Chapter 13

Spreading Activation Connectivity Based Approach to Network Clustering

Alexander Troussov
The Russian Presidential Academy of National Economy and Public Administration, Russia

Sergey Maruev
The Russian Presidential Academy of National Economy and Public Administration, Russia

Sergey Vinogradov
The Russian Presidential Academy of National Economy and Public Administration, Russia

Mikhail Zhizhin
Colorado University of Boulder, USA

ABSTRACT

Techno-social systems generate data, which are rather different, than data, traditionally studied in social network analysis and other fields. In massive social networks agents simultaneously participate in several contexts, in different communities. Network models of many real data from techno-social systems reflect various dimensionalities and rationales of actor’s actions and interactions. The data are inherently multidimensional, where “everything is deeply intertwingled”. The multidimensional nature of Big Data and the emergence of typical network characteristics in Big Data, makes it reasonable to address the challenges of structure detection in network models, including a) development of novel methods for local overlapping clustering with outliers, b) with near linear performance, c) preferably combined with the computation of the structural importance of nodes. In this chapter the spreading connectivity based clustering method is introduced. The viability of the approach and its advantages are demonstrated on the data from the largest European social network VK.

DOI: 10.4018/978-1-5225-2814-2.ch013
1. PROLIFERATION OF TECHNO-SOCIAL SYSTEMS POSES NEW CHALLENGES TO THE STRUCTURAL ANALYSIS OF MASSIVE NETWORKS

The proliferation of techno-social systems has lead to the emergence of massive networks connecting people and various digital artifacts. These networks become more and more multidimensional. Network models of the data have nodes, representing actors, abstract concepts, projects etc. Abstract ideas become foci of interactions leading to the formation of communities of actors, where the same actor typically belongs to many communities simultaneously. This phenomenon requires certain considerations on the approaches to study the structure of such networks, including understanding of local topology – centralization, degree distribution, as well as mezzo- and macro-level structures, such as communities. In this paper the focus is on multidimensional networks, where the nodes represent actors and various artifacts they create and do; however the approach could be useful for other types of networks dimensionalities.

Clustering is one of the most frequently used method in Data Science-related applications; for example, in a recent poll among data scientists clustering is the second most used technique after regression, coming ahead of decision trees/rules, visualization, K-nearest neighbors, principal component analysis, statistics, and text mining (Kdnuggets, 2016).

The sheer volume of Big Data and multidimensionality of the data poses novel challenges for graph-mining. This chapter focuses on the tasks traditionally associated with mining of social networks, but one can expect that the same approach is applicable to analyze the structure of network models of data.

- Centralization, which is ranking the nodes according their structural prominence.
- Communities, such as on-line communities exhibiting some properties of communities of practice. Lave & Wenger (1998) define communities of practice as “groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.” This learning that takes place is not necessarily intentional.

Two questions regarding clustering that need to be addressed by future research are:

- Centralization and community detection - how these tasks are related? If useful centralization is possible without community detection?
- Strict partitioning for detection of communities in massive social graphs with multidimensional links is not needed and might be not computationally feasible. Instead, local partitioning or local fuzzy overlapping clustering with outliers might be more appropriate.

Requirements for clustering could be summarized as follows:

- Local overlapping clustering, that is a node, could belong to several clusters. Overlapping, as contrasted with fuzzification, is an unavoidable consequence of multidimensionality. To explain why fuzzification is not an adequate approach for structural discovery in multidimensional networks, let us consider a simple example. Suppose that a person on a social network is classified as a mathematician. When the classification is based on fuzzification, new evidences that a person is interested, for instance, in classical music should decrease the value of her membership function in the group of mathematicians, while in reality there may be a positive correlation between two interests mentioned above.

www.igi-global.com/e-resources/library-recommendation/?id=109

Related Content

Managing E-Relationships in a Supply Network
www.igi-global.com/chapter/managing-relationships-supply-network/49805?camid=4v1a

An Exploration of the Critical Need for Formal Training in Leadership for Cybersecurity and Technology Management Professionals
www.igi-global.com/article/an-exploration-of-the-critical-need-for-formal-training-in-leadership-for-cybersecurity-and-technology-management-professionals/210628?camid=4v1a

Peer-to-Peer Network-Based Image Retrieval
Chun-Rong Su and Jiann-Jone Chen (2013). Multimedia Networking and Coding (pp. 377-399).
www.igi-global.com/chapter/peer-peer-network-based-image/73145?camid=4v1a

Medium Access Protocols for Cooperative Collision Avoidance in Vehicular Ad-Hoc Networks
www.igi-global.com/chapter/medium-access-protocols-cooperative-collision/43167?camid=4v1a