3N3-OS-28a-2

マルチスライスネットワークにおける制約付きコミュニティ抽出法

Constrained Community Detection in Multislice Networks

江口 幸司　　村田 剛志
Koji Eguchi　　Tsuyoshi Murata

東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Computing, Tokyo Institute of Technology

Community detection is one of the methods for network analysis. It is useful for understanding, visualizing and compressing networks. Constrained community detection, which takes given constraints into account to improve the accuracy, is a variation of community detection. Nakata proposed a method for constrained community detection based on the optimization of constrained Hamiltonian by extended Louvain method. In this paper, we propose a new method for constrained community detection in multislice networks. Multislice networks are the combinations of multiple individual networks, which have abilities of representing temporal networks and those with several types of edges. While optimizing Mucha's modularity is popular for community detection in multislice networks, our method optimizes the constrained Hamiltonian which we extend for multislice networks. By using proposed method, we successfully detect communities taking constraints into account.

1. はじめに

与えられたネットワークにおいて，エッジのつながりが密な部分ネットワークを見つけることでネットワークの構造を解析する方法はコミュニティ抽出と呼ばれる．これは，ネットワークの理解や可視化，圧縮などに用いられる．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考える．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つなぐことにより，複数のネットワークを組にしたものである．

従来のコミュニティ抽出には，モジュラリティ[Newman 04] を最適化する方法がよく用いられてきた．モジュラリティは，ネットワークのコミュニティへの分割を評価する関数である．その最適化手法の一つである Louvain 法 [Blondel 08] は，貪欲法で，高速にかつ高い精度でモジュラリティを最適化することができる．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では，Mucha らによるマルチスライスネットワークのモジュラリティ[Mucha 10] を Louvain 法を拡張した Gen Louvain 法 [Jutla 11] で最適化する方法がしばしば用いられる．

背景知識やユーザからのフィードバックを考慮したコミュニティ抽出を制約付きコミュニティ抽出といい，その方法として，制約付きハミルトニアン [Eaton 12] を最適化するものがある．制約付きハミルトニアンは，モジュラリティを一般化したハミルトニアン [Reichardt 06] に制約項を付加したもので，それを最適化する方法には擬似焼きなまし法や拡張された Louvain 法が用いられる．しかしながら，マルチスライスネットワークにおける制約付きコミュニティ抽出の研究は見受けられない．

そこで本稿では，マルチスライスネットワークにおける制約付きコミュニティ抽出を実現するために，ハミルトニアンをマルチスライスネットワーク用に拡張し，制約項を付加して制約付きハミルトニアンを構成する．それを最適化することにより，制約付きコミュニティ抽出をおこなう方法を提案する．

提案したハミルトニアンを用いて，制約付きコミュニティ抽出をおこなったところ，制約を考慮したコミュニティ構造が得られた．また，制約を逐次増やして制約付きコミュニティ抽出を繰り返すと，コミュニティ抽出の精度が高くなった．

連絡先: 江口幸司，東京工業大学 情報理工学院 情報工学系，東京都目黒区大岡山 2-12-1 W8-59，eguchi.k.ac@m.titech.ac.jp

2. 関連研究

本節では，提案法の基礎となる手法や概念について述べる．

2.1 マルチスライスネットワーク

マルチスライスネットワークは，図 1 のように異なるネットワークのノード同士をエッジでつなぐことにより，複数のネットワークを組み合わせたものである．マルチスライスネットワークを構成するそれぞれのネットワークをスライスと呼ぶ．すべてのスライスは同じノードの集合を持ち，スライス間エッジは異なるスライス上の同じノード間だけに存在しうる．

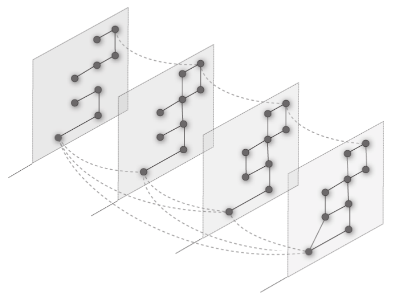

図 1: マルチスライスネットワークの例．([Mucha 10] から引用.)

