

ラベルつきノードを含むネットワークの3次元可視化の一手法

佐々日向子[†] 伊藤 貴之[†]

[†] お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚2丁目1番1号

E-mail: †{hinako,itot}@itolab.is.ocha.ac.jp

あらまし ネットワークの可視化に関する研究は数多く発表されてきたが、大規模なデータに対する可視化結果の複雑さを回避するための課題はまだ残っている。一方で、近年では複合的なデータから構築されたネットワークの可視化が増えており、例としてノードに1個以上のラベルが付与されたネットワークの可視化が課題となっている。本研究では、ネットワークの接続構造とラベル情報の関係を表現するために二つの手法を組み合わせるネットワークの可視化手法を開発している。一つは3次元空間にノード群を配置し、視点移動のための回転操作によって、ユーザが興味をもつ部分に対話的かつ効果的に可視化する手法。二つ目は同一ラベル情報を有するノードの集合をドロネー三角分割法を用いて描画した多角形で表現して重ね描きする手法である。それらに加え本報告では、各ノードに付与されているラベルが一目でわかるように、各ノードのラベルの帰属度を円グラフのように描画する追加機能を報告する。この追加機能により、ネットワークの接続構造とラベルの帰属度分布を同時に可視化できる。本報告では論文共著関係ネットワークを題材として、本手法による可視化の実行例を示す。

キーワード ネットワーク, 3次元可視化, 集合可視化, ラベル

1 はじめに

ネットワークの可視化において、ネットワークの接続構造とラベル情報を同時に可視化したい場合や、一つのノードに複数のラベルを付与して可視化をしたい場合がある。例えばソーシャルネットワークでは、ユーザ間の友人関係が複雑なネットワークを形成している。多くの場合において、それぞれの人物は、職場、親戚、趣味などの多種多様なコミュニティに同時に身を置いている。このコミュニティと友人関係との相関を可視化することで、ソーシャルネットワークの分析に大いに役に立つことが期待される。また、コミュニティの関係を分析することで、ソーシャルネットワークにおける情報の拡散の分析やレコメンド機能の精度の向上に利用できると期待される。本研究では、コミュニティと友人関係の相関の可視化を実現するために、ネットワークの3次元可視化手法と集合の2次元可視化手法を組み合わせる新たな可視化手法を開発している。

ネットワーク可視化の研究には長い歴史があり、小規模なネットワークであれば既に視認性の高い可視化を実現できている。一方でその多くの手法では、ネットワークの接続構造にもとづいたクラスタリングによって抽出されたコミュニティを可視化している。さらに、その多くの手法では、ノードが排他的に1つのコミュニティに属するようなクラスタリングが適用されている。このような手法では、各ノードが複数のコミュニティに属するような状況を表現できない。さらに、ネットワークの接続構造に現れない潜在的なコミュニティを表現することもできない。

また、ノード数やエッジ数の多い大規模なネットワークを可視化する際には、ノードの重なりやエッジの交差が生じて視認性が低下するという難点がある。この問題に対して旧来からの

解決策として、力学モデルや次元削減の導入によってノードの重なりやエッジの交差を避けたレイアウトを実現する手法が議論されている。しかしそれらの手法をもってしても、2次元空間であるディスプレイ上でのネットワーク可視化での視認性には限界がある。また、ネットワーク可視化に関する習熟度や専門知識のないユーザを想定した場合には、ネットワークのどこを読み取ったら欲しい情報が読み取れるのかという可読性の面でも課題が残っている。

これらの問題を解決するために、我々は以下の2種類の可視化を組み合わせる手法を提案している [1]。

- ネットワークを3次元空間に配置して対話的な視点操作とともに可視化する手法。
- ネットワークの接続構造にもとづいたクラスタリングを適用せず、ノードにあらかじめ付与されたラベル情報にもとづいて2次元空間上でノードの集合を可視化する手法。

