

マルチスライスネットワークにおけるコミュニティ抽出法の改良

Improvement of Community Detection in Multislice Networks

名部井 康博

Yasuhiro Nabei

村田 剛志

Tsuyoshi Murata

東京工業大学大学院 情報理工学研究科 計算工学専攻

Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

Community detection is one of the methods for network analysis. It is used for simplifying networks that are easy to understand. Temporal networks and networks with multiple types of links can be expressed by multislice networks which are combinations of individual networks. A method for detecting communities in multislice networks was proposed by Mucha. However, Mucha's method cannot be used for multislice networks of arbitrary coupling because this method considers only two special couplings. In this paper, we propose and discuss a new method for detecting communities in multislice networks of arbitrary coupling. Proposed method is the same as Mucha's one for the multislice networks of the two special couplings. By executing community detections, we confirm that proposed method is superior to Mucha's one for the multislice networks of arbitrary coupling.

1. はじめに

ネットワーク解析の一つにコミュニティ抽出がある。コミュニティ抽出とは、ネットワークからエッジの繋がりが密な部分ネットワークを見つけることであり、ネットワークの可視化などに用いられる。本稿では、マルチスライスネットワーク [Mikko 14] におけるコミュニティ抽出をあつかう。マルチスライスネットワークとは、異なるネットワーク上のノード同士をエッジで結ぶことで、複数のネットワークを組み合わせたネットワークのことである。たとえば、動的ネットワークや複数種類のエッジを持つネットワークなどはマルチスライスネットワークで表すことができる。

よく用いられるコミュニティ抽出手法として、モジュラリティを用いる手法がある。モジュラリティとは、抽出したコミュニティの質を測る評価関数で、コミュニティ内部のエッジが密であるほど、またコミュニティ間のエッジが疎であるほど値が大きくなる。この手法では、モジュラリティを最適化することでコミュニティ抽出をおこなう。マルチスライスネットワークにおけるコミュニティ抽出手法として、Mucha らが作ったマルチスライスネットワークのモジュラリティ [Mucha 10] を最適化することでコミュニティ抽出をおこなう手法が提案されている。

しかし、Mucha のモジュラリティは二つの特殊なカップリングで表されるマルチスライスネットワークを想定している。そのため、任意のカップリングで表されるマルチスライスネットワークでコミュニティ抽出をおこなうと、多くの場合にネットワーク間のエッジが疎なコミュニティも抽出してしまう。そこで本稿では、任意のカップリングのマルチスライスネットワークでネットワーク間のエッジが密なコミュニティが抽出できるように、Mucha のモジュラリティを改良したモジュラリティを提案する。

提案モジュラリティは、Mucha のモジュラリティが想定している二つのカップリングのマルチスライスネットワークの場合、Mucha のモジュラリティと同じ働きをする。実験では、そ

の二つ以外のカップリングのマルチスライスネットワークでコミュニティ抽出をおこない、提案モジュラリティの方がネットワーク間のエッジが密なコミュニティを抽出できることを確認した。

2. 関連研究

本節では、マルチスライスネットワークとマルチスライスネットワークにおけるコミュニティ抽出の従来手法について説明する。

2.1 マルチスライスネットワーク

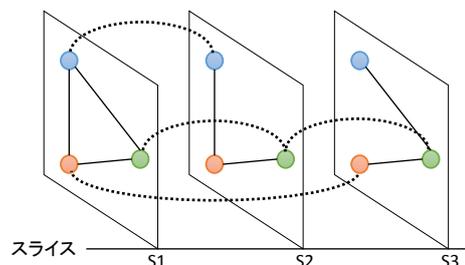


図 1: 3つのスライスからなるマルチスライスネットワーク。破線はスライス間エッジ。

マルチスライスネットワークとは、図 1 のように複数のネットワークを組み合わせたネットワークである。マルチスライスネットワークを構成するネットワークをスライスと呼ぶ。スライス間エッジは異なるスライス上の同じノード同士の間のみ存在する。たとえば、図 1 において S1 上の青のノードと S2 上の赤のノードとをエッジで結ぶことはできない。

2.2 カップリング

カップリングとはスライス間エッジの集合のことである。実データをマルチスライスネットワークで表す場合、各スライス上のエッジリストは与えられているが、スライス間エッジのエ

