

「デマの壁」仮説の検証

篠田 孝祐*2
Kosuke Shinoda

鳥海不二夫*1
Fujio Toriumi

*2理化学研究所/慶應義塾大学
RIKEN/KEIO University

*1東京大学
The University of Tokyo

1. はじめに

「デマの壁」仮説とは、ソーシャルメディア内の情報伝達を阻害・修正する「壁」の存在を指摘する仮説である。「6次の隔たり」と言われるように、人のつながりなど社会に存在するネットワークは少ない隔たりでつながっていると考えられる。それが、確かならば、我々はすべての情報へ接触可能であるように思える。だが、現実にはそういうことはない。例えば、Twitterを介して、「ある出来事が起きたようなツイートがでているがそれは嘘である」という否定情報を知ること、そのような「デマ」の存否を間接的にそのような知ることがある。もしくは、そもそも知る機会のない情報が多数存在するのではないだろうか。特に、先の大震災の時には、『デマ』でなく『「デマ」の否定』を先に知ることや、Twitter上でのデモ活動の呼びかけを報道で知ることではないだろうか？

この様な、出来事は単なる情報伝達経路やタイミングの偶然がもたらした結果でしかないとも考えられる。だが、筆者らの経験だけでしかないかもしれないが、「デマ」や「つまらない」情報が、自分のタイムラインには、ほぼ現れなかったという経験を「気のせい」で済ますことはできない。そこで、我々は、情報伝達を担っている、Twitterをはじめとしたソーシャルメディアには、「正しくない」情報の拡散を食い止めて、時には「正しくない」と情報を訂正する『壁』のような存在があるのではないかと考え、先の大震災時のTwitterデータを用いて「デマの壁」が存在するかどうかの検証を行うことにする。

本論文では、まず、Twitterでの情報流通を考えるうえで前提となるネットワークの特徴を述べ、検証に用いるデータの特徴を述べる。続いて「デマの壁」を検証するうえで、まず取っ掛かりとなるTwitterのネットワークの構造にある特徴的な状況と、その特徴が生じた理由を検討する。そして、今後の課題を含めて、仮説の真偽を証明するプロセスを検討する。

2. TwitterのMentionNetworkの構造的特徴と利用するデータセットの概要

2.1 Twitter上での情報流通構造

Twitterでは、フォロワー・フォロワーから形勢されるネットワークは、2011年1月31日時点のデータでは、その次数分布はベキとなっている[小出12]。そして、その上でやりとりされるMentionNetworkもベキ分布を示す[鳥海11, 小出12]。なお、MentionNetworkとは、Twitterの機能であるMention Tweetをエッジとして形勢されるネットワークである。Mention

Tweetは、Tweetに@User名を含んだものであり、@User名で指定されたユーザはそのTweetを自分宛のTweetとして確認できる。このような場合には、MentionTweetを行ったユーザと@User名で指定されたユーザの間にメッセージのやり取りという形でコミュニケーションが行われたと考えられ、それをエッジとしている。

2.2 対象データセットと東日本大震災の経過概要

本研究は、2011年3月に発生した東日本大震災前後の期間のTweetデータを分析対象とする。本節では、対象期間の主な出来事などを述べ、続いてデータ収集方法ならびにコミュニケーションの分析手法を説明する。

2.2.1 震災の経過

対象期間の主な出来事を以下に記す。

3/9	未明	宮城沖で地震
11	14:46	東日本大震災
	16時前	三陸沖沿岸に大津波到来
	18時頃	首都圏の主な鉄道の運行停止決定
		首都圏のコンビナートで火災が発生
12	未明	福島原発一号機 水素爆発, 余震頻発
14	昼	再度水素爆発, 計画停電開始
16	未明	富士山麓での大規模な地震
17	昼	原発への放水, 最大規模の計画停電実施

