

中心性に基づく大規模ネットワーク粗視化法の提案

Proposing a Coarse-Graining Method based on Centrality for Large-scale Networks

伏見 卓恭 齊藤 和巳 池田 哲夫 武藤 伸明
Takayasu FUSHIMI Kazumi SAITO Tetsuo IKEDA Nobuaki MUTOH

静岡県立大学経営情報イノベーション研究科

Graduate School of Management and Information of Innovation, University of Shizuoka

Many methods for visualizing networks have been proposed as a key technique of qualitatively evaluating the whole network structure. However in case of large-scale networks, drawing all nodes makes it difficult to understand the anatomy of networks and requires large computation time. Extracting essential portions of large networks plays an important role in overcoming these shortcomings. In this paper, we propose a coarse-graining method based on important nodes derived from one of representative centrality concepts. Concretely, our method first extracts well-balanced important nodes referred to as pivots, and then adds the links between the relatively neighbor pivots. We qualitatively evaluate whether our method can reveal the skelton of the network by using some road networks, and quantitatively evaluate whether important nodes are extracted by using some social networks. By experimental results, we show our method can extract the well-balanced important nodes and make it possible to uncover the backbone of networks.

1. はじめに

近年、Web 上でのハイパーリンクネットワークや、SNS 内でのユーザ関係ネットワークにとどまらず、電力網や道路網などにネットワーク分析手法を適用するなど、幅広い分野で複雑ネットワークが注目を浴びている。これら大規模なネットワークから、重要ノードを抽出するための指標がいくつか提案されている [Freeman 79]。

一方、重要ノードだけでなくネットワークの全体構造を概観するために、多くのネットワーク可視化法が提案されている [Yamada 03, Kamada 89]。大規模なネットワークを対象とする場合、全てのノードの低次元座標を求め布置することは多大な計算時間を要し限界がある。布置された結果が得られる場合でも、多くのノードが姦めき合い、ネットワークの骨格構造を概観することは困難なことが多い。

そこで、全てのノードを布置するのではなく、ノードを絞り込み、あるいは、グルーピングすることで、ノード数を減らし可視化する、粗視化のスキームが多く提案されている。既存の粗視化法として、 k -core 抽出によるコア度に着目し、高いコア度のノードを中心に、徐々にコア度の低いノードを周辺に描画する方法があげられる [Alvarez-Hamelin 05]。他の手法として、Newman らのクラスタリング法 [Newman 04] を代表とするコミュニティ抽出手法を土台にし、抽出したコミュニティを 1 つのノードとする方法、ネットワークの自己相似性に着目し、Box-Covering 法、Cluster-Growing 法によりノードをグルーピングする方法 [Song 05]、ネットワークの特性を表すスペクトラル構造を保存しながらノードをグルーピングする方法 [Gfeller 07] などが挙げられる。

しかし、ノードをグルーピングし 1 つのノードに併合する手法では、併合された新ノードが元のノードの性質を失うなどの問題がある。本研究では、上述した既存法とは異なり、元のネットワークにおいて重要なノードをピボットとして抽出し、相対的に近傍するピボット間にリンクを付与し可視化する方

法を提案する。重要ノードの指標として、次数中心性、近接中心性、媒介中心性の概念に基づいてピボットを抽出するが、ある種のネットワークに適用した場合、中心性上位ノード（重要ノード）同士は近傍に位置することがあり、全体構造を概観する粗視化には適さない場合もある。そこで、ノード同士の互いに協調する点を考慮した集合中心性 [伏見 12] によりピボットを抽出する。集合中心性は、ノード部分集合を対象とし、集合としての中心性値が高くなるノード群を抽出する手法であり、抽出したノード同士が協調して働く点を考慮し、既存の中心性を拡張定義したものである。

