

重複コミュニティ発見のための重み付き線グラフ

吉田 哲也^{1,a)}

受付日 2012年4月17日, 再受付日 2012年6月1日,
採録日 2012年7月9日

概要: 本稿では, 複数のコミュニティへの所属を許容する重複コミュニティの発見を実現するために, ネットワークの重みを反映する重み付き線グラフを提案する. 従来のノード分割に基づくコミュニティ発見手法ではノードは1つのコミュニティに割り当てられるため, 複数のコミュニティには所属できないという課題がある. この課題に対し, 本稿ではネットワークをその線グラフに変換し, 変換後の線グラフにノード分割手法を適用することにより重複コミュニティ発見を実現する. 従来の線グラフはネットワークの接続関係のみから定義されるが, ネットワークの重みを活用したリンク分割を実現するため, 重みに基づいて拡張した重み付き線グラフを提案し, その性質を示す. さらに, ノード分割に基づくモジュラリティを拡張し, 重複コミュニティ発見におけるソフトなノード分割に対するモジュラリティを提案する. 提案法を人工ネットワークと実世界のネットワークに適用し, 他手法との比較を通じてその有効性を示す.

キーワード: コミュニティ発見, 重複コミュニティ, 線グラフ, モジュラリティ

Weighted Line Graphs for Overlapping Community Discovery

TETSUYA YOSHIDA^{1,a)}

Received: April 17, 2012, Revised: June 1, 2012,
Accepted: July 9, 2012

Abstract: We propose an approach for overlapping community discovery via weighted line graphs of networks. For undirected connected networks without self-loops, we propose weighted line graphs by: 1) defining weights of a line graph based on the weights in the original network, and 2) removing self-loops in weighted line graphs, while sustaining their properties. By applying some off-the-shelf node partitioning method to the weighted line graph, the node in the original network can be assigned to more than one community based on the community labels of its adjacent links. Various properties of the proposed weighted line graphs are clarified. Furthermore, we propose a generalized quality measure for soft assignment of nodes in overlapping communities. Preliminary experiments are conducted over synthetic and real-world networks, and the results indicate that the proposed approach can improve the quality of discovered overlapping communities.

Keywords: community discovery, overlapping community, line graph, modularity

1. はじめに

ネットワークからのコミュニティ発見は主に社会科学などの分野で研究されてきたが, 近年の計算資源の発展と普及にともない, 計算機科学の立場からの研究も活発に行われている. たとえばネットワークの構造的な性質の探求な

どが活発に行われている [11], [21]. 従来の研究ではネットワークからのコミュニティ発見を密な部分ネットワークを同定する問題ととらえ, 各頂点を1つのコミュニティに割り当てられることが多かった [15], [16], [17], [18], [23]. しかし, 実世界のネットワークでは1つのノードが複数のコミュニティに属することがある. たとえば Facebook などのソーシャルネットワーク [12] では, ネットワークのノードに対応するユーザは音楽や映画などの好みに応じて複数のグループに所属する場合がある.

本稿では, 複数のコミュニティへの所属を許容する重複

¹ 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

^{a)} yoshida@meme.hokudai.ac.jp

コミュニティの発見を実現するために、ネットワークの重みを反映する重み付き線グラフを提案する。自己ループのない無向ネットワークに対して、従来の線グラフ [5], [6] を下記のように拡張する：1) オリジナルのネットワークにおける重みに基づいて線グラフにおけるリンクの重みを定義する、2) ネットワーク（グラフ）の性質を保存しながら重み付き線グラフにおける自己ループを削除する。オリジナルのネットワークを重み付き線グラフに変換し、既存のノード分割に基づくコミュニティ発見手法を変換後のネットワークに適用してリンク分割を行う。接続するリンクのコミュニティラベルをノードに割り当てることにより、ノードに対する重複コミュニティ発見を実現する。

提案する重み付き線グラフの様々な性質を示し、この提案がグラフ理論における線グラフ [5] や先行研究 [6] の自然な拡張であることを示す。さらに、ノード分割に基づく従来のモジュラリティ [14] を拡張し、重複コミュニティ発見におけるソフトなノード分割に対するモジュラリティを提案する。提案する指標はオリジナルのネットワークを線グラフなどの別のネットワークに変換するか [6], [8], あるいはネットワークの辺を直接コミュニティに割り当てるか [2] に不変なため、重複コミュニティ発見に対し広く適用可能であると考えられる。

提案法を人工ネットワークおよび実ネットワークに適用して評価し、他手法との比較を通じてその有効性を確認した。特に、本稿で提案する、1) オリジナルのネットワークにおける重みの活用、2) 自己ループの削除、がそれぞれ重複コミュニティ発見の性能向上に役立つことを確認した。

2章で関連研究を紹介し、3章で先行研究を説明する。4章で提案法を述べ、5章で評価実験を報告して提案法の有効性を議論する。6章でまとめと今後の展望を述べる。

2. 関連研究

ネットワークからのコミュニティ発見に対する従来の多くの研究ではコミュニティを密な部分ネットワークととらえ、ネットワークのノード分割によりコミュニティ発見を行うものが多い [15], [16], [17], [18]。また、発見したコミュニティの評価には 4.4 節で述べるモジュラリティと呼ばれる指標 [14] が標準的に用いられてきた。

しかし、実世界のネットワークではノードが複数のコミュニティに属することがあるため、コミュニティ間でノードが重複する場合がある。重複コミュニティ発見に対してはこれまで主に以下のアプローチがとられてきた。

