

ネーミングゲームを用いたコミュニティ検出法の 一考察

大阪府立大学大学院 工学研究科 手柳勇紀, 北條仁志
Yuki Teyanagi, Hitoshi Hohjo
Graduate School of Engineering,
Osaka Prefecture University

1 はじめに

実世界には、ソーシャルネットワーク、脳機能ネットワーク、タンパク質ネットワークなど、さまざまなネットワークが存在する。多くの場合、ネットワークは、同じような性質を持つノードの集まりであるコミュニティを持つ。このようなコミュニティの検出には、ノードクラスタリング、ネットワーク粗視化、ネットワーク機能分析などのさまざまな利点がある。ノードをクラスタリングすることにより、各ノードがどのクラスに属しているかを分析することができる。たとえば、マーケティングでは、同様の購入履歴を持つ顧客向けの自動推薦システムとして利用できる。ネットワークの粗視化により、検出されたコミュニティをネットワーク全体ではなく単一のネットワークとしてとらえ、さまざまな視点から分析することができる。ネットワークの機能分析が可能となることで、ネットワーク全体に対する特定のノードやコミュニティの影響を分析することができるようになる。したがって、コミュニティ構造の検出と分析は、さまざまな実世界のネットワークの構造を理解するための重要な方法論であり、社会的コミュニティのコンセンサス形成などのさまざまな問題に適用されてきた。

これまでにコミュニティの検出に関する多くの研究が行われてきた。しかしながら、従来の方法は、計算が複雑であったり、事前に設定しなければならないパラメータが多かったりなどの問題がある。

そこで本論文では、ゲーム理論ベースのネーミングゲームを用いることで、事前情報を必要としないシンプルなコミュニティ検出法を検討する。

2 ネーミングゲーム

ゲーム理論の分野では、個人の経済行動、意見の広がり进行分析していくためには、考える多くの戦略的状況下においてどのような収束プロセスを経るのかを観察することが必要になってくる。その経済行動分析手法の1つであるネーミングゲームは、エージェントベースモデルを用いて収束プロセスを観察するものである。これまでのネーミングゲームの研究では、その多くがエージェントダイナミクスに焦点を当てており、以下のような研究が行われてきた。Baronchelli[1]は話し手専用ネーミングゲーム (Speaker-Only Naming Game:SO-NG) や聞き手専用ネーミングゲーム (Hearer-Only Naming Game:HO-NG) を提案した。専用ネーミングゲームでは、話し手、または聞き手のみがメモリの更新を行う。Liら [9] は、聞き手が複数存在するネーミングゲーム (Naming Game with Multiple Hearer:NGMH) を提案した。ここでは、1人の話し手と複数の聞き手全てがコンセンサスに達したときに、初めてローカルコンセンサスとなるというルールをHO-NGモデルに追加した。Gaoら [6] が提案したグループ内でのネーミングゲーム (Naming Game in Groups:NGG) では、母集団をいくつかのグループに分割し、その各グループ内で全エージェントは話し手と聞き手の双方の役割を担う。上記のモデルではいずれも、関係するエージェントが多いほうが、収束スピードが速くなるという結果が現れた。Wangら [16] は有限メモリ言語ゲーム (Finite-Memory Language Games:FMLG) を提案し、ここではエージェントのメモリサイズを制限している。上記の研究によって、様々な特徴を持つ社会において、1つの意見を共有するプロセスが理論的にも実験的にも解明されてきている。また、ネーミングゲームに用いるネットワークに焦点を当てている研究も見られる。Baronchelliら [1, 4] はランダムグラフ、完全グラフ、スモールワールド、スケールフリーネットワークといったネットワークトポロジーで最小ネーミングゲームを適用した。この研究では、複雑ネットワークの違いが収束プロセスにどのように影響をもたらすかという点に着目している。

Louら [12]は、複数のコミュニティが存在するマルチローカルワールドネットワークにネーミングゲームを適用した。Louの研究によると、ネットワークのコミュニティ構造が顕著な場合には、グローバルなコンセンサスに到達することなく、コミュニティごとに異なる状態に収束する。つまり、ネーミングゲームはコミュニティ構造に大きく影響を受ける。