粗視化はネットワークの全体構造の概観、すなわち骨格構造の鮮明化を目的にするため、ピボットはネットワーク全体から万遍なく抽出されることが望ましい。また、ネットワークを代表する重要なノードを押さえていることも望まれる。これらの点から提案法を座標情報を有する道路ネットワークを用いて定性的に、ソーシャル・ネットワークを用いて重要ノード抽出ならびに抽出バランスの観点から定量的に評価する。

2. 提案法

提案法は、集合中心性によりピボットと呼ぶ重要ノードを抽出し、グラフ距離において相対的に近傍するピボット間にリンクを付与することで、骨格構造を表すネットワークを構築し、可視化する。提案法のアルゴリズムを以下に示す。ネットワーク $G = (V, E)$ が与えられ、抽出するピボット集合 $R \subset V$ 、その要素数を $K = |R|$ とする。

1. 全ノード集合 V から集合中心性によりピボット集合 R を抽出する (2.1 節);
2. ピボット集合 R に対して、相対近傍法によりリンク集合 E' を求める (2.2 節);
3. $G' = (R, E')$ に対して可視化手法を適用し、低次元座標ベクトル群 X を計算、出力する;

以下にピボット選択法とリンク付与法について説明する。

連絡先: 伏見卓恭, 静岡県立大学経営情報イノベーション研究科, 静岡県静岡市駿河区谷田 5 2 - 1, 054-264-5436, j11507@u-shizuoka-ken.ac.jp

2.1 ピボット選択法

集合中心性は、 K 個のピボットを抽出する際、抽出したノード同士が互いに協調して働く点を考慮し拡張定義された中心性である [伏見 12] .

集合次数中心性

個々のノードの隣接度に注目する次数中心性の概念をノード集合に対する概念に拡張すると、集合 R 内のいずれかのノードと隣接するノード数 (協調隣接度) によって定義される .

$$F(R) = \left| \bigcup_{v \in R} \Gamma(v) \right| \quad (1)$$

ここで、 $\Gamma(v)$ は、ノード v の隣接ノード集合を表す . 個々のノードに対する次数中心性、すなわち “多くの隣接ノードを持つノードは重要である” という概念の自然な拡張となっている .

集合近接中心性

個々のノードの近接度に注目する近接中心性の概念をノード集合に対する概念に拡張すると、集合 R 内の最も近いノードいずれかとの距離 (協調近接度) によって定義される .

$$F(R) = \sum_{u \in V} \max_{v \in R} \rho(u, v) \quad (2)$$

ここで、 $\rho(u, v) = 1/(1 + d(u, v))$ は、ノード u, v 間の類似度、 $d(u, v)$ はグラフ距離を表す . 個々のノードに対する近接中心性、すなわち “他のノードとの距離が小さいノードは重要である” という概念の自然な拡張となっている .

集合媒介中心性

個々のノードの媒介度に注目する媒介中心性の概念をノード集合に対する概念に拡張すると、集合 R 内のいずれかのノードを媒介する割合 (協調媒介度) によって定義される .

$$F(R) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(R)}{\sigma_{s,t}} \quad (3)$$

ここで、 $\sigma_{s,t}(R)$ は、ノード s からノード t への経路で、 $r \in R$ を通る最短パス数を表す . 個々のノードに対する媒介中心性、すなわち “多くのノード間を媒介するノードは重要である” という概念の自然な拡張となっている .

各集合中心性において、それぞれ目的関数 1, 2, 3 を最大にするようなピボット集合 R を求めるためのアルゴリズムを以下に示す (貪欲法) .

1. 初期化 : $R = \emptyset, k = 0$ とする ;
2. 選定 : 追加した際の増分が最大となるノード $\hat{v} = \arg \max_{v \in V} \{F(R \cup \{v\}) - F(R)\}$ を求める ;
3. 更新 : $R \leftarrow R \cup \{\hat{v}\}, k \leftarrow k + 1$ とする ;
4. 判定 : $k = K$ なら終了、さもなければ 2. へ ;

また、求める集合の要素数 $K = 1$ のとき、各集合中心性により選定されるノードは、各既存中心性トップノードと等しくなる .

