

意見形成モデルの影響度曲線における類似ノード群の特性分析

鈴木 優伽[†] 齊藤 和巳[†]

[†] 静岡県立大学経営情報イノベーション研究科 〒422-8526 静岡県静岡市駿河区谷田 52-1

E-mail: †{j16105,k-saito}@u-shizuoka-ken.ac.jp

あらまし Web 空間内に存在する大規模なソーシャルネットワーク等において、オピニオンや情報が形成されるため、意見形成や情報拡散の研究がされている。我々はこれまでの研究で、生物学的知見を基にした意見形成過程のモデル化を試み、異なるネットワークごとの意見形成過程の特性を示した。しかしながら、どのようなネットワーク構造が、それらの特性を有するのか十分な検討はされていない。そのため、本研究では、より詳細な分析の第一歩として、各ノードの意見形成過程における他ノードへの影響度に着目する。具体的には、意見形成過程における影響度が類似したノード群の特性から、ネットワーク構造を分析し、ネットワーク構造と意見形成過程の関係性について示す。

キーワード 影響度曲線, 意見形成, 類似影響度クラス

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] [3] や情報拡散 [4] [5] の研究が盛んに行われている。また、従来研究 [6] [7] [8] 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。そのため、従来研究 [6] [7] [8] では、“ソーシャルネットワーク”等のネットワークの種類で分類される大域的なネットワーク構造と意見形成過程の分析がされている。我々も、これまでの研究において、生物学的知見を基にした意見形成過程のモデル化を試み、ネットワークの種類で分類した、大域的なネットワーク構造と意見形成過程の特性について示してきた。しかし、ネットワーク中には、密な関係のノード群の存在や、疎なノード群の存在など、大域的な視点では捉えられない構造が存在する。従来研究と同様、ネットワーク構造がノードの意見形成に影響を与えるとすれば、それら構造を考慮することでより詳細な分析が可能であると考えられる。

よって、本研究では“類似影響クラス”の視点を用いたネットワーク構造分析法を提案し、従来とは異なる視点で捉えたネットワーク構造と意見形成過程の関係について示す。以下に、“類似影響度クラス”の概要をしめす。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ノード群、ネットワークの端におり孤立傾向をもつノード群が存在すると仮定する。密なコミュニティに属するノードが意見形成を行えば、コミュニティのメンバーの多くに影響を及ぼすが、端に存在するノードでは、他のノードに与える影響は少ない。すなわち、各ノード

が意見形成を行った際に、他のノードにどう影響を与えるかは、自身の立ち位置（構造）によって多様に変化し、影響度が類似したノードは、類似した構造をもつと考えられる。そこで、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、他のノードへの影響度を図る影響度曲線を導入し、影響度曲線が類似したノード群を“類似影響クラス”とする。

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て、評価実験の結果を詳細に述べる。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 $\mathcal{C}(v) = \{u \in \mathcal{V}; (u, v) \in \mathcal{E}\}$ とする。VM では、各ノード v は、親ノード $u \in \mathcal{C}(v)$ 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