ネットワークの接続構造とは独立なラベル情報によって集合を形成し、これとネットワークの接続構造を重ね描きするように表示することで、集合を構成する各ラベルがどのような接続構造につながっているか、といった点を視認できる。

本報告ではこの可視化手法の改良について報告する。本手法ではノードに複数のラベルを付与できること、そして各々のラベルには帰属度が実数で与えられていることを想定する。言い換えれば、 m 種類のラベルがあるデータにおいて、個々のノードには各々のラベルへの帰属度が m 次元ベクトルで付与されているとする。このとき本手法では、ノードに付与されているラベルの帰属度を円グラフのように描画する。また本手法では、特定のラベルを付与されたノードの集合を多角形で描画するために、2次元のスクリーン座標系で当該ノードを接続するドロネー三角メッシュを生成し、その領域境界を描くことで集合を表現することにした。この実装により、特定のラベルを付与さ

れたノードが集中する領域を以前よりも的確に表現できるようになった。

本報告の構成は以下のとおりである。2章では関連研究について述べる。3章では提案手法について、4章では本手法の実行結果と考察を述べる。そして5章では本研究のまとめと今後の課題、展望について述べる。

2 関連研究

2.1 ノードに追加情報を付与されたネットワークの可視化

タグや特徴量などの追加情報をノードに付与されたネットワークの2次元可視化手法は、従来からいくつか提案されている。

伊藤ら[2]による可視化手法は、ネットワークを構成するノードをクラスタリングする際に、多数のエッジを持つ重要なノードをクラスタから分離しやすくするようなノード間距離を定義することで、このような重要なノードに着目しやすい可視化結果を得る。従来の手法ではネットワークをクラスタリングすることで重要なノードが大きなクラスタに分類され、見つけにくくなってしまったことがあったが、この手法ではあえて重要なノードを独立させることで重要なノードを見つけやすくしている。また、この手法ではノードに特徴量ベクトルが付与されていることを想定し、そのベクトル間距離をクラスタリングに加味することもできる。

伊藤ら[3]による別の可視化手法では、力学モデルと空間充填モデルを併用することにより、各ノードに与えられたアイテムとネットワークの接続構造を同時に、かつ良好に配置できる。この手法ではノードに階層的クラスタリングを適用することで共有のアイテムを所持するノードを近くに配置するようにしている。この手法では共通のアイテムを付与されたノードが近くに配置されるような引力を与えるため、アイテムの数の増加とともに急激に計算時間が増加するという問題点がある。また、この手法では本報告の提案手法と同様に、ノードに付与されたアイテムに応じて円形のノードを円グラフのように描画している。しかし、伊藤らの手法ではノードに付与された各アイテムの帰属度は考慮せず、ノードに付与されたアイテムの数だけ円を均等に分割し、等しい大きさの扇形で表現している。

2.2 ネットワークの2次元可視化と3次元可視化

ネットワークの可視化手法は2次元空間にノードを配置する手法が圧倒的に多いが、3次元以上の空間にノードを配置する手法もいくつか発表されている。ネットワークの2次元可視化と3次元可視化の可読性を実験によって比較した論文もいくつか発表されている。Wareら[4]の実験では、ネットワークを読み取る所要時間とその誤り率から可読性を検証した。結果として、ノード数が100~140個の場合は3次元可視化が2次元可視化よりも所要時間が短かったことを示している。また、2次元可視化では被験者による読み取りの誤り率が最大40%であったのに対して、3次元可視化では最大20%であった。さらにWareらは3次元可視化における回転操作についても検証し

ており、2次元可視化の操作に加えて自動回転、手動回転、頭の動きに同期した回転、そしてそれぞれ立体視メガネを装着しているか否かの要素を加えての8つの実装を比較した。その中でも最もネットワークの読み取り時間が短かったのは、立体視メガネをかけずに手動回転させる方法だった。このことから、手動操作の3次元可視化に有用な面があることがわかる。