2.2 リンク付与法

上述した方法により抽出されたピボットペアに対して、相対的に近傍するピボット間にリンクを付与する方法 (相対近傍法) [Supowit 83] を説明する .

ピボット集合 $R = \{p_1, p_2, \dots, p_K\}$ とすると、相対近傍法は任意のピボットペア p_i と p_j が相対近傍である場合に p_i, p_j 間にリンクを付与する . 詳細なアルゴリズムを以下に示す .

1. 任意のピボットペア $\{p_i, p_j\}$ について、ピボット間の距離 $d(p_i, p_j)$ を半径とする円をピボット p_i, p_j のそれぞれを中心として描く ; (円の重なった部分を Lune と呼ぶ)
2. Lune に他のピボットが存在しない場合に、ピボット p_i, p_j 間にリンクを付与する ;
3. (2) を全ピボットペアに対して実行する ;

道路ネットワークに関していうと、最小全域木やガブリエルグラフなどの近接グラフと比較して、相対近傍グラフが道路網に対する良い近似となるという結果が報告されている [Watanabe 10] . また、ソーシャル・ネットワークにおいても相対的に近傍な (少ないステップ数で辿りつける) ピボット間にリンクを付与するのは、妥当なリンク付与法であると考えられる .

3. 評価実験

可視化法に依存しない元の座標情報のある道路ネットワークを対象に、提案法が骨格構造を浮き彫りにできるかを定性的に評価する . また、ソーシャル・ネットワークを対象に、重要なノードをネットワーク全体からバランスよく抽出できるかを定量的に評価する .

3.1 ネットワークデータ

1つ目のネットワークは、日本の「ウィキペディア」内の「人名一覧」に登場する人物において、ウィキペディア内の記事中に 6 回以上共起する 2 人の人物をリンクで結ぶことによってネットワークを構築した、共起人名ネットワークである . ノード数は 9,481、リンク数は 122,522 であり、ソーシャル・ネットワークの特徴を有する . 以下このネットワークを Wiki ネットワークと呼ぶ .

2つ目のネットワークは、ブログのトラックバックネットワークである . あるブロガーから他のブロガーへトラックバックを通して情報が拡散していく . “goo”^{*1} というサイトの「JR 福知山線脱線事故」というテーマからトラックバックを 10 段辿ることにより 2005 年 5 月に収集した . ノード数は 12,047、リンク数は 39,960 であり、ソーシャル・ネットワークの特徴を有する . 以下このネットワークを TrackBack ネットワークと呼ぶ .

3つ目のネットワークは、全国デジタル道路地図データベース静岡市版 [shi 11] をもとに作成した静岡市の道路網データであり、交差点をノード、交差点間の道路をリンクとしたネットワークである . ノード数は 31,963、リンク数は 44,467 である . 以下のこのネットワークを ShizuokaR ネットワークと呼ぶ .

4つ目のネットワークは、全国デジタル道路地図データベース浜松市版 [ham 11] をもとに作成した浜松市の道路網データであり、ノード数は 26,208、リンク数は 39,323 である . 以下このネットワークを HamamatsuR ネットワークと呼ぶ .

*1 <http://blog.goo.ne.jp/usertheme/>

3.2 評価方法

ピボットがネットワーク全体からバランスよく抽出されているかを, Newman クラスタリング [Newman 04] よりノード群を N 個のクラスタに分割し, 各クラスタのサイズ分布 P と各クラスタ内のピボット数分布 $Q(K)$ の KL ダイバージェンスにより評価する. Newman クラスタリングによる n 番目のクラスタのノード集合を P_n , n 番目のクラスタに含まれるピボットの集合を Q_n とし, それぞれ $p_n = |P_n|/|V|$, $q_n = |Q_n|/K$ のように正規化すると,

$$KL(P||Q(K)) = \sum_{n=1}^N p_n \log_N \frac{p_n}{q_n} \quad (4)$$

KL ダイバージェンスが小さければ, 分布間の距離が小さく, ピボットがネットワーク全体からバランスよく抽出されていることを示す.

