

反復的なグラフサンプリングによる巨大ネットワークの局所可視化

鳥山 菜海子 (指導教員：伊藤貴之)

1 はじめに

情報可視化の研究では、“Overview first, then zoom and fileter, details on demand” とあるように、最初に全体を可視化し、徐々に局所構造に着目していく手法が一般的とされている。大規模ネットワークの可視化においても、最初にネットワーク全体を可視化したのち、局所構造に注目していくという操作手順を用いた論文が多数発表されている。しかし、全ての可視化でこの手法が最良であるとは限らない。例えばソーシャルネットワークのデータを可視化する場合、データの所有者にとって興味深い現象は、ソーシャルネットワークデータの全体像よりも、局所部分の構造や属性であることが多い。そこで本研究では、ネットワークのサンプリングを反復して多数の部分ネットワークを作成し、それぞれのネットワークを可視化する手法を研究している。具体的には、巨大ネットワークデータからグラフサンプリングを反復し、多数の部分ネットワークを抽出したのち、ネットワークを小さなアイコンとして表現する。

実行例として、各政党の Twitter 公式アカウントおよびそれらの発言をフォローしたアカウントで構成されるネットワークを可視化した事例を示す。

2 関連研究

大規模ネットワークを効果的に可視化するために、重要な局所ネットワークをアイコン化する可視化手法がいくつか知られている。Yoghourdjian ら [1] は、タンパク質結合ネットワークの大局的構造をアイコン表現する試みを発表している。また Chen ら [2] の手法では、ユーザ定義による局所ネットワークを基準にして、グラフの大局的構造を抽象化して可視化する手法を提案している。この手法では、ノードの構造情報をベクトル化してノードでエンコードすることで、局所ネットワークの多様な表現を可能にしている。他にもいくつかのネットワーク可視化手法が知られているが、我々の手法は「数十個・数百個単位のアイコンで表現された多数の部分ネットワーク群をその特徴によって分類表示する」ことに重点を置いたという点でこれらと異なる。

3 処理手順

3.1 グラフサンプリング

本手法ではグラフサンプリングを反復することで多数の部分ネットワークを抽出する。現在の実装では Hu ら [3] が紹介する手法のうち、単純な幅優先探索にもとづくグラフサンプリングを採用している。探索の初期位置となるノードには、多数のノードからリンクされている重要度の高いノードを採用している。こちらはリンク情報の入ったデータを一つずつ探索していき、リンク数が多いノードをまとめたデータをあらかじめ作成している。これにより、重要度の高いノードの周辺に構築される重要なネットワーク構造が高い確率で抽出される。

3.2 部分ネットワークのアイコン表示

続いて、グラフサンプリングを反復して生成した多数の部分ネットワークを、アイコンとして可視化する。現状の実装では単純に、特徴量から円グラフを生成している。本手法においてアイコンが表現したい情報は、主に「部分ネットワークの接続構造」と「特徴量の大小関係」である。しかし、円グラフでは部分ネットワークの接続構造がわからない問題がある。そのため、アイコンデザインの例として、(1) 円グラフ以外の形で特徴量を表現する、(2) 部分ネットワークの接続構造を概略的にあらわすアイコンを生成する、などの検討を今後の課題としたい。

3.3 アイコン集合の画面表示

上述の手法により算出された各アイコンの特徴量に対して、次元削減とクラスタリングを適用し、その結果を階層型データ可視化手法によって画面配置する。現時点での実装では、各ノードの特徴量の平均値をアイコンの特徴量とし、これに k-means 法によるクラスタリングを適用している。続いて、クラスタリング結果から得られた階層構造を、伊藤らのデータ可視化手法 [4] で画面配置している。我々の実装では図 1 に示されるように、ウィンドウの左半分に階層構造を表示している。

3.4 部分ネットワークのグラフ配置

我々の実装では、アイコン集合のうち一つをクリック操作によって指定すると、そのアイコンに対応する部分ネットワークをウィンドウの右半分に表示する。部分ネットワークの表示に関しては、現在の実装では、伊藤らの画面配置手法 [5] を採用して、ネットワークを画面配置している。この手法では、一般的なグラフ構造に加えて、各ノードは特徴量に相当する多次元ベクトルを有するものとする。そして

- リンクで連結された共通ノードを多く有する 2 ノードが小さい距離を有する
- 特徴量の非類似度が小さい 2 ノードが小さい距離を有する

という 2 条件に沿って算出されたノード間距離にもとづいて、ノードのクラスタリングを実行し、クラスタを単位として力学モデルおよびラプラシアンスムージングの 2 段階処理によってノードを画面配置する。このクラスタリング手法は、多くのノードに連結された重要なノードをクラスタから分離する働きが強いことから、ソーシャルメディアの可視化においてインフルエンサーなどの重要なノードとそれ以外のノード群の関係を表示するのに向いている。

4 実行例と考察

本章では実行例として、各政党の Twitter 公式アカウントおよびそれらの発言をフォローしたアカウントで構成されるネットワークを可視化した事例を示す。

ネットワークのノードは Twitter アカウント, エッジはフォロー関係を表しており, 政党公式アカウントは幸福実現党・自由民主党・公明党・新党大地・自由党・社会民主党・日本共産党・日本のこころ・日本を元気にする会・維新の党・民進党・希望の党・立件民主党の 13 党で, 可視化時には各党には固有の色が与えられている. 各アカウントは 13 政党の各々の公式発言を何回リツイートしたかによって得られる 13 次元ベクトルが特徴量として与えられ, 各アイコンにはこれらの平均値を特徴量として用いる.

