocacy* 35, 2 (1998), 61–75. https://doi.org/10.1080/00028533.1998.11951621 arXiv:https://doi.org/10.1080/00028533.1998.11951621

[31] Andreas Peldszus. 2014. Towards segment-based recognition of argumentation structure in short texts. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, 88–97. https://doi.org/10.3115/v1/W14-2112

[32] Andreas Peldszus and Manfred Stede. 2016. Rhetorical structure and argumentation structure in monologue text. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*. Association for Computational Linguistics, 103–112. https://doi.org/10.18653/v1/W16-2812

[33] Peter Potash and Anna Rumshisky. 2017. Towards Debate Automation: a Recurrent Model for Predicting Debate Winners. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2455–2465.

[34] Felicia Pratto and Oliver P. John. 1991. Automatic Vigilance: The Attention-Grabbing Power of Negative Social Information. *Journal of personality and social psychology* 61 (10 1991), 380–91. https://doi.org/10.1037//0022-3514.61.3.380

[35] Daniel M Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A Huberman. 2011. Influence and passivity in social media. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 18–33.

[36] John R Rossiter and Larry Percy. 1987. *Advertising and promotion management*. McGraw-Hill Book Company.

[37] Christian Stab and Iryna Gurevych. 2014. Identifying Argumentative Discourse Structures in Persuasive Essays. In *EMNLP*.

[38] Christian Stab and Iryna Gurevych. 2017. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics* 43, 3 (2017), 619–659. https://doi.org/10.1162/COLI_a_00295

[39] Alessandra Tasso and Strada Cappuccini. 2005. Frame Effects in Persuasive Messages Against Smoking. (01 2005).

[40] Henning Wachsmuth, Martin Potthast, Khalid Al Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. 2017. Building an Argument Search Engine for the Web. In *Proceedings of the 4th Workshop on Argument Mining*. Association for Computational Linguistics, 49–59. https://doi.org/10.18653/v1/W17-5106

[41] Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin. 2017. Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes. *Transactions of the Association for Computational Linguistics* 5 (2017), 219–232. http://aclweb.org/anthology/Q17-1016

[42] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. 2009. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*. Acm, 205–218.

[43] Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational Flow in Oxford-style Debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, 136–141. https://doi.org/10.18653/v1/N16-1017