6J-1

社会ネットワークにおける中心性の定量的な頑健性評価法の提案*

伏見　卓恭†　　斉藤　和巳†　　武藤　伸明†　　池田　哲夫†

静岡県立大学経営情報学研究科†

1 はじめに

複雑ネットワークに関する研究が自然科学，社会科学を問わず広い分野で注目を浴びている．大規模なネットワークを対象とする際に，その全体の構造を把握することは困難な場合がある．特に，友人関係などの社会ネットワークはプライバシー等の問題から全体構造を正確には把握できない．社会調査などから得られるエゴセントリック情報（友人数など）をもとに全体構造を推定し，構造分析や重要ノードを抽出することは重要な研究課題である．

重要ノード抽出のための中心性指標として PageRank や HITS，近接中心性などが挙げられる．これら中心性指標値は，ネットワークにリンクの張替や missing リンクなどのノイズが混入すると，ランキングが必然的に変化する．ノイズが含まれていてもランキングがあまり変化しない中心性が頑健な指標とみなすことができる．構造推定などの場面では，ネットワークを 100%正確に推定することは不可能であり，ノイズが混入することは不可避である．このような状況下での重要ノード抽出のために，頑健な中心性が必要となる．Ng ら[1]や Borgatti ら[2]の研究ではノードやリンクを追加・削除することによりノイズを混入させ，ピアソン相関係数を用いて頑健性を評価している．

本研究では，ノイズを混入したネットワークでの頑健な中心性を評価するための直観的かつ適切な定量的評価尺度を提案する．

2 分析手法

観測したネットワークデータ（以後実ネットワーク）からノイズ混入率 p に従ってノイズを混入させたネットワーク（以後張替ネットワーク）を生成する．実ネットワークと張替ネットワークに対して各種中心性値を計算し，その値で降順にランキングする．ランキングの m 位までのノード集合の一致度を Jaccard 係数を用いて定義する．

$X(m)$：実ネットワークにおける中心性 m 位までのノード集合．

$Y_P(m)$：ノイズ混入率 p の張替ネットワークにおける中心性 m 位までのノード集合．

$$J(X(m),Y_P(m)) = \frac{|X(m) \cap Y_P(m)|}{|X(m) \cup Y_P(m)|}$$

Jaccard 係数の値を縦軸にとり，m を横軸にとり，各張替確率ごとの値をプロットする(図 1)．各曲線は，実ネットワークと異なるノイズ混入率の張替ネットワークとの Jaccard 係数を表す．Jaccard 係数が大きいことは，実ネットワークでの重要ノードを張替ネットワークでも重要ノードとして抽出している割合が高いことを示す．グラフによる定性的評価[3]に加え，本研究ではグラフすなわち中心性の頑健性を直観的かつ適切に示す定量的な尺度を導入する．

2.1 提案尺度

各ノイズ混入率 p の描く曲線と横軸で囲まれた領域の面積 S_p を求める（以後面積尺度と呼ぶ）．この時横軸に対数をとると次式が得られる．

$$S_p = \sum_{m=1}^{N} (\log(m+1) - \log(m)) * J(X(m),Y_p(m))$$
$$= \sum_{m=1}^{N} \log\left(\frac{m+1}{m}\right) * J(X(m),Y_p(m))$$

この面積尺度は重要ノードの観点から，ランキングの上位の重みを大きく，下位の重みを小さくした Jaccard 係数の重み和になっている．実際には，ネットワークの規模による違いをなくすために， log(N+1) で除して正規化を行う．この値が 1 に近い値を維持しているほど頑健な中心性指標であることを示す．

3 評価実験

提案尺度の評価のために，ブログのトラックバックネットワーク（ノード数 12,047，リンク数 39,960）を採用する．このネットワークに対してリンクの張替というノイズを与える．リン

* 「Proposing the Quantitative Measure of Centrality Robustness in the Social Networks」

†「Takayasu Fushimi, Kazumi Saito, Nobuaki Muto and Tetsuo Ikeda, Graduate School of Management and Information, University of Shizuoka」

Copyright ©2011 Information Processing Society of Japan.
All Rights Reserved.

