

ネットワーク構造による意見形成モデルの影響度分析

鈴木 優伽†

斉藤 和巳†

静岡県立大学経営情報イノベーション研究科

1 はじめに

ソーシャルメディアの登場により、Web空間内で大規模なソーシャルネットワークが形成されている。また、ソーシャルメディアの登場以前より、Web空間内の多様な電子ネットワーク上で情報の形成や拡散が行われている。それらネットワーク上のオピニオン形成や情報拡散過程のモデル化は、将来流行する話題の特定や、影響度の大きいノードの抽出を可能とする。そのため、ネットワーク上での意見形成 [1, 2] や情報拡散 [3, 4] の研究が盛んに行われている。また、従来の研究 [5, 6] において、意見形成や情報拡散過程は、対象のネットワーク構造に影響を受けるとされ、それらの関係が分析されている。

一般にネットワーク構造は、ノード間の繋がり（リンク）で表現されるため、ネットワーク全体そのものを単適に示していることが多い。しかし、ネットワーク全体ではなく、以下で提案する“類似影響クラス”の視点では、ネットワーク構造は異なる形に捉えられる。例えば、密なコミュニティをもつノード群、多くのノード間の橋渡しの位置に存在するノード群、ネットワーク構造の端に存在するノード群を“類似影響クラス”とすれば、ネットワーク構造は異なる視点で捉えられる。さらに、構造の違いは、ネットワーク上の意見形成過程に影響を与え、それらを多様に特徴づける。すなわち、意見形成過程における影響度の多様さを基に、類似影響クラスを構成すれば、従来とは異なる観点で、ネットワーク構造の分析が可能と考えられる。

よって本研究では、各ノードの意見形成の影響度を基にしたネットワーク構造の分析法を提案する。詳細には、基本的な意見形成モデル [1] から、あるノードが意見形成した際の際他ノードへの影響度を図る影響度曲線を導入する。その際、影響度曲線が類似するノード群を類似影響クラスとし、それらクラスの特徴からネットワーク構造を分析していく。

本稿の構成は以下とする。まず、提案するネットワーク構造分析法について説明する。次に、データセットを用いた評価実験の結果について詳細に述べる。最後に、実験結果から本分析法の有望性を示す。

2 提案分析法

本分析法の手順は以下の通りである。

- St1: 大域オピニオン採用付き VoterModel を用いて、意見形成過程におけるノードの影響度曲線を算出;
 - St2: 影響度曲線からノード群を K -medoids 法でクラス分け;
 - St3: クラス毎のネットワーク指標の Z -score を計算し、推測されるネットワーク構造と意見形成過程の関係を分析;
- 以降、各ステップの詳細な説明を述べる。

2.1 意見形成モデル

ノード集合 $\mathcal{V}$ とリンク集合 $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ からなる有向ネットワークを $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ とし、局所オピニオン採用でノード v に影響を与える親ノード集合を $C(v) = \{u \in \mathcal{V}; (u, v) \in \mathcal{E}\}$ と

する。VM では、各ノード v は、親ノード $u \in C(v)$ または任意のノード $u \in \mathcal{V}$ をランダムに 1 つ選択し、時刻 t でのノード u のオピニオンを時刻 $t+1$ での v のオピニオンとして採用することを繰り返す。本稿では、時刻 $t=0$ で、ノード v のみがオピニオン 1 を持ち、他のノード $u \in \mathcal{V} \setminus \{v\}$ はオピニオン 0 のとき、時刻 t の経過とともに、ネットワーク上でノード v の初期オピニオンの影響を受けるノード数の期待値を影響度曲線として分析する。

いま、時刻 t でノード u がノード v の初期オピニオンを採用する確率を $p_v(u; t)$ で表す。問題設定より、 $p_v(v; 0) = 1$ で、 $u \neq v$ なら $p_v(u; 1) = 0$ となり、確率 $p_v(u; t+1)$ は次式で求まる。

$$p_v(u; t+1) = \frac{\alpha}{|C(u)|} \sum_{w \in C(u)} p_v(w; t) + \frac{1-\alpha}{|\mathcal{V}|} \sum_{w \in \mathcal{V}} p_v(w; t)$$

ここで、 α は親ノードからの局所オピニオン採用確率を表し、 $1-\alpha$ は大域オピニオン採用確率となる。よって、時刻 t でノード v の初期オピニオンの影響を受けるノード数の期待値は $y_v(t) = \sum_{u \in \mathcal{V}} p_v(u; t)$ となる。本稿では、最終時刻 T までのオピニオン形成過程を考え、ステップ St1 としてベクトル $\mathbf{y}_v = (y_v(1), \dots, y_v(T))$ を影響度曲線として算出する。

