

多属性グラフへのエッジバンドリングの適用 Edge Bundling in Multi-Attributed Graphs

山下 敬史[†] 佐賀 亮介^{*}
Takafumi Yamashita Ryosuke Saga

1. はじめに

1.1 研究背景

今日様々な市場において、サービスの質の向上などを最終目的とした情報可視化の重要性が増大している[1][2]. 情報可視化とは、データにクラスタリング等の前処理を加えた後、図形などを用いて人の目に分かりやすく表現することである. これによりデータ間の関係性などを直感的に読み取ることができるようになる.

情報可視化には様々な手法が存在し、その一つにデータをノードとリンクによって表現するネットワーク可視化が挙げられる[3]. データをノードとリンクとして表現することで、データを視覚的に捉えることができる. しかしこの手法では、データの数、すなわちノードとリンクの数が増加するにつれ“visual clutter”(視覚的な散らばり、大量のエッジが雑然とした状態)によってグラフの視認性が低下するという問題を抱えている.

この visual clutter を低減させる手法として、初めにグラフのレイアウトを変えるアプローチが提案された[4]. ノードの座標を適切に配置しなおすことで一定の成果が挙げられてきたが、このアプローチではノードが固定されたグラフに適用できなかつたり、大量にリンクを持つ場合改善が見られない事がある. そこでエッジを束ねるエッジバンドリングという新たなアプローチが提案されてきた[5]. これは一定の規則に基づきエッジ同士を束ね、エッジの主な流れを把握しやすくする手法である. ノードの階層構造に基づく手法[6]や平行座標に基づく手法[7]などが提案されており、その中にグラフを力学モデルに当てはめてエッジを束ねる手法がある[8]. 後述する関連研究で提案された力学的バンドリング手法により、エッジの束を明確にすることで煩雑なグラフの視認性を向上することに成功してきた.

一方で、グラフには何らかの属性が付与されることがある. 例えば空港をノード、航路をリンクとした空路図において、エッジには航空会社の違いが属性として付与され得る. また FACT-Graph においてトレンドの情報などが属性として考えられる[9]. しかし従来のバンドリング手法では例に挙げたようなグラフの属性は考慮されておらず、グラフが持つ情報を表現しきれないと考えられる.

そこで本研究では属性の違いを考慮したバンドリング手法を提案し、異なる属性ごとにエッジを束ねグラフの視認性を向上させることを目的とする. 関連研究では空路図が用いられていたが、本論文では新聞の社説データから抽出したキーワードの共起グラフに対し提案手法を適用することで、その有用性を示す.

1.2 本論文の貢献

本論文の貢献を以下に示す.

- 従来の力学的手法を拡張し、属性を考慮したバンドリングが可能になった点で新規性がある.
- 新聞の社説など属性を持つ様々な実データに対して適用可能性があり、ネットワーク可視化を行う際に有用である.

2. エッジバンドリング

2.1 Force Directed Edge Bundling(FDEB)

代表的なエッジバンドリング手法として、Holten らによる Force-Directed Edge Bundling が挙げられる[10]. この手法では各エッジをばねとみなし、フックの法則に基づく力を考えることで力学的にエッジを束ねていく. エッジをいくつかの分割点に分割し、着目している分割点とその前後の分割点間に互いに引っ張り合うばねの力、そして別のエッジの対応する分割点間に引力が働く. エッジの分割数は各座標の移動後に更新され、徐々にバンドリングが進んでゆく. 分割点に働く力を図 1 に示す.

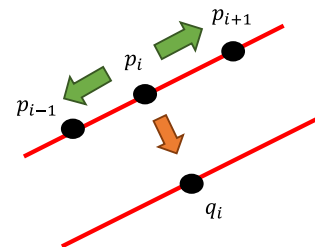


図 1 FDEB の概念図

一方でバンドリングを過度に行ってしまうと、少ないエッジの束に集約されてしまい、本来グラフが持つ関係性を表現できなくなってしまう. そこで互換性 Compatibility という指標を導入し、エッジの組合せのうち不適切なものに対して力のかかり具合に重み付けをすることで、元のグラフが表現していた関係性が損なわれないようにしている.