- a) 重複コミュニティに対する望ましい性質を定義し、その性質を満たす手法を考案する。
- b) ネットワークのノードを直接複数のコミュニティに割り当てる。
- c) ネットワークのノードではなくリンクをコミュニティに割り当てる。

a) の例としては、コミュニティをクリークの連なりと定義し、クリークの重複行列に基づいて k -クリークコミュニティ^{*1}を発見する手法がある [1]。ノードは複数の k -クリークコミュニティに所属できるため、コミュニティの重複が可能である。

b) の例としては、オリジナルのネットワークをノードのコピーが複数存在するような別のネットワークに変換し、変換したネットワークに既存のノード分割手法を適用してコミュニティ発見を行うものがある [8]。変換後のネットワークで発見したコミュニティラベルをオリジナルのネットワークのノードに割り当てることで重複コミュニティ発見を行う。

c) の例としては、ノードではなくリンクに既存のコミュニティ発見手法を適用するアプローチがある [2], [6]。リンクに対するコミュニティラベルに基づき、隣接するリンクのラベルをノードに割り当てることで重複コミュニティ発見を行う。

3. 先行研究

3.1 準備

本稿では、行列は太字の大文字、ベクトルは太字のイタリック小文字で表記し、 \mathbf{A}_{ij} で行列 \mathbf{A} の第 ij 要素を表す。tr は行列のトレースを表し、 \mathbf{A} の転置を \mathbf{A}^T で表す。要素がすべて 1 である n 次元ベクトルを $\mathbf{1}_n$ と表記する。また、ベクトル \mathbf{a} から生成される対角行列を $\text{diag}(\mathbf{a})$ と表記する^{*2}。逆に、正方行列 \mathbf{A} の対角成分から生成されるベクトルを $\text{diag}(\mathbf{A})$ と表記する。

ノード（頂点）の集合 V とリンク（辺）の集合 E から構成されるネットワーク（グラフ）を $G = (V, E)$ とする^{*3}。ネットワーク分析では無向で自己ループのない単純グラフを扱うことが多いため [12], [16]、本稿でもこのクラスのネットワークを扱う。

ネットワーク G のノード数を n とし、リンク数を m とする。 G の接続関係を表現する隣接行列を $\mathbf{A} \in \{0, 1\}^{n \times n}$ と表記し、 G のノード i, j がリンクで接続する場合に $\mathbf{A}_{ij} = 1$ であり、接続しない場合は $\mathbf{A}_{ij} = 0$ である。単純グラフに対する隣接行列の対角要素はすべて 0 である。ネットワーク G の隣接行列 \mathbf{A} に対して $\mathbf{k} = \mathbf{A}\mathbf{1}_n$ を次数ベクトルと呼び、要素 k_i は i 番目のノードの次数を表す。また、 G の接続行列 $\mathbf{B} \in \{0, 1\}^{n \times m}$ は、ノード i がリンク α の端点である場合に $\mathbf{B}_{i\alpha} = 1$ であり、そうでなければ $\mathbf{B}_{i\alpha} = 0$ と定義される [5]。

*1 k -クリークコミュニティは $k-1$ 個のノードを共有して接続する k -クリークの和集合と定義される。

*2 行列 $\text{diag}(\mathbf{a})$ の第 i 対角要素はベクトル \mathbf{a} の第 i 要素である。

*3 本稿ではノード、リンク、ネットワークをそれぞれ頂点、辺、グラフとも呼ぶ。

3.2 線グラフに基づくリンク分割

2章のc)におけるリンク分割を実現する手法として、単純グラフに対する線グラフの活用が提案された[6]。グラフ理論における線グラフは以下で定義される。

定義 1 (線グラフ [5])。単純グラフ $G = (V, E)$ の線グラフ $L(G)$ とは、 G 中の隣接する辺の組 $x, y \in E$ をそのグラフの頂点として辺で結んで得られる E 上のグラフである。

ネットワーク G のリンクは $L(G)$ のノードに対応するため、既存のノード分割手法を $L(G)$ に適用することで G に対するリンク分割が行われる。

連結な単純グラフであるネットワーク G に対し、 G 上の酔歩に基づいた線グラフの拡張として、 G の隣接行列に基づく重み付き線グラフが提案された[6]。もともと、定義1における従来の線グラフ $L(G)$ の隣接行列 \mathbf{C} は、 G の接続行列 \mathbf{B} を用いて以下で表される。

$$\mathbf{C} = \mathbf{B}^T \mathbf{B} - 2\mathbf{I}_m \quad (1)$$

ここで $\mathbf{I}_m \in \{0, 1\}^{m \times m}$ は単位行列である。先行研究[6]ではこれを拡張し、以下の行列を変換後のネットワークに対する(非負実数値の)隣接行列として提案した。

$$\mathbf{E} = \mathbf{B}^T \mathbf{D}^{-1} \mathbf{B} \quad (2)$$

$$\mathbf{E}_1 = \mathbf{B}^T \mathbf{D}^{-1} \mathbf{A} \mathbf{D}^{-1} \mathbf{B} \quad (3)$$

ここで \mathbf{D} はネットワークの次数ベクトル \mathbf{k} から $\mathbf{D} = \text{diag}(\mathbf{k})$ として生成される対角行列である。上記はともに G のリンク間での酔歩に基づく隣接行列であるが、式(2)の \mathbf{E} は G 上のリンク-ノード-リンクに対する酔歩に基づくものであり、式(3)の \mathbf{E}_1 は G 上のリンク-リンク-リンクに対する酔歩に基づくものである[6]。行列 \mathbf{E} , \mathbf{E}_1 の要素は整数とは限らないため、変換後のネットワークではリンクに重みが付与されることになる。

4. 重み付き線グラフに基づく重複コミュニティ発見

4.1 先行研究での課題

3.2節で述べたように、先行研究ではオリジナルのネットワークを式(2), (3)などの隣接行列を持つネットワークに変換し、既存のノード分割手法を適用した結果を報告していたが、以下のような課題がある。