先の震災時、本震災直後は普段より多くのユーザがTwitterを利用し[鳥海11]、多数の情報のなかでデマも拡散されていた[鳥海12]。以降の分析では、上記の出来事を踏まえて議論する。

2.2.2 Tweetデータの収集

対象データは、2011年3月7日0時から3月23日24時までの期間に以下の手順で収集した約4億の日本語Tweetである。

1. 主に日本語で発言している1.3百万人のユーザIDのリストを準備
2. 対象期間のTweetを順次クローリング

ただし、TwitterAPIの制限などから、該当期間のすべての日本語Tweetを収集できておらず、8割程度*1である。

3. 距離の中心性の分布の特異性

震災前後のMentionNetworkを分析すると、そのユーザ数やリンク数に違いはあり、そこに参加する利用者がコミュニ

*1 東日本大震災ビッグデータワークショップ - Project 311 - にて提供されたTweetデータとの比較から

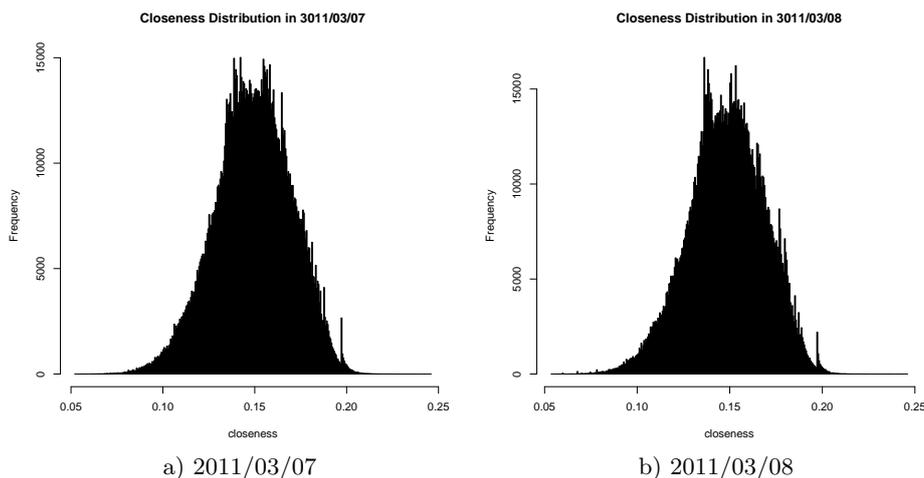


図 1: 平時の Mention Network の距離の中心性分布 (1 日単位)

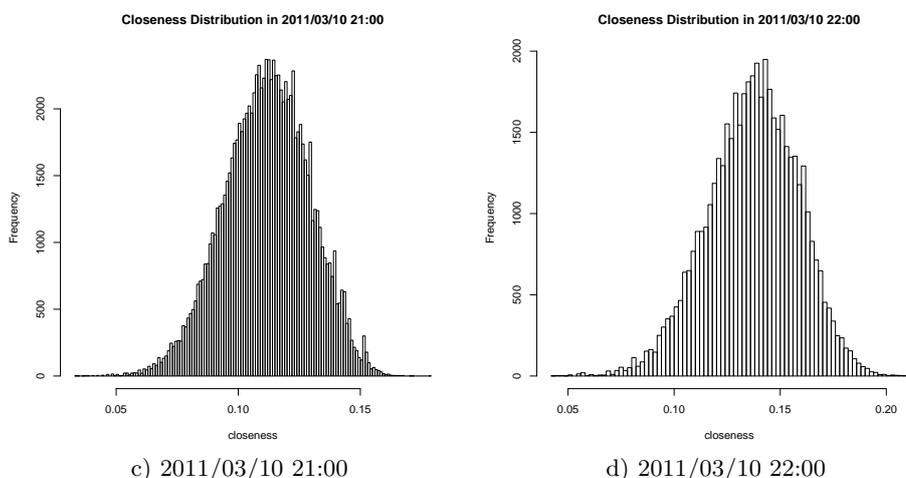


図 2: 平時の Mention Network の距離の中心性分布 (1 時間単位)