2.2 マルチスライスネットワークのモジュラリティ

マルチスライスネットワークにおけるモジュラリティとしては，Mucha らによって提案されたものがある．これは式 (1) で表される：

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

ここで，μ は全エッジ数，A_{ijs} はスライス s における隣接行列 A の (i, j) 要素，γ_s はスライス s における分解能，k_{is} は

スライス s 上のノード i の次数，m_s はスライス s におけるスライス内エッジの数，δ はクロネッカーのデルタ，g_{is} はスライス s 上のノード i が属するコミュニティのインデックスである．C_{jsr} はスライス s と r 上のノード j にスライス間エッジがある場合は ω，ない場合は 0 をとる．ω はスライス間エッジの重みを決めるパラメータである．

2.3 ハミルトニアン

ハミルトニアンは，Reichardt と Bornholdt によって提案されたコミュニティ抽出の評価関数である．ハミルトニアン $\mathcal{H}$ は式 (2) で表される：

$$\mathcal{H} = -\sum_{i\neq j} a_{ij}A_{ij}\delta(g_i, g_j) + \sum_{i\neq j} b_{ij}(1-A_{ij})\delta(g_i, g_j) + \sum_{i\neq j} c_{ij}A_{ij}(1-\delta(g_i, g_j)) - \sum_{i\neq j} d_{ij}(1-A_{ij})(1-\delta(g_i, g_j)) \tag{2}$$

式 (2) は，同じコミュニティに属するノード間にエッジがあれば報酬，なければ罰則を，異なるコミュニティに属するノード間にエッジがあれば罰則，なければ報酬を与えている．

式 (2) の各パラメータを調整して整理し，定数項を無視して係数を調整することで，モジュラリティと等価な式が得られる．このことから，ハミルトニアンはモジュラリティを一般化したものであるといえる．

2.4 制約付きコミュニティ抽出

制約付きコミュニティ抽出は，背景知識やユーザからのフィードバックを考慮したコミュニティ抽出のことである．その方法の一つに，制約付きハミルトニアンを最適化するものがある．制約付きハミルトニアンは，Eaton らによって提案された制約付きコミュニティ抽出の評価関数で，ハミルトニアンに制約項を付加したものである．制約項 U は，式 (3) で表される：

$$U = \sum_{i\neq j} \left(u_{ij}\left(1-\delta(g_i, g_j)\right) + \bar{u}_{ij}\delta(g_i, g_j)\right) \tag{3}$$

u_{ij} はノード i と j が同じコミュニティに属するときに 1，そうでなければ 0 となり，$\bar{u}_{ij}$ はノード i と j が異なるコミュニティに属するときに 1，そうでなければ 0 となる．

式 (2) のハミルトニアンと式 (3) の制約項を用いて，制約付きハミルトニアン $\mathcal{H}'$ は式 (4) で表される：

$$\mathcal{H}' = \mathcal{H} + \mu U \tag{4}$$

式 (4) の制約付きハミルトニアンを最適化する手法には，擬似焼きなまし法がよく用いられていた．また，仲田によって提案された，最適化手法に拡張された Louvain 法を用いる方法によって，精度を維持して高速化することができる [仲田 15].