2.2 K -medoids クラスタリング

ノード群を影響度曲線 $\mathbf{y}_v (v \in \mathcal{V})$ を基に K -medoids 法でクラスタリングする。 K -medoids クラスタリングは、ノード集合 $\mathcal{V}$ とその要素 $v, u \in \mathcal{V}$ 間の類似度 $\rho(u, v)$ が与えられたとき、目的関数 $J(\mathcal{P}) = \sum_{v \in \mathcal{V}} \max_{u \in \mathcal{P}} (\rho(v, u))$ を最大にするような代表ノード集合 $\mathcal{P}$ を求める。 $K = |\mathcal{P}|$ 個の代表ノード集合を抽出し、残りのノード群を最も類似する代表ノードのクラスに割り当てることで、ノード集合 $\mathcal{V}$ を K 個のクラス $\mathcal{V}^1, \mathcal{V}^2, \dots, \mathcal{V}^K$ に分割する。ここでは、類似度 $\rho(v, u)$ を次式で定義する。

$$\rho(v, u) = \frac{\sum_{t=1}^T \mathbf{y}_v(t) \mathbf{y}_u(t)}{\sqrt{\sum_{j=1}^T \mathbf{y}_v(t)^2} \sqrt{\sum_{j=1}^T \mathbf{y}_u(t)^2}}$$

この時、対象ネットワーク内において、密なコミュニティに属するノード、他のノード間の橋渡しの位置に存在するノード、中心から離れた位置に存在するノードでは、他ノードへ与える影響度が異なる傾向をもつ。よって、以降では $\mathcal{V}^1, \mathcal{V}^2, \dots, \mathcal{V}^K$ を類似影響クラスとする。

2.3 Z -score

得られた類似影響クラスを分析するために、 Z -score を導入する。ノード $v \in \mathcal{V}$ に対し、次数、クラス係数、近傍次数、近接度、媒介度を属性とし、 $D(v)$, $CC(v)$, $ND(v)$, $CL(v)$, $BW(v)$ とする。すると、全ノードでの属性値の平均 $\mu_D, \dots, \mu_{BW}$ と標準偏差 $\sigma_D, \dots, \sigma_{BW}$ が計算できる。一方、ノード集合のクラス分類結果を $\mathcal{V}^1, \mathcal{V}^2, \dots, \mathcal{V}^K$ とすれば、各 $\mathcal{V}^k$ での各 $\mu_D^1 \dots \mu_D^K$, $\sigma_D^1 \dots \sigma_D^K$ が求まる。このとき、各クラスにおける属性 $m \in \{D, CC, ND, CL, BW\}$ の Z -score を $z_m^k = \sqrt{|\mathcal{V}^k|}(\mu_m^k - \mu_m) / \sigma_m$ として算出する。

Influence Analysis of Voter Model Focusing on Network Structure

†Yuka SUZUKI †Kazumi SAITO

†Graduate School of Management and Information of Innovation, University of Shizuoka