連絡先: 名部井康博, 東京工業大学大学院 情報理工学研究科 計算工学専攻, 東京都目黒区大岡山 2-12-1 W8-59, y-nabe@ai.cs.titech.ac.jp

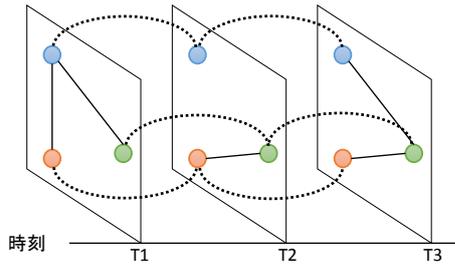


図 2: 動的ネットワークを, Ordinal カップリングを使って表わすマルチスライスネットワーク.

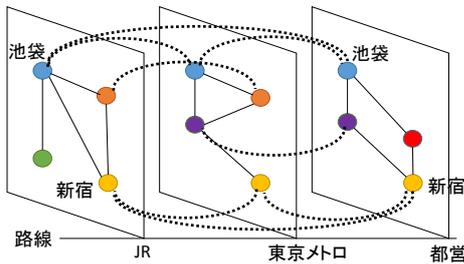


図 3: 複数の路線のネットワークを, Categorical カップリングを使って表わすマルチスライスネットワーク. 駅をノードとして, 各路線における隣接駅をエッジで結ぶ.

ジリストは与えられていないことがほとんどである. つまり, 扱う実データに対して適切なカップリングを自分で定義しなければならない. 代表的なカップリングとして Ordinal カップリングと Categorical カップリング [Mikko 14] がある. Ordinal カップリングでは, 隣合うスライス上の同じノード同士を全て, 重みの同じエッジで結ぶ. Categorical カップリングでは, 各スライス上の同じノード同士を全て, 重みの同じエッジで結ぶ. 図 2 は動的ネットワークを Ordinal カップリングを使ってマルチスライスネットワークで表した例であり, 図 3 は複数種類のエッジを持つネットワークを Categorical カップリングを使ってマルチスライスネットワークで表した例である.

この二つのカップリングがよく用いられるが, 実データをこの二つのカップリング以外で表したい場合もある. たとえば路線のネットワークの場合, 図 3 のように Categorical カップリングによって, 同じ駅を全てエッジで結ぶのではなく, 東京メトロ副都心線渋谷駅と東急東横線渋谷駅のように直通運転で乗り換えの必要がない場合のみエッジで結ぶことや, 乗り換えにかかる時間によってスライス間エッジの重さを変えることも考えられる.

2.3 Mucha のモジュラリティ

Mucha のモジュラリティはマルチスライスネットワークにおけるモジュラリティであり, 以下の式で表される.

$$Q_{mucha} = \frac{1}{2\mu} \sum_{ijsr} [(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s})\delta_{sr} + \delta_{ij}C_{jsr}]\delta(g_{is}, g_{jr}) \quad (1)$$

ここで, 添え字 i, j はノードを, s, r はスライスを表している. μ, m_s, k_{is} はそれぞれ, 全エッジの重みの合計, スライス s 上のエッジの重みの合計, スライス s 上のノード i から出るスライス内のエッジの重みの合計である. δ_{ij} は $i=j$ のとき 1 となり, $i \neq j$ のとき 0 となる. A は隣接行列で, A_{ijs} はスライス s における A の (i, j) 成分を表す. γ_s はパラメータで, スライス s 上でコミュニティの大きさを調節する. γ を大きくすると得られるコミュニティは小さくなり, γ を小さくすると得られるコミュニティは大きくなる. C はスライス間のエッジを表す行列で, C_{jsr} はノード j についての C の (s, r) 成分である. C の全ての要素は 0 か ω のいずれかの値をとる. ω はパラメータで, スライス間のエッジの重さを表す. ω を大きくすると複数のスライスをまたがるコミュニティがでやすくなる.

2.4 Gen Louvain

モジュラリティを用いるコミュニティ抽出では, モジュラリティを最適化することでコミュニティ抽出をおこなう. Gen Louvain [GenLouvain] は, Louvain 法 [Blondel 08] をもとにした, マルチスライスネットワークのモジュラリティの最適化を実現する MATLAB コードである. 以下のような手順で最適化をおこなう.

