CS224W Project Report: Characterizing and Predicting Dogmatic Networks

Emily Alsentzer, Shirbi Ish-Shalom, Jonas Kemp

1. Introduction

Increasing polarization has been a defining feature of the 21st century.[1] Systematic evidence shows that elevated dogmatism, a tendency to assert opinions as truths and ignore opposing viewpoints, has increasingly polarized discourse in topics ranging from the environment, to health, politics, and guns.[2,3,4,5] Some researchers attribute the immense polarization between groups to stagnation in the pace and consistency of reform.[6] Other large bodies of research have investigated how social, economic, or psychological factors contribute to elevating dogmatism, with a primary focus on individual behavior.

However, the past decade has seen fundamental changes in the structure of social interactions with the advent of the Digital Age. Today, people can control who, how, when, and where they interact with others. At the click of a button, they can unfollow people with whom they disagree. Evidence suggests that customized social networks can exacerbate certain behaviors, as people are more likely to adopt behaviors popular among their immediate social connections.[7,8] Indirect connections can also play a role in influencing behavior through cascades; for example, Miller et al. found that sentiment becomes increasingly polarized as the depth of a cascade increases in hyperlink networks.[9] The influence of those we directly *and* indirectly interact with online can have a significant impact on our personal behavior. This influence can have (sometimes severe) real-world consequences, as seen in the recent #PizzaGate scandal where Reddit conspiracy theories about Hillary Clinton ultimately led to an armed man attacking a Washington, DC pizzeria.[10]

Given the profound role of social media in shaping people's opinions and actions, we therefore propose that dogmatism is not a phenomenon resulting from individual behavior, but rather results from the customized structure of the social network with whom a user is communicating. With the new age of *information consumption personalization,* we expect that investigating the structure of social networks will uncover information about how an individual's interactions with their social network instigate or perpetuate dogmatism. In particular, we will investigate the network characteristics of dogmatic Reddit communities with the ultimate goal of predicting the formation of dogmatic groups online.

2. Background

2.1 Predicting Positive and Negative Links in Online Social Networks[11]

Leskovec, Huttenlocher, and Kleinberg demonstrate that machine learning methods can effectively predict the sign of links in online social networks using information about local structure, such as node degree and triads. Comparing their results to the classical theories of balance and status in signed social networks, the authors find that their model better captures the subtleties of signed connectivity in large real-world networks. This result clearly demonstrates how structural features of a network can provide information about the nature of interactions between community members. In our investigation, we hypothesize that the same concept can be extended to more complex metrics of interaction than sign, including those related to dogmatism. In keeping with the idea of customized social networks, we also propose that global network properties rather than just individual-level behavior may be predictive of such metrics, particularly in the case of self-selecting online social networks such as Reddit communities.

2.2 Identifying Dogmatism in Social Media: Signals and Models[12]

Fast and Horvitz present a statistical model for binary classification of online comments to identify dogmatism in social media. Linguistic features predictive of dogmatism in the model, derived from the Linguistic Inquiry and Word Count (LIWC) lexicon, include certainty, "you" pronouns, "they" pronouns, and negative emotion. On the other hand, features predictive of non-dogmatic comments include tentativeness, relativity, "I" pronouns, and the past tense. While not explicitly a network analysis project, Fast and Horvitz trained their classifier using Reddit comment data, offering a natural avenue to connect to the study of dogmatism in online networks. Though the model has significant limitations and only achieved a test accuracy of about 80%, it presents a useful starting point for understanding how to quantitatively characterize dogmatism based on language and online textual interaction.

2.3 A Measure of Polarization on Social Media Networks Based on Community Boundaries[13]

As discussed in the introduction, polarization is highly related to dogmatism, with more dogmatic discourse tending to increase polarization between opposing groups. Guerra et al. argue that the traditional metric of modularity is not a sufficiently direct measure of polarization, and propose a new metric based on network boundary conditions between the communities. Their model's explicit consideration of boundary structure is a compelling advance over previous models, but also limits analysis to existing polarization between adjacent communities. We take a predictive rather than descriptive approach, hypothesizing that intrinsic features of a community should in fact predict a propensity for antagonism and polarization, irrespective of actual relationships to other communities.

