Design of an ACO algorithm for Solving Community Finding Problems

Antonio Gonzalez-Pardo
Computer Science Department
Universidad Autónoma de Madrid
Email: antonio.gonzalez@uam.es

David Camacho
Computer Science Department
Universidad Autónoma de Madrid
Email: david.camacho@uam.es

***Abstract*—The amount of data generated by social media users is increasing exponentially mainly produced by the high number of users connected everyday that interacts with each other through the Social Network (SN). As a result, SNs has become an interesting domain for research due to the wide variety of problems that can be solved. Among these problems, this work is focused on Community Finding Problems (CFP) whose goal is to group the different users in several clusters in such a way users belonging to the same cluster are similar (according to a specific metric) whereas they are different from the users of the other clusters. In this work, we describe the algorithm proposed in [1]. This algorithm for CFP is based on Ant Colony Optimization (ACO) algorithm, and it uses the information regarding the topology of the network, i.e. the connections of the users in the SN. For the experimental phase, we have compared the performance of the described algorithm against the performance of some well-known algorithms extracted from the State-of-the-Art. The results reveal that the proposed Topology-based ACO algorithm is a good approach to solve the community finding problem and it provides competitive results against the analyzed algorithms.**

I. Introduction

Nowadays, Social Networks (SNs) has become a powerful tool for a wide variety of purposes. The main reason lies in the number of users that connects to these SNs everyday. Initially, the goal of any SN is to put people in touch with each other and also to publish different media contents. In this sense, for example Instagram is mainly used for sharing photos, whereas Twitter is used for sharing small text messages called tweets.

Nevertheless, the number of users connected to these SNs generates such quantity of information that its analysis has attracted the attention of both, the research community and the industry. The main reason is the wide variety of applications that can be developed using these data.

In this sense, the most classical problem faced when working with Twitter is to analyze the content of the tweets and perform sentiment analysis over the tweets published by the users [2]. Other works try to identify the key players in a social network [3]. This last application domain is motivated by the usage that radical groups, as terrorist supporters, are using the SNs, such as Islamic State of Iraq and Syria (ISIS), to disseminate their propaganda. Joining both applications appear works like [4], [5], [6] where authors try to determine the radicalization level of different users based on the content of their tweets. Finally, other works tries to determine the structure hidden by the relations of users in the SNs, like [7], what is called Community Finding Problems (CFPs).

This work provides a summary of our current results presented in [1] on the detection of communities within a given SN. Although there are several valid approaches to perform this task such as Iterated Greedy [7], or Genetic Algorithms [8], in this work the algorithm selected to perform this task has been Ant Colony Optimization (ACO) [9].

II. The ACO model for CFP

This section provides a short description of the ACO model designed to find the different communities in the SN [1].

Any CFP can be modeled as a Constraint Satisfaction Problem (CSP) where the goal is to assign a cluster to each user of the SN whereas some constraints are satisfied, such as: each cluster must contain at least one user, or users belonging to the same are similar. Traditionally, solving any CSP using an ACO algorithm implies the creation of a decision graph where the ants are executed. This graph is composed by $X \cdot k$ nodes, where X is the number of elements of the dataset and k is the number of clusters. As it can be understood, the main disadvantage of this representation is related to the size of the resulting graph. If the number of nodes increases, it could be difficult to work with the corresponding graph. And also, this representation implies the definition of the number of clusters before execution.

The model described in this paper is focused on the reduction of the resulting graph. In such a way, the resulting graph does not depend on the number of clusters, but on the number of elements in the dataset (i.e. the number of users in the SN). The task of assigning a cluster to the different users is carried out by the ants, while they are moving through the graph. In this case, the ants travel through the graph visiting the different nodes that represents the users of the dataset. The goal of each ant is to visit all the nodes, because the need to analyze all the users of the dataset. Each time, any ant arrives to a node, the ant has to decide the cluster that will contain this user using the standard equation of the ACO algorithm.

In order to do that, the ants use a modified heuristic function τ_{ij} that measures how good is to incorporate the user e_i to the cluster C_j by analyzing the different connections between e_i and the other users that compose the cluster C_j. This heuristic function is shown in Eq. 1.

$$\tau_{ij} = Topology\ (e_i, C_j) = \frac{|\mathcal{N}_i \cap C_j|}{|\mathcal{N}_i|} \quad (1)$$

Where $\mathcal{N}_i$ represents the set of elements connected to e_i and $|\cdot|$ is a function that compute the number of elements contained in a specific set. Therefore, this function takes into account how many connections of e_i belongs to the cluster C_j.