道路ネットワークに対する Newman クラスタリングの結果は, クラスタが地域の近似になっており, 各地域からピボットをバランスよく抽出できるかという解釈となる.

ピボットとして選択されたノードが重要ノードである割合を定量的に評価する. 元のネットワークにおいて, ある中心性ランキング上位 k の重要ノード集合を I_k , ピボット数 k のピボット集合を R_k とすると, 両集合の Jaccard 係数により評価する.

$$J(k) = \frac{|I_k \cap R_k|}{|I_k \cup R_k|} \quad (5)$$

Jaccard 係数が大きければ, 重要ノード集合の要素とピボット集合の要素の一致率が高く, より多くの重要ノードがピボットとして抽出されていることを示す.

ピボット数は多ければ多いほど可視化の計算量が大きくなるが, 少なければ少ないほど骨格構造の説明力が低下するため, ピボット数を変動させて評価する.

提案する 3 手法の有効性を評価するために, 次数中心性 (dgc), 近接中心性 (clc), 媒介中心性 (bwc), ネットワーク全体からのランダム選択 (rds) によるピボット抽出およびリンク付与と比較する.

3.3 定性評価

元の座標情報のある道路ネットワークを用いて, 粗視化結果を定性的に評価する. 粗視化結果のノードの色は各クラスタを表し, 本稿では 10 クラスタに分割している. ShizuokaR ネットワークに対する結果を図 1 に示す. 全ノードをプロットした結果 (オリジナル) および, それぞれの方法により粗視化した結果と比較すると, 集合近接中心性やランダム選択による粗視化結果は, 全体構造を概観でき, オリジナルの骨格構造を反映しており, また, 各地域の近似となっている各コミュニティから万遍なくピボットが選択されていることがわかる. 一方, 近接中心性や媒介中心性による粗視化結果は, 抽出されたピボットが互いに近傍しているため, 全体構造を概観することが困難である. 道路ネットワークでは, 多くのノード (交差点) の次数は高々 3, 4 であり, 次数中心性ならびに集合次数中心性の結果は, 市街地に点在している結果となった. 誌面の都合上詳細は省略するが, HamamatsuR ネットワークは ShizuokaR ネットワークとほぼ同様な結果が得られた.

これら定性的評価の結果より, 提案する集合近接中心性および集合媒介中心性による粗視化法は, 全体構造を概観でき, 骨格構造をうまく反映できる結果が得られることがわかった.

3.4 定量評価

重要ノードを抽出可能なソーシャル・ネットワークを用いて, 重要ノード抽出率およびピボット選択バランスにより定量的に評価する. なお, 既存中心性は重要ノードを抽出するための指標であるため, 重要ノード抽出率の評価結果は明らかであるため, ここでは省略する. 横軸にピボット数, 縦軸にそれぞれの評価値を示す.

Wiki ネットワークに対する評価結果 (図 2) を見ると, (a) から既存中心性と比較して, 集合中心性およびランダム選択は KL ダイバージェンス (分布間距離) が小さく, 各コミュニティからバランスよくピボットを抽出できていることがわかる. ランダム選択は, ピボットをネットワーク全体からランダムに抽出しているため, KL ダイバージェンスが極めて小さくなるのは自明である. (b) ~ (c) からランダム選択と比較して, 集合中心性は重要ノードをピボットとして抽出できていることがわかる. 実際, 人名ネットワークの各コミュニティにおける重要人物 (例: 徳川家康, 明石家さんま, 野村克也, 緒方賢一, 手塚治虫, ベートーヴェン, 小泉純一郎ほか) を抽出できている. すなわち, 集合中心性による粗視化は, ランダム選択に匹敵するほどネットワーク全体からバランスよくピボットを抽出しており, かつ各種中心性の概念における重要ノードを抽出できていることになる. 誌面の都合上詳細は省略するが, TrackBack ネットワークは Wiki ネットワークとほぼ同様な結果が得られた.

