A Behavioral Focused Method for Community Extraction*

No Author Given

No Institute Given

Abstract. This paper proposes a community extraction method adapted to social networks based on follower & followee graphs.This method extracts communities first by behavior and then by interest to provide a natural way to understand the users' behaviors within a community. This paper introduces a method based on a hierarchical classification of community decomposition which remedies to the common problem of determining the number of extracted communities and their related sizes. We will point out how the hierarchical classification has improved the community discovery based on a mutuality aware method. We have been applying this method to the Flavorlens social platform, a dish tasting experience sharing platform.

Keywords: Community Extraction · Social Network · Graph Theory.

1 Introduction

Community extraction in social media graphs has become an important topic of research since most of social media platform offer nowadays a recommendation service [1]. Recommendation purposes vary widely from forums and groups to friends, events or ads but they are now a staple of most social platforms. Therefore, so is clustering the users into communities with common behavioral characteristics.

However, community discovery is still a challenging task. Several methods [2] differ in key aspects and steps, leading to different communities. In this paper we propose an innovative clustering method for a directed followers graph: the Behavioral Focused (so-called BF) method. This method has been designed for social networks such as Flavorlens[3] where the users post observations about dishes. One user can follow any other user (this is not a mutual relationship), a user can also like, share, comment or add to his eating out list any observations from any of other users. Moreover, this method doesn't only take the follower links into account but also the interactions between users (such as liking, commenting, sharing and adding observations to their eating out list).

We can indeed observe very different kinds of behaviors from the users. For instance we can identify 'social' users who highly interact with each other and contrariwise, 'lurker' users who are interested in the content but do not really

* Supported by organization x.

interact with other users nor post a lot of observations. This example is great to illustrate that users with different behaviors would be interested in recommendations of different nature. Flavorlens social users would probably be interested in finding new users to interact with whereas lurker users would most likely be interested in new observations. Once we know which nature of recommendations is most fitted to a group of users it is advantageous to group them by interest so we can determine what content would be the best recommendation. The rest of this paper is structured as follows. Section 2 formalizes the community management in Flavorlens. The Behavioral Focused Method is described in Section 3. We conclude and explore future research directions in Section 4.

2 Formalisation of the Community Management in Flavorlens

Let O be the set of dish Observations and M the set of Flavorlens members, we define:

- The set of followers F_m, which includes the members followed by m, $m \in M$.
- The set of dishes O_m, which includes the dish observations collected by m.
- The strength of the relationship between $A \in M$ and $B \in F_A$ at time t can be defined by[1]:

$$St_{A \to B} = \frac{(t - t_f)}{(t - t_i)} \times \frac{(2Cut_{A,B} + L_{A,B} + Com_{A,B} + Sh_{A,B})}{5} \tag{1}$$

  With:
  - t_i : time when A joined Flavorlens
  - t_f : time when A started to follow B
  - $L_{A,B}$: percentage of observations O_B that A liked
  - $Com_{A,B}$: percentage of observations O_B that A commented
  - $Sh_{A,B}$: percentage of observations O_B that have been shared by $B \in F_A$ or someone in $A's$ cluster
  - $Cut_{A,B}$: percentage of observations O_B that A puts in his eating out list.

- The strength of a potential relationship between $A \in M$ and $B \notin F_A$ can be defined by:

$$St^{(p)}_{A \to B} = \frac{(2Cut_{A,B} + L_{A,B} + Com_{A,B} + Sh_{A,B})}{5} \tag{2}$$

From the Flavorlens social network, we extract the directed Follower graph, $FLFG^2 = (V, E_f)$, where E_f is the set of edges representing the connections given by F_m. In Section 2, We will illustrate by applying our method to the directed FLFG Graph as it is shown in Fig.1.

[1] The weight of $Cut_{A,B}$ is higher because it means that the person wants to try the same dish, they have therefore been influenced

[2] FLFG : FlavorLens Follower Graph

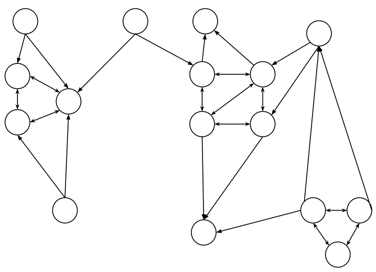

Fig. 1: Sample of the directed Follower FLFG Graph extracted from Flavorlens.

