

有向グラフ可視化のためのバンドリングとノード配置

十枝 菜穂子[†] 中澤 里奈^{††} 伊藤 貴之[†] 齋藤 隆文^{†††}

[†]お茶の水女子大学理学部情報科学科 〒112-0012 東京都文京区大塚 2-1-1

^{††}お茶の水女子大学大学院人間文化創成科学研究科 〒112-0012 東京都文京区大塚 2-1-1

^{†††}東京農工大学大学院生物システム応用科学研究所 〒183-0057 東京都府中市晴見町 3-8-1

E-mail: [†]{g1220530, itot}@is.ocha.ac.jp, ^{††}leena@itolab.is.ocha.ac.jp ^{†††}txsaito@cc.tuat.ac.jp

あらまし SNS 上の人間関係や遺伝子のネットワーク、論文の参照関係など、さまざまなデータが有向グラフの形で表現される。これらのデータは情報量が膨大になることも多いため、全体の構造を短時間で理解するための可視化手法が有用である。本報告では非階層型の有向グラフを可視化する一手法を提案する。具体的には、エッジのバンドリング（束化）に加えて、なす角が小さい束の合流、ノードクラスタ間で双方向に接続されたエッジ群のバンドリングを提案するものである。また、それらを効果的にするノード配置手法も提案する。

キーワード 可視化, グラフ可視化, 有向グラフ, エッジバンドリング, ノード配置, クラスタリング

1. 概要

ノードとそれを連結するエッジから構成されるグラフ構造は、SNS 上の人間関係や遺伝子のネットワーク、論文の参照関係などのさまざまなデータを表現するために用いられる。遺伝子ネットワークや論文参照関係については大規模なデータベースが公開されており、また SNS 上の人間関係は Web API などを通して入手することがある程度可能である。これらのデータは情報量の規模が大きく、さらに日々情報量が増加したり内容が変化したりする性質をもつものが多い。

このようなグラフ構造の全体像をユーザが一目で把握できるような形で可視化することは有用である。現にグラフの可視化は、情報可視化という研究分野の中でも最も多くの論文が発表されている課題のひとつである。SNS 上で稼働するアプリの中には人間関係をグラフとして可視化するアプリもあり、また生命情報分野のオープンソースソフトウェアの中には遺伝子のネットワークやパスウェイを可視化するパッケージも知られている。しかしこれらのグラフ可視化ソフトウェアの多くは、グラフを構成するエッジの方向性を考慮しないアルゴリズムを採用している。エッジが方向性を有するグラフ（有向グラフ）に特化したグラフ可視化手法を開発することで、これらのグラフ構造の可視化がさらに効果的になると考えられる。

そこで本報告では、グラフの中でも特にエッジが方向性を有し、さらに双方向にエッジが接続される可能性のある非階層型の有向グラフを対象として、エッジのバンドリングとノードのクラスタリングを用いて可視化する一手法を提案する。提案手法ではまずノードを 1 階層または 2 階層にクラスタリングし、クラスタごとにノードを配置した後エッジのバンドリング(束化)を用いたエッジ描画をおこなう。エッジのバンドリ

ングでは、同一クラスタ間を接続するエッジ群のバンドリングだけでなく、なす角が小さい束の合流、クラスタ間で双方向に接続されたエッジ群の方向の情報を保持したバンドリングの 3 つを同時におこなう。また、ノード配置手法については、このバンドリングを効果的にする配置手法を目指す。

2. 関連研究

グラフ可視化におけるエッジのバンドリング（束化）とは、近隣するノードを接続する複数のエッジが束として可視化されるように変形または統合する処理を指す。Holten は、同一クラスタペアに属する 2 ノードを連結するエッジ群をバンドリングすることでクラスタ間の関係性を強調する手法[1]を提案しているが、これは無向グラフを対象とした手法である。

Selassie らの手法[2]や Holten らの手法[3], Ersoy らの手法[4]では向きと距離が近いエッジをバンドリングすることでグラフ全体の特徴を概観することが可能であるが、エッジをその中央付近で合流させているため各エッジの接続するノードの読み取りに曖昧性が生じる。

これに対しエッジをその両端（特定のノードの近く）でバンドリングすることで曖昧性を排除しながらもエッジの交差を減らす手法[5]も提案されているが、これはノードのクラスタリングを前提としていない。

