

ネットワークの崩壊と構造の関係についての分析

Analysis of relation between network structures and collapse

臼井 翔平 *1*2

Shohei Usui

鳥海 不二夫 *1

Fujio TORIUMI

*1 東京大学大学院工学系研究科

School of Engineering The University of Tokyo

*2 日本学術振興会

Japan Society for the Promotion of Science

The ability to spread information on social media is expected to be useful in emergencies like disasters. On the other hand, it is difficult for some users to use social media in the event of a disaster, and the nodes may be missing. We clarify the structure of the network that can robustly spread information against such node missing. In this research, when analyzing this, we adopt a network dataset covering the network structure feature space. As a result of the analysis, as a result of the analysis, it is difficult for the network to be divided into multiple components when a node is missing, it is found that the network has a small mean path length and a poor distribution of the degree distribution. We also clarified that a robust network for information spreading among networks that are difficult to divide into multiple components has a small average path length and a small cluster coefficient.

1. 序論

近年では、ソーシャルメディアの発展に伴い、人間関係ネットワークが目に見える形で具現化され、ネットワーク上での情報の伝達が明確になった。ソーシャルメディア上の多くのユーザは、ソーシャルメディアによって情報収集・発信をしているため、このソーシャルメディア上の情報拡散が注目を集めている。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情報にまで至る。従って、ソーシャルメディア上の情報拡散は、マーケティング、政治的意見形成等、様々な影響を現実の社会現象与えている事は明らかである。

ソーシャルメディアの情報拡散能力は、災害のような非常事態時での活躍が期待されている。2010年の the 2010 Chile earthquake [1] や 2011年の the Great East Japan Earthquake [2] では、Twitter を使った情報拡散が有用であった事が報告されている。

一方で、災害時には一部のユーザがソーシャルメディアを利用する事が困難となる事が想定される。このような場合、ネットワーク上ではノードの欠損とみなす事が出来る。ノードが欠損すると、ネットワークが分断される可能性が考えられる。本研究では、このようなノードの欠損に対して、どのようなネットワークにおいてネットワークの分断が起きるのかを分析する。

2. ネットワークデータセット

2.1 ネットワークデータセットの構築

様々なネットワークを含むデータセットを構築するために、貪欲なネットワーク生成手法を用いる [3]。まず、以下の特徴の範囲内から特徴点を抽出する。

- 平均経路長 L ($2 \leq L \leq 6$) [5],
- クラスタ係数 C ($0 \leq C \leq 0.6$) [5],

- 次数相関 r ($-0.5 \leq r \leq 0.5$) [6],
- ベータ分布のパラメータ $\log \alpha$ ($0 \leq \log \alpha \leq 3$)
- ベータ分布のパラメータ $\log \beta$ ($0 \leq \log \beta \leq 4$)

なお、ベータ分布のパラメータ α 及び β は、ネットワークに対してベータ分布を最適化したパラメータである。

ベータ分布 $q(k)$ は以下の式で表される。

$$q(k) = \frac{x^\alpha (1-x)^\beta}{B(\alpha, \beta)}, \quad (1)$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1}, \quad (2)$$

なお、 α 及び β はフィッティングパラメータである。ここで、次数分布における x は以下の式で表される。

$$x = \frac{k}{k_{max} + 1}. \quad (3)$$

また、ベータ分布を確率分布とする (すなわち、 $\sum_k^{k_{max}} q(k) = 1$) ために、以下の処理を行う。

$$q'(k) = \frac{q(k)}{\sum_{k'}^{k_{max}} q(k')}. \quad (4)$$

ベータ分布は α 及び β を調整する事によって、様々な分布の形状を表現する事が可能である。ネットワークの次数分布に対して α 及び β を、焼きなまし法 (Simulated Annealing: SA) [4] によって最適化する。最適化の際の目的関数として、ネットワークの次数分布 $p(k)$ とベータ分布 $q(k)$ の分布間二乗距離を、この特徴に関する距離として定義する。

次に、抽出された点を基に、貪欲なネットワーク生成手法を用いてネットワークを生成する。この時、生成されたネットワークと抽出された特徴点の距離が $D \geq 0.1$ である時、そのネットワークは破棄するものとする。これを繰り返し行う事によって、ネットワーク構造特徴空間上を網羅するようなネットワークデータセットを構築する事が出来る。