- i) 変換後のネットワーク上の重みはオリジナルのネットワークの接続関係(次数)のみに基づいて定義される。
- ii) 行列 \mathbf{E} , \mathbf{E}_1 が重み付き隣接行列として推奨されたが、変換後のネットワークにおける接続関係は定義1の従来の線グラフと異なる。
- iii) 変換後のネットワークから発見したコミュニティはオリジナルのネットワークにおけるノードのコミュニティ割当てに対して評価されていない。

グラフやネットワークを扱う研究ではノード間の類似度

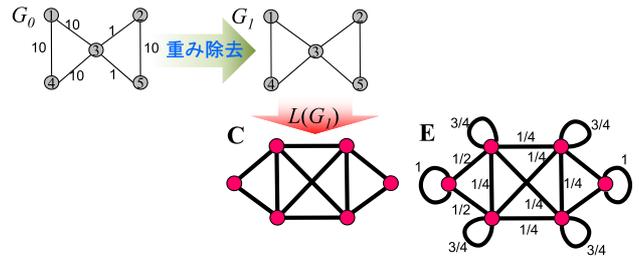


図 1 従来の線グラフの例 [5], [6]

Fig. 1 Previous line graphs [5], [6].

をリンクの重みとして表現することが多いが[20], [23], 上記 i) よりネットワークでノードがどの程度関連するかを変換後のネットワークに反映できないことになる。たとえば図1の重み付きネットワーク G_0 で、ノード $\{1, 3, 4\}$ とノード $\{2, 5\}$ は大きな重みを持つリンクで接続しあうためコミュニティを形成すると考えられるが、 G_0 の重みは線グラフ上での重みを定義する際に無視されてしまう。これは、 G_0 を重みなしネットワーク G_1 に変換し、変換後の G_1 に基づいて線グラフを考えることに対応する(図1参照)。

グラフ理論における定義1の線グラフの興味深い性質として、グラフ G とその線グラフ $L(G)$ に対する Whitney の一意性定理[22]があり、 G が三角形や4頂点からなる星形グラフでなければ G の構造(接続関係)を $L(G)$ から復元できることが保証される。しかし、上記 ii) で述べたように \mathbf{E} や \mathbf{E}_1 では対角要素が非零で自己ループを含み、グラフ理論における線グラフの構造と異なるため、上記の定理に基づいて G の構造を復元することは容易ではない[19]。

また、変換後のネットワークから発見したコミュニティは、オリジナルのネットワークのノードではなく変換後のネットワークにおけるノードの割当てに基づいて評価されていた。さらに、評価関数が変換後のネットワークにおける隣接行列に対して定義されており、同じオリジナルのネットワークに対しても \mathbf{E} や \mathbf{E}_1 などは行列自体が異なるため単純に評価値のみで比較できなかった。

本稿では上記の課題に対するアプローチを提案する。4.2節は i), 4.3節は ii), 4.4節は iii) に対応する。

4.2 重み付きネットワークに対する重み付き線グラフ

ノード数 n とリンク数 m の単純グラフ(ネットワーク) G がリンクに対して非負の重みを持つとする。ノード i と j の間のリンクの重みを w_{ij} で表し、 G のリンクの重みを並べたベクトルを $\mathbf{w} \in \mathbb{R}^m$ で表す。以下ではリンクの重みはノード間の類似度に対応するものとする[20], [23]。

4.2.1 重み付きネットワークの表現行列

重み付きネットワーク G の隣接行列を非負実数値行列 $\tilde{\mathbf{A}} \in \mathbb{R}_+^{n \times n}$ で表し、 $\tilde{\mathbf{A}}_{ij} = w_{ij}$ とする。行列 $\tilde{\mathbf{A}}$ に基づき、以下のベクトルと行列を定義する。

$$\tilde{\mathbf{k}} = \tilde{\mathbf{A}} \mathbf{1}_n \quad (4)$$

$$\tilde{\mathbf{D}} = \text{diag}(\tilde{\mathbf{k}}) \quad (5)$$

式 (4) のベクトル $\tilde{\mathbf{k}}$ は各ノードに接続するリンクの重みの和を表し、次数ベクトル \mathbf{k} の拡張に対応する。式 (5) の対角行列 $\tilde{\mathbf{D}}$ は式 (2), (3) での対角行列 \mathbf{D} に対応する。

次に、 G に対する重み付き接続行列 $\tilde{\mathbf{B}}$ を定義する。ノード i, j を端点とするリンク $\alpha = (i, j)$ に対し、 $\tilde{\mathbf{B}}$ の第 α 列で $\tilde{\mathbf{B}}_{i\alpha}$ と $\tilde{\mathbf{B}}_{j\alpha}$ を w_{ij} とし、それ以外の要素は 0 とする。これは、 G における重みに基づいて 0-1 行列である従来の隣接行列 \mathbf{A} を $\tilde{\mathbf{A}}$ に拡張したように、従来の接続行列 \mathbf{B} を $\tilde{\mathbf{B}}$ に拡張することに対応する。

命題 1. 接続行列 \mathbf{B} , $\tilde{\mathbf{B}}$ に対し以下の性質が成り立つ。

$$\mathbf{B}\mathbf{1}_m = \mathbf{k}, \quad \tilde{\mathbf{B}}\mathbf{1}_m = \tilde{\mathbf{k}} \quad (6)$$

$$\mathbf{1}_m^T \mathbf{B} = 2\mathbf{1}_m^T, \quad \mathbf{1}_m^T \tilde{\mathbf{B}} = 2\mathbf{w}^T \quad (7)$$

