論文 / 著書情報
Article / Book Information

題目(和文)	ソーシャル・ネットワーク分析に関する研究: サンプリングと高次の相互作用に焦点を当てて
Title(English)	Studies on Social Network Analysis: Sampling and Higher-order Interactions
著者(和文)	中嶋一貴
Author(English)	Kazuki Nakajima
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第12170号, 授与年月日:2022年9月22日, 学位の種別:課程博士, 審査員:三好 直人,高安 美佐子,南出 靖彦,脇田 建,村田 剛志,首藤 一幸
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12170号, Conferred date:2022/9/22, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	数理・計算科学 系 数理・計算科学 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of （ 理学 ）
学生氏名： Student's Name	中嶋 一貴	指導教員（主）： Academic Supervisor(main)	三好 直人
		指導教員（副）： Academic Supervisor(sub)	首藤 一幸

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

In this thesis, I present four works of my contributions to the field of social network analysis to highlight the opportunities to more understand the structure and dynamics of social interactions. The first two works to be presented in Chapters 2 and 3 are contributions to the subfield of the measurement of structural properties of huge online social networks. The last two works to be presented in Chapters 4 and 5 are contributions to the subfield of analyses of social networks involving higher-order interactions.

In Chapter 2, we aim to present a practical framework for estimating the property of interest based on a random walk on a social network involving private nodes. Most existing random walk-based algorithms assume a social network in which each node publishes its all neighbors if the node is queried. However, a certain percentage of private nodes, that do not publish their own neighbors' data when they are queried, is present in practical scenarios. In this work, we develop a sampling algorithm by extending a simple random walk to the case of a social network involving private nodes, and then, we propose estimators with reduced biases induced by private nodes for a few network properties. Our experimental results show that the proposed estimators reduce biases induced by private nodes in the existing estimators by up to 92.6% on empirical social network datasets involving private nodes.

In Chapter 3, we introduce the social graph restoration problem. A number of random-walk-based algorithms that estimate structural properties using a small number of queries have been developed over the last decade. However, most of the existing algorithms enable analysts or researchers only to estimate local structural properties in principle. On the other hand, analysts' or researchers' interests in the characteristics of social networks are generally diverse; these characteristics include local structural properties, global structural properties, and visual graph representations. To address this gap, we propose a method for restoring the original social graph from its small sample obtained by a random walk. The proposed method generates a graph that preserves the estimates of local structural properties and the structure of the subgraph sampled by a random walk. Our experimental results show that the proposed method more accurately reproduces the local and global structural properties on average and the visual representation of the original graph than existing methods.

In Chapter 4, we aim to develop randomized baseline models for investigating the structure and dynamics of empirical social networks involving higher-order interactions. Such networks can be represented by hypergraphs, in which hyperedges encode higher-order interactions among an arbitrary number of nodes. To analyze structures and dynamics of given hypergraphs, a solid practice is to compare them with those for randomized hypergraphs that preserve some specific properties of the original hypergraphs. Here we propose a family of such reference models for hypergraphs, called the hyper dK-series, by extending the so-called dK-series for dyadic networks to the case of hypergraphs. The hyper dK-series preserves up to the individual node's degree, node's degree correlation, node's redundancy coefficient, and/or the hyperedge's size depending on the parameter values. Furthermore, we numerically find that higher-order hyper dK-series more accurately preserves the shortest path length and degree distribution of the one-mode projection of the original hypergraph, which the method does not intend to preserve. We also apply the hyper dK-series to numerical simulations of epidemic spreading and evolutionary game dynamics on empirical social hypergraphs.

In Chapter 5, we aim to reveal the properties of collaborations between two or more institutions in research grants and the impact of those collaborations on research outputs. The reliance on teamwork in scientific work has increased over the past decades. Funded research projects are often collaborative between institutions, and leading institutions tend to be densely connected to each other, which is known as the rich-club phenomenon in networks of research grants. While it is not uncommon that more than two institutions to collaborate on one funded project, properties of such higher-order collaborations in research grants are little known. In this work, we investigate higher-order rich-club phenomena in collaborative research grants among funded institutions and the impact of rich clubs on research outputs. Using publicly available data from the National Science Foundation, we constructed a bipartite network of funded institutions and collaborative grants. We find rich clubs both in the entire network and the subnetwork induced by the collaborative grants involving a given number of institutions from two to five. Furthermore, our analyses using a randomized baseline indicate that the collaborative grants within rich clubs tend to achieve higher research productivity than the average research productivity for those member institutions.

In Chapter 6, I conclude the thesis. I discuss the position of this thesis in the field of social network analysis and future possible directions.