

静的ネットワークに対する新しいスケルトン抽出法

平野未来[†] 白山晋^{††}

ネットワークの規模が大きくなると、その可視化結果は煩雑なものになりやすい。結果として、ネットワークの全体像の把握が難しくなる。このため、ネットワーク全体を簡素化して表現する方法が模索されている。その一つがネットワークのスケルトンである。本稿では、静的ネットワークからスケルトンを抽出する新しい手法について考察する。そして、任意性が少なく、コミュニティ構造の抽出に適した手法を提案する。

A New Method of Extracting a Skeleton from Static Complex Networks

MIKU HIRANO[†] SUSUMU SHIRAYAMA^{††}

It becomes more difficult to interpret visualizations of a complex network as the scale of network becomes larger. Therefore, some methods to simplify a representation of a whole structure of the large-scale network have been developed. To extract a skeleton of complex networks is one approach to simplifying the structure. In this paper, we consider a new method of extracting a skeleton from static complex networks. And then, the method that has little arbitrariness in extracting the skeleton and is appropriate to reveal community structures of the network is proposed.

1. はじめに

人文・社会科学の分野において、人、社会、さらには現象や事象における様々な関係性をネットワーク構造として捉え、分析することが行われてきた。近年、そのような社会科学におけるネットワーク分析の研究にグラフ理論を発展させた手法や統計物理の手法が導入され、複雑ネットワーク科学という新しい領域が生まれている¹⁾。

複雑ネットワークにおいては、ネットワークの統計的指標（大局的構造）と注目領域の局所的リンク構造の両者にコミュニティ構造を加えることで、ネットワークの全体像が示されることが多い²⁾。しかし、それらの3つの構造を組み合わせただけでは、大きな構造から局所的な構造を直感的に把握することが難しいことも事実である。例えば、局所的に似通った構造を持ち、同じようなコミュニティ構造が内在する2つのネットワークにおいても、統計的指標のいくつか異なる場合は少なくない。例えば、三井は、SNS (Social Networking Service) のネットワークを分析する際に、リンク形成メカニズムが同じでありコミュニティ構造も類似しているが、クラスター係数が異なるものがあることを見出している³⁾。

このため、3つの構造以外でネットワークの特徴を調べる方法が模索されている。隣接行列を線形代数的に分析する方法やネットワーク可視化などがある。特に、可視化はネットワークの全体像を直感的に把握する上で有用とされ、古くから多くの検討がなされている^{4),5)}。しかしながら、ネットワーク可視化には確立された方法はない。距離空間の定義が難しいこと、及びデータの大規模性が主な理由である。

ネットワークは、データを示すノードと、データの接続関係を示すリンクによって表され、無向グラフ、あるいは有向グラフとして扱われる。ネットワークの可視化はレイアウト（配置）が基本になる。具体的には、ノードを示すオブジェクト（頂点、ノード）を決め、それらの関係性を、リンクを示すオブジェクト（辺、エッジ、アーク）で表し、表示空間上で適当な物理座標が与えられ表示される。図1で示すA, B, C, Dがノード、AB, AC, BD, CDがエッジである。

図1の位置情報を持たない3つのネットワークは同じものである。図のように、ノードとエッジを表示させること自体は単純ではあるが、物理座標に対応する表示空間の中での配置が、必ずしもノードの配置を適切に表現しえないという困難がともなう。

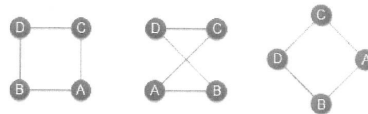


図1 ネットワークの可視化例

このように、ネットワーク可視化にはレイアウトの問題があり、“表示空間におけるノードやエッジの位置関係や距離をネットワークの大きな構造から局所的な構造までを直感的に把握できるように決めること”が課題になっている。ネットワークが大規模化するとこの課題は非常に難しいものになる。

一方、ネットワークの全体像を直感的に把握する上での可視化の有用性は増している。このため、大規模ネットワークに対する効果的な可視化手法が検討されている。代表的なものに、最小全域木をガイドとしてエッジを加えていく方法⁶⁾とコミュニティ構造に基づくネットワークの可視化表現がある。前者の最小全域木は、ネットワークの骨格を示すものとしてネットワークの分析という観点で注目されてきたものである^{6),7)}。後者は、コミュニティ構造を利用したネットワークの縮約表現である¹¹⁾。

