

Detecting Overlapping Communities in Two-mode Data Networks using Formal Concept Analysis

Abir Messaoudi, Rokia Missaoui
Mohamed-Hamza Ibrahim

LARIM, Université du Québec en Outaouais, Québec, Canada
{mesa08,missro01,ibrm05}@uqo.ca

Abstract. Social networks frequently feature complex structures such as two-mode data expressed by bipartite graphs. Most research work on community detection in bipartite graphs focus on either finding non-overlapping communities or identifying overlapping ones by first projecting two-mode (bi-dimensional) data into two one-mode tables which are further analyzed. However, this often leads to a loss of information and produces inaccurate communities. Therefore, efficiently detecting communities in such two-mode data networks often remains a key challenge in social network analysis. In this paper, we introduce a novel three-step strategy to detect overlapping as well as hierarchically nested communities in bipartite graphs. First, we extract the formal concepts that represent potential groups in the social network. Then, we rank and filter the obtained groups to keep only core ones that have a high mean of stability and separation. Finally, we detect communities by refining the core groups using a Silhouette Analysis. Our experiments on real-world social networks show that our method can accurately identify overlapping communities.

1 Introduction

Although numerous community detection methods have been proposed, relatively few ones are designed for heterogeneous or multi-layer networks or even two-mode data ones (*i.e.*, two types of nodes and one type of links). The main focus is generally on homogeneous (commonly called one-mode data) networks, which are extracted from real-world networks by considering one type of nodes (objects) and one type of links. Moreover, most of the studies are concerned with disjoint communities rather than overlapping and hierarchically nested ones which are very common in real-life situations and applications. For the reasons cited above, the number of research studies on identifying overlapping and nested communities in heterogeneous networks is growing up. This is an issue we seek to address in the following paper. Our method is completely unsupervised and can automatically determine both the number of communities and their description. It relies on the inherent structure discovered from data using Formal Concept Analysis (FCA), and two relevancy metrics associated with formal concepts, namely stability and separation. Finally, it makes use of Silhouette analysis to refine the process of community detection.