D-034

Twitter の reply と retweet 関係からなる成長ネットワークの比較分析

Comparative analysis of growing networks from reply and retweet relations on Twitter

加藤翔子 † 大久保誠也 † 斉藤和巳 †
Shoko Kato Seiya Okubo Kazumi Saito

1. はじめに

ソーシャルメディア上における人間関係から構築されるネットワークは，ユーザーが個人の興味に基づいて人間関係を追加・削除することにより，絶えず成長している．このような複雑な構造を有するネットワークの成長をモデル化することは，人間関係における友好関係の発見や，バイラルマーケティングなどへの応用が期待され，盛んに研究がおこなわれている [1, 2].

小出らの研究 [3] では，人間関係のように複雑なネットワーク成長モデルを構築するための第一歩として，Twitter 上における @-message ネットワークの時間変化に伴う連結成分構造を分析した．これに対し，本研究では，@-message に内包される機能である reply と retweet を，それぞれ別のネットワークとして構築する．Twitter において，reply とは@user で指定した相手と会話をする機能であるので，reply ネットワークはユーザー間のコミュニケーションによって構成される．一方，retweet は@user で指定した相手の Tweet を，他のユーザーにも拡散する機能であるので，retweet ネットワークはユーザー間の情報の流れによって構成される．この 2 つのネットワークを，時間変化に伴うノード数，リンク数，連結成分数，Gini 係数の変化を分析することで，その構造にどのような違いがあるのかを比較する．

また，東日本大震災前後の Tweet データを分析に用いることで，社会的な出来事が 2 つのネットワークに及ぼす影響も比較する．なお，東日本大震災を例とする社会的な出来事を，以降本研究ではソーシャルイベントと呼ぶ．

本実験で用いたデータにおいては，ノード数，リンク数，連結成分数，Gini 係数の時間変化の分析結果より，reply ネットワークと比較したとき，retweet ネットワークはソーシャルイベントによる影響を受けやすいことを示す．

2. 分析手法

与えられた成長ネットワークに対し，時刻 t でのノード数を $n(t)$，リンク数を $e(t)$，連結成分数を $c(t)$ と定義する．これらの値が，時間の経過によってどのように変化するのかを分析する．

2.1.Gini 係数

連結成分数により，時間変化に対する連結成分数を得ることはできるが，各連結成分に属するノード数のばらつきの程度はわからない．これを定量的に評価するため，本分析では，Gini 係数 [4] を用いる．いま，時刻 t での各連結成分に属するユーザ数を $n_1(t), \cdots, n_{c(t)}(t)$ で表す．この時，時刻 t での連結成分 $c(t)$ の Gini 係数 $G(t)$ は，次式で定義される．

$$G(t) = \frac{\sum_{i=1}^{c(t)-1} \sum_{j=1+1}^{c(t)} |n_i(t) - n_j(t)|}{(c(t)-1) \sum_{c=1}^{c(t)} n_i(t)}$$

この数式は，1 つの巨大な連結成分になるほど，$G(t)$ の値が大きくなり，同程度のノード数を持つ複数の連結成分によって構成されるネットワークでは，その値は小さくなる．

3. 実験

3.1. データ概要

本研究では，2011 年 3 月 5 日 00:00:00 から同月 24 日 23:59:59 までの日本語で投稿された Tweet の中から，文頭が”@user”から始まる Tweet を reply，”RT @user” で始まる Tweet を retweet として収集し，ネットワークを構築する．各ユーザをノードとし，各ユーザから @user で指定されたユーザへリンクを張ることで，各ユーザのリンク関係のデータを作成する．各リンクには時刻を付与し，時間経過によるネットワークの成長過程を分析する．

最終的に生成されたネットワークのノード数と多重を考慮したリンク数については，表 1 の結果となる．

表 1: ネットワークの最終ノード数と多重リンク数

	ノード数	多重リンク数
reply ネットワーク	4,023,145	30,170,550
retweet ネットワーク	2,157,981	30,932,951

3.2. 分析結果

時間遷移に対するノード数 $n(t)$，リンク数 $e(t)$，連結成分数 $c(t)$，Gini 係数 $G(t)$ を，図 1，図 2，図 3，図 4 にそれぞれ示す．横軸は各日付の 0 時の時点を表す．

図 1，図 2，図 3，図 4 のいずれにおいても，3 月 11 日の昼以降から夜にかけての retweet ネットワークに顕著な動きが見られる．

図 1 では，11 日の昼以降から retweet ネットワークのノード数が急激に増加し，その後は 11 日以前と同じペースで増えていることがわかる．これは，震災前には retweet をしなかったが，震災のあった 11 日昼ごろから夜にかけて初めて retweet した，というユーザーが多く存在することを示唆している．

図 2 でも同様に，11 日の昼以降から retweet ネットワークのリンク数が急激に増加し，その後は 11 日以前と同じペースで増えていることがわかる．これは，こ

†静岡県立大学大学院

の期間内における retweet の回数が，震災前や 12 日以降より多いことを意味する．

図 3 では，retweet ネットワークにおける連結成分数が 11 日昼ごろから夜にかけて減少し，その後は 11 日以前と同様に横ばいになっている．これにより，複数の連結成分間を繋ぐリンクがこの期間内に生成され，一つの連結成分になったことが示唆される．