Kimらは、部分グラフによるネットワーク構造全体の簡略化について検討し、新しい最小全域木の構成法を提案した^{8),9)}。彼らは、その部分グラフをスケルトン(skeleton)と呼び、スケルトンの持つネットワーク構造を調べ、全体の(元の)ネットワークの性質を十分に表し示している。

今藤らは、Max-Flow コミュニティ抽出法¹²⁾によってコミュニティを抽出後、コミュニティをノードに、コミュニティ間をリンクとしたネットワークの縮約表現を提案した¹¹⁾。彼女らはそれをコミュニティグラフと称している。湯田らは、SNS に対して Newman コミュニティ抽出法¹⁴⁾によってコミュニティを抽出し、コミュニティグラフによるネットワーク構造の分析を試みている¹³⁾。

van Ham と Wattenberg は、Kimらのスケルトンに注目し、ネットワークの概略を示すと同時にコミュニティ構造を顕在化するような可視化法を提案している¹⁰⁾。彼らの可視化手法は、スケルトンに含まれるノードとエッジを基本と

*[†] 東京大学大学院工学系研究科環境海洋工学専攻
Department of Environmental and Ocean Engineering, School of Engineering, The University of Tokyo

^{††} 東京大学人工物工学センター
Research into Artifacts, Center for Engineering, The University of Tokyo

し、ネットワークの統計的指標に基づいてエッジを追加するというものである。接続性やコミュニティ構造などのネットワークの特徴を保ったまま、元のネットワークが粗視化できることを示している¹⁰⁾。

しかしながら、スケルトンを一意に定めることはできず、スケルトン抽出に任意性が含まれる。また、ネットワークによっては、コミュニティ構造が十分に表出されないという問題がある。とはいえ、スケルトン抽出は、“表示空間におけるノードやエッジの位置関係や距離をネットワークの大きな構造から局所的な構造までを直感的に把握できるように決める”という課題の一つの解決策として期待されている。

本研究においては、スケルトン抽出に注目し、任意性が少なく、コミュニティ構造の抽出に適した手法を提案する。

2. 提案手法

提案手法を示す前に基本的な定義を与えておく。

ネットワークをグラフ G として扱う。頂点 (ノード) を V 、辺 (エッジ) を E とすると、 $G=(V, E)$ で表わされる集合として取り扱う。ノードを v 、あるいは識別子を付けて示す (例えば、 v_i)。場合によっては識別子のみでノードを示す (例えば、ノード i)。エッジは e で表す。ノードの総数を n 、エッジの総数を m とする。また、ノード v の表示空間 (物理空間) での位置を $\mathbf{p}(v)$ とする。

ネットワークとして、接続されている要素を 1、そうでない要素を 0 とする隣接行列も併用する (隣接行列の対角成分は 0 とする)。隣接行列を \mathbf{A} 、その成分を a_{ij} とし、

$$a_{ij} = \begin{cases} 1 & j \text{ から } i \text{ へのエッジがあるとき} \\ 0 & i \text{ と } j \text{ にエッジがないとき} \end{cases} \quad (1)$$

とする。

提案手法では、以下で示すいくつかのネットワークの統計的指標を利用する。

ネットワークの平均的距離として、グラフレイアウトまでを考慮し、グラフ距離とユークリッド距離 (物理空間での距離) を用いる。グラフ距離において d_{ij} をノード v_i と v_j の間の最短経路長とする。平均頂点間距離 L を、

$$L = 2 \sum_{i>j} d_{ij} / n(n-1) \quad (2)$$

とする。近接中心性 $C_c(v)$ を、

$$C_c(v_i) = (n-1) / \sum_{j \in V, j \neq i} d_{ij} \quad (3)$$

とする。また、物理空間での距離は、 $r_{ij} = |\mathbf{p}(v_i) - \mathbf{p}(v_j)|$ で表し、平均距離 D を、

$$D = 2 \sum_{i>j} r_{ij} / n(n-1) \quad (4)$$

とする。

次数を k で示し、ノード v_i の次数を k_i とする。また、次数分布を $p(k)$ で示す。 v_i の隣接ノード間に存在するリンクの数を E_i とし、あるノード v_i のクラスタ率 $C_2(v_i)$ を、