4. おわりに

本研究では, 大規模なネットワークの全体構造の概観と骨格構造の鮮明化を目的に, 集合中心性によるピボット抽出および相対近傍法によるリンク付与からなる粗視化法を提案した. 粗視化の要件として, 重要ノードをネットワーク全体からバランスよく抽出することに着目し提案法を評価した. 評価実験の結果, 既存の中心性のように互いの依存関係を考慮しない方法では, ピボットとして重要ノードを抽出できるが, ネットワーク構造によってはバランスが偏る場合がある. ランダム選択では, ピボットをバランスよく抽出できるが重要ノードを網羅できない点で問題があることが示された. 一方提案法では, 両者の利点を活かした粗視化の要件を満たす結果が得られることがわかった. 提案法は, 道路ネットワークに対してはカーナビにおける交差点表示, ソーシャル・ネットワークに対してはソーシャル・グラフでのキーパーソン表示などへの応用が期待できる.

今後はさらに多様なネットワークデータを用いて, 提案法の有効性を検証していきたい.

謝辞 本研究は, 豊田中央研究所との共同研究および, 科学研究費補助金基盤研究 (C) (No. 23500128) の補助を受けた.

参考文献

- [Alvarez-Hamelin 05] Alvarez-Hamelin, J. I., Dall'Ásta, L., Barrat, A., and Vespignani, A.: k-core decomposition: a tool for the visualization of large scale networks, *World Wide Web Internet And Web Information Systems* (2005)
- [Freeman 79] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, No. 3, pp. 215-239 (1979)