1. 各ノードをそれぞれ一つのコミュニティとして初期化する.
2. 各コミュニティについて, それぞれの隣接コミュニティとマージした場合のモジュラリティの最大値を求める.
3. 各コミュニティについて, 最大値がもとのモジュラリティ値より高ければ, 最大値を与える隣接コミュニティとマージする.
4. モジュラリティの値が増加しなくなるまで 2 と 3 を繰り返す.

3. Mucha のモジュラリティの改良

本節では, Mucha のモジュラリティの問題点と, その問題点を改良した提案モジュラリティについて説明する.

3.1 Mucha のモジュラリティの問題点

Mucha のモジュラリティの問題点は, スライス間エッジが疎になるコミュニティも抽出してしまうことである. Mucha のモジュラリティ式 (1) において

$$(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s})\delta_{sr} \quad (2)$$

$$\delta_{ij}C_{jsr} \quad (3)$$

項 (2) はスライス内に関する評価, 項 (3) はスライス間に関する評価である. 項 (2) は, Newman と Girvan によって提案されたモジュラリティ [Newman 04] と同じ評価法で, エッジの重みから, ノードの次数を保持したままエッジをランダムに張り替えた場合の重みの期待値を引いている. こうすることで, コミュニティ内部のエッジが密であり, コミュニティ間のエッジが疎になるようなコミュニティが抽出されることが知られている. 一方で項 (3) はスライス間のエッジの存在自体をプラスに評価する. 図 4 のマルチスライスネットワークでは, S1 上の 4 つのノードが属するコミュニティと他のすべてのノードが属するコミュニティとの二つのコミュニティが抽出されることが期

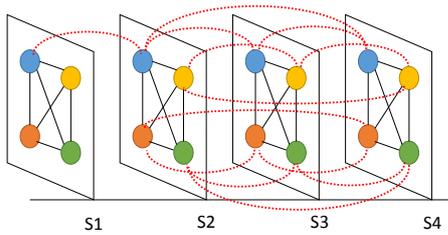


図 4: Mucha のモジュラリティと提案モジュラリティとで、コミュニティ抽出結果が異なるマルチスライスネットワークの例.

待される。しかし、Mucha のモジュラリティを用いたコミュニティ抽出では、すべてのノードを含んだ一つのコミュニティを抽出してしまう。なぜなら、項 (3) はスライス S1 の青のノードとスライス S2 の青のノードとを結んだエッジの存在自体をプラスに評価するので、コミュニティ抽出の際にマージが起こるからである。

図 2 や図 3 のような、Ordinal カップリングか Categorical カップリングで表されるマルチスライスネットワークの場合、Mucha のモジュラリティを用いても、スライス間エッジが疎なコミュニティは抽出されない。なぜならこの二つのカップリングは、すべてのノードについてスライス間エッジが密に繋がっているからである。一方、この二つ以外のカップリングで表されるマルチスライスネットワークでは、図 4 での例のようにスライス間エッジが疎なコミュニティも抽出してしまう。

3.2 提案モジュラリティ

提案モジュラリティは以下の式で表される。

$$Q_{proposed} = \frac{1}{2\mu} \sum_{ijsr} [(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s})\delta_{sr} + (C_{jsr} - \Gamma_j \frac{c_{js}c_{jr}}{2M_j})\delta_{ij}]\delta(g_{is}, g_{jr}) \quad (4)$$

M_j , c_{js} はそれぞれ、ノード j の全てのスライス間エッジの重みの総和、スライス s 上のノード j から出るスライス間エッジの重みの総和を表す。 Γ_j はパラメータで、ノード n に関してコミュニティの大きさを調節する。 Γ を大きくすると得られるコミュニティが小さくなり、 Γ を小さくすると得られるコミュニティが大きくなる。 C は隣接行列で C_{jsr} はノード j における C の (s, r) を表す。その他の項は Mucha のモジュラリティのものと同じである。