Proof. 行列 \mathbf{B} , $\tilde{\mathbf{B}}$ の定義より明らか。 \square

命題 1 より、上記で定義した $\tilde{\mathbf{B}}$ は従来の接続行列 \mathbf{B} の自然な拡張になっていることが分かる。

4.2.2 重み付きネットワークに対する重み付き線グラフ

上記の行列に基づき、定義 1 に対する隣接行列である式 (1), および先行研究の式 (2), (3) の拡張として、重み付き線グラフに対する以下の隣接行列を提案する。

$$\tilde{\mathbf{C}} = \tilde{\mathbf{B}}^T \tilde{\mathbf{B}} - 2\text{diag}(\mathbf{w}_2) \quad (8)$$

$$\tilde{\mathbf{E}} = \tilde{\mathbf{B}}^T \tilde{\mathbf{D}}^{-1} \tilde{\mathbf{B}} \quad (9)$$

$$\tilde{\mathbf{E}}_1 = \tilde{\mathbf{B}}^T \tilde{\mathbf{D}}^{-1} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1} \tilde{\mathbf{B}} \quad (10)$$

ここで $\mathbf{w}_2 = \mathbf{w} \odot \mathbf{w}$ であり、 \odot は行列の要素積を表す [10]。

定理 2. 重み付き線グラフに対し以下の性質が成り立つ。

$$\mathbf{1}_m^T \mathbf{C} = (\mathbf{k} - \mathbf{1}_n)^T \mathbf{B}, \quad \mathbf{1}_m^T \tilde{\mathbf{C}} = (\tilde{\mathbf{k}} - \mathbf{1}_n)^T \tilde{\mathbf{B}} \quad (11)$$

$$\mathbf{1}_m^T \mathbf{E} = 2\mathbf{1}_m^T, \quad \mathbf{1}_m^T \tilde{\mathbf{E}} = 2\mathbf{w}^T \quad (12)$$

$$\mathbf{1}_m^T \mathbf{E}_1 = 2\mathbf{1}_m^T, \quad \mathbf{1}_m^T \tilde{\mathbf{E}}_1 = 2\mathbf{w}^T \quad (13)$$

証明は付録 A.1 に示す。定理 2 は提案する隣接行列がその列和の観点から自然な拡張になっていることを示す。式 (12), (13) の意義は 4.4 節で述べる。

4.3 自己ループを除去した重み付き線グラフ

定義 1 より、単純グラフに対する従来の線グラフも単純グラフとなる。しかし、式 (2), (3) での \mathbf{E} , \mathbf{E}_1 では対角要素が非零で自己ループを含むため単純グラフではない。このため、グラフ G とそれを変換したグラフの間で構造 (接続関係) の対応が成り立たないことになる。

上記の課題に対し、連結ネットワークを対象として自己ループを持つ重み付き無向ネットワークを自己ループを持たない重み付き無向ネットワークに変換する手法を提案する。連結ネットワークに対しては定義 1 での線グラフも連結であることに注意されたい。

正方対称行列 $\mathbf{M} \in \mathbb{R}_+^{\ell \times \ell}$ を自己ループを持つネットワークの (重み付き) 隣接行列とする。提案法は行列 \mathbf{M} の性質を保存しながら対角要素を非対角要素に分配する。これを実現するため、以下の行列とベクトルを定義する。

$$\mathbf{m} = \text{diag}(\mathbf{M}) \quad (14)$$

$$\mathbf{D}_M = \text{diag}(\mathbf{m}) \quad (15)$$

$$\mathbf{M}_{wo} = \mathbf{M} - \mathbf{D}_M \quad (16)$$

$$\mathbf{m}_{wo} = \mathbf{M}_{wo} \mathbf{1}_\ell \quad (17)$$

\mathbf{m} は行列 \mathbf{M} の対角要素から生成されるベクトルであり、 \mathbf{D}_M はこの \mathbf{m} を要素とする対角行列である。行列 \mathbf{M}_{wo} は \mathbf{M} の非対角要素のみを持つ行列であり (対角要素はすべて零)、 \mathbf{m}_{wo} はその行和に対応する。

上記に基づき、提案法は行列 $\mathbf{M} \in \mathbb{R}_+^{\ell \times \ell}$ を以下の行列 $\mathbf{N} \in \mathbb{R}_+^{\ell \times \ell}$ に変換する。

$$\mathbf{N} = \mathbf{M}_{wo} + \mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{M}_{wo} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2} \quad (18)$$

ここで $\text{diag}(\mathbf{m}_{wo})$ は \mathbf{m}_{wo} を要素とする対角行列である。

変換後の行列 \mathbf{N} に対して以下の性質が成り立つ。

定理 3. 非負実数値を持つ正方対称行列 \mathbf{M} に対し、式 (18) で定義する行列 \mathbf{N} では以下の性質が成り立つ。

$$\text{diag}(\mathbf{N}) = \mathbf{0}_\ell \quad (19)$$

$$\mathbf{N}^T = \mathbf{N} \quad (20)$$

$$\mathbf{N}\mathbf{1}_\ell = \mathbf{M}\mathbf{1}_\ell \quad (21)$$

証明は付録 A.2 に示す。式 (19) より行列 \mathbf{N} の対角成分はすべて零であり、式 (20) より対称行列である。このため、行列 \mathbf{N} を (重み付き) 隣接行列とするネットワークは無向で自己ループのない単純グラフとなる。

