

社会ネットワーク上での情報拡散シミュレーションの高速化

李 孝岩 齊藤 和己

神奈川大学大学院 理学研究科

1. はじめに

近年、情報拡散の観点から社会ネットワークを研究することが注目されており、中でも最も多く研究されている課題が影響力最大化問題である [1]。影響力最大化問題とは、影響度の高いノードの組み合わせを見つけ出すことである。この問題は NP 困難な最適化問題であるため、一般にはその近似解を効率的に求めることが目的となり、これまでに多くの手法が提案されてきた [1]。この問題に関する現在の既存研究においては、確率に基づく基本的な情報拡散モデルである独立カスケード (IC: Independent Cascade) モデルや線形閾値 (LT: Linear Threshold) モデルが多く用いられている [1]。先行研究では、影響力最大化問題を解くための前提となる課題であるノードの可到達ノード数の計算において、高速化手法が提案されているが [2] [3]、近年では TikTok や Instagram といったショート動画アプリの急速な普及により、社会ネットワークの規模および構造はさらに大規模かつ複雑になってきている。そこで本研究では、さらなる効率向上を目的として、新たな高速手法を提案する。

あるネットワーク $G = (V, E)$ において、先行研究では ボンドパーコレーション法 (Bond Percolation, BP 法) [2] [3] に基づいて部分ネットワーク G_m を生成し、その後、強連結成分 (Strongly Connected Components, SCC) の分解によって商グラフ Q_m を構築するという手法が提案されている。さらに、この商グラフ Q_m 上において、REP (辺削除) および MCP (ノード削除) という二種類の枝刈技術を導入しており、本稿ではこの従来法を SRM 法と呼ぶ。SRM 法の導入により、可到達ノード数の計算時間は大幅に削減された。しかしながら、ネットワークの規模が拡大するにつれて、計算時間も増加する傾向にある。そこで本研究では、SRM 法を基盤としつつ、新たな枝刈技術を導入し、これを BOP と呼ぶ。これにより、さらなる計算時間の削減を実現した。本稿では、この新手法を SRMB 法と称し、SRM 法と比較して計算効率が大幅に向上したことを確認する。

2. 提案手法

初期ネットワークに対して強連結成分分解を行い、得られた商グラフを Q_m とする。 Q_m 内で最大の強連結成分 (MAXSCC) を C_m とし、 C_m の要素数を $|C_m|$ と表す。 C_m から可到達な他の強連結成分を集合 D と定義する。このとき、 Q_m における C_m から到達可能な強連結成分集合を

$$N_{Q_m}(C_m) = \{D_1, D_2, D_3, \dots\}$$

とする、集合 D の要素数を $|N_{Q_m}(C_m)|$ と表す。 C_m に向かって指す任意の強連結成分 B に対して、次が成り立つ：

$$\{B \mid C_m \in N_{Q_m}(B)\}$$

そして集合 B の要素数を $|L_{Q_m}(C_m)|$ と表す。

$$|L_{Q_m}(C_m)| = |\{B \mid C_m \in N_{Q_m}(B)\}|$$

次に、 Q_m から $N_{Q_m}(C_m)$ を除去した誘導部分グラフを S_m と定義する。このとき、任意の強連結成分 $B \in N_{Q_m}(C_m)$ に対して、次の式が成り立つ：

$$|N_{Q_m}(B)| = |N_{Q_m}(C_m)| + |N_{S_m}(B)|$$

このとき、 Q_m において B および C_m を除く任意の強連結成分 E に対して、次の関係が成り立つ：

$$|N_{Q_m}(E)| = |N_{S_m}(E)|$$

以下に BOP 手法の図を示す

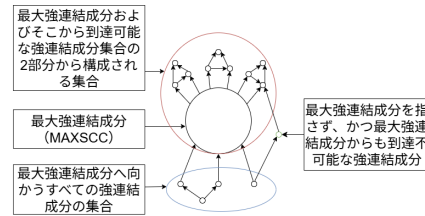


図 1: BOP 手法の説明図

3. 実験

本実験では、異なるネットワーク環境下における SRM 法と SRMB 法の実行時間を比較し、高速化の効果を検証する。

3.1 実験設定

異なるネットワークにおいても期待される高速化効果が得られることを確認するため、本実験では 3 種類の異なる有向ネットワークを用いた。第 1 のデータセットには先行研究と同様に twitter のデータを使用し、第 2 および第 3 のデータセットは Stanford Network Analysis Project より取得したものである。詳細は表 1 に示す。

表 1: データ

データ名	ノード数	リンク数
twitter	1,088,040	15,737,1628
pokec	1,632,803	30,622,564
LiveJournal	4,847,571	68,993,773

本研究では IC モデルの拡散倍率を人為的に変更し、異なるネットワーク環境を模擬している。本実験の目的は、実行時間における高速化効果を検証することである。多様な拡散確率での結果を評価するため、拡散倍率は全部で 6 種類設定し、それぞれ 0.125, 0.25, 0.5, 1, 2, 4 とした。実

行結果は秒単位で記録し、安定した結果を得るため、各拡散倍率に対してランダムに生成した 10 個のネットワークに対してそれぞれ 10 回ずつ実行した。すなわち、1 つの拡散倍率につき合計 100 回の実行を行い、その平均値を比較用の結果として採用した。

