HOKKAIDO UNIVERSITY

Title	社会的ネットワークにおける効率的な重要ノードの発見の手法とその応用［論文内容及び審査の要旨］
Author(s)	田畑, 公次
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12816号
Issue Date	2017-06-30
Doc URL	https://hdl.handle.net/2115/66526
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Koji_Tabata_abstract.pdf, 論文内容の要旨



Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　田畑 公次

学 位 論 文 題 名

社会的ネットワークにおける効率的な重要ノードの発見の手法とその応用
（Efficient methods for finding important nodes in social networks and their applications）

近年, インターネット上でのソーシャルネットワーキングサービスの普及により, 人と人との様々なつながり方がネットワークデータとして得られるようになってきた. ネットワーク上での重要なノードを見つける研究は従来から行われてきたが, 大規模なデータも入手可能となってきたという状況の中, 効率的に重要ノードを発見する手法の需要が高まってきている.

ネットワーク上でのノードの重要性を測る尺度として, 距離中心性 (closeness centrality), 媒介中心性 (betweenness centrality), 離心中心性 (eccentricity centrality), 次数中心性 (degree centrality), 固有ベクトル中心性 (eigenvector centrality) など様々な中心性が用いられている. さらに近年ネットワーク上での情報伝播に基づいてノードの重要性を評価する影響度 (influence) なども提案されている. これらの尺度は発見するノードの使用目的により使い分けられる.

例えば, あるノードの距離中心性は, そのノードから他のノードまでの距離の逆数で定義される. 距離中心性が最大のノードは, 言い換えると他のノードに至るまでの平均距離が最小となるノードである. したがって, 距離中心性を尺度として用いると, どのノードからも比較的アクセスしやすく重要な位置にあるノードを見つけることが可能である. ネットワーク及びユークリッド空間上で距離中心性が最大のノードを見つける問題は 1-median 問題と呼ばれる古典的な最適化問題であり, 最適施設配置問題としても扱われてきた.

また, あるノード集合の影響度は, その集合に属するノードからネットワークを介して情報伝達が行われると仮定した場合に, 情報が伝わるノード数の期待値で定義される量である. インターネット上でのソーシャルネットワークを介した方が, オフラインでのソーシャルネットワークを介するより高速かつ広範囲に情報が広がりやすいため, インターネット上での口コミを利用してマーケティングを行う動きが高まっている. このようなマーケティングは, 情報がウイルスが拡散するように指数関数的に広範囲に広がっていく様子に例えてバイラルマーケティングと呼ばれている. 影響度の尺度は, バイラルマーケティングにおいて効果的なノード集合を見つけるのに使うことができるため, 近年注目されている.

本研究では, ネットワークにおけるノードの重要度を測る尺度として, 距離中心性と影響度の 2 つに焦点を絞り, それらの指標を用いた重要ノードの効率的な発見法を提案し, それらの有効性を示す. 距離中心性の尺度においては, 1-median 問題に対する反復法を用いた効率的な近似アルゴリズムを提案する. また, 影響度の尺度に関しては, Independent Cascade (IC) モデルと呼ばれる辺毎に伝播確率が与えられる情報伝達モデルにおいて, 二端子直並列グラフに対して影響最大化問題を正確 (exact) に解くアルゴリズムを提案する. 以下にそれぞれの概要と, 本研究における特徴について述べる.

[1-median 問題に対する反復法を用いた効率的な近似アルゴリズム]

グラフにおける 1-median 問題とは, 辺に長さのある無向グラフ $G = (V, E)$ において, ノード

$v \in V$ から他のノードまでの距離の和を $d(v)$ としたとき, $d(v)$ を最小化するようなノード v を見つける問題である. 提案手法では, ノード $v_0 \in V$ に対して v_0 を根とする全域木を含む部分グラフ G' を考えると, G' においてあるノード v から他のノードまでの距離の和 $d'(v)$ が $d'(v) \geq d(v)$ を満たすことを利用して, 反復的により良い解を求める手続きを行うことで近似解を求める. 全域木を含む部分グラフ G' として単純なグラフを用いることにより, 効率的に $d'(v)$(またはその上界) を求めることができるため, 1 回の反復にかかる時間を短くでき, 更に実験によれば少ない反復回数で収束するため, 高速な計算が可能である. 提案法では 1 回の反復を $O((n+m)\log n)$ で行うことができる部分グラフを用いた.

すべてのノードに対して最短経路木を作成することにより $d(v)$ を計算すれば, $O(n^2 \log n + mn)$ 時間で問題を正確 (exact) に解くことが可能であるが, これは巨大なグラフにおいて実行が困難である. 実データや人工データを用いた実験によれば, 本手法はこの手法より 100 倍程度高速に動作する. また DTZ と呼ばれる従来の近似手法と比較してもスケールフリーネットワークに対して精度と実行速度の両方において優れていることが実験により示されている.

[二端子直並列グラフに対して影響最大化問題を正確に解くアルゴリズム]

影響最大化問題は, 与えられた小さな整数 k に対し, k 要素の情報源ノード集合 S で, その影響度 $\sigma(S)$ を最大化するものを見つける問題である. この問題は, IC モデルなどの一般的な情報伝達モデルでは NP 困難であることが知られているため, 貪欲法による効率的な近似アルゴリズムの研究が盛んに行われている.

本研究では,IC モデルにおいて, 扱うグラフを二端子直並列グラフに制限したときに影響最大化問題を正確 (exact) に解くアルゴリズムを提案する. 従来知られている一般グラフに対する貪欲法による近似解法では得られる解集合の影響度が最適解に比べ $(1-1/e)$ 倍以上になることが理論的に保証されているものの, ランダムに生成した二端子直並列グラフにおいては最適解の 0.8 倍程度の近似解しか得られない.

従来, 効率的に影響最大化問題の最適解を求める手法は, 木に対しては動的計画法を用いた手法が知られていたが, 二端子直並列グラフなどの木よりも複雑なクラスに対しては知られていなかった. 提案アルゴリズムでは, ある種の極大性を満たすものしか解に成り得ないという性質を利用して候補を絞ることにより効率化を図っている. 分割統治法を用いて, 部分グラフの解の候補を合成することにより全体のグラフに対する解の候補を計算するが, 解の候補の合成をする際に極大性のチェックを行うことにより解の候補を常に少なく保つことができる. 実験によれば, ランダムに生成した 800 ノードの二端子直並列グラフに対し, 提案法を用いれば $k = 7$ の影響最大化問題の最適解を 1 分程度で求めることが可能である.