3 The Behavioral Focused Method

Definition 1. *(Ghost user) A ghost user is a user who doesn't interact with other users, its node in the follower graph have no edges.*

Definition 2. *(Satellite node) A satellite node is a node which is not included in a cluster*[3] *but this node still has several edges in and/or out connected to one or more clusters. There are two categories of satellite nodes: the in-satellites and the out-satellites. The in-satellites have in-edges and the out-satellites have out-edges (one node can be both).*

The BF method is broken down into 4 steps:

- Step 1: Extracting communities with a mutuality focused method[4] to get a graph with very cohesive clusters with a high number of mutual connections inside the clusters and as little as possible between the clusters (see Fig.2 (a)). The general performance of the mutuality focused method is optimal except in case of ghost or satellite nodes.

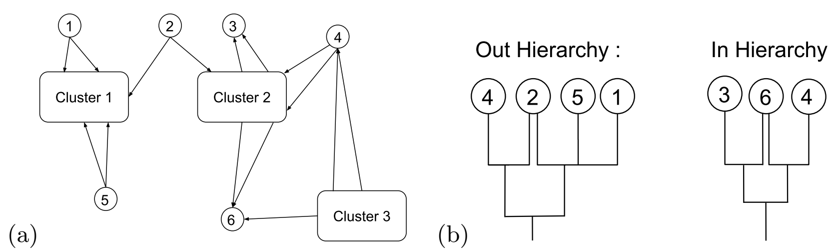

Fig. 2: The Follower's Graph present in Fig.1 After the Mutuality Focused Clustering (a) and The Hierarchical Classification of the Satellite Communities (b)

[3] node not strongly mutually linked to any cluster

- Step 2: Grouping the ghost and satellite nodes respectively into a ghost and a satellite community since the users express similar behaviors.
- Step 3: Separating the satellite community in two communities which respectively contains the in and out-satellite nodes.
- Step 4: Creating hierarchical classifications in two steps for the two communities generated in step 3 (See Fig.2 (b)). First, by creating communities based on which clusters the satellite nodes are connected with, all satellites nodes connected to the same cluster will form a new community. Secondly, by iteratively merging the communities created by merging the two most similar communities at each iteration. The similarity between two communities is defined using equations (1), (2) and their overlapping. The iterative process ends when all the satellite communities have been merged into a single community. Since it's often impossible to know beforehand when to stop the merging process to get the most interesting and relevant communities, we save each iteration into a hierarchical classification[5] (see Fig.2 (b)) to access all different possible communities for the satellite communities.

4 Conclusion and Future Work

Extracting communities from social networks has become a reoccurring problem since the thrive of social media platforms. In this article, our innovation is the BF method because it gives indications on both the behavior of the user and their interests. Moreover, the hierarchical classification allows us to really tailor the clustering to the community detection requirements. As future work, the fuzzy logic especially in the hierarchical clustering process should be considered to gain precision in the community understanding, as well as trying to verify the scalability of the method on larger graphs.

References

1. I. Guy, et al.: Social Media Recommendation based on People and Tags. In Proc. of the 33^{rd} int. ACM SIGIR pp. 194-201. Geneva, Switzerland (2010). https://doi.org/10.1145/1835449.1835484
2. F.D. Malliaros and M. Vazirgiannis: Clustering and Community Detection in Directed Networks: A Survey. In: Physics Reports, Vol. 533, Issue 4, pp. 95-142 (2013). https://doi.org/10.1016/j.physrep.2013.08.002
3. A. Fritzen, F. Andres, and M. Leite: Introducing Flavorlens: A Social Media Platform for Sharing Dish Observations. In Proc. of the 3^{rd} Int. Workshop ACM MHFI'18. NY, USA,(2018). https://doi.org/10.1145/3279954.3279961
4. Y. Li, Z.L. Zhang, and J. Bao: Mutual or unrequited love: identifying stable clusters in social networks with uni- and bi-directional links. In Proc. of the 9^{th} int. WAW'12. In: A. Bonato and J. Janssen (Eds.). Springer-Verlag. Berlin, Heidelberg, pp. 113–125, (2012). https://doi.org/10.1007/978-3-642-30541-2_9
5. R. Toujani and J. Akaichi: An approach based on mixed hierarchical clustering and optimization for graph analysis in social media network: toward globally hierarchical community structure. In: J. Knowl Inf Syst (2019). https://doi.org/10.1007/s10115-019-01329-2