エッジ P 中の分割点 p_i に対し働く力 F_s は以下の式で与えられる.

$$F_s = k_p \cdot (\|p_{i-1} - p_i\| + \|p_i - p_{i+1}\|) + \sum_{Q \in E} \frac{C_e(P, Q)}{\|p_i - q_i\|} \quad (1)$$

ただし、 k_p はばね定数、 i は分割点の番号 ($1 < i < n$)、 $C_e(P, Q)$ はエッジ P, Q 間の互換性、 q_i は p_i と対応するエッジ Q 中の分割点である. 第一項はエッジ内の前後の分割点間で引っ張り合うばねの力であり、エッジの硬さを定めるものである.

また、エッジ P, Q 間の互換性 $C_e(P, Q)$ は 4 種類の互換性を

大阪府立大学大学院 工学研究科
Graduate School of Engineering, Osaka Prefecture University

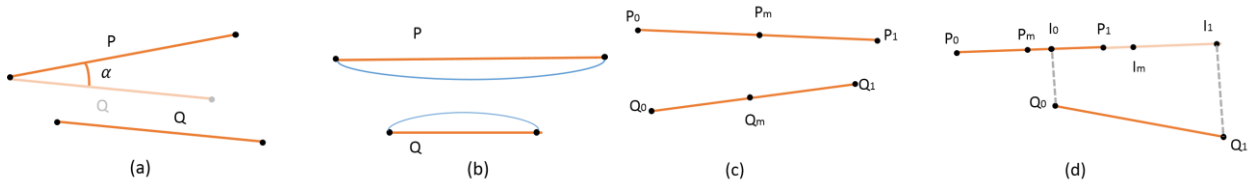


図 2 各互換性の概念図

(a) angle compatibility (b) scale compatibility
(c) position compatibility (d) visibility compatibility

各項とする以下の式で与えられる。

$$C_e = C_a \times C_s \times C_p \times C_v \quad (2)$$

各互換性はそれぞれ[0,1]の値をとる。各互換性の概念図を図 2 に示し、説明を以下に述べる。

- C_a (angle compatibility) : エッジがなす角度に関する互換性。平行なエッジほど束ねても本来の表現力を損なわないという考えに基づいており、エッジがなす角度が大きくなるほど値は 0 に近づく。図 2(a)で示した角度を α とすると、 C_a は以下の式で定義される。

$$C_a = |\cos(\alpha)| \quad (3)$$

- C_s (scale compatibility) : エッジの長さに関する互換性。エッジの長さが同じであるほどバンドリングしても本来の表現力は失われないという考えに基づいている。長さの差が大きい場合は図 2(b)で示したようにエッジの湾曲の度合いが大きくなり不自然なバンドリングとなるため、長さの差が少ないほど C_s は大きな値をとる。 C_s は以下の式で与えられる。

$$C_s = \frac{2}{I_{avg} \cdot \min(|P|, |Q|) + \max(|P|, |Q|)/I_{avg}} \quad (4)$$

$$I_{avg} = (|P| + |Q|)/2 \quad (5)$$

- C_p (position compatibility) : エッジの距離に関する互換性。近い距離に位置するエッジの組合せほど束ねられるべきという考えにより、エッジの midpoint 間の距離が大きくなればなるほど値は小さくなる。図 2(c)に示した P_m , Q_m をそれぞれエッジ P, Q の midpoint とするとき、 C_p は以下の式で与えられる。

$$C_p = I_{avg} / (I_{avg} + \|P_m - Q_m\|) \quad (6)$$

- C_v (visibility compatibility) : エッジの重なり合いに関する互換性。一方のエッジをもう一方のエッジがなす直線状に投影したときの重なり合いが多いほど値は大きくなる。図 2(d)に示した、エッジ Q を投影した線分 I_0I_1 の midpoint を I_m とすると、 C_v は以下の式で与えられる。

$$C_v = \max(V(P, Q), V(Q, P)) \quad (7)$$

$$V(P, Q) = \max\left(1 - \frac{2 \cdot \|P_m - Q_m\|}{\|I_0 - I_1\|}, 0\right) \quad (8)$$