連絡先: 東京大学大学院工学系研究科

〒 113-8654 東京都文京区本郷 7-3-1

E-mail: usui@crimson.q.t.u-tokyo.ac.jp

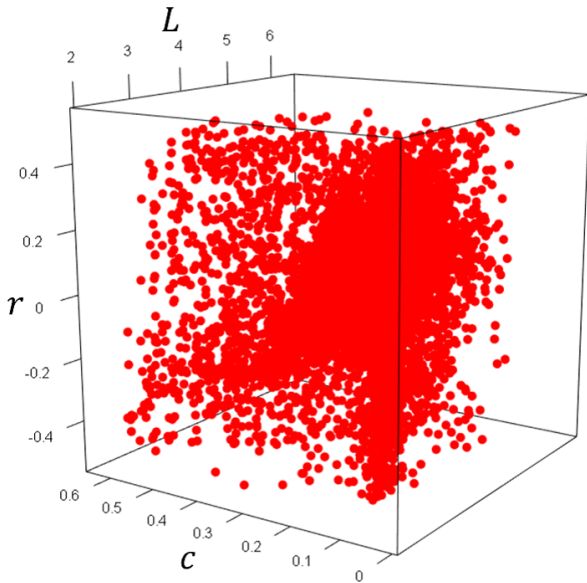


図 1: 提案手法によるネットワーク群

本研究では、ノード数 1000、リンク数 10000 のネットワークを 10000 個生成し、生成したネットワークに関して、全ての特徴量を計算し直した。さらに、類似したネットワーク構造を除外するために、半径 $D_F = 0.1$ の多次元球内にネットワークが高々1つとなるように無作為抽出した。その結果として、7028 個のネットワークを含むデータセット G_n を得た。図 1 に 3 次元特徴空間 (L, C, r) への射影した図を示す。

2.2 ネットワークの頑健さ δ

本研究では、ネットワーク上に存在するノードが 30% 損失した時の、最大コンポーネントの全体における割合 δ を、ネットワークの頑健さとする。本研究において、損失は意図的でないものを対象としているため、損失するノードはランダムに選ばれるものとする。図 2 にデータセット内のノード 30% 損失時の、最大コンポーネントの全体における割合 δ をのヒストグラムを示す。この図より、多くのネットワークの最大コンポーネントの全体における割合 δ はほぼ 1 であるが、いくつかのネットワークでは大きく減少している事がわかる。本分析では、このように大きく減少するネットワークがどのようなネットワークであるかを明らかにする。

3. 頑健性の強いネットワーク構造の分析

本章では、提案手法によるデータセットに含まれる各ネットワークの頑健性を測定し、ネットワークの頑健性とネットワーク構造特徴の間の関係について分析を行う。本研究では、様々な構造のネットワークを用いて統計的に分析を行う事によって、ネットワークの頑健性に起因するネットワーク構造特徴を明確にする。

