The Journal of Systems and Software 94 (2014) 30–38

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Efficient unveiling of multi-members in a social network

M. Nidhal Jelassi[a,b,c,d,*], Christine Largeron[b,c,d], Sadok Ben Yahia[a,b,e]

[a] University of Tunis El Manar, Faculty of Sciences of Tunis, Lipah, Tunis, Tunisia
[b] Université de Lyon, Saint-Étienne, F-42023, France
[c] CNRS, UMR 5516, Laboratoire Hubert Curien, Saint-Étienne, F-42023, France
[d] Université de Saint-Étienne, Jean-Monnet, Saint-Étienne, F-42023, France
[e] Institut Mines - Telecom, Telecom SudParis, UMR CNRS Samavor, 91011 Evry Cedex, France

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 22 May 2013
Accepted 24 June 2013
Available online 13 July 2013

Keywords:
Social network
Minimal transversal
Essential itemset
Hypergraph

ABSTRACT

With the rapid growth of the Web 2.0, the discovery of key actors in social networks, called influencers, mediators, ambassadors or experts, has recently received a renewed of attention. In this article, we consider a particular type of actor that we call a multi-member since he belongs to several communities. We introduce a methodological framework to identify these actors in a hypergraph, in which the vertices are the actors and the hyperedges are the communities. We also show that detecting such multi-members is similar to the problem of the determination of a subset of minimal transversals of a hypergraph. An efficient algorithm that relies on the connection between the definition of a multi-member and that of an essential itemset is also introduced. Experiments done on several datasets showed that the introduced algorithm outperforms the pioneering ones of the literature.

1. Introduction

A social network can be defined as a set of entities interconnected with each other. The entities are generally actors or organizations and the relations depict their interactions. The analysis of the strategic position that some actors may have in the network and their identification were mainly studied in the field of sociology notably using measures like degree, closeness, betweenness, designed to highlight the "most important" actors in the network (Freeman, 1979; Wasserman and Faust, 1994; Borgatti, 1992). The appearance of Web 2.0 with online social networking sites has led to a renewed interest for the Social Networks Analysis (SNA), notably to identify actors called, depending on the authors, influencers, mediators, ambassadors or experts (Everett and Borgatti, 2005; Domingos, 2005; Agarwal et al., 2008; Opsahl and Hogan, 2010).

In this regard, the community structure of the network has also been considered (Scripps et al., 2007a; Forestier et al., 2012). Forestier et al. have notably pointed out that there exist roles that are defined by the communities (Forestier et al., 2012). Generally, a community structure corresponds to a partition of the network actors where each actor belongs to one community. However, in real-life, an actor can often belong to several groups. In this article, we focus on such actors who participate in more than one community. Moreover, we suppose that these overlapping communities are given or that they can be detected by suited algorithms (Baumes and Magdon-Ismail, 2005; Palla et al., 2005; Lancichinetti et al., 2010). Thus, we know to which community(ies) each actor belongs even if the relationships between each pair of actors are not known.

In this context, the aim of our work is to determine the smallest subset of actors, called "*multi-members*" in the remainder of the article, who are able to represent as best as possible the different communities. For instance, in project management, we may know the skills required to carry out a project and those of the actors. A community is composed of the subset of actors who share a same skill and the communities are not disjoint because an actor can have several skills. In that case, the multi-members correspond to the smallest subset of actors who have the necessary skills for the project. In another example related to a marketing analysis, we can consider a product for which there are several models or versions introduced at different times. We consider that each community is formed of the people who bought the same model and we search for the smallest number of people who bought several models and who are able to compare these models, ideally one user who bought all the models.

In order to determine the smallest subset of actors who are able to represent as best as possible the different communities, we advocate in this paper the use of concepts drawn from the hypergraph theory. The hypergraph is used to represent the communities. Then we introduce a new efficient approach for the discovery of

* Corresponding author at: University of Tunis El Manar, Faculty of Sciences of Tunis, Lipah, Tunis, Tunisia. Tel.: +216 71 872 600.
E-mail address: nidhal.jelassi@fst.rnu.tn (M.N. Jelassi).

0164-1212/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2013.06.061

multi-members in the hypergraph. To sum up, the main aim of this paper is threefold:

1. We introduce a methodological framework to identify multi-members even without the adjacency matrix. To do so, we suggest the use of a hypergraph, defined as an extension of a graph, in which the vertices are the actors and the hyperedges are the communities.
2. We show that the detection of multi-members in a hypergraph is similar to the problem of the determination of a subset of minimal transversals (Berge, 1989). Interestingly, even if the extraction of minimal transversals is a NP-Hard problem, it has many applications in various fields of computer science, e.g., game theory, databases, artificial intelligence, data mining, machine learning, to cite but a few (Hagen, 2008).
3. We introduce a new algorithm to extract the multi-members from a hypergraph and we show experimentally its efficiency compared to those of the literature. This algorithm exploits the connection between the definitions of a multi-member and of an *Essential itemset* (Casali et al., 2005), notion that we borrow from the data mining field.

This article is organized as follows. In Section 2, we scrutinize the related work.