2.4 Discussion

Our project addresses a similar question to the sign prediction problem: can we predict the nature of discourse and interactions in a network based on structural properties? However, we extend the problem in two important ways.

First, rather than edge sign we adopt dogmatism as our measure of interest, per the work of Fast and Horvitz. While this is a much more complex and challenging measure to accurately quantify, it captures a dimension of human interaction that goes beyond mere positive or negative sentiment, and one that is especially relevant in the current political climate.

Second, we choose to focus on the properties of a community as a whole rather than individual links. We aim to characterize the dogmatism of groups rather than individuals, because as the model proposed by Guerra et al. suggests, group-level interactions ultimately define polarization. Our approach can potentially offer a predictive complement to the polarization model, insofar as we hypothesize that high community-level dogmatism might by proxy indicate the likelihood of a community developing polarized relationships.

3. Methods

3.1 Data Set

In order to better understand how online interactions can influence dogmatism, we studied online communities on Reddit, an online content aggregator and discussion forum. We acquired data from over 2000 subreddit communities, courtesy of TA Will Hamilton, to understand the relationship between network properties, community dogmatism, and sentiment. Each subreddit is encapsulated in a list of all comments from 2014, where each comment links with 1) the parent comment id, 2) the author's name, and 3) the score (net upvotes versus downvotes). Furthermore, "monthly interaction networks" for four week periods are available for each subreddit during 2014, totaling 13 networks per subreddit over the course of the year. In the interaction networks, each node is a user, and nodes are connected if the users replied in the same linear thread within three comments of one another. Only users who commented at least 50 times in 2014 were included in the networks.

3.2 Sentiment and Dogmatism Statistics

To characterize the nature of discourse in each network, we ran sentiment and LIWC text analysis tools on all comments for each community in our dataset. For sentiment analysis, we ran the TextBlob[14] classifier on each comment to extract polarity, which ranges from -1 (negative sentiment) to 1 (positive sentiment), and subjectivity, which ranges from 0 (objective) to 1 (subjective). We then calculated average polarity and subjectivity for each monthly network.

For dogmatism analysis, we used a LIWC counter developed by Will Hamilton to count the frequency of LIWC words and generate normalized scores for each monthly subreddit. Although we were unable to use Fast and Horvitz's classifier to directly generate a dogmatism score for each subreddit, we constructed two proxy scores based on LIWC categories, of the form

$$d = \log \sum_{i=1}^{P} p_i - \log \sum_{j=1}^{Q} q_i \tag{1}$$

where p_i, q_i are scores for LIWC features found by Fast and Horvitz to be positively or negatively correlated with dogmatism, respectively. In the first simplified score, we chose only the single features with the highest and lowest odds ratios for dogmatism (negative emotion and "I" pronouns). In the second full score, we incorporated 12 features identified as having a significant relationship with dogmatism: certainty, "you" pronouns, "they" pronouns, present tense, negation, negative emotion (positively related); and tentativeness, insight, perception, relativity, "I" pronouns, and past tense (negatively related).

3.3 Network Statistics

For each monthly network, we extracted several features we hypothesized to be predictive of sentiment or dogmatism in the network. Features extracted included average and standard deviation of degree, average excess degree, average clustering coefficient, diameter, average degree centrality, average closeness centrality, number and average size of connected components, number and average size of communities (computed using the Clauset-Newman-Moore algorithm), average PageRank score, and proportion of triads (i.e. number of triads over the number of possible triads in a complete network). All network statistics were computed using SNAP.[15] Due to time and computational constraints, we were unable to compute all of these statistics for 25 of the largest subreddits and thus excluded them from our analysis.

We hypothesized two possible scenarios for network structure that might correlate with high levels of dogmatism. In one scenario, the entire subreddit falls under a single "groupthink" ideology, leading to one or a few large communities and connected components, and a relatively complete network with a high proportion of triads and a low diameter. In the other scenario, the subreddit contains multiple hotly contested ideologies, corresponding to several smaller communities or connected components, a lower proportion of triads (relative to a more complete network), and a higher diameter. In either case, we expect to see high clustering and likely high average PageRank scores (corresponding to important "thought leaders" driving the dogmatic discourse). If polarity and subjectivity provide a proxy for dogmatism (e.g. through similarity to LIWC scores for negativity or certainty), we might expect more negative or more objective networks to exhibit similar features to highly dogmatic networks.

