

有向ネットワークの構造が情報拡散に与える影響の分析

Influence analysis of information diffusion focusing on directed networks

臼井 翔平*2
Shohei Usui鳥海 不二夫*1
Fujio TORIUMI平山 高嗣*2
Takatsugu HIRAYAMA間瀬 健二*2
Kenji MASE*1 東京大学大学院工学系研究科
School of Engineering The University of Tokyo*2 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University

Online social networks(OSNs) are attractive for their information sharing capabilities. In this paper, we analyze information diffusion focusing on directed network structures. We focus on 13 features and analyze which feature is effective for information diffusion. We found that the reachability, node assortativity, in-degree assortativity, and in-out-degree assortativity are effective features. In addition, we produced network with high information-diffusion performance. Finally We analyzed the change in the real network and we showed that the changed network have high information-diffusion performance.

1. はじめに

現在、ソーシャルメディア上での情報拡散が注目を集めており、多くのユーザがソーシャルメディアによって情報を収集・発信を行っている。この情報収集・発信は、個人的な趣味趣向に関する情報だけに留まらず、政治や経済といった公の情報にまで至っている。これに伴って日本では、2013年4月に改正公選挙法によって、インターネットを用いた選挙活動の解禁を定めた。今後このような動きは拡大される事は明らかである。

2011年3月に発生した東日本大震災の際はTwitterが情報拡散に貢献したとされている。この時、関連ネットワークの構造が変化している[1]。しかし、有向ネットワークについてどのようなネットワーク構造が情報拡散に影響しているかは明らかになっていない。そこで本研究では、有向ネットワークに関してどのような構造特徴が情報拡散に影響を与えるかを明らかにすることを目的とする。これにより、情報拡散しやすい構造を明らかにする事で、広く情報拡散することに貢献できると考える。

本論文ではまず、ネットワークから複数の構造特徴を抽出し、それぞれの構造特徴が情報拡散にどのような影響を与えるかを分析する。その後、無向ネットワークとの比較を行い、最後に実際のTwitterの関連ネットワークに対してどのような変化が起きているかを分析する。

2. ネットワーク指標

本研究では、ネットワーク構造の差異を定量的に分析するために以下に示す13個の指標を用いて、ネットワークから特徴を抽出する。

2.1 相互性 ρ [2]

相互性は、全体のリンクの中での相互リンクの割合を示している。相互性 ρ は、相互リンクの数 M_{rec} を用いて以下の式(1)によって定義される。

$$\rho = \frac{M_{rec}}{M} \quad (1)$$

連絡先: 名古屋大学大学院情報科学研究科
〒464-8603 愛知県名古屋市千種区不老町
E-mail: usui@cmc.ss.is.nagoya-u.ac.jp

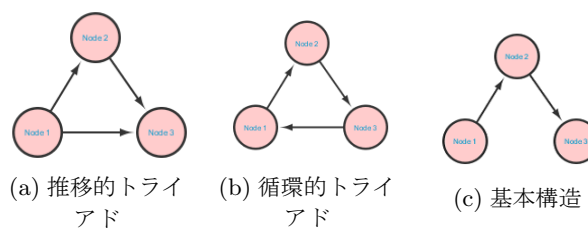


図 1: Triad structures.

なお、 M は総リンク数である。

2.2 到達可能率 α [3]

ノード i からリンクをたどってノード j まで到達可能な時、 $a_{ij} = 1$ とし、それ以外を $a_{ij} = 0$ とした行列 T を到達可能性行列と呼ぶ。この時、到達可能率は α は総ノード数 N を用いて以下の式(2)のように定義される。

$$\alpha = \frac{\sum_{i \neq j} a_{ij}}{N(N-1)} \quad (2)$$

2.3 推移性 τ , 循環性 c [2]

有向ネットワークでは、相互リンクを考えない場合、トライアドの構造は図1(a)の推移的トライアド構造と図1(b)の循環的トライアド構造の2種類のみとなる。この時、図1(c)の構造を基準とし、推移性は推移的トライアド構造の割合を、循環性は循環的トライアド構造の割合を示す。