3. 提案法

本節では，本稿で提案するマルチスライスネットワークにおける制約付きコミュニティ抽出法について述べる．

3.1 マルチスライスネットワークのハミルトニアン

ハミルトニアンをマルチスライスネットワーク用に拡張するにあたって，式 (1) のモジュラリティを考える．スライス内については，シングルスライスネットワークの場合と同様である．一方，スライス間については，エッジの有無を 0 または ω の値で示しているだけなので，スライス間エッジがある場合だけにそれを考慮すればよい．したがって，マルチスライスネットワークのハミルトニアン $\mathcal{H}_{\mathrm{multi}}$ を式 (5) のように定める：

$$\begin{aligned}\mathcal{H}_{\mathrm{multi}} = &-\sum_{ijsr} a_{ijs}A_{ijs}\delta(g_{is}, g_{js})\delta_{sr} \\ &+\sum_{ijsr} b_{ijs}\left(1-A_{ijs}\right)\delta(g_{is}, g_{js})\delta_{sr} \\ &+\sum_{ijsr} c_{ijs}\, A_{ijs}(1-\delta(g_{is}, g_{js}))\delta_{sr} \\ &-\sum_{ijsr} d_{ijs}\left(1-A_{ijs}\right)(1-\delta(g_{is}, g_{js}))\delta_{sr} \\ &-\sum_{ijsr} e_{jsr}\, C_{jsr}\delta(g_{js}, g_{jr})\delta_{ij} \\ &+\sum_{ijsr} f_{jsr}\, C_{jsr}(1-\delta(g_{js}, g_{jr}))\delta_{ij}\end{aligned} \tag{5}$$

式 (5) において，第一項から第四項は式 (2) のハミルトニアンと同じである．また，第五項と第六項によって，異なるスライス上の同じノード間にエッジが存在するとき，それらのノードが同じコミュニティに属する場合には報酬，異なるコミュニティに属する場合には罰則を与える．

また，式 (5) において $a_{ijs} = c_{ijs} = 1 - \gamma_s k_{is}k_{js}/2m_s$，$b_{ijs} = d_{ijs} = \gamma_s k_{is}k_{js}/2m_s$，$e_{jsr} = f_{jsr} = 1$ と定めて整理すると，式 (6) のように書き直せる：

$$\begin{aligned}\mathcal{H}_{\mathrm{multi}} = &-2\sum_{ijsr}\left[\left(A_{ijs} - \gamma_s\frac{k_{is}k_{js}}{2m_s}\right)\delta_{sr} + \delta_{ij}C_{jsr}\right]\delta(g_{is}, g_{jr}) \\ &+\sum_{ijsr}\left[\left(A_{ijs} - \gamma_s\frac{k_{is}k_{js}}{2m_s}\right)\delta_{sr} + \delta_{ij}C_{jsr}\right]\end{aligned} \tag{6}$$

式 (6) の第二項（定数項）を無視し，係数を調整すると，式 (1) と等価な式を得ることができる．したがって，マルチスライスネットワークにおいて，ここで構築したハミルトニアンはモジュラリティの一般化であるということができる．

3.2 マルチスライスネットワークの制約付きハミルトニアン

マルチスライスネットワークのハミルトニアンに制約項を加えることで制約付きハミルトニアンを構築する．制約項は，ノードのペアに対して，同じコミュニティに属するべきであるという制約（正の制約），もしくは異なるコミュニティに属するべきであるという制約（負の制約）を与えるように構成する．スライス間のノードのペアに対する制約は，同じノードだけに与えることにする．したがって，制約項 U_{multi} は，式 (7) で表される：

$$\begin{aligned}U_{\mathrm{multi}} = &\sum_{ijsr}\left(u_{ijs}\left(1-\delta(g_{is}, g_{js})\right) + \bar{u}_{ijs}\delta(g_{is}, g_{js})\right)\delta_{sr} \\ &+\sum_{ijsr}\left(v_{jsr}\left(1-\delta(g_{js}, g_{jr})\right) + \bar{v}_{jsr}\delta(g_{js}, g_{jr})\right)\delta_{ij}\end{aligned} \tag{7}$$

ここで，u_{ijs} と $\bar{u}_{ijs}$ は，スライス s 上のノード i と j に正の制約がある場合には前者が 1 となり，負の制約がある場合には後者が 1 となる．v_{jsr} と $\bar{v}_{jsr}$ は，スライス s と r 上のノード j に正の制約がある場合には前者が 1 となり，負の制約がある場合には後者が 1 となる．それ以外の場合は 0 となる．