Section 3 briefly recalls all the mathematical background required in Section 4, notably the notions at the crossroads of graph theory and data mining. Section 4 formally introduces our definition of a multi-member into a hypergraph. Section 5 describes in detail our contributions in order to identify the multi-members into a hypergraph. In Section 6, we present the promising results of the empirical study that we carried out on different datasets obtained notably from social tagging activities. Finally, Section 7 concludes this paper and gives avenues for future work.

2. Scrutiny of related work

In this article, we consider a social network and we suppose that there is an overlapping community structure in other words that the actors belong at least to one community and possibly to more than one community. If the community structure is not defined but the relations between pairs of actors are fully known, the communities can be detected using for example the clique percolation algorithm of Palla et al. implemented in CFinder (Palla et al., 2005), or the algorithms, respectively, proposed by Baumes and Magdon-Ismail (2005) and OSLOM (Order Statistics Local Optimization Method) introduced by Lancichinetti et al. (2010). The communities can also be given by the context. Indeed, in various applications, the relations between the pairs of actors are not available, in other words the adjacency matrix of the graph associated to the network is not available but, the belonging of the actors to communities is known. In that last case, the algorithms and indicators, mentioned in introduction such as the centrality or the prestige, cannot be used to identify actors which have a strategic position. One advantage of our approach is to provide an answer to this problem even in such case. Moreover, it consists in determining key actors in relation to this community structure. Indeed, as highlighted by Scripps et al. or Wang et al., it would be of benefit for such a task to have a finer understanding of the position already occupied by the actors within the communities into the network (Scripps et al., 2007a,b; Wang et al., 2010). However, as highlighted previously, in real networks it is common that communities overlap with each other. Our aim consists precisely in searching actors who participate in more than one community.

The key actors that we want to identify, correspond to the smallest subset of actors who are able to represent as best as possible the different communities. Such actors are called *multi-members* since a majority of them belong to several groups. They are, in a certain way, related to the *ambassadors* defined by Scripps (Scripps et al., 2007b). In the trivial case, where it exists one or several actors who belong to all the communities, the objective is to find one of them. To solve this problem in its general form, we propose to represent the communities by a hypergraph and to define the set of multi-members as a minimal transversal which fulfils certain conditions (Berge, 1989).

Several algorithms have been presented in the literature for the extraction of a minimal transversal of a hypergraph. Chronologically, Berge was the first author who proposed an algorithm dedicated to this task but this algorithm requires the storage of intermediate cross whose number may be exponential (Berge, 1989). Thus, this algorithm is not applicable even in the case of hypergraphs consisting of more than ten vertices. Kavvadias et al. have proposed a new algorithm, based on that of Berge, that proceeds incrementally, but this algorithm employs a depth-first traversal of the graph where edges are computed by a Cartesian product of the generalized vertex (set of vertices belonging to the same hyperedges) (Kavvadias and Stavropoulos, 2005). An alternative is MTminer, proposed recently by Hebert et al., which extracts the minimal transversal in a level-wise manner and which is based on the anti-monotone property (Hébert et al., 2007). In a previous work, Jelassi et al. (2012), we introduced a first algorithm M2D which exploits essential itemset property to detect the hypergraph minimal transversal multi-members (Tmm).

In this article, we propose O-M2D an optimized version of this M2D which uses the minimal transversality value to directly jump to the level in which the Tmms stand (Berge, 1989). Moreover we detailed the proofs of the propositions underlying these algorithms.

In the next section, we present the theoretical framework that is useful in the article. We provide notably an original connection between the notion of minimal transversal and the notion of essential itemset that we would use to represent the multi-members.

3. Key notions

In the following, we briefly recall the definitions of hypergraph and minimal transversal and we show their connection with notions from the data mining field, namely: extraction context and essential itemset.

Definition 1 (hypergraph *(Berge, 1989)*). Let $H = (\mathcal{X}, \xi)$ with $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ a finite set of elements and $\xi = \{e_1, e_2, \ldots, e_m\}$ a family of subsets of $\mathcal{X}$. H is a hypergraph on $\mathcal{X}$ if:

1. $e_i \neq \emptyset, i \in \{1, \ldots, m\}$;
2. $\bigcup_{i=1,\ldots,m} e_i = \mathcal{X}$.

The elements of $\mathcal{X}$ are called *vertices* of the hypergraph and they correspond to the entities. Elements of ξ, called *hyperedges* of the hypergraph, correspond to the communities. Fig. 1 depicts a hypergraph $H = (\mathcal{X}, \xi)$ such that $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $\xi = \{\{1, 2\}, \{2, 3, 7\}, \{3, 4, 5\}, \{4, 6\}, \{6, 7, 8\}, \{7\}\}$.

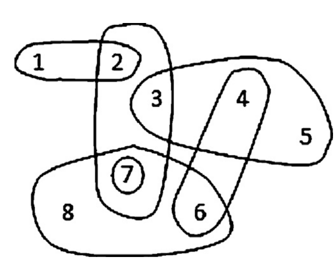

Fig. 1. An example of a hypergraph.