2.2 Divided Edge Bundling

Selassie らは FDEB を改良した Divided Edge Bundling を提案した[11]。この手法は有向・無属性グラフに対し適用可能な手法である。ばね力に加え、Selassie らはポテンシャル関数に基づくクーロン力を提案した。ポテンシャル関数

とは着目している 2 点間の距離を変数とする潜在的エネルギーである。ポテンシャル関数が最小の値をとる点をポテンシャル最小点 m_j とすると、 p_i は m_j に引き付けられるようにクーロン力が働く。また着目しているエッジの組合せが異なる向きである場合、 m_j はエッジ Q の進行方向に対し右側に移動する。属性の組合せに応じて m_j が移動することにより、働くクーロン力の大きさに差が出て同じ方向のエッジ同士が束ねられる。Divided Edge Bundling の概念図を図 3 に示す。

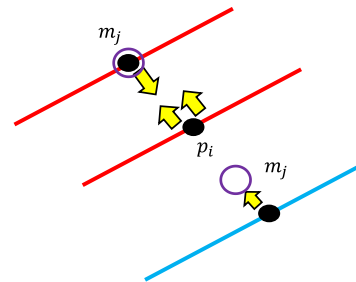


図 3 Divided Edge Bundling の概念図

m_j の座標とエッジ P 中の分割点 p_i に対し働くクーロン力 F_c は以下の式で与えられる。

$$m_j = \begin{cases} q_j & (\text{if } P \cdot Q > 0) \\ q_j + lN_j & (\text{if } P \cdot Q < 0) \end{cases} \quad (9)$$

$$F_c = \frac{-sk_c}{\pi C (s^2 + |p_i - m_j|^2)^2} \quad (10)$$

ただし l , s はパラメータ、 N_j は移動方向を表すベクトル、 k_c はクーロン定数、 C はエッジの分割数である。

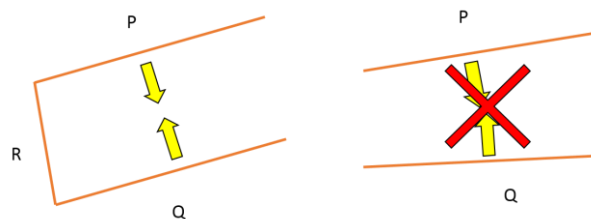


図 4 connectivity compatibility の概念図

また互換性にも新たな指標 C_c (connectivity compatibility)が提案されている。 C_c は[0,1]の値をとり、これは 2 エッジ間の結ぶエッジ数、すなわち最短経路に基づいている。最短経路が少ないエッジの組合せほど値は大きくなるが、図 4

に示すように 2 エッジ間を結ぶ経路が存在しなかった場合、値が 0 となり力が働かなくなる。グラフが分断された部分を持つ場合バンドリングに厳しい制限が課せられるため、各データセットに対して C_c を導入するのが適切かどうか判断する必要がある。 $D_{min}(P, Q)$ をエッジ P, Q 間の最短経路数とすると、 C_c は以下の式で与えられる。

$$C_c(P, Q) = \frac{1}{1 + D_{min}(P, Q)} \quad (11)$$

2.3 従来手法の問題点

1 節で述べた FDEB により、グラフに力学モデルを当てはめることで自然なバンドリングが可能になった。2 節で述べた Divided Edge Bundling により、エッジが持つ方向を考慮し同じ方向のエッジ同士を束ねることに成功した。しかし 1 章で述べたとおり、これらの手法ではエッジに付与され得る属性を考慮していない。様々な属性が混在するグラフにおいて、属性を考慮し同じエッジを束ねることで、本来グラフが持つ情報を活かしつつ視認性を向上させ、表現力を高めることができると考えられる。