2.4 出次数相関 r_{oo} , 入次数相関 r_{ii} , 出入次数相関 r_{oi} , 入出次数相関 r_{io} , ノード内次数相関 r_n [4]

有向ネットワークにおけるノードは、出次数と入次数の2つを持っている。この時、方向のあるリンクで結ばれたノード間の次数相関は以下の4つである。

- リンク元の出次数とリンク先の出次数の相関
- リンク元の入次数とリンク先の入次数の相関
- リンク元の出次数とリンク先の入次数の相関
- リンク元の入次数とリンク先の出次数の相関

これをそれぞれ、出次数相関 r_{oo} , 入次数相関 r_{ii} , 出入次数相関 r_{oi} , 入出次数相関 r_{io} とする. 次数相関は以下の式によって定義する [4]. その際, M はネットワーク上の全リンク数を表し, k_m はリンク元の次数を, l_m はリンク先の次数を表す. この時, 次数にはそれぞれ上記の条件に合わせて入次数, または出次数を用いる.

$$r = \frac{M^{-1} \sum_m k_m l_m - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2}{\frac{1}{2} M^{-1} \sum_m ((k_m)^2 + (l_m)^2) - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2} \quad (3)$$

これに加えて, ノード内の入次数と出次数の相関をノード内次数相関 r_n とする. ノード内次数相関は入次数 k_n^{in} , 出次数 k_n^{out} として, 同様に定義する.

$$r_n = \frac{N^{-1} \sum_n k_n^{in} k_n^{out} - [N^{-1} \sum_n \frac{1}{2}(k_n^{in} + k_n^{out})]^2}{\frac{1}{2} N^{-1} \sum_n ((k_n^{in})^2 + (k_n^{out})^2) - [N^{-1} \sum_n \frac{1}{2}(k_n^{in} + k_n^{out})]^2} \quad (4)$$

ここで, N は全ノード数を表し, k_n^{in} , k_n^{out} はノード n の入次数及び出次数を表す.

2.5 出次数分布ベキ指数 γ_o , 入次数分布ベキ指数 γ_i [5]

ベキ指数はノードの次数分布をベキ分布とした時のベキ指数である. ノードの次数分布は, 次数 k を持つノードの存在確率分布 $p(k)$ として定義される. すなわち, $p(k)$ はネットワークからランダムにノードを選択した場合に, そのノードの次数が k である確率を表す. 現実世界に存在する多くのネットワークでは, $p(k)$ が k のベキ乗に比例し,

$$p(k) \propto k^{-\gamma} \quad (5)$$

を満たす例が多く確認されている. 本研究では, 出次数分布のベキ指数 γ_o 及び入次数のベキ指数 γ_i をそれぞれ用いる.

2.6 出次数分布ベキ分布決定係数 R_o^2 , 入次数分布ベキ分布決定係数 R_i^2 [5]

ベキ分布決定係数 R は, 次数分布がどれだけベキ分布に従っているかを表現する値である. ここで, $p(k)$ の対数をとった値を y_i , y_i の平均を \bar{y} , ベキ指数に従った予測値を f_i としたときベキ分布決定係数は以下の式で求まる.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (6)$$

本研究では, 出次数分布のベキ分布決定係数 R_o^2 及び入次数のベキ分布決定係数 R_i^2 をそれぞれ用いる.

3. 情報拡散シミュレーション

本研究では, 第 2 章で挙げた指標が変化することで, 情報拡散がどのように変化するかを分析する. そのために, 様々なネットワークを用意し, それらの上で情報拡散モデルを用いて情報拡散シミュレーションを行う.

3.1 分析手法

本章では, 各指標を対象とし, 対象とした指標のみが異なるネットワークを用いて情報拡散シミュレーションを行う. これらのネットワーク上での情報拡散の度合いを見る事によって, 指標による情報拡散への影響を見ることができると考える.

