

複素数による有向グラフの行列表現とその応用

三戸圭史 総合研究大学院大学 統計科学コース 博士後期課程 4年

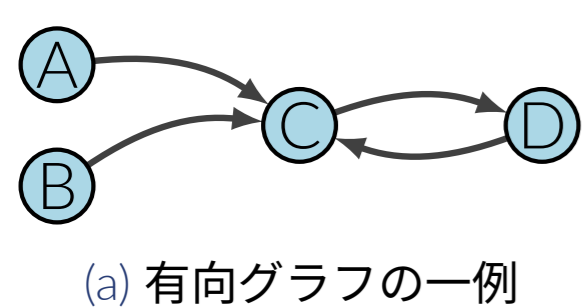
背景：有向グラフの行列表現

無向グラフを対象としたデータ解析手法の研究は古くから行われている一方、有向グラフに対しては単純な対称化を行った上で無向グラフの手法を適用する取り組みや大規模化が困難なベイズ推論によるアプローチが主な手法であった。本研究では、有向情報を保持しつつ利便性の高い行列表現による解析手法を開発し、その有効性を示すことを目的とする。

エルミート隣接行列

隣接行列 $A \in \mathbb{R}^{N \times N}$ に代わる有向グラフの行列表現として、複素数 $\alpha \in \mathbb{C}$ を用いたエルミート隣接行列 $\mathcal{A}_\alpha \in \mathbb{C}^{N \times N}$ を以下で定める。

$$(\mathcal{A}_\alpha)_{uv} := \begin{cases} 1 & u \rightsquigarrow v, \\ \alpha & u \rightsquigarrow v, \\ \bar{\alpha} & u \leftarrow v. \end{cases}$$



(a) 有向グラフの一例

$$\begin{array}{c} \begin{matrix} & A & B & C & D \\ A & 0 & 0 & 1 & 0 \\ B & 0 & 0 & 1 & 0 \\ C & 0 & 0 & 0 & 1 \\ D & 0 & 0 & 1 & 0 \end{matrix} \\ \text{(b) 隣接行列} \end{array} \quad \begin{array}{c} \begin{matrix} & A & B & C & D \\ A & 0 & 0 & i & 0 \\ B & 0 & 0 & i & 0 \\ C & -i & -i & 0 & 1 \\ D & 0 & 0 & 1 & 0 \end{matrix} \\ \text{(c) エルミート隣接行列} \end{array}$$

エルミート隣接行列の性質

- エルミート対称ゆえに固有値が実数
- \mathcal{A}_α から作られるエルミートグラフラプラシアン L_α は半正定値
- エルミート正規化ラプラシアン \mathcal{L}_α の固有値について $\forall i = 1 \dots N, 0 \leq \lambda_i \leq 2$.
- 重み付き有向グラフ $G = (V, E, w)$ において、 $|\alpha| = 1$ とすると、 L_α の二次形式は無向グラフの場合の自然な拡張となる

$$\mathbf{x} \in \mathbb{C}^N, \mathbf{x}^* L_\alpha \mathbf{x} = \sum_{u \rightarrow v \in E} w(u, v) |x_u - h(u, v)x_v|^2,$$
$$\text{where } h(u, v) := \begin{cases} 1 & u \rightsquigarrow v, \\ \alpha & u \rightsquigarrow v, \\ \bar{\alpha} & u \leftarrow v. \end{cases}$$

隣接行列の複素表現は発展途上ゆえ、様々な表現が提案されている

$$(\mathcal{A}_\alpha)_{uv} := w(u, v)\alpha + w(v, u)\bar{\alpha},$$

$$(L_q)_{uv} := \delta_{uv}d(u, u) - \frac{1}{2}e^{i2\pi q(w(u,v)-w(v,u))}(w(u, v) - w(v, u)).$$