式 (7) の制約項と式 (6) のハミルトニアンを用いて，式 (8)

の制約付きハミルトニアン $\mathcal{H}'_{\mathrm{multi}}$ を得る：

$$\begin{aligned}\mathcal{H}'_{\mathrm{multi}} =& \mathcal{H}_{\mathrm{multi}} + \mu U_{\mathrm{multi}} \\ =& -2\sum_{ijsr}\left(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s} + \mu\Delta U_{ijs}\right)\delta(g_{is}, g_{js})\delta_{sr} \\ & -2\sum_{ijsr}(C_{jsr} + \mu\Delta V_{jsr})\,\delta(g_{js}, g_{jr})\delta_{ij} \\ & + \sum_{ijsr}\left(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s} + \mu u_{ijs}\right)\delta_{sr} \\ & + \sum_{ijsr}(C_{jsr} + \mu v_{jsr})\delta_{ij}\end{aligned} \tag{8}$$

ここで，$\Delta U_{ijs} = (u_{ijs} - \bar{u}_{ijs})/2$，$\Delta V_{jsr} = (v_{jsr} - \bar{v}_{jsr})/2$ である．また，式 (8) の第三項と第四項は，コミュニティ抽出の結果によらない定数である．以下，本稿では式 (8) の制約付きハミルトニアンを提案ハミルトニアンと呼ぶ．提案ハミルトニアンを Gen Louvain 法で最適化することで，マルチスライスネットワークにおける制約付きコミュニティ抽出をおこなう．

4. 実験

本節では，第 3 節で提案したコミュニティ抽出法を用いておこなった実験について述べる．

4.1 制約項の有効性の検証

本実験では，制約なしのコミュニティ抽出と制約付きコミュニティ抽出をおこない，その結果がどのように異なるかを調べた．データセットには “PADGETT FLORENTINE FAMILIES”[Padgett 93] を用いた．これは，ノード数 16，エッジ数 35，スライス数 2 の実ネットワークである．パラメータは $\omega = 1$，$\gamma_s = 1$ とした．パラメータ μ は，[仲田 15] にならって，制約なしのコミュニティ抽出をおこなう場合は $\mu = 0$，制約付きコミュニティ抽出をおこなう場合は $\mu = 2$ とした．

本実験では，以下の三つの条件の場合において実験をおこない，その結果として得られたコミュニティ構造を比較した．

1. 制約なし
2. スライス 1 のノード 4 とノード 11 に正の制約を与える
3. スライス 1 とスライス 2 のノード 6 に負の制約を与える

それぞれのコミュニティ抽出の結果として得られたコミュニティ構造を MuxViz[Porter 15] を用いて可視化したものを図 2 から図 4 に示す．なお，同じ色のノードは同じコミュニティに属していることを示す． これらを比較すると，制約を与えた場合にはコミュニティ構造が変化し，制約を考慮したコミュニティ構造が得られていることがわかる．したがって，提案ハミルトニアンを最適化する方法によって，制約付きコミュニティ抽出を実現することができるといえる．

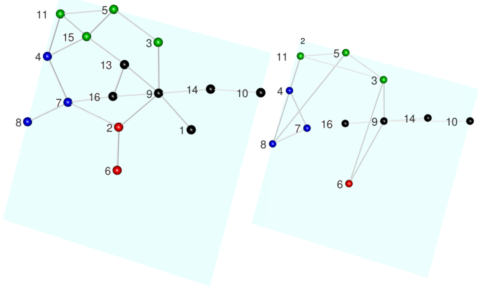

図 2: “PADGETT FLORENTINE FAMILIES” における制約なしのコミュニティ抽出結果.

