

# ネットワーク構造の推移性に着目した 局所的クラスタリング手法の提案

松島裕<sup>†</sup> 加藤竜丸<sup>‡</sup> 武藤敦子<sup>†</sup> 犬塚信博<sup>†</sup>

<sup>†</sup>名古屋工業大学大学院情報工学専攻

<sup>‡</sup>名古屋工業大学工学部情報工学科

1. はじめに 近年、現実ネットワーク内からより密な部分構造であるコミュニティを抽出する手法(クラスタリング手法)について関心が集まりつつある。特に、WWW など全体の構造情報を得ることが困難なネットワークに対しクラスタリングが可能な局所的な手法の研究が盛んに行われている。しかし、従来の局所的な手法は、同じネットワークに対してでもクラスタリング結果が異なるという問題があり、より安定性の高い手法が望まれている。

本稿では、コミュニティの内部構造に着目することで、より安定性の高い局所的なクラスタリング手法を提案する。第2章でネットワーク構造の推移性を用いた局所的なクラスタリング手法を提案する。第3章では、提案手法と既存手法を比較することで、抽出されたコミュニティ構造の特徴、クラスタリング結果の精度や安定性について比較、分析を行う。

2. 既存クラスタリング手法の特徴 クラスタリング手法には、ネットワーク構造全体からコミュニティを抽出する大局的な手法と、ある頂点から辺に沿ってコミュニティを探索していく局所的な手法がある。例えば、大局的な手法であるGN法[3]は媒介中心性の最も高い辺を順次取り除くことでできた連結成分をコミュニティとして出力している。そのため、ネットワーク全体の情報が必要であり、WWW など常に成長を続ける巨大なネットワークへの適応は困難である。

これに対し、局所的な手法ではある頂点から枝に沿ってコミュニティを広げていくため、ネットワーク全体のトポロジーを必要とせず、上記の問題を克服している。しかし、局所的Newman法[4]やL-Shell法[5]はコミュニティの内や外への辺の割合のみを基準として

コミュニティを広げているため、制約が弱く、また最初に選択した頂点によってコミュニティ構造が変化するという問題点がある。そのために、多くの研究では、全頂点に対しクラスタリング処理を用いた後に、すべての結果から最も尤もらしい結果を出力しているが、完全に局所的なクラスタリング手法とはいえない。

3. 提案手法 本章では、ネットワーク構造の推移性[1]に着目したクラスタリング手法を提案する。

3.1. ネットワーク推移性 ネットワーク推移性とは、“ある頂点の隣接頂点間にも辺がある割合が多い”という性質である[1]。推移性の強弱を図る尺度としてクラスタ係数を用いる。頂点 $v$ の隣接頂点集合を $N(v)$ 、ネットワークの隣接行列を $A$ とすると、 $v$ のクラスタ係数 $C(v)$ は次のように表すことができる。

$$C(v) = \frac{1}{|N(v)|C_2} \sum_{v_i \neq v_j \in N(v)} A(v_i, v_j) \quad (1)$$

本手法では、 $C(v)$ を頂点 $v$ をコミュニティに追加するかを決める評価関数として用いる。

3.2. アルゴリズム 頂点 $v$ の所属するコミュニティ( $v$ と同じコミュニティに所属する頂点の集合)を、図1の手順で抽出する。

$Com$ をコミュニティを形成する頂点集合(最初は $Com=\{v\}$ で初期化)とし、すべての外側隣接頂点( $Com$ に含まれない隣接頂点)の集合 $OutNeighbors(Com)$ を候補頂点集合 $Can$ に追加する。次に、すべての候補頂点 $v_i$ に対してクラスタ係数 $F(v_i, Com \cup Can)$ を計算し、閾値 $\alpha$ 以上ならば $Com$ に追加、 $Com$ の外側隣接頂点を $Can$ に追加する。この手順を $Com$ が更新されなくなるまで繰り返すことで $v$ の所属するコミュニティ集合 $Com$ を出力する。

図2は、名古屋工業大学の学科Aの友人ネットワーク[2]に対して、提案手法を用いて全頂点をクラスタリングした結果である。

Local network clustering method using transitivity  
Hiroshi Matsushima<sup>†</sup>, Ryumaru Kato<sup>‡</sup>, Atsuko Mutoh<sup>†</sup> and  
Nobuhiro Inuzuka<sup>†</sup>

<sup>†</sup>Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, 466-8555 Nagoya, Japan