3.1 頑健性と各ネットワーク特徴の相関関係

ここでは一つ一つの構造特徴と頑健性の関係を見るために、今回採用した 5 つの特徴と最大コンポーネントの全体における割合 δ の相関を分析する。

表 1 に、各特徴と最大コンポーネントの全体における割合 δ との相関係数を示す。表 1 より、平均経路長 L 、クラスタ係数

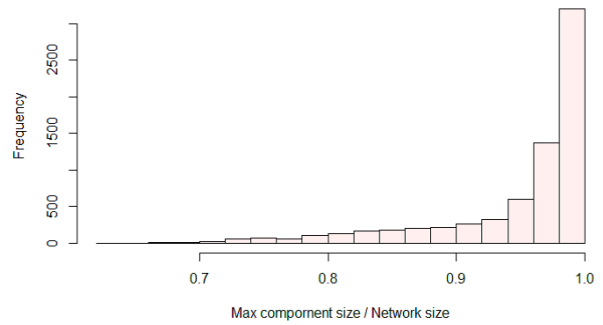


図 2: ノード 30% 欠損時の最大コンポーネントの全体における割合

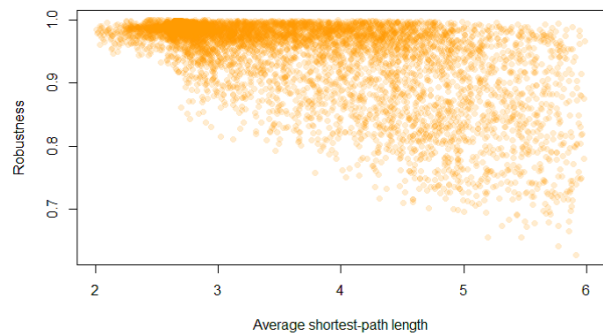


図 3: 平均経路長 L と頑健性の関係

C に負の相関があり、次数分布のパラメータ $\log\alpha$ に正の相関がある事が明確となった。また、次数分布のパラメータ $\log\beta$ と次数相関 r は大きな相関がない事がわかる。以降では、相関のあった平均経路長 L 、クラスタ係数及び次数分布のパラメータ $\log\alpha$ についてそれぞれ見ていく事とする。

まず、平均経路長 L と頑健性の関係について図 3 に示す。平均経路長 L と頑健性は相関係数 -0.613 と負の相関である。図より、平均経路長が短い時、非常に強い頑健性を持つが、平均経路長が長くなると、頑健性が低い可能性がある事が示されている。しかし、平均経路長が長くても高い頑健性を持つネットワークが存在し、平均経路長の他にも要素が存在する事が示唆される。

続いて、クラスタ係数 C と頑健性の関係について図 4 に示す。クラスタ係数 C と頑健性は相関係数 -0.281 と負の相関である。しかし、図からクラスタ係数が高い程、頑健性が低いという事は読み取れない。すなわち、クラスタ係数の値に関わらず、頑健性が高い場合も低い場合も存在する。従って、クラスタ係数の値は大きな影響を与えていない事が示唆された。

表 1: 頑健性と各特長の相関係数

L	C	r	$\log\alpha$	$\log\beta$
-0.613	-0.281	0.111	0.239	-0.0634

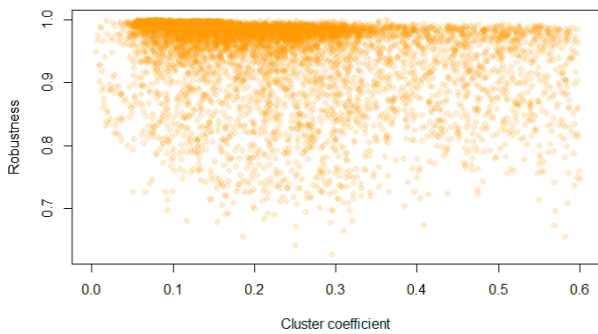


図 4: クラスタ係数と頑健性の関係

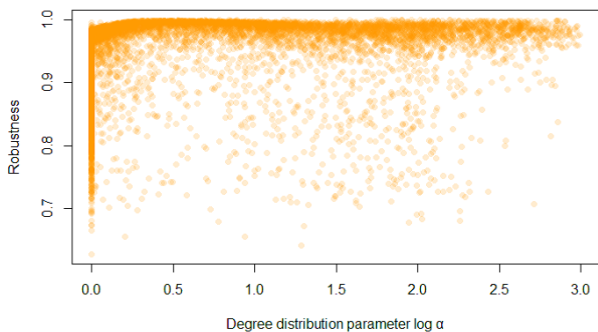


図 5: 度数分布のパラメータ $\log \alpha$ と頑健性の関係

最後に、度数分布のパラメータ $\log \alpha$ と頑健性の関係について図 5 に示す。度数分布のパラメータ $\log \alpha$ と頑健性は相関係数 0.239 と正の相関である。図から、度数分布のパラメータ $\log \alpha$ が $\log \alpha \approx 0$ の時に頑健性が非常に低い値を取っている事が伺える。 $\log \alpha \approx 0$ の時、度数分布はべき分布となる。従って、ネットワークがべき分布を取る時、頑健性が損なわれる事が示されている。