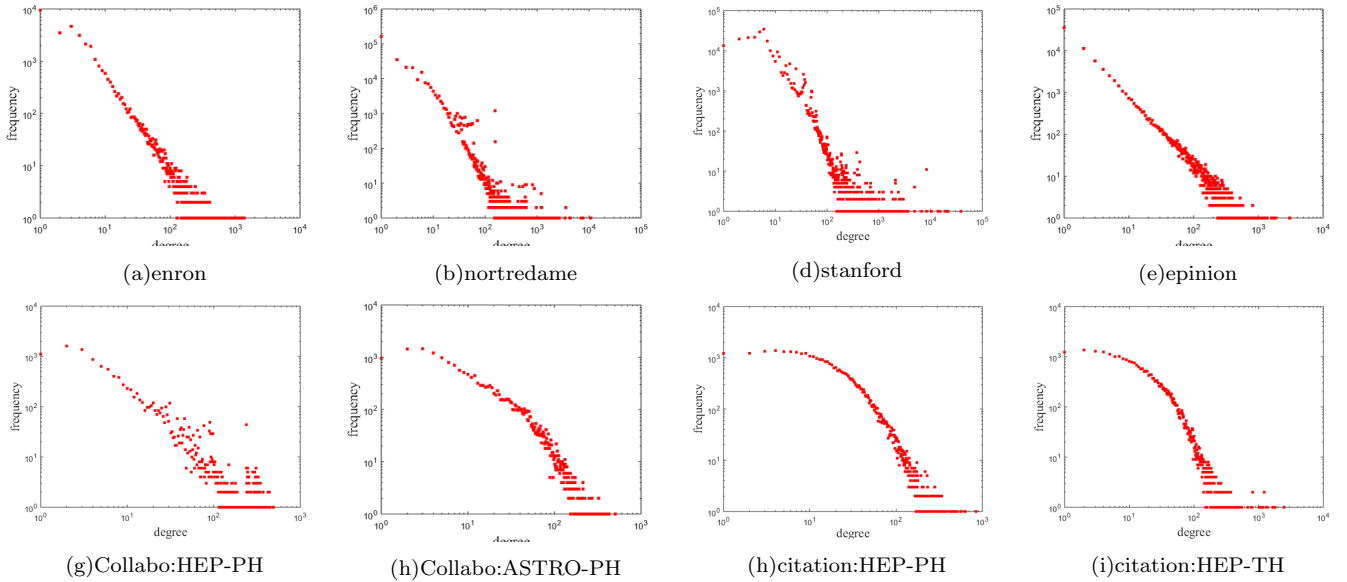


図 1 各ネットワークの度数分布

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、以下の目的関数を最大にするような代表ノード集合 \mathcal{P} を求める。

$$J(\mathcal{P}) = \sum_{v \in \mathcal{V}} \max_{u \in \mathcal{P}} \{\rho(v, u)\}$$

$K = |\mathcal{P}|$ 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。

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

また、ノード集合 \mathcal{V} の部分集合 $\mathcal{Q} \subset \mathcal{P} \subset \mathcal{V}$ に対し、目的関数 $J(\mathcal{P})$ は以下のサブモジュラ性 [9] を有する。

$$J(\mathcal{Q} \cup \{w\}) - F(\mathcal{Q}) > J(\mathcal{P} \cup \{w\}) - F(\mathcal{P})$$

$w \in \mathcal{P}$ であることに注意。サブモジュラ性を有する関数の最適化問題は、貪欲解法を用いることで最悪ケースの解品質を得られることが保証されている。[Leskovec] よってここでは、近似

解法として貪欲法を採用する。

貪欲法では、既に選定した代表ノード集合 \mathcal{P} を固定し、以下に示す目的関数を増加させるノード u を代表ノードとして求める。

$$J(\mathcal{P} \cup \{u\}) = \sum_{v \in \mathcal{V} \setminus \mathcal{P}} \max\{\gamma(v; \mathcal{P}), \rho(v, u)\}$$

ここで、 $\gamma(v; \mathcal{P})$ は既に選定された代表ノードとの類似度の最大値を示し、 $\gamma(v; \mathcal{P}) = \max_{w \in \mathcal{P}} (\rho(v, w))$ で算出される。以下に、貪欲法による K -modoids 法のアルゴリズムを示す。

- A1. $k = 1, \mathcal{P}_0 \leftarrow \emptyset, \gamma(v; \emptyset) \leftarrow 0$ で初期化;
- A2. $\hat{w}_k = \arg \max_{w \in \mathcal{V} \setminus \mathcal{P}_{k-1}} \{J(\mathcal{P}_{k-1} \cup \{w\})\}$ を求める;
- A3. $\mathcal{P}_k = \mathcal{P}_{k-1} \cup \{\hat{w}_k\}$ とする;
- A4. $k = K$ ならば \mathcal{P}_K を出力して終了;
- A5. 各ノード $v \in \mathcal{V}$ に対し、 $\gamma(v; \mathcal{P}_k)$ を求め、 $k = k + 1$ として A2 へ戻る;

