

多重有向コア抽出法による Twitter データの震災時と通常時の特性比較

加藤 翔子† 齊藤 和巳† 風間 一洋†† 佐藤 哲司†††
 †静岡県立大学 ††和歌山大学 †††筑波大学

1 はじめに

現実社会の人間関係や、Web サービス上のユーザ間のつながりなど、社会ネットワークに対する関心が高まっている。特に、ソーシャル・ネットワーク・サービスの一つである Twitter(<http://twitter.com/>)は、多くの研究者に注目され、様々な知見が得られている [1][2]。

このような現実のネットワークはコア構造を持つために、場所によってリンク密度が異なることが指摘されている [3]。このコミュニティ構造を自動的に抽出できれば、同じ興味のユーザの抽出や、スパムなどの問題があるノード群の除外が可能となり、社会ネットワーク分析の上で有用な手法となることが期待できる。

論文 [4] では、単純無向ネットワークのコア抽出を行う SR(Spectral Relaxation) 法を多重有向に拡張した MDSR(Multi-Directed-SR) 法を提案し、大規模多重有向ネットワークのコア抽出を行った。本稿では、東日本大震災直後と平常時における Twitter の reply ツイート [6][7] に MDSR 法を適応し、抽出された各コアに含まれるトライアドを俯瞰することで、2つのネットワークの構造を比較する。

2 分析手法

与えられたグラフ G の全ノード集合を $V = \{1, \dots, N\}$ とし、その隣接行列を \mathbf{A} とする。隣接行列の第 (i, j) 成分 $A(i, j)$ は、ノード i から j へ張られたリンクの本数を表す。本稿では、自己リンクなしの多重有向グラフを対象とするので、 $A(i, i) = 0$ かつ $A(i, j)$ は非負整数となる。このグラフについて、以下の手法でコア抽出と可視化を行う。

2.1 MDSR 法

MDSR 法 [4] では、多重有向隣接行列の右固有ベクトルと左固有ベクトルを 2 値に量子化してコア部を抽出する。さらに、隣接行列から抽出したコア部に含まれるリンクを削除した後に上記の処理を適用し、再帰的にコア部を抽出する。右固有ベクトルは HITS アル

ゴリズムのオーソリティ度ベクトル、左固有ベクトルはハブ度ベクトルに対応する。

V の 2 つの部分集合 W と X に対し、その間に張られたリンク密度 $G(W, X)$ を次式で定義する。

$$G(W, X) = \frac{1}{\sqrt{|W||X|}} \sum_{i \in W} \sum_{j \in X} A(i, j). \quad (1)$$

ここで、 $|W|$ や $|X|$ は、集合 W や X の要素数を表し、 $G(W, X)$ はこれらの幾何平均である。この式 (1) を最大にするノード部分集合ペア W と X の探索問題を考えることで、リンクが密集しているコア部分を抽出する。

2.2 トライアド

トライアドとは、有向グラフにおいて、任意に選択した 3 ノードの接続の状態を表すものである。Wasserman ら [5] によって定義された 16 種類のトライアドには、3 ノード全てが非接続であるものから、3 ノード全てが相互接続しているものまで含まれるが、我々は 3 ノードが連結である 13 種類のトライアド (図 1) を用いて、2.1 節の手法で抽出した 100 個のコアのうち、オーソリティノードとハブノードの総数が 3 以上のコアに含まれるトライアドの頻度を観察する。

2.3 k -近傍法

k -近傍法は、パターン認識の分野でよく用いられる単純な分類アルゴリズムである。本稿では、抽出したコア構造をノードとし、トライアドの頻度ベクトルのコサイン類似度が高い関係のあるコア間のみリンクを張ることでネットワークを作成し、ばねモデルを用いて可視化する。ネットワーク作成の詳細な手順を以下に示す。

1. 全ノードペア間のコサイン類似度を計算する
2. 各ノードに対して、自身との類似度で降順に相手ノードを並び替える
3. $k \leftarrow 1$ とする
4. 各ノードは、自身との類似度が k 位タイのノード全てとリンクを結ぶ
5. 全ノードが単一連結成分になれば終了する。そうでなければ $k \leftarrow k + 1$ とし、4 に戻る

なお、本稿のネットワーク可視化におけるリンクは、各コア間の類似度構造の複雑さをより明確化するための補助線として扱う。

Comparative analysis of Twitter data when the earthquake occurs and normal period using MDSR method

