

## 二部グラフのクラスタリング手法の比較実験

学習院大学 計算機センター 久保山 哲 二  
学習院大学 計算機センター 城 所 弘 泰  
学習院大学 計算機センター 磯 上 貞 雄  
学習院大学 計算機センター 村 上 登志男

### 1 研究の概要

現実世界のデータは、語と文書、購買者と商品、ツイートとハッシュタグなど、2種類の頂点からなる2部グラフによって表現できる対象が数多く存在する。2部グラフを対象とした知識発見のためのデータ分析手法は、様々な実データの分析で必要とされている。その要素技術として、2部グラフ中で辺密度の高い部分構造を抽出する2部グラフの構造的クラスタリングは、異なる研究分野において異なるアプローチで提案されてきた。たとえば、2部グラフから、辺密度の高い構造を抽出するために、2部クリーク状の構造を列挙する手法が提案されている。一方で、同じデータは2部グラフの隣接行列として表現でき、特異値分解や非負値行列分解等の行列分解を用いて少数の隠れ頂点を中間に持つ2つの2部グラフとして2頂点間の関係を近似する手法が多数提案されている。同種の問題はそのほとんどがNP困難であることが知られている。

本研究では、2部クラスタリングの様々な手法(Newman-Girvan モジュラリティの概念を元に開発されてきたクラスタリング手法、2部クリーク抽出や交差辺最小化問題として開発されてきた手法、行列分解として開発されてきた手法など)を比較し、現実のデータに適用することで、その有効性を比較検討した。

### 2 報告

本研究に関連して、2部グラフを用いたTwitterからのトピック抽出に関する共同研究を行った。これらは、以下の国際会議ワークショップで採択され報告済みである。

1. Takako Hashimoto, Hiroshi Okamoto, Tetsuji Kuboyama, Kilho Shin: Topic life cycle extraction from big Twitter data based on community detection in bipartite networks. BigData 2017: 2740-2745
2. Takako Hashimoto, Tetsuji Kuboyama, Hiroshi Okamoto, Kilho Shin: Topic Extraction on Twitter Considering Author's Role Based on Bipartite Networks. DS 2017: 239-247
3. Takako Hashimoto, Tetsuji Kuboyama, Hiroshi Okamoto, Kilho Shin: Topic Extraction from Millions of Tweets Based on Community Detection in Bipartite Networks. EJC 2017: 395-408