2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)

How to Optimize Social Network Influence

Felix Zhan
USAOT
felixzhan@usaot.or

Abstract— Online social networks are rapidly becoming an identifying feature of modern society. As online communities continue to grow, business networks of different domains also grow. As online social network begins to become increasingly integrated in our daily lives, the optimization of social network influence becomes increasingly important. In this paper, I propose to optimize influence using a modified linear threshold model. Under this model, I will focus on strategically selecting a set of users that will optimize our influence within the network. I call this type of problem as influence optimization. In this paper, I will propose a model to optimize social network influence.

Keywords—Influence Optimization, Community Detection, Big Data

I. INTRODUCTION

With the rapid expansion of online social networks, online social networks have become a critical medium of communication and interactions between individuals. Online social networks have provided ways to communicate strangers and friends alike. With online social networks, communication is faster than ever, and the scope of communication is much more diverse. Due to the implications of social networks, it has become necessary for companies and researchers to study and gain understanding of them. In particular, it has become very beneficial for researchers to study the diffusion of information within an online social network. While online social networks retain many properties of a regular networks such as local influence, online social networks have allowed to expand influence beyond that of the local network. Ultimately, online social networks have expanded individual scopes of influence.

An important topic for online social networks is influence maximization. The objective of influence maximization is to maximize the number of individuals involved with a single idea within a network, which can range from news, gossip, and products, to recommendations and etc. This problem within networks has various applications such as marketing, advertising, and even the medical field. For example, a company might want to promote a new product and needs to select a set of users that can maximize the exposure of the advertised product. By selecting an optimal set of users, a company can maximize its profits and minimize its losses. Some popular influence diffusion models include the independent cascade (IC) and the linear threshold (LT). Both models are used to characterize how influence propagates throughout the network starting from the initial seed nodes. In this paper, I will use a propagation model that will have characteristics of both.

Independent cascade and linear threshold usually only take in consideration of neighboring nodes, but I will also introduce a credibility factor in my model that will modify the influence of the neighbors. My goal is to maximize the total influence of a target product by selecting a set of users to distribute retail samples. The users can be influenced by neighboring nodes, rated by their neighbors. The rating will serve as the credibility factor. The higher the credibility of the node, the more influence it will have on the rest of its neighbors. In order to select users, there must be guidelines for specific types of users. To solve this, I will use an algorithm to find all communities within the network and sort them by size. An important difference is that I divide the graph into communities rather than sub-graphs. This distinction is important is because communities are well connected; consequently, the spread of influence will increase. After establishing the communities that I want, I will measure the closeness centrality (connectivity) of each node within the selected communities, taking the node with the highest rating from each chosen community to form a set of seed users.

II. RELATED WORK

A. Influence Maximization in Networks

There has been a lot of research done specifically on the topic of influence maximization in networks. Chen et al. first introduced the idea of using a linear threshold model [6]. Later on, Chen et al. also discussed the scalability of the model [7]. Along the same lines, Pathak et al. proposed a generalized linear threshold model for multiple cascades. Goyal et al. took an alternate route and used the greedy algorithm for influence maximization [11]. Heidari et al. built on that and proposed a scalable version of the algorithm [12]. Similarly, Selim and Zhan created an algorithm to identify the shortest paths on large networks, especially those of social media. On a different note, Kempe et al. discussed the impact of influence maximization in social networks specifically [13][33]. Kostka et al. compared the spread of influence on social media to rumors [14].

B. Other Research on Networks

As I previously explained, networks are an important and highly influential topic to study; therefore, there has been a great quantity of research done on improvements of them. Ahn and Zhan used proxies for node immunization identification on large graphs [1][24][28]. Blosser and Zhan proposed a collaborative social network preserving privacy [2]. Chiu et al. improved network security by detecting suspicious activity from partially paired data [3]. Chopade et al. used a variety of techniques for community detection [7][8][9]. Building off of Chopade et al., Pirouz et al. added ranking to community detection methods [16][17]. Xu et al. used community detection and social networks to organize researcher communities [20].

978-1-7281-1488-0/19/$31.00 ©2019 IEEE
DOI 10.1109/AIKE.2019.00029

Zhan et al. researched the prediction of social network community behaviors and friends [21][22][30][31]. On the other hand, Zhan et al. researched social network computing, especially of trust [23][25][29][32].

C. Applications

There are various applications of networks in a multitude of fields including [4-5,10,15,18-19,26-27-28,34-49].

