Debiasing Community Detection: The Importance of Lowly Connected Nodes

Ninareh Mehrabi, Fred Morstatter, Nanyun Peng, and Aram Galstyan

Information Sciences Institute

University of Southern California

{ninarehm, fredmors, npeng, galstyan}@isi.edu

Abstract—

Community detection is an important task in social network analysis, allowing us to identify and understand the communities within the social structures provided by the network. However, many community detection approaches either fail to assign low-degree (or lowly connected) users to communities, or assign them to trivially small communities that prevent them from being included in analysis. In this work we investigate how excluding these users can bias analysis results. We then introduce an approach that is more inclusive for lowly connected users by incorporating them into larger groups. Experiments show that our approach outperforms the existing state-of-the-art in terms of F1 and Jaccard similarity scores while reducing the bias towards low-degree users.

I. INTRODUCTION

Community detection is a fundamental task in social network analysis [1] that identifies subgroups within social networks. Existing community detection approaches suffer from a major flaw—the inability to assign lowly connected users to communities. Despite the fact that lowly connected users are definitionally not well-integrated into the social network, the information they provide can be crucial for a better understanding of the motivations and beliefs of the community, especially considering that there could be a long-tail of lowly connected users. Failure to incorporate the lowly connected users may result in biased results. For instance, studying groups within a social network can be biased towards users whose tweets get a large number of retweets. This will lead to a biased analysis of the data in which not all users' voices are heard.

We demonstrate the existence of this bias by showing what is omitted by existing community detection approaches. We use two state-of-the-art approaches, CESNA [2], and the Louvain method [3]. Louvain uses only the network when assigning communities, while CESNA uses both the network

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '19, August 27-30, 2019, Vancouver, BC, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6868-1/19/08...\$15.00

<http://dx.doi.org/10.1145/3341161.3342915>

TABLE I

COUNTS AND EXAMPLES OF INFORMATION MISSED BY TRADITIONAL COMMUNITY DETECTION APPROACHES. THESE STATISTICS ARE TAKEN BY COMPARING THE GROUND TRUTH LABELS WITH THE LABELS ASSIGNED BY THE COMMUNITY DETECTION ALGORITHM.

Method	Gamergate Dataset		U.S. Election Dataset	
	Discarded Unique Hashtags	Examples	Discarded Unique Hashtags	Examples
CESNA	597 (38.6%)	#gamerignorance, #ethicsinjournalism, #feministgate, #sexismiswickedcool	1518 (66.9%)	#refugeeswelcome, #Guns, #OscarsStillSoWhite, #TrumpDumped
Modularity	625 (40.4%)	#misandricfeminists, #violenceagainstwomen, #gamersagainstgamergate, #stopgamejournalism2014	302 (13.3%)	#Christians4Hillary, #Killary, #FakeHateCrimes, #TrumpRiots

and user attributes. We apply these methods to two separate datasets, one based on Gamergate and one based on the 2016 U.S. Presidential Election. Both of these datasets contain two major contingents and ground truth information. Table I shows the information that is omitted by putting the lowly connected users into insignificant communities. The “#refugeeswelcome” hashtag is excluded from analysis if we use CESNA for community detection since the user who tweeted this hashtag was put in a small insignificant community. By not including this user tweeting this hashtag into a significant community, we lose this piece of information when we analyze the significant communities in the U.S. Election dataset. In other words, degree is not correlated with relevancy to the topic, and therefore degree and connections in social networks should not be the only indicative factors for community membership. Other examples are demonstrated in Table I, such as the “#Christians4Hillary,” “#gamersagainstgamergate,” and many other hashtags as reported in the table.

The goal of this work is twofold. First, we demonstrate empirically the existence of biases in existing community detection approaches using real world datasets. Through this analysis, we show that current state-of-the-art community detection methods ignore lowly connected users that have few links—by putting them into insignificant small or singleton communities which are then ignored in the study. We define **insignificant communities** throughout this paper as follows:

$$C_{insig} = \{c_i \mid \text{where } |c_i| < \theta\} \quad (1)$$

Where $|c_i|$ denotes the size of the community c_i , and θ is a

threshold value that can be treated as a hyperparameter and will be used in our proposed algorithm.

We call node n an **insignificantly labeled node** if:

$\exists C_{insig}$ such that $n \in C_{insig}$ where C_{insig} is an insignificant community.