3.4 Classification Models

To assess whether network structures can accurately predict sentiment or dogmatism in communities, we developed four machine learning classification models: logistic regression, random forest, gradient boosting machine, and support vector machine with Gaussian kernel. To determine which network structures most strongly indicated dogmatic communities, we conducted feature importance analysis for the best-performing model, assessed by evaluating each feature's contribution to the model's final classification decision.

Our data preprocessing included generating balanced classes for prediction, centering and scaling network statistics to mean 0 and standard deviation 1 (for ease of interpretation, particularly for logistic regression), and splitting into 70% training and 30% test sets. For subjectivity and polarity, we defined the top and bottom quartiles to be the positive and negative classes and excluded the middle quartiles, in order to ensure a balanced dataset and to help emphasize the separation between the classes. For our dogmatism metrics, we defined scores over zero to belong to the positive class, and then used oversampling to balance the classes during training. (Note that these preprocessing steps result in a different dataset for each outcome of interest.)

Each of the classifiers listed above offers a different approach to our problem. Logistic regression provides a baseline model with easily interpretable coefficients, though the linear decision boundary may not be sufficient to achieve good separation of the data. By contrast, the support vector machine (SVM) uses the Gaussian kernel to map to a high-dimensional space and thereby fit a nonlinear decision boundary in our lower-dimensional feature space. Random forests and gradient boosting machines are both ensemble methods that make use of multiple decision trees, though the boosted ensemble is explicitly constructed in a stepwise fashion such that each tree "learns" from the mistakes of earlier trees.[16] Hyperparameters for all models (for example, regularization parameters or number of boosting steps) were selected using 10-fold cross-validation. Models were developed in R using the caret package.[17]

3.5 K-means Clustering

We performed *k*-means clustering using Python's Scikit-Learn library[18] in order to assess the difficulty of distinguishing between dogmatic and non-dogmatic subreddits. We clustered subreddits, according to the 12 LIWC dogmatism features used in the full score, into $k = 2$ clusters. We analyzed the quality of the clustering by calculating a silhouette score, which measures intra-cluster distance compared to nearest-cluster distance. We then plotted histograms for each of the network features and performed either two-sample *t*-tests or Mann Whitney tests, depending on normality assumptions, to determine whether there was a significant difference in each of the 14 network statistics between the clusters. We used the Bonferroni correction to account for multiple hypothesis testing, and we calculated effect sizes using Cohen's *d* statistic.

3.6 Temporal Analysis

Finally, in an attempt to account for temporal trends in dogmatism within each subreddit, we took a repeated-measures approach, treating each network as a monthly "measurement" on its respective subreddit. We fit a linear mixed-effects model in R with random intercepts and slopes over time for each subreddit, to account for within-subreddit correlations. We modeled the full dogmatism score as our outcome variable, and predictors included time and each of the network features listed in section 3.3.

4. Results

4.1 Summary Statistics

We calculated network statistics, average comment scores, sentiment analysis, and dogmatism scores of all monthly networks for 2022 subreddits. Table 1 summarizes the mean and standard deviation of the features and the outcomes of interest over all networks.

Metric	Mean	Standard Deviation
Average Clustering Coefficient	0.30	0.0850
Proportion of Triads	0.000310	0.00508
Average Excess Degree	0.18	18.0
Diameter	6.55	1.99
Average Degree Centrality	0.00933	0.0188
Average Closeness Centrality	0.20	0.0931
Average PageRank	0.00214	0.00267
Average Connected Components	0.07	0.129
Number of Connected Components	58.3	97.7
Number of Communities	80.4	114
Average Community Size	18.5	13.7
Number of Users	1470	2726
Average Degree	4.78	2.89
Std Dev of Degree	7.84	6.40
Average Polarity	0.106	0.0406
Average Subjectivity	0.403	0.0503

Full Dogmatism	-0.210	0.0662
Simplified Dogmatism	-0.384	0.221

Table 1. Mean and standard deviation of the network properties, average sentiment, and average of two dogmatism scores across all subreddits.