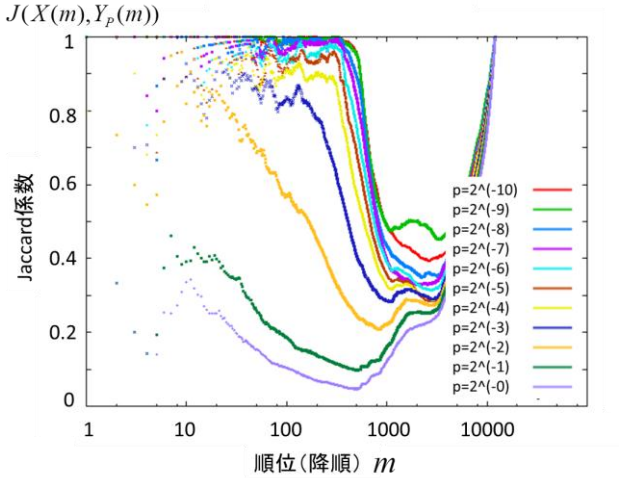

図 1 Blog HITS ランキングの Jaccard 係数

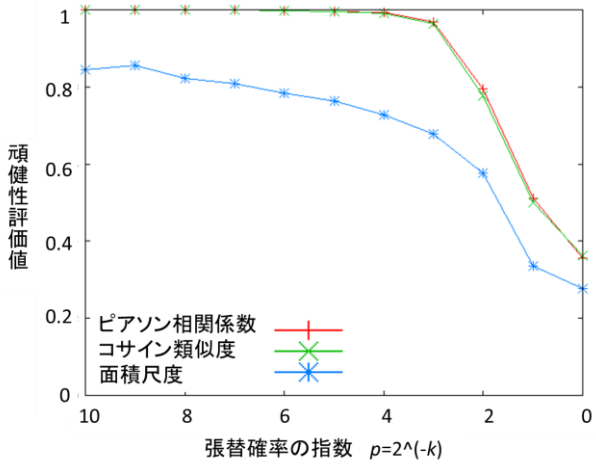

図 2 Blog HITS 頑健性評価値

ク張替確率 p = 2^(-k) (k = 0, …, 10)に従い次数不変でリンクを張り替え，11 個の張替ネットワークを生成する．実ネットワークを加えた 12 個のネットワークに対して中心性値を計算し降順にランキングする．図 1 は各張替確率での張替ネットワークと実ネットワークとの m 位までの Jaccard 係数をプロットしたものである．このグラフの各曲線を提案尺度により定量的に表す（図 2）．

ナイーブな尺度として，ピアソン相関係数，コサイン類似度が考えられる．ネットワークのノード数を N，ノード i の実ネットワークでの中心性値 を x_i，張替ネットワーク（張替確率 p ）での中心性値を $y_{p,i}$ とする．ピアソン相関係数 r_p およびコサイン類似度 θ_p は以下になる．

$$r_p = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_{p,i} - \bar{y}_p)}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_{p,i} - \bar{y}_p)^2}}$$

$$\theta_p = \frac{\sum_{i=1}^{N} x_i y_{p,i}}{\sqrt{\sum_{i=1}^{N} x_i^2}\sqrt{\sum_{i=1}^{N} y_{p,i}^2}}$$

ここで，オーバーバーは平均値を示す．これらの値が大きければ，実ネットワークと張替ネットワークでの中心性指標値が高い相関関係にあることを示す．図 2 は横軸に張替確率を対数軸でとっている．図 1 の各曲線が図 2 の各点に対応する．実験結果では図 1 のように張替確率に関係なく 1000 位あたりで Jaccard 係数が低下しているが，図 2 のピアソン相関係数とコサイン類似度では，張替確率 10^(-4)まで反映されず，ほぼ 1 の値を示している．一方提案する面積尺度では，図 1 の Jaccard 係数低下状態が反映されている．相関係数やコサイン類似度で反映できない原因として，ランキングが変化しても中心性値の差が小さいことが考えられる．また，張替確率の増大に伴う Jaccard 係数の低下も図 2 の面積尺度では反映できている．図示はしていないが，同様の実験をいくつかの大規模ネットワークに対して行い，既存尺度に比べて面積尺度が直観的に反映されている結果が得られている．

4 おわりに

本研究では，ネットワークの中心性の頑健性を評価するための定量的な尺度を提案した．結果として提案尺度は，ピアソン相関係数やコサイン類似度では反映されていなかったランキングの変化を直観的かつ適切に反映できることが実験により示唆された．また上位の重みを相対的に大きくする直観的な重み和尺度として横軸に対数をとるという自然な変換により実現した．

謝辞　本研究は，科学研究費補助金基盤研究(C) (No. 22500133) の補助を受けた．

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] 伏見卓恭，入月卓也，永田大，斉藤和巳，池田哲夫，武藤伸明，“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”，第 72 回情報処理学会全国大会，(2010).

Copyright ©2011 Information Processing Society of Japan.
All Rights Reserved.