提案モジュラリティは Mucha のモジュラリティのスライス間に関する項 (3) を次のように改良している。

$$C_{jsr} - \Gamma_j \frac{c_{js}c_{jr}}{2M_j} \quad (5)$$

スライス間に関する評価もスライス内に関する評価と同じように、エッジの重みから、ノードの次数を保持してエッジをランダムにつなぎ直した場合の重みの期待値を引いている。こうすることで、スライス間エッジも密になるようにコミュニティを抽出することができる。

提案モジュラリティは Ordinal カップリングか Categorical カップリングのマルチスライスネットワークでは、Mucha のモジュラリティと同じ働きをする。スライス数が N 個、スライ

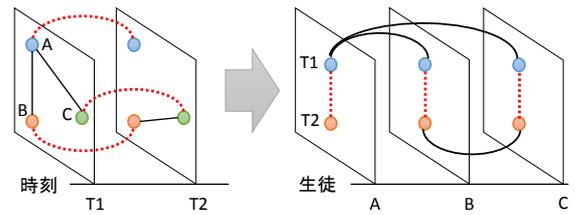


図 5: 左は Ordinal カップリングを使って highschool をマルチスライスネットワークで表した例。右は、左のネットワークのノードとスライスを入れ替えた例。

ス間エッジの重さが 1、カップリングが Categorical のとき、項 (5) は以下ようになる。

$$\begin{aligned} C_{jsr} - \Gamma_j \frac{c_{js}c_{jr}}{2M_j} &= 1 - \Gamma_j \frac{(N-1)(N-1)}{2 \cdot N C_2} \\ &= 1 - \Gamma_j \frac{N-1}{N} \end{aligned} \quad (6)$$

Γ を定数とすれば項 (6) は定数になる。よってカップリングが Categorical の場合、提案モジュラリティは Mucha のモジュラリティにおいて $\omega = 1 - \Gamma_j(N-1)/N$ とすることと変わらない。カップリングが Ordinal の場合も同様である。この二つ以外のカップリングでは、コミュニティ内部でスライス間エッジが密であり、コミュニティ外部とスライス間エッジが疎になるように評価するので、Mucha のモジュラリティと異なる。提案モジュラリティを用いて図 4 のネットワークでコミュニティ抽出をおこなうと、期待通りスライス S1 の 4 つのノードが属するコミュニティとその他のノードが属するコミュニティの二つに分割できる。

4. 実験

Mucha のモジュラリティと提案モジュラリティとで差が生じる、Ordinal カップリングと Categorical カップリング以外のマルチスライスネットワークでコミュニティ抽出をおこなった。そして、抽出したコミュニティのスライス間エッジの密度を比較した。

データセットは動的ネットワークとして表される highshool[SocioPattens] を用いた。このデータセットでは、ある高校の生徒をノードとして、ある時刻で生徒同士が会うことをそのノード間のエッジで表す。各時刻におけるネットワークをスライスとして Ordinal カップリングを使うことで、マルチスライスネットワークで表すことができる。さらに、Ordinal と Categorical 以外のカップリングのマルチスライスネットワークで実験をおこなうため、ノードとスライスを図 5 のように入れ替えた。入れ替える前の各スライス上のネットワークが、入れ替え後のカップリングになるので、入れ替え後は Ordinal と Categorical 以外のカップリングのマルチスライスネットワークになる。入れ替え後の highshool は表 1 のとおりである。#nodes, #slices, #intra, #inter はそれぞれ、ノード数、スライス数、スライス内のエッジ数、スライス間エッジ数を表す。

表 1: 実験で用いたデータセット

	#nodes	#slices	#intra	#inter
highschool	86	180	2974	6891

スライス間エッジの密度を以下のように定義する.

$$\text{密度} = \frac{\text{全コミュニティに存在するスライス間エッジ数}}{\text{全コミュニティに存在するスライス間エッジ数}} \quad (7)$$

図 6 は, 提案モジュラリティと Mucha のモジュラリティと両方のコミュニティ抽出結果である. $\gamma_s = 1$, $\Gamma_j = 1$ とし, スライス内エッジの重さを 1 とし, パラメータ ω を図 6 の横軸のように設定している. 本来, 提案モジュラリティにはパラメータ ω は存在しないが, Mucha のモジュラリティと同様にスライス間エッジの重さを ω とした. 図 6 のように提案モジュラリティの方が Mucha のモジュラリティよりも密度が高くなっている. 図 7 は提案モジュラリティの Γ_j の値を図 7 の横軸のように動かした場合の結果である. $\gamma_s = 1$ とし, スライス内エッジとスライス間エッジの重さを 1 とした. 提案モジュラリティは Γ_j の値を動かすことで, スライス間エッジの密度を変えることができる.

