

# SCALABLE COMMUNITY DETECTION IN MASSIVE NETWORKS USING AGGREGATED RELATIONAL DATA

Timothy Jones<sup>1</sup>, Owen G. Ward<sup>2</sup>, Yiran Jiang<sup>3</sup>  
John Paisley<sup>1</sup> and Tian Zheng<sup>\*1</sup>

<sup>1</sup>*Columbia University*, <sup>2</sup>*Simon Fraser University* and  
<sup>3</sup>*Purdue University*

*Abstract:* The mixed membership stochastic blockmodel (MMSB) is a popular Bayesian network model for community detection. Fitting such large Bayesian network models quickly becomes computationally infeasible when the number of nodes grows into hundreds of thousands and millions. In this paper we propose a novel mini-batch strategy based on aggregated relational data that leverages nodal information to fit MMSB to massive networks. We describe a scalable inference method that can utilise nodal information that often accompanies real-world networks. Conditioning on this extra information leads to a model that admits a parallel stochastic variational inference algorithm, utilising stochastic gradients of bipartite graphs formed from aggregated network ties between node subpopulations. We apply our method to a citation network with over two million nodes and 25 million edges, capturing explainable structure in this network. Our method recovers parameters and achieves better convergence on simulated networks generated according to the MMSB.

*Key words and phrases:* Aggregated relational data, community detection, mixed-membership, network data.

## 1. Introduction

Relational data between objects is commonly represented by a graph or network that encodes pairwise interactions, and has been studied across the natural and social sciences. Due to this prevalence in modern applications, analysis of such data is vitally important. Among the tasks commonly considered for such data, community detection stands out as being one of the most crucial for practitioners and has been widely studied. Community detection algorithms aim to identify groups of nodes that exhibit similar connective behaviors. More specifically, nodes in networks often cluster into small communities, where nodes within a community show a similar propensity to form ties with other nodes (Bickel and Chen, 2009). These clusters are often assortative, where nodes within a community interact more than those in different communities (Fortunato and Hric, 2016).

---

<sup>\*</sup>Corresponding author. E-mail: [tian.zheng@columbia.edu](mailto:tian.zheng@columbia.edu)