The distribution of dogmatism scores (Fig. 1) presented a challenge for analysis. For both the full and simplified scores, the distribution was fairly narrow, with little separation between the top and bottom quartiles. However, when using a threshold of zero to assign networks to the positive or negative class, just 5% of networks were classified as dogmatic using the simplified score, and just 0.3% using the full score.

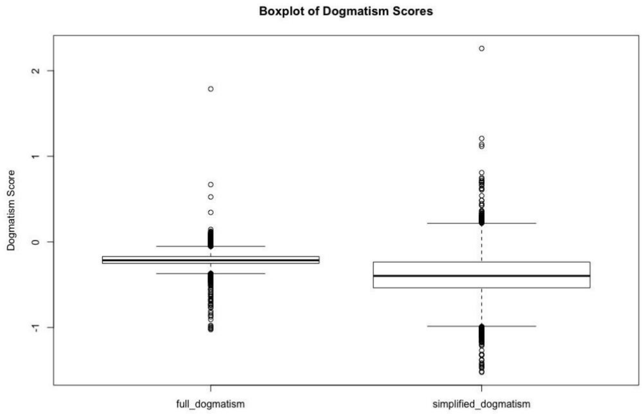

Figure 1. Boxplot of full and simplified dogmatism metrics.

The least dogmatic subreddits using the full score were Scandinavian communities: norge (Norwegian), Suomi (Finnish), and Denmark. Using the simplified score, the least dogmatic communities were related to finding fellow video game players (3dsFCswap, smiteLFM) or to Snapchat. Under both scores, the most dogmatic subreddits were for hockey (NY Rangers, LA Kings) or football (Atlanta Falcons, St. Louis Rams).

4.2 Classifier Results

Classifier predictions for polarity and subjectivity produced very poor results, with test AUC values in the range of 0.5-0.6. Initially, we assigned dogmatism class labels in the same fashion as for sentiment (1 if in the top quartile, 0 if in the bottom quartile) but achieved similarly poor outcomes. As our research question primarily relates to predicting dogmatism, with sentiment merely a proxy, we chose to focus our efforts on improving the dogmatism classifiers. For each of the four classifiers, we refit with upsampling using the classification scheme described in section 3.4 (1 if dogmatism score greater than zero, 0 otherwise) and assessed sensitivity, specificity, and AUC on both the training and test sets (Tables 2 and 3). We did not report accuracy due to the class imbalance, as a classifier that simply assigns the majority class trivially achieves high accuracy. Sensitivity and specificity were assessed at the typical 0.5 decision threshold for the training set, and at the threshold closest to the upper-left corner of the ROC curve for the test set. The gradient boosting machine model performed best in predicting both the full dogmatism and simplified dogmatism metrics, with AUC values of 0.72 and 0.76 respectively.

TRAIN	Sensitivity	Specificity	AUC	**TEST**	Sensitivity	Specificity	AUC
Logistic Regression	0.70	0.81	0.84		0.55	0.64	0.62

Gradient Boosting Machine	0.68	0.80	0.83		0.85	0.49	0.72
Random Forest	0.14	1.00	0.78		0.60	0.76	0.66
SVM	0.53	0.92	0.79		0	1	0.50

Table 2. Train and test sensitivity, specificity, and AUC for models predicting the full dogmatism metric.

TRAIN	Sensitivity	Specificity	AUC	**TEST**	Sensitivity	Specificity	AUC
Logistic Regression	0.78	0.73	0.82		0.75	0.37	0.57
Gradient Boosting	0.78	0.77	0.85		0.78	0.63	0.76
Random Forest	0.87	0.16	0.99		0.75	0.66	0.76
SVM	0.87	0.80	0.80		0	1	0.5

Table 3. Train and test sensitivity, specificity, and AUC for models predicting the simplified dogmatism metric.

