Clustering in Feature-Rich Networks Using Data Recovery Approach

Dr. Soroosh Shalileh (PhD in Computer Science)
Center for Language and Brain, HSE University, Myasnitskaya Ulitsa, 20, Moscow, 101000, Russia

Abstract

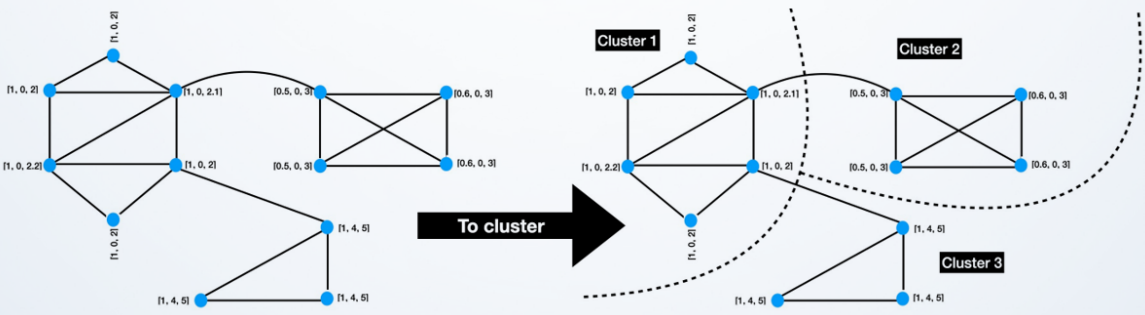
A feature-rich network is a network whose nodes are associated with categorical or quantitative features. This talk reviews our eight recently proposed cluster detection methods in feature-rich networks.

We categorize our methods based on: (A) (1) whether the features are used directly, or (2) they are converted to similarity; (B) (1) whether network/similarity data are comparable across the entire data table (summability mode), or (2) network/similarity data are comparable only within individual columns (nonsummability mode), and (C) (1) clusters are extracted sequentially, or (2) all clusters are extracted simultaneously.

Methods in category (C-1) are rather time-consuming, but they allow for automated derivation of the number of clusters. In contrast, methods of the category (C-2)[1] are fast analogue to the k-means clustering method to be run in the (B-2) mode [2]. The latter method, however, works in the joint node/feature space and, thus, may be subject to the so- called dimensionality curse; to tackle this, we apply, on par with, the squared Euclidean distance, the so-called Cosine distance.

We compare and evaluate our proposed methods' performance with the SOTA algorithms over 800 synthetic and eight real-world datasets. Our experiments with real-world and synthetic datasets show that these algorithms are valid and competitive.

1. Shalileh, S. and Mirkin, B., 2022. Community Partitioning over Feature-Rich Networks Using an Extended K-Means Method. Entropy, 24(5), p.626.
2. Shalileh, S. and Mirkin, B., 2021. Least-squares community extraction in feature-rich networks using similarity data. Plos one, 16(7), p.e0254377.



Biography

Dr Soroosh Shalileh received his PhD. in Computer Science, from NRU HSE University, Moscow RF. Currently, he is working as an AI researcher and the head of the computation group in the center for language and brain NRU HSE University, Moscow RF.