Statistica Sinica **35** (2025), 1-24
doi:https://doi.org/10.5705/ss.202023.0061

# ORTHOGONAL SYMMETRIC NON-NEGATIVE MATRIX FACTORIZATION UNDER THE STOCHASTIC BLOCK MODEL

Subhadeep Paul and Yuguo Chen*

*The Ohio State University and University of Illinois Urbana-Champaign*

*Abstract:* We present a method based on the Orthogonal Symmetric Non-negative matrix Tri-Factorization (OSNTF) of the adjacency and the normalized Laplacian matrices for community detection in networks. We establish the connection of the factors obtained through this factorization to a non-negative basis of an invariant subspace of the approximating matrix, drawing parallel with the spectral clustering. Since the exact OSNTF may not exist or may not be computable for a given matrix like many non-negative matrix factorization methods, we study the approximate OSNTF that solves an optimization problem. We show that the global optimizer of the OSNTF objective function is consistent for community detection in networks generated from the stochastic block model as well as its degree corrected version. We compare the method with several state-of-the-art methods for community detection, including regularized spectral clustering, SCORE and SCOREplus, and spectral clustering followed by likelihood-based refinement, in both simulations and real datasets with known ground truth community assignments. These results show the excellent performance of the OSNTF under a wide variety of simulation setups and for real datasets obtained from disparate fields.

*Key words and phrases:* Community detection, degree corrected stochastic block model, invariant subspace, network data, non-negative matrix factorization.

## 1. Introduction

Over the last two decades, there has been a surge in interest in the statistical inference of network data motivated by their applications in information sciences, biology, social sciences, and economics. A network consists of a set of entities called nodes or vertices and connections among them called edges or relations. The problem of community detection in networks has received considerable attention in the literature. A community is often defined as a group of nodes that are more "structurally similar" to each other than the rest of the network. Therefore nodes that belong to a community have similar patterns of connection to the rest of the network.

Several methods have been proposed in the literature for the efficient detection of network communities. These methods include modularity maximization

*Corresponding author. E-mail: yuguo@illinois.edu