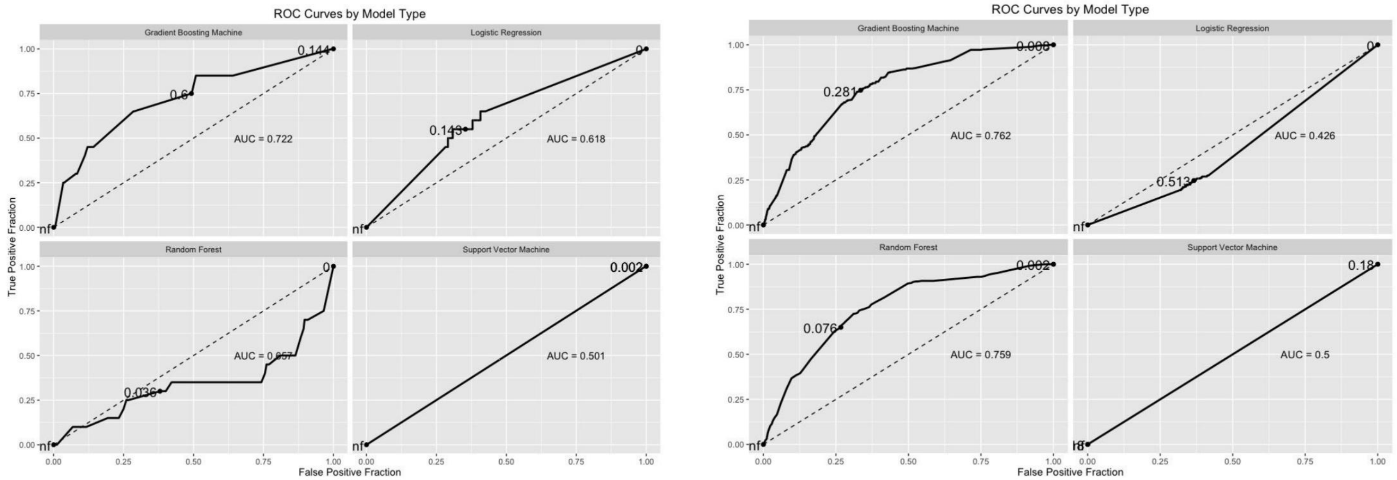

Figure 2. Test set ROC curves for models predicting the full dogmatism score (left) and simplified dogmatism score (right).

We analyzed the contributions of the network features to prediction of the full and simplified dogmatism scores in the gradient boosting model (Fig. 3). The most important network feature for the full score was diameter, followed by the number of communities, degree centrality, and clustering coefficient; the most important feature for the simplified score was clustering coefficient, followed by closeness centrality, average degree, and average community size.

4.3 K-means Clustering

K-means clustering of the subreddits according to the 12 dogmatism features yielded a silhouette score of 0.24 when clustering into $k = 2$ clusters. We performed a Mann-Whitney test to compare the number of connected components between the two clusters (due to violation of the normality assumption), and used *t*-tests to compare all other network features due to the approximate normality of the data and the large sample size. We found a significant difference between the clusters on every single network feature (p-values << 0.05), and two of the network features, average clustering coefficient and average closeness centrality, produced medium effect sizes with Cohen's *d* values of 0.74 and 0.54 respectively. All remaining network features were associated with small to negligent effect sizes.