3.2 節での行列 \mathbf{E} と \mathbf{E}_1 , および 4.2.1 項で定義した $\tilde{\mathbf{B}}$ に基づいて提案する行列 $\tilde{\mathbf{E}}$ と $\tilde{\mathbf{E}}_1$ を式 (18) の \mathbf{M} に代入することで、自己ループを除去した \mathbf{F} と \mathbf{F}_1 , および $\tilde{\mathbf{F}}$ と $\tilde{\mathbf{F}}_1$ を定義できる。特に、式 (12), (13), (21) より、これらの行列に対して以下の性質が成り立つ。

系 4. 自己ループを除去した重み付き線グラフの隣接行列に対して以下の性質が成り立つ。

$$\mathbf{1}_m^T \mathbf{F} = \mathbf{1}_m^T \mathbf{E} = 2\mathbf{1}_m^T, \quad \mathbf{1}_m^T \tilde{\mathbf{F}} = \mathbf{1}_m^T \tilde{\mathbf{E}} = 2\mathbf{w}^T \quad (22)$$

$$\mathbf{1}_m^T \mathbf{F}_1 = \mathbf{1}_m^T \mathbf{E}_1 = 2\mathbf{1}_m^T, \quad \mathbf{1}_m^T \tilde{\mathbf{F}}_1 = \mathbf{1}_m^T \tilde{\mathbf{E}}_1 = 2\mathbf{w}^T \quad (23)$$

上記の性質の意義は 4.4 節で述べる。

4.3.1 提案する重み付き線グラフの例

図 1 に示したネットワーク G_0 に対して提案法で構築するネットワークの例を図 2 に示す。 G_0 の重みは左の部分グラフの方が右より大きいため、これを反映して図 2 ではネットワーク上の重みは対称とはなっていない。また、 $\tilde{\mathbf{F}}$ では \mathbf{C} や $\tilde{\mathbf{C}}$ と同様に自己ループを持たない。

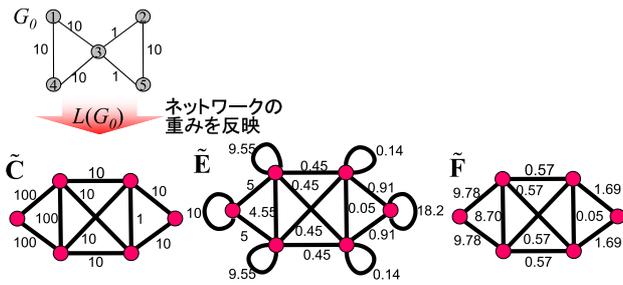


図 2 提案する重み付き線グラフの例
Fig. 2 Proposed weighted line graphs.

4.4 重複コミュニティ発見に対するモジュラリティ

4.4.1 ノード分割に対するモジュラリティ

ノード分割に対してはモジュラリティと呼ばれる指標が標準的に用いられてきた。ネットワーク G のノード分割 \mathcal{P} に対し、モジュラリティはその隣接行列 \mathbf{A} に基づいて下記で定義される [14]。

$$Q = \frac{1}{2m} \sum_{C \in \mathcal{P}} \sum_{i,j \in C} (\mathbf{A}_{ij} - \frac{k_i k_j}{2m}) \quad (24)$$

ここで m は G のリンク数であり、分割 \mathcal{P} の要素 C はコミュニティに対応し、 i, j は C に属するノードを表す。式 (24) の値が大きいほどネットワークの分割が良いとされる。なお、式 (24) は以下のようにも表現できる [15]。

$$Q = \frac{1}{\mathbf{1}_n^T \mathbf{A} \mathbf{1}_n} \text{tr}(\mathbf{S}^T (\mathbf{A} - \mathbf{P}) \mathbf{S}) \quad (25)$$

ここで $\mathbf{P}_{ij} = k_i k_j / 2m$ であり、0-1 行列である行列 $\mathbf{S} \in \{0, 1\}^{n \times c}$ は各ノードのコミュニティへのハードな割当てを表現する指示行列である (c はコミュニティ数)。

重み付きネットワークを扱うため隣接行列を非負実数値 $\tilde{\mathbf{A}}$ に拡張した場合でも、モジュラリティは式 (24) で $\tilde{\mathbf{A}}$ を用いて同様に定義される。その場合、式 (24) での正規化係数に対応する $2m$ は重みの総和 ($\sum_{i,j} \tilde{\mathbf{A}}_{ij}$) となる。

4.4.2 ソフトなノード分割に対するモジュラリティ

式 (24) のモジュラリティはノード分割に基づくという意味でノードのハードな割当てに基づき定義されるが、重複コミュニティ発見ではノードは複数のコミュニティに割り当てられる。本稿では、ノードのソフトな割当ての観点から従来のモジュラリティの拡張を考える。

ノードの複数コミュニティへの割当てを表現するため、式 (25) の指示行列を $\tilde{\mathbf{S}} \in \mathbb{R}_+^{n \times c}$ に拡張する。その際、行列 $\tilde{\mathbf{S}}$ の各行が EM 法 [4] などのようにノードの確率的な割当てを表すという意味でソフトな割当てを表現するためには、以下の制約を満たす必要がある。

$$\tilde{\mathbf{S}}_{ij} \geq 0, \quad \forall i, j \quad (26)$$

$$\tilde{\mathbf{S}} \mathbf{1}_c = \mathbf{1}_n \quad (27)$$

上記の $\tilde{\mathbf{S}}$ に基づき、式 (25) のトレースに比例する関数として下記の指標 Q_s を考える。

$$Q_s \propto \text{tr}(\tilde{\mathbf{S}}^T (\mathbf{A} - \mathbf{P}) \tilde{\mathbf{S}}) \quad (28)$$

式 (24) と同様、上記の隣接行列 \mathbf{A} は重み付き隣接行列 $\tilde{\mathbf{A}}$ に拡張可能である。式 (28) はネットワーク G の隣接行列 \mathbf{A} (および $\tilde{\mathbf{A}}$) とノードの割当てに基づく関数であり、 G の情報のみに基づいて定義されることに注意されたい。

