

F-013

SIS モデルの影響度とコミュニティ構造の関係分析

Relationship Analysis of Influence Level of SIS Model and Community Structure

小出明弘†

水本 嗣留†

斉藤和巳†

木村昌弘‡

元田浩¶

Akihiro Koide

Tsuguru Mizumoto

Kazumi Saito

Masahiro Kimura

Hiroshi Motoda

1. はじめに

近年、ソーシャルネットワークを通じた情報拡散現象に対する関心が急速に高まっている。一般に、ソーシャルネットワーク上の情報拡散過程は、ノードが非アクティブ(情報が伝わっていない状態)からアクティブ(情報が伝わった)に変化する過程としてモデル化され、IC モデル・LT モデルを用いた影響最大化問題や汚染最小化問題等の研究がおこなわれている[1, 2].

さて、上記研究の情報拡散モデルとして用いた IC モデルと LT モデルでは、ノードが一度アクティブになると、再び非アクティブには戻らないとした。これは、疫学で知られた SIR(Susceptible/Infective/Recovered)モデルである。ここで S は感染可能状態、I は感染状態、R は回復状態を意味する。しかし、現実のソーシャルネットワークの情報拡散現象には、非アクティブに戻って再度アクティブになる SIS モデルとして考えるべき例が少なくないと考えられる[3]。SIS モデルでは、S 状態にあるブログユーザーが、記事を読んで触発され、記事を投稿するまでの I 状態に移行し、その後同じトピックの記事を読む S 状態に戻り、また投稿するという過程が繰り返される。こうした SIS モデルにおける情報拡散過程は、SIR のものと大きく異なるものである。本論文では、SIS モデルの時間とともに移り変わる影響度とネットワークのコミュニティ構造との関係を分析する。

2. 分析データ

ブログのトラックバックネットワークのデータで分析を行う。ブログすなわちWeblogのトラックバック機能とは、あるブログの内容が自身のブログの内容と関連性が高い時、またはあるブログの内容を参照・引用した場合などに、そのブログにトラックバックというリンクを張るが、その際にリンクを張った相手に通知する機能のことである。このリンク関係によるネットワークを利用し、あるブログ著者から別のブログ著者へと情報が伝播しうると考えられるので、ブログのトラックバックネットワークを用いて実験を行った。本論文ではこのネットワークを、ブログネットワークと呼ぶことにする。このデータは、(<http://blog.goo.ne.jp/usertheme/>)の「JR 福知山線脱線事故」というテーマからトラックバックを10段階辿ることにより、2005年5月に収集したものである。このネットワークは、12,047 ノードと79,920 リンクをもつ有向ネットワークである。多くの大規模なネットワークと同じように、自分に向かうリンク次数の分布(入次数分布)も自分から向かうリンク次数の分布(出次数分布)もべき則分布に従うという特徴を有す。

† 静岡県立大学経営情報学部経営情報学科

‡ 龍谷大学

¶ 大阪大学

3. SIS モデルを用いた影響度分析

3.1 分析方法

本分析では、期間 $t=0$ でネットワーク上の各ノード v を順番に単一情報源に設定し、期間 $t=1$ から 100 まで変化させて、各ノード v と各期間 t の影響度 $\sigma(v, t)$ を求めた。ここで影響度とは、アクティブとなったノード数の期待値で、10,000 回のシミュレーションで求めた。

一方、期間 t ごとに $\sigma(v, t)$ が大きい上位 3 位までのノードの推移をプロットし、よく用いられるヒューリスティック指標である次数中心性、近接中心性、媒介中心性の 3 つの基準を用いて推定したものとの関係を分析する。具体的には、まず、各基準でのランクに基づき、それぞれノード番号を 1 から順に振り直す。そして、横軸を期間とし、縦軸を各基準でランク付けしたノード番号として、影響度 $\sigma(v, t)$ において上位 3 位に入ったノードの推移をプロットして分析する。例えば、次数中心性を基準とし、次数最大のノードが全期間内で影響度においても最大となれば、縦軸を y としたとき、 $y=1$ の直線がプロットされる。この際、各リンクを通して情報が伝わる確率 p はすべてのリンクで同じとし、 $p=0.2, 0.1, 0.75, 0.05, 0.025$, および 0.01 とした。