3. 属性を考慮した手法

3.1 アプローチ

本論文では、エッジが持ちうる属性を A, B, C の 3 属性と定義する。属性 A と属性 B は互いに対立する 2 属性であり、属性 C はそのどちらにも属しうる属性である。全てのエッジの組合せに対し、フックの法則と引力が働く。これによりエッジ全体の大まかな流れを表現する。そしてエッジの組合せが同じ属性同士、すなわち “A と A” , “B と B” , “C と C” の場合、クーロン力が引力として更に強く引き付けるように働く。エッジの組合せが異なる、すなわち “A と B” の場合は、クーロン力が斥力としてエッジが反発しあうように働く。これらの力によってエッジを A · B · C の属性ごとに束ねてゆく。

3.2 互換性

2.1 節で述べたとおり、過度のバンドリングによって本来エッジが表現しているノード間の繋がりが把握しづらくなる可能性がある。不適切なバンドリングを防ぎこれらの表現力を損なわずバンドリングを実施するため、先行研究で提案された互換性を導入した。属性を考慮したバンドリングであっても、エッジの角度や長さなどに関して不適切な組合せに対し重みを付けるのは妥当と考えられるからである。また本論文で用いたデータセットにおいては、グラフの分断部は意味的なつながりを持たずバンドリングを行わない方が適切と考えられるため Selassie らの提案した互換性も導入している。

すべて考慮した互換性は以下の式で表される。

$$C_e = C_a \times C_s \times C_p \times C_v \times C_c \quad (12)$$

3.3 ポテンシャル最小点の決定

本論文で用いるクーロン力もポテンシャル最小点に基づいている。2.2 節で述べた Divided Edge Bundling ではエ

ッジの向きに着目し m_j が移動していたが、本手法ではエッジの属性の組合せが異なる場合に、着目している分割点と対応する分割点がなすベクトルと反対の方向に移動する。 m_j と移動方向ベクトル T_j は以下の式で与えられる。

$$m_j = \begin{cases} q_j & (\text{if } P \text{ and } Q \text{ are the same attribution}) \\ q_j + lT_j & (\text{else}) \end{cases} \quad (13)$$

$$T_j = \begin{cases} p_i - q_j & (\text{if } P \text{ and } Q \text{ are the same attribution}) \\ q_j - p_i & (\text{else}) \end{cases} \quad (14)$$

ここで、 T_j は次節で述べる分割点に働く力の方向も表している。ポテンシャル最小点の移動の概念図を図 5 に示す。

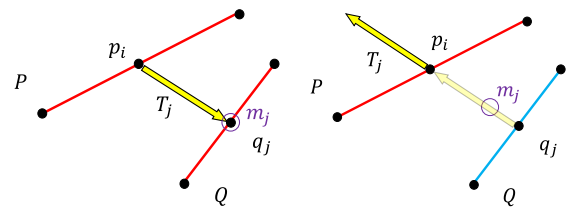


図 5 属性の組合せに応じたポテンシャル最小点の移動の概念図

3.4 分割点に働く力

エッジ中の全ての分割点に対して、FDEB で提案された式(1)によるばね力と引力が働く。それに加え、1 節で述べたようにエッジの属性の組合せに応じたクーロン力が働く。また、Selassie らが提案した逆ローレンツ関数に基づくクーロン力によって、エッジの方向を考慮したバンドリングに一定の成果が見られたため、本論文においてもポテンシャル関数に逆ローレンツ関数を採用した。

ポテンシャル関数に基づく、属性を考慮したクーロン力 F_c' は以下の式で表される。

$$F_c' = f(J_p) \cdot \frac{T_j}{|T_j|} \cdot \frac{-C_e(P, Q)sk_c|p_i - m_j|}{\pi C (s^2 + |p_i - m_j|^2)^2} \quad (15)$$

$$f(J_p) = \alpha \cdot J_p + \beta \quad (16)$$

ここで J_p はエッジ P の重み、 α , β はパラメータである。式(16)で表される関数を導入し、重みの大きいエッジほど力が強く働くように考慮している。エッジに重みが付与されている場合、重みの大きいエッジほど重要なエッジであるため、バンドリングした際のエッジの束の中心に位置すべきと考えられる。ここでエッジの重みを考慮する関数は式(16)の形である必要はなく、データセットに合わせて適切に設定することが望まれる。

