コミュニティ構造を制御可能な属性付きグラフ生成

前川 政司 †　　佐々木 勇和 ‡
大阪大学情報科学研究科 †　　大阪大学情報科学研究科 ‡
George Fletcher§　　鬼塚 真 ¶
Eindhoven University of Technology§　　大阪大学情報科学研究科 ¶

1 イントロダクション

研究者は自身の設計した手法を評価するために, 大規模かつ多様なデータを必要としている. しかし，利用可能な属性付きグラフは多くの場合はコミュニティの正解ラベルは付与されていない上に小規模であり，手法評価に十分な量の収集は困難である．そのため，実グラフと似た特性を示す正解（クラス）ラベルを持つ属性付きグラフの生成は重要である．実グラフの主要な特徴として，ノード次数分布がべき乗則に従うことが挙げられる．構造と属性の両方に関しては, core/border と homophily/heterophily に注目する.

本研究では属性の観点から似た属性を持つノードの集合としてクラスを導入する．クラス内にはクラス平均に近い属性値を持つ core と複数クラスの属性を混合した border が存在する [2]．これらの現象を捉えるために，グラフ生成器はノードごとにクラスへの所属割合（**ノード-クラス所属割合**）を仮定する必要がある．また，同じクラスに属するノードは互いに接続する傾向があることが知られており，homophily と呼ばれている [4]. コミュニティによっては似ていない属性を持つノードを内包するものがあり，反対の概念として heterophily と呼ばれている．これらの現象を捉えるために所属割合とエッジの傾向を切り分けて，ノードごとのクラスへの接続傾向（**ノード-クラス接続割合**）を導入する.

グラフ生成 [3, 1] は広く研究されているが，ノードレベルでクラスとの関係を考慮する手法はない．そこで，ノード-クラス所属/接続割合を表す潜在変数を用いることで，ノードレベルでの制御が可能なグラフ生成器である acMark を提案する．acMark は高精度かつ効率的な人工属性付きグラフの生成を可能とする．acMark の特徴は，属性付きグラフの特性による制約を表す潜在変数を導入することと，効率的なグラフ生成アプローチを取ることである.

2 事前準備

クラスラベルを持つ属性付きグラフは $G = (\boldsymbol{S}, \boldsymbol{X}, \boldsymbol{C})$ と表すことができ，隣接行列 $\boldsymbol{S} \in \{0,1\}^{n\times n}$，属性行列 $\boldsymbol{X} \in \mathbb{R}^{n\times d}$，クラスラベル $\boldsymbol{C} \in \{1,\ldots,k\}^n$ から構成される．ここでの n, d, k はそれぞれノード数，属性数，クラス数を表す.

グラフ生成器がサポートすべき特徴として，グラフ特徴とクラス特徴がある．グラフ特徴はノード次数分布と属性値の分布である．これらについて，はユーザが指定した分布を実現する必要がある．クラス特徴は，homophily/heterophily および core/border を表す統計量であるクラス接続平均，クラス接続分散と，それらにクラスの大きさを補完するクラスサイズ分布からなる．クラス接続平均の各要素 $\boldsymbol{M}_{l_1,l_2}$ はクラス l_1 のノードからクラス l_2 のノードへの接続割合の平均を表し，以下のように定式化される：

$$\boldsymbol{M}_{l_1 l_2} = \frac{1}{|\Omega_{l_1}|} \sum_{i\in\Omega_{l_1}} \Big(\sum_{j\in\Omega_{l_2}} \boldsymbol{S}_{ij} / \sum_{j\in N} \boldsymbol{S}_{ij} \Big). \tag{1}$$

クラス接続分散の各要素 $\boldsymbol{D}_{l_1,l_2}$ はクラス l_1 のノー

On generating Attributed Graphs with Controlled Community Structure
† Seiji Maekawa, Osaka University
‡ Yuya Sasaki, Osaka University
§ George Fletcher, Eindhoven University of Technology
¶ Makoto Onizuka, Osaka University