3.2 頑健なネットワークの構造

前節では、一つ一つの特徴に着目した分析を行った。次に、複数の特徴の複合的な影響に着目した分析を行う。ここでは、決定木を用いて頑健なネットワークの条件を分析する。ここでは、最大コンポーネントの大きさが全体の 90% 以上であるネットワークを頑健であるとする。決定木の説明変数としては、前節で用いた 5 特徴を用いる。図 6 に構築された決定木を示す。図における Yes は最大コンポーネントが全体の 90% 以上であるネットワークの割合であり、No はそうではないネットワークである。

本決定木より、平均経路長 $L < 4.27$ であるネットワークの 93.1% は頑健であるという結果となった。従って、最も重要な構造特徴は平均経路長 L である事が伺える。次に、平均経路長 $L > 4.27$ であるネットワークの内、度数分布のパラメータ $\log \alpha \approx 0$ であるネットワークの 77.7% は頑健なネットワークである。一方で、平均経路長 $L > 4.27$ であるネットワークの内、度数分布のパラメータ $\log \alpha > 0$ であるネットワークの

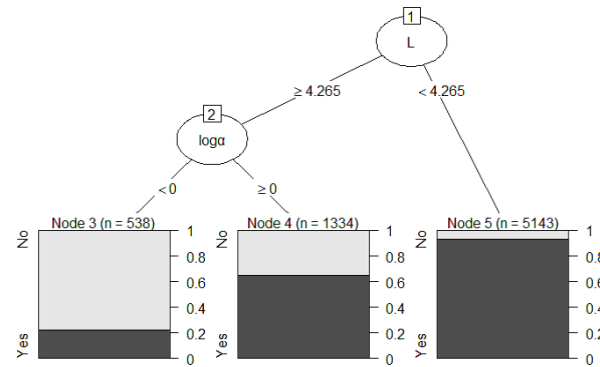


図 6: 頑健なネットワークに関する決定木 (Yes:頑健なネットワーク, No:頑健でないネットワーク)

64.8% は頑健ではない。従って、平均経路長 L が長く、かつ、べき分布を持つネットワークが頑健ではないと考えられる。この結果は前節の結果と同様である。

4. ノードが欠損したネットワーク上での情報拡散

本節では、前節で頑健である事を示したネットワークのノードが欠損した時の情報拡散について分析する。ノードが 30% 欠損した際の、最大コンポーネントが全体の 90% を超えるネットワークについて、基になったネットワークとノードが 30% 欠損したネットワークのそれぞれについて、拡散率を求める。拡散率は独立カスケードモデル [7] を用いる。

4.1 独立カスケードモデル (Independent Cascade model:IC model)

IC model は情報の送り手を主体とした情報拡散モデルである。IC model において、リンクで結ばれたノード v, u は情報拡散率 $s_{vu} (0 \leq s_{vu} \leq 1)$ をもつ。

ある時刻 t において、アクティブになったノード v は、リンクで結ばれた非アクティブなノード u をアクティブにする機会を一度だけ与えられる。すなわち、時刻 t においてアクティブになったノード v は時刻 $t+1$ に一度だけ情報の発信を行う。その際、ノード v が情報を発信する確率、つまり、ノード v の隣接ノード u がアクティブになる確率は s_{vu} となる。その施行が成功した場合、ノード u は時刻 $t+1$ においてアクティブとなる。この情報拡散過程は、アクティブなノードが増加しなくなった時点で終了する。

各モデルの $t=0$ において、初期アクティブノード v_s を用意する。この時、収束後のアクティブノードの割合を $\sigma(v_s)$ とし、拡散率 σ を以下のように定義する。

$$\sigma = \frac{1}{N} \sum \delta(v_s) \quad (5)$$

なお、 N はネットワークのノード数とする。各ネットワークで 10 回拡散率 σ を算出し、平均をそれぞれのネットワークのモデルにおける拡散率 σ とする。

4.2 拡散シミュレーション結果

図 7 に基となったネットワーク上での拡散率とノードが 30% 欠損したネットワーク上での拡散率を示す。また、 $y = x$ の直

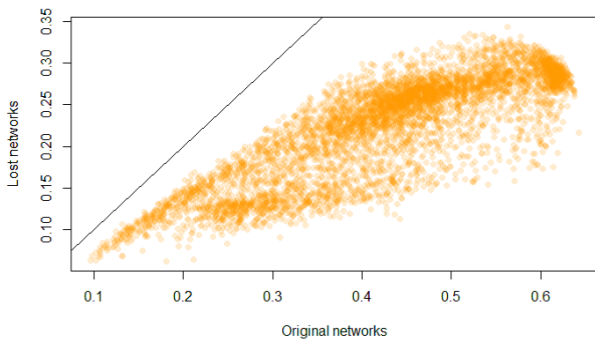


図 7: 基になったネットワーク上での拡散率とノードが 30% 欠損したネットワーク上での拡散率 (直線は $y = x$)

