

ノード選択確率を考慮した中心性指標の拡張

Extending centrality measures considering probabilities of node selection

伏見卓恭[†]

Takayasu Fushimi

斉藤和巳[‡]

Kazumi Saito

池田哲夫[‡]

Tetsuo Ikeda

武藤伸明[‡]

Nobuaki Mutoh

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など, 現実のあらゆる場面で複雑ネットワークが観察されるようになっている. ソーシャルネットワーク分析の分野では, 個々のノードの性質に着目し, ノードをランキングすることで膨大な数のノード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 [1, 2]. 重要ノード抽出の指標として次数中心性, 近接中心性, 媒介中心性などが広く知られている. これら中心性指標は, すべてのノードを対等に扱い, ノードの性質に着目する手法である.

しかし, 現実ネットワークにおいてすべてのノードは様ではなく, ノードやリンクに多重度などの重みをもつ場合が多い [3]. 道路網における交差点付近の人口比率, ソーシャルメディアにおけるユーザの活発度, Web サイトにおけるユーザの滞在確率, 企業間取引ネットワークにおける企業の売上高など, 重みや確率といった有用な情報が付与されている. すなわち, ノードを均等に扱うより不均一さを考慮した分析手法の方が, より現実性に則した結果が得られると期待できる.

重みを考慮した中心性の既存研究として, リンクの多重度を考慮した [4] などがあげられる. このように, リンクの多重度や重みを考慮する手法が多く提案されているが, ノードの選択確率を考慮した手法は見られない.

ソーシャルネットワークにとどまらず, 電力網や道路網などに複雑ネットワーク分析手法を適用することで, 現実問題への応用も考えられている [5, 6, 7]. 媒介中心性を拡張した集合媒介中心性は, 道路網における看板配置問題への応用が期待できる. さらに, 媒介中心性はネットワーク上でのトラフィックのボトルネックを抽出する指標としても知られているため, ノードの選択確率を考慮する方が現実的であると自然に想定される.

本研究ではノードが潜在的に有する選択確率を考慮し, 選択確率の高いノードとの関係性をより焦点に当てる中心性指標 (以下, 確率中心性) へと拡張する. たとえば, ノードを交差点, リンクを交差点間の道路とした道路ネットワークにおいて, 交差点付近の人口比率をノードの選択確率とした場合を想定する. 任意の最短パスを媒介する割合により重要ノードを選定する媒介中心性において, 人口比率を考慮しない場合では, 大きな都市を結ぶノードだけでなく小さな村間を結ぶ橋などに該当するノードもランキング上位になる場合がある. 一般的に, 小さな村間を結ぶ交差点より, 大きな都市間の媒介となる交差点の方が重要な役割を果たすというのが直観的である. 現実への応用を見据えれば, ノード選択確率を加味した拡張手法の方が有用性が期待できる.

本研究では現実問題への応用を視野に入れ, 上述した具体例のようなケースを目論み, オリジナルの中心性に選択確率を考慮し, 確率的中心性へと拡張する. 紙面の都合上, 媒介中心性について言及するが, 次数中心性, 近接中心性に関しても同様の拡張が可能である.

また, ネットワークの重心が人口分布の重心と近い位置にあれば, 確率媒介中心性と通常の媒介中心性のランキング結果に大きな違いはないが, 2つの重心が大きく異なる場合, 両中心性による上位集合は異なるかと推測できる. この推測を人工データを用いて, 簡単な例で実証的に検証する. 通常の媒介中心性との比較による評価実験の結果, 地域人口を考慮したより現実的な重要ノードを抽出可能なことを示す.

2. 媒介中心性および確率媒介中心性

以下に通常の媒介中心性および, 本稿で拡張する確率媒介中心性について, よく知られた文献 [1] の説明に準じて説明

する. ノード集合 V , リンク集合 E からなる無向ネットワークを $G = (V, E)$ と表記する. 文献 [1] と同様に, ネットワークとしては連結な単純無向ネットワークを前提とする.

ノード v の媒介度とは, 任意のノードペアを結ぶパスを, どの程度媒介しているかを示す指標である. 媒介中心性とは, 多くのノード間の橋渡しをしているノードは重要であるという直観に基づいた中心性であり, 任意のノードペア間の最短パスのうち, 媒介しているパスの割合によりノードをランキングするものである. ノード v の媒介中心性 $bwc(v)$ を以下のように定義する.

$$bwc(v) = \frac{1}{|V|^2} \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \quad (1)$$

ここで, $\sigma_{s,t}$ は始点ノード s と 終点ノード t 間の最短パス数であり, $\sigma_{s,t}(v)$ はノード v を通るノード s, t 間の最短パス数を表す. 媒介中心性は, 始点と終点ノードをランダムに選んだ際, ノード v が最短パス上に存在する確率を表していると言い換えられる.