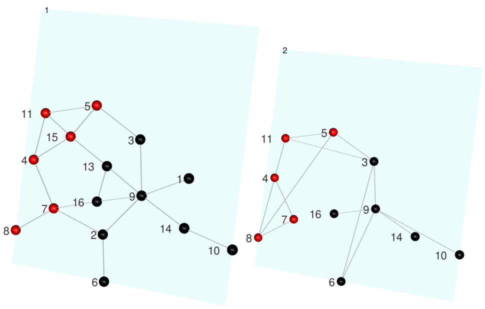

図 3: “PADGETT FLORENTINE FAMILIES” においてスライス 1 のノード 4 とノード 11 に正の制約を与えたときのコミュニティ抽出結果.

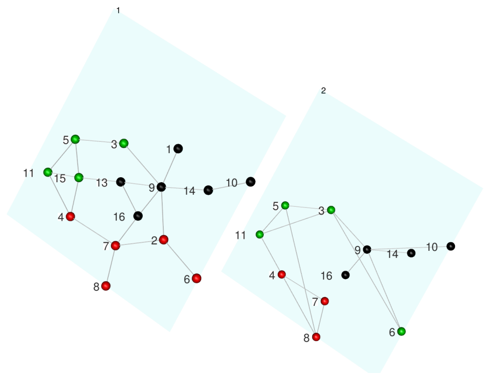

図 4: “PADGETT FLORENTINE FAMILIES” においてスライス 1 とスライス 2 のノード 6 に負の制約を与えたときのコミュニティ抽出結果.

4.2 制約を逐次与えて制約付きコミュニティ抽出を繰り返す実験

ここでは，コミュニティ抽出の精度を評価するために，Domenico らによる人工ベンチマーク [Domenico 15] を用いた．また，精度を測る指標にはマルチスライスネットワークの Normalized Mutual Information (NMI)[Domenico 15] を用いた．本実験では，隣接コミュニティが多いノードに制約を一つ与えて，提案法を用いて制約付きコミュニティ抽出をおこない，その精度 (NMI) を調べることを NMI = 1 となるまで繰り返した．この方法によって，表 1 に示すようなスライス数は同じでノード数が異なる複数のベンチマークを用いた実験（実験 A）と，表 2 に示すようなノード数が同じでスライス数が異なる複数のベンチマークを用いた実験（実験 B）をおこなった．パラメータは $\omega = 1$，$\gamma_s = 1$，$\mu = 2$ と定めた．この実験の結果を図 5 と図 6 に示す． これらを見ると，制約を逐次与

表 1: 実験 A で用いたベンチマーク.

Benchmark Network	#Nodes	#Edges	#Slices
Sample1-1	10	11	3
Sample1-2	20	36	3
Sample1-3	30	54	3
Sample1-4	40	67	3
Sample1-5	50	78	3

表 2: 実験 B で用いたベンチマーク.

Benchmark Network	#Nodes	#Edges	#Slices
Sample2-1	30	55	2
Sample2-2	30	54	3
Sample2-3	30	57	4
Sample2-4	30	55	5

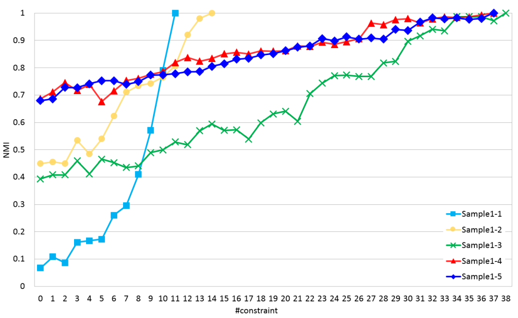

図 5: 実験 A の結果．横軸は制約数，縦軸はコミュニティ抽出の精度 (NMI).

