

自己中心トライアドに基づくソーシャルメディア特性分析 User Analysis Based on Egocentric Triads

山田 卓見*
Takumi Yamada

大久保 誠也*
Seiya Okubo

斉藤 和巳†
Kazumi Saito

1. はじめに

ソーシャルメディア上には、多様なソーシャルネットワークが構築されている。このようなネットワークの構造分析を目的として、小出らは、高次数ユーザの隣接関係の類似度をデンドログラムなどで分析した [1, 2]。また西らは、小出らと同様に、レビュー、ブログ、マイクロブログの各サイトの高次元ユーザに着目し、自己中心ネットワークのトライアド分布による各サイトの分析を行い特徴を明らかにしてきた。 [3]。

本研究では、新たにソーシャルニュースサイトの特徴をトライアドに関して分析し明らかにする。レビュー、ブログ、マイクロブログ、ソーシャルニュースサイトの各サイトとして、@cosme, Ameblo, Twitter, digg を用いる。実験には、中心ユーザと入出力リンク関係を持つユーザ群から自己中心ネットワークを構築して各ユーザのトライアド分布を求め、これら分布間の類似度に基づく k -近傍グラフを可視化して評価する。各サイトの実験結果と digg の実験結果を比較し、ソーシャルニュースサイトの特徴を明らかにしていく。

2. 分析手法

本研究では、[3]と同様の分析を行う。

トライアドとは、有向グラフにおいて、任意に選択した連結3ノードの接続パターンを表すもので [4]、図1に示すように、可能なパターン数は13となる。本研究では、次節で説明する3つのネットワークにおいて、次数が上位 M 位以内のノードとその隣接ノードのトライアドを分析する。

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だけにリンク

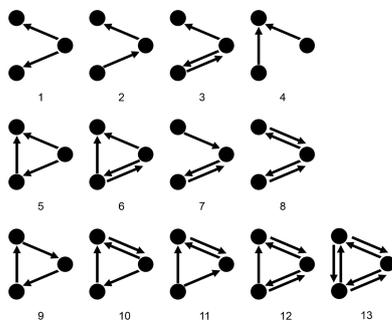


図1: トライアドの種類

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ネットワーク可視化を行う。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

ネットワークを $G = (V, E)$ とし、分析対象とする高次数ノードの数を M とすれば、提案分析法の処理手順は以下となる。

1. $m = 1$ から M まで以下の処理を繰り返す
 - (a) 第 m 位の高次数ノードと入出力リンク関係を持つノード集合を $V_m \subset V$ とし、 V_m 間の全リンク集合 E_m を求め、自己中心ネットワークを $G_m = (V, E_m)$ として構築する。
 - (b) ネットワーク G_m に出現するトライアドのパターンを数え上げ、13次元のトライアド頻度ベクトル \mathbf{x}_m を構築する。
2. トライアド頻度ベクトルの集合 $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ から k -近傍ネットワークを構築し、このネットワークをバネモデルで可視化する。

ここで各ノード m に対し、高頻出トライアドパターンを $\arg \max_{1 \leq i \leq 13} \{x_{m,i}\}$ として求め、そのパターンに応じてマーカーと色を指定して可視化する。

3. 実験データ

本稿で用いたデータは、以下の4つである。

一つ目は、化粧品の口コミサイト“@cosme”* から2009年12月に取得したお気に入りネットワークである。サイトに登録しているあるユーザ X がユーザ Y をお気に入り登録をすると、ユーザ X からユーザ Y に対してリンクが張られる。このネットワークはランダムに選択したユーザから10段まで辿って収集したもので、ノード数は45,024ノード、リンク数は546,930本である。

二つ目は、日本のブログサービスサイト“Ameba”†の読者ネットワークである(以下、ameblo)。“Ameba”の各ブログには読者のブログへのハイパーリンクが張ってある読者リストがある。ここで、ブログ X から

