

次数不変のネットワーク張り替えでの中心性の頑健性評価*

伏見 卓恭[†] 入月 卓也[‡] 永田 大[†] 斉藤 和巳^{†‡} 池田 哲夫^{†‡} 武藤 伸明^{†‡}

静岡県立大学経営情報学研究科[†] 静岡県立大学経営情報学部[‡]

1. はじめに

近年社会ネットワークの構造に関する分析が自然科学や社会科学分野で注目を浴びている。しかし、現実社会のネットワークにおいては、プライバシー等の問題から全体構造を把握することは難しい。社会調査などから得られるエゴセントリック情報（次数やクラスタ係数）をもとに、ネットワーク全体の構造を推定することは重要な研究課題である。ネットワークの構造推定の研究では真の全体構造が既知であることが必要である。実社会のネットワークの全体構造を入手することは一般に困難であるので、本研究ではブログや SNS に代表されるようなウェブ上でのネットワークを対象にネットワークの構造推定の実験を行う。

本研究では、上記推定結果を評価するための適切な尺度を探索するために、社会ネットワークにおける中心性の頑健性について分析する。ここで、上記のように社会調査から得られる情報をもとにすることから、次数一定と仮定する。詳細には、取得した実ネットワークに対して、各ノードの次数を一定にし、任意の確率でリンクを張り替え、張替ネットワークを生成する。この時張替確率が変わっても、ランキングの変動が少ない（頑健な）中心性を探究する。そこで、実ネットワークと張替ネットワークの中心性指標によるランキングの一致度により各中心性の頑健性を評価し、張替確率が大きくても頑健な中心性を確認する。中心性の頑健性に関する先行の研究としては、Ng[1] や Borgatti[2]の研究があるが、それらの研究では次数不変と仮定しておらず、この点が本研究と異なる。

2. 分析手法

本研究では、取得した実ネットワークに対してリンクを任意の張替確率に従い張り替え、張替ネットワークを生成し、これらネットワークに対して各種の中心性の値を求め頑健性を比較する。

* The Evaluation of Robustness of Centrality Measures by Edge Rewiring with Invariable Degree

[†] Takayasu Fushimi, Dai Nagata, Kazumi Saito, Tetsuo Ikeda and Nobuaki Muto. Graduate School of Administration and Informatics University of Shizuoka

[‡] Takuya Iriduki, School of Administration and Informatics University of Shizuoka

2. 1 張替ネットワークの生成

今回利用する張替ネットワークの生成方法として、Watts と Strogatz[3]のsmall-world実験と同様に、分析対象の実ネットワークのリンクを、張替確率に従い複数の張替ネットワークを生成する。具体的には、張替確率 $P=2^{-k}$ ($0 \leq k \leq 10$) で張替ネットワークを 11 個生成する。これらと実ネットワーク ($P=0$) を含め 12 個のネットワークについて分析する。張り替えを行う際、以下の 2 点に留意する。実ネットワークから個々のノードの次数を変えずにリンク張り替えを行う。また、自分から自分へのリンク (self-link) や同じ相手ノードに対して複数のリンク (multiple-link) を張ってはならない。こうして張り替えたネットワークは、実ネットワークと比較して構造が異なることから、実ネットワークにリンク張り替えによるエラーが入ったと仮定することができる。

2. 2 頑健性の評価方法

上記の方法で生成した全てのネットワークに対して中心性指標の値を計算する。ネットワークの中心性指標として、従来研究で用いられている次数中心性、固有ベクトル中心性、近接中心性、媒介中心性を用いて分析する。固有ベクトル中心性は、HITS アルゴリズムにより求めた固有ベクトルおよび PageRank スコアを用いる。全てのネットワークにおいて、ノードを中心性指標の値で降順に並び替える。当然エラーを入れた張替ネットワークと実ネットワークとでは、ランキングが変化している。しかし変化の傾向は中心性指標により異なる。そこで、リンク張り替えというエラーに耐えられるような（頑健な）中心性を確認するために、実ネットワークの中心性によるランキングとの一致度を用いてその頑健性を評価する。一致度には Jaccard 係数を用いるが詳細な定義は下記に示す。

$A(r)$: 実ネットワークにおける中心性 r 位までのノード集合。

$A_p(r)$: 張替確率 P の張替ネットワークにおける中心性 r 位までのノード集合。

$$J(A(r), A_p(r)) = \frac{|A(r) \cap A_p(r)|}{|A(r) \cup A_p(r)|}$$

なお、次数不変で実験を行っているので、任意

の張り替えで次数中心性のランキングに対する Jaccard 係数は常に 1 である。

3. 評価実験

3.1 実験データ

本研究では 2 つのネットワークデータを用いて分析する。

1 つ目は、ブログのトラックバックネットワーク (以下ブログネットワークと呼ぶ) のデータで、ノード数は、12,047 でリンク数は、79,920 である。

2 つ目は、日本の「ウィキペディア」内の「人名一覧」からの人物ネットワーク (以下ウィキペディアネットワークと呼ぶ) で、ノード数は、9481 でリンク数は、245,044 である。これらのネットワークには、次数分布がべき則分布に従うなど多くの大規模ネットワークと同様な特徴が見られる。