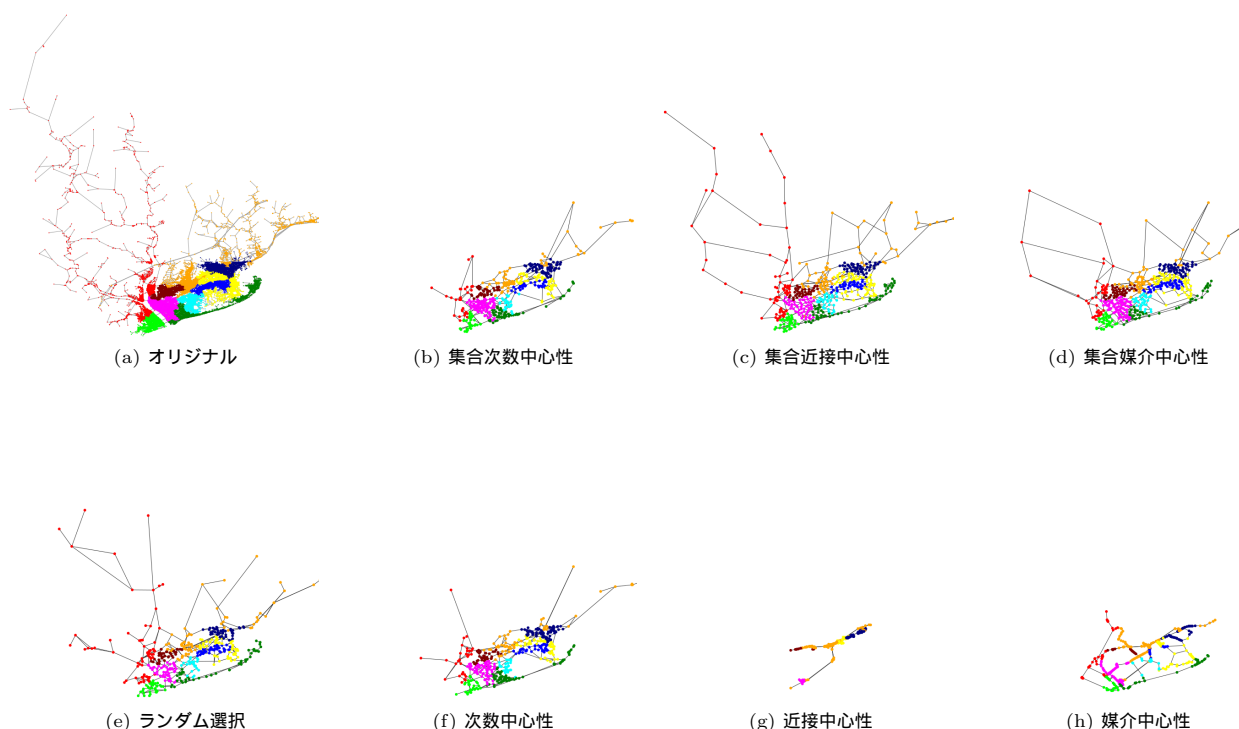


図 1: ShizuokaR ネットワーク 粗視化結果 $K = 500$

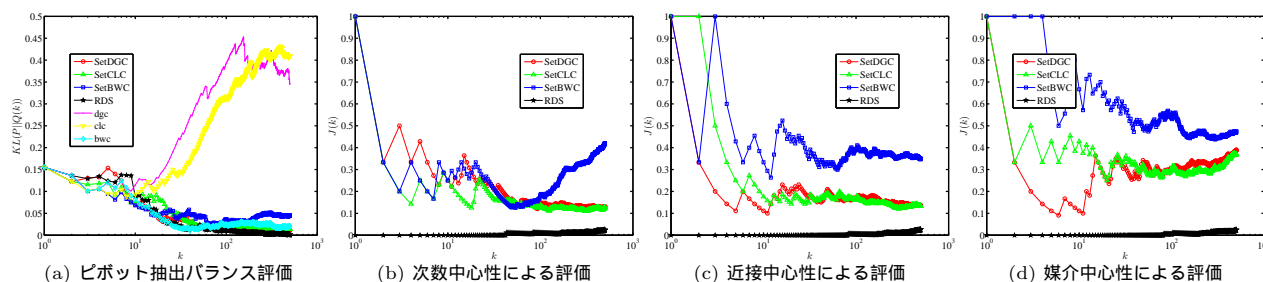


図 2: Wiki ネットワーク

[Gfeller 07] Gfeller, D. and Paolo, : Spectral Coarse Graining of Complex Networks, *Physical Review Letters*, Vol. 99, No. 3 (2007)

[ham 11] 全国デジタル道路地図データベース浜松市版 (2011)

[Kamada 89] Kamada, T. and Kawai, S.: An algorithm for drawing general undirected graphs, *Inf. Process. Lett.*, Vol. 31, pp. 7–15 (1989)

[Newman 04] Newman, M. E. J.: Detecting community structure in networks, *The European Physical Journal B - Condensed Matter and Complex Systems*, Vol. 38, No. 2, pp. 321–330–330 (2004)

[shi 11] 全国デジタル道路地図データベース静岡市版 (2011)

[Song 05] Song, C., Havlin, S., and Makse, H. A.: Self-similarity of complex networks, *Nature*, Vol. 433, No. 7024, pp. 392–395 (2005)

[Supowit 83] Supowit, K. J.: The Relative Neighborhood Graph, with an Application to Minimum Spanning Trees, *J. ACM*, Vol. 30, No. 3, pp. 428–448 (1983)

[Watanabe 10] Watanabe, D.: A Study on Analyzing the Grid Road Network Patterns using Relative Neighborhood Graph, in *The Ninth International Symposium on Operations Research and Its Applications (ISORA'10)*, pp. 112–119 (2010)

[Yamada 03] Yamada, T., Saito, K., and Ueda, N.: Cross-entropy directed embedding of network data, in *Proceedings of the 20th International Conference on Machine Learning (ICML03)*, pp. 832–839 (2003)

[伏見 12] 伏見 卓恭, 斉藤 和巳, 武藤 伸明, 池田 哲夫: ノード集合に対する媒介中心性の提案, 第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012) (2012)