分割点に働く力を総括すると以下のように表される。

$$F = \begin{cases} F_s & (\text{if } P \text{ or } Q \text{ is attribution } C) \\ F_s + F_c' & (\text{if else}) \end{cases} \quad (17)$$

エッジの組合せの内、一方が属性 C である場合はエッジ内ばね力と引力のみが働く。同じ属性同士の組合せである場合、クーロン力が引力として作用する。そして異なる属性同士の組合せである場合には、クーロン力が斥力として働く。

4. 実行例

4.1 予備実験

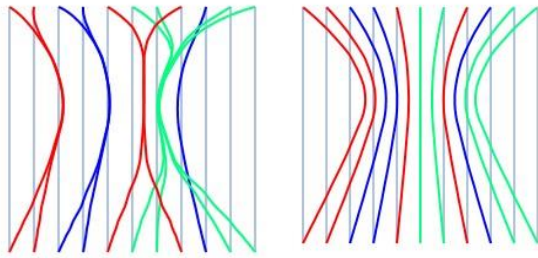


図 6 サンプルに対するバンドリング実行結果
(左)提案手法 (右)FDEB

予備実験として、サンプルグラフに本手法を適用した。サンプルは 11 本のエッジを等間隔に平行に引いたものとし、属性は赤・青・緑のいずれかを付与した。このうち赤と青が互いに対立し、緑はどちらにも属しうる属性とした。またエッジの重みは $[0,1]$ の値でランダムに決定した。実行結果を図 6 に、実行時の各パラメータの値を表 1 に示す。

表 1 バンドリング実行時のパラメータの値

parameters	k_p	k_c	s
default value	10	4.0×10^4	50

parameters	l	α	β
default value	0.7	0.5	0.5

FDEB によるバンドリングでは全てのエッジに等しく引力が働いており、エッジ全体の束は表現できているが属性ごとに分離してはいない。提案手法によるバンドリングでは同じ属性同士が束ねられつつも赤と青のエッジは分離している。一方で赤と青のエッジ間にも引力が働いていることにより、エッジ全体の大まかな流れを表現できており、提案手法の有用性が伺える。

4.2 パラメータによる変化

本手法においてユーザが任意に定める必要のあるパラメータは 1 節で示した通り k_p , k_c , s , l , α , β の 6 個である。表 1 の値を基準値とし、それに対して各パラメータを変化させた時の結果の違いを表 2 に示す。

k_p はばね定数であり、エッジ内部で引っ張り合う力の大きさを決定するパラメータである。この値が大きくなるとばね力が強くなり、引力や斥力の影響を受けにくくなる様子が表から見て取れる。

k_c はクーロン力を決定するパラメータである。この値が大きくなると属性 C を含まないエッジの組合せに対して働くクーロン力の大きさが引力・斥力を問わず大きくなる。提案手法において、このパラメータが 0 に近づくと、FDEB と同様の結果となる。

s は本来ポテンシャル関数の傾斜を表すものである。本手法におけるクーロン力の式(15)において、 s が大きくなるにつれてクーロン力は小さくなる。 s を小さくすることによって、表のようにエッジ同士が接する範囲が広がる。

