

Chapter 37

Data Mining Techniques for Communities' Detection in Dynamic Social Networks

Céline Robardet
Université de Lyon, France

ABSTRACT

Social network analysis studies relationships between individuals and aims at identifying interesting substructures such as communities. This type of network structure is intuitively defined as a subset of nodes more densely linked, when compared with the rest of the network. Such dense subgraphs gather individuals sharing similar property depending on the type of relation encoded in the graph. In this chapter we tackle the problem of identifying communities in dynamic networks where relationships among entities evolve over time. Meaningful patterns in such structured data must capture the strong interactions between individuals but also their temporal relationships. We propose a pattern discovery method to identify evolving patterns defined by constraints. In this paradigm, constraints are parameterized by the user to drive the discovery process towards potentially interesting patterns, with the positive side effect of achieving a more efficient computation. In the proposed approach, dense and isolated subgraphs, defined by two user-parameterized constraints, are first computed in the dynamic network restricted at a given time stamp. Second, the temporal evolution of such patterns is captured by associating a temporal event types to each subgraph. We consider five basic temporal events: the formation, dissolution, growth, diminution and stability of subgraphs from one time stamp to the next one. We propose an algorithm that finds such subgraphs in a time series of graphs processed incrementally. The extraction is feasible thanks to efficient pruning patterns strategies. Experimental results on real-world data confirm the practical feasibility of our approach. We evaluate the added-value of the method, both in terms of the relevancy of the extracted evolving patterns and in terms of scalability, on two dynamic sensor networks and on a dynamic mobility network.

DOI: 10.4018/978-1-4666-2455-9.ch037

INTRODUCTION

Social network analysis conceives social relationships in terms of graphs of interactions whose nodes represent individual actors within the networks and links social interactions such as ideas, friendship, collaboration, trade, etc. Virtual communities, and particularly online communities, are peculiar social networks whose analysis is facilitated by the fact that the network is in some sense monitored continuously. Social networks have attracted a large amount of attention from epidemiologists, sociologists, biologists and computer scientists that have shown the ubiquitous role played by social networks in determining the way problems are solved or organizations are run. The study of such networks has attracted much attention in the recent years and has proceeded along two main tracks: the analysis of graph properties, such as degree distribution, diameter or simple graph patterns such as cliques (Scherrer et al., 2008, Leskovec et al., 2005), and the identification of communities, which are loosely defined as collections of individuals who interact unusually frequently (Newmann, 2004, Palla et al., 2005). Communities reveal properties shared by related individuals. However, most of the interesting real-world social networks that have attracted the attention of researchers in the last few years are intrinsically time dependent and tend to change dynamically. As new nodes and edges appear while some others disappear over time, it seems decisive to analyze deeply the evolution of such dynamic graphs. Furthermore, there is a crucial need for incremental methods that enable to find groups of associated nodes and detect how these structures change over time.

Communities are loosely defined as highly connected subgraphs that are also isolated from the rest of the graph. Such properties can be captured by measures such as modularity (Newman, 2004) used to find disjoint communities forming a partition. The modularity of a given partition of nodes is the number of edges inside clusters (as opposed

to crossing between clusters), minus the expected number of such edges if the graph was random conditioned on its degree distribution. Community structures often maximize the modularity measure. However, this measure has an intrinsic resolution scale, and can therefore fail to detect communities smaller than that scale and favor in general communities of similar size (Fortunato et al., 2007). Moreover, it has been shown (Brandes et al., 2008) that finding the community structure of maximum modularity for a given graph is NP-complete and thus heuristics have been proposed that approximate this optimization problem.

Instead of directly looking for a global structure of the graph, such as a partition of the vertices, it can be more efficient to proceed in two steps. One might first compute subgraphs that capture locally strong associations between vertices and then use these local patterns to construct a global model of the graph's dynamics. Such a framework provides more interesting patterns when the analyst can specify his inclination by means of constraints. Many pattern mining under local constraints techniques (e.g., looking for frequent patterns, data dependencies) have been studied extensively the last decade (Morik et al., 2005). One crucial characteristic of local pattern mining approaches is that the interestingness of a pattern can be computed independently of the other patterns. Such framework enables the analyst to specify a priori relevancy of pattern by means of constraints. The constraints have been identified as a key issue to achieve the tractability of many data mining tasks: useful constraints can be deeply pushed into the extraction process such that it is possible to get complete (every pattern which satisfies the user-defined constraint is computed) though efficient algorithms.

Specific subgraphs defined by constraints have already been examined. Fully connected subgraphs, also called cliques, are a local pattern type that has been considered as communities. Palla et al. (2005) consider that communities rely on several complete (fully connected) subgraphs of size

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/data-mining-techniques-communities-detection/73467

Related Content

Clubhouse Experience: Sentiment Analysis of an Alternative Platform From the Eyes of Classic Social Media Users

Ipek Deveci Kocakoçand Pnar Özkan (2022). *Data Mining Approaches for Big Data and Sentiment Analysis in Social Media* (pp. 244-264).

www.irma-international.org/chapter/clubhouse-experience/293159

A Taxonomy for Distance-Based Spatial Join Queries

Lingxiao Liand David Taniar (2017). *International Journal of Data Warehousing and Mining* (pp. 1-24).

www.irma-international.org/article/a-taxonomy-for-distance-based-spatial-join-queries/185656

Parallel Real-Time OLAP on Multi-Core Processors

Frank Dehneand Hamidreza Zaboli (2015). *International Journal of Data Warehousing and Mining* (pp. 23-44).

www.irma-international.org/article/parallel-real-time-olap-on-multi-core-processors/122514

A Novel Method for Classifying Function of Spatial Regions Based on Two Sets of Characteristics Indicated by Trajectories

Haitao Zhang, Chenguang Yuand Yan Jin (2020). *International Journal of Data Warehousing and Mining* (pp. 1-19).

www.irma-international.org/article/a-novel-method-for-classifying-function-of-spatial-regions-based-on-two-sets-of-characteristics-indicated-by-trajectories/256160

A Proactive Approach for BI

Stanley Loh (2015). *Integration of Data Mining in Business Intelligence Systems* (pp. 83-97).

www.irma-international.org/chapter/a-proactive-approach-for-bi/116809