$$C_2(v_i) = 2E_i / k_i(k_i - 1) \quad (5)$$

とする。

エッジに対する媒介中心性 (EBC: Edge Betweenness Centrality) を次式とする。

$$C_e(e) = \sum_{i \neq j \in V} \frac{\rho_{ij}(e)}{\sigma_{ij}} \quad (6)$$

ここで、 $\rho_{ij}(e)$ はノード i から j までのリンク e を含む最短経路の数、 σ_{ij} はノード i と j をつなぐ最短経路の数である。ノードに対する媒介中心性 (NBC: Node Betweenness Centrality) を次式とする。

$$C_v(v) = \sum_{i \neq v \neq j \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \quad (7)$$

ここで、 $\sigma_{ij}(v)$ はノード i から j までのノード v を通る最短経路の数である。

本稿では、Newman コミュニティ抽出法¹⁴⁾によってコミュニティ構造を抽出する。Newman コミュニティ抽出法 (以降、Newman 法と呼ぶ) は切断法¹⁵⁾と結合法に分かれる。切断法では、エッジに対する媒介中心性が最大のエッジの切断を繰り返しながら、以下のモジュラリティ Q を最大するようにコミュニティを抽出する。

$$Q_\alpha = e_{\alpha\alpha} - a_\alpha^2, \quad (8)$$

$$Q = \sum_\alpha Q_\alpha$$

ここで、 $e_{\alpha\beta} = (\sum_{i \in \alpha} \sum_{j \in \beta} a_{ij}) / 2m$ 、 $a_\alpha = \sum_\beta e_{\alpha\beta}$ である。 α と β はあるコミュニティを示す。

また、部分グラフとして木 T を利用する。特に、最小全域木 (MST: Minimum Spanning Tree) が鍵となる。最小全域木とは、連結なグラフの辺に重みが定義されているとき、すべての頂点を含む木で辺の重みの和が最小のものである。以上を用いて提案手法を説明する。

なお、本稿では、重複辺やループのない重み無しの単純グラフを考える。また、連結グラフを基本とする。提案手法は有向グラフに対しても適用できるが、本稿では無向グラフ用のコミュニティ抽出法を用いるため、断りのない限り無向グラフを対象とする。

2.1 任意性の少ないスケルトン抽出法

Kim らはエッジの重みとして EBC を与え、クルスカルのアルゴリズム (Kruskal's algorithm) によって最小全域木 (MST) を抽出し、それをスケルトンと呼び、スケルトンによって元のネットワークの概略を示した⁸⁾⁹⁾。得られる MST は EBC の大きなエッジが削除されたものになる。先述した切断法によるコミュニティ抽出に倣ったものである。van Ham らは、この MST がコミュニティ構造を反映したものになることを後述する媒介測度 (bm: betweenness metric) によって示している¹⁰⁾。このように Kim らのスケルトン抽出法はネットワークの直感的な把握に有効な方法であるが、いくつかの問題がある。一つは同じ EBC の値を持つ複数のエッジが存在する場合、MST が一意に定まらないというものである。本節では、Kim らの方法を補完することによって、任意性の少ないスケルトン抽出法を構築する。

2.1.1 最小全域木の構成

与えられたネットワークに対して、EBC を算出し、それを重みとしてエッジに与える。Kim らの方法と同様に、EBC によって重みづけられたネットワークから最小全域木 (MST) を抽出する。ただし、本稿では、プリムのアルゴリズム (Prim's algorithm) によって連結性を保ちながら最小全域木を求める。

2.1.2 ルートノード選択

図 2 に示すように、グラフレイアウトによってスケルトンを表現する際に、ルートノード (ルート) の選び方が問題になる。

そこで、ルートを決めて、MST を構成することにする。ルートとしては様々なものが考えられる。また、後述するグラフレイアウトにおいてルートを表示空間の中心に配置するのでその点も考慮する。

本稿では、スケルトン抽出に媒介中心性という考えを用いている。この点では、視覚的認知特性を考慮し、周りのノードの最短経路上にルートが存在するように、NBC が最大のノードをルートとすることが考えられる。