III. Experimental evaluation

For testing the designed algorithm, we have used the dataset published by SNAP[1]. This dataset is composed by 10 different Ego Networks, and it has been selected because the dataset contains the groundtruth with the communities that compose the networks.

For the experimental phase, we have compared the performance of the ACO model against some of the well-known algorithms extracted from the State-of-the-Art. Table I shows for each algorithm how many networks each algorithm provides the best solution. Note that the best solution is the one with the highest Omega Index.

	Facebook	
	Total	**Percentage**
Topology-based ACO	5	50
Walktrap	1	10
Infomap	3	30
Clauset Newman Moore	1	10
Label Propagation	0	0

TABLE I

This table shows for each algorithm (first column), how many Ego Networks each algorithm has provided the best solution and the related percentage (%) for each dataset analyzed. The best solution is the one that provides the highest Omega Index score because it is the closest to the groundtruth.

IV. Conclusions and Future Work

This work provides a summary of the work published in [1]. In that work, we described an implementation of a new Ant Colony Optimization (ACO) algorithm to perform the community finding tasks in Social Networks (SNs), based on the topology information of the network.

In the experimental phase, we compared the performance of the described algorithm against some of the state-of-the-art algorithms. The results reveal that the proposed Topology-based ACO algorithm is a good approach to solve the community finding problem because it provides competitive results against the analyzed algorithms.

As future work, we are working on the adaptation of novel bio-inspired algorithm to perform community finding tasks. These algorithms are Artificial Bee Colony [3] and a new algorithm based on the calling behavior of male Japanese tree frogs [10].

[1] http://snap.stanford.edu/data/egonets-Facebook.html

Acknowledgment

This work has been co-funded by the following research projects: Spanish Ministry of Science and Education and Competitivity (MINECO) and European Re-gional Development Fund (FEDER) under grant TIN2017-85727-C4-3-P (DeepBio); and the Justice Programme of the European Union (2014-2020) 723180 – RiskTrack – JUST-2015-JCOO-AG/JUST-2015-JCOO-AG-1. The contents of this publication are the sole responsibility of their authors and can in no way be taken to reflect the views of the European Commission.

References

[1] A. Gonzalez-Pardo, J. J. Jung, and D. Camacho, "Aco-based clustering for ego network analysis," *Future Generation Computer Systems*, vol. 66, pp. 160 – 170, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X16302278

[2] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 28:1–28:41, Jun. 2016. [Online]. Available: http://doi.acm.org/10.1145/2938640

[3] M. Lozano, C. Garca-Martnez, F. J. Rodrguez, and H. M. Trujillo, "Optimizing network attacks by artificial bee colony," *Information Sciences*, vol. 377, pp. 30 – 50, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025516312075

[4] M. Fernandez, M. Asif, and H. Alani, "Understanding the roots of radicalisation on twitter," in *In WebSci ?18: 10th ACM Conference on Web Science*. ACM (Association for Computing Machinery), May 2018. [Online]. Available: http://oro.open.ac.uk/54344/

[5] R. Lara-Cabrera, A. Gonzalez-Pardo, and D. Camacho, "Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in twitter," *Future Generation Computer Systems*, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X17308348

[6] R. Lara-Cabrera, A. G. Pardo, K. Benouaret, N. Faci, D. Benslimane, and D. Camacho, "Measuring the radicalisation risk in social networks," *IEEE Access*, vol. 5, pp. 10 892 – 10 900, 2017.

[7] J. Sánchez-Oro and A. Duarte, "Iterated greedy algorithm for performing community detection in social networks," *Future Generation Computer Systems*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X17323932

[8] C. Pizzuti, "Ga-net: A genetic algorithm for community detection in social networks," in *Parallel Problem Solving from Nature – PPSN X*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1081–1090.

[9] M. Dorigo, "Optimization, learning and natural algorithms (in Italian)," Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy, 1992.

[10] C. Blum, B. Calvo, and M. J. Blesa, "Frogcol and frogmis: new decentralized algorithms for finding large independent sets in graphs," *Swarm Intelligence*, vol. 9, no. 2, pp. 205–227, Sep 2015. [Online]. Available: https://doi.org/10.1007/s11721-015-0110-1