そして各ノード v を最も類似する代表ノードのクラスに割り当てる。以降では割り当てたクラス $\mathcal{V}^1, \mathcal{V}^2, \dots, \mathcal{V}^K$ 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、クラスタ係数、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 μ_D, \dots, μ_{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 $\mu_D^k \dots \mu_{BW}^k, \sigma_D^k \dots \sigma_{BW}^k$ 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。

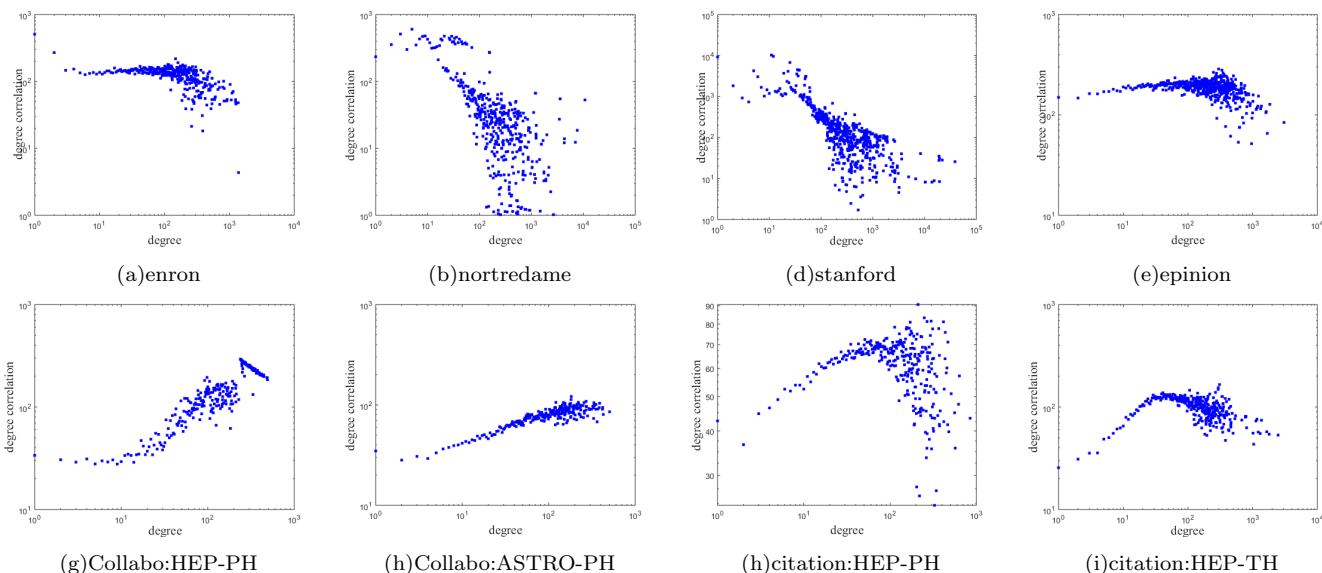


図2 各ネットワークの次数と近傍次数の分布

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^(注1)のうち、エンロンのEmailネットワーク、ノートルダム大学・スタンフォード大学の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^(注2)のユーザ関係ネットワーク、論文の共著者・論文引用データからなる研究者ネットワークを用いる。論文の共著者ネットワークデータは、High Energy Physics・AstroPhysicsのデータを用い、それぞれCollabo:HEP-PH・Collabo:ASTRO-PHと表記する。また、論文引用ネットワークデータは、High Energy Physics Phenomenology・High Energy Physics Theoryのデータを用い、それぞれcitation:HEP-PH・citation:HEP-THとする。ノード数は265214, 325729, 281903, 75879, 12008, 18772, 34546, 27770であり、リンク数は420045, 1497134, 281903, 508837, 118521, 198110, 421578, 352807である。