グラフ信号処理への応用

L_α の対称性と固有値が実数になる性質から、無向グラフでのグラフフーリエ変換を有向グラフへ自然に拡張でき、グラフ上の信号のdenoising等も提案されている。

ローパスフィルタによるdenoising [1]

low-pass filter kernel を $\hat{h}(\lambda; c) := \frac{1}{1+c\lambda}$ としてノイズのあるグラフ信号 $\mathbf{g} \in \mathbb{R}^N$ から下記によって信号 \mathbf{f} を復元する。

$$\tilde{\mathbf{f}} = U\hat{H}U^*\mathbf{g}, \text{ where } L_\alpha = U\Lambda U^*.$$

対称化により無向化したグラフでのdenoised信号では単に隣接ノード間で平滑化した結果になる一方、エルミート行列でのdenoisingでは誤差が同等かそれ以下になり、特にsink/source側のノードでの誤差の改善が、数値実験的に確認されている。

クラスタリングへの応用

無向グラフと有向グラフでは抽出したいクラスタ構造が異なる場合があり、対称化 ($A + A^T, AA^T + A^T A$, etc...) によるアプローチでは有向グラフ特有の構造を捉えられない場合もある

無向グラフ クラスタ間で疎結合、内部で密結合する分割

有向グラフ 流入元や流出先が類似する分割

⇒ \mathcal{L}_α を用いたスペクトルクラスタリングでは、 \mathcal{L}_α の複素固有ベクトルでノード埋め込みをしてクラスタリング

\mathcal{L}_α での有向グラフクラスタリングの性能 [2]

$$\Phi(\{S_k\}_{k=0}^{K-1}) := \sum_{k=1}^{K-1} \frac{w(S_k, S_{k-1})}{\text{vol}(S_k) + \text{vol}(S_{k-1})}, \theta_K := \max_{\{S_k\}_{k=0}^{K-1}} \Phi(\{S_k\}_{k=0}^{K-1})$$

上記の分割指標“flow ratio” $\Phi(\{S_k\})$ を最大化する分割を $\{S_k\}$ とすると、 \mathcal{L}_α の最小固有値・固有ベクトル λ_1, \mathbf{f}_1 では以下が成立する。

- ある値 $a \in \mathbb{C}$ が存在し次を満たす： $\|\mathbf{y} - a\mathbf{f}_1\|^2 \leq \frac{1}{\gamma_K}$
- ある値 $b \in \mathbb{C}$ が存在し次を満たす： $\|\mathbf{f}_1 - b\mathbf{y}\|^2 \leq \frac{1}{\gamma_{K-1}}$

ここで $\mathbf{y} \in \mathbb{C}^N$ は θ_K を達成する K 分割を表現する指示ベクトル、 $\gamma_K := \frac{\lambda_2}{1-(4/K)\theta_K}$ とする。

⇐ 固有ベクトルによる埋め込み誤差が、ある上界で抑えられる。

研究テーマ：スケールフリーに頑健なクラスタ手法

- 先行手法はSBM (どのノードも平均的に同じ次数を持つ状況) では有効に機能する
 - 一方、次数分布がべき乗で近似されるグラフデータ (スケールフリーネットワーク) ではクラスタ推定性能が劣化
 - スケールフリーネットワークは実問題でよく観測される性質
- ⇒ クラスタ構造を持つスケールフリーネットワークでも頑健にクラスタ推定できる手法の開発