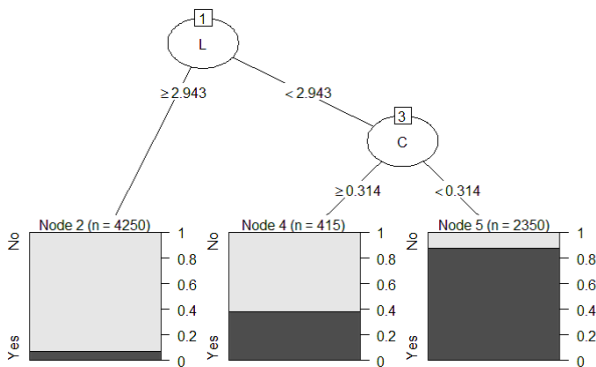


図 8: ノード 30% 欠損したネットワーク上での拡散率が 0.25 以上であるネットワークについての決定木 (Yes: 拡散率 0.25 以上, No: 拡散率 0.25 未満)

線を同時に示す。これより、拡散率が高いネットワーク程、拡散率の低下が大きい事がわかる。さらに、ネットワークによって低下率が違う事がわかる。

次に、ノードが 30% 欠損したネットワーク上での拡散率が高いネットワークについて考える。拡散率が 0.25 以上であるネットワークについて、決定木を構築する。図 8 に構築された決定木を示す。図の Yes は拡散率が 0.25 以上であるネットワークの割合を、No は拡散率 0.25 未満であるネットワークの割合を示している。平均経路長 $L < 2.94$ でクラスタ係数 $C < 0.314$ であるネットワークの 87.5% が拡散率 0.25 以上である。一方で、平均経路長 $L \geq 2.94$ であるネットワークの 92.6% が拡散率 0.25 未満である事がわかる。従って、平均経路長とクラスタ係数が小さいネットワークが頑健なネットワークである事がわかる。

5. 結論

本研究では、頑健なネットワーク構造を発見するために、頑健なネットワークとネットワーク構造の関係について分析した。分析する際に、ネットワーク構造特徴空間を網羅するようなネットワークデータセットを用いた。このネットワークデータセットからノードがランダムに 30% 欠損した際のネットワー

クについて分析した。ネットワークの最大コンポーネントの全体に占める割合が高いネットワークは、平均経路長が小さく、次数分布がべき分布ではないネットワークであった。さらに、最大コンポーネントの全体に占める割合が高いネットワークについて、情報拡散の拡散率を求め、平均経路長が小さく、クラスタ係数が小さいネットワークが頑健である事を示した。

今後の課題としては、本研究で用いた 5 つの特徴以外の特徴についての拡張を挙げる。また、本研究ではランダムな欠損のみを対象としたため、他の欠損についても分析する事を挙げる。

参考文献

- [1] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics, SOMA '10*, pp. 71–79, New York, NY, USA, 2010. ACM.
- [2] Mai Miyabe, Asako Miura, and Eiji Aramaki. Use trend analysis of twitter after the great east japan earthquake. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion, CSCW '12*, pp. 175–178, New York, NY, USA, 2012. ACM.
- [3] 白井 翔平, 鳥海 不二夫. 情報拡散に影響するネットワーク構造特徴. *人工知能学会論文誌*, Vol. 30, No. 1, pp. 195–203, 2015.
- [4] Scott Kirkpatrick, Daniel C. Gelatt, and Mario P. Vecchi. Optimization by simulated annealing. *SCIENCE*, Vol. 220, No. 4598, pp. 671–680, 1983.
- [5] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, No. 393, pp. 440–442, 1998.
- [6] Mark E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, Vol. 67, No. 2, p. 026126, February 2003.
- [7] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '03*, pp. 137–146, New York, NY, USA, 2003. ACM.