

氏 名 三戸 圭史

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2679 号

学位授与の日付 2026 年 3 月 24 日

学位授与の要件 先端学術院先端学術専攻 (統計科学コース)  
学位規則第6条第1項該当

学位論文題目 Incorporating Graph Domain Structure into Matrix  
Representations and Optimal-Transport Kernels

論文審査委員 主 査 間野 修平  
統計科学コース 教授  
日野 英逸  
統計科学コース 教授  
相馬 輔  
統計科学コース 准教授  
原 聡  
電気通信大学 大学院情報理工学研究科 教授

# 博士論文の要旨

氏 名：三戸 圭史

論文題目：Incorporating Graph Domain Structure into Matrix Representations and Optimal-Transport Kernels

Graph-structured data provides a powerful way to model entities and their relations in diverse domains, from social and information networks to molecules and biological systems. However, the non-Euclidean nature of graphs, such as the variability in the number of nodes, the absence of canonical node ordering, and the presence of asymmetric interactions, makes it difficult to apply standard machine learning techniques. As the scale and complexity of data continue to grow with the advancement of the information society, three practically serious challenges arise. First, as graph data becomes larger, it becomes harder to grasp their latent structures, necessitating appropriate visualization and metrics. Second, with the spread of the internet, graph structures that concentrate popularity and demand on a few nodes, called hubs, have become more prevalent, requiring methods that can make robust inferences even on such heterogeneous graphs. Third, in addition to the ability to identify complex graph structures, methods that are scalable to data size are required. This dissertation addresses these challenges through two topics.

First, to robustly capture the underlying structure even in the presence of hub nodes, we propose a new matrix representation that incorporates key properties of the Hermitian adjacency matrix and the non-backtracking matrix, and develop a clustering method that leverages their eigenvalue and eigenvector information. The Hermitian adjacency matrix has recently been introduced to mitigate the difficulties of handling adjacency matrices for directed graphs, by encoding edge directions using complex numbers and their conjugates. While community detection methods based on the Hermitian adjacency matrix have been shown to outperform conventional approaches for directed graphs, their performance degrades in sparse graphs, where the number of edges is small. In contrast, in sparse undirected graphs, it is known that leveraging the eigenvalue information of the non-backtracking matrix enables robust community detection. In this study, by combining the features of these two matrix representations, we propose a method that allows for robust community detection even in sparse directed graphs. Our formulation constitutes a natural extension in the sense that the relationship established between the adjacency matrix and the non-backtracking matrix in undirected graphs also holds between the Hermitian adjacency matrix and our proposed representation in directed graphs. Through numerical experiments on

community detection tasks, we confirm that our method demonstrates robust performance compared to existing methods, especially in heterogeneous graphs with hubs. Furthermore, by investigating its connection to belief propagation, we obtain insights into the community structures that can be detected in sparse settings.

Second, with the aim of achieving both discriminative power and scalability for regression and classification on graph data, we propose an acceleration of the Categorical Wasserstein Weisfeiler-Lehman (WWL) kernel, a recently proposed graph kernel that leverages optimal transport. Problems such as predicting physicochemical properties of chemical compounds represented as graphs (e.g., toxicity and efficacy) and attribute prediction in social networks centered on individuals are commonly formulated as regression and classification tasks within the framework of kernel methods, after designing a kernel function for graph data. Although graph kernels based on optimal transport have been reported to be highly discriminative, they suffer from very high computational costs and thus do not scale with increasing graph size or with the number of graphs in a dataset. In this work, we focus on the Categorical WWL graph kernel, one of the optimal transport-based graph kernels, and clarify the underlying structure of the features it employs. By leveraging this structure, we propose a fast and accurate computational algorithm that does not require approximations. While conventional algorithms require quadratic computational cost with respect to the number of graph nodes, we theoretically show that our proposed algorithm can compute the kernel in linear time. Moreover, through numerical experiments, we confirm that our algorithm scales well with both graph size and the number of graphs in a dataset, and achieves performance that is comparable to or better than existing methods.

This dissertation is organized as follows. Chapter 1 reviews prior work on graph data analysis and positions the present study within the literature. Chapter 2 introduces the notation and provides preliminaries on related methods for handling graph data. In Chapter 3, we introduce a new matrix representation for directed graphs that incorporates properties of the Hermitian adjacency matrix and the non-backtracking matrix, and propose a clustering method based on it. We discuss not only its relationship with existing matrix representations but also its connection to belief propagation. Chapter 4 proposes an acceleration method for the Categorical WWL kernel that is a graph kernel based on optimal transport. By focusing on latent relationships among the features used in the Categorical WWL kernel, we show that the entire feature set can be equivalently represented as a tree structure, and demonstrate that this enables fast and exact computation. Finally, Chapter 5 summarizes the dissertation and outlines directions for future work.