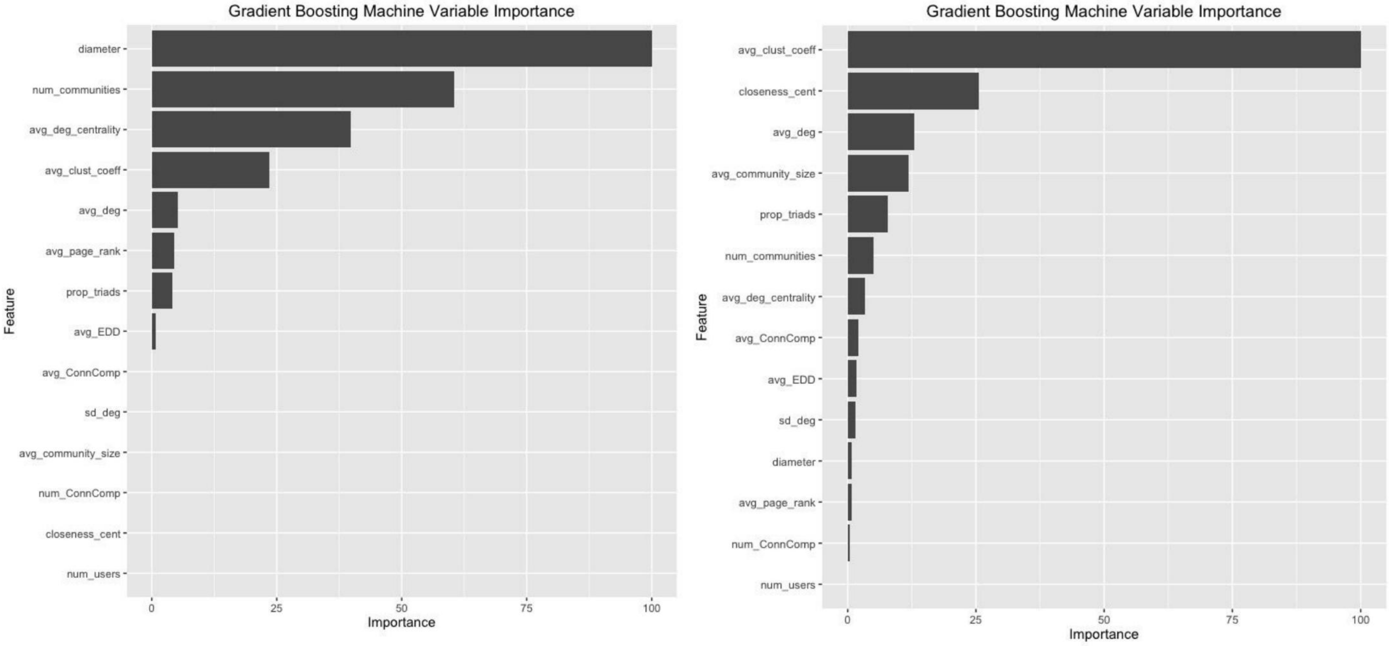

Figure 3. Ranking of features by importance in the prediction of full dogmatism scores (left) and simplified dogmatism scores (right) in the gradient boosting model.

4.4 Temporal Analysis

Fitted coefficients in the mixed model (Table 4) suggest a significant but very small increase in average dogmatism over time in all networks. After accounting for temporal correlations within subreddits, we found five network features to be significantly related to the full dogmatism score. However, in each case the effect of the relevant network feature on dogmatism was quite small.

	Month	Avg. Degree	Avg. Clustering Coefficient	Proportion of Triads	Avg. Degree Centrality	Avg. PageRank
Coeff.	0.000348	0.0129	0.00818	0.00497	-0.0168	0.00721
Std. Error	0.000113	0.00402	0.00110	0.00172	0.00381	0.00179
p-value	0.0021	0.0014	<0.0001	0.0039	<0.0001	0.0001

Table 4. Coefficients for significant network features (p < 0.05) in the mixed-effects model.

5. Conclusions

Our machine learning approach achieved modest AUC values of 0.72 and 0.76 in predicting the full and simplified dogmatism metrics, respectively, using gradient boosting. The moderate performance of the more powerful ensemble used in the gradient boosting algorithm, coupled with the few number of important network features (Figure 3) and the poor performance of the other three models (Figure 2), suggests that prediction of dogmatism using network features alone is a challenging problem.

Both our *k*-means clustering and mixed model analysis provide further evidence for the difficulty of this problem. The low silhouette score for the *k*-means clusters suggests that there is not a clear separation between a high-dogmatism and low-dogmatism cluster, at least based on LIWC features. Although there was a significant difference in all of the network features between the two clusters, the small effect size of 12 out of the 14 network features further suggests that network properties alone are insufficient in differentiating dogmatic from non-dogmatic networks. Similarly, in the mixed model, we found that both time and network features relate to only small changes in dogmatism score. These multiple lines of evidence support the idea that our chosen network features cannot by themselves explain the observed variation in dogmatism between networks, and suggest our models' observed overfitting on these relatively few features during training may result from fitting a large amount of noise in addition to the small variation actually explained by the network features. Fitting noise in training could explain the smaller than 0.5 AUC values for logistic regression and random forest in the prediction of simplified dogmatism and full dogmatism respectively.