以下ではまず, 対象とした指標のみが異なるネットワークについて説明し, その後情報拡散モデルについて解説する. さらに分析に際して用いる, ネットワークの情報拡散能力を表す指標を定義する.

3.2 シミュレーション設定

本研究では, 各指標を対象とし, 対象とした指標のみが異なるネットワークを用いる. ネットワークは, 一つの指標値だけを少しずつ変化させ, その他の指標値は全て固定したネットワークを用いる. 固定する指標値には実ネットワークから抽出したものを扱い, 各指標値を表 1 に示す.

情報拡散シミュレーションは情報拡散モデルを用いて行う. 本研究では, 既存の IC モデルを用いる [6]. この時, ネットワーク上のあるノード v_i から情報拡散を開始し, 情報が伝わったノードの割合を $\sigma(v_i)$ とする. さらに, 全てのノードから情報拡散を開始し, $\sigma(v_i)$ の平均をとった値を Ability of information diffusion(AID) σ として以下の式で定義する.

$$\sigma = \frac{\sum \sigma(v_i)}{N} \quad (7)$$

情報拡散シミュレーションを 100 回行い, 平均を算出した.

3.3 シミュレーション結果

3.3.1 指標値が拡散能力に与える影響

全てのネットワークに対して, 変化させている指標と, AID の変化の間の相関係数をそれぞれ表 2 に示す. 相関係数には Spearman の順位相関を用いた.

ここから, ノード内次数相関 r_n に最も強い正の相関が出ていることがわかる. したがって, ノード内次数相関 r_n が高い程, AID の値が高く, 情報拡散に優れたネットワークである事が示された. 次に, 到達可能率 α 及び, 出入次数相関 r_{io} に強い正の相関が出ている. したがって, 到達可能率 α 及び入出次数相関 r_{io} が高い程, AID の値が高く, 情報拡散に優れたネットワークであると言える. さらに, 入次数相関 r_{ii} に強い負の相関が出ている. したがって, 入次数相関 r_{ii} が低い程, AID の値が高く, 情報拡散に優れたネットワークである. また, 他の 9 指標には上記の 4 指標程強い相関は見られなかった. 以下では強い相関が得られた 4 つの指標について見る.

3.3.2 到達可能率 α

図 2 に到達可能率 α が異なるネットワークの到達可能率 α と AID の関係を示した. ここからも, 到達可能率 α が高い程, AID の値が高くなっていることが読み取れる.

到達可能率 α は, あるノード v から情報拡散を開始した際の到達可能なノード数の割合である. したがって到達可能なノードが少なくなれば, $\sigma(v)$ が小さな値となり, その平均である AID の値が小さくなるのは明らかである. すなわち, より情報拡散能力の高いネットワークにするためには, 高い到達可能率 α が必要である.

3.3.3 入次数相関 r_{ii}

図 3 に入次数相関 r_{ii} が異なるネットワークの入次数相関 r_{ii} と AID の関係を示した. ここからも, 入次数相関 r_{ii} が低い程, AID の値が高くなっていることが読み取れる.

入次数相関 r_{ii} はリンクで結ばれたノード間の入次数の相関である. すなわち, リンクで結ばれたノード同士の入次数が離れている程, 情報拡散能力が高いネットワークである事を示唆している. 入次数が高いノードは情報を取得する確率が高い.

表 1: 固定指標値

| Node | Link | ρ | α | τ | c | r_{oo} | r_{ii} |
|----------|----------|--------|----------|---------|------------|------------|----------|
| 175 | 1739 | 0.384 | 0.785 | 0.228 | 0.157 | 0.509 | 0.335 |
| r_{oi} | r_{io} | r_n | R_o^2 | R_i^2 | γ_o | γ_i | |
| 0.389 | 0.490 | 0.329 | 0.784 | 0.684 | 0.897 | 0.871 | |

表 2: 各指標値と AID の相関

| | | | | | | |
|----------|----------|---------|---------|------------|------------|----------|
| ρ | α | τ | c | r_{oo} | r_{ii} | r_{oi} |
| -0.0534 | 0.766 | -0.0456 | 0.0556 | -0.327 | -0.636 | -0.357 |
| r_{io} | r_n | R_o^2 | R_i^2 | γ_o | γ_i | |
| 0.756 | 0.966 | 0.0932 | 0.128 | 0.0593 | 0.0449 | |