2.3 集合可視化

各ノードが1個以上の集合に属するデータの可視化を集合可視化(Set Visualization)という。集合の2次元可視化手法として、CollinsらによるBubbleSets[5]、AlperらによるLineSets[6]などがあげられる。

BubbleSetsは特定の集合に属さないノードを避けるように仮想エッジを配置することにより、ノードの配置にかかわらず連続した輪郭で集合を描画できる手法である。対象のノードを曲線で囲むことで集合を表現をするBubbleSetsに対して、LineSetsは同一集合に属するノードを滑らかな曲線で接続することで集合を表現する。両手法ともに、散布図、木構造、ネットワーク、地図上などの多種多様な可視化画面への重ね描きが可能である。さらに、特定のラベルを持つノードを最大限に含み、かつそのラベルを持たないノードが含まれる数を最小限に抑えて1つの集合として描画することができる。しかしノードの数が増えれば集合は複雑な形になり、視認性に欠けてしまう。加えてノード数の増加とともに加速的に計算量が増え、大きな計算時間を要するため、対話操作をとまうリアルタイムな集合可視化に向かない面がある。

Bubble SetsやLine Setsよりも対話性の高い集合可視化を実現する一手段として、もっと単純な形状で集合を表現することが考えられる。我々は凸包を用いた集合可視化を試みている[1]。凸包は対象となる全てのノードを囲む最小の凸多角形である。凸包の生成にはギフト包装法やQuickHull[7]といった数多くのアルゴリズムが知られており、BubbleSetsやLineSetsよりも計算時間が圧倒的に小さい。しかし、本研究が対象とするネットワークの可視化結果に凸包を適用すると、集合に含まれないノードも内包する大きな図形で集合を描画してしまうことが多く、適切な集合表現とならない場合が多い。

渡邊ら[8]は散布図状の点群の形状評価にドロネー三角分割法を用いている。ドロネー三角分割法とは与えられた点群から三角メッシュを生成する手法の一つである。ドロネー三角分割法は生成する三角形の最小角度が最大になるように三角メッシュを構成するという特徴を持っている。渡邊らの手法では、散布図上の全ての点を含む三角メッシュを生成し、閾値以上の辺を取り除くことで散布図の形状の計算をしている。本報告の提案手法ではこの考えを適用し、ドロネー三角分割法によって生成される多角形領域によって集合を表現する。

3 提案手法

本章では提案手法の処理手順について述べる。図1に本手法の処理手順を示す。本手法ではまずネットワークを3次元空

間上に構築し、ネットワークを見る視点を決定する。その視点から得られるネットワークを2次元に投影し、その上に集合を重ねて描画する。また、視点を変更するごとに集合の描画を反復して行う。3.1節ではネットワークの3次元可視化について、3.2節では集合の描画についての詳細を説明する。

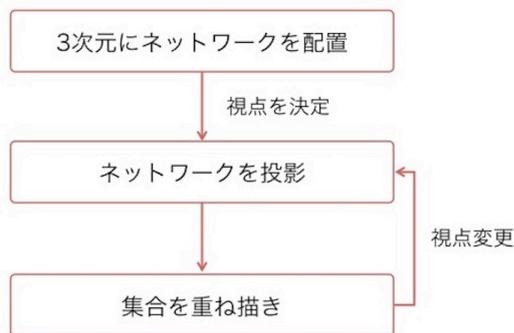


図1 処理手順

3.1 ネットワークの3次元可視化

本手法では与えられたノード、エッジ、ラベルのデータセットからネットワークを構築し、これを3次元空間に配置する。本手法では伊藤ら[2]の手法に実装されたクラスタリング手法を適用し、他ノードへの接続が多い重要なノードを他のノード群から距離をおいて配置されるようなクラスタリングを施す。続いて本手法ではネットワークの接続構造からクラスタ間距離を算出し、これに多次元尺度構成法(MDS: Multi-Dimensional Scaling)を適用することで各クラスターの3次元空間上の位置を算出する。