3.2 分析結果

分析結果を図 1(a), (b) に示す。ここで 1 位からそれぞれ実線、破線、点線で表している。図より以下の三つの特徴がわかる。特徴(1)確率が 0.2 程度まで大きくなると、早い段階から順位変動が起こらなくなる。特徴(2) $1 \leq t \leq 5$ の期間を除き、段状に結果が推移している。特徴(3)早い期間の間に上位を取ったノードが、後半で再び上位を取り、その後は変化があまり起こらない。

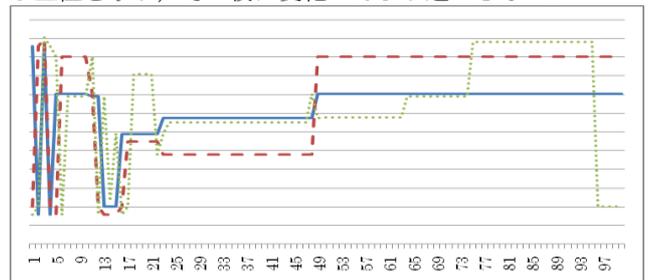


図 1(a) 確率 0.1, 次数中心性と影響度との関係分析

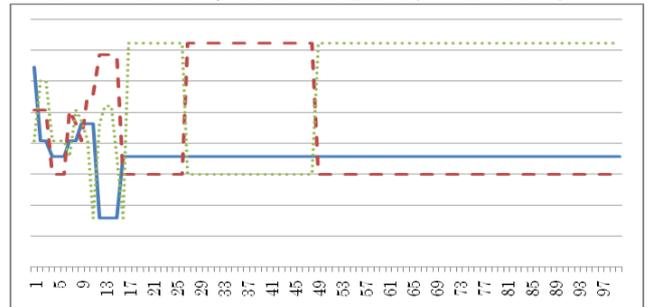


図 1(b) 確率 0.2, 媒介中心性と影響度との関係分析

また、今回3つの基準でプロットし、分析をしたが、どの基準をとってもプロットした結果に大きな違いが見られなかった。分析結果の例は二つのみ示している。

3.3 考察

特徴(1)については、確率が高すぎるとある期間まで情報伝播が行われるとすべてのノードに情報がいきわたってしまい、その結果、順位変動に全く変化がなくなったと考えられる。

特徴(2), (3)の結果からは、同じようなノードが上位を取り合っており、これらのノードが今回使用したブログネットワークにおいて、コミュニティの中心的位置にあるノードであることが推測できる。

この考察から「今回利用したブログネットワーク内のコミュニティ構造において、各々のコミュニティの中の影響度の高いノードが、 t の期間の中で上位を取り合っている。」と仮説をたて、可視化によるコミュニティ構造の分析を行った。