These insignificantly labeled users have value and should be included in the communities, as they offer a more diverse, nuanced representation of those communities. Second, to overcome this issue, we introduce a new community detection method—Communities with Lowly connected Attributed Nodes (CLAN)—that will mitigate the existence of this bias toward low-degree nodes which are more likely to be insignificantly labeled. We call method M **bias towards node** n if n is determined to be an insignificantly labeled node by M .

Our contributions are as follows:

- 1) We introduce CLAN, a community detection approach that categorizes low-degree users into communities.
- 2) We show that CLAN is able to outperform the existing state-of-the-art community detection methods in terms of predictive accuracy, while still being able to classify more users.
- 3) We demonstrate the existence of bias in community detection approaches that is introduced from ignoring low-degree users. We show that CLAN is able to overcome this challenge by classifying these low-degree users.

II. RELATED WORK

A. Non-Attributed Methods

One of the widely used community detection methods is the Louvain method, which utilizes the modularity value to obtain partitions by optimizing an objective function discussed in [3]. However, this method tends to create many small and singleton communities and assign low-degree users into them—which can be biased towards low-degree people who are relevant by context to the significant groups, but not connected through the network structure. Other examples are BigCLAM [4] and DEMON [5]. More information can be found in [6] which is a user guide for community detection and its methods. Using node attributes in addition to the network structure will add additional information, as discussed in [7], and can be helpful in reducing these types of biases in social networks. There are methods that use node attributes in combination with the network structure to create communities.

B. Attributed Methods

These are methods that utilize node attributes in addition to the network structure. Some attributed community detection methods are discussed in [8]. From this, one of the widely known methods that uses node attributes in addition to the network structure is a generative model called CESNA [2]. This method restructures the network by incorporating node attributes. Experiments show that this method performs well by having a high precision value; however, it suffers from having a significantly low recall value. Low recall value may be indicative that this method also discards some of the

TABLE II
STATISTICS OF DATASETS DISCUSSED ALONG WITH THE NUMBER OF TOTAL USERS AND NUMBER OF USERS WITH GROUND TRUTH LABELS.

Datasets	Total Users	Labeled Users
Gamergate	21,441	7,320
U.S. Presidential Election	1,089,974	10,074

important users—which we are trying to avoid in order to minimize bias. In addition to CESNA, other methods have been proposed, such as Block-LDA [9], PAICAN [10], shared latent space models such as CLSM [7], and embedding-based approaches like LANE [11], which utilize node attributes in order to detect communities, or ELAINE [12], which utilizes edge attributes. However, none of these methods try to target the existing bias in social networks.

III. COMMUNITY DETECTION DATASETS

Studying community assignments requires datasets where the underlying communities are explicitly labeled. To satisfy this, we use two different datasets—Gamergate and U.S. Presidential Election.

A. Gamergate

The Gamergate dataset consists of tweets posted in 2014 between the months of August through October. The tweets surround the Gamergate controversy [13]. It contains 21,441 users who collectively produced 104,914 tweets. These users fall into one of the two groups surrounding the controversy. One group consists of Gamergate supporters who are tweeting about ethics in journalism and believe that regardless of the relationship between journalists and game developers, journalists should give honest reviews to game developers. The other group, Gamergate opposers, argues that Gamergate supporters attack female game developers and also feminist critics, and that they are not concerned with ethics in journalism, but are using the opportunity to attack women in the gaming industry. We collected ground truth labels indicating the stance (belonging group) of each user for this dataset by performing a Mechanical Turk experiment from which we obtained 7,320 labeled users.

B. U.S. Presidential Election

This dataset contains 10,074 users who discuss the U.S. presidential election of 2016. This dataset consists of two major groups which indicate the political party of each user. This dataset comes from [14]. We only utilized the seed users from the whole dataset.

