

複雑ネットワーク上のコスト付き汚染拡散最小化問題に対する 解法の提案と評価

和泉元 陽介 †

小野 景子 †

熊野 雅仁 †

木村 昌弘 †

† 龍谷大学理工学部電子情報学科

1 はじめに

複雑ネットワークにおいて好ましくない情報の拡散を防ぐ有効な戦略の開発は重要課題であり、これまでにいくつかの研究がなされてきたが [1]、単位時間あたりに除去できるノード数やリンク数の制限については考慮されていなかった。しかしながら、インフルエンザ感染において単位時間あたりに接種できるワクチン数などに制限があるように、現実問題ではそれらに制限がある場合が多い。本論文では、IC モデル [2] に対して、単位時間あたりに除去できるノード数やリンク数に制限を設けたコスト付き汚染拡散最小化問題を定式化し、次数、直径、ページランクおよび、クラスター係数に基づいた除去法を提案する。大規模ネットワークを用いた実験により、それらの性能を比較評価する。

2 問題の定式化

2.1 汚染拡散モデル

ある好ましくない情報がネットワーク $G = (V, E)$ を通じて拡散する過程の数理モデルとして、Kempe [2] らの研究にしたがって IC モデルを仮定する。そのような好ましくない情報によって汚染されたノードを、“アクティブノード”と呼ぶことにする。初期アクティブノード集合が V_0 である時の最終的なアクティブノード総数 (感染ノードの総数) を、 $\sigma(V_0; G)$ とする。本研究では、 $\sigma(V_0; G)$ を最小化するようなリンク除去法について考える。

2.2 コスト付き汚染拡散最小化問題

ある好ましくない情報が IC モデルに基づいてネットワーク上を拡散していくとする。このとき、指定された数のリンク群を適切に除去することにより、その好ましくない情報に感染するノード数を最小化することを目指す。コスト付き汚染拡散最小化問題とは、汚染拡散最小化問題に除去コストの考えを組み入れた問題であり、各時刻で除去可能なリンク数 (リンクコスト) を指定して、それ以下の数のリンク群を除去することにより汚染ノード数の最小化を目指すものである。

3 提案法

各時刻 t において、 t で新たにアクティブとなったノードの集合を V_t とし、各 $v \in V_t$ の非アクティブな近傍ノードを、時刻 t における“フロンティアノード”と呼ぶ。また、ノード $v \in V_t$ とフロンティアノード w を結ぶリンク $(v, w) \in E$ を、時刻 t における“フロンティアリンク”と呼ぶ。時刻 t における非アクティブノード集合 $W_t = V \setminus \bigcup_{s=0}^t V_s$ に対して、グラフ G のノード集合 W_t への誘導グラフを $G_t = (W_t, E_t)$ とする。

3.1 汚染周り次数法

各時刻 t に対して、 t におけるフロンティアノード群をグラフ G_t 内で次数の高い順にソートし、さらに、それらに繋がっているフロンティアリンク群を抽出することにより、 t におけるフロンティアリンク群をソートする。汚染周り次数法では、このソート順にしたがって、指定されたリンク数だけフロンティアリンクを除去する。

3.2 汚染周り直径法

各時刻 t に対して、 t におけるフロンティアノード群を、グラフ G_t 内での他ノードとのグラフ間距離の平均値が小さい順にソートし、さらに、それらに繋がっているフロンティアリンク群を抽出することにより、 t におけるフロンティアリンク群をソートする。汚染周り直径法では、このソート順にしたがって、指定されたリンク数だけフロンティアリンクを除去する。

3.3 汚染周りページランク法

各時刻 t に対して、 t におけるフロンティアノード群をグラフ G_t 内でページランク値の高い順にソートし、さらに、それらに繋がっているフロンティアリンク群を抽出することにより、 t におけるフロンティアリンク群をソートする。汚染周りページランク法では、このソート順にしたがって、指定されたリンク数だけフロンティアリンクを除去する。実験では、ページランク法のダンピング・ファクター値は 0.85 を用いた。

3.4 汚染周りクラスター係数法

各時刻 t に対して、 t におけるフロンティアノード群をグラフ G_t 内でクラスター係数の高い順にソートし、さらに、それらに繋がっているフロンティアリンク群を抽出することにより、 t におけるフロンティアリンク群をソートする。汚染周りクラスター係数法では、こ