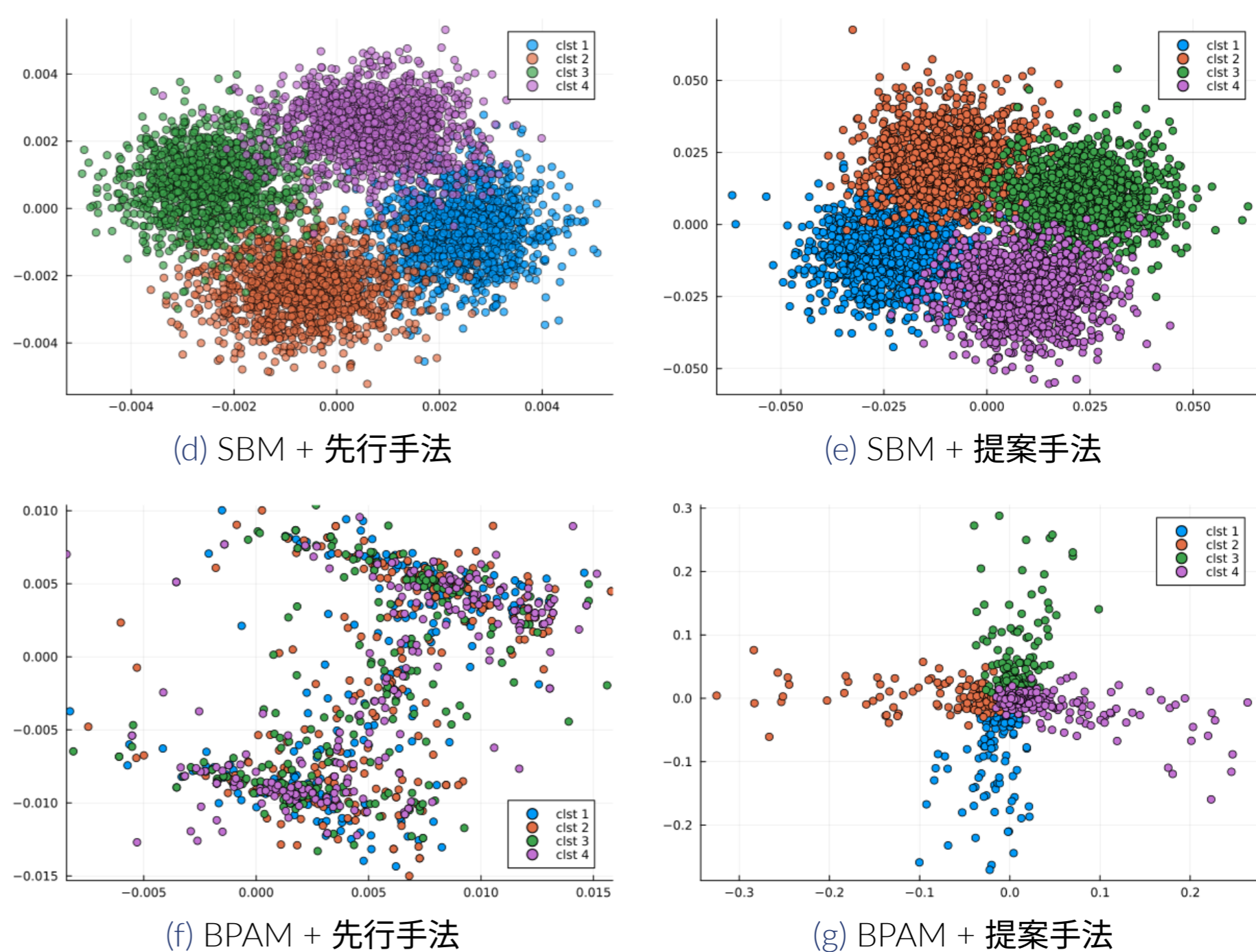


Figure 1. 2つのグラフ生成モデル (SBM, BPAM) を各手法でノード埋め込みした結果。真のクラスタでノードを色分け。

参考文献

- [1] Satoshi Furutani et al. “Graph Signal Processing for Directed Graphs Based on the Hermitian Laplacian”. In: *Machine Learning and Knowledge Discovery in Databases*. Springer International Publishing, 2020, pp. 447–463.
- [2] Steinar Laenen and He Sun. “Higher-Order Spectral Clustering of Directed Graphs”. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS’20. Vancouver, BC, Canada: Curran Associates Inc., 2020. ISBN: 9781713829546.