Statistica Sinica **35** (2025), 1-26
doi:https://doi.org/10.5705/ss.202023.0081

A PSEUDO-LIKELIHOOD APPROACH TO COMMUNITY DETECTION IN WEIGHTED NETWORKS

Andressa Cerqueira* and Elizaveta Levina

Federal University of São Carlos and University of Michigan

Abstract: Community structure is common in many real networks, with nodes clustered in groups sharing the same connections patterns. While many community detection methods have been developed for networks with binary edges, few of them are applicable to networks with weighted edges, which are common in practice. We propose a pseudo-likelihood community estimation algorithm derived under the weighted stochastic block model for networks with normally distributed edge weights, extending the pseudo-likelihood algorithm for binary networks, which offers some of the best combinations of accuracy and computational efficiency. We prove that the estimates obtained by the proposed method are consistent under the assumption of homogeneous networks, a weighted analogue of the planted partition model, and show that they work well in practice for both homogeneous and heterogeneous networks. We illustrate the method on simulated networks and on a fMRI dataset, where edge weights represent connectivity between brain regions and are expected to be close to normal in distribution by construction.

Key words and phrases: Community detection, pseudo-likelihood, weighted networks.

1. Introduction

Network models have been a useful general tool for understanding and modeling interactions between objects in many domains. The relationships (edges) between objects (nodes) can represent many things depending on the application: social interactions, trading partnerships, web links, packets sent between computers, disease contagion, neural connectivity, and so on. Some settings result in binary networks, where only the presence or absence of an edge is recorded, and other settings lead to weighted networks, where an edge is associated with a weight which typically quantifies the strength of the connection.

The probabilistic modeling of networks has traditionally focused on binary networks, starting from the classical Erdös-Rényi graph (Erdös and Rényi, 1960) which models edges as i.i.d. Bernoulli random variables. Yet many networks encountered in practice are weighted, and many of the binary networks in the literature are obtained by thresholding raw edge weights. For instance, the arguably most studied network with communities, the karate club dataset

*Corresponding author. E-mail: acerqueira@ufscar.br