次に, 確率媒介中心性について説明する. ノード s の選択確率とは, 始点と終点ノードをランダムに選ぶ際, ノード s が選択される任意の確率である. すなわち, 始点ノードとしてノード s がランダムに選ぶ際の選ばれやすさを表す. ノード s の選択確率を以下のように設定する.

$$p(s) = \frac{w_s}{\sum_{u \in V} w_u} \quad (2)$$

ここで, w_s はノード s に付与されている重みとする.

式 (1) に式 (2) の確率を導入する. ノード v の確率媒介中心性 $pbwc(v)$ を以下のように定義する.

$$pbwc(v) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \cdot p(s) \cdot p(t) \quad (3)$$

式 (3) より, 一様でないノード選択確率分布において, 確率が大きい s と t を媒介するようなノード v の値が大きくなり, 重要ノードとして抽出できるようになる. ノード選択確率分布が一様の場合, 通常の媒介中心性に帰着できるため, 自然な拡張になっている.

3. 評価実験

この節では, 評価実験に関する詳細を説明する.

3.1. データセット

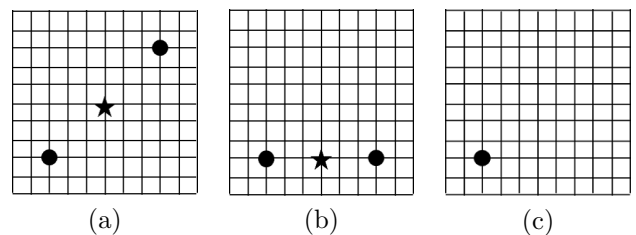


図 1: ガウシアンカーネルの中心ノード

人工ネットワークとして, 2次元平面上の正方格子を用い, 縦に 100, 横に 100 でノード数が 10,000 のネットワークを作る. このネットワーク上に, ガウシアンカーネルを用いて

[†]静岡県立大学大学院経営情報イノベーション研究科, 日本学術振興会特別研究員 DC2

[‡]静岡県立大学大学院経営情報イノベーション研究科

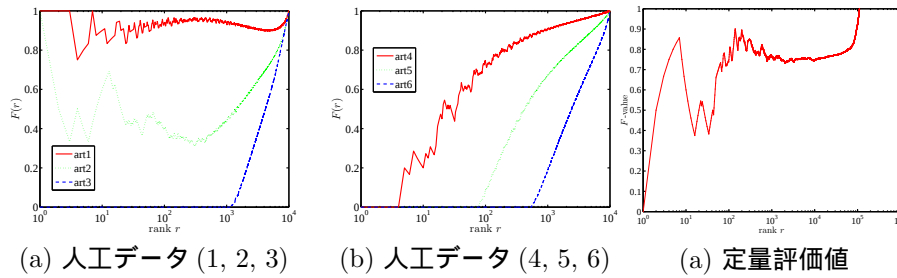


図2: 人口分布を変えた際の定量評価値

(a) 人工データ (1, 2, 3) (b) 人工データ (4, 5, 6) (a) 定量評価値 (b) ランキング上位 40

図3: 静岡県道路網に対する処理結果

人工的な人口分布を付与する．人口の拠点となる中心ノードを設定し，パラメータ σ に従い中心ノードからのグラフ距離が離れるほど人口比率が小さくなるように付与する．以下に，ガウシアンカーネルの中心になるノードおよびパラメータの設定を変えた人工データを示す．図1において，●はガウシアンカーネルの中心ノード，すなわち人口の拠点となる．はそれら人口の重心を示す．人口重心とガウシアンカーネルの中心が一致する人口の少ないでは を図示していない．

1. 中心ノード：ネットワークの重心ノード， $\sigma = 25$
2. 中心ノード：図1(a) 参照， $\sigma = 20$
3. 中心ノード：図1(b) 参照， $\sigma = 20$
4. 中心ノード：図1(c) 参照， $\sigma = 200$
5. 中心ノード：図1(c) 参照， $\sigma = 100$
6. 中心ノード：図1(c) 参照， $\sigma = 50$

