

複雑ネットワークの可視化手法について

Method of making to visible of complex network

中村 将人^{*1}
Masato Nakamura

木村 昌臣^{*2}
Masaomi Kimura

^{*1} 芝浦工業大学大学院工学研究科電気電子情報工学専攻

^{*2} 芝浦工業大学工学部情報工学科

Shibaura Institute of Technology Graduate School of Engineering Electrical Engineering and Computer Science #1
Shibaura Institute of Technology College of Engineering Information Science and Engineering #2

A complex network is composed of a lot of nodes and the edges. Since the structure of the network is too complex, it is not easy to be visually understood. In the present study, Fruchterman Reingold method is combined with the coherent hierarchy clustering method. This is continuously present the view from the entire to the details.

1. 背景と目的

複雑ネットワークには多くのノードとエッジが含まれ、またその構造は複雑であるため、これを視覚的に把握するのは容易ではなく、これまで多くの可視化技術が研究されてきた。各ノードの二次元配置方法についての研究は枚挙にいとまがないが、先行研究[1]では、効率的な配置方法を提案した上である属性が同じノードの集合をグループとし、さらにグループのグループを定義することにより階層構造を作成することによりネットワーク構造の把握を支援する仕組みが提案されている。しかし、予めノードに属性が与えられない場合には階層構造を作るのが難しく、また各階層が同じレベルのものであるかは自明でない。本研究では、大量の情報をサマライズする方法として地図の拡大縮小方法を参考にした。国レベルから地域の詳細までを縮尺に応じて表示し分けることと同様に、ネットワークの全体から詳細までを少ないパラメータ操作で行えるとネットワークの任意の詳細度による把握が容易になると考えられる。本研究では Fruchterman Reingold 法 (以下、FR 法) と凝集型階層的クラスタリング手法を組み合わせることでこの可視化を実現する手法を提案する。

2. 提案手法

本手法では、複雑ネットワークのノードの二次元配置を FR 法によって行う。互いに近いノードのまとまりを見るためにはクラスタリングを行う必要があるが、このときに二次元配置で得られた座標を直接用いてクラスタリングを行うと、グラフ構造上距離が遠い 2 ノードが近いものとして扱われ同じクラスタに含まれてしまうおそれがある(図 1)。そこで、直接エッジでつながれていない 2 ノード間の距離については無限大とみなせる大きな距離をとるものとした。

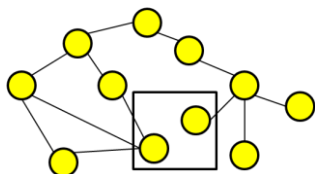


図 1 グラフ構造上は遠い 2 点の接近

凝集型階層的クラスタリングを用いるのは、このクラスタリング

中村 将人, 芝浦工業大学大学院 工学研究科 電気電子情報工学専攻 東京都江東区豊洲 3-7-5
m110098@shibaura-it.ac.jp

手法が、それらの距離に応じて近隣のノードをまとめるか否かを定めることができ、デンドログラムの高さに相当するパラメータに応じておおざっぱに構造をみるか、詳細に構造をみるかを決定することができるかと期待できるからである。

階層型クラスタリングを行った後にグラフを描画する。その際、クラスタを長方形で表現し、上位のクラスタは下位のクラスタを含むように表示する。描画されたクラスタ同士が重なって表示される事を防ぐため、各クラスタの重心とネットワーク全体の重心の位置関係から長方形の位置を調整する。

2.1 配置およびノード間距離の決定

はじめに、ネットワークのノードとエッジの繋がりを隣接行列によって与える。次にこの隣接行列から、FR 法を用いて全ノードの座標を決定する。これによって配置されたノードの座標とネットワークの隣接行列によるエッジの有無の情報をもとに各ノード間の距離をもとめ凝集型階層的クラスタリングを行う。

(1) FR 法によるノードの配置

ネットワークの隣接行列から FR 法によって各ノードの座標位置を決定する。FR 法はノード間にクーロン力に見立てた斥力とバネによる引力が働くことを仮定し各ノードの配置を決定する方法である。

(2) 凝集型階層的クラスタリングによるノードの分類

FR 法によって配置された座標から、凝集型階層的クラスタリングを行う。この時、ネットワークの隣接行列から直接エッジによる繋がりが無い 2 点間の距離に無限大とみなせる大きな数を置いた。クラスタ間の距離のとりかたについてはセントロイド法をもちいて計算を行う。なお、クラスタを併合する過程で計算されたクラスタ間距離、クラスタの内包関係、クラスタ内のノードの重心および二次元配置における x 座標と y 座標の各最大値・最小値が保存される。

(3) クラスタの長方形の頂点座標の決定

クラスタに含まれるノードの座標が全て含まれるように、クラスタを表現する長方形の頂点の座標を決定する。長方形の頂点の位置は、下位のクラスタに含まれるノードの座標の最も小さい $X(x_{min})$, $Y(y_{min})$ の値, 最も大きい $X(x_{max})$, $Y(y_{max})$ の値から決定される。長方形は (x_{min}, y_{min}) , (x_{min}, y_{max}) , (x_{max}, y_{max}) , (x_{max}, y_{min}) の 4 つの頂点で決定される。

(4) クラスタの座標の調整

クラスタ同士が重なって表示されることを避けるため、クラスタを表す長方形の位置の調整を行う。ネットワーク上の全てのノードの重心を算出し、これをネットワーク全体の重心とする。ネットワーク全体の重心からクラスタに含まれるノードの重心へのベクトルを求め、これに定数 α をかけたものを各クラスタの重心座標に加え、クラスタが重心から遠ざかるように座標の調整を行った。

