

2部グラフの可読性向上のためのクラスタ構造の動的描画手法

佐藤 修治[†]

三末 和男[†]

田中 二郎[†]

[†] 筑波大学大学院 コンピュータサイエンス専攻

1 はじめに

本論文では2部グラフで表現されるネットワークを可読性の高い表現で描画する手法を述べる．ノードをクラスタリング手法を利用して分類し，先行研究である「アンカーマップ [1]」の描画手法に適用して，クラスタの情報をインタラクティブに表現することで可読性の向上を狙う．

2 前提知識

2.1 2部グラフ

グラフとは，ノードの集合とその接続関係を表すエッジの集合から構成される論理構造であり，物の間の関係を表すのに適している．グラフの内，ノードの集合を二つの排他的な集合 V_1 と V_2 に分割することができ，エッジの集合 E が $V_1 \times V_2$ の部分集合であるようなグラフ G を2部グラフといい， $G = (V_1 \cup V_2, E)$ として表現される．2部グラフは実世界の様々な場面で現れ，「顧客」と「購買商品」の関係，「コミュニティ」と「メンバー」の関係などが例に挙げられる．

2.2 アンカーマップ

グラフの可視化に良く用いられる描画法の一つとして Eades のスプリングモデル [2] がある．スプリングモデルは，エッジにバネを埋め込み安定状態を計算することでノードの配置を求める，グラフ描画のレイアウト手法である．

アンカーマップはスプリングモデルを発展させ，2部グラフの2種類のノードの集合の一方に位置の制約を課した描画スタイルである．本研究での位置の制約は次の2点である．

- 集合 V_1 の要素は円周上に等間隔に配置される
- 集合 V_2 の要素は自由配置で，集合 V_1 の要素との関係を適切に表現する位置に配置される

このとき，集合 V_1 を「アンカー」と呼び，集合 V_2 を「フリーノード」と呼ぶ．エッジはアンカーとフリーノードを直線で接続する形式で表現される．都道府県

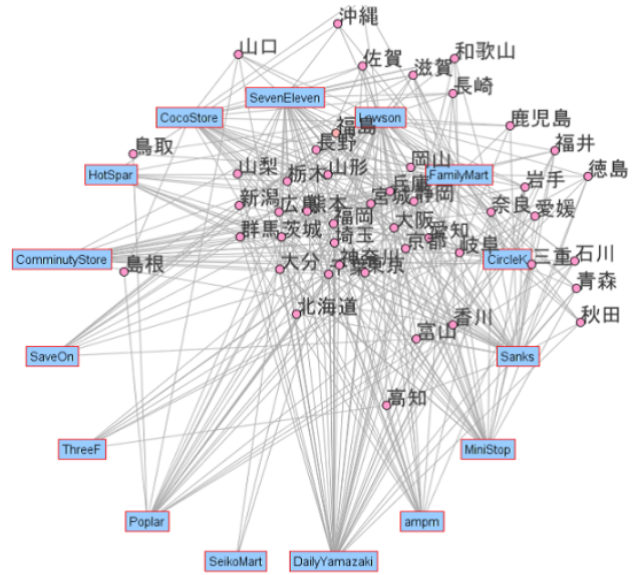


図1 アンカーマップ

とコンビニ店舗数の関係をアンカーマップで表現した図を (図1) に示す．コンビニ店舗をアンカー，都道府県をフリーノードとして描画し，店舗数とエッジのバネの強さを比例させて店舗数が多い都道府県に近くレイアウトされるよう設定している．

グラフ右上にフリーノードが集まっていることから，右上アンカーである「SevenEleven」や「Lawson」は多くの都道府県に店舗を構えていることが読み取れる．また，「東京」と「神奈川」などのように近くにレイアウトされているノード同士は，似た店舗立地傾向を持つと推測可能である．

3 等類似度線によるクラスタ構造の描画

グラフの可視化において，ノード同士の関連性という重要な情報がある．通常ノード同士はエッジにより関連性を把握することが出来るが，2部グラフにおいては片方のノード集合の内部にエッジは存在しない．そのため，これらを適切に可視化することでグラフの意味を読み取りやすくなると考えた．

人間の注視特性より，重要な部分を囲み線などで表す方法が有効であることは良く知られている．そこで，ノードをクラスタリングし，その結果を等高線のように表示することでノード間の関連性を可視化する手法を取った．