ケーションツールとしてというよりはリンクが増加したユーザのタイプから情報源として利用したいという意図が見て取れた [鳥海 11]。それは、ネットワークの構造的特徴量や中心性の変化からも、同様の解釈可能であるという示唆を得ている [篠田 13]。だが、その中で、本論文が目的とする「デマの壁」という仮説の検証はできていない。様々なデータを検討するなかで、デマに限らず一部での情報共有・拡散が行われた時間帯に、ネットワーク中心性の一つである距離の中心性 (closeness) [Freeman 79, Newman 03] の分布に特徴的なデータが得られたため、その特徴的なデータの説明とデマとの関係性を議論する。

3.1 Mention Network の距離の中心性分布

ネットワーク中心性は様々な指標が定義されているが、本研究では代表的な指標である、距離の中心性 (closeness centrality) に着目する。一般的に、ネットワークの構造的特徴を議論する際には、次数中心性 (degree centrality) に着目し、その分布による違いを議論することが多いが、我々が対象とする Twitter の Mention ネットワークは、基本的にベキ分布に近い分布であり、大きな違いはない。

一方で、Mention ネットワークの距離の中心性の分布をみたとき、大きくことなる場合がある。具体的には、図 1 から図 4 は、ある時間帯に行われた Mention ネットワークにおける距離の中心性の分布である。これらは、大きく震災前の時点の分布を示す図 1, 2 と、震災後の状況を示す図 3, 4 に別れ、

それぞれ 1 日単位の分布, 1 時間単位の分布を 2 ケースづつ図示している。

Wasserman ら [Wasserman 94] をはじめとして、距離の中心性の分布は一般的に正規分布になると認識されている [Dorogovtsev 02, Anthony, Hormozdiari 07]。図 1, 2 は、ほぼ正規分布を示しており*2, これらの知見と一致する。それに対して、図 3, 4 は、全体的には正規分布に近い状態を示しているが、分布の何か所かにスパイク状の分布が確認できる。

この様な、スパイク状の分布が明確に確認できるのは、主に震災直後の一定期間と、ユーザ数が減っている早朝の時間帯であり、平時の夕方やよるの時間帯にはほとんど見ることができない。Twitter の Mention ネットワークに、なぜ特定の期間だけ現れるのかを、既知のネットワーク構造の距離の中心性の分布から推定する。

3.2 Small World ネットワークと Scale-Free ネットワーク

一般的に、ネットワークを構成するノード群の距離の中心性の分布は正規分布であるとされているが、特定の構造によって実際にどうなるのかを確認するために、Small World ネットワーク [Watts 98] と Scale-Free ネットワーク [Barabási 99] における距離の次数分布を図 5, 6 に図示した。