一方、表示空間では、物理空間での距離という概念が発

生し、次数がリンクの集中度合いを意味するようになる。そして、全体がコンパクトに表示されることが望ましい場合も多い。この場合は、適当なルートからMSTを構成し、抽出されたスケルトンに対して近接中心性を計算し、その最大ノードをルートにすることや、最大次数のノードをルートにすることが考えられる。

ルートノードの候補が複数ある存在する場合は、候補の中で他の統計的指標が最も大きなものをルートとする。この処理でルートが一意に選択できない場合、候補をすべて表現し、ユーザが選ぶものとする。

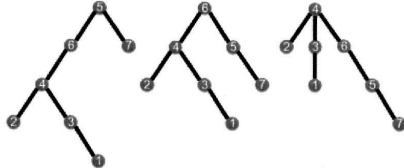


図2 同一MSTの複数の表現例(ルートノードが左から5, 6, 4)

2.1.3 エッジ選択

ネットワークのすべてのエッジがすべて異なるEBCをもつことは稀である。そのため、MSTの抽出の際に、図3に示すエッジ選択の任意性が生じる。

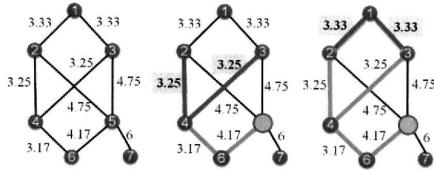


図3 エッジ選択の任意性

図3の左図の重み付きネットワーク(エッジの横にEBCの値を示す)からMSTを抽出する場合、図中央や図右のように同じEBCを持つエッジの選択の必要性が生じる。図中央の場合はどちらを先に選んでも次のステップで同じ結果になるが、右図の場合は異なる結果となる。

そこで、選択の対象となるエッジの端点ノードに対する統計的指標を用いて選択するエッジを決定する。

本稿では、ノードの属性として、 $1/NBC$ 、近接中心性、次数をこの優先順位で用いることにする。なお、次数1のノードはNBCの値が0になり、必ず全域木に含まれることを付記しておく。

また、これらの値が同じ場合、候補をすべて表現する。あるいは、2パス以上離れたエッジの属性値を用いて一意になるようにエッジを定める。詳細については講演時に述べる。

2.2 コミュニティを利用したスケルトン抽出法

Kimらが原理を述べ、van Hamらがいくつかのネットワークを用いて明らかにしたように、EBCによって重み付けられたネットワークに対するスケルトンはNewmanらのコミュニティ構造を表出するものでもある^{8),9),10)}。しかしながら、切断法のように、切断を繰り返しながらEBCを求めるわけではないのでコミュニティ構造が明確に表出される保証はない。そこで、コミュニティを抽出し、その情報を用いてスケルトンを抽出する方法を考察する。

2.2.1 最小全域木の構成

はじめに、Newman法によってコミュニティを抽出し、コミュニティに番号を振る。番号を α とし、コミュニティ

を C_α で示す。次に、ノードの属性に属するコミュニティ番号を与える(図4では色によってコミュニティの違いを示している)。

図5のように、コミュニティをノードとするネットワークを作成する(今藤ら¹⁰⁾と同様にコミュニティグラフと呼ぶ)。エッジの重みを、 $e_{\alpha\beta} = (\sum_{i \in \alpha} \sum_{j \in \beta} a_{ij}) / 2m$ で算出する。

このコミュニティグラフに対して最小全域木(MST)を抽出すると、 $e_{\alpha\beta}$ の小さなエッジが選択される。式(8)より、大きな $e_{\alpha\beta}$ はQを小さくするように作用するので、コミュニティ構造が明確になるものと考えられる。

