

動的変化を伴う複雑ネットワークに適した コミュニティ発見手法の提案

小林俊介[†]木村昌臣[‡]芝浦工業大学大学院工学研究科[†]芝浦工業大学工学部情報工学科[‡]

1. はじめに

近年、現実世界の事象における繋がりを表現し、それを分析する手段として複雑ネットワークの研究が盛んに行われている。コミュニティ構造分析もその一つであり様々な手法が提案されてきたが、そのほとんどはある時点でのネットワークに対してコミュニティ構造を同定するものであった。現実におけるネットワークはその構造を時間とともに変化させることがあり、それを想定していない既存手法を用いると計算量が膨大なものになってしまう。これを解決するためにより計算量を減らすことができ、またオンラインな構造同定法を目指すための手段としてラベル伝播による方法論が提案されている [1]。しかし、この方法ではコミュニティを結合する方向にしか伝播を行えないため、ある時点で一つのコミュニティを複数に分けたほうが適切な場合であっても、分割が行えない欠点が存在する。そこで本研究はこの問題を解決できる手法の提案を行う。

2. Label Propagation Algorithm (LPA)

ラベル伝播によるコミュニティ構造同定法 (LPA) では、全ノードがそれぞれ別のラベルを持った初期状態から、個々のノードが近傍のラベルをある基準に基づき選択し、自身のラベルにするというステップを繰り返す。これにより図 1 の様なクリークにおいてノードのラベルが一致しやすいため、密なつながりを持つ部分でのラベルの流行によってネットワーク全体のコミュニティ構造が発見される。この手法の最大の長所はラベル決定の基準に局所量を用いることでラベル決定を近傍ノードの情報だけで行うことができる点である。大規模ネットワークに適したコミュニティ発見手法として知られる Clauset 等の手法 [2] が大域的な最適化指標を用いるのに対して、LPA はより少ない計算量ですむことになる。さらに大域的な指標を用いる場合、個々のノードのラベルの組み合わせで評価値が決まることから、2 つ以上のラベルを同時に変えてしまうと解

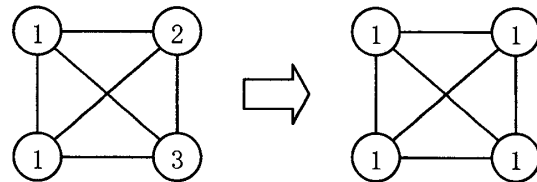


図 1 コミュニティラベルの流行

は収束しないが、LPA ではノードそれぞれのラベルを非同期で変更した場合でも、解が収束するという特徴がある。これによりノード毎に並列処理を行うと最大で計算時間を $1/(\text{ノード数})$ まで削減できるという利点がある。

3. 提案手法

本研究の対象は動的変化を伴うネットワークである。これに対しある時点の解 (コミュニティ構造) を求める際、前の時点の解を有効活用する「連続的」な方法により効率的なコミュニティ同定ができるのではないかと我々は考えた。ネットワーク構造の変化はコミュニティ構造に影響を与えるが、その範囲は局所的なものであり、繋がりの変化したノード周辺のみとなる場合が多いと予測できる。そこで、局所パラメータだけでコミュニティ同定を行える LPA をもとにしたネットワーク構造の変化に追従してコミュニティを発見する手法を本研究では提案する。

LPA の解は伝播の初期段階における伝播順序に強く依存する特徴があるため、時点毎に全ラベルが異なる状態から始める既存の LPA は良い解を得るために複数回の試行が必要となる。これに対し本研究が提案する方法 (連続的手法) では、初期時点で複数回の試行によって良い解を求めておけば、それ以降の解探索は良い状態から再伝播ステップを開始できるので各ステップにおける試行回数の削減が可能となる。

先述の通り LPA では伝播プロセスにおいて強いラベルが弱いラベルを淘汰してゆくが、ここで問題となるのは淘汰されたラベルが二度と出現しないことである。つまりある時点のネットワークに対し次の時点でエッジが増加・減少し、かつその変化により前の時点で存在したコミュニティを分割したほうが解として適切である場合であっても、コミュニティ数が増えることがないため良い解が求まらないことがある。これを解決するために本手法ではネットワ

A proposal of a community detection technique for dynamic complex networks

[†]Shunsuke Kobayashi [‡]Masaomi Kimura

[†]Graduate School of Engineering, Shibaura Institute of Technology

[‡]Shibaura Institute of Technology

ーク変化後, 出現したエッジの両端に共通の新規ラベル, 消滅したエッジの両端に別々の新規ラベルを割り当てる. これはノードが出現・消滅した場合でも, それに伴うエッジの増減が発生するため同様に扱うことが可能である.