3.2 実験結果

まず、SRM 法および SRMB 法の実行時間を計測し、平均偏差平均と標準偏差を算出した。次に、異なる拡散倍率における実行時間の差異を明らかにするために、各実験条件において最大強連結成分 C_m のノード数を $|C_m|$ 、可到達成分数 $|N_{Q_m}(C_m)|$ 、被化到達成分数 $|L_{Q_m}(C_m)|$ をそれぞれ集計した。これらの実験結果は、図 2、図 3、図 4 に示す。

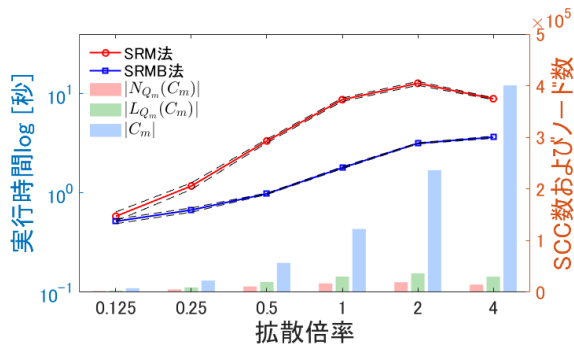


図 2: Twitter の実行時間の比較図

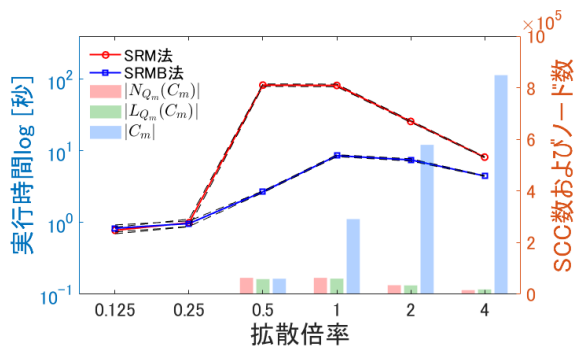


図 3: pokec の実行時間の比較図

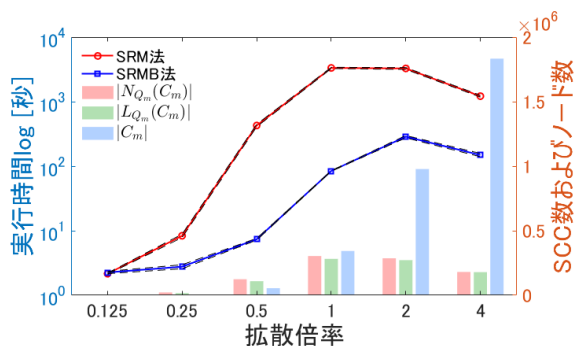


図 4: LiveJournal の実行時間の比較図

3 つの異なるネットワークデータに対する実験結果を観察すると、拡散倍率の増加に伴い、SRM 法および SRMB 法の実行時間は当初増加した。しかし、拡散倍率がさらに大きくなると、両手法の実行時間はむしろ減少する傾向が見られた。これは、拡散倍率の上昇により最大 SCC (青色のバー) が拡大し続けたためである。SCC 内では、1 つのノードについて到達可能ノード数を計算すれば、その SCC 全体の到達ノード数が得られるため、処理時間が逆に短縮されたと考えられる。

3 つのデータセットにおける高速化の結果を観察したところ、多くの場合において、従来法 SRM と比較して新手法 SRMB の実行時間は全体的に短縮されたことが確認できた。しかし、pokec における拡散倍率 0.125 および 0.25、ならびに LiveJournal における拡散倍率 0.125 では、実行時間の短縮は見られず、むしろわずかに増加する結果となった。この原因として、新たに導入した枝刈手法 BOP の効率は、 C_m から到達可能な SCC 数 (赤色のバー) および C_m へ到達可能な SCC 数 (緑色のバー) の 2 つの要素に依存している点が挙げられる。これらの値が大きいくほど BOP の効果が高まる一方、上記の条件下では両者の値が極めて小さく、BOP がほとんど機能しなかった。さらに、SRMB は従来法 SRM に比べて追加の処理ステップを含むため、その分実行時間が増加する結果となったと考えられる。図の折れ線の下に示されている破線は、それぞれの平均値に対して平均偏差の 2 倍を加減した範囲を示している。この結果から、SRM 法の偏差は SRMB 法に比べてやや大きい傾向が見られる。

4 おわりに

本研究では、先行研究における SRM 法を基盤とし、到達可能ノード数を計算する新たな手法として SRMB 法を提案した。本手法は多くの場合において高速化を実現した。今後の課題としては、本研究で用いた SRMB 手法を基盤とし、さらなる高速化を追求することである。また、本手法の実用性が今後さらに向上することも期待される。

参考文献

- [1] D. Kempe, J. Kleinberg and E. Tardos, "Maximizing the spread of influence through a social network," In Proceedings of KDD' 03, pp. 137 - 146, 2003.
- [2] K. Ohara, K. Saito, M. Kimura, and H. Motoda, "情報拡散モデルに基づく社会ネットワーク上の影響度分析," in Proc. 2014 Int. Conf. Data Science and Advanced Analytics (DSAA 2014), Tokyo, Japan, 2014, vol. 1, pp. 216 - 222.
- [3] Kimura, M., Saito, K., Ohara, K., & Motoda, H. (2016). Speeding-up node influence computation for huge social networks. International Journal of Data Science and Analytics, 1(1), 3 - 16.