コミュニティグラフに対するMSTの抽出後、各々のコミュニティに対してMSTを抽出し、ネットワーク全体のMSTを構成する。

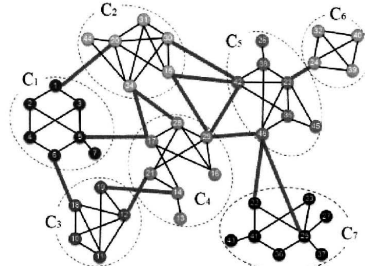


図4 コミュニティ抽出

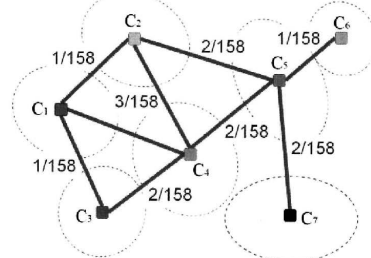


図5 コミュニティグラフの作成

2.2.2 ルートノード選択

はじめに、コミュニティグラフのノードの属性値として、式(8)の Q_α 、 $e_{\alpha\alpha}$ (コミュニティ内のリンク数に相当)、コミュニティ内のノード数を与える。これらのいずれかの値が最大のものをルートを含むコミュニティと定める(いずれかはユーザが決める)。このコミュニティ内で、NBCが最大のノードをルートノードとする。

なお、表示空間では、全体がコンパクトに表示されることが望ましい場合は、2.1.2項で述べた方法によってルートノードを選択する。

2.2.3 エッジ選択

コミュニティグラフに対するMSTの抽出において、同じ $e_{\alpha\beta}$ の値を持つエッジの選択は、2.1.3項で述べた方法によって行う。コミュニティ内も同様である。

2.2.4 コミュニティ間のエッジ選択

コミュニティ間に複数のエッジが存在する場合はいずれかのエッジを選択する。

はじめに、コミュニティ α に属するノード i とコミュニティ β に属するノード j の属性値として次式の P_{ij} を与える。

$$P_{ij} = a_{ij} \frac{k_{i\beta} k_{j\alpha}}{k_i k_j} \quad (9)$$

ここで、 $k_{i\beta}$ はノード*i*のもつエッジの中でコミュニティβに繋がるエッジの数である。

ノード間にエッジがあり、他方のコミュニティに繋がるエッジの比率が大きいノード同士を選べば P_{ij} の値が大きくなる。本稿では、 P_{ij} が最大となるエッジを選択する。

2.3 評価基準

下記の3つの項目で、抽出されたスケルトンの評価を行う。

2.3.1 最小全域木の総数

元のネットワークを重み無し無向グラフとすると、キルヒホッフの行列木定理によって全域木の数が求められる。その値を $\tau_1(G)$ とする。また、Kimらと van Hamらの最小全域木の数を $\tau_2(G)$ とする。提案手法で得られる最小全域木の数を $\tau_3(G)$ とし、これらと比較することで任意性が少ないことを示す。

2.3.2 媒介測度

スケルトンにおけるコミュニティの明確さを、van Hamらが提案した媒介測度 (bm) によって数値化する。

両端のノードが異なるコミュニティに属するエッジの数を χ とする。コミュニティの数を c とすると、最小全域木において明確にコミュニティが分離している場合は、 $\chi = n - c$ が成立する¹⁰⁾。 $n = c$ の場合を除き、媒介測度を、

$$bm = \chi / (n - c) \quad (10)$$

とする。なお、2.2節で示したコミュニティを利用した方法では、必ず1となるのでこの指標での評価は行わない。

2.3.3 統計的指標の再現度

スケルトンが元のネットワークの統計的指標をどの程度再現できるかを評価値とする。本稿では、Kimら⁸⁾⁹⁾のように抽出したスケルトンの度数と EBC の分布が冪乗則に従っていることを再現性の基準と考える。

2.4 グラフレイアウト

Radial graph drawing によってスケルトンを表現する。

3. 実験結果と考察

紙面の都合で、ここでは一例のみ結果を示す。

元になるネットワークは CNN モデルで作成したノード数 75 のネットワークとする。

2.2節で述べた方法での結果を示す (図6)。ノードはコミュニティによって色付けている。

表示を考慮し、ルートノードを NBC が最大であるノードとした。コミュニティグラフを構築した後、各コミュニティの MST を抽出し統合した。また、MST の抽出においてエッジが重複した場合には、1/NBC、近接中心性、度数の順にエッジを選択することにした。

表1に van Ham らの方法との比較を示す。MST の総数は、van Ham らが 75、提案手法では 1 であり、任意性の問題は解決できている。スケルトンの度数と EBC の分布は冪乗則に従っている。