*2 視覚的には正規分布のように見えるが、コルモゴロフ・スミノフ検定では、 p 値はほぼ 0 であるため、正規分布とは言えないと思われる

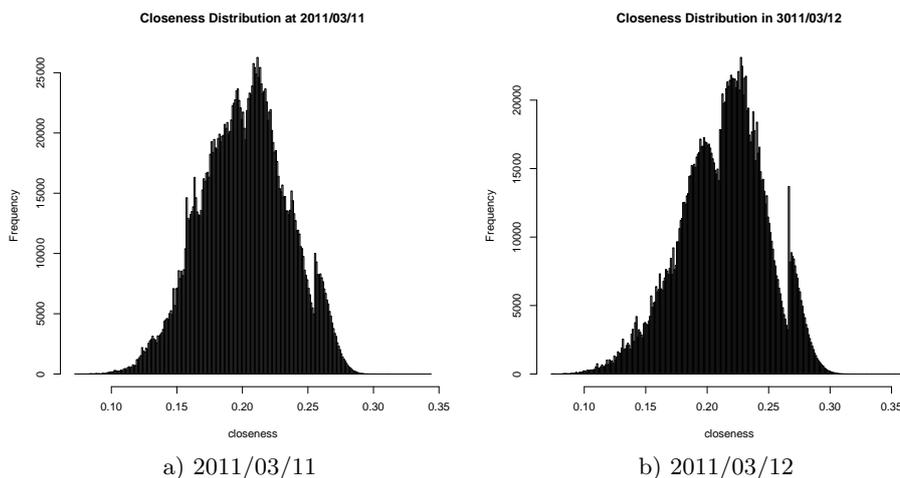


図 3: 震災直後の Mention Network の距離の中心性分布 (1 日単位)

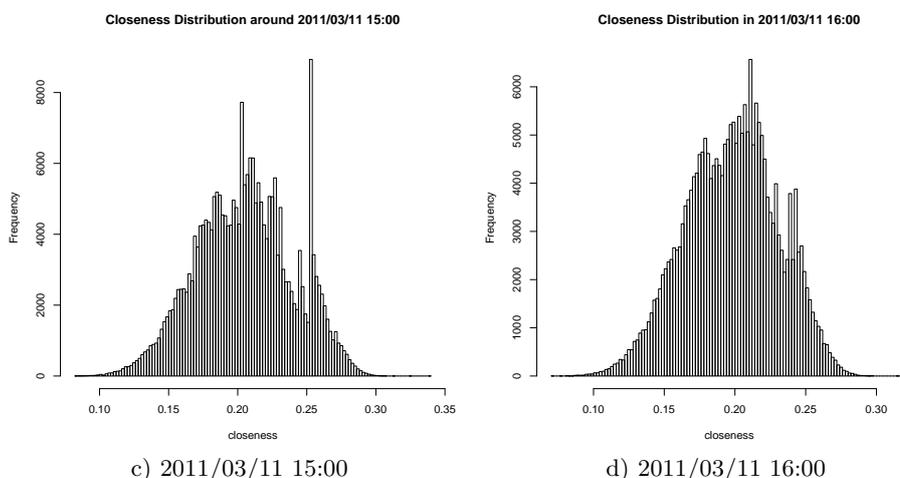


図 4: 震災直後の Mention Network の距離の中心性分布 (1 時間単位)

図 5 の SmallWorld ネットワークは Watts-Strogaz (WS) モデルで正則格子グラフからリンクを組み替えてランダムなグラフへと近づけている。このヒストグラムは、視覚的には正規分布に近いが分布状況であり^{*3}, WS モデルのベースとなる正則格子グラフがすべてのノードの距離の中心性の値が同一であることをからすると、明確な特徴はないとも言える。

一方、図 6 の ScaleFree ネットワークは、BA モデルで作成し、ノードを追加したときの接続するリンクの数による距離の中心性の分布をグラフにしている。視覚的には、SmallWorld ネットワークと違いがあまりないが、距離の中心性の値が高い方に、図 3, 4 と似たスパイクが出ている。このスパイク上の箇所は、その多くが、ハブノードのみに接続した隣接ノードである。このことから、図 3, 4 でのスパイクの数は、多くのノードがリツイートなりで情報を拡散させた形跡だと考えられる。

4. 結論ならびに今後の課題

本論文では、震災前後の Twitter の Mention Tweet から構成されるコミュニケーションネットワークを対象として、一定区間で形成されたネットワークのノード群の距離の中心性の

分布に着目した。そして、基本的な特徴をもつネットワークのノード群の距離の中心性の分布と比較することで、主に震災後のコミュニケーションが活性化したとくに現れる特異な分布に関して、比較した。本来の目的である「デマの壁」の検証までは至らなかったが、従来、距離の中心性の分布は、正規分布となるという見解とは異なる結果であり、Twitter の Mention Tweet が形成するネットワークの特異性を示せた。

今後は、フォローフォロワーネットワークデータを活用して、既存の情報拡散モデルを用いて、仮想的な Mention ネットワークとの比較などを通して「デマの壁」の仮説を検証したい。