Devising an appropriate dogmatism score presented a major challenge in our analysis in and of itself. In particular, with the full score we attempted to capture variability in 12 different LIWC features within a single number; yet much of the 12-dimensional variance in these scores was lost when collapsed into one value (and indeed, our *k*-means results suggest that variation in these 12 dimensions was low to begin with). The simplified score gives a somewhat wider distribution, but it is not a truly satisfying metric given that it fails to account for so many of the dimensions of dogmatism identified by Fast and Horvitz. Thus, when gradient boosting finds different network features to be more predictive of the simplified rather than the full score, it is challenging to to interpret which of these inconsistent results truly captures the essence of the phenomenon. Notably, however, nearly every approach we took found average clustering coefficient to relate significantly to dogmatism, an intuitive result that lends at least some face validity to our analysis.

Our analyses have demonstrated the possibility of using network features in the prediction of dogmatism, though further work is clearly needed to develop a more rigorous dogmatism metric. Fast and Horvitz achieved AUC scores of 0.80 in predicting dogmatism using linguistic features; we achieve similar performance with only network features, suggesting that combining both better NLP and network approaches may offer an even better classifier for this complex social problem. Social media platforms are becoming increasingly important in the creation and polarization of individuals' ideologies, as demonstrated by the polemics and fake news circulated during the 2016 presidential election. Reddit communities contributed to the acceleration of the #PizzaGate conspiracy theory, which ultimately resulted in a man shooting off a rifle in a pizzeria filled with children. Further work to elucidate the development and perpetuation of dogmatism online is needed to understand how such extreme ideologies can form in the age of social media.

References

1. Doherty, Carroll. "7 things to know about polarization in America." *Pew Research Center* (2014).

2. Jacobson, Gary C. "Partisan polarization in American politics: A background paper." *Presidential Studies Quarterly* 43.4 (2013): 688-708.

3. Guber, Deborah Lynn. "A cooling climate for change? Party polarization and the politics of global warming." *American Behavioral Scientist* (2012): 0002764212463361.

4. Baker, Jeffrey P. "Mercury, vaccines, and autism: one controversy, three histories." *American Journal of Public Health* 98.2 (2008): 244-253.

5. Wozniak, Kevin H. "American public opinion about gun control remained polarized and politicized in the wake of the Sandy Hook mass shooting."*USApp–American Politics and Policy Blog* (2015).

6. Frye, Timothy. *Building states and markets after communism: the perils of polarized democracy*. Cambridge University Press, 2010.

7. Young, H. Peyton. "The diffusion of innovations in social networks." *The economy as an evolving complex system III: Current perspectives and future directions* 267 (2006).

8. Centola, Damon. "The spread of behavior in an online social network experiment." *Science* 329.5996 (2010): 1194-1197.

9. Miller, Mahalia, et al. "Sentiment Flow Through Hyperlink Networks." *ICWSM*. 2011.

10. Aisch, Gregor, Jon Huang, and Cecilia Kang. "Dissecting the #PizzaGate Conspiracy Theories." *New York Times* 10 Dec. 2016: n. pag. Print.

11. Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg. "Predicting positive and negative links in online social networks." *Proceedings of the 19th International Conference on the World Wide Web*. ACM, 2010.

12. Fast, Ethan, and Eric Horvitz. "Identifying Dogmatism in Social Media: Signals and Models." *arXiv preprint arXiv:1609.00425* (2016).

13. Guerra, Pedro Henrique Calais, et al. "A Measure of Polarization on Social Media Networks Based on Community Boundaries." *ICWSM*. 2013.

14. Loria, Steven. "Textblob: Simplified text processing." Python release v0.11.1 (2016).

15. Leskovec, Jure and Sosic, Rok. "SNAP: A General-Purpose Network Analysis and Graph-Mining Library." *ACM Transactions on Intelligent Systems and Technology (TIST),* **8** (2016).

16. James, Gareth, et al. *An introduction to statistical learning*. Vol. 6. New York: Springer, 2013.

17. Kuhn, Max, et al. caret: Classification and Regression Training. R package version 6.0-73 (2016).

18. Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, **12** (2011): 2825-2830.