III. Proposed Appraoch

A. Problem Definition

Given a social network, we apply a community detection algorithm to find community set. We then sort the community set from the highest to the lowest according to the cardinality of the set. We will then select k highest cardinality community. From those communities, we will find the closeness centrality of each node within the community and return the node with the highest closeness centrality. The k returned nodes will be used to propagate the social network by a modified mixed model of independent cascade and linear threshold. The problem is how to find a set of seed nodes to optimize influence [1-14].

B. Propagation Model

The propagation model describes how influence spreads throughout the network. The propagation model is a combination of the Linear Threshold Model (LT) and the Independent Cascade Model(IC). The LT model describes that the users' opinions are mostly influenced by friends, meaning that a user is more likely to be influenced by certain idea when peers and family share a similar idea. In our model, the node's threshold will sample from a uniform distribution and the sum of the neighbors' influence adds up to one.

$$\theta \sim Uniform\ (0,1)$$

Another factor I add is the probability of acceptance, given that the threshold is met. The purpose of using probability is to model a possibility of rejecting an idea even when others have accepted it. In addition, I introduce another component: a credibility factor. Credibility models the validity of the individual's opinion. Depending on the credibility of the user, its influence on others will vary, and consequently, so will the spread of ideas. In our model, users with high credibility can maximize their own influences on others, while users with low credibility do not have that potential. Credibility is bound the weight of the influence. Credibility suggests that a user's idea is more valid if there are others that share a similar idea [15-25].

The final component in the propagation model is the state of change. Every node in a network has a threshold, and once the sum of its neighbor influence meets this threshold, the node can accept reject the idea. In a standard linear threshold model, when the threshold is met, the node changes its state. Instead of basing the probability off of edge weight, I use the sum of the influence being propagated towards the node.

For a given social network $G(V, \varepsilon, W)$ and given V_j element of V I define the opinion state of V_j to be

$State(V_j) = \sum_{V_k \in N(V_j)} W_{kj}(C_k)$, where W_{kj} is the influence between two users and C_k is the credibility factor of user k. The credibility factor is modeled with a linear equation where α the percentage that the neighbors make the user credible is and β is the base credibility of the user. β is sampled from a beta distribution where α is significantly larger than β. This creates a left-skewed distribution for calculating β, showing that base credibility is more important than reputation.

$$C_j = \alpha \left(\sum_{V_k \in N(V_j)} State(V_k) \right) + \beta$$

α is calculated by subtracting β from 1. Since β is calculated from a left skewed distribution, α must be small in order to satisfy the condition that the sum of α and β is less than or equal to one.

B. Community Detection

In a network structure, a network is said to have some form of community structure if there is a group of nodes that are significantly more connected than they are to the rest of the graph. Communities are important because network expansion often reveals scalability limitations. Scalability is often an issue with problems such as graphs and networks that can grow exponentially. Introducing the idea of communities not only helps us with scalability but also in calculating influence within a network. Individuals will frequently fall within certain communities and consequently have more interaction with people inside those communities. Therefore, it is clear that selections of individuals should be made through communities.

Recognizing the relevance of a community, I will use a community detection algorithm to find all communities within the network. By using a community detection algorithm, I can divide the network into smaller subgraphs where it will be easier to identify influential users in each community. By dividing the network into different communities, it will be an important task for us to recognize the communities that will have most impact on the network. To determine the aforementioned communities, I will look into the number of users within each community. The community with the largest number of users will most likely have the most impact on the network. That will be one of my underlying conditions within our method. The choice of community detection algorithm is not important as long as one can find the communities within the network. The community detection algorithm I have decided to use in this paper is the Girvan-Newman Algorithm.

1) Girvan-Newman Algorithm: The Girvan-Newman Algorithm is a community detection algorithm that finds communities by constantly removing edges from the original network. This algorithm progressively removes edges from the network until only the connected components remain. The remaining components will be the communities of the original complex network. The concept behind Girvan-Newman Algorithm is the existence of inter-community edges. Inter-community edges are edges that link nodes from one community to nodes from another community. This indicates that an intercommunity edge is usually not part of a community, but rather an edge that keeps the network connected. By detecting these inter-community edges and removing them repeatedly, the result should provide a network of true communities. In order to find these inter-community edges, I need look into the concept of edge closeness. Using edge

closeness, I can detect inter-community edges, which in turn allows us to use the Girvan-Newman Algorithm [7-9].