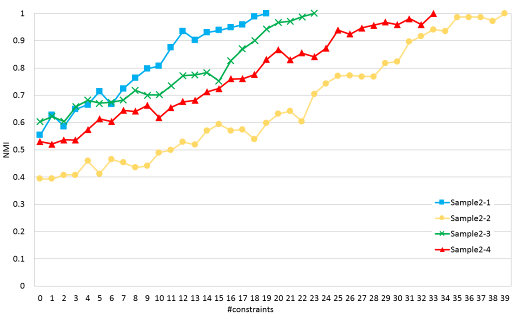

図 6: 実験 B の結果．横軸は制約数，縦軸はコミュニティ抽出の精度 (NMI).

えて制約付きコミュニティ抽出を繰り返しおこなうことによって，コミュニティ抽出の精度が高くなっていることがわかる．したがって，ユーザからのフィードバックを考慮してコミュニティ抽出をおこなうことで，ユーザの望むコミュニティを抽出することができるといえる．

5. おわりに

本稿では，ハミルトニアンをマルチスライスネットワーク用に拡張し，制約項を付加して制約付きハミルトニアンを構築した．また，それを目的関数として Gen Louvain 法を用いて，マルチスライスネットワークにおける制約付きコミュニティ抽出をおこなう方法を提案した．

制約なしの場合と制約付きの場合のそれぞれにおいて提案法を用いてコミュニティ抽出をおこなったところ，制約付きの場合には制約を考慮したコミュニティ構造が得られていることを確認した．また，人工ベンチマークを用いて，制約を逐次与えて制約付きコミュニティ抽出を繰り返しおこなったところ，コミュニティ抽出の精度は高くなった．

残された課題としては，ユーザの負担を減らすために与える制約数が少なくて済むようにコミュニティ抽出アルゴリズムを改良することが挙げられる．また，大規模ネットワークなどのさまざまなネットワークに適用するために，制約を自動的に付与できるようにして，様々なネットワークにおいて実験をおこなうことも挙げられる．今後はこれらの課題を解決することで，ユーザの負担軽減や提案法の質の向上を目指したい．

参考文献

[Mikko 14] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter: Multilayer Networks, *Journal of Complex Networks*, Vol. 2, No. 3, pp. 203-271 (2014).

[Newman 04] M. E. J. Newman and M. Girvan: Finding and evaluating community structure in networks. *Phys. Rev. E*, Vol. 69, p. 026113 (2004).

[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. 2008, No. 10, p. P10008 (2008).

[Mucha 10] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela: Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science*, Vol. 328, No. 5980, pp. 876-878 (2010).

[Jutla 11] Inderjit S. Jutla, Lucas G. S. Jeub, and Peter J. Mucha: A generalized Louvain method for community detection implemented in MATLAB. `http://netwiki.amath.unc.edu/GenLouvain` (2011-2014).

[Eaton 12] Eric Eaton and Rachael Mansbach: A Spin-Glass Model for Semi-Supervised Community Detection. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12)*, pp. 900-906. AAAI Press, July 22-26 (2012).

[Reichardt 06] Jörg Reichardt and Stefan Bornholdt: Statistical mechanics of community detection. *Phys. Rev. E*, Vol. 74, p. 016110 (2006).

[Padgett 93] John F. Padgett and Christopher K. Ansell: Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, Vol. 98, No. 6, pp.1259-1319 (1993).

[Domenico 15] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall: Identifying modular flows on multilayer networks reveals highly overlapping organization in social systems. *Phys. Rev. X*, Vol. 5, p. 011027 (2015).

[Porter 15] Manlio De Domenico, Mason A. Porter, and Alexandre Arenas. MuxViz: A Tool for Multilayer Analysis and Visualization of Networks. *Journal of Complex Networks*, Vol. 3, No. 2, pp. 159-176 (2015).

[仲田 15] 仲田圭佑, 村田剛志: ハミルトニアンに基づいた制約付きコミュニティ抽出の高速化. 人工知能学会論文誌, Vol. 30, No. 1, pp.96-101 (2015).