l は式(13)で示した通り、 m_j の移動する幅を決定するパラ

表 2 パラメータの変化による結果の変化

Parameters	High	Low
k_p		
k_c		
s		
l		
α		
β		

メータである。 l の値によって属性ごとの束の距離が変化する様子が見える。

α , β は重み付けに関するパラメータである。エッジの重みは様々な値をとるので、これらのパラメータを変化させて重みの影響を調節する。

これらのパラメータは扱うデータセットによって適切な値が異なると考えられるため、結果を見ながらチューニングする必要がある。

5. 実験

5.1 使用データ

本論文では属性を持つデータを扱う必要がある。そこで同時期には同じような話題が選ばれやすい新聞に着目し、かつ新聞社毎の主張の違いは属性として付与しやすいと考え、朝日新聞と読売新聞の 2 紙における 2008 年度分の社説データを用いた。TF-IDF の値の上位 200 語を各紙の社説からキーワードとして抽出し、グラフのノードとした。このキーワード間の Jaccard 係数が閾値を超えたノード間にエッジを引いた各紙の共起グラフを作成・合成した。閾値は 0.25 とした。各紙から抽出されたキーワード数とエッジ数を表 3 に示す。

表 3 抽出されたキーワード・エッジ数

	朝日新聞	読売新聞	合成
キーワード	200	200	138
エッジ	269	264	412

エッジの属性は、対立する 2 属性は「朝日新聞に出現するエッジ」「読売新聞に出現するエッジ」、どちらにも属する属性は「両紙に出現するエッジ」とし、いずれかをグラフを合成する際に付与した。

また本実験で扱う社説データから抽出したキーワードは位置情報を持っていないため、グラフを描画する際にノードの配置を任意に定める必要がある。本実験では Kamada-Kawai アルゴリズムによってグラフのレイアウトを決定した[12]。

5.2 実行結果

前節で作成した合成グラフに提案手法を適用した結果を図 7 に、右側の密集部を拡大したものを図 8 に示す。

バンドリング実施前である図 7(a)は大量のエッジが存在しており、どのノード間にエッジがあるのか視認しづらい。また赤・青・緑それぞれのエッジが混在しており、非常に煩雑なグラフとなっている。

図 7(b)は FDEB を実施したもので、エッジが集約することである程度エッジの束が確認できるようになっている。しかし属性によって色分けしたものが全く考慮されずバンドリングされているため、各属性の繋がりがどのようなものなのか確認できない。

一方、提案手法を実施した図 7(c)においては、エッジが全体的に集約され流れが把握しやすくなりつつも、明確に赤と青のエッジの束が分離している。特に図 8 に示した密集部において赤と青の異なる太い束が顕著に表現されており、提案手法を用いることで属性ごとのエッジの束をわかりやすく表現することができたといえる。

5.3 考察

本手法を適用することにより、属性ごとにエッジを束ねることは可能になった。しかし、今回はエッジの属性は考慮したがノードが持ちうる属性については考慮していない。ノードの属性を考慮した再配置などを行うことで、visual clutter をより低減させることができるのではないかと考えられる。

また今回グラフの重みを考慮した関数を導入したが、本実験の結果ではその効果が表れているのか確認しにくい。異なる関数を導入することによってエッジの重みをより効果的に結果に反映させたり、異なるコンセプトに基づいて重みを表現することが可能になると考えられる。

さらに、グラフが持ちうる属性は今回定義した 3 属性とは限らない。属性数の増加に伴って属性間の条件が異なってくる場合でも、バンドリングを実施できるよう手法を拡張する価値はあると考えられる。

また今回の結果に対する考察が定性的評価であることは否めない。エッジバンドリングにおける定量的な評価指標は提案されていないため、指標を導入して従来手法と比較し、どのようなケースにどの手法が適するか判断する必要がある。

6. おわりに

情報可視化手法の 1 つであるノードとリンクによるネットワーク可視化では、ノード数・リンク数が増加するにつれグラフが煩雑なものとなり特徴が把握しづらくなるとい

う問題があった。この問題を解決するために、エッジを束ねて見やすくするという観点からエッジバンドリングという手法が提案されており、力学モデルにグラフを当てはめる等のアプローチで改善が図られてきた。しかし従来のバンドリング手法ではグラフが持ちうる属性を考慮しておらず、データが持つ情報を十分に表現しきれていないという問題があった。

そこで本論文では、エッジの属性に着目し、属性の組合せに応じてエッジ間に作用する力を変化させることで属性ごとにエッジを束ねる手法を提案した。また、新聞 2 紙から作成した、出現した新聞社の違いをエッジの属性とするキーワードの共起グラフに適用し、従来手法と提案手法を適用してその結果を比較した。