Solving contamination minimization problems with costs on complex networks

Yosuke IZUMIMOTO†, Keiko ONO‡, Masahito KUMANO‡ and Masahiro KIMURA ‡

†Division of Electronics and Informatics, Ryukoku University

‡Department of Electronics and Informatics, Ryukoku University

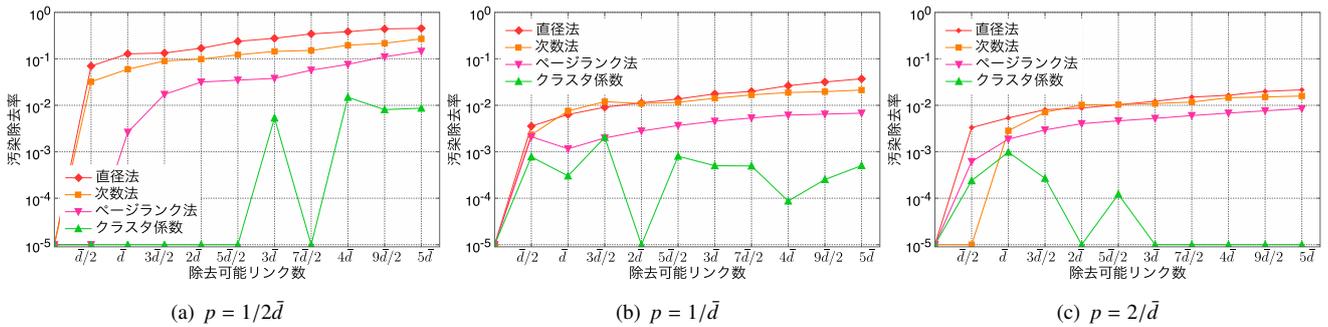


図 1: Wikipedia ネットワークの結果

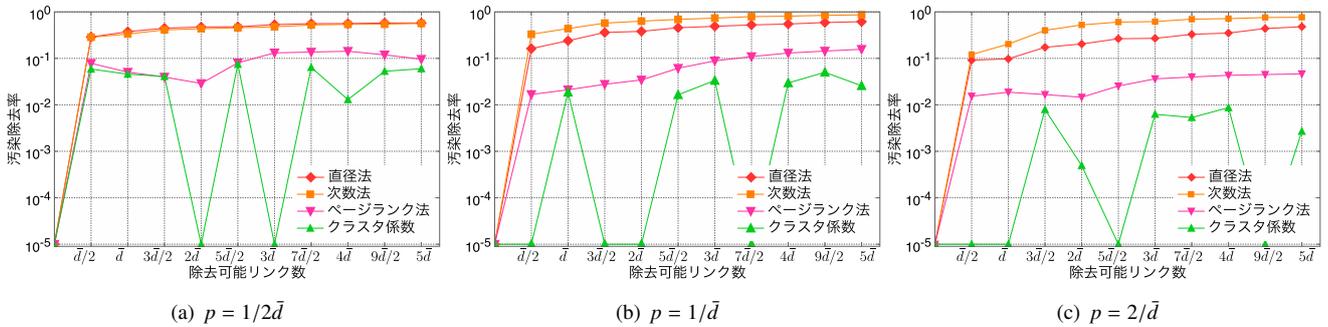


図 2: ブログネットワークの結果

のソート順にしたがって、指定されたリンク数だけフロンティアリンクを除去する。

4 実験

4.1 実験データ

実験では、社会ネットワークの顕著な特徴を多く持つ大規模な実世界ネットワークの実データとして、Wikipedia ネットワーク [1] とブログネットワーク [1] を用いた。各ネットワークの統計量を表 1 に示す。IC モデルでは、拡散確率 p を指定する必要がある。実験では、拡散確率が通常、低い、高い場合として、 $p = 1/\bar{d}$, $p = 1/2\bar{d}$, $p = 2/\bar{d}$ の 3 つの場合を調べた。ここに、 \bar{d} はネットワークの平均次数である。単位時間あたりに除去可能なリンク数 L については、 $L = \bar{d}/2, \bar{d}, \dots, 5\bar{d}$ の 10 パターンを調べた。また、初期アクティブノード集合 V_0 は、最大次数のノードから時刻 $t = 1$ で感染したノード集合とした。

表 1: ネットワークデータの統計量

ネットワーク	Wikipedia	ブログ
ノード数	9481	12047
最大次数	585	222
平均次数	12.9	3.3
リンク総数	122,522	39,960
平均ノード間距離	4.7	8.17

4.2 実験結果

実験結果を図 1, 2 に示す。縦軸は“汚染除去率”(= $1 - \{ \text{感染ノードの総数} \} / \sigma(V_0; G)$)であり、1,000 回

の試行での平均値をプロットしている。まず、拡散確率が低い場合 ($p = 1/2\bar{d}$) では、汚染周り直径法の性能が良いことが観察される。汚染周りクラスタ係数法は、常に最も性能が悪かった。また、Wikipedia ネットワークでは汚染周り直径法の性能が良く、ブログネットワークでは汚染周り次数法の性能が良かった。拡散確率がある程度大きくなると、どの手法が効果的かは、ネットワーク構造に依存することが示唆される。提案法での $L = 5\bar{d}$ における汚染除去率は、最大で 86%、最小で 3% であった。

5 まとめ

IC モデルに対して、単位時間あたりに除去可能なリンク数に制限を設けたコスト付き汚染拡散最小化問題を定式化した。その解法としてネットワーク中心性に基づく 4 手法を提案し、大規模な実ネットワークを用いた実験でそれらの性能を比較検証した。より高品質な解を求めることのできる効率的手法の構築が、今後の課題である。

参考文献

- [1] M. Kimura et al, Blocking links to minimize contamination spread in a social network, ACM Trans. Knowl. Discov. Data, Vol.3, No.9, 1-23, 2009.
- [2] D. Kempe et al, Maximizing the spread of influence through a social network, KDD'03, 137-146, 2003.