本報告では、ノードがクラスタリングされた有向グラフのクラスタ間の関係を表現するためのバンドリング手法を提案する。なお、本研究ではノードをクラスタリングして配置し、同一クラスタ間に生成されるエッジ群の束化を可能とした無向グラフ可視化手法 Koala[6]を拡張する形で提案手法を実装する。

3. 提案手法

本章では本手法の提案手法について、データ構造および処理手順について説明する。

3.1 データ構造

提案手法では入力データとして有向グラフを想定する。また本研究ではノードの接続性だけでなく任意の数値属性を加味してグラフ配置を実現できるようにするために、各ノードに対して多次元ベクタとなる特徴量を付与できるものとする。

3.2 ノードクラスタリング

提案手法では任意の2ノード間の距離を算出し、その距離に基づいてノードをクラスタリングする。以下にノード間距離の算出、ノードのクラスタリングそれぞれについて述べる。

3.2.1 ノード間距離の算出

本手法では任意の2ノード間の距離 d を

d_{vec} : 特徴量ベクタに基づくノード間距離

d_{adj} : 隣接ノードの共通性に基づくノード間距離

の一次結合

$$d = \alpha d_{vec} + (1 - \alpha) d_{adj}$$

により算出する[6]。以下に2種類の距離の定義を述べる。

特徴量ベクタに基づくノード間距離

各ノードに付与された特徴量ベクタを a_i および a_j としたとき、

$$inner = a_i \cdot a_j / |a_i| |a_j|$$

により特徴量ベクタの内積 $inner$ を求め、この値を用いて以下の式

$$d_{vec} = 1.0 - inner$$

によって特徴量ベクタに基づくノード間距離 d_{vec} を算出する。

隣接ノードの共通性に基づくノード間距離

ノードの両方にエッジで接続されている隣接ノードの個数を n_{adj} とする。この値を用いて

$$d_{adj} = 1.0 / (1.0 + n_{adj})$$

によって隣接ノードの共通性に基づくノード間距離 d_{adj} を算出する。

3.2.2 クラスタリング

提案手法では階層のないクラスタ群と2階層のクラスタ群を同時に生成するために、以下の手順によってクラスタリングする。

1. 最長距離法に基づく階層型クラスタリングにより、閾値 α までノードをクラスタリングする。これを今後クラスタ群 A とする。
2. クラスタ群 A の情報を保持したまま閾値 β ($\alpha < \beta$) までクラスタリングを続ける。

以上の処理により、クラスタ群 A のクラスタを内部に持つ二階層のクラスタ群ができる。これをクラスタ群

B とする。

3.3 ノード配置

提案手法では、クラスタごとにノードを配置するため、以下のアルゴリズムにより各ノードの配置場所を決定する。階層をもたないクラスタ群 A の場合のアルゴリズムと階層をもつクラスタ群 B の場合のそれぞれについて述べる。

階層をもたないクラスタ群 A の場合

以下のアルゴリズム[6]によりノードの配置場所を決定する。

1. クラスタをノードに置き換えたグラフを生成し、これに対しグラフ配置アルゴリズムを適用する。現時点ではエッジにバネの力学モデルを適用した Spring-force モデルによる配置計算方法を使用している。また、グラフ生成の際にエッジ束を構成する本数でエッジ束に重みをつけることで、双方向にエッジを有するクラスタペアが近くに配置されることが期待される。
2. 各クラスタのノード数からクラスタ半径を計算し、エッジ長が両端のクラスタの半径の合計値にできるだけ近づくようにスムージング処理を適用する。
3. クラスタ内にそのクラスタを構成するノードを配置する。

階層をもつクラスタ群 B の場合

クラスタ群 A の場合のアルゴリズムを拡張した以下のアルゴリズムによりノードの配置場所を決定する。

1. 小さいクラスタ(以下小クラスタとする)を内包する大きなクラスタ(以下大クラスタとする)群に対して、クラスタをノードに置き換えたグラフを生成し、グラフ配置アルゴリズムを適用する。この各大クラスタの半径は一定とする。
2. 各大クラスタ内部について小クラスタをノードとしたグラフを生成し、これに対してグラフ配置アルゴリズムを適用する。
3. 各小クラスタのノード数からクラスタ半径を計算する。
4. 小クラスタの半径に合わせたスムージング処理を適用する。
5. 各小クラスタ内にそのクラスタを構成するノードを配置する。