IV. PROPOSED METHOD

We have confirmed that ignoring low-degree users introduces bias into the resulting analysis of the data. Placement of low-degree people into insignificant groups may create the illusion that these small communities are communities that are not particularly relevant to the major groups that need to be

studied, and therefore can be ignored. We observed that these low-degree nodes had relevant and informative content to the study, based on our ground truth labels, that it is crucial to keep them in the significant communities. Low-connectivity in social networks may be a result of several factors. Consider the retweet network as an example: there may be different reasons for low-connection, such as users having unique ways of expressing ideas that are yet relevant to the majority group's opinion, the tweet content is long which tends to repel other users from reading and retweeting, or in case of introverted users with small friendship circles, tweets may not get enough attention or retweets. Network sampling can also change network measures and create unrealistic low-degree users [15], [16]. These reasons are not indicative of irrelevancy of topic and should not cause a bias toward low-degree users of such types and their exclusion from significant communities. In light of this, we propose a community detection approach that addresses the following issues:

- 1) Bias against low-degree nodes.
- 2) Bias from low recall: Having low recall means exclusion of users; this is not desirable since we want to be able to maintain all the users for a comprehensive study.

$$R = \frac{1}{n} \sum_{i=1}^{i=n} \frac{|x_i \cap y_i|}{|y_i|} \quad (2)$$

We define recall as shown in Equation 2, where n denotes the total number of communities found by the method, x_i denotes the labels for community i found by the community detection method, and y_i the ground truth labels for community i .

- 3) Predictive accuracy of the model.

We introduce CLAN, Communities from Lowly connected Attributed Nodes, which addresses these issues. CLAN uses node attributes, such as text, alongside the network to classify the insignificantly labeled users into their correct belonging significant communities. By not changing the network structure and only adding information to this structure through the additional node attributes, such as text utilization, results obtained from CLAN would tend to have higher recall values. Through experimentation will show in our results that CLAN has a superior performance and is in a higher agreement with the ground truth communities. Algorithm 1 provides a formalization of our approach which shows how we achieve the above solutions.

The CLAN algorithm is a two-step process in which we first use unsupervised learning to develop communities using network attributes and modularity value. Once we have the communities, we will then turn the problem into a supervised classification problem where we will classify the insignificantly labeled users from insignificant communities into the major communities using additional node attributes that were held out in the first step—such as text attributes associated to a user. The first step of this straightforward process is where we utilized the Louvain method discussed in [3]. Note that any community detection task that only utilizes the network, such as BigCLAM and DEMON [4], [5], can replace this

step, which makes our algorithm flexible, in general, and also capable of handling overlapping communities.

Algorithm 1 CLAN

Input:

1: Network

2: Threshold

Output: Communities

// Step 1: Finding the communities

 $C = \text{find_communities}(\text{Network})$

// Step 2: Classifying minority users into significant communities

for C_i in C **do** **if** $\text{count}(C_i) > \text{Threshold}$ **then**

//These are the significant communities

 Add C_i to training set **else**

//These are the insignificant communities

 Add C_i to test set **end if****end for**

model = train(training set)

predictions = model.inference(test set)

for p in predictions **do** **if** p.label == C_i **then** Add p.data to C_i **end if****end for**

return C

Once this method is applied and the communities are obtained, we train a classifier on the majority communities using node attributes as features. The introvert users will then be classified into the majority groups that were used in our training process. The features can be any held out node features, such as text or hashtags, that each user used in his tweets.

V. COMMUNITY DETECTION RESULTS

In this section we will iterate over the results obtained from CLAN compared to two widely used community detection methods. Thus, we report two numerical values one of which shows how our method is less biased towards having insignificantly labeled nodes and the other showing our method's ability to maintain the high performance in terms of detecting the communities compared to the state of the art methods.

A. Quantitative Results

Through our quantitative experiments, we will first report the scores for the F1 and Jaccard similarity scores between the ground truth labels and three different methods. Table III contains results for the F1 and Jaccard similarity scores obtained from comparisons made between the ground truth labels and labels obtained by applying each of the methods on the two datasets.

TABLE III

THE QUANTITATIVE RESULTS OBTAINED FROM CALCULATING THE F1 AND JACCARD SIMILARITY SCORES WITH REGARDS TO THE GROUND TRUTH LABELS FOR EACH OF THE METHODS ALONG WITH THE NUMBER OF SIGNIFICANTLY LABELED USERS.