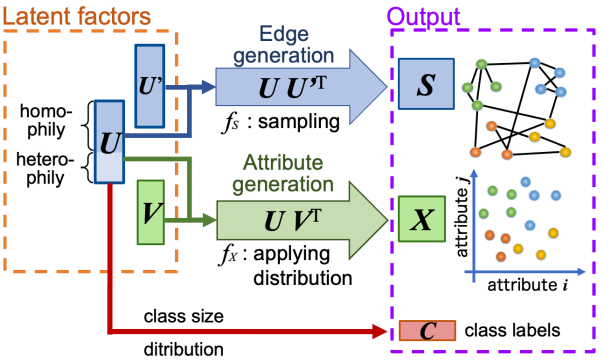
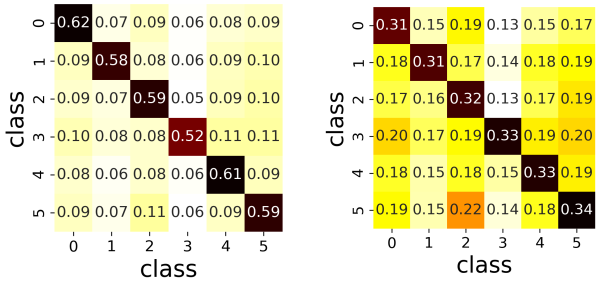


図 1: acMark の概要図



(a) 各セルは生成グラフのクラス接続平均を表す.

(b) 各セルは生成グラフのクラス接続分散を表す.

図 2: クラス接続平均とクラス接続分散の可視化.

ドからクラス l_2 のノードへの接続割合の分散を表し，以下のように定式化される：

$$\boldsymbol{D}_{l_1 l_2} = \sqrt{\frac{1}{|\Omega_{l_1}|}\sum_{i\in\Omega_{l_1}}\Big(\sum_{j\in\Omega_{l_2}}\boldsymbol{S}_{ij}\Big/\sum_{j\in N}\boldsymbol{S}_{ij} - \boldsymbol{M}_{l_1 l_2}\Big)^2}. \tag{2}$$

3 提案手法：acMark

提案手法である acMark では，グラフ特徴およびクラス特徴，ノード数，エッジ数，属性数を入力し，次に入力情報から潜在変数を作成する．最後に潜在変数を用いて表現される制約に従うグラフを生成する．潜在変数は三つの行列，ノード-クラス所属割合を表す $\boldsymbol{U} \in \mathbb{R}^{n\times k}$，ノード-クラス接続割合を表す $\boldsymbol{U}' \in \mathbb{R}^{n\times k}$，属性-クラス相関割合を表す $\boldsymbol{V} \in \mathbb{R}^{d\times k}$ からなる．それぞれがノード-クラス間および属性-クラス間の射影を意味するため，隣接行列 $\boldsymbol{S}$ は $\boldsymbol{U}\boldsymbol{U}'^{\top}$，属性行列 $\boldsymbol{X}$ は $\boldsymbol{U}\boldsymbol{V}^{\top}$ をもとに生成を行う．クラスラベルについては，入力であるクラスサイズ分布をもとに生成を行う．提案手法の概要を図 1 に示す．

クラス構造を示すユーザ入力である $\boldsymbol{M}, \boldsymbol{D}$ の制約を表現するために潜在変数 $\boldsymbol{U}, \boldsymbol{U}'$ を設計し，グラフ生成手順の中で潜在変数で表されるロスを最小化する．指定したノード次数分布に従うグラフを生成することは NP-complete [1] であるので，acMark は高次数ノードから生成を行うヒューリスティックな手順を提案する．

4 実験

図 2 を用いて，acMark がノードとクラス間の接続割合を制御可能かを示す．図 2a では，入力 $\boldsymbol{M}$ の対角要素を全て 0.6 として生成したグラフのクラス接続平均を示しており，高い精度で生成できていることがわかる．図 2b では，入力 $\boldsymbol{D}$ の対角要素をそれぞれ $[0.2, 0.2, 0.25, 0.25, 0.3, 0.3]$ として生成したグラフのクラス接続分散を示している．対角要素の左上から右下にかけて値が大きくなっていることから，傾向を制御できていることが確認できる．

以上のことから，acMark は既存手法が制御できないノード-クラス所属/接続割合をサポートすることで，ユーザが指定した特性を持つグラフを柔軟に生成可能である．

謝辞

本研究は JSPS 科研費 JP20H00583 の助成を受けたものです。

参考文献

[1] Guillaume Bagan, Angela Bonifati, Radu Ciucanu, George Fletcher, Aurélien Lemay, and Nicky Advokaat. gMark: Schema-driven generation of graphs and queries. *IEEE TKDE*, 2017.

[2] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of KDD*, 1996.

[3] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 2011.

[4] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 2001.