また, 前時点の解を利用した連続的な解探索を行う際のラベル選択基準はネットワーク構造の変化に影響を受けないよう, ネットワーク構造の傾向に依存しないものが良いと考えられる. これを踏まえ本手法では以下の式でノード i のラベル L_i' を決定する.

$$L_i' = \operatorname{argmax}_L \sum_{j \in N_i, L_j=L} \left(1 - \frac{k_i \cdot k_j}{2m}\right) \cdot w_{ij} \quad (1)$$

ここで N_i はノード i の近傍集合, k は次数, w はエッジの重みを表す. この式は基本的に近傍ノードの持つラベルの中で最も数が多い, 流行しているラベルを選択することを表す. また $k_i \cdot k_j / 2m$ 部分により初期段階において近傍のラベルが全て異なる場合などは i 以外の近傍が少ないノード, つまり i のみと繋がり強いノードのラベルを優先的に選択するものとなる. なおこの値は 1 に対して小さなものであるため伝播初期段階にのみ大きな影響を与える. アルゴリズムの流れは以下ようになる.

- (i) ネットワーク中の全ノードに対して順にそのノードの近傍のもつラベルから式 (1) を用いて新規ラベルを決定
- (ii) (i) のステップをコミュニティ構造の変化がなくなるまで繰り返した結果をその時点の解とする
- (iii) ネットワーク変化後, 変化したエッジの両端のノードが持つラベルを新規ラベルに変更し, (i) に戻る

4. 実験・結果

本手法により動的ネットワークにおける連続的なコミュニティ構造の発見が行えるかを検証するため, 本手法 (連続 LPA) と式 (1) を用いたステップ毎にラベルを初期化する LPA (非連続 LPA), 大域的最適化による手法 [3] (RB 法) を比較した. 対象となるネットワークは CNN モデルを用いて生成した. このモデルはあるノードの近傍間に優先してエッジを張ることでクラスタ性の高いネットワークを作ることができる. 成長ステップでは 1 ステップ毎に, 1 つの新規ノードおよびそれと既存ノードを結ぶエッジ 1 つが, または既存ノード間にエッジ 1 つが出現するものとした. なおエッジは重みなし無向とした. 連続 LPA では初回適用時のみ良い解を得るために 500 回の試行を行う方法 (連続 LPA1) と, RB 法の結果を使用する方法 (連続 LPA2) をとった. 連続的は適用を各

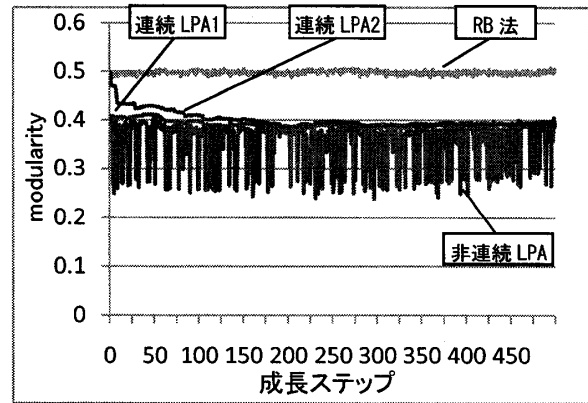


図 2 適用結果

ステップ 1 回ずつ行い, 非連続 LPA も同様に各ステップ 1 回ずつ適用した. 図 2 は 1000 ステップの成長をさせた状態を初期のネットワークとし, さらに 500 ステップ成長させる過程に対しコミュニティ同定を行った結果である. このグラフの通り, 非連続 LPA は結果の良さが大きく上下するのに対して, 連続 LPA1 は概ね上下が少なく高い水準を保っている.

modularity 最適化に基づく RB 法の最適解と, 本手法で得られる解は異なっていることが図 2 から見てとれるが, modularity 最適解から出発した連続 LPA2 の結果が成長ステップを追って漸的に連続 LPA1 の結果に近付いていることから, 連続 LPA の解は RB 法と異なるコミュニティ分割を与えているものと考えられ, それが RB 法の結果と連続 LPA1 の modularity の差として現れていると思われる. 実際, RB 法の解と本手法の解を比較するとコミュニティの平均的な大きさが異なる等の違いが見られる.

5. まとめ

実験からノード, エッジが増加方向に成長するネットワークに対して本手法によるコミュニティ同定が可能であることがわかった. また modularity には解像限界と呼ばれる問題が存在しており, 本研究で分かった modularity の差との関係について考察する必要がある. 加えて実在するネットワークに対しての実験も行いたい.

参考文献

- [1] U. N. Raghavan, R. Albert, S. Kumara: Near linear time algorithm to detect community structures in large-scale networks Phys. Rev. E 76, 036106 (2007).
- [2] A. Clauset, M. E. J. Newman, C. Moore: Finding community structure in very large networks, Phys. Rev. E 70, 066111 (2004).
- [3] J. Reichardt, S. Bornholdt: Statistical mechanics of community detection, Phys. Rev. E 74, 016110 (2006).