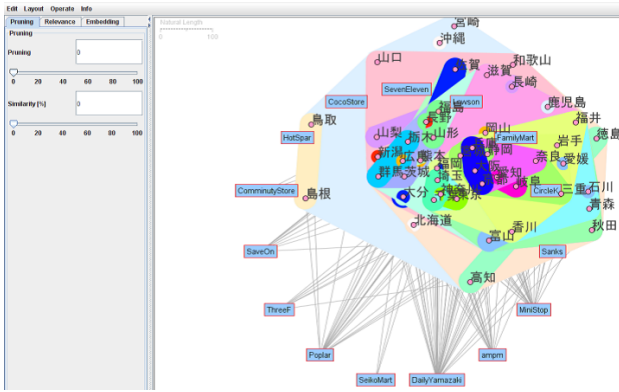


図2 等類似度線描画 (閾値 $t=0.5$)

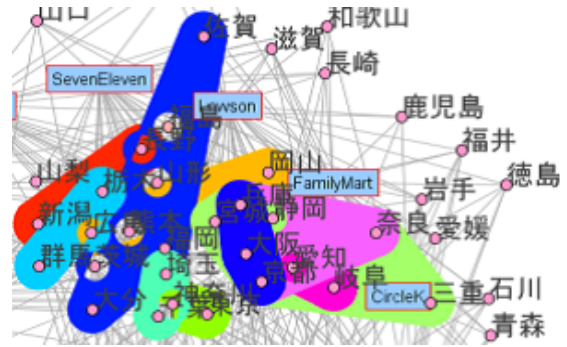


図3 等類似度線拡大

3.1 クラスタリング

クラスタリングとは、異なる性質のもの同士が混在している集合の中から互いに類似したものを集めてクラスタを形成することで、集合を分類する手法である。

本手法で用いたクラスタリングは階層型クラスタリングで、距離の指標にジャカード係数、クラスタリング方法に最短距離クラスタリングを利用している。

類似度 S をジャカード係数を用いて定義した。ノード x が接続するノードの集合を $A(x)$ としたとき、2つのノード $x, y (\in V)$ の類似度 $S(x, y)$ は式 (1) で表される。

$$S(x, y) = \frac{|A(x) \cap A(y)|}{|A(x) \cup A(y)|} \quad (1)$$

全ノードの組み合わせの類似度を求めた後、類似度の高い順に最短距離クラスタリングでクラスタを形成していく。

3.2 等類似度線描画

クラスタリングの情報を効果的に表す手法として、「等類似度線描画手法」の開発を行った。従来のクラスタリングを用いた手法ではノードの集約表示が一般的であるが、それらの手法でノード間の関連性の可読性を向上させることは難しかった [3]。我々の手法はクラスタ構築の際に用いた類似度を利用し、ある一定の閾値 t 以上の類似度で構築されたクラスタに属するノードを閉曲線で描画する [4]。閾値 t の変更はスライダにより動的に可能である。読み手自身の読みやすい度合いで、等高線のような閉曲線をインタラクティブに表示することが出来る。

図2は図1を閾値 $t = 0.5$ で等類似度描画表現を適用した例である。閉曲線内部の色の彩度は類似度と比例させており、色が濃いほど類似度が高く表示される。

4 考察

等類似度線描画手法の特徴は、アンカーとフリーノード間の関連性とフリーノード同士の関連性が同時に認知し易くなることである。既存手法のアンカーマップにおいても、フリーノードの関連性は配置される場所によってその関連性の度合いを得ることが可能であった。本手法ではクラスタリングの結果を可視化することにより、位置は近いが似ていない、もしくは位置は遠いが似ているノード集合を読み取ることが可能であった (図3)。

佐賀県と大分県は位置は異なるが類似度線が100%で描画される。ここから、接続ノードであるコンビニの店舗は同一であるが、店舗数の比率が異なるという新たな知見が得られた。

類似度線の変更のインターフェースにスライダを導入することで、インタラクティブ操作による読み手への適応度の向上がなされている。また、ノード間類似度による数値化表示で、ノード関係が「どれ位似ているか」を読み取る支援が行われていると考える。

5 まとめ

2部グラフの可視化における可読性向上のため等類似度線描画手法を開発した。本手法を用いることでノードの関連性が読み取りやすくなることを示し、グラフの可読性向上において有効であると述べた。

参考文献

- [1] K. Misue. "Drawing Bipartite Graphs as Anchored Maps," In *Proc. of APVIS2006*, pp.169-177, 2006.
- [2] P. Eades. "A heuristic for graph drawing," *Congressus Numerantium*, Vol.42, pp.149-160, 1984.
- [3] Y. Frishman, A. Tal. "Dynamic Drawing of Clustered Graphs," In *Proc. of InfoVis2004*, pp.191-198, 2004.
- [4] 佐藤 修治, 三末 和男, 田中 二郎. "大規模2部グラフの可読性向上のためのクラスタ構造の動的描画," In *Proc. of WISS2007*, pp.19-24, 2007.