ノードとエッジを描画する際に、全てのエッジを描画するとエッジが絡まり合って視認性を損なうことが多い。そこで我々の現時点での実装では、対話操作によってユーザーに指定されたノードに接続されたエッジのみを描画している。また、ノードに付与されたラベルを視認しやすいように、ラベルに対応した色でノードを色付けする。このときノードを円形で表示し、各ラベルの帰属度に比例した中心角を各ラベルに与えることで、各ラベルに対応する色の扇形を描画する。

3次元可視化を採用した理由として、前章で紹介したWareら[4]の実験結果として3次元可視化が読み取り時間と読み取り誤差の面で優れていたから、という点があげられる。別の理由として、3次元可視化のほうが計算時間をかけることなく多様な配置を試すことができる。ネットワークや集合の可視化結果が視認性に欠けるので可視化をやり直したいと判断した場合、2次元可視化ではネットワークの再配置が必要となるが、3次元可視化では視点移動によって異なる配置結果を試すことができる。

3.2 集合の描画

続いて、特定のラベルを付与されたノードの集合をドロネー三角分割法を用いて描画された多角形で表現し、可視化されたネットワーク上に重ねて描画する。まず特定のラベルを付与さ

れたノードの集合を抽出し、それらを接続するドロネー三角メッシュを生成する。三角メッシュの生成には、スクリーン座標系での2次元座標値を用いる。続いて、三角メッシュを構成する辺の中から、閾値を上回る長さの辺を削除し、さらに削除された辺を所有する三角形を削除する。この処理のあとに残留した三角形群の領域境界を描くことで、集合に相当する多角形を描画する。また、以上の処理によって他のノードとの接続関係が切れて孤立状態となったノードに対しては、そのノードの上に円を重ねて描画する。我々が以前に実装した凸包を用いた手法[1]では、必要以上に大きな多角形を描画してしまうという問題点があったが、この手法によりそれが改善されている。

なお本手法では、ネットワークの3次元可視化結果に対して手動回転操作によって視線方向を制御することを想定している。そこで我々の実装では、回転操作のたびに集合可視化のためのドロネー三角分割法を反復する。そのため、対象となるノード数が増加しても高速に集合を描画できる必要がある。ドロネー三角分割法の計算時間はノード数 n に対して $O(\log n)$ となることが知られており、我々が想定する数千ノード規模のデータにおいて不自由ない回転操作が可能である。

ここで、全てのラベルに対応する全ての多角形を同時に描画すると、複数の多角形が重なることで視認性が低下することがある。ゆえに現時点での実装では、ユーザーが選択した特定の多角形のみを描画することで視認性を維持する。

4 実行結果

我々は本手法をJava Development Kit (JDK) 1.8により実装した。なおネットワークの3次元可視化については伊藤ら[2]が提案したKoala[2]の実装を拡張する形で実装した。

本報告では一例としてNERC Biomolecular Analysis Facility (NBAF) 2の出版書誌に基づいて作られたデータセットを可視化した例を示す。このデータセットは、1821人の著者をノード、11097個の共同著者関係をエッジとした論文共著ネットワークである。伊藤ら[2]はこのデータセットに対して、各著者が執筆した論文のタイトルに含まれる頻度の高い11個の単語を選択し、それらの各単語の頻度から各ノードについて11次元の特徴ベクトルを算出した。そして特徴量ベクトルを構成する11個の値のうち最大値となる次元に対応する単語にもとづいて、各ノードに色を割り当てた。本実験ではこの11個の値のうち、一定以上の大きさとなる値の次元に対応する単語をラベルとして扱う。