3.1 ネットワーク全体の特徴

対象ネットワークの全体的な特徴を明らかにするために、図1に、ネットワークごとの次数分布を示す。また図2に、x軸を自身の次数、y軸を近傍ノードの総次数を自身の次数で除した値とした分布を示す。図1から、どのネットワークにおいても次数分布はスケールフリー性を持つことが確認できる。図2をみると、自身の次数と近傍ノードの総次数は、右肩上がり・右肩下りの関係になることがわかる。特に、(b)nortredame, (c)stanfordでは右肩下がり、(g)Collabo:HTP-PH, (g)Collabo:HTP-PHでは右肩上がりの傾向が強いことがわかる。すなわち、右肩下りのネットワークでは、ノード同士の繋がり方に差がみられ、右肩上がりのネットワークでは、どのノード同士も比較的同程度に繋がっていると考えられる。

3.2 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 $w \in \mathcal{P}$ 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し曲折する曲線。(a2)比較的なだらかな曲線。図1から、(a)enron, (b)nortredame(c), stanfordの情報ネットワークでは(a1)の傾向が強く、(c)epinionsのソーシャルネットワーク、(d)(f)(g)(h)(i)の研究者ネットワークでは(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。

3.3 Z-scoreによる抽出クラスの特徴分析

図4に、各属性のZ-scoreを類似影響クラスごと示す。彩色には、図3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3と4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り的な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・ソーシャルネットワーク・研究者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。

