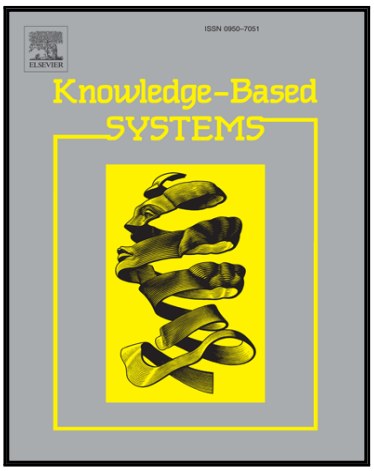

Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach

Wenjun Wang, Pengfei Jiao, Dongxiao He, Di Jin, Lin Pan, Bogdan Gabrys

PII: S0950-7051(16)30235-0
DOI: 10.1016/j.knosys.2016.07.021
Reference: KNOSYS 3609

To appear in: *Knowledge-Based Systems*

Received date: 10 December 2015
Revised date: 11 July 2016
Accepted date: 15 July 2016

Please cite this article as: Wenjun Wang, Pengfei Jiao, Dongxiao He, Di Jin, Lin Pan, Bogdan Gabrys, Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach, *Knowledge-Based Systems* (2016), doi: 10.1016/j.knosys.2016.07.021

Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach

Wenjun Wang[a], Pengfei Jiao[a], Dongxiao He[*,a], Di Jin[a], Lin Pan[a], Bogdan Gabrys[b]

[a]*School of Computer Science and Technology, Tianjin University, Tianjin 300072, China*
[b]*Data Science Institute, Bournemouth University, Poole House, Talbot Campus, Fern Barrow Poole, Dorset, BH12 5BB, UK*

Abstract

A wide variety of natural or artificial systems can be modeled as time-varying or temporal networks. To understand the structural and functional properties of these time-varying networked systems, it is desirable to detect and analyze the evolving community structure. In temporal networks, the identified communities should reflect the current snapshot network, and at the same time be similar to the communities identified in history or say the previous snapshot networks. Most of the existing approaches assume that the number of communities is known or can be obtained by some heuristic methods. This is unsuitable and complicated for most real world networks, especially temporal networks. In this paper, we propose a Bayesian probabilistic model, named Dynamic Bayesian Nonnegative Matrix Factorization (DBNMF), for automatic detection of overlapping communities in temporal networks. Our model can not only give the overlapping community structure based on the probabilistic memberships of nodes in each snapshot network but also automatically determines the number of communities in each snapshot network based on automatic relevance determination. Thereafter, a gradient descent algorithm is proposed to optimize the objective function of our DBNMF model. The experimental results using both synthetic datasets and real-world temporal networks demonstrate that the DBNMF model has superior performance compared with two widely used methods, especially when the number of communities is unknown and when the network is highly sparse.

Key words: community detection, temporal networks, Bayesian nonnegative matrix factorization, gradient descent, model selection.

1. Introduction

Complex networks, such as social networks, biological networks and information networks, are very common in real life. The analysis of complex networks has been becoming more and more

*Corresponding author.
Email address: hedongxiao@tju.edu.cn (Dongxiao He)