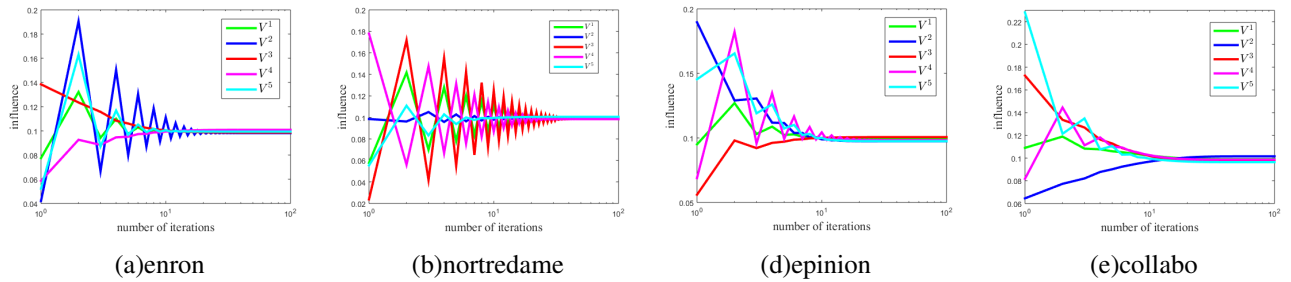


図 1: 影響度曲線の図

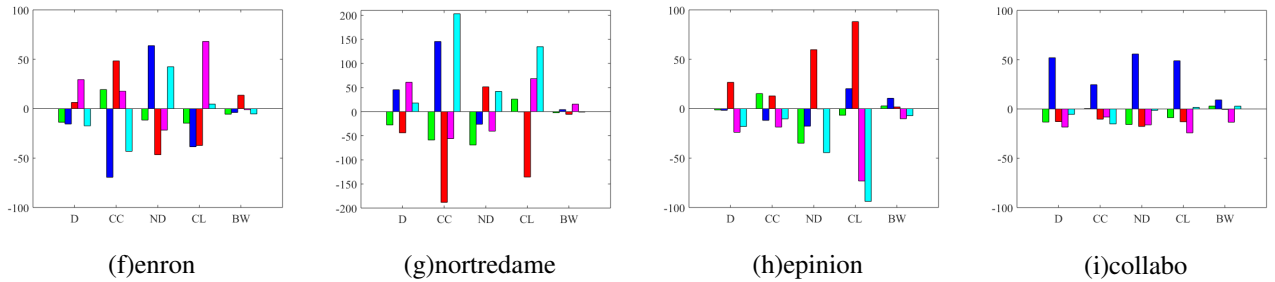


図 2: 各クラスターのネットワーク統計量

3 評価実験

スタンフォード大学が提供するデータセット*のうち、エンロンの Email ネットワーク、ノートルダム大学の Web サイトのハイパーリンクからなる Web ネットワーク、オンラインソーシャルサイトのエピニオン†のユーザ関係ネットワーク、論文の共著者ネットワークのデータを用いる。ノード数は 265214, 325729, 75879, 12008 であり、リンク数は 420045, 1497134, 508837, 118521 である。

3.1 影響度曲線に基づく分析

図 1 に、各ネットワークデータでの、類似影響クラスにおける代表ノード $u \in \mathcal{P}$ の影響度曲線を示す。ただし、代表ノードごとに異なる色で彩色する。図 1 から、影響度曲線は以下の 2 つに分類可能である。(a1) 影響度がステップごとに大きく変化する曲線。(a2) 比較的なだらかな曲線。図 1 から、(a)enron, (b)nortredame の情報ネットワークでは (a1) の傾向が強く、(c)epinions, (d)collabo のソーシャルネットワークでは (a2) の傾向が強いことが見て取れる。しかるに、(a1) の傾向が強い場合は、情報ネットワークのような少数の情報発信源的ノードと多数の情報受信的ノードから構成されるネットワーク構造、すなわち、リンクの受け渡し数に差異がある 2 部グラフ的ネットワーク構造であると考えられ、(a2) の傾向が強い場合は、共著者ネットワーク内で見られるような、比較的多数のノードが均等に繋がっているネットワーク構造であることが示唆される。

3.2 Z-score による抽出クラスの特徴分析

図 2 に、各属性の Z-score を類似影響クラスごと示す。彩色には、図 1 の代表ノードに用いた色と、同じ色を対応させていることに注意。図 1 と 2 から、(a1) の傾向が強いクラスは、クラス係数が低く近傍次数の値が高く、(a2) の傾向が強いクラスは、次数、近傍次数、近接度の値が高いことが分かる。すなわち、(a1) 傾向のクラスは、自身は周りとは密な関係を持たないが、友人が沢山のオピニオン発信先をもつよう

な橋渡しの構造であると考えられる。一方、(a2) 傾向のクラスは、ネットワークの中心性指標の値が特に高いため、ネットワーク全体の中心に位置していると考えられる。また両者は、情報ネットワークとソーシャルネットワークで共通の特徴である。これより本分析法が、ネットワークのタイプや大域的な構造に囚われない視点からの分析が可能であることが示唆される。

4 おわりに

本研究では、ネットワーク構造がオピニオン形成に影響を与えるという先行研究から、逆説的に、オピニオン形成過程からネットワーク構造を分析する方法を提案した。評価実験では、本分析法が一定の有用性を示す事を確認した。今後は、さらに多様なネットワークへ適応した提案法の有用性を評価する。

謝辞 本研究は、科研費 (No.16H02904) の補助を受けた。

参考文献

- [1] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social network. In *WINE 2007*, pp. 281–286, 2007.
- [2] F. Wu and B.A. Huberman. How public opinion forms. In *WINE 2008*, pp. 334–341, 2008.
- [3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influential through a social network. In *WINE 2007*, pp. 137–146, 2007.
- [4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Van-Briesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD 2007*, pp. 420–429, 2007.
- [5] K.R. Varshney. Bounded confidence opinion dynamics in social network of bayesian decision makers. arXiv, 2013.
- [6] F. Gargiulo and S. Huet. Opinion dynamics in a ground-based society. In *EPL*, Vol. 91, 2010.

*<https://snap.stanford.edu/data/index.html>

†<http://www.epinionglobal.com/>