実験では朝日新聞と読売新聞の 2008 年度分の社説データからキーワードの共起グラフを作成し、合成したグラフに対して提案手法を適用した。その結果、従来手法では表現できなかったエッジの属性ごとの束を表現することに成功し、データが持つ情報をより表現できるエッジバンドリングが可能であることを確認した。

今後の課題として、ノードの属性を考慮すること、対立する 3 属性以上の属性を持つグラフに対してバンドリングが可能になるように本手法を拡張すること、各エッジバンドリング手法を定量的に比較・検討できるような指標を導入することが挙げられる。

謝辞

本論文は MEXT/JSPS 科研費 25420448 の助成を受けたものです。ここに謝意を表します。

参考文献

- [1] Ward, M., Grinstein, G., Keim, D., “Interactive data visualization”, A. K. Peters (2010).
- [2] Herman, I., Melançon, G., Marshall, M. S., “Graph Visualization and Navigation in Information Visualization”, IEEE Transactions on Visualization and Computer Graphics, Vol.6, No.1 (2000).
- [3] Gansner, E. R., Koutsofios, E. S., North, C., Vo, K. -P., “A Technique for Drawing Directed Graphs”, IEEE Transactions on Software Engineering, Vol.19, No.3 (1993).
- [4] Mueller, C., Gregor, D., Lumsdaine, A., “Distributed Force-Directed Graph Layout and Visualization”, Proceedings of the 6th Eurographics conference on Parallel Graphics and Visualization, Vol.6 (2006).
- [5] Zhou, H., Xu, P., Yuan, X., Qu, H.: Edge Bundling in Information Visualization: Tsinghua Science and Technology, Vol.18, No.2 (2013).
- [6] Holten, D., “Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data”, IEEE Transactions on Visualization & Computer Graphics, Vol.12, No.5 (2006).
- [7] Zhou, H., Yuan, X., Qu, H., Cui, W., Chen, B., “Visual Clustering in Parallel Coordinates”, Computer Graphics Forum, Vol.27, No.3 (2008).
- [8] Telea, A., Ersoy, O., “Image-Based Edge Bundles: Simplified Visualization of Large Graphs”, Eurographics/ IEEE-VGTC Symposium on Visualization, Vol.29, No.3 (2010).
- [9] Saga, R., Terachi, M., Tsuji, H., “FACT-Graph: Trend Visualization by Frequency and Co-occurrence”, Electronics and Communications in Japan, Vol.95, No.2 (2012).
- [10] Holten, D., Wijk, J. J. van., “Force-Directed Edge Bundling for Graph Visualization”, Computer Graphics Forum, Vol.28, No.3 (2009).
- [11] Selassie, D., Heller, B., Heer, J., “Divided Edge Bundling for Directional Network Data”, IEEE Transaction Visualization & Computer Graphics, Vol.17, No.12 (2011).
- [12] Kamada, T., Kawai, S., “An algorithm for drawing general undirected graphs”, Information Processing Letters, Vol.31, No.1 (1989).

