Chapter 45
Graph Mining and Its Applications in Studying Community–Based Graph under the Preview of Social Network

Bapuji Rao
VITAM, India

Anirban Mitra
VITAM, India

ABSTRACT

One of the fundamental tasks in structured data mining is discovering of frequent sub-structures. These discovered patterns can be used for characterizing structure datasets, classifying and clustering complex structures, building graph indices & performing similarity search in large graph databases. In this chapter, the authors have discussed on use of graph techniques to identify communities and sub-communities and to derive a community structure for social network analysis, information extraction and knowledge management. This chapter contributes towards the graph mining, its application in social network using community based graph. Initial section is related literature and definition of community graph and its usage in social contexts. Detecting common community sub-graph between two community graphs comes under information extraction using graph mining technique. Examples from movie database to village administration were considered here. C++ programming is used and outputs have been included to enhance the reader's interest.

INTRODUCTION

Social network can be defined as the set of relationships between individuals where each individual is a social entity. It represents both the collection of ties between people as well as the strength of those ties (Mitra, Satpathy, Paul, 2013). In a general way, Social network is used as a measure of social "connectedness", within the social networks for observing and calculating the quality and quantity of informa-

tion flow within individuals and also within groups. Hence, from author's point of discussion, a social network can be defined as structure comprises of social actors (consisting of either group of individuals or organization) and the connectivity among them. Further, such structured network can also be called as social structure.

Network comprising of social entities becomes active when the relationships get established in course of regular interaction in the process of daily life and living, cultural activities such as marriage, thread ceremonies, different communities yearly celebrations, engagements, etc. and so on. Among vast examples, regular interaction may be a household requesting another for help, support or advice, creation of new friendship or choice of individuals to spend leisure time together; and many more. Sometimes a relationship can be a *negative*, i.e. hostility or alliance, alienation as against mutuality or integration with even having the security aspect as an important factor (Tripathy & Mitra, 2012).

To extract information or pattern of interaction between two or more social entities or between two or more social group, one needs to look deep into the properties of social network and the interactions. To analyse the above said process, the authors have put the network into mathematical model using the concept and properties of graph theory. About one of the properties on interaction, the Social networks show strong community relationships for interaction of people, such that, these interaction may be limited with in specific group or community or exceed outside the virtual boundary of the community or group. Relationships between communities can be analyzed using the basic algebraic concept of *transitivity*. Considering a simple example of three actors (say A^1, A^2 and A^3), there is a high possibility that if A^1 and A^2 are friends, A^2 and A^3 are friends then most likely A^1 and A^3 are also friends (however, the degree of their friendship may differs and can be easily demonstrate using weighted graph). Further, the property of transitivity can be used to measure the *Clustering coefficients*.

Due to the property of strong community effect, the actors or the social entities in a network form a group which is closely connected. These groups are termed as modules or clusters or communities or even sub-groups. The authors have observed that individuals interact more frequently within a group. To detect similar groups within a social network, which is also known as similar *community detection* is a major challenge in analyzing the social network. To extract such type of communities helps in solving some more tasks, that are associated with the analysis of social network. Several works in terms of definitions and approaches are available in the area of community detection.

One of the most important processes in structured data mining is to discover frequent sub-structures. Authors have initiated the work with a discussion on various available techniques. Once we find at least one similar vertex between two sub-structures, then it is easy to merge both the sub-structures to form a larger structure. For this purpose the authors follow the simple graph techniques to identify at least one community between two villages community sub-structure and merge both to produce a larger community structure. The authors propose a new algorithm which merges two community sub-graphs in an efficient, easier and faster way. As each community sub-graph is represented as an adjacency matrix form which only contains 0s and 1s. Though, the adjacency matrix of both the community sub-graph contains 0s and 1s, which can be represented as a bit-matrix which substantially occupies less space for a larger community graph. In later section the authors discuss about the algorithm, memory management, and some examples related to the proposed algorithm which shows merging of two community sub-graph and produces one community graph having with at least one common community between those community sub-graphs.

Related Content

Super Computer Heterogeneous Classifier Meta-Ensembles
Anthony Bagnall, Gavin Cawley, Ian Whittley, Larry Bull, Matthew Studley, Mike Pettipherand F. Tekiner (2007). *International Journal of Data Warehousing and Mining (pp. 67-82).*
www.irma-international.org/article/super-computer-heterogeneous-classifier-meta/1785

New Information Technologies and Other Pertinent Issues Impacting the Strategic Dimension of CRM for Business Excellence
Sudhakar Kuppurajuand Girish Subramanian (2003). *Managing Data Mining Technologies in Organizations: Techniques and Applications (pp. 149-173).*
www.irma-international.org/chapter/new-information-technologies-other-pertinent/25764

A Solution to the Cross-Selling Problem of PAKDD-2007: Ensemble Model of TreeNet and Logistic Regression
Wei Mingjun, Chai Lei, Wei Renyingand Huo Wang (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments (pp. 285-289).*
www.irma-international.org/chapter/solution-cross-selling-problem-pakdd/40410

TBSGM: A Fast Subgraph Matching Method on Large Scale Graphs
Fusheng Jin, Yifeng Yang, Shuliang Wang, Ye Xueand Zhen Yan (2018). *International Journal of Data Warehousing and Mining (pp. 67-89).*
www.irma-international.org/article/tbsgm/215006

An Experimental Approach and Monitoring Tools for Evaluating a Dynamic Cubing System
Anne Tchounikine, Maryvonne Miqueland Usman Ahmed (2016). *International Journal of Data Warehousing and Mining (pp. 1-19).*
www.irma-international.org/article/an-experimental-approach-and-monitoring-tools-for-evaluating-a-dynamic-cubing-system/171117