謝辞

本研究を行うにあたりデータの収集にご協力いただいた、クックパッド株式会社の兼山元太氏及び株式会社ホットリンクに感謝いたします。また、本研究は、科研費 (24300064) の助成を受けたものです。

参考文献

[Anthony] Anthony, D. and Dekker, H.: Centrality in Social Networks: Theoretical and Simulation Approaches
 [Barabási 99] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, No. 286, pp. 509–512 (1999)

*3 こちらも、Mention Network と同様にコルモゴロフ・スミノフ検定では、 p 値は正規分布とでないことを否定するほどの値とはならない。しかしながら、ランダムグラフに近づくにつれ p 値は 0.5 に近づく

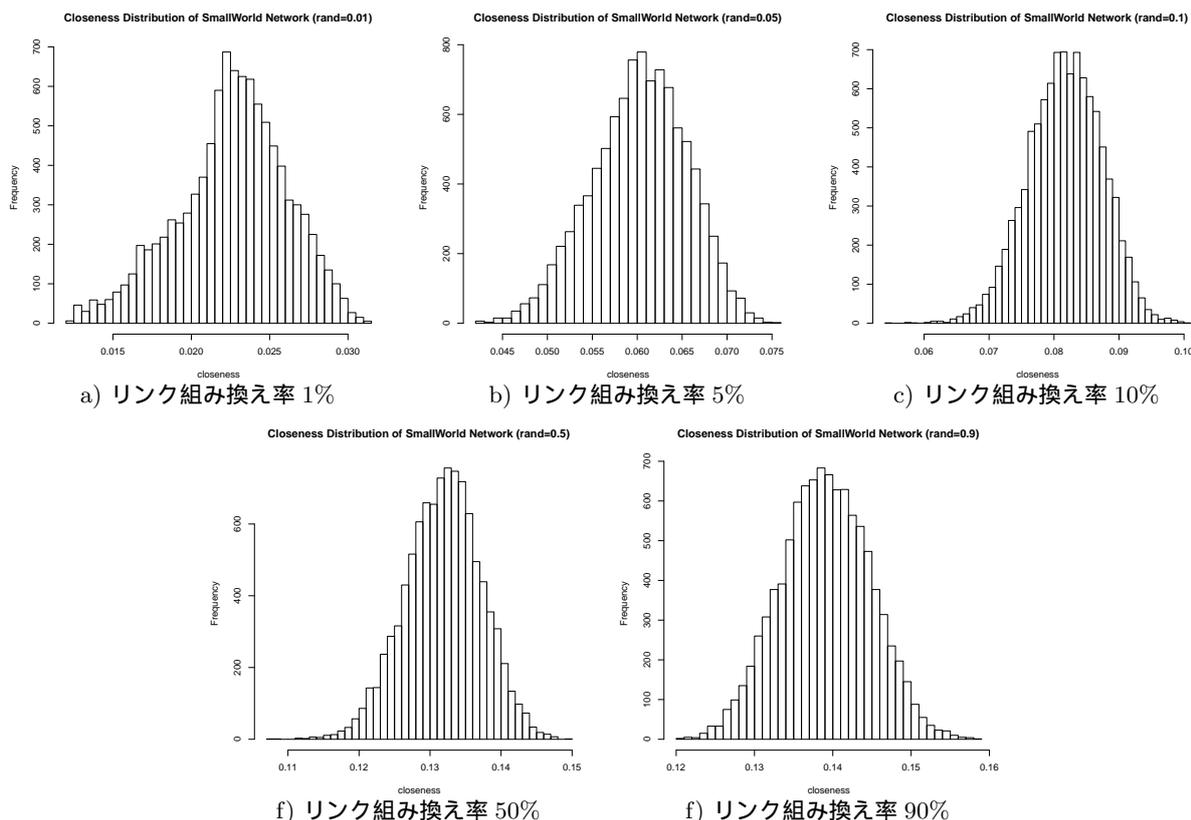


図 5: Small World Network の平均の中心性分布 (正則グラフからランダムグラフまでの分布例)

