3I3-OS-14a-7

Twitterにおける情報フローネットワークの提案と分析

Proposal and Analysis of the Information Flow Network in Twitter

倉持 俊也 Toshiya Kuramochi　土方 嘉徳 Yoshinori Hijikata　西田 正吾 Shogo Nishida

大阪大学大学院 基礎工学研究科
Graduate School of Engineering Science, Osaka University

Recently, a lot of researchers have focused on the Twitter network. One of the hottest field of research is influencer extraction. The basic approach is to apply centrality measures (e.g., degree-centrality or ranking algorithms such as PageRank) to the network. But we already know that the Twitter network is not just a social network—it's also an information sharing network. We think considering the actual information diffusion path is essential for measuring the actual influence. We propose a novel network model *information flow network* and investigate its property through the real data experiment.

1. はじめに

近年，Facebook や Twitter をはじめとするソーシャルメディアの普及により，ユーザ間で情報を共有する行為が一般的となっている．例えば，日頃の活動を撮影した写真を友人間で共有し，また，インターネット上で発見した面白いブログ記事を同じ趣味の仲間で共有している．情報の共有は直接の知人関係や限られたコミュニティ内に留まらない．価値の高い情報や大勢が関心を持っている情報はネットワーク上のエッジを伝わって大規模に拡散される．Twitter では，情報はユーザ間の伝播を繰り返し，合計で 10 万以上のユーザに到達することがある [Bakshy 11]．このような特徴を有する Twitter のネットワークは，多くの SNS が表現する社会的な知人関係を表したソーシャルネットワークではなく，情報共有のネットワークであると言われている [Kwak 10, Wu 11]．

一方，メディアコミュニケーションに関する一連の研究において，古くからオピニオンリーダーの存在や二段階流れの仮説 [Katz 55] などが提唱されてきた．近年では，ブログや Twitter などのネットワークを対象として，これらの理論の検証が行われている [Wu 11]．特に，他者に強い影響力を持つ少数のユーザ群 [Rogers 62] を発見することに注目が集まっている．例えば，Twitter のフォロー関係のネットワークにおいて次数や PageRank などの中心性尺度によりユーザの影響力を調査する研究 [Kwak 10] や，リツイートやメンション (会話) の数に着目して影響力を調査する研究 [Cha 10] などがある．これらの研究は，Twitter のフォロー関係における中心性尺度は，ユーザの人気度は表すが，必ずしも影響力を表しはしないことを明らかにした．さらに，ユーザのフォロー関係にトピック情報を付与し，PageRank ベースのアルゴリズムで影響力の強いユーザを抽出する方法 [Weng 10] が提案されている．また，あるユーザを始点とした情報伝播がどれだけ広く拡散されるかに着目し，機械学習により始点のユーザの特徴から拡散規模を予測する研究 [Bakshy 11] も存在する．

我々は，Twitter のような情報共有のネットワークにおいて，様々な種類の情報拡散の経路を観測できることに注目する．実際に起こった複数の情報拡散の経路を重ね合わせることで，実際に情報の伝播しやすいエッジを検出できると考えられる．本研究では，Twitter のネットワークにおいて，図 1 のように実際の情報拡散の経路を重ね合わせ，さらに情報の意味的情報を用いて生成される情報フローネットワークの提案を行う．

連絡先: 倉持 俊也，大阪大学大学院 基礎工学研究科，大阪府 豊中市 待兼山町 1-3，06-6850-6383，kuramochi@nishilab.sys.es.osaka-u.ac.jp

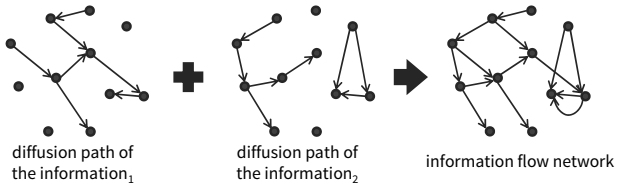

図 1: 情報フローネットワーク

2. 情報フローネットワーク

2.1 情報伝播

一般に，ネットワーク上を情報が伝播したかどうかを判断することは困難である．本研究では，既存研究 [Bakshy 11, Wu 11] で行われているように，ツイートに含まれる URL に着目することで，情報伝播を明確に定義する．すなわち，あるユーザが URL を含むツイートを発信し，それ以降の時間にそのユーザのフォロワーが同じ URL を含むツイートを行った場合に情報が伝播したと見なすこととする．

情報として URL を用いることで，ネットワーク中の情報の流れを追跡できるというメリットがある．しかしながら，URL が情報の全てではない．ブログ間の情報伝播を調査した [Matsumura 08] のように，ツイート中に出現する単語がどれだけ重複するかに注目することで，さらに多様な情報の伝播を扱うことができると考えられる．これは今後の課題とする．

2.2 情報フローネットワーク

Twitter のユーザ間のフォロー関係のネットワーク $G^0 = (V^0, E^0)$ は有向グラフである．隣接行列 A^0 は以下の通りである．

$$A^0 = \{a^0_{v,w}\}_{v,w \in V^0}, \quad a^0_{v,w} \in \{0, 1\} \tag{1}$$

ネットワーク G_0 中の任意のユーザがツイートした URL の集合を U とする．全ての URL は，意味的情報により N 次元

のベクトルとして，

$$u_i = (u_{i,1},\, u_{i,2},\, \cdots,\, u_{i,N}) \tag{2}$$

のように表現されているとする．なお，意味的情報とはサイト上でのカテゴリや付与されたタグのような離散値か，トピックモデルや *tf·idf* によって算出された連続値を指す．また，ノード v から w に伝播した URL の集合を $U_{v,w} \subset U$ と表記する．

