**Semisupervised Clustering via Community Detection Algorithms**

Luca Frigau, Giulia Contu, Claudio Conversano, Francesco Mola
University of Cagliari, Italy

**Abstract:**

Semisupervised clustering extends standard clustering methods to the semisupervised setting, in some cases considering situations when clusters are associated with a given outcome variable that acts as a "noisy surrogate" that is a good proxy of the unknown clustering structure. In this article, a novel approach to semisupervised clustering associated with an outcome variable named network-based semisupervised clustering (NeSSC) is introduced. It combines an initialization, a training and an agglomeration phase. In the initialization and training a matrix of pairwise affinity of the instances is estimated by a classifier. In the agglomeration phase the matrix of pairwise affinity is transformed into a complex network, in which a community detection algorithm searches the underlying community structure. Thus, a partition of the instances into clusters highly homogeneous in terms of the outcome is obtained. We consider a particular specification of NeSSC that uses classification or regression trees as classifiers and the Louvain, Label propagation and Walktrap as possible community detection algorithm. NeSSC's stopping criterion and the choice of the optimal partition of the original data are also discussed. Several applications on both real and simulated data are presented to demonstrate the effectiveness of the proposed semisupervised clustering method and the benefits it provides in terms of improved interpretability of results with respect to three alternative semisupervised clustering methods.

**Keywords:** Community detection; complex networks; label propagation; Louvain; tree-based classifiers; Walktrap.