式 (28) の値を確定するためには正規化項に対応する比例定数を決める必要がある。式 (25) の比例定数 ($1/\mathbf{1}_n^T \mathbf{A} \mathbf{1}_n$) はノード分割を行うネットワークにおけるリンク情報に基づいて導入されたものであるため、 G を $\tilde{\mathbf{E}}$ や $\tilde{\mathbf{F}}$ などで表現される (重み付き) 線グラフに変換してからノード分割を行う場合には、変換後のネットワークにおけるリンク情報に基づく比例定数が自然な正規化項であると考えられる。

系 4 の式 (22) と (23) は行列 \mathbf{E} と \mathbf{E}_1 [6]、および提案する \mathbf{F} と \mathbf{F}_1 や $\tilde{\mathbf{F}}$ と $\tilde{\mathbf{F}}_1$ を隣接行列とする重み付き線グラフの性質を示している。この性質により、線グラフのような変換後のネットワークにおけるリンクに基づいて正規化項を導入しても、 G の隣接行列に基づいて正規化項を表現することが可能となる。たとえば G が 0-1 隣接行列 \mathbf{A} で表現される場合には $\mathbf{1}_n^T \mathbf{A} \mathbf{1}_n = \mathbf{1}_m^T \mathbf{E} \mathbf{1}_m = \mathbf{1}_m^T \mathbf{F} \mathbf{1}_m = 2m$ となる。同様に、重み付きネットワークに対しても、 $\mathbf{1}_n^T \tilde{\mathbf{A}} \mathbf{1}_n = \mathbf{1}_m^T \tilde{\mathbf{E}} \mathbf{1}_m = \mathbf{1}_m^T \tilde{\mathbf{F}} \mathbf{1}_m = \sum_{i,j} \tilde{\mathbf{A}}_{ij}$ となる。このため、変換後のネットワークの隣接行列が式 (22) や (23) を満たす限り、正規化項は G をどの隣接行列で表現されるネットワークに変換するかに対して不変となる。

上記の議論に基づき、ソフトなノード分割に対するモジュラリティとして以下の指標を提案する。

$$Q_s = \frac{1}{\mathbf{1}_n^T \tilde{\mathbf{A}} \mathbf{1}_n} \text{tr}(\tilde{\mathbf{S}}^T (\mathbf{A} - \mathbf{P}) \tilde{\mathbf{S}}) \quad (29)$$

式 (28) と同様、式 (29) はオリジナルのネットワーク G の表現行列に基づき定義されるため、 G を別のネットワークに変換するか、あるいは直接リンクをコミュニティに割り当てるかに対して不変な定義となっている。また、従来のモジュラリティと同様に、重み付きネットワークに対しては非負実数値の隣接行列 $\tilde{\mathbf{A}}$ を式 (29) で用いる。

4.5 リンク分割に対する指示行列

本稿のアプローチはオリジナルのネットワークにおけるリンクをコミュニティに割り当てるため、式 (29) の Q_s を求めるにはリンク分割で得られるリンクのコミュニティラベルから指示行列 $\tilde{\mathbf{S}}$ を定義する必要がある。

0-1 隣接行列で表現される重みなしネットワークもリンクの重みを 1 と見なすことができるため、以下では非負実数値行列 $\tilde{\mathbf{A}} \in \mathbb{R}_+^{n \times n}$ で表現される重み付きネットワークに対して述べる。リンク分割で得られるコミュニティラベルに基づき、行列 $\tilde{\mathbf{H}} \in \mathbb{R}_+^{m \times c}$ を以下で定義する。

$$\tilde{\mathbf{H}}_{\alpha k} = \begin{cases} w_{\alpha} & \text{if link } \alpha \text{ (with weight } w_{\alpha}) \\ & \text{is assigned to community } k \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

式 (30) の $\tilde{\mathbf{H}}$ から指示行列 $\tilde{\mathbf{S}}$ を以下のように定義する.

$$\tilde{S}_{ik} = \frac{\sum_{\alpha \text{ adjacent to } i} \tilde{\mathbf{H}}_{\alpha k}}{\sum_{k'} \sum_{\alpha \text{ adjacent to } i} \tilde{\mathbf{H}}_{\alpha k'}} \quad (31)$$

最尤法などと同様に式 (31) は各ノードを同じラベルを持つ辺の重みの和に比例してそのコミュニティに割り当てることに対応する. $\tilde{\mathbf{A}}$ は非負実数値行列であるため, 式 (31) で定義する $\tilde{\mathbf{S}}$ は式 (26) を満たし, 同様に式 (27) も満たす. このため, 式 (31) の $\tilde{\mathbf{S}}$ はリンク分割に対する指示行列と見なすことができる.

5. 評価

5.1 実験設定

5.1.1 ネットワーク

提案法を重みを持つ人工ネットワークおよび実ネットワークに適用して評価した. 人工ネットワークはコミュニティ構造を人為的に埋め込んだネットワークであり, スケールフリー性を持つ Barabási-Albert (BA) モデル [3] を用いて付録 A.3 に示す手順で生成した. 付録 A.3 の手順により, 比較的小さな重みで連結するネットワークの中に, 大きな重みを持つ疎なコミュニティに対応する部分ネットワークが埋め込まれるとともに, 各ノードは他のコミュニティのノードとも大きな重みのリンクを持つという意味で重複コミュニティ構造を持つものとなっている.