C. Closeness Centrality

Closeness in a network is the measure of average distance between nodes. In this method, I use the idea of closeness as the main form of selection for the set of seed nodes. In a connected graph, closeness centrality is the sum of the length of the shortest paths between the nodes and all other nodes in the graph. The relevance of this calculation and selection is due to the scope of closeness centrality, which is the entire graph. By selecting a node with a high closeness centrality score, I can ensure on average that there is a chance of spreading influence to a large portion of the network.

This idea is important because if one is in a large community, but they are not close to majority of the group, their influence as a whole is greatly reduced. By having relatively close ties to the major, users with a higher closeness centrality score are more likely to spread influence than their counterparts [30-31].

D. Greedy Approximation Algorithm

As networks continue to grow, computation and finding the set of seed individuals to maximize our influence on the graph becomes a difficult problem. In order to solve it, I propose a greedy approximation algorithm that will provide a solution with a set of seed individuals that will theoretically maximize influence. The algorithm will start by using the Girvan-Newman Algorithm to find all sets of communities within the given network [2-14].

From the given sets of communities, it will sort the communities from the highest number of users to the lowest. From the largest community, I will create a sub-graph, and approximate the closeness centrality inside the sub-network. This calculation will return the node that has the highest closeness, becoming one of the seed individuals. I then input the seed set into the propagation model, which will diffuse the influence of the seed nodes and return a list of nodes that have been influenced during the round of diffusion. The return list of active nodes will change states, going from inactive to active. The community will then be modified by removing the user that had their state change to active. The remaining nodes in the community will all be inactive. I will then recalculate the number of nodes within each community. If the current community is still the largest, the procedure described above is repeated. If another community is larger, I will use the larger community to select our next seed, repeating the procedure. This procedure will repeat until either all nodes within the network have been influenced or propagation has stop after a set number of iterations [18-38].

IV. CONCLUSION

In this paper, I proposed a propagation model considering credibility and community detection. The proposed model has a potential to provide a more efficient method of information propagation and calculates influence on online social networks. In the future, I will conduct extensive experiments to comparing this model with existing models.

REFERENCES

[1] Raymond Ahn, Justin Zhan, Using proxies for node immunization identification on large graphs, IEEE Access, Vol. 5, pp. 13046-13053, 2017.

[2] Gary Blosser, Justin Zhan, Privacy preserving collaborative social network, International Conference on Information Security and Assurance, pp. 543-548, 2008.

[3] C Chiu, J Zhan, F Zhan, Uncovering suspicious activity from partially paired and incomplete multimodal data, Vol. 5, pp. 13689-13698, IEEE Access, 2017.

[4] Brittany Cozzens, Richard Huang, Maxwell Jay, Kyle Khembunjong, Sahan Paliskara, Felix Zhan, Mark Zhang, Shahab Tayeb, Signature Verification Using a Convolutional Neural Network, University of Nevada Las Vegas AEOP/STEM/REAP/RET Programs Technical Report, 2018.

[5] Wei Chen, Yajun Wang, and Siyu Yang. "Efficient influence maximization in social networks". In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM. 2009, pp. 199–208.

[6] Wei Chen, Yifei Yuan, and Li Zhang. "Scalable influence maximization in social networks under the linear threshold model". In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. IEEE. 2010, pp. 88–97.

[7] Pravin Chopade, Justin Zhan, Marwan Bikdash, Node attributes and edge structure for large-scale big data network analytics and community detection, 2015 IEEE International Symposium on Technologies for Homeland Security, pp. 1-8, 2015.

[8] Pravin Chopade, Justin Zhan, A framework for community detection in large networks using game-theoretic modeling, IEEE Transactions on Big Data, Vol. 3, Issue 3, pp. 276-288, 2017.

[9] Pravin Chopade, Justin Zhan, Structural and functional analytics for community detection in large-scale complex networks, Journal of Big Data, Vol. 2, Issue 1, , 2015

[10] Amit Goyal, Francesco Bonchi, and Laks. A data-based approach to social influence maximization. .In Proceedings of the VLDB Endowment, 5.1 (2011), pp. 73–84.

[11] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. "Celf++: optimizing the greedy algorithm for influence maximization in social networks". In: Proceedings of the 20th international conference companion on World wide web. ACM. 2011, pp. 47–48.

[12] Mehdi Heidari, Masoud Asadpour, and Heshaam Faili. "SMG: Fast scalable greedy algorithm for influence maximization in social networks". In: Physica A: Statistical Mechanics and its Applications 420 (Feb. 2015), pp. 124–133. DOI: 10.1016/j.physa.2014.10.088.

