

社会ネットワークの構造と情報拡散モデルの関係分析 Analysis of Relationships Between Social Network Structure and Information Diffusion Models

伏見 卓恭† 吉川 友也† 齊藤 和巳† 元田 浩‡ 木村 昌弘§
Takayasu Fushimi Yuya Yoshikawa Kazumi Saito Hiroshi Motoda Masahiro Kimura

1. はじめに

社会ネットワーク上での「クチコミ」による情報拡散や、新型ウイルスの伝染が世間を賑わせているように、近年社会ネットワークにおける情報の拡散は注目を浴びている。このような情報拡散が起こる代表的な社会ネットワークには、ウェブ上でのブログのトラックバックネットワークなどが考えられる。すなわち、各人や各ブログなどをノードとし、ブログ間トラックバックなどその繋がりをリンクとする社会ネットワークで情報拡散現象を分析できる。このような社会ネットワークでは、どのようなネットワーク構造において、より多くまたはより広く情報が拡散するか(影響最大化)ということが重要になってくる。すなわちネットワーク構造の違いによる情報拡散の変化を分析することは重要な研究テーマである。

そこで我々は、情報拡散の基本的なモデルである独立カスケードモデル(Independent Cascade model)と線形閾値モデル(Linear Threshold model) [1]を用いて分析することとする。対象とするネットワークデータにはブログのトラックバックネットワークとウィキペディアの人名ネットワークを用いる。これら2つのネットワークに対して、張替確率を設定し、各ノードの次数を変えずにオリジナルネットワークのリンクを張替確率に従ってランダムに張替したネットワークを生成して分析する。

我々は以前の研究で、オリジナルのネットワークと比較して確率1でランダム張替をしたネットワークの方が期待影響度が大きくなることを、さらにICモデルの方がLTモデルよりネットワーク構造の影響を受けやすいことを実験で確認している[2]。

本論文では、ネットワークの構造を表す指標であるクラスタ係数と平均ノード間距離と情報拡散がどのように関係しているかを分析する。また、情報拡散の期待影響度との関係を実験で評価する。

2. 基本的な情報拡散モデル

本論文では、情報拡散の基本モデルを有向リンクのネットワーク $G=(V,E)$ 上で適用して分析する。基本モデルでは、“アクティブ(情報を持っている)”状態から“非アクティブ(情報を持っていない)”状態には状態遷移することはないとする。リンクを通じた情報拡散の試行が成功しようと失敗しようと、非アクティブなノードをアクティブにできるのは一度きりである。これらの前提のもとにICモデルとLTモデルにより分析する。

3. 分析手法

分析手法は、次の2つのステップから構成されている。

第1ステップは、WattsとStrogatzのsmall-world実験[3]と同様に、分析の対象となるオリジナルのネットワークのリンクを、張替確率に従いランダム張替ネットワークを作成する。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個生成した。これとオリジナルのネットワーク($P=0$)を含め12個のネットワークについて分析する。張替を行うに際して、2つのことに留意する必要がある。オリジナルネットワークから個々のノードの次数を変更せずに張り変えなければならない。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。具体的な作成方法は以下の通りである。オリジナルのネットワークからリンクリスト $L_E=(e_1, \dots, e_{|E|})$ を準備する。それぞれのリンクは $e=(u,v)$ のような from-part と to-part の順序対である。これから2つのノードリスト L_{From} と L_{To} を作る。先ほどの事項に留意し、張替確率 P に従い、 L_{To} をランダムに並び替える。そして L_{From} と L_{To} を結合し新たなリンクリストを作り、これに基づいて張替ネットワークを作成する。こうして生成したすべての張替ネットワークについて、クラスタ係数

$$C = \frac{1}{|V|} \sum_{u \in V} \frac{| \{ (v \in A_G(u), w \in A_G(u)) : v \neq w, w \in A_G(v) \} |}{|A_G(u)|(|A_G(u)| - 1)}$$

と平均ノード間距離

$$L = \frac{1}{|V|(|V| - 1)} \sum_{u \neq v} l(u,v)$$

を算出しネットワークの構造の違いについて分析する。ここで、 $A_G(u)$ はノード u の隣接ノード集合を、 $l(u,v)$ はノード u と v 間の最短パス長を表す。

第2ステップは、情報拡散モデルに基づいて、上で生成した張替ネットワークに対してシミュレーションを行う。情報拡散のシミュレーションの結果から、期待影響度を算出し、上記のクラスタ係数、平均ノード間距離との関係性を相関係数、回帰分析により調査する。

4. 評価実験

4.1 実験データと第1ステップの結果

1つ目は、あるプログラマーから他のプログラマーへと情報が拡散しうると考えられるブログのトラックバックネットワークのデータである。このネットワークをブログネットワークと呼ぶ。ノード数は、12,047であり、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である。このネットワークをウィキペディアネットワークと呼ぶ。ノード数は9,481であ

† 静岡県立大学経営情報学部 University of Shizuoka

‡ 大阪大学 Osaka University § 龍谷大学 Ryukoku University

り、リンク数は 245,044 である。これらのネットワークで、次数分布がべき則分布に従うなど多くの大規模なネットワークと同じような特徴を有す。詳細は文献[2]参照。

これら 2つのネットワークデータに対して、2つの指標クラスタ係数 C と平均ノード間距離 L の張替ネットワークごとの変化を調べた。ブログネットワークでのクラスタ係数と平均ノード間距離の変化はどちらも張替確率を大きくすると減少する、すなわち右肩下がりになることが図1より見てとれる。しかしその減少の仕方には違いがあり、クラスタ係数(■)は $P=2^{-1}$ 付近で減少率が大きくなり、急激にその値を小さくしている。一方平均ノード間距離(●)は、最初から減少率がほぼ一定で緩やかに減少している。