元のネットワークの再現性については講演時に述べる。

4. まとめ

本研究では、ネットワーク可視化における“表示空間におけるノードやエッジの位置関係や距離をネットワークの大きな構造から局所的な構造までを直感的に把握できるように決める”という課題の解決策として、スケルトン抽出に注目し、任意性が少なく、コミュニティ構造の表出に適した手法を提案した。

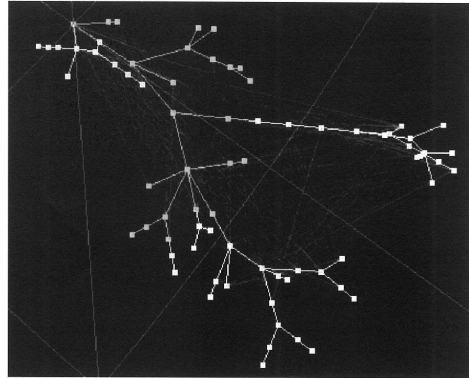


図6 コミュニティを利用したスケルトン抽出法による結果

表1 コミュニティを利用したスケルトン抽出法の評価

	提案手法(2.2節)	van Ham ら
MST の総数	1	75
度数の冪乗則	-0.491	-0.657
EBC の冪乗則	-1.191	-0.611

参考文献

- 1) ダンカン・ワッツ著, 辻竜平, 友和政樹訳; スモールワールド・ネットワーク 世界を知るための新科学的思考法, 阪急コミュニケーションズ, 2004.
- 2) S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang: Complex networks: Structure and dynamics, Physics Reports 424, pp.175-308, 2006.
- 3) 三井一平, 内田誠, 白山晋: コミュニティ構造を有するネットワーク成長モデル, 情報処理学会研究報告-知能と複雑系, Vol.2006, No.2, pp.17-24, 2006.
- 4) I. Tollis and G. D. Battista: Graph Drawing: Algorithms for Geometric Representations of Graphs, Prentice Hall, 1998.
- 5) 杉山公造: グラフ自動描画法とその応用, 計測自動制御学会学術図書, コロナ社, 1993.
- 6) A.T. Adai, V. Shailesh, S. Wieland, and E. M. Marcotte: LGL: Creating a Map of Protein Function with an Algorithm for Visualizing Very Large Bio- logical Networks, Journal of Molecular Biology, Vol.340, No.1, pp.179-190, 2004.
- 7) B. Cheswich, H. Bursh, and S. Branigan: Mapping and visualizing the internet, Proc. of Usenix Annual Technical Conference, pp.1-12, 2000.
- 8) D.-H. Kim and H. Jeong: Scale-Free Spanning Trees of Complex Networks, Journal of the Korean Physical Society, Vol. 44, No.3, pp.624-627, 2004.
- 9) D.-H. Kim, J.-D. Noh, and H. Jeong: Scale-free trees: The skeletons of complex networks, Phys. Rev. E 70, 046126, 2004.
- 10) F. van Ham and M. Wattenberg: Centrality Based Visualization of Small World Graphs, In Proceedings of Eurographics, Vol. 27, No.3, pp. 975-982, 2008.
- 11) 今藤紀子, 喜連川優: Max-Flow コミュニティグラフとその特徴分析, 日本データベース学会 Letters, Vol.3, No.1, pp.69-72, 2004.
- 12) N.Imafuji and M.Kitsuregawa.: Finding Web Communities by Maxi- mum Flow Algorithm using Well-Assigned Edge Capacities, IEICE Trans. on Inf. & Syst., Vol.E87-D, No.2, pp.407-415, 2004.
- 13) 湯田聰夫, 小野直亮, 藤原義久: ソーシャル・ネットワーク・サービスにおける人的ネットワークの構造, 情報処理学会論文誌, Vol. 47, No. 3, pp. 865-874, 2006.
- 14) 林幸雄編著, 大久保潤, 藤原義久, 上林 憲行, 小野 直亮, 湯田 聰夫, 相馬亘著: ネットワークの道具箱, 近代科学社, 2007.
- 15) M. E. J. Newman and M. Girvan: Finding and evaluating community structure in networks, Phys. Rev. E 69, 026113, 2004.