[13] David Kempe, Jon Kleinberg, and ´ Eva Tardos. "Maximizing the Spread of Influence through a Social Network". In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 137-146 (July 2003). DOI: 10 . 1145 / 956750 .956769.

[14] Jan Kostka, Yvonne Anne Oswald, and Roger Wattenhofer. "Word of mouth: Rumor dissemination in social networks". In International Colloquium on Structural Information and Communication Complexity. Springer. 2008, pp. 185–196.

[15] Nishith Pathak, Arindam Banerjee, and Jaideep Srivastava. A Generalized Linear Threshold Model for Multiple Cascades. In: Dec. 2010, pp. 965–970.

[16] Matin Pirouz, Justin Zhan, Shahab Tayeb, An optimized approach for community detection and ranking, Journal of Big Data, Vol. 3, Issue 1, pp. 22, 2016.

[17] Matin Pirouz, Justin Zhan, Optimized relativity search: node reduction in personalized page rank estimation for large graphs, Journal of Big Data, Vol. 3, Issue 1, 2016.

[18] Shahab Tayeb, Matin Pirouz, Brittany Cozzens, Richard Huang, Maxwell Jay, Kyle Khembunjong, Sahan Paliskara, Felix Zhan, Mark Zhang, Justin Zhan, Shahram Latifi, Toward data quality analytics in signature verification using a convolutional neural network, 2017 IEEE International Conference on Big Data, pp. 2644-2651, 2017.

[19] Haysam Selim, Justin Zhan, Towards shortest path identification on large networks, Journal of Big Data, Vol. 3, Issue 1, pp. 10, 2016

[20] Xian-Ming Xu, Justin Zhan, Hai-tao Zhu, Using social networks to organize researcher community, International Conference on Intelligence and Security Informatics, pp. 421-427, 2008.

[21] Felix Zhan, Gabriella Laines, Sarah Deniz, Sahan Paliskara, Irvin Ochoa, Idania Guerra, Shahab Tayeb, Carter Chiu, Matin Pirouz, Elliott Ploutz, Justin Zhan, Laxmi Gewali, Paul Oh, Prediction of online social networks users' behaviors with a game theoretic approach, pp. 1-2, 2018 15th IEEE Annual Consumer Communications & Networking Conference.

[22] Felix Zhan, Brandon Waters, Maria Mijangos, LeAnn Chung, Raghav Bhagat, Tanvi Bhagat, Matin Pirouz, Carter Chiu, Shahab Tayeb, Elliott Ploutz, Justin Zhan, Laxmi Gewali, An efficient alternative to personalized page rank for friend recommendations, pp. 1-2, 2018 15th IEEE Annual Consumer Communications & Networking Conference.

[23] Justin Zhan, Xing Fang, A novel trust computing system for social networks, IEEE Third International Conference on Social Computing, pp. 1284-1289, 2011.

[24] Justin Zhan, Secure collaborative social networks, IEEE Transactions on Systems, Man, and Cybernetics, Part C, Vol. 40, Issue 6, pp. 682-689, 2010.

[25] Justin Zhan, Xing Fang, Social computing: the state of the art, International Journal of Social Computing and Cyber-Physical Systems, Vol. 1, Issue 1, pp. 1-12, 2011.

[26] Huiyuan Zhang et al. “Least cost influence maximization across multiple social networks”. In: IEEE/ACM Transactions on Networking (TON) 24.2 (2016), pp. 929–939.

[27] Justin Zhan, Xing Fang, A computational trust framework for social computing (a position paper for panel discussion on social computing foundations), IEEE Second International Conference on Social Computing, pp. 264-269, 2010.

[28] Justin Zhan, Gary Blosser, Chris Yang, Lisa Singh, Privacy-preserving collaborative social networks, International Conference on Intelligence and Security Informatics, pp. 114-125, 2008.

[29] Justin Zhan, Xing Fang, Trust maximization in social networks, International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 205-211, 2011.

[30] Justin Zhan, Vivek Guidibande, Sai Phani Krishna Parsa, Identification of top-K influential communities in big networks, Journal of Big Data, Vol. 3, Issue 1, pp. 16, 2016.

[31] Justin Zhan, Sweta Gurung, Sai Phani Krishna Parsa, Identification of top-K nodes in large networks using Katz centrality, Journal of Big Data, Vol. 4, Issue 1, pp. 16, 2017.