図 4 では，retweet ネットワークにおける連結成分の Gini 係数が，11 日昼ごろから夜にかけて顕著に高まっており，巨大な連結成分が生成されたことがわかる．これを図 3 と合わせると，11 日以前に存在した複数の連結成分が，この期間内で最大連結成分に統合され，その結果最大連結成分がネットワークのほとんどを占める状態となったと考えられる．

reply ネットワークにおいては，図 1 と図 2 においてノード数，リンク数の単調増加，図 3 において連結成分数の単調減少が確認できる．また，図 4 より，retweet より早い段階で最大連結成分がネットワークのほとんどを占める状態となることがわかる．いずれの分析においても，retweet ネットワークのような震災による影響は顕著に見られない．

4. おわりに

本研究では，Twitter 上の成長モデルを構築するための第一歩として, reply と retweet をネットワークとして構築し，時間変化に伴う連結成分構造を分析した．その結果，retweet ネットワークは reply ネットワークよりソーシャルイベントによる影響を受けやすく，reply ネットワークは retweet ネットワークより早い段階で最大連結成分がネットワークのほとんどを占める状態となることを確認した．

今後は，本分析で得られた知見を用いて，さらに大規模な Twitter データや，ソーシャルイベントのない定常状態での Twitter データを用いて分析し，ネットワークの成長モデルを構築し，評価をしていく．

謝辞

本研究は，株式会社豊田中央研究所との共同研究，および，科研費 (No.23500312) の助成を受けた．

参考文献

[1] R.Albert, and A.L.Barabasi. ”Statistical mechanics of complex networks”, Reviews of Modern Physics ,vol. 74, pp. 47-97,2002.

[2] M.E.J.Newman. “ The structure and function of complex networks. ”, SIAM Review, 45:167—256,2003.

[3] 小出明弘, 斉藤和巳, 大久保誠也, 鳥海不二夫, 風間一洋, 「Twitter の@-message で構成される成長ネットワークの分析」, 情報処理学会 第 74 回全国大会, 2012.

[4] M.J.Salganik, P.S.Dodds, and D.J.Watts. “ Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market”, Science 10 February 2006,pp.854-856

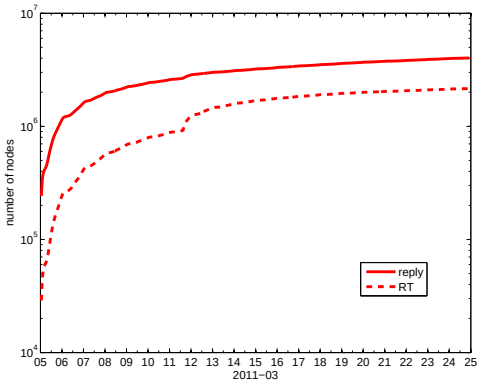

図 1: 時間遷移に対するノード数 $n(t)$

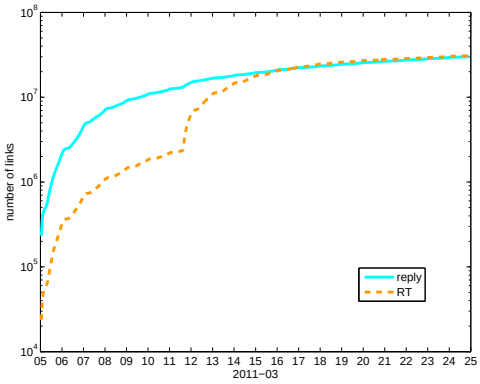

図 2: 時間遷移に対するリンク数 $l(t)$

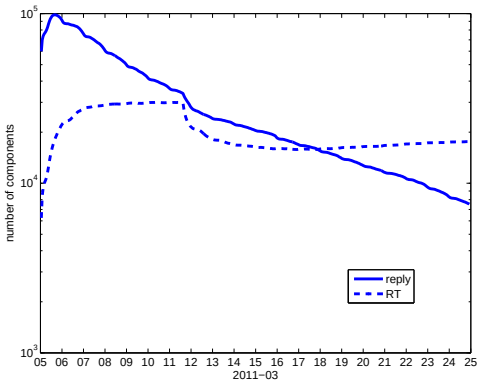

図 3: 時間遷移に対する連結成分数 $c(t)$

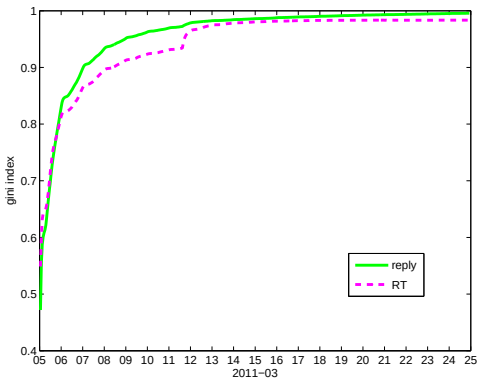

図 4: 時間遷移に対する Gini 係数 $G(t)$