

MASTER'S THESIS

Overlapping community detection exploiting direct dependency structures in complex networks

Liang, Fengfeng

Date of Award:
2017

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

Many important applications in the social, ecological, epidemiological, and biological sciences can be modeled as complex systems in which a node or variable interacts with another via the edges in the network. Community detection has been known to be important in obtaining insights into the network structure characteristics of these complex systems. The existing community detection methods often assume that the pairwise interaction data between nodes are already available, and they simply apply the detection algorithms to the network. However, the predefined network might contain inaccurate structures as a result of indirect effects that stem from the nodes' high-order interactions, which poses challenges for the algorithms upon which they are built. Meanwhile, existing methods to infer the direct interaction relationships suffer from the difficulty in identifying the cut point value that differentiates the direct interactions from the indirect interactions. In this thesis, we consider the overlapping community detection problem with determination and integration of the structural information of direct dependency interactions. We propose a new overlapping community detection model, named direct-dependency-based nonnegative matrix factorization (DNMF), that exploits the Bayesian framework for pairwise ordering to incorporate the structural information of the underlying network. To evaluate the effectiveness and efficiency of the proposed method, we compare it with state-of-the-art methods on benchmark datasets collected from different domains. Our empirical results show that after the incorporation of a direct dependency network, significant improvement is seen in the community detection performance in networks with homophilic effects.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
List of Symbols	ix
Chapter 1 Introduction	1
1.1 Motivation	1
1.1.1 The Importance of Community Detection	1
1.1.2 The Inaccurate Structures in Predefined Networks	2
1.1.3 The Limitations of Direct Interaction Inference	4
1.2 Main Contributions	6
1.3 Outline of the Thesis	7
Chapter 2 Literature Review	8
2.1 Community Structures and Properties	8
2.1.1 Community Structures	8
2.1.2 Properties of Community Structures	9

2.2	Community Detection	12
2.2.1	Hierarchical Clustering	12
2.2.2	Graph Partitioning Methods	13
2.2.3	Spectral Methods	14
2.3	Overlapping Community Detection	16
2.3.1	Clique Percolation Model	16
2.3.2	Nonnegative Matrix Factorization Model	17
2.4	Direct Interaction Inference	19
2.4.1	Correlation-based Approach	19
2.4.2	Information-Theoretic-based Approach	20
2.4.3	Tie-Strength-based Approach	21
2.4.4	Global Approach	22
Chapter 3 Problem Statement		25
3.1	Problem Definition	25
3.1.1	Direct Dependency Structures	25
3.1.2	Community Detection	28
3.2	Key Observations	30
Chapter 4 Overlapping Community Detection Exploiting Direct De- pendency Structures		33
4.1	Preliminary	33
4.2	Model Formulation	35
4.2.1	Modeling Direct Dependency Structures	35
4.2.2	Model Assumptions	37
4.2.3	Objective Function	38
4.3	Parameter Learning	43
Chapter 5 Experiments		46
5.1	Evaluation on Existing Benchmark Social Networks	46
5.2	Comparison Study with State-of-the-art Methods	49

5.2.1	Evaluation Metric	50
5.2.2	Comparison Results	50
5.2.3	Discussion	53
Chapter 6 Conclusions and Future Work		54
6.1	Conclusions	54
6.2	Future Work	55
Bibliography		57
Curriculum Vitae		