[32] Justin Zhan, Xing Fang, Peter Killion, Trust optimization in task-oriented social networks, 2011 IEEE Symposium on Computational Intelligence in Cyber Security, pp. 137-143, 2011.

[33] Tian Zhu, Bai Wang, Bin Wu, Chuanxi Zhu. “Maximizing the spread of influence ranking in social networks”. In: Information Sciences 278 (2014), pp. 535–544.

[34] Carter Chiu and Justin Zhan, Deep Learning for Link Prediction in Dynamic Networks Using Weak Estimators, IEEE Access, Volume 6, Issue 1, pp., 2018.

[35] Moinak Bhaduri and Justin Zhan, Using Empirical Recurrences Rates Ratio for Time Series Data Similarity, IEEE Access, Volume 6, Issue 1., pp.30855-30864, 2018.

[36] Jimmy Ming-Tai Wu, Justin, Zhan, and Sanket Chobe, Mining Association Rules for Low Frequency Itemsets, PLoS ONE 13(7): e0198066. , 2018.

[37] Payam Ezatpoor, Justin Zhan, Jimmy Ming-Tai Wu, and Carter Chiu, Finding Top-k Dominance on Incomplete Big Data Using MapReduce Framework, IEEE Access, Volume 6, Issue 1, pp. 7872-7887, 2018.

[38] Pravin Chopade and Justin Zhan, Towards A Framework for Community Detection in Large Networks using Game-Theoretic Modeling, IEEE Transactions on Big Data, Volume: 3, Issue: 3, pp.276-288, 2017.

Moinak Bhaduri, Justin Zhan, and Carter Chiu, A Weak Estimator For Dynamic Systems, IEEE Access, Volume 5, Issue 1, pp. 27354-27365, 2017.

[39] Matin Pirouz and Justin Zhan, Toward Efficient Hub-Less Real Time Personalized PageRank, IEEE Access, Volume 5, Issue 1, pp. 26364-26375, 2017.

[40] Moinak Bhaduri, Justin Zhan, Carter Chiu, and Felix Zhan, A Novel Online and Non-Parametric Approach for Drift Detection in Big Data, IEEE Access, Volume 5, Issue 1, pp. 15883-15892, 2017.

[41] Carter Chiu, Justin Zhan, and Felix Zhan, Uncovering Suspicious Activity from Partially Paired and Incomplete Multimodal Data, IEEE Access, Volume 5, Issue 1, pp. 13689 - 13698 ,2017.

[42] Jimmy Ming-Tai Wu, Justin Zhan, Jerry Lin, Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets, IEEE Access, Vol. 5, No. 1, pp. 10024–10039, 2017.

[43] Justin Zhan and Binay Dahal, Using Deep Learning for Short Text Understanding, Journal of Big Data, 4: 34, pp. 1-15, 2017.

[44] Wensheng Gan, Jerry Chun-Wei Lin, Philippe Fournier-Viger , Han-Chieh Chao, Justin Zhan, and Ji Zhang, Exploiting Highly Qualified Pattern with Frequency and Weight Occupancy, Knowledge and Information Systems, Knowledge and Information Systems, pp. 1-32, 2017.

[45] Justin Zhan, Timothy Rafalski, Gennady Stashkevich, Edward Verenich, Vaccination Allocation in Large Dynamic Networks, Journal of Big Data, 4:2, 2017.

[46] Zhan, J., Oommen, J., and Crisostomo, J., Anomaly Detection in Dynamic Systems Using Weak Estimator, *ACM Transaction on Internet Technology*, Vol. 11, No. 1, pp. 53-69, 2011.

[47] Zhan, J., Hsieh C., Wang, I., Hsu, T., Liau, C., and Wang D., Privacy-Preserving Collaborative Recommender Systems, *IEEE Transaction on Systems, Man, and Cybernetics*, Part C, Volume 40, Issue 4, pp. 472-476, 2010.

[48] Wang, I., Shen, C., Zhan, J., Hsu, T., Liau, C. and Wang, D., Empirical Evaluations of Secure Scalar Product, *IEEE Transactions on Systems, Man, and Cybernetics*, Part C, Vol. 39, Issue 4, pp. 440-447, 2009.

[49] Jerry Lin, Tzung-Pei Hong, Philippe Fournier-Viger, Qiankun Liu, Jia-Wei Wong, and Justin Zhan, Efficient Hiding of Confidential High-utility Itemsets with Minimal Side Effects, Journal of Experimental & Theoretical Artificial Intelligence, Vol. 0, Iss. 0, pp. 1-21, 2017.