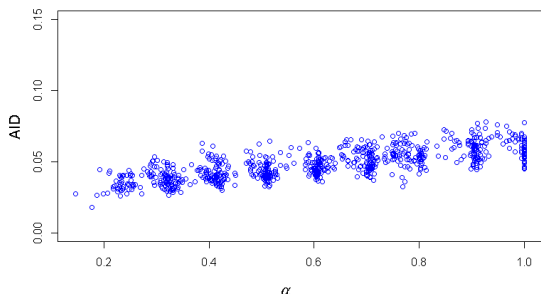


図 2: AID と到達可能率 α の関係

入次数が高いノードからリンクされているノードは、このようなノードから情報を受け取る事ができるので、入次数が低くても高確率で情報を取得することが可能である。従って、入次数が高いノードからリンクされているノードは入次数が高い必要がないため、入次数相関 r_{ii} が低い方が情報拡散に適していると考えられる。

3.3.4 入出次数相関 r_{io}

図 4 に入次数相関 r_{io} が異なるネットワークの入次数相関 r_{io} と AID の関係を示した。ここからも、入次数相関 r_{io} が低い程、AID の値が高くなっていることが読み取れる。

入出次数相関 r_{io} はリンク元ノードの入次数とリンク先ノードの出次数の相関を取った値である。すなわち、リンク元ノードの入次数とリンク先ノードの出次数が同程度である程、情報拡散能力が高いネットワークである事を示唆している。上記と同様に入次数が高いノードは情報を取得する確率が高く、そのノードからリンクされているノードは同様に情報を取得する確率が高く、出次数が高ければ拡散は行われやすくなる。したがって、リンク元ノードの入次数が高く、リンク先の出次数が高い時に情報拡散が広く行われると考えられる。つまり、入出次数相関 r_{io} が高い程、情報拡散に適していると考えられる。