*静岡県立大学 経営情報学部

†神奈川県立大学 理学部 情報科学科

*<http://www.cosme.net/>

†<http://ameblo.jp/>

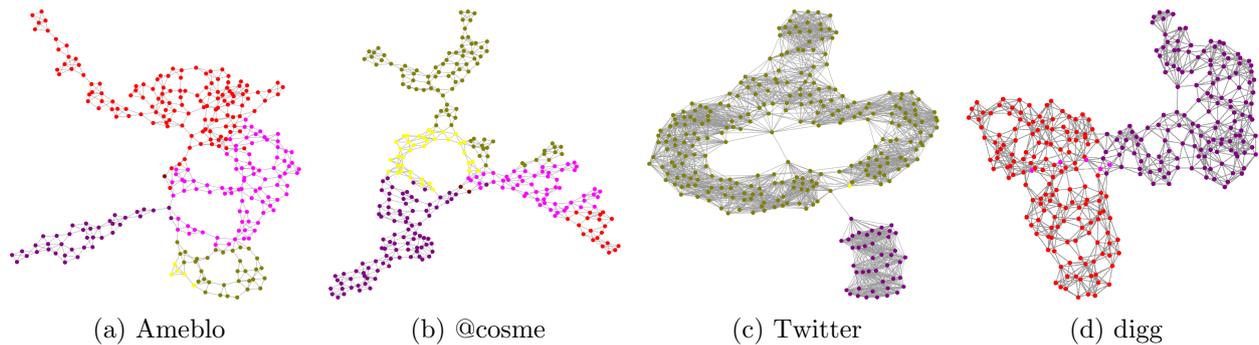


図 2: 各 SNS・ユーザ繋がり

ブログ Y への読者リンクは、ブログ X がブログ Y をお気に入りブログとして登録した時に生成される。このネットワークのノード数は 56,604 ノード、リンク数は 1,071,080 本である。

三つ目は、Twitter 検索[‡]で収集された日本語で書かれたツイートを公開しているユーザのフォローネットワークである。ここで、ユーザ X からユーザ Y へのリンクは、ユーザ X がユーザ Y をフォローしたときに生成される。本稿では、2011 年 1 月 31 日時点において累積 200 回以上ツイートしているユーザのフォロー関係から、ユーザをノード、フォロー関係をリンクとした有向ネットワークを作成した。ネットワークのノード数は 1,079,986 ノード、リンク数は 157,371,341 本である。

四つ目は、ソーシャルニュースサイト digg[§]で、ユーザがインターネット上で見つけた興味深いニュースをリンクによって紹介し、参加型のユーザが共有した際のネットワークである。digg は閲覧したユーザ X が公開されているニュースに対し vote(投票) することで、公開しているユーザ Y の評価をしている。公開しているユーザ Y が閲覧しているユーザ X にフォローされているため、ブログサイトと同じようなリンクになるユーザが多くなった。しかし、ほかのサイトと比べ、機能種類がシンプルなためにユーザの繋がり方もよりシンプルなものとなった。このネットワークのノード数は 336,225 ノード、リンク数は 1,731,658 本である。

4. 分析結果

分析結果を図 2 に示す。@cosme, Ameblo では一方的結合のトライアド 1(赤), 4(紫) と相互結合のトライアド 8(オリーブ), また中間的なトライアド 3(ピンク), 7(黄) のユーザが存在した。Twitter には、相互結合のトライアド 8 のユーザが支配的に存在しており、ある一点から一方的なトライアド 4 のユーザが存在するようになっている。digg には、一方的なトライアド 1, 4 のユーザのみしか存在しなかった。

Ameblo, @cosme, digg では一方的に結合するユーザが多数存在した。情報の発信源となるユーザと、読み手側のユーザに別れるメディア上の機能により、このようなネットワークが構築されている。しかし Ameblo,

@cosme は一つのユーザが情報発信と読み手どちらの立場にもなりうることで、中間的なトライアドのユーザ出現によりわかる。digg は発信側と読み手が分かれるのが特徴である。Twitter はフォロー機能により、相方向的に繋がるユーザが多くなる。お気に入り機能で一方的な繋がりはあるが、相互結合が特徴なユーザが支配的であるとわかった。

5. おわりに

本研究では、西らの研究のレビュー、ブログ、マイクロブログのソーシャルメディア分析に加えて、digg というソーシャルニュースサイトを加えた。そうすることで従来の研究で分析されていた各サイトの実験結果と digg の結果を比較し、ソーシャルニュースサイトの特徴を明らかにしようとした。その結果、ソーシャルニュースサイトでは情報発信専門ユーザと読み手専門ユーザによって構築されるネットワークであることが明らかになった。今後は、本分析で得られた知見を用いて、さらに大規模なソーシャルメディアデータや自己中心ネットワークのトライアド分布の分析法を評価していく。

謝辞 本研究は、科学研究費補助金基盤研究(C)(No.18K11441)の助成を受けた。

参考文献

- [1] 小出明弘, 齊藤和巳, 風間一洋, 鳥海不二夫, "ネットワーク分析による Twitter ユーザのフォロー形成に関する一考察," 情報処理学会論文誌, 数理モデル化と応用-ネットワークが創発する知能論文特集-, Vol.6, No.2, pp.164-173, 2013.
- [2] 小出明弘, 齊藤和巳, "マイクロブログ, ウェブログ, レビューサイトのネットワーク構造比較," 第 97 回 知識ベースシステム研究会 (SIG-KBS), 2012.
- [3] 西可南子, 加藤翔子, 大久保誠也, 齊藤和巳, "レビュー, ブログ, マイクロブログサイトの高次数ユーザの自己中心トライアド分析" 第 13 回情報科学技術フォーラム (FIT2014), 2014
- [4] S.Wasserman, K.Faust, "Social Network Analysis," pp566, 1994.

[‡]<http://yats-data.com/yats/>

[§]<http://digg.com/>