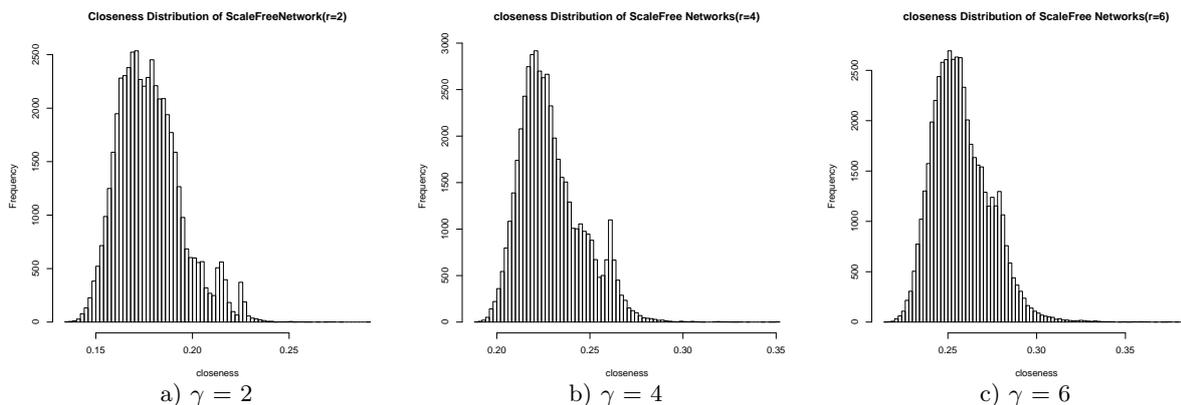


図 6: Scale-Free Network の距離の中心性分布 (接続する次数の違い)

[Dorogovtsev 02] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F.: Pseudofractal scale-free web, *Phys. Rev. E*, Vol. 65, p. 066122 (2002)

[Freeman 79] Freeman, L. C.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, pp. 215–239 (1979)

[Hormozdiari 07] Hormozdiari, F., Berenbrink, P., and Sahinalp, S. C.: Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution, *PLoS Computational Biology*, Vol. 118, (2007)

[Newman 03] Newman, M. E. J.: The structure and function of complex networks, *SIAM Review*, Vol. 45, pp. 167–256 (2003)

[Wasserman 94] Wasserman, S. and Faust, K.: *Social Network Analysis: Methods and Applications*, Cambridge University Press (1994)

[Watts 98] Watts, D. J. and Strogatz, S. H.: Collective dynamic of small-world networks, *Nature*, Vol. 393, pp. 440–442 (1998)

[篠田 13] 篠田 孝祐, 榊 剛史, 鳥海 不二夫, 風間 一洋, 栗原 聡, 野田 五十樹, 松尾 豊: 東日本大震災時における Twitter の活用状況

とコミュニケーション構造の分析, 知能と情報, Vol. 25, No. 1, pp. 598–608 (2013)

[小出 12] 小出 明弘, 斉藤 和巳, 風間 一洋, 鳥海 不二夫: コリンク構造に着目した多重グラフの特性分析, 日本データベース学会論文誌, Vol. 11, No. 2, pp. 13–18 (2012)

[鳥海 11] 鳥海不二夫, 篠田孝祐, 栗原聡, 榊剛史, 風間一洋, 野田五十樹: 震災がもたらしたソーシャルメディアの変化, ネットワークが創発する知能研究会 (JWEIN'11) (2011)

[鳥海 12] 鳥海不二夫, 篠田孝祐, 兼山元太: ソーシャルメディアを用いたデマ判定システムの判定精度評価, 情報処理学会デジタルプラクティス, Vol. 3, No. 3, pp. 201–208 (2012)