†Shoko KATO †Kazumi SAITO ††Kazuhiro KAZAMA

†††Tetsuji SATOH

†University of Shizuoka ††Wakayama University

†††University of Tsukuba

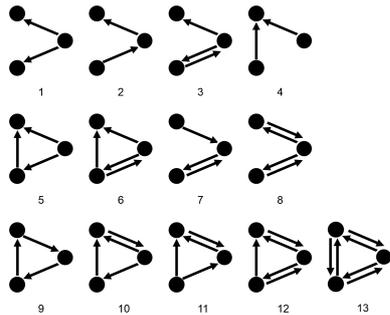


図 1: トライアドの種類

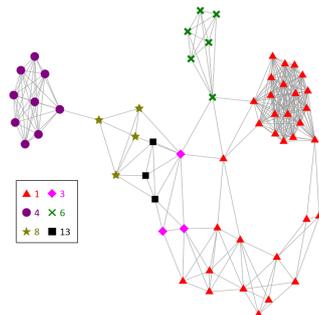


図 2: 平常時

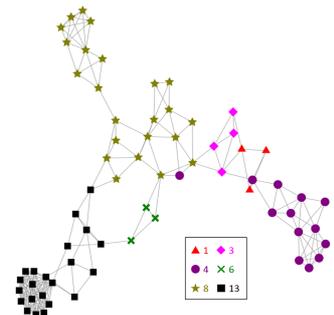


図 3: 震災時

3 評価実験

東日本大震災直後と平常時における Twitter の reply ツイート [6][7] の多重有向ネットワークについて、2 節で説明した手法でコア抽出と可視化を行った結果を、図 2, 3 に示す。なお、凡例中の数字は図 1 に対応し、各コアにおいて最も頻出なトライアドを表す。

これらの結果より、平常時は 1 番のトライアドが頻繁に見られるが、震災時は 13 番や 8 番のトライアドが多く見られる。また、各コアを形成するアカウントについて調査したところ、1 番トライアドが頻出なコアのハブは bot アカウントが多く、8 番や 13 番が頻出なコアは一般ユーザと bot アカウントが混在していた。

したがって、平常時は少数の bot が多数のアカウントに機械的に reply を送る片方向リンクが多く見られるが、震災時は互いにコミュニケーションを取り合う双方向リンクが増えることが示唆される。これにより、震災時には、安否確認や不安から、通常より頻繁に連絡を取り合うケースが多かったことが推測できる。

4 おわりに

震災時と平常時における Twitter の reply データを、多重有向ネットワークとしてコア抽出分解を行った。抽出したコアのうち、トライアドを持つものについて k -近傍法で可視化することで、2 つのネットワークの構造を比較した。その結果、平常時には bot アカウントによる一方通行の reply が多く見られるが、震災時には一般ユーザが互いにコミュニケーションを取り合う場合が増えることが明らかとなった。

今後は、他のデータを用いての同様の実験や、単純有向ネットワークや重み付き有向ネットワークとしてのコア抽出分解を検討し、さらに定量的な評価をおこなう予定である。

謝辞 本研究は、科学研究費補助金基盤研究 (B) (No.25280110) の助成を受けた。

参考文献

- [1] B.A.Huberman, D.M.Romero and F.Wu, "Social networks that matter: twitter under the microscope", First Monday, Volume 14.Number 1.January 5 2009.
- [2] H.Kwak, C.Lee, H.Park, and S.Moon, "What is twitter, a social network or a news media?" In Proceedings of the 19th international conference on World wide web, pp.591-600.ACM, 2010.
- [3] E.J.Newman, and J.Park, "Why social networks are different from other types of networks", Phys. Rev.E, Vol.68, No.3, p.036122, 2003.
- [4] 加藤翔子, 齊藤和巳, 風間一洋, 佐藤哲司, "MDSR 法を用いた reply ツイートネットワークの特性分析", 第 6 回 Web とデータベースに関するフォーラム (WebDB Forum 2013), 2013.
- [5] S.Wasserman, K.Faust, "Social Network Analysis", pp566, 1994.
- [6] 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, 2011.
- [7] 山口 裕太郎, 山本 修平, 佐藤 哲司, "投稿活動の変化に着目したマイクロブログユーザの可視化手法の提案", 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2013) シンポジウム, 1C-3, pp.72-79, 2013.