Mucha のモジュラリティの場合 ω の値を小さくすると, コミュニティがスライス間で分断されやすくなるので密度が大きくなる. しかし, Mucha のモジュラリティの問題点で指摘したようにスライス間エッジが疎なコミュニティも抽出するため, 提案モジュラリティよりも密度が小さくなる. 提案モジュラリティは, Γ_j の値を調節することで, スライス間に関してコミュニティの大きさを調節できる. よって, スライス間エッジの密度が大きいコミュニティを得ることができる.

5. おわりに

提案モジュラリティは Ordinal と Categorical 以外のカップリングのマルチスライスネットワークにおいて, Mucha のモジュラリティよりスライス間エッジが密なコミュニティを抽出できた. そして, Ordinal カップリングか Categorical カップリングのマルチスライスネットワークにおいて, 提案モジュラリティは Mucha のモジュラリティと同じ働きをする.

課題として, Ordinal カップリングや Categorical カップリング以外で表されるデータセットは多く考えられるが, 実際を作るのが難しいということがあげられる. ほとんどの場合, 与えられるデータセットは各ネットワークのエッジリストであり, カップリングのエッジリストは自分で作る必要があるからである. 別の課題としてコミュニティ抽出の精度の検証がある. 今回の実験では, 抽出したコミュニティのスライス間エッジの密度のみを比較した. 正解コミュニティ付きの人工ネットワークを作り, 正確なコミュニティ抽出の精度を比較することが今後の課題である.

参考文献

- [Mikko 14] Mikko Kivela, Alexandre Arenas, Marc Barthélemy, James P. Gleeson, Yamir Moreno, Mason A. Porter: Multilayer Networks. J. Complex Netw. 2(3): 203-271 (2014)
- [Mucha 10] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela:

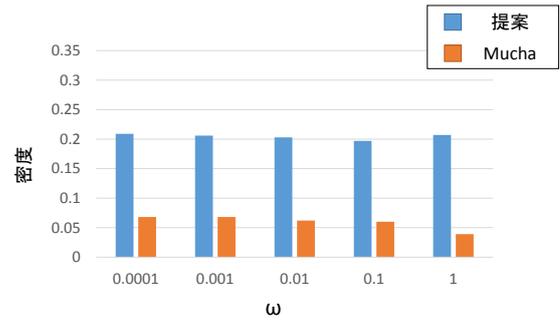


図 6: データセット highschool で抽出したコミュニティのスライス間エッジの密度 (縦軸). 横軸はパラメータ ω . 青が提案モジュラリティの場合, 赤が Mucha のモジュラリティの場合.

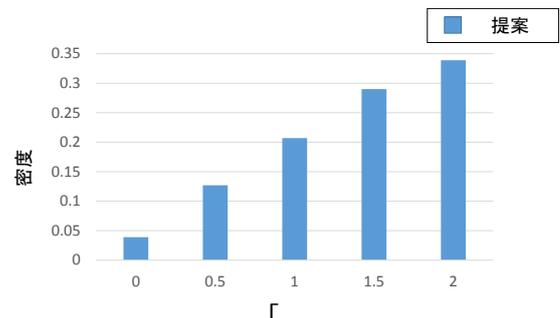


図 7: データセット highschool で抽出したコミュニティのスライス間エッジの密度 (縦軸). 横軸は提案モジュラリティのみ存在するパラメータ Γ_j .

Communiy Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328(5980): 876–878 (2010)

[Blondel 08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, Vol.P10008, 2008

[Newman 04] M.E.J.Newman: Fast algorithm for detecting community structure in networks. Physical Review E, Vol.69, No.066133, pp.1-5, 2004

[GenLouvain] <http://netwiki.amath.unc.edu/GenLouvain/GenLouvain>

[SocioPattens] <http://www.sociopatterns.org/datasets/>