3.3.5 ノード内次数相関 r_n

図 5 にノード内次数相関 r_n が異なるネットワークのノード

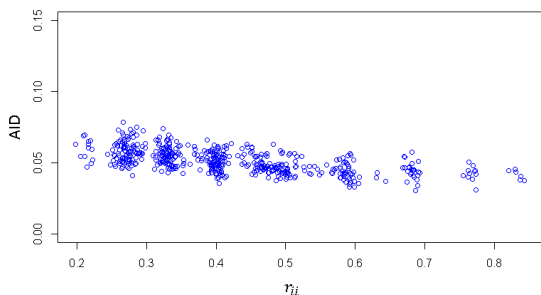


図 3: AID と入次数相関 r_{ii} の関係

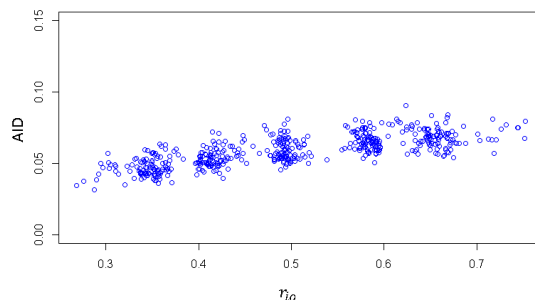


図 4: AID と入次数相関 r_{io} の関係

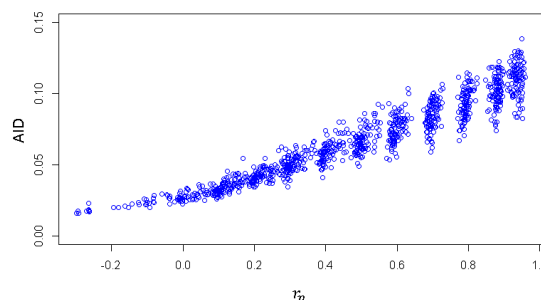


図 5: AID とノード内次数相関 r_n の関係

内次数相関 r_n と AID の関係を示した。ここからも、ノード内次数相関 r_n が高い程、AID の値が高くなっていることが読み取れる。

ノード内次数相関 r_n はノードの出次数と入次数の相関である。つまり、ノードの持つ出次数と入次数が同程度であれば、ノード内次数相関 r_n は正の大きな値となるため、出次数と入次数の数が同程度であるネットワークの方が情報拡散能力は高いと考えられる。ノード内次数相関が低い時、出次数もしくは入次数の一方がもう一方に対して非常に高いことが想定される。ノード内次数相関が低いネットワークでは出次数が高く入次数が低いノード、または出次数が低く入次数が高いノードが多い。それに対して、ノード内次数相関の高いネットワークでは出次数と入次数が共に高いノードが存在している。出次数が入次数に対して非常に高いノードの場合、出次数の高いノード、つまり情報発信能力の高いノードの入次数が低いため、情報拡散が行われにくいと考えられる。逆に入次数が出次数に対して非常に高いノードの場合、入次数の高いノード、つまり情報収集能力の高いノードの出次数が低いため、情報拡散が行われにくいと考えられる。したがって、出次数と入次数が共に高いノードが多い程、つまり情報発信能力と情報収集能力が共に高いノードが多く存在するようなネットワークが、情報拡散能力が高いネットワークであると考えられる。

3.4 情報拡散に適したネットワーク

ここまでの分析によって、情報拡散における重要な指標が明らかになった。図 2 から図 5 を比較すると、ノード内次数相関 r_n が最も AID を変化させている事がわかる。そこでノード内次数相関が高い事を優先的に考える。ここで、ノード内次数相関が高い値をとっている時、ノードの出次数と入次数はほぼ同値になるため、低い入次数相関 r_{ii} と高い入出次数相関

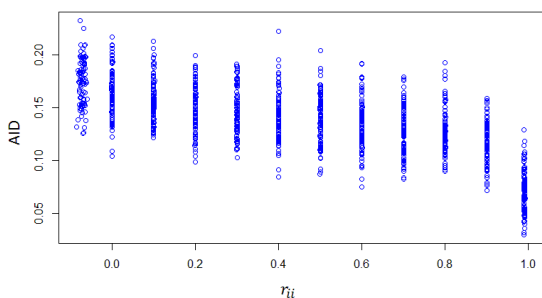


図 6: 情報拡散に適したネットワークにおける入次数相関 r_{ii} と AID の関係

r_{io} は両立しない。そこで、高い到達可能率 α と低いノード内次数相関 r_n を満たした状態で、入次数相関 r_{ii} の値を変化させる。図 6 に入次数相関 r_{ii} と AID の値を示す。尚、ノード内次数相関 $r_n \approx 1$ であるため、 $r_{ii} \approx r_{io} \approx r_{oi} \approx r_{oo}$ が成り立つ。これより、ノード内次数相関が高い時、入次数相関は低い情報拡散性能に優れている事がわかる。また、入次数相関 $r_{ii} = -0.0769$ の時に AID は最大値 0.232 をとっており、前節までの実験と比較しても最大値であることがわかる。この時の各指標値を表 3 に示す。これより、高い到達可能率 α 及び高いノード内次数相関 r_n が実現できている事がわかる。

3.5 無向ネットワークとの比較

無向ネットワークでは平均経路長が短く同類選択性が高いネットワークが情報拡散性能に優れている事がわかっている [7]。有向ネットワークにおいては各ノードから各ノードへ必ずしもたどり着けないため、平均経路長は意味をなさないと言われている。さらに無向ネットワークでは、無向ネットワークにおける同類選択性 r は $r = r_{oo} = r_{ii} = r_{oi} = r_{io}$ であり、入次数相関 r_{io} の性質が強くなるために同類選択性が高いネットワークが情報拡散性能に優れている事が示唆される。また、無向ネットワークでは、相互性は $\rho = 1$ であり、到達可能率やノード内次数相関も $\alpha = r_n = 1$ であり、到達可能率とノード内次数相関が共に高い値であるため、情報拡散性能に優れていると言える。