重み付き実ネットワークとしては表 1, 表 2 に示すものを用い, ネットワークの最大連結成分を用いた. 表 1 のネットワークは GML (graph markup language) で表現され公開されている*4. 表 1 での lesmis はレ・ミゼラブルの共起ネットワークであり, ノードは登場人物である. celegans は線虫の神経ネットワーク [21], netscience はネッ

表 1 実ネットワーク

Table 1 Weighted networks.

dataset	#nodes	#links
lesmis	77	254
celegans	297	2148
netscience	379	914
hep-th	5835	13815

表 2 共著ネットワーク

Table 2 Co-authorship.

dataset	# nodes	#links
Pascal	65	114
IV'04	112	255

*4 <http://www-personal.umich.edu/~mejn/netdata/> (実験では celegans を無向グラフに変換した)

トワーク理論における共著ネットワーク, hep-th は高エネルギー理論のプレプリントを投稿した科学者の共同ネットワークである. 表 2 は共著ネットワークであり*5, 隣接行列で \mathbf{A}_{ij} は著者 i, j の共著論文数を表す.

5.1.2 コミュニティ発見手法

代表的なノード分割に基づくコミュニティ発見手法として以下の手法を用いた.

- 1) labelPropagation [18]
- 2) leadingEigenvector [15]
- 3) walktrap [17]

LabelPropagation [18] はノード間でのラベル伝播を通じて各ノードにコミュニティラベルを割り当てる. Walktrap [17] はネットワーク上での短いランダムウォークから計算される遷移確率に基づいてノード間の距離に基づく手法である. LeadingEigenvector [15] は, グラフカット [20] と同様に式 (25) のようにモジュラリティを行列の固有値として定式化し, 最大固有値に対応する固有ベクトルを用いてコミュニティの分割を行う.

表 1, 表 2 の重み付き無向グラフを変換した重み付き線グラフに上記の手法を適用してリンク分割を行った.

5.1.3 評価指標

線グラフ上でのノード分割から式 (31) の指示行列を計算し, リンク分割に対して式 (29) で提案した Q_s を評価した*6. なお, 5.1.2 項の手法は同じネットワークに対しても分割結果が異なることがあるため, 以下では各ネットワークに対する 10 回平均の結果を報告する.

5.1.4 線グラフに対する隣接行列

オリジナルのネットワークから変換したネットワークに対する隣接行列として以下を用いた.

- 0-1 行列に基づくもの [6]: 式 (2) の \mathbf{E} , 式 (3) の \mathbf{E}_1
- 重み行列に基づくもの: 式 (9) の $\tilde{\mathbf{E}}$, 式 (10) の $\tilde{\mathbf{E}}_1$
- 0-1 行列に基づき, 式 (18) により自己ループを除去したもの: 式 (18) の \mathbf{F} , 式 (23) の \mathbf{F}_1
- 重み行列に基づき, 式 (18) により自己ループを除去したもの: 式 (22) の $\tilde{\mathbf{F}}$, 式 (23) の $\tilde{\mathbf{F}}_1$

5.2 人工ネットワークに対する結果

コミュニティあたりのノード数 $n_c = 50$ として, 付録 A.3 に示す手順でコミュニティ数 c を変えてネットワークを生成した. BA モデルではランダムにネットワークを生成するため同じコミュニティ数 c に対して 10 個ネットワークを生成し, 各ネットワークに対して 5.1.2 項の手法を 10 回適用したため, 計 100 回の平均を報告する.

結果を表 3 に示す. 表 3 の各列は 5.1.4 項の隣接行列, 行は 5.1.2 項の手法に対応し, 行ごとに Q_s の最大値を太

*5 Pascal: <http://analytics.ijs.si/~blazf/pvc/data.html>
IV'04: <http://iv.slis.indiana.edu/ref/iv04contest/>

*6 Q_s の値が大きいかほど良いリンク分割に対応する.

表 3 人工ネットワークに対する結果
Table 3 Results of synthetic weighted networks.

c	method	\mathbf{E}	$\tilde{\mathbf{E}}$	\mathbf{F}	$\tilde{\mathbf{F}}$	\mathbf{E}_1	$\tilde{\mathbf{E}}_1$	\mathbf{F}_1	$\tilde{\mathbf{F}}_1$
3	labelPropagation	0.001	0.063	0.025	0.294	0.000	0.102	0.000	0.045
	leadingEigenvector	0.025	0.171	0.025	0.170	0.029	0.134	0.030	0.135
	walktrap	0.057	0.279	0.064	0.353	0.073	0.365	0.073	0.378
4	labelPropagation	0.001	0.061	0.024	0.289	0.000	0.087	0.000	0.053
	leadingEigenvector	0.024	0.155	0.023	0.156	0.020	0.116	0.020	0.117
	walktrap	0.048	0.287	0.058	0.351	0.072	0.366	0.071	0.387
5	labelPropagation	0.001	0.059	0.024	0.288	0.000	0.063	0.000	0.027
	leadingEigenvector	0.023	0.150	0.023	0.148	0.015	0.100	0.015	0.100
	walktrap	0.048	0.276	0.053	0.353	0.074	0.355	0.077	0.385

字で示す。ほぼすべてのネットワークと手法の組合せ（各行）で、提案する重みを活用し自己ループを除去した隣接行列 ($\tilde{\mathbf{F}}$ および $\tilde{\mathbf{F}}_1$) で最良の結果が得られた。

生成したネットワークでは、リンク数（次数）の観点からはコミュニティ内では疎なため、リンクの重みを考慮しないとコミュニティが全体の中で埋もれてしまう。このため、既存の \mathbf{E} , \mathbf{E}_1 [6] ではコミュニティ構造をとらえることができず、 Q_s は非常に小さな値となった。他方、提案法ではリンクの重みを活用することにより、 \mathbf{E} と $\tilde{\mathbf{E}}$, \mathbf{E}_1 と $\tilde{\mathbf{E}}_1$ を比較すると Q_s を約 1 桁（10 倍）程度向上させることができた。さらに、自己ループの除去も効果的であった。