3.4 エッジ描画

提案手法では図1の3つの条件を満たすエッジ描画を目標とする。このうち条件 A は従来のバンドリング手法と同じ目標であり、条件 B は有向グラフ描画のために条件 A を拡張したと考えられる。条件 C を満たすために提案手法ではまず合流するエッジ束を選択する。続いて、各エッジの条件に合わせて曲線を描画する。以下、合流するエッジ束の選択手法、曲線の描画手法

のそれぞれについて述べる。

またエッジの終点側にマークをつけることでエッジの向きを表現するとノード上にマークが重なってしまうため、現在の我々の実装では青からオレンジのグラデーションで曲線を描画することによってエッジの向きを表現している。

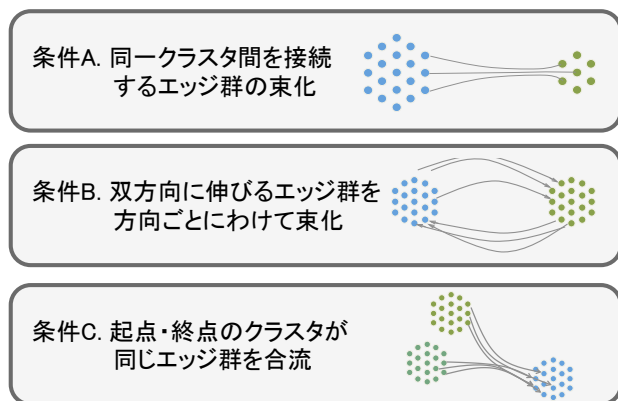


図 1. エッジ描画の条件

3.4.1 合流するエッジ束の選択手法

提案手法では以下のアルゴリズムによって合流するエッジ束を選択する。

1. 1つのクラスタを取り出す。
2. 時計回りに1つずつエッジ束を参照する。
3. 左隣のエッジ束が同じ向きの場合、合流するかどうかの評価値を算出し、ユーザの入力した閾値以上ならば合流すると判定する。
4. 全ての束およびクラスタに対して同じ処理を適用する。

ここで評価値の算出は以下の2つの評価方法から算出する。

i. エッジ束の角度

2本のエッジ束がなす角度が閾値よりも大きい場合は、これらが合流されないように評価値を0とする。

ii. 合流できる部分の長さ

2本のエッジ束の角度の二等分線に両クラスタから垂線を伸ばし、その交点と基準となるクラスタとの距離のうち短い方の長さを評価値とする(図2)。

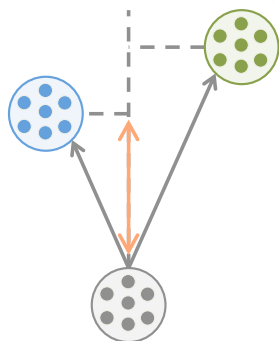


図 2. 合流できる部分の長さの算出
オレンジの矢印が求める長さとなる。

3.4.2 エッジの曲線描画手法

本手法では全てのエッジを以下の3パターンに分類し、それぞれの曲線をBezier曲線によって描画する。

i. 他の束と合流するエッジ

2本のエッジの垂直二等分線上にBezier曲線の制御点を設定する(図3(左))。

ii. 他の束と合流はしないが双方向に伸びているエッジ

2つのクラスタ間の垂直二等分線上にBezier曲線の制御点を設定する(図3(右))。

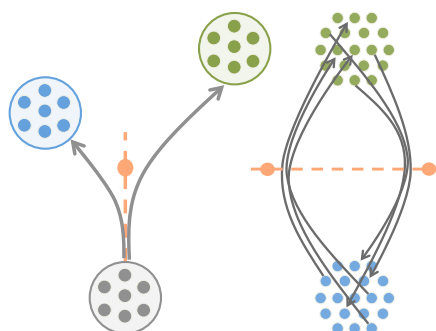


図 3. i, ii での Bezier 曲線の制御点の決定手法

(左)他の束と合流するエッジ(i)の制御点

(右)合流せず双方向に伸びているエッジ(ii)の制御点

iii. i, ii のどちらでもないエッジ

以下の a または b によって制御点を決定する[6]。

- a. クラスタの中心点を連結する線分上の2点を制御点とする(図4(左))。
- b. 2ノードを連結する線分とクラスタの中心点を連結する線分について、それぞれを3分割した点を生成し、これらの点を連結する線分上に2点の制御点を生成する(図4(右))。

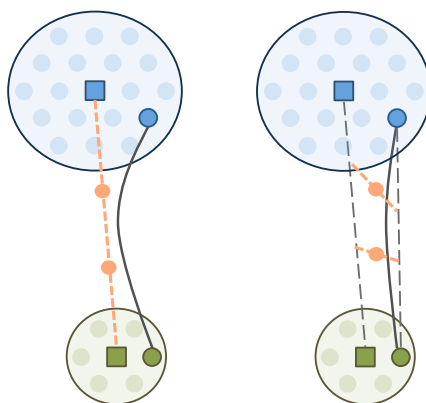


図 4. iii での Bezier 曲線の制御点の決定手法

4. 適用事例

本章では適用事例として、1990 から 94 年、2000 から 2010 年までに ACM SIGGRAPH で発表された論文に対して、論文をノード、参照関係をエッジとしたグラフを可視化した事例[7]を示す。このグラフにおいてノ

ード(論文)は 1072 個, エッジ(参照関係)は 5498 本である. また各ノードに付与される特徴量として, 10 の研究分野についてそれぞれの程度当てはまるかを各論文の内容によって分類した 10 次元の特徴量を用いる.

また, この 10 次元の特徴量の中で最も値の高い次元ごとに各ノードの色付けを行っている.

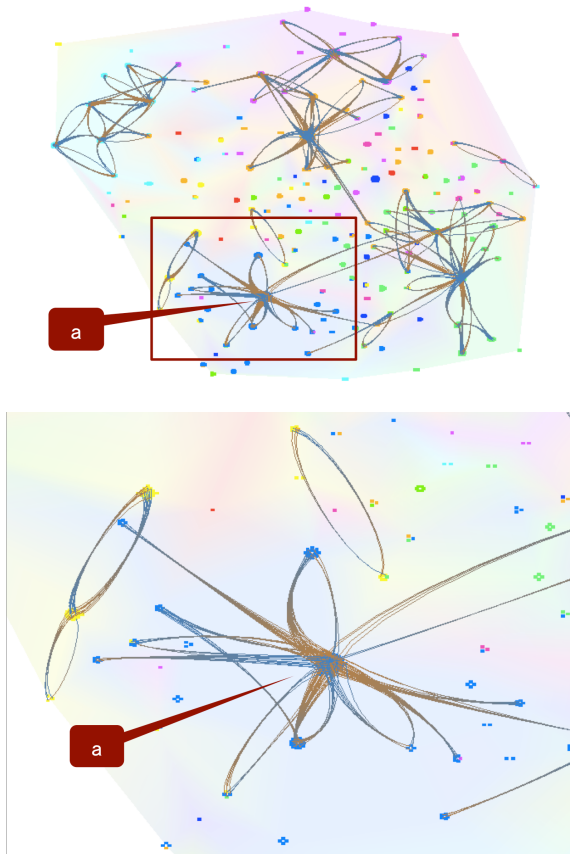


図 5. (上)提案手法による可視化例
(下) クラスタ a 付近の拡大図

図 5 がクラスタ群 A(非階層型)を用いた可視化結果の一例である. クラスタ A に接続されているエッジがバンドリングされ太い束になっており, このクラスタの論文が多数の論文から参照されていることを読み取ることができる.

図 6 はクラスタ群 B(階層型)を用いた可視化結果の一例である. 多くのクラスタが同分野(同色)のクラスタとの参照関係で結ばれている中, テクスチャの自動生成・自動加工に関する論文のクラスタ b が他の多くの分野のノードから参照されていることを読み取ることができる. レンダリング分野である c では, 3D モデルに貼り付けるテクスチャの生成のために, クラスタ b で提案されたテクスチャ生成が用いられている. また, テクスチャ・形状特徴分野の d や形状処理・変形分野の e では, 画像補完に使うテクスチャの生成や 3D モデルに貼り付けるテクスチャの生成のために, クラ

スタ b の技術が用いられている. これらの参照関係から, クラスタ b の論文の内容が, さまざまな分野に適用される汎用的な内容であると読み取ることができる.

ただし現時点ではクラスタ群 d, e において, 論文が内容に沿って適切に分類されておらず, 雑多な内容の論文が同一クラスタに混ざっている. そこで, 各論文への特徴量の付与手法およびノードクラスタリング手法を今後再検討する必要がある.

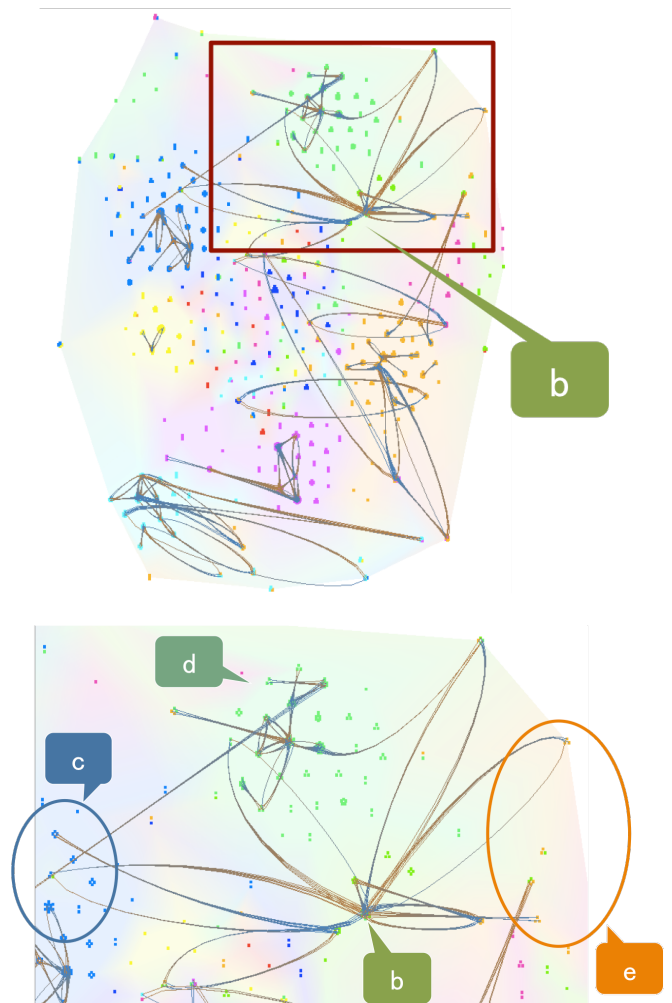


図 6. (上)提案手法による可視化例
(下) クラスタ b 付近の拡大図

5. まとめと今後の課題

本報告では有向グラフ可視化に特化したエッジバンドリング手法, およびそれをより効果的にするノード配置を提案した. 本手法では双方向に接続されたエッジの視覚的な分離を容易にする上に, クラスタ間のエッジ束の合流という概念の適用により, クラスタ間の関係性の概観がさらに容易になったと考えられる.

今後の課題として, 起点と終点の両方で合流するエッジ束の曲線のなめらかな描画方法, 合流させるエッ

ジ束の選択手法，重要なエッジ束を合流させることに
適したクラスタの配置方法などを再検討したい．

参 考 文 献

- [1] Danny Holten, “Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data”, IEEE Trans. on Visualization and Computer Graphics, Vol. 12, No. 5, pp. 741-748, 2006
- [2] David Selassie, Brandon Heller and Jeffrey Heer, “Divided Edge Bundling for Directional Network Data”, IEEE Trans. on Visualization and Computer Graphics, Vol. 17, No. 12, pp. 2354-2363, 2011
- [3] Danny Holten, Jarke J. van Wijk, “Force-Directed Edge Bundling for Graph Visualization”, Computer Graphics Forum, Vol. 28. No. 3. Blackwell Publishing Ltd, 2009
- [4] Ozan Ersoy, Christophe Hurter, Fernando V Paulovich, Gabriel Cantareira, Alexandru Telea, “Skeleton-based edge bundling for graph visualization”, IEEE Trans. on Visualization and Computer Graphics, Vol. 17, No. 12, pp. 2364-2373, 2011
- [5] Sheng-Jie Luo, Chun-Liang Liu, Bing-Yu Chen, Member, IEEE, and Kwan-Liu Ma, Senior Member, IEEE, “Ambiguity-Free Edge-Bundling for Interactive Graph Visualization”, IEEE Trans. on Visualization and Computer Graphics, Vo.18, No. 5, pp. 810-821, 2011
- [6] T. Itoh, K. Klein, “Key-node-Separated Graph Clustering and Layout for Human Relationship Graph Visualization”, IEEE Computer Graphics and Applications, Vol. 35, No. 6, pp. 30-40, 2015.
- [7] R. Nakazawa, T. Itoh, T. Saito, “A Visualization of Research Papers Based on the Topics and Citation Network”, 18th International Conference on Information Visualisation (IV2015), pp. 283-289, 2015.