(注1) : <https://snap.stanford.edu/data/index.html>

(注2) : <http://www.epinionglobal.com/>

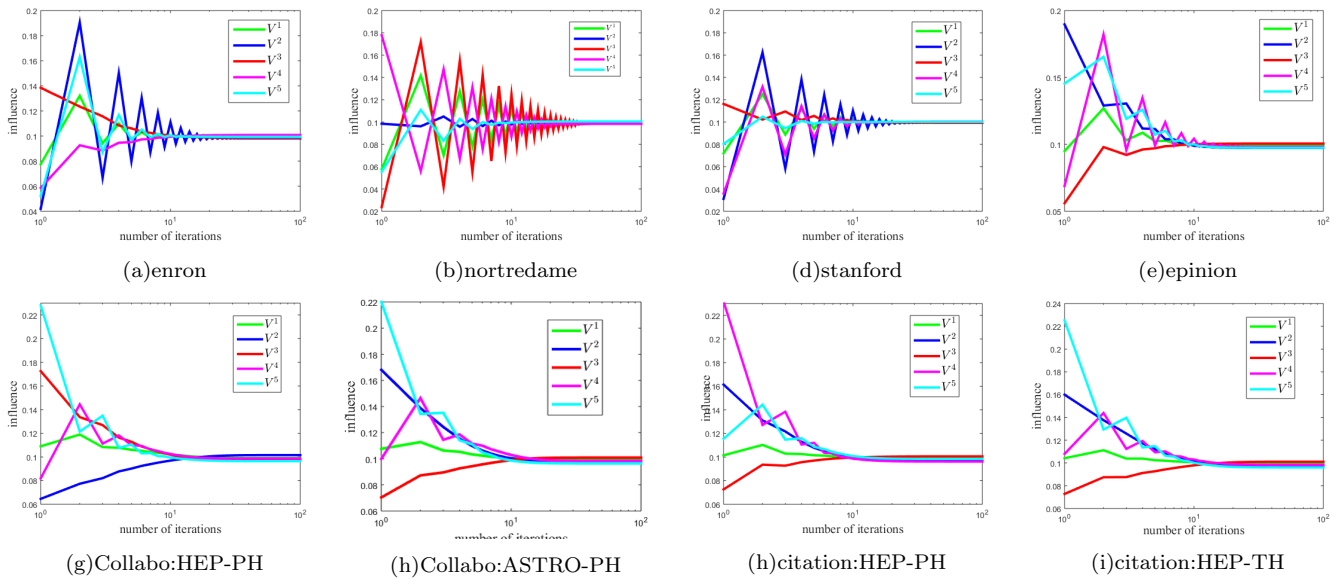


図3 類似影響度クラスごとの影響度曲線

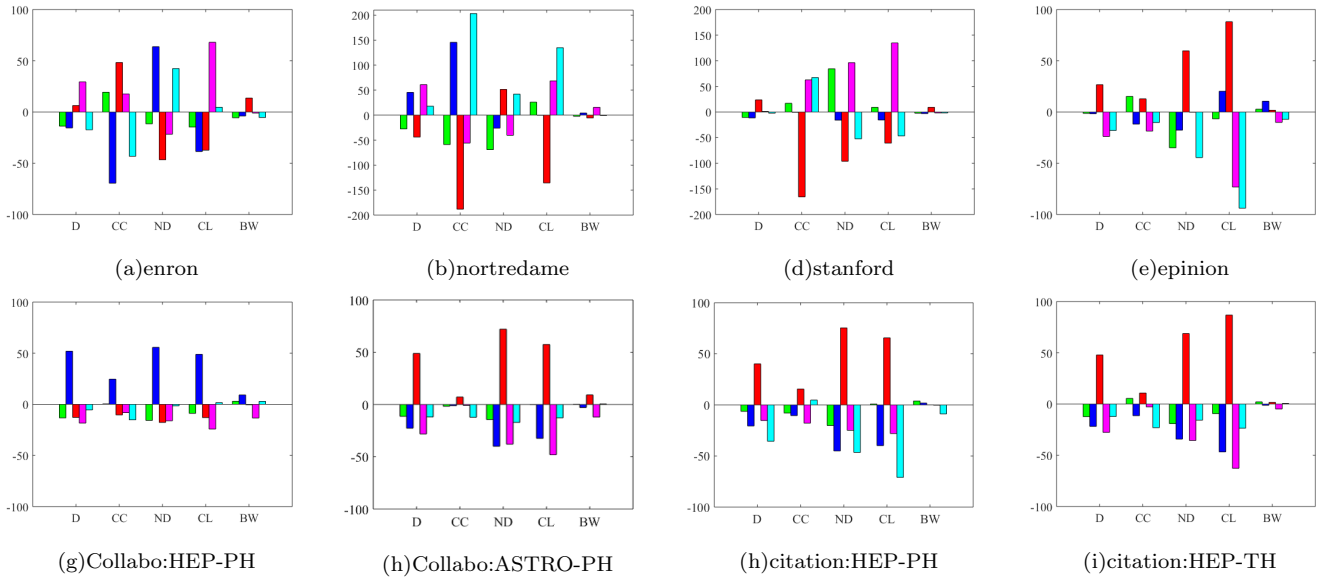


図4 類似影響度クラスごとのネットワーク統計量の Z-score

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、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、ネットワーク構造と意見形成過程の関係について分析することを試みた。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。

文献

[1] Even-Dar, E. and Shapira, A.: A note on maximizing the spread of influence in social network, *In WINE 2007*, pp. 281–286 (2007).
 [2] Wu, F. and Huberman, B.: How public opinion forms, *In WINE 2008*, pp. 334–341 (2008).
 [3] Zhen, W., Yi, L., Lin, W. and Yan, Z.: Freezing period

strongly impacts the emergence of a global consensus in the voter model, *Scientific Reports* (2013).
 [4] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *In WINE 2007*, pp. 137–146 (2007).
 [5] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. and Glance, N.: Cost-effective outbreak detection in networks, *In KDD 2007*, pp. 420–429 (2007).
 [6] Varshney, K.: Bounded confidence opinion dynamics in social network of bayesian decision makers, arXiv (2013).
 [7] Gargiulo, F. and Huet, S.: Opinion dynamics in a ground-based society, *EPL*, Vol. 91, No. 5 (2010).
 [8] Mahdi, J.: Social power and opinion formation in complex networks, *Physica A: Statistical mechanics and its applications*, Vol. 392, No. 4, pp. 959–966 (2013).
 [9] G. L. and Wolsey, L. A. and Fisher, M. L.: An analysis of approximations for maximizing submodular set functions, *Mathematical programming*, Vol. 14, pp. 265–294 (1978).