ウィキペディアネットワークでの結果もほぼ同様のものとなった。

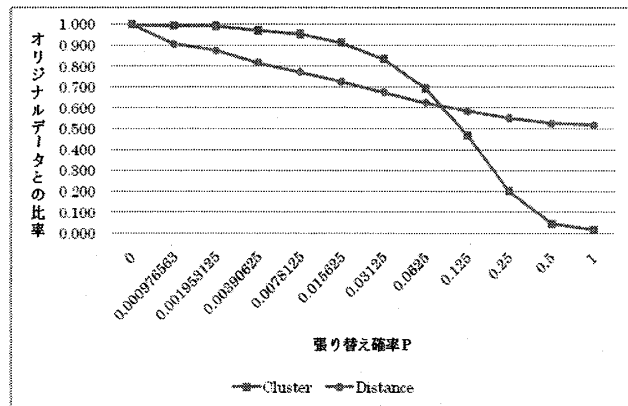


図1 ブログネットワークの C,L の推移

4.2 シミュレーション設定と第2ステップの結果

4.1 節で求めたネットワークの張替による特性と情報拡散の期待影響度との関係を分析するために、情報拡散シミュレーションを行った。使用した 2つの確率モデル IC モデルと LT モデルの設定内容を説明する。まず IC モデルでは、全ての有向リンク (u,v) に拡散確率 $\beta_{u,v}$ の付与が必要である。平均次数の逆数である $\beta_{u,v} = \beta = 1/|V|$ を用いた。次に LT モデルでは、ノード v の親ノード $u \in \Gamma(v)$ から受けるリンクに重みの付与が必要である。重みは次のように定義した: $\omega_{u,v} = 1/|\Gamma(v)|$ 。以上のように定義されたパラメータを用いて実験を行った。

これらのモデルを 2つのネットワークデータの 12 個の張替ネットワークに対して適用した。その情報拡散の結果得られる期待影響度を張替確率ごとにプロットした。図2、図3はブログネットワークの結果を示す。

ブログネットワークのデータに関して、図1、2、3より、クラスタ係数は、LT モデルの期待影響度の変化の仕方の逆の動きをしていることが見てとれる。一方平均ノード間距離は、IC モデルの期待影響度の変化の仕方の逆の動きをしていることがわかる。このことを定量的には評価するために、相関係数を求めた。LT モデルによる期待影響度の変化とクラスタ係数の変化は非常に高い負の相関係数($r=-0.9966$)が見られた。一方の IC モデルによる期待影響度の変化と平均ノード間距離にも非常に高い負の相

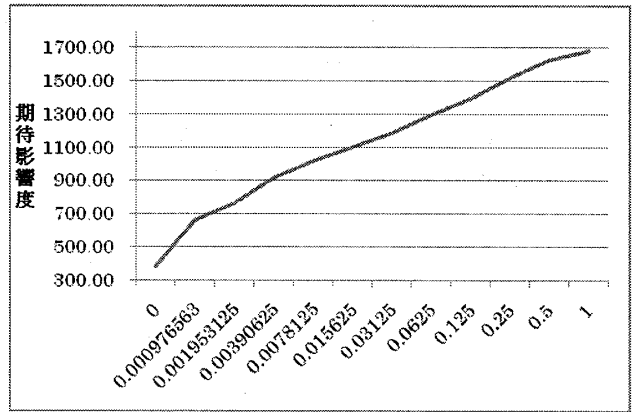


図2 ブログネットワーク期待影響度の推移(ICモデル)

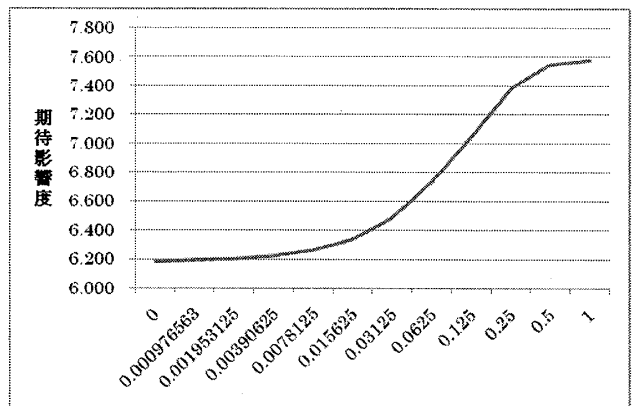


図3 ブログネットワーク期待影響度の推移(LTモデル)

関関係($r=-0.9946$)が見られた。ウィキペディアネットワークに対しての分析結果でもほぼ同様の結果を得ることができた。さらに重回帰分析を行った結果でも同様の結果を得た。

5. おわりに

本論文では、構造の異なるネットワーク間での情報拡散による期待影響度の変化について分析した。結果として、IC モデルでは平均ノード間距離の値が小さくなることと期待影響度が大きくなること、また LT モデルではクラスタ係数の値が小さくなることと期待影響度が大きくなることに強い相関があるという実験結果を得た。

参考文献

- [1] Kempe, D., Kleinberg, J. and Tardos, E., "Maximizing the spread of influence through a social network", *proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2003)*, pp.13-146, (2003).
- [2] Fushimi, T., Kawazoe, T., Saito, K., Kimura, M. and Motoda, H., "What Does an Information Diffusion Model Tell about Social Network Structure?", *Proceedings of the 2008 Pacific Rim Knowledge Acquisition Network Workshop (PKAW2008)*, pp.288-299, (2008).
- [3] Watts, D.J. and Strogatz, S.H., "Collective dynamics of 'small-world' networks", *Nature* 393, pp.440-442, (1998).