4. コミュニティ構造の可視化分析

上述した仮説を検証するため、ブログネットワークを可視化し分析する。

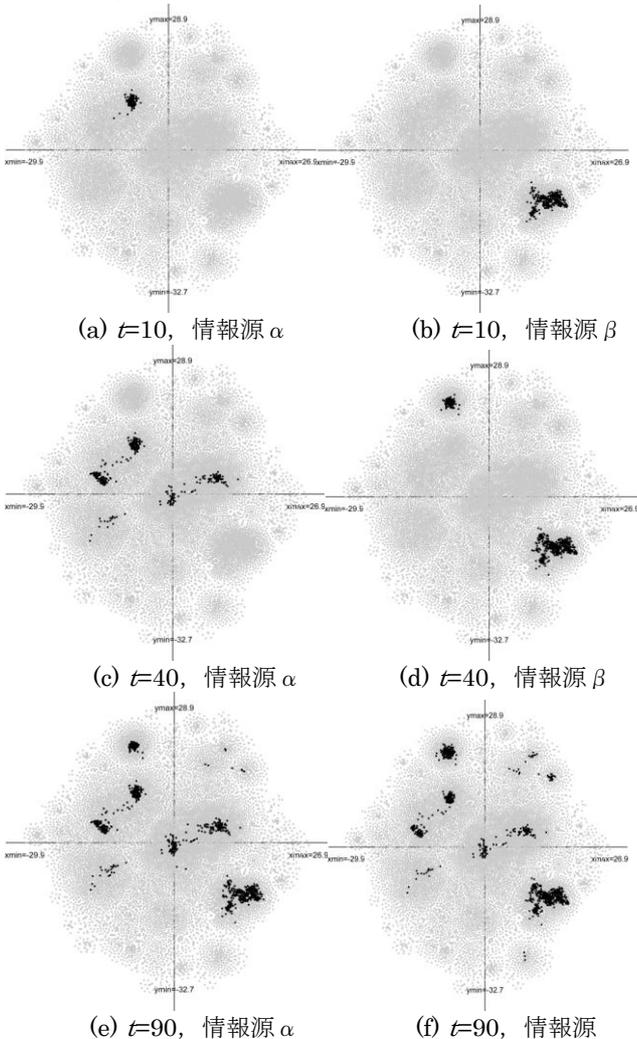


図2 情報源 α , β , $t=10, 40, 90$ における可視化結果

4.1 分析方法

SIS モデルによる影響度分析により、一定の期間内で上位を取り続けるノードを二つ抽出した。この二つのノードとは、図1(a)において $20 \leq t \leq 40$ で1位を取ったノードと $5 \leq t \leq 10$ と $t \geq 45$ で上位を取り続けたノードのことで、便宜上このノードの名前をそれぞれ α , β とする。 α , β を情報源とし、 $t=10, 40, 90$ と設定したとき、各々の期間で情報がどのように伝わっていくのかを可視化した。

ノードからノードへ情報が伝わる確率を $p=0.1$ とした。また、試行回数を10000回とし、50%以上の確率で情報が伝わった場合にはノードを黒く表示し、そうでなければ灰色で表示した。

4.2 分析結果

分析結果を図(a)~(f)に示す。(a), (b)は $t=10$ に設定したときの可視化結果である。まだ期間が短く、それほど情報伝播が起っていないため、コミュニティが大きいと予想される(b)の方が確実に情報を受け取っているノードが多い。

(c), (d)は $t=40$ に設定したときの可視化結果である。(c)では、複数のコミュニティを巻き込みながら情報が伝播していることがわかる。一方、(d)では情報源周辺のコミュニティ自体は大きいですが、周辺のコミュニティと距離があると予想され、なかなか情報が伝わらず、この二つの情報源ノード α と β 間で影響度の逆転が起ったと考えられる。

(e), (f)は $t=90$ と設定したときの可視化結果である。 $t=100$ に近づいていくとほとんどのコミュニティの中心部に情報が伝わっていることが分る。それにより、情報が新たに他のコミュニティに伝わるものがなくなるので、より大きなコミュニティの中心にあるノードの方が影響度が大きくなり、再び影響度に逆転が起きたと考えられる。

これらの結果から、仮説「今回利用したブログネットワーク内のコミュニティ構造において、各々のコミュニティの中の影響度の高いノードが、 t の期間の中で上位を取り合っている。」の状況が起っていることが示唆される。

5. おわりに

本論文では、SIS モデルの時間とともに移り変わる影響度とネットワークのコミュニティ構造との関係を分析した。SIS モデルによる影響度分析とコミュニティ構造の可視化結果より、これらの関係についての幾つかの知見が得られた。今後は、ブログネットワーク以外でも分析を行いどのような関係があるのかを研究していきたい。

参考文献

- [1] Kempe, D., Kleinberg, J. and Tardos, E., "Maximizing the spread of influence through a social network", *Proc. KDD'03*, pp. 137-146, (2003).
- [2] Albert, R., Jeong, H. and Barabashi, A.-L., "Error and attack tolerance of complex networks", *Nature*, Vol. 406, pp. 378-382, (2000).
- [3] Leskovec, J., McGlohon, M. and Faloutsos, C., "Cascading behavior in large blog graphs", *Proc. SDM'07*, pp. 551-556, (2007).