また本実験では, 部分ネットワークのノード数を 5000 に固定している. 今後の課題として, 接続構造に応じて部分ネットワークのノード数を可変にしたい.

これらの部分ネットワークの集合をクラスタリングしてアイコンの集合で表示した結果を図 1 に示す. この可視化例ではアイコンを 6 個のクラスタに分割している. クラスタ数を動的に変更して, その都度クラスタリングを再実行することもできる. アイコンの 1 個をマウスオーバーすると, 対応する部分ネットワークを画面右側に表示する.

図 1 の左側をみると, 各クラスタが同様な配色の円グラフ群で構成されていることがわかる. 例えば左下には, 立憲民主党と共産党の発言を熱心に拡散するユーザーで構成された部分ネットワークが多数あることがわかる. このような表現によって, 部分ネットワーク群の特徴量 (この事例では政党別の拡散回数で構成されたベクトル) の分布を把握し, 拡散の構造を画面右側に表示された部分ネットワークで確認することができる. 図 2 にいくつかの例を示す. この例を見ただけでも, 立憲民主党と共産党の発言の拡散には多様なパターンが存在することが示唆される.

5 まとめと今後の課題

本研究では, ネットワークのサンプリングを反復して多数の部分ネットワークを抽出し, これらをアイコンで表現し, 階層的に可視化する手法を提案した.

今後の課題として, アプリケーション固有のデータの特性や可視化の工程などを考慮したサンプリング手法の導入を検討したい. また, 本手法で抽出している局所ネットワークのノード数が固定となっている点も, 適切な判断基準で探索を打ち切ることで, ネットワーク中の興味深い部位を適切なサイズで抽出できるようにしたい.

もう一つの課題として, ネットワークの接続構造と特徴量を両方同時に表現するアイコンを表現できるようにしたい.

また, 部分ネットワークのグラフ配置の実装についても課題がある. 現状の実装では, クラスタリングと画面配置に計算時間を要することから, あらかじめ前処理としてこれらの計算を済ませ, 配置結果をキャプチャ画像として保存している. そして実行時には, クリック操作に応じてキャプチャ画像を表示させている. こちらも対話的操作になるよう改良を進めたい.

これらの改良を進めた上で, データ領域の専門家 (例えば Twitter データ分析の専門家) と議論を進め, 巨大ネットワークの分析ツールとしてふさわしい方向に研究を進め, また Twitter の発言拡散ネットワーク以外にも適用事例を拡げたい.

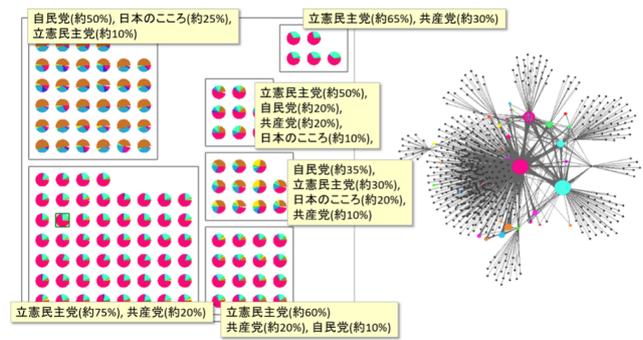


図 1: 可視化例.

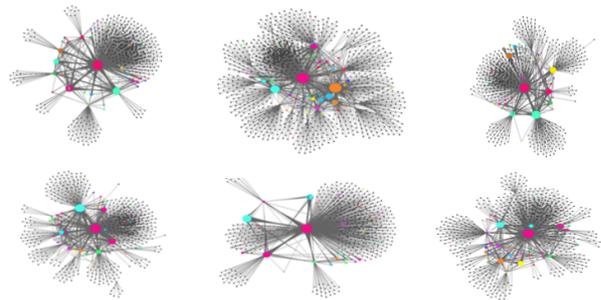


図 2: 「立憲民主党 (約 75%), 共産党 (約 20%)」の特徴量を持つ部分ネットワーク 6 個を表示した例.

6 謝辞

Twitter ネットワークデータを提供して下さった豊橋技術科学大学吉田光男氏に感謝の意を表します.

参考文献

- [1] V. Yoghoudjian, et al., Graph Thumbnails: Identifying and Comparing Multiple Graphs at a Glance, IEEE Transactions on Visualization and Computer Graphics, Vol. 24, No. 12, pp. 3081-3095, 2018.
- [2] W. Chen, et al., Structure-Based Suggestive Exploration: A New Approach for Effective Exploration of Large Networks, IEEE Transactions on Visualization and Computer Graphics, Vol. 25, No. 1, pp. 555-565, 2019.
- [3] P. Hu, W. C. Lau, A Survey and Taxonomy of Graph Sampling, arXiv, 1308.5865, 2013.
- [4] T. Itoh, et al., Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, IEEE Computer Graphics and Applications, Vol. 26, No. 2, pp. 40-47, 2006.
- [5] 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.