図2は各ノードの色とラベルとなった用語の対応表である。赤がGenetic(遺伝子)、オレンジがMolecular(分子)、黄色がLoci(遺伝子座)、黄緑がMicrosatellites(マイクロサテライト)、緑がIsolation(隔離)、青緑がInbreeding(同型交配)、水色がTranscriptions(DNAの転写)、青がExpression(発現)、藍色がBacterial(微生物)、紫がBreeding(育種)、そしてピンクがPolymorphic(多型)をそれぞれ示している。また、黒色のノードはラベルを一つも付与されていないノードである。

色	単語
■ (赤)	Genetic
■ (オレンジ)	Molecular
■ (黄色)	Loci
■ (黄緑)	Microsatellites
■ (緑)	Isolation
■ (青緑)	Inbreeding
■ (水色)	Transcriptomics
■ (青)	Expression
■ (藍色)	Bacterial
■ (紫)	Breeding
■ (ピンク)	Polymorphic

図2 各ノードの色とラベルの対応表

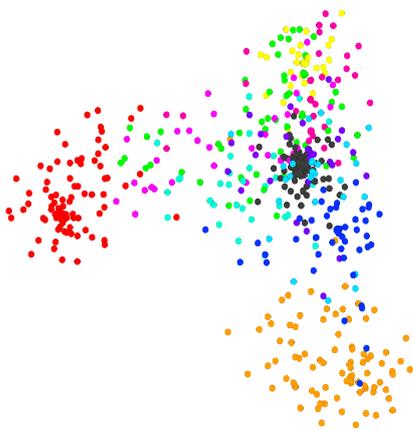


図3 ラベルの描画前のネットワーク

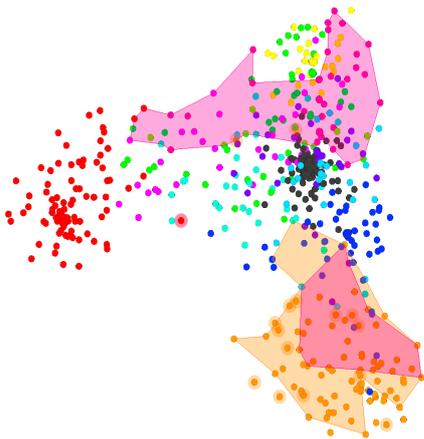


図4 ラベルの描画後のネットワーク

図3はラベルの集合を描画する前のネットワーク可視化結果である。また、図4は図3のネットワーク可視化結果の上にラベルの集合を重ねて描画した例である。なお、多角形の外周上にあるノードは集合に含まれるノードである。図3のような描画ではノードの色でしかラベルを判別できない。そのため、黄緑と緑、青緑と水色や、紫色とピンクなど、近い色相で描画さ

れた小さなノードのラベルを識別するのは難しく、誤解を招きやすい可視化になっている。また、この図からは各ノードに対して単一のラベルしか読み取れない。しかし図4のように本手法の集合可視化を施すことで、図3と比べてオレンジ色とピンク色の分布に注目することができる。さらにこの二つの集合が重なっていることで、この二つの分野の関係が深いことが推測できる。

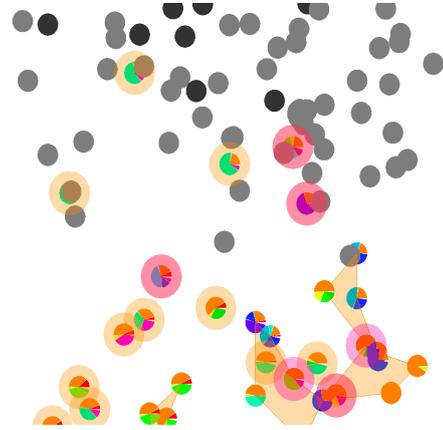


図5 円グラフのように描画されたノード

図5は図4を拡大し、付与されている帰属度によって円グラフのようにノードを描画した結果である。灰色のノードは選択されたラベルを付与されていないノードである。この機能を使用することにより、複数のラベルを選択表示していてもどのノードがどんなラベルを付与されているのが一目でわかる。例えば、図5のピンク色の円で囲われたノードは円グラフを見るとピンク色とオレンジ色両方のラベルを付与されているのがわかる。さらに、それらのほとんどがピンク色の占める面積よりもオレンジ色の面積が大きく、これらのノードにあたる著者の論文では「分子」(オレンジ色に対応)の方が「多型」(ピンク色に対応)よりも出現頻度が高いことが推測できる。しかし、手前側のノードに遮蔽されているノードがいくつか発生してしまった。