以下，これらの人工ネットワークを art1～art6 とよぶ．道路ネットワークとして，全国デジタル道路地図データベース静岡県版 [8] をもとに作成した静岡県の道路網データを用い，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 110,925，リンク数は 162,322 である．このネットワークを Shizuoka ネットワークと呼ぶ．人口については，平成 22 年国勢調査に関する地域メッシュ統計 [9] より抽出した，交差点の属する 2 分の 1 地域メッシュ (500m 四方) の人口比率を交差点に付与した．

3.2. 評価方法

通常の媒介中心性 (original) と確率媒介中心性 (proposed) のランキングの一致度を F 値を用いて定量的に評価する．

$$F(r) = \frac{2|B(r) \cap PB(r)|}{|B(r)| + |PB(r)|} = \frac{|B(r) \cap PB(r)|}{|B(r)|} \quad (4)$$

ここで $B(r)$, $PB(r)$ は，それぞれ媒介中心性，確率媒介中心性のランキング r 位までのノード集合である．

3.3. 実験結果

図2(a) から，art1 や art2 のようにネットワークの重心と人口分布の重心が近いときには，両手法のランキング結果の一致度を示す F 値から，大きく異ならないことがわかる．art3 では，ネットワーク重心と人口分布の重心が大きく異なるため，ランキング結果も大きく異なる．

図2(b) から，art4, art5, art6 とガウシアンカーネルのパラメータを大きくしていくと，人口分布の裾が広がるため，ネットワークの重心付近にも比較的人口が分布し，両手法の違いが小さくなる傾向にあることがわかる．以上のように人工データに対する結果より，1 節のはじめにでの推測が確認できた．

図3(a) から，両手法のランキング結果は全体を俯瞰すると大きくは異ならないが，ランキング上位に着目すると，抽出されたノード集合に差があることがわかる．実際に抽出された上位 40 ノードをプロットした図3(b)を見ると，通常の媒介中心性 (▲) では，新東名高速道路上に多くプロット

されている．新東名高速道路は，静岡県内の山間部（人口が少ない）を通過しているため，このような結果が得られたと考えられる．一方確率媒介中心性 (●) では，旧東名高速道路上に多くプロットされているのが見て取れる．旧東名高速道路は，静岡県内の市街地付近（人口が多い）を通過しているためこのような結果が得られたと考えられる．

4. おわりに

本稿では，現実ネットワークのノードが様々な点で一様でないことに着目し，ノード選択確率の概念を提案した．さらに，既存の中心性指標にノード選択確率を導入し確率中心性へ拡張した．オリジナルの媒介中心性と確率媒介中心性のランキング結果より，確率媒介中心性の方がより現実にもれた結果が得られ，さらに現実問題への応用が期待できることが示唆された．今後は，ソーシャルネットワークをはじめとした様々なデータに適用し，確率中心性の有用性を検証していきたい．

謝辞 本研究は科学研究費補助金 (No.25・10411) の補助を受けた．

参考文献

- [1] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, No. 3, pp. 215–239 (1979).
- [2] Newman, M. E. J.: A measure of betweenness centrality based on random walks, *Social Networks* (2005).
- [3] Thadakamalla, H. P., Albert, R. and Kumara, S. R. T.: Search in weighted complex networks, *Physical Review E*, Vol. 72, p. 066128 (2005).
- [4] Zhang, Y., Zhang, Z., Wei, D. and Deng, Y.: Centrality measure in weighted networks based on an amoeboid algorithm, *Journal of Information and Computational Science*, Vol. 9, No. 2, pp. 369–376 (2012).
- [5] Kalapala, V., Sanwalani, V., Clauset, A. and Moore, C.: Scale Invariance in Road Networks, *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, Vol. 73, No. 2 Pt 2, p. 6 (2006).
- [6] 池田哲夫，斉藤和巳，武藤伸明，伏見卓恭：複雑ネットワーク分析手法による道路網の分析，第 20 回地理情報システム学会研究発表大会 (GISA2011) (2011).
- [7] 伏見卓恭，斉藤和巳，池田哲夫，武藤伸明：ノード群の協調的振舞いに着目した集合媒介中心性の提案と応用，電子情報通信学会和文論文誌 D，Vol. J96-D, No. 5, pp. 1158–1165 (2013-05).
- [8] “全国デジタル道路地図データベース静岡県版,” 2012.
- [9] “平成 22 年国勢調査に関する地域メッシュ統計,” (公財) 統計情報研究開発センター (S i n f o n i c a) 2010.