

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

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。本報告では、ノードに付与されたラベル情報の可視化に着目し、ネットワークの3次元可視化と集合を表す凸包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情報を同時に可視化する手法を提案する。本手法ではまず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を対話的かつ効果的に可視化する。加えて、同一ラベル情報を有するノードの集合を凸包で表現して重ね描きすることで、ネットワークの接続構造とラベル情報の関係性を表現する。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, 凸包, 集合可視化

1. はじめに

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た新たな可視化手法を提案する。

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。

そこで本報告では以下の2種類の可視化を組み合わせた手法を提案する。

- 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。なお、提案手法ではノードに複数のラベルを付与できることを想定する。これにより、ノードが複数のコミュニティに属している状況を表現できる。現時点での実装では各集合を単純に凸包で描いているが、集合の可視化手法は多数提案されているので、それらを適用することも可能である。

さらに我々の実装では、ネットワーク中のユーザが注目したい局所を重点的に可視化し、それ以外の集合やエッジの描画を対話的に省略できるようにする。これによって可視化結果を単純化できるので、視認性を維持しやすくなる。このような機能によって、可視化の工程に慣れていないユーザが、特定の集合に関する分布や接続構造との関係を理解しやすくなると考えられる。

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。

Shneiderman ら [1] による手法では、ネットワークの大局的な接続構造を可視化するのではなく、ノードに付与されているタグの種類や数値にもとづいてノードを配置することでネットワークを可視化している。ユーザの対話的操作によりノードのタグなどを指定することで、興味をもったノード周辺の接続構造のみを描画することを可能にしている。これにより、従来のネットワーク可視化が抱えるノードとエッジの描画結果の複雑さを緩和している。しかしこの手法では、ある程度の大規模なネットワークも描画が可能である反面、タグによってノードの配置領域を決定するため、タグの種類が多くなるにつれて視認性を損なうことがある。特に、各ノードに複数のタグが付与されているようなデータセットはこの手法は不向きである。また、この手法は注目ノードをユーザ自身の操作で選択するため、データの内容についての知識がないユーザにはどれが重要なノードなのかをこの可視化からは読み取ることができない。

伊藤ら [2] による手法は、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分離させることで、このような重要なノードに着目しやすくする可視化手法である。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。

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 [3] 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 [4]、Alper らによる LineSets [5] 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という問題点もある。

我々の現時点での実装では暫定的に、凸包によって集合を描画している。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。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ように集合を描画してしまうという難点はある。一方で、突合の生成にはギフト包装法や QuickHull [6] 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。

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。本手法の手順は大きく分けて2つの段階で構成される。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説明する。

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。

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 [3] 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凸包で表現し、可視化されたネットワーク上に重ねて描画する。現時点での実装では暫定的に、集合の可視化に凸包を適用している。なぜなら本研究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おり、回転操作のたびに集合可視化を反復する。そのため、対象となるノード数が多くなっても高速に集合を描画できる必要があるからである。現時点での実装では、QuickHullのアルゴリズムにより生成された凸包を描画している。

ここで、全てのラベルに対応する全ての凸包を同時に描画すると、複数の凸包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、ユーザが選択した特定の凸包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が提案したKaola [2]の実装を拡張した。

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12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12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12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12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1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。

図2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3は図2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、凸包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2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

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1 各ノードの色とラベルの対応表

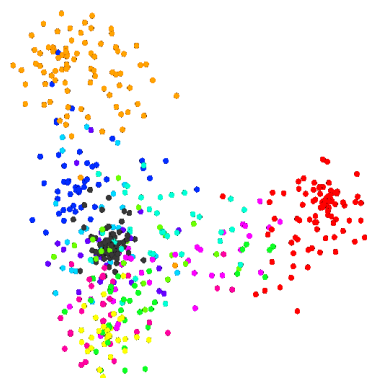


図2 ラベルの描画前のネットワーク

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3のように凸包による集合可視化を施すことで、図2と比べて赤いノードの分布を把握しやすくなる。さらにこの凸包の内部にはオレンジ色、ピンク色、水色などのノードが含まれており、赤色に対応する論文の分野がこの3色に対応する分野とも深く関係することが示唆される。

図4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、奥行きが把握が容易になる。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得られなかった。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に一因があると考えられる。

5. まとめと今後の課題

本報告では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凸包を描画することで、ネットワークの接続構造

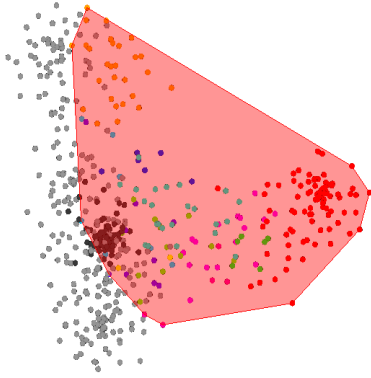


図3 ラベルの描画後のネットワーク

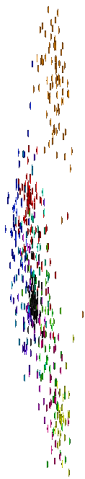


図4 図3を回転した図

とラベル分布を同時に可視化する手法を提案した。

今後の課題として、各ノードに複数のラベルを付与する機能が実装できていないので実装したい。複数のラベルを付与し複数の凸包を重ねて表示することで、各ノードが複数の集合に所属していることを明示することができる。さらには、複数の集合にまたがる重要なノードを手掛かりにしてラベル間の関係を視認しやすくなるであろう。これに関連する描画上の工夫として、ノードを単色の円で表現するのではなく、ラベルの帰属度に沿って円グラフ表示する手法を考えている。

ネットワークの3次元可視化において、一番効率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。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凸包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。例として、TwitterなどのSNSから取得したデータでネットワークを構築し、人物間の友人関係と所属コミュニティの可視化を実現する方向で実装を進めたい。

文 献

- [1] B. Shneiderman, A. Aris, "Network Visualization by Semantic Substrates," IEEE Transactions on Visualization and Computer Graphics, Volume 12, No. 5, pp. 733 - 740, 2006.
- [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] 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.
- [4] 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.
- [5] 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.
- [6] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," ACM Transactions on Mathematical Software, Vol. 22, No. 4, pp. 469-483, 1996.