図6 図3を回転した図

図6は回転操作により別視点で見たときの可視化結果である。

手動で回転操作しながら目視観察することで、奥行きへの把握が容易になる。しかし、本実験においてはネットワークの配置が薄くなってしまい、別視点から集合を見る利点がありませんでした。ネットワークの配置が薄くなる原因は現在利用している MDS のライブラリの実装、もしくはプログラムのミスがあると考えられる。

5 まとめと今後の課題

我々は 3 次元空間に配置されたネットワークの上にラベル情報を表す多角形を重ねて描画することで、ネットワークの接続構造とラベル分布を同時に可視化する手法を開発している。本報告では、ラベル情報の表示に関する改良について報告した。

ネットワークの 3 次元可視化において、最も効果的な可視化結果を得られる最適な視点を探さないといけない、という問題が生じる。その問題を克服するための課題として、視点推薦の機能を開発したいと考えている。ここで可視化結果の効果を測る一手段として、集合を表現する多角形の形状に関する評価関数を設定することを考える。そして例えば特定のラベルの集合を可視化したい場合、特定の視点から描画される多角形の形状評価値を計算する。これらの操作を複数の視点に対して反復し、最も形状評価値が高かった視点からの可視化結果を描画する。この機能を実装できれば、より少ない操作で有益な可視化結果が得られると考える。

これらの機能を実装した後に、多様なデータセットを本手法に適用したい。現在ユーザのフォロー関係とユーザがリツイートした政党の公式アカウントによるツイートの情報のデータを使った可視化を試みている。これにより、特定の政党のツイートをリツイートしたユーザが他にどんな政党のツイートをリツイートする傾向があるのか、また特定の政党のツイートをリツイートした人はどんなユーザをフォローしているのかを分析し、何か知見が得られるのではないかと考える。

文 献

- [1] 佐々, 伊藤, ネットワークの 3 次元可視化と集合可視化の重ね描きの一手法, 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM), E5-4, 2018.
- [2] 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.
- [3] T. Itoh, C. Muelder, K. Ma, J. Sese, “力学モデルと空間充填モデルの併用によるアイテム集合付きネットワークの可視化手法,” *研究報告グラフィクスと CAD (CG)*, Vol.137, No. 14, pp. 1 - 6, 2009.
- [4] C. Ware and G. Franck, “Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions,” *ACM Transactions on Graphics*, Vol. 15, No. 2, pp. 121-140, 1996.
- [5] C. Collins, G. Penn, S. Carpendale, “Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations,” *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1009 - 1016, November/December 2009.
- [6] B. Alper, N. H. Riche, G. Ramos, M. Czerwinski, “Design Study of LineSets, a Novel Set Visualization Technique,”

- IEEE Transactions on Visualization and Computer Graphics*, Vol. 17, No. 12, pp. 2259 - 2267, December 2011.
- [7] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, “The Quickhull Algorithm for Convex Hulls,” *ACM Transactions on Mathematical Software*, Vol. 22, No. 4, pp. 78 - 83, 1996.
 - [8] A. Watanabe, T. Itoh, M. Kanazaki, K. Chiba, “A Scatterplots Selection Technique for Multi-Dimensional Data Visualization Combining with Parallel Coordinate Plots,” *International Conference Information Visualisation (IV)*, Vol. 21, pp. 469-483, 2017.