3.6 震災によるネットワークの変化

2011 年 3 月 11 日に発生した東日本大震災の際に Twitter の関連ネットワークの構造が大きく変化しているという報告がある [1]。本研究では、情報はコミュニケーションをとりあうような関係の上を流れると考える。したがって Twitter では、この関連ネットワークの上を情報が流れるとみなす。そこで、1 日毎に関連ネットワークを生成し、震災前の 2011 年 3 月 10 日及び震災後の 2011 年 3 月 12 日の情報拡散に関する指標値を表 4 及び AID を示す。震災後のネットワークは震災前のネットワークと比較して AID が高い事が窺える。震災後ではノード数に比べてリンク数が非常に増大している。すなわち、リンク密度が 1.59 倍になっていることがわかる。リンク

表 3: AID の高いネットワークの各指標値

| Node | Link | ρ | α | τ | c | r_{oo} | r_{ii} |
|----------|----------|--------|----------|---------|------------|------------|----------|
| 175 | 1739 | 0.513 | 0.994 | 0.0816 | 0.0897 | -0.0510 | -0.0769 |
| r_{oi} | r_{io} | r_n | R_o^2 | R_i^2 | γ_o | γ_i | |
| -0.0605 | -0.0683 | 0.980 | 0.537 | -0.703 | 0.462 | -0.663 | |

表 4: 震災前後の Twitter 関連ネットワークの各指標

| | Node | Link | α | r_{ii} | r_{io} | r_n | AID |
|-----------|---------|----------|----------|----------|----------|--------|---------|
| 2011/3/10 | 1520524 | 5308675 | 0.370 | 0.0998 | 0.00373 | 0.273 | 0.00298 |
| 2011/3/12 | 1734187 | 10954466 | 0.436 | 0.221 | -0.0184 | 0.0106 | 0.00689 |

密度が高まれば当然情報は拡散しやすくなる。また、到達可能率 α も震災後のネットワークの方が高くなっている事がわかる。これらが要因となって、より多くのユーザに情報が届くと考えられる。しかし、ノード内次数相関 r_n は震災後に小さくなり、また入次数相関 r_{ii} も高くなっている。これらは情報拡散に不利にはたらくと考えられる。これらの値を維持することができれば、より情報拡散しやすいネットワークとなる事が考えられる。

4. 結論

本研究では、有向ネットワーク構造に着目し、情報拡散のしやすさを分析した。複数の指標に対し、対象指標値のみを変化させたネットワークを用い、情報拡散シミュレーションを行った。その結果、到達可能率、ノード内次数相関、入次数相関、入出次数相関の 4 つの指標が情報拡散に影響していることを明らかにした。最後に、実際の Twitter 関連ネットワークの震災による変化を分析し、震災後のネットワークがより高い情報拡散能力を有している事を示した。

今後の課題としては、2 つ以上の指標による影響を複合的に分析することを挙げる。

謝辞

本研究は科研費 (24300064) の助成を受けた。

参考文献

- [1] Fujio Toriumi, Takeshi Sakaki, Kosuke Shinoda, Kazuhiro Kazama, Satoshi Kurihara, and Itsuki Noda. Information sharing on twitter during the 2011 catastrophic earthquake. In *Proceedings of the 22nd international conference on World Wide Web companion, WWW '13 Companion*, pp. 1025–1028, 2013.
- [2] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, Vol. 8. Cambridge university press, 1994.
- [3] Gunther Schmidt. *Relational Mathematics*, Vol. 132 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2011.
- [4] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, Vol. 67, No. 2, p. 026126, February 2003.
- [5] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, Vol. 286, No. 5439, pp. 509–512, 1999.
- [6] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.
- [7] Shohei Usui, Fujio Toriumi, Takatsugu Hirayama, and Kenji Mase. Analysis of influential features for information diffusion. In *SocialCom 2013*, Washington, D.C., September 2013.