2.1 最小ネーミングゲーム

ネーミングゲームでは多くのモデルが研究されているが、その中でも最もシンプルなモデルを最小ネーミングゲームという。最小ネーミングゲームは、ネットワーク上に存在する複数のエージェントが、1つのオブジェクトに1つの名前を付けていくゲームであり、Baronchelli[2]らによって提案された。1回のゲームは以下の手順で行われる。

まずネットワークを生成し、そのノードはエージェントを表す。また2つのエージェントを接続するエッジは、そのエージェントたちがコミュニケーションを取れることを示している。そして1回のゲームの初めに、1人のエージェントがネットワークの中から話し手としてランダムに選ばれる。次に話し手と接続されているエージェントの中から1人、聞き手としてランダムに選ばれる。もし話し手がオブジェクトを表す名前を保持していなければ、メモリ内で新たに名前を生成し、それを聞き手に伝える。話し手が名前を保持していれば、メモリ内の名前からランダムに1つ選択し、聞き手に伝える。そして、聞き手は送られてきた名前を受け取ると、自分のメモリ内に同じ名前が存在するか検索する。もし存在しなければ、そのゲームは失敗となり、聞き手はその名前を自分のメモリに記憶する。もし存在していれば、ゲームは成功となり、話し手と聞き手は共にそれぞれのメモリ内にその名前を残し、他の名前はすべて消去する。

このペアワイズコミュニケーションでの成功をローカルコンセンサスと呼ぶ。また、ゲームを繰り返し行って最終的にすべてのエージェントが、オブジェクトに対し共通の名前を持っている状態に収束することをグローバルコンセンサスと呼ぶ。図1(a)には、ゲームが失敗し聞き手が名前を学習した場合の例を、図1(b)にはゲー

ムが成功し、話し手、聞き手の両者のメモリがただ1つの名前のみを保持している場合の例を示している。(a)の場合では「サル」が、(b)の場合では「キジ」が、話し手が聞き手に伝達した名前となっている。最小ネーミングゲームの収束性は理論的に証明されている。

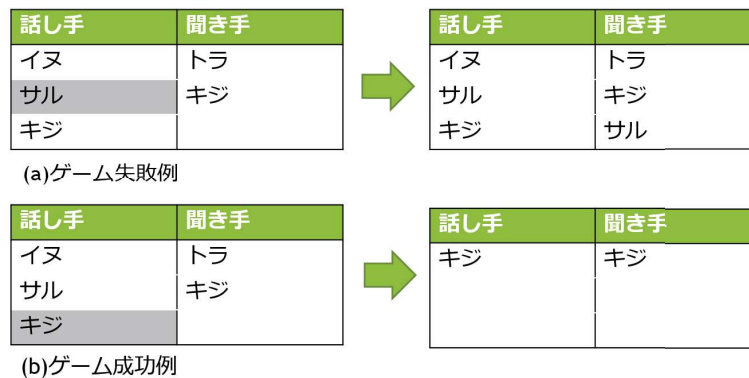


図 1: 最小ネーミングゲームの失敗例と成功例

3 コミュニティ検出

3.1 コミュニティの定義

コミュニティの定義は1つではない。本稿では、一般的に受け入れられているコミュニティの定義を採用する。その定義によると、コミュニティとは「グループ内のエッジの数がグループ外のノードと接続しているのエッジの数よりも多いノードのグループの集まり」である。この定義は以下のように定式化できる。

$$\sum_{i=1}^k |\{(u, v) | (u, v) \in E, u \in C_i, v \in C_i\}|$$

$$\gg \sum_{i, j} |\{(u, v) | (u, v) \in E, u \in C_i, v \in C_j, i \neq j\}|$$

ここで u, v はネットワークのノード、 E はエッジの集合、 C_i はコミュニティ i に属するノードの集合を表している。

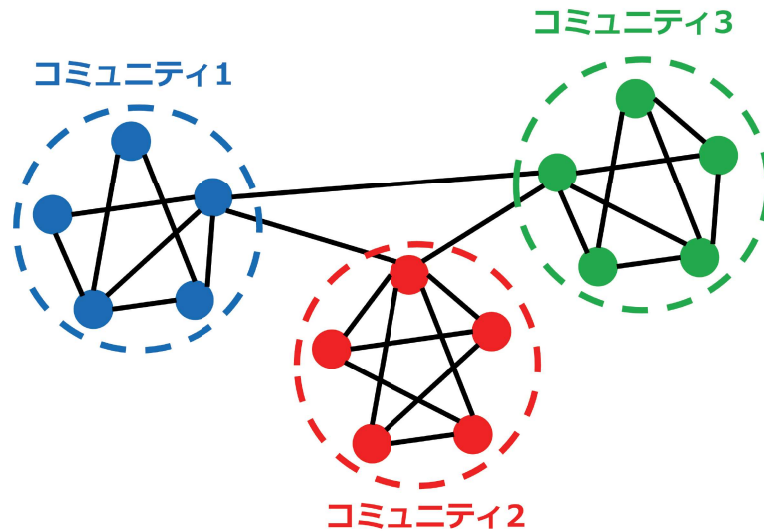


図 2: コミュニティの例

3.2 従来のコミュニティ検出法

コミュニティ検出法に関してはさまざまな研究が行われている。

Newman と Girvan[7] が提案した、Girvan-Newman アルゴリズムは、元のグラフから、「最も重要度の高い」エッジを徐々に削除することでコミュニティを検出する手法である。このアルゴリズムでは、まずグラフ全体を1つのコミュニティとして考える。次にすべてのエッジの中間中心性を計算し、最も中間中心性の高いエッジを切断する。エッジを切断することによって分割されたグラフをそれぞれのコミュニティと考え、同様の操作を繰り返す。そしてモジュール性が最も高くなる際のコミュニティ分割を採用する。Girvan-Newman アルゴリズムはコミュニティに関する事前情報を必要としないものの計算量が大きくなってしまいう課題があった。

この課題を解決するために開発された手法が Newman アルゴリズム [3] である。Newman アルゴリズムでは、まず個々のノードをそれぞれコミュニティとみなす。次にすべての2つのコミュニティの組み合わせに対し、それらのコミュニティを併合したときのモジュール性の増加量 ΔQ を計算し、 ΔQ の最も大きいコミュニティの組み合わせをマージする。マージされた2つのコミュニティに対してエッジを持つ他のコミュニティとの間の ΔQ を計算する。これをコミュニティが1つになるまで繰り返し、モジュール性が最も高くなる時のコミュニティ分類を採用する。

Newman アルゴリズムは、Girvan-Newman アルゴリズムと比べ計算時間の観点では向上したが、精度の点では同程度か少し劣っている。

ラベル伝播アルゴリズム [15] も、コミュニティ検出法としてよく知られている。ラベル伝播アルゴリズムは、多くの隣接ノードが所属しているコミュニティに自分も所属しようという考え方に基づいている。ラベル伝播アルゴリズムでは、まずネットワークの全ノードにそれぞれ異なるラベルを与える。ラベルを更新していくノードの順番をランダムに決定した後、その順番に従ってノードを選択する。選ばれたノードは、隣接するノードが持つラベルの中で最大多数を占めるラベルを自身のラベルとする。すべてのノードが隣接ノードの多数が占めるラベルを持つ状態になると終了し、同じラベルを持つノードのグループをコミュニティとみなす。この手法では、比較的短い計算時間でコミュニティを検出することができるが、ノードの選択順序をランダムに決定するため、得られる結果が毎回同じとは限らないという問題がある。

4 本研究のモデル

本研究では、ネーミングゲームを用いたコミュニティ検出法を提案する。本モデルは以下のように構成される。まず始めに、生成した実世界ネットワーク中のノードから、媒介中心性にしがって話し手を1人選ぶ。次に、話し手として選ばれたノードと隣接しているノード全てを聞き手とする。話し手と聞き手が選ばれると、話し手は自分のメモリ内に単語が存在するか確認する。単語がメモリ内に存在していれば、話し手はその中から1つ単語を選び、聞き手全員に伝える。話し手のメモリ内に単語が存在していなければ、話し手はあらかじめ設定した外部単語リストからランダムに1つ単語を選択し、その単語を自分のメモリ内に保存し、聞き手に伝達する。

聞き手は単語を受け取ると、その単語が自分のメモリ内の単語と一致するか調べていく。メモリ内に存在しない場合、ゲームは失敗となり聞き手は受け取った単語を自分のメモリに保存しゲームは終了となる。メモリ内に存在する場合、ゲームは成功となり、話し手と成功した聞き手のメモリ内から、一致した単語以外の単語を

消去し終了となる。またゲーム成功の際は、ローカルコンセンサスに達したノード間のエッジを切断する。

今回の実験では、この提案モデルを用いてシミュレーションを行っていく。

STEP1: ネットワークを生成し、各ノードの媒介中心性を計算し各エージェントの選択順序を決定する。さらに全エージェントのメモリを空の状態にする。

STEP2: ネットワークの中からエージェント $i(1 \leq i \leq n)$ を、選択順序に基づいて1人選び、話し手とする。 i と接続しているエージェント全てを聞き手として選ぶ。

STEP3: 選ばれた話し手 i がメモリ内に単語を保持していたら、その中からランダムに単語を1つ選択し聞き手全員に送る。 i がメモリ内に単語を保持していなければ、外部の単語リストから1つランダムに選び i のメモリ内に保存し、その単語を聞き手に送る。

STEP4: 聞き手はそれぞれ、送られてきた単語が自分のメモリ内の単語と一致するか確認する。もし一致していればゲームは成功となり、話し手 i と成功した聞き手はそれぞれメモリを更新し、一致した単語以外の単語をメモリ内から削除する。このとき、成功した聞き手の中からノードを1つ選び、そのノードと話し手 i とのエッジを切断する。もし一致していなければゲームは失敗となり、失敗した聞き手は自分のメモリ内に送られてきた単語を追加する。

STEP5: ネットワーク内のすべてのエッジが無くなるまで2~4を繰り返す。

アルゴリズム終了後に同じ単語を持つノードの集まりを1つのコミュニティとみなす。

5 数値実験

5.1 実世界ネットワーク

シミュレーションでは次のような実世界ネットワークを用いた。いずれもコミュニティ検出法の精度を評価する際に、よく利用されるネットワークである。表1は使用した各ネットワークのプロパティである。

5.1.1 ザカリーの空手クラブ

このデータセットは、34のノードと78のエッジで構成されている [17]。各ノードはアメリカの大学の空手クラブのメンバーを表しており、各エッジは空手クラブの外部でコミュニケーションをとっているかを表している。

5.1.2 イルカのコミュニケーションネットワーク

このデータセットでは、62のノードのそれぞれが、ニュージーランドの海域に生息するイルカを表している [13]。エッジは2頭以上のイルカが並んで観察されたことを示している。

5.1.3 アメリカンカレッジフットボール

このデータセットは、12の地区に所属するチームのアメリカンフットボールの試合のスケジュールである [7]。各ノードがチームを、各エッジが2チームが試合を行ったことを表している。

表 1: 各ネットワークのプロパティ

ネットワーク	ノード数	エッジ数	平均次数
空手クラブ	34	78	4.59
イルカ	62	318	5.13
アメリカンフットボール	115	1232	10.72

5.2 評価指標

コミュニティ検出アルゴリズムによって検出されたコミュニティの品質を評価する際に、一般的に用いられるのがモジュール性 [8] である。

モジュール性は検出されたコミュニティを $C_i, (i = 1, \dots, k)$ とし、 $e(C_i)$ 、 $a(C_i)$ をそれぞれコミュニティ内のエッジ、コミュニティ外のエッジとすると、モジュール性は以下のように定義できる。

$$Q = \sum_{C_i} e(C_i) - a(C_i)^2$$

5.3 結果

今回は、外部単語リストの単語数を 50 とした。また提案した手法によって得られた結果のモジュール性 Q の値を、既存手法である Girva-Newman (GN) アルゴリズム、Newman アルゴリズム、ラベル伝播アルゴリズム (LPA) と比較した。得られた結果を表 2 に示す。提案手法とラベル伝播アルゴリズムは 50 回の試行を行いその平均をとった値である。いずれのグラフに対しても、提案手法により得られた Q 値は既存手法よりも低いという結果になった。しかし、空手クラブネットワークでは、提案手法は平均値こそ、他手法を下回っているものの最大値は 0.35 と実際の Q 値に近い数値が得られている。

表 2: 各手法による Q 値の比較

ネットワーク	実際の値	提案手法	GN	Newman	LPA
空手クラブ	0.37	0.23	0.36	0.34	0.26
イルカ	0.38	0.29	0.39	0.40	0.49
アメリカンフットボール	0.55	0.28	0.40	0.40	0.58

6 おわりに

コミュニティ構造が顕著なネットワークではコミュニティ毎に収束しやすいという特徴を持つ NG を用いてコミュニティ検出を行った。ネットワークサイズが小さ

い空手クラブネットワークに対しては、実際の Q 値と近い値になることもあったがランダム性の影響を受けており、安定して高い精度を出すことができなかった。また本研究で使用したネットワークは、マルチローカルワールドネットワークほど顕著なコミュニティ構造を持たないため、エッジを切断してもグローバルなコンセンサスに近づいてしまった。今後の課題は、ネーミングゲームのランダム性を抑えつつより大規模なネットワークに対しても高い精度を保てるようにしていくことである。

参考文献

- [1] Baronchelli,A., Role of feedback and broadcasting in the naming game, Physical Review E 83, 046103 (2011) 1-6.
- [2] Baronchelli,A., M.Felici, V.Loreto, E.Caglioti, and L.Steels, Sharp transition towards shared vocabularies in multi-agent systems, Journal of Statistical Mechanics 2006, P06014 (2006) 1-12.
- [3] Clauset,.A, M.E.J.Newman, C.Moore, Finding community structure in very large networks, Physical Review E 70, 066111 (2004) 1-6.
- [4] Dall ' Asta,L., A.Baronchelli, A.Barrat, and V.Loreto, Physical Review E 74, 036105 (2006) 1-13.
- [5] Fu,G., Y.Cai, W.Zhang, Analysis of naming game over networks in the presence of memory loss, Physica A 479 (2017) 350-361.
- [6] Gao,Y., G.Chen, R.H.M.Chan, Naming game on networks: Let everyone be both speaker and hearer, Science Report 4, 6149 (2014) 1-9.
- [7] Girvan,M., M.E.J.Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 7821-7826.

- [8] Girvan,M., M.E.J.Newman, Finding and evaluating community structure in networks, *Physcal Review E* 69, 026113 (2004) 1-15.
- [9] Li,B., G.Chen, T.W.S.Chow, Naming game with multiple hearers, *Communications in Nonlinear Science and Numerical Simulation*18 (2013) 1214-1228.
- [10] Lou,Y., G.Chen, Analysis of the "naming game" with learning errors in communications, *Science Report* 5 12191 (2015) 1-12.
- [11] Lou,Y., G. Chen, J. Hu, Communicating with sentences: A multi-word naming game model, *Physica A* 490 (2018) 857-868.
- [12] Lou,Y., G. Chen, Z.Fan, L.Xiang, Local communities obstruct global consensus: Naming game on multi-local-world networks, *Physica A* 492 (2018) 1741-1752.
- [13] Lusseau,D., K.Schneider, O.J.Boisseau, P.Haase, E.Slooten, S.M.Dawson, The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations, *Behavioral Ecology and Sociobiology*, 54 (2003) 396-405.
- [14] Lu,Q., G.Korniss, B.K.Szymanski, The Naming Game in social networks: community formation and consensus engineering, *Journal of Economic Interaction and Coordination* 4, 221 (2009) 1-16.
- [15] Raghavan,U.N., R.Albert, S.Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Physcal Review E* 76, 036106 (2007) 1-11.
- [16] Wang,W.X., B.Y.Lin, C.L.Tang, G.R.Chen, Agreement dynamics of finite-memory language games on networks, *The European Physical Journal B* 60 (2007) 529-536.
- [17] Zachary,W.W., An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* 33 (1977), 452-473.