3.2 分析結果と考察

これら 2 つの各ネットワークデータに対して、2.1 の方法により 11 個の張替ネットワークを生成し、各種中心性指標の値を求める。そして、中心性ごとに値の降順にノードを並び替え、Jaccard 係数の値をプロットする。図 1 はブログネットワークの媒介中心性のランキングに対する Jaccard 係数、図 2 は HITS のランキングに対する Jaccard 係数をプロットしたものである。媒介中心性は、張替確率が小さくても Jaccard 係数は小さくなりランキングの値は不安定である。図示はしていないが、近接中心性は媒介中心性と同様の傾向が見られる。一方 HITS は張替確率をある程度大きくしても、上位数百位くらいまでの Jaccard 係数は 1 に近い値でランキングの値は安定的であることが示唆される。図示はしていないが、PageRank は HITS と同様の傾向が見られる。ウィキペディアネットワークでもほぼ同様の結果が得られた。

この結果は、先行研究である [1] や [2] とは異なるものである。Ng ら [1] は、HITS の値は不安定であると主張している。[1] では、eigengap の小さいネットワークデータに対して 30% のノードを取り除くというエラーを入れている。また Borgatti ら [2] は、媒介中心性や近接中心性も同様に頑健であると主張している。[2] ではノードの追加・削除、リンクの追加・削除のエラーを入れる実験をしている。一方我々は、次数不変でリンクを張り替えている。このようなエラーの下では、媒介中心性などの中心性指標より HITS の値は安定であることが実験より示唆された。

4. おわりに

本研究で我々は、次数不変という条件下でリンクをランダムに張り替えるというエラーを入れ、ネットワークの各種中心性の頑健性を評価した。

実験の結果、近接中心性、媒介中心性より固有ベクトル中心性の方がより頑健であることが示唆された。我々の今後の研究としては、中心性ランキング間の Jaccard 係数や eigengap、ネットワーク構造と頑健な中心性指標との関係などを分析し、固有ベクトル中心性が頑健な指標である原因を追究していくつもりである。

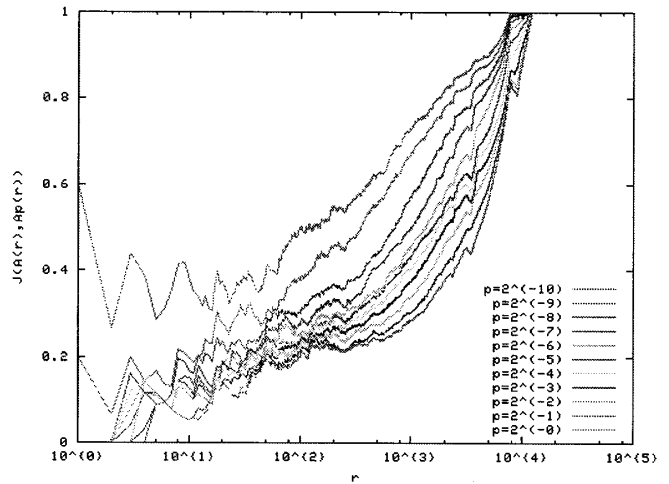


図 1 Betweenness Centrality of Blog Network

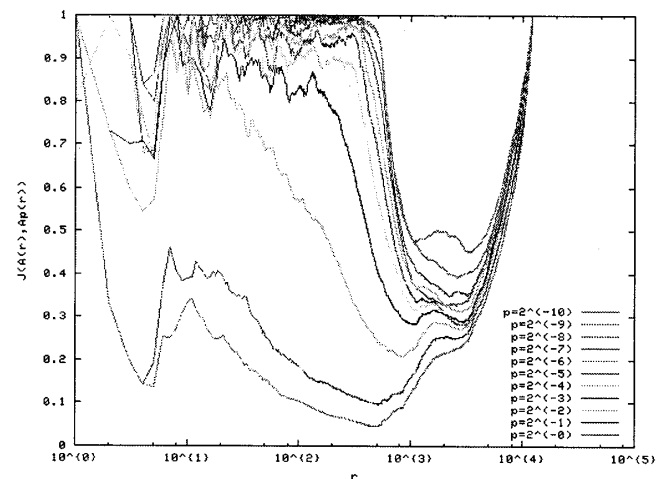


図 2 Eigenvector Centrality (HITS) of Blog Network

参考文献

- [1] A. Y. Ng, A. X. Zheng and M. I. Jordan, "Stable Algorithms for Link Analysis", *24th Ann. Int ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 258-266, (2001).
- [2] S. P. Borgatti., K. M. Carley. and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data", *Social Networks* 28, pp. 124-136, (2006).
- [3] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature* 393, pp. 440-442, (1998).