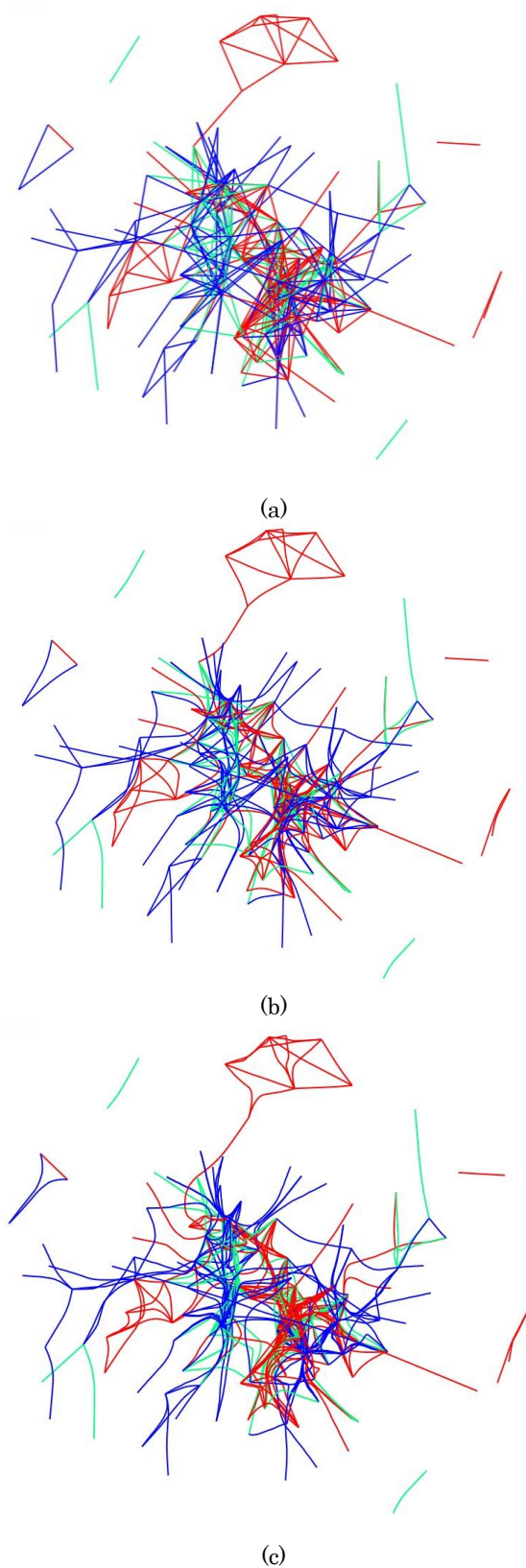


図7 バンドリング結果
 (a) 初期図 (b) FDEB 適用後
 (c) 提案手法適用後

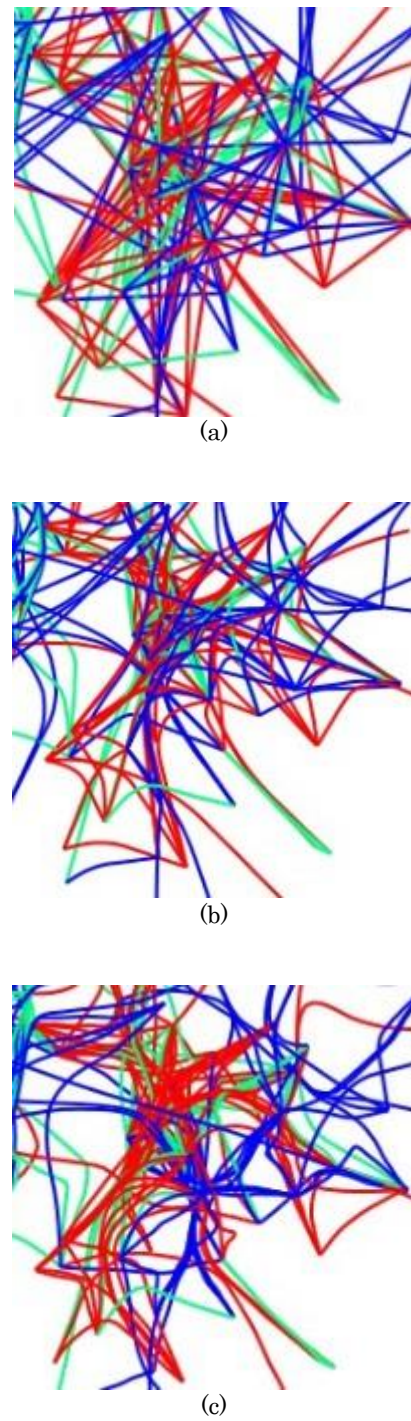


図8 密集部拡大図
 (a) 初期図 (b) FDEB 適用後
 (c) 提案手法適用後