Results of the Doctoral Thesis Defense

博士論文審査結果

Name in Full

氏名 三戸 圭史

Title

論文題目 Incorporating Graph Domain Structure into Matrix Representations and Optimal-Transport Kernels

出願者は、現代の大規模かつ複雑なグラフデータにおける構造抽出および類似度評価の高度化を目的として、(i) 有向グラフに対する複素非バックトラック行列 (Complex Non-Backtracking Matrix; CNBT) の提案と、クラスタリングへの応用、及び (ii) Categorical Wasserstein Weisfeiler-Lehman (WWL) グラフカーネルの計算の高速化の、互いに関連はあるが独立したテーマを研究した。

提出された論文は英文で執筆されており、全 5 章・96 頁から構成されている。第 1 章では、背景の説明として、グラフデータ解析に共通した困難さ (頂点の次数の偏り・有向辺の非対称性・非自明な計量など) を整理し、グラフを特徴づける行列のスペクトルに基づく解析方法と、グラフ同士の類似性の尺度であるグラフカーネルに焦点を当てて、本研究の問題意識を提示している。前者に関しては、従来の手法では、有向グラフの非対称性やスパース性、ハブの存在によりクラスタリングの性能が著しく劣化すること、後者に関しては、大規模グラフ同士のカーネルの計算に要するコストの高さを指摘している。これらに基づき、本論文で扱う 2 つのテーマを明確に説明している。

第 2 章は準備である。スペクトルグラフ理論、スペクトルクラスタリング、非バックトラック行列、確率伝播法、最適輸送、グラフカーネル、最適輸送と Sinkhorn アルゴリズムなど、本論文に現れる概念を丁寧に整理している。特に、1 章で述べられた問題意識に対応して、非バックトラック行列やエルミート隣接行列の導入の動機が説明されている。

第 3 章では、本論文の第一の貢献である CNBT 行列を提案している。CNBT 行列は、有向辺の向きを複素位相で表し、グラフ上のパスにおける有向辺の向きにより定義される回転と呼ぶ概念を導入することで定義された、有向グラフの新しい行列表現である。エルミート隣接行列との関係をグラフゼータの伊原表示と橋本表示の関係として示し、行列の固有ベクトルを頂点ごとに定義されるベクトルに還元する in/out-vector と呼ぶ操作を通じて、CNBT 行列に基づくクラスタリングのアルゴリズムを提案し、スパースな有向グラフに対しても従来法よりも高いクラスタの再現性を示すことを数値実験によって実証している。さらに、CNBT 行列と線形化された確率伝播法との関連付けについても議論され、CNBT 行列の発展的応用への理論的洞察が与えられている。

第 4 章では、本論文の第二の貢献として、Categorical WWL グラフカーネルの計算の高速化が議論されている。Categorical WWL グラフカーネルの定義における Weisfeiler-Lehman のラベル付けを木構造で表示し、グラフカーネルの計算に現れる

Wasserstein 距離が、効率的に計算できる木の上で定義された Wasserstein 距離と等しいことを示すことで、グラフカーネルを厳密かつ効率的に計算する新手法を提案している。提案手法により、グラフカーネルの計算に現れる計算を単体法により厳密に解いた場合と同じ結果が、単体法や近似解法である正則化 Sinkhorn アルゴリズムによる計算よりも高速に得られることを、実データ解析により確認している。

第 5 章は総括であり、本論文における提案手法の意義と、今後の展望について述べられている。

本論文は、有向グラフのクラスタリングに向けた新たなグラフの行列表現の構築と理論的整理、及び最適輸送に基づくグラフカーネルの計算の高速化という、グラフデータ解析における二つの重要課題に対して新しい方法論を提示している。CNBT 行列の提案は本論文の新規性として高く評価でき、グラフゼータや確率伝搬法に関連する理論的考察も興味深い。さらにクラスタリングのアルゴリズムの提案も、グラフデータ解析の新しい手法を提供したのものとして評価できる。加えて、スパースかつハブを含むような現実的なグラフに対して既存手法よりも高いクラスタリング性能を示すことを実証しており、実用性も十分認められる。Categorical WWL グラフカーネルの高速化において、木構造の活用により計算量を劇的に削減した点は、大規模データセットへの適用を可能にしたという点で重要な意義を持ち、グラフデータ解析における大きな貢献であると認められる。

全体として、理論的整合性が丁寧に議論されており、アルゴリズム設計・理論的性質の証明・数値実験が適切に行われている。特にスパースな有向グラフにおけるクラスタリングの性能向上と categorical WWL kernel の正確かつ高速な実装は、グラフデータの機械学習分野における今後の応用と展開を期待させる成果といえる。第 3 章、第 4 章の内容は、それぞれ、査読付き国際論文誌 *Journal of Complex Networks* 及び *Transactions on Machine Learning Research* にいずれも三戸氏が筆頭著者として採択されている。

以上の理由により、審査委員会は、本論文が統計科学・機械学習分野における学位の授与に値すると判断した。