ユーザ間のフォロー関係のネットワーク $G^0 = (V^0, E^0)$ と URL 集合 U から，実際に情報が伝播したエッジに注目することで情報フローネットワーク $G = (V, E)$ を生成する．情報フローネットワークのノード集合 V は，フォロー関係のネットワークのノード集合 V^0 のうち，URL を含むツイートを行ったユーザであるので，常に $V \subset V^0$ である．情報伝播の方向はフォローの方向とは反対であるため，エッジ集合 E は，$E \subset (E^0)^T$ である．ただし，右肩の T は転置行列を表す．また，G^0 は有向グラフとして表されるが，G は有向多重グラフとして表され，隣接行列 A の (v, w) 成分は，

$$a_{v,w} = \sum_{u_i \in U_{v,w}} u_i \tag{3}$$

のように，各 URL のベクトルの和により計算される．

情報フローネットワークは，実際の情報伝播の経路に着目したネットワークである．さらに，エッジごとに流れやすい URL の性質が異なると考えられるため，エッジには URL の意味的情報に基づくベクトルを付与する．

3. 評価実験

多くの研究でされてきたように，次数中心性に基づくユーザランキングを作成することで，(1) 情報フローネットワークとフォロー関係のネットワークを比較し，(2) 意味的情報の効果を検討する．

我々は，3 月 28 日に New York Times の Web サイトに投稿された記事のうち，World，Business，Opinion，Sports，Arts の 5 カテゴリの記事を全て収集した．さらに，4 月 4 日までにそれらの URL を含むツイートを行ったユーザを全て収集し，また，それらのユーザ間のフォロー関係を取得した．記事の数は，World，Business，Opinion，Sports，Arts のカテゴリごとにそれぞれ 33，56，29，71，66 である．また，取得したユーザは合計で 13,266 人である．URL は，それぞれが属するカテゴリに応じて 5 次元のベクトルで表現される．例えば，World カテゴリの記事は $u = (1, 0, 0, 0, 0)$ となる．

これらのデータを用いてフォロー関係のネットワークと情報フローネットワークを作成し，ノードの次数に基づくランキングの比較を行う．フォロー関係のネットワークにおいては入次数 (どれだけフォローされているか)，情報フローネットワークにおいては出次数 (どれだけ情報を伝えたか) を用いる．ここで，次数を調べるために，情報フローネットワークのエッジに付与されたベクトルをスカラ値に変換する．本実験では単純なベクトルとの内積により計算することとする．ベクトル (1, 1, 1, 1, 1) との内積を用いた場合を *All*，ベクトル (1, 0, 0, 0, 0) との内積を用いた場合を *World* と呼ぶ．また，*Business*，*Opinion*，*Sports*，*Arts* も *World* と同様に，単位ベクトルを用いて計算する．

フォロー関係のネットワーク (*Follow*) と，6 種類の情報フローネットワークにおいて，次数に基づくノードのランキングを作成した．表 1 に，それぞれのランキング間の Spearman の順位相関係数を示す．*Follow* は *All* を除く全ての情報フローネットワークと低い相関にあることが分かる．フォロー関係の一部が実際の情報の流れを表していることを示唆している．さらに，*Sports* と *Arts* は他のいずれのネットワークとも相関を示さない．これは，Sports と Arts のカテゴリに属するニュースの伝播経路が，他カテゴリのニュースとは異なることを示しており，情報の種類によって伝播の経路が異なるという我々の仮説を支持している．

表 1: 各ネットワークの順位相関係数

	All	*World*	*Business*	*Opinion*	*Sports*	*Arts*
Follow	0.5907	0.2380	0.2895	0.3269	0.0809	0.1521
All		0.4237	0.4800	0.5401	0.2112	0.2850
World			-0.0054	-0.0634	0.0536	-0.0166
Business				-0.0621	0.0204	-0.0127
Opinion					-0.0385	-0.0544
Sports						0.0322

4. おわりに

本研究では，実際の情報拡散の経路と，その情報の質に基づいて生成される情報フローネットワークの提案を行った．さらに，小規模な実験によりその特性を調査した．今後は大規模なデータセットでの実験により詳細な調査を行うと共に，アプリケーション応用の可能性を検証する．

参考文献

[Bakshy 11] Bakshy, E., Hofman, J.M., Mason, W.A., and Watts, D.J., “Everyone’s an Influencer: Quantifying Influence on Twitter,” *Proc. of WSDM’11*, pp.65–74, 2011.

[Cha 10] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K.P., “Measuring User Influence in Twitter: The Million Follower Fallacy,” *Proc. of ICWSM’10*, pp.10–17, 2010.

[Katz 55] Katz, E. and Lazarsfeld, P.F., *Personal Influence: the Part Played by People in the Flow of Mass Communication*, Free Press, 1955.

[Kwak 10] Kwak, H., Lee, C., Park, H., and Moon, S., “What is Twitter, a Social Network or a News Media?,” *Proc. of WWW’10*, pp.591–600, 2010.

[Matsumura 08] Matsumura, N., Yamamoto, H., Tomozawa, D., “Finding Influencers and Consumer Insights in the Blogosphere,” *Proc. of ICWSM’08*, pp.76–83, 2008.

[Rogers 62] Rogers, E.M., *Diffusion of Innovations*, Free Press, 1962.

[Weng 10] Weng, J., Lim, E.-P., Jiang, J., and He, Q., “TwitterRank: Finding Topic-sensitive Influencial Twitterers,” *Proc. of WSDM’10*, pp.261–270, 2010.

[Wu 11] Wu, S., Hofman, J.M., Mason, W.A., and Watts, D.J., “Who Says What to Whom on Twitter,” *Proc. of WWW’11*, pp.705–714, 2011.