2.2 グラフ描画処理

以上の処理によって得られたクラスタの配置を画面に描画する。描画されるグラフについて、ユーザが画面に出現するスライダーの値を変化させることによってクラスタを分割する距離を指定することができ、クラスタリングの詳細の程度(粒度)を決定できる。

なお、ユーザの要求した度合でクラスタリングされたクラスタだけが表示され、クラスタに含まれる下位のクラスタ・ノードは表示されないようになっている。この時、クラスタに含まれるノードの数が多い程、そのクラスタを表す長方形は赤く表示され、ノードの数が少なくなる程、そのクラスタを表す長方形は黒く表示される。(図 2～図 7)

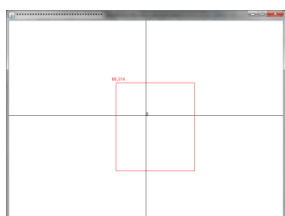


図 3 閾値 100

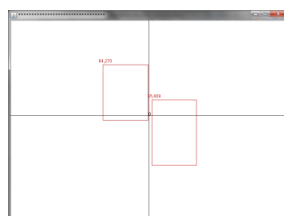


図 2 閾値 90

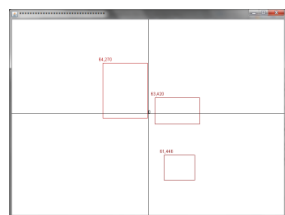


図 5 閾値 70

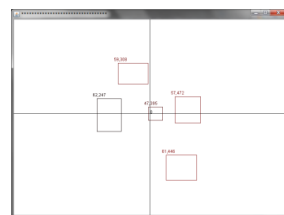


図 4 閾値 50

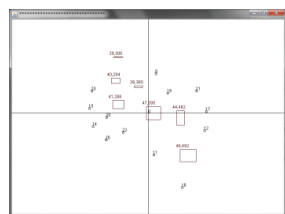


図 6 閾値 30

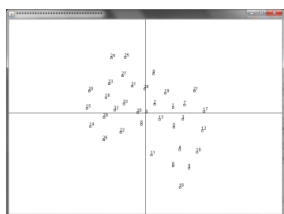


図 7 閾値 0

また、グラフの表示方法を設定するためのスライダーがプログラムの起動時にグラフとともに表示される。(図 8)

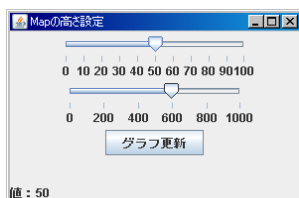


図 8 表示設定用スライダー

上のスライダーはクラスタを分割する距離を粒度として 100 分位で表したものであり、下のスライダーはグラフそのものの拡大率を表したものである。スライダーを動かした後に“グラフ更新”ボタンを押すことにより、スライダーの値に則したグラフが描画される。

3. 実験

3.1 対象ネットワーク

本手法における処理時間のノード数依存性を計測するため、“Gallery of network images”[3]から入手できるネットワークデータのうち、Zachary's karate club (ノード数 34), American College football (ノード数 115), Neural Network (ノード数 297) および Coauthorships in network science (ノード数 1594)を対象に実験を行った。

3.2 処理時間の計測

計測はプログラムの起動から全ノードの座標決定・階層的クラスタリング終了までの処理時間と、クラスタの分割の距離をもとにした粒度が最大値の 60% のときおよび 10% のときの描画時間を計測した。ネットワークごとに 5 回の計測を行い、平均値を取得した。

3.3 結果・考察

計測時間の結果を表 1 に示す。

プログラムの起動から全ノードの座標決定・階層的クラスタリング終了までの処理時間の平均値はネットワークのノード数 N の増加に応じて大きく増加していることがわかる。これは使用している計算手法がいずれも $O(N^2)$ であることに起因していることが考えられる。

また、粒度が 60% のときおよび 10% のときの描画時間については、細かいクラスタがでてくる 10% の場合にかかる時間は 60% のおよそ 2~5 倍程度であり、ノード数についての依存性もおよそ $O(\log N)$ 程度であることが伺える。

表 1 ノード数と処理時間

ノード数	34	115	297	1596
配置・クラスタ計算時間(ms)	688.0	2952	15717	496994
グラフ描画(60%)(ms)	11.2	12.4	21.8	342.4
グラフ描画(10%)(ms)	24.4	65.2	109	781

これにより、一度、配置・クラスタリング等の処理が終わってしまえば、様々な粒度でネットワークを描画することが容易な方法であるといえる。

4. まとめ

本研究では、複雑ネットワークの全体から詳細までを連続的に提示する可視化手法を提案した。本手法では FR 法によりノードの座標を決定し、凝集型階層的クラスタリングを行うことにより、ユーザが見たい粒度でネットワークを見ることができるようになった。今後はトピックマップやセマンティック Web の可視化への応用などが考えられる。

参考文献

- [1] 土井 敦, 伊藤 貴之: 力学モデルを用いた階層型グラフデータ画面配置手法の改良手法とウェブサイト視覚化への応用, 芸術科学会論文誌 Vol. 3 No. 4 pp250-263
- [2] 山口 裕美, 伊藤 貴之: 長方形の入れ子構造を用いた階層型データ視覚化手法の拡張, 情報処理学会論文誌, Vol. 44, No. 10, pp. 2469-2477, 2003.
- [3] Mark Newman, “Gallery of network images”, <http://www-personal.umich.edu/~mejn/networks/>