Method	Gamergate			U.S. Election		
	Significantly Labeled	F1 Score	Jaccard	Significantly Labeled	F1 Score	Jaccard
CESNA	0.31	0.343	0.211	0.05	0.253	0.149
Modularity	0.79	0.434	0.282	0.80	0.753	0.604
CLAN	1.00	0.478	0.318	1.00	0.787	0.649

In addition to reporting the F1 and Jaccard score results from the ground truth comparisons, we conducted another experiment in which we report the results indicating the number of significantly labeled users by each of the methods. These show the number of users that the method has appropriately labeled into their significant communities preventing from being ignored in insignificant communities. The more significantly labeled users a method has, the less susceptible it is to bias based on our definition of bias. In Table III, we reported the percentage of significantly labeled users in each of the methods from the two datasets. These results confirm that our method, CLAN, has mitigated the bias towards these users by incorporating them into the significant communities.

The results reported in Table III confirm that not only is our method able to achieve superior predictive accuracy, but it also mitigates bias against insignificantly labeled users by assigning them labels that prevent them from exclusion.

VI. CONCLUSIONS AND FUTURE WORK

We introduce CLAN, a new community detection method that mitigates the bias in existing detection methods that fail to properly account for sparsely connected nodes in social networks. CLAN minimizes such biases by including the lowly connected nodes in their true communities. Our empirical results demonstrate that inclusion of those users enables CLAN to achieve overall superior performance in terms of F1-score and Jaccard similarity. We reported these results by providing evidence through our qualitative and quantitative experiments. Through qualitative analysis, we are able to show that these lowly connected users, in aggregate, offer information that can be of use for analysis of social network data.

We see two promising directions for future work. First, we would extend this method to introduce a new hierarchical community detection method capable of detecting subgroups within larger communities. Second, we would like to consider edge attributes simultaneously with node attributes to enable our method to consider different types of connections.

VII. ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR0011890019. We thank Greg Ver Steeg and Palash Goyal for helpful insights and discussions. The full version of this paper can be found here: <https://arxiv.org/abs/1903.08136>.

REFERENCES

- [1] L. Tang and H. Liu, "Community detection and mining in social media," *Synthesis lectures on data mining and knowledge discovery*, vol. 2, no. 1, pp. 1–137, 2010.
- [2] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *2013 IEEE 13th International Conference on Data Mining*, Dec 2013, pp. 1151–1156.
- [3] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [4] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 587–596.
- [5] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Demon: a local-first discovery method for overlapping communities," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 615–623.
- [6] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics reports*, vol. 659, pp. 1–44, 2016.
- [7] Y.-S. Cho, G. Ver Steeg, E. Ferrara, and A. Galstyan, "Latent space model for multi-modal social data," in *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 447–458.
- [8] I. Falihi, N. Grozavu, R. Kanawati, and Y. Bennani, "Community detection in attributed network," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, pp. 1299–1306. [Online]. Available: <https://doi.org/10.1145/3184558.3191570>
- [9] R. Balasubramanyam and W. W. Cohen, "Block-lda: Jointly modeling entity-annotated text and entity-entity links," in *SIAM International Conference on Data Mining*. SIAM, 2011, pp. 450–461.
- [10] A. Bojchevski and S. Günnemann, "Bayesian robust attributed graph clustering: Joint learning of partial anomalies and group structure," in *AAAI*, 2018.
- [11] X. Huang, J. Li, and X. Hu, "Label informed attributed network embedding," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017, pp. 731–739.
- [12] P. Goyal, H. Hosseinmardi, E. Ferrara, and A. Galstyan, "Embedding networks with edge attributes," in *Proceedings of the 29th on Hypertext and Social Media*, ser. HT '18. New York, NY, USA: ACM, 2018, pp. 38–42. [Online]. Available: <http://doi.acm.org/10.1145/3209542.3209571>
- [13] T. E. Mortensen, "Anger, fear, and games: The long event of# gamergate," *Games and Culture*, vol. 13, no. 8, pp. 787–806, 2018.
- [14] A. Badawy, A. Addawood, K. Lerman, and E. Ferrara, "Characterizing the 2016 russian ira influence campaign," *arXiv preprint arXiv:1812.01997*, 2018.
- [15] S. González-Bailón, N. Wang, A. Rivero, J. Borge-Holthoefer, and Y. Moreno, "Assessing the bias in samples of large online networks," *Social Networks*, vol. 38, pp. 16–27, 2014.
- [16] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose," in *7th International AAAI Conference on Weblogs and Social Media, ICWSM 2013*. AAAI press, 2013.