<sup>‡</sup>Department of Computer Science, Faculty of Engineering, Nagoya Institute of Technology 466-8555, Nagoya, Japan

```

1. Com = {v} //頂点 v のみからなるコミュニティ
2. while(Com の更新がある)
3.   Can = Can ∪ OutNeighbors(Com)
4.   for(v_i ∈ Can) //閾値 α 以上の v_i を Com に追加
5.     if(F(v_i, Com ∪ Can) > α)
6.       Com = Com ∪ {v_i}
7.       Can = Can - {v_i}
8. return(Com)
    
```

$OutNeighbors(Com)$  :  $Com$  の外側隣接頂点の集合

$F(v, X)$  : 評価関数 ( $X$ :頂点集合)

$$F(v, X) = \frac{1}{|N(v_i) \cup X| C_2} \sum_{v_i \neq v_j \in (N(v) \cup X)} A(v_i, v_j)$$

図 1: 頂点  $v$  と同じコミュニティに所属する頂点集合  $Com$  を抽出する手順

4. 実験と考察 本章では、提案手法の特徴、優位性について検証するために、(1) 抽出されるコミュニティの詳細分析、(2) クラスタリング結果の精度、(3) クラスタリングの安定性について検証する。

4.1. コミュニティの詳細分析 図 2 と同様のネットワークに提案手法と GN 法 [3] を適用し、抽出されたコミュニティの密度とクラスタ係数 (コミュニティ内の各頂点のクラスタ係数の平均値) の平均値について調べた。抽出されたコミュニティの平均密度はともに 0.5 前後だが、平均クラスタ係数については GN 法は 0.56 であるのに対し、提案手法は 0.89 と明らかに高く、より推移性の高いコミュニティを抽出されている。

4.2. クラスタリングの精度 次の手順で人工的に生成したネットワークに対し提案手法と L-Shell 法 [5] を適用し、クラスタリング結果の精度を調べる。

1. 任意の個数のコミュニティを Holme-Kim モデル [6] を用いて生成
2. 各頂点間にランダムに辺を繋げる

Holme-Kim モデルとは、スケールフリー性 [1]、クラスタ性を満たすネットワークモデルである。

ランダムに繋げた辺の割合の変化させたときの精度の違いについて調べた結果が図 3 である。提案手法は L-Shell 法よりも正解率が高いことが分かる。また、提案手法はランダムな辺の割合をさらに増やしても 0.3 前後で推移し続けた。

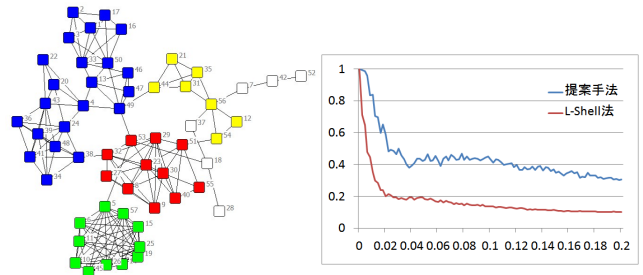


図 2: 提案手法適応結果 各頂点の色は所属する頂点を表す  
 図 3: クラスタリングの精度推移

4.3. クラスタリングの安定性 クラスタリング結果からコミュニティ行列  $A$  (頂点  $i$  と  $j$  が同じコミュニティに属しているならば  $A[i, j] = 1$ ) を複数個作成し、値の異なるセルの割合の平均を算出した。

その結果、L-Shell 法は 0.686 であったのに対し、提案手法は 0.837 だった。そのため、L-Shell 法に比べてより安定性が高いといえる。

5. おわりに 本稿では、ネットワーク構造の推移性に着目し、局所的にクラスタリングする手法を提案した。提案手法はクラスタ係数に基づいてコミュニティを探索することで、より推移性の高いコミュニティを抽出することができた。また、他の局所的な手法よりもより精度の高いコミュニティ構造を抽出できること、クラスタリング結果が安定していることが分かった。

今後の課題としては、アルゴリズムの効率性の改善、特殊な構造に対しても適用できるように改良することである。

参考文献

- [1] Wasserman, S., Faust, K. Social network analysis - Methods and applications, Cambridge University Press, 1994.
- [2] N.Inuzuka, T.Nakano, K.Shimomura. Friendship Analysis Using Attendance Records to University Lecture Classes. IASK International Conference Teaching and Learning, 478-486, 2008.
- [3] M.E.J.Newman, M.Girvan. Finding and evaluating community structure in networks. Phys. Rev. E, 2004
- [4] A. Clauset. Finding local community structure in networks. Phys. E, 2005.
- [5] Bagrow J p, Bollt E M. A Local Method for detecting communities. Phys. Rev. E, 2005.
- [6] Holme, P., Kim, B. J. Physical Review E, Vol. 65, article No.026107, 2002.