5.3 実ネットワークに対する結果

重複コミュニティ発見の例として、表 2 の共著者ネットワーク (Pascal) に対し先行研究の \mathbf{E} と提案法の $\tilde{\mathbf{F}}$ で構築した重み付き線グラフに labelPropagation を適用した結果を図 3, 図 4 に示す。図ではリンクの種類と色の組でコミュニティラベルを表している。

先行研究の \mathbf{E} を用いた場合 (図 3) にはリンクが多数のコミュニティに細分化され、たとえば中央のノード 20 や左下の部分グラフに属するノード 8 は多くのコミュニティに割り当てられた。他方、提案法の $\tilde{\mathbf{F}}$ を用いた場合 (図 4), これらのノードは単一のコミュニティに割り当てられたが、ノード 24 ノード 34 などは図 3 と同様に複数のコミュニティに割り当てられた。多くのネットワークでは正解に対応するコミュニティラベルがないため抽出したコミュニティの評価は難しい問題であるが、提案法ではノード 8 を含む左下の部分グラフやノード 20 を含む中央の部分グラフなどの過度な細分化は見られず、ノードの（確率的な）集まりとしてのコミュニティの抽出が行われていた。

表 1, 2 の実ネットワークに対する結果を表 4 に示す。表 3 と同様、表 4 の各列は 5.1.4 項の隣接行列、行は 5.1.2 項の手法に対応し、行ごとに Q_s の最大値を太字で示す。たとえば表 4 では Pascal に対する labelPropagation の行における \mathbf{E} , $\tilde{\mathbf{F}}$ のセルが図 3, 図 4 に対応する。

表 4 の結果から、提案する重み付き線グラフを用いる

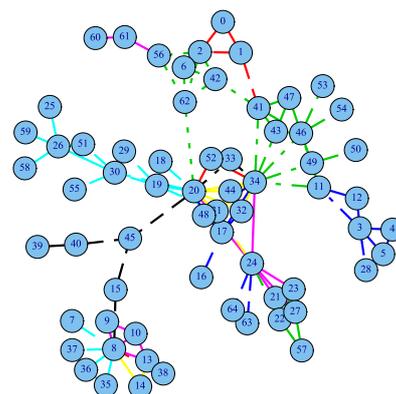


図 3 Pascal に対する \mathbf{E} への labelPropagation の適用結果
Fig. 3 Communities over \mathbf{E} of Pascal with labelPropagation.

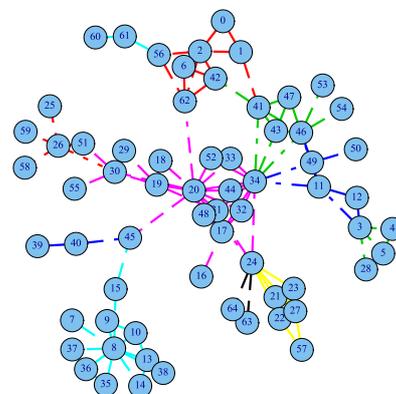


図 4 Pascal に対する $\tilde{\mathbf{F}}$ への labelPropagation の適用結果
Fig. 4 Communities over $\tilde{\mathbf{F}}$ of Pascal with labelPropagation.

ことにより、ノード分割手法の適用を通じて重複コミュニティを発見する際の性能向上が実現できたことが分かる。オリジナルのネットワークにおける重みを活用した隣接行列 ($\tilde{\mathbf{E}}$, $\tilde{\mathbf{E}}_1$), および自己ループを除去した隣接行列 (\mathbf{F} , \mathbf{F}_1 , および $\tilde{\mathbf{F}}$, $\tilde{\mathbf{F}}_1$) で表現される線グラフを用いることにより、従来の隣接行列 (\mathbf{E} , \mathbf{E}_1) と比較してほぼすべてのネットワークと手法に対して性能向上が見られた。

(\mathbf{E} , $\tilde{\mathbf{E}}$) との比較, および (\mathbf{E}_1 , $\tilde{\mathbf{E}}_1$) との比較により, オリジナルのネットワークにおける重みの活用が効果的であることが分かる。同様に, (\mathbf{E} , \mathbf{F}) との比較, および (\mathbf{E}_1 ,

表 4 実ネットワークに対する結果
Table 4 Results of real-world weighted networks.

dataset	method	\mathbf{E}	$\tilde{\mathbf{E}}$	\mathbf{F}	$\tilde{\mathbf{F}}$	\mathbf{E}_1	$\tilde{\mathbf{E}}_1$	\mathbf{F}_1	$\tilde{\mathbf{F}}_1$
lesmis	labelPropagation	0.018	0.131	0.432	0.449	0.410	0.400	0.418	0.374
	leadingEigenvector	0.283	0.343	0.302	0.361	0.390	0.426	0.387	0.419
	walktrap	0.398	0.440	0.458	0.495	0.383	0.488	0.395	0.470
celegans	labelPropagation	0.004	0.084	0.057	0.250	0.000	0.059	0.000	0.032
	leadingEigenvector	0.111	0.174	0.112	0.179	0.207	0.256	0.207	0.256
	walktrap	0.343	0.343	0.335	0.362	0.305	0.316	0.305	0.397
netscience	labelPropagation	0.039	0.131	0.533	0.690	0.727	0.717	0.737	0.763
	leadingEigenvector	0.702	0.731	0.711	0.735	0.762	0.775	0.746	0.755
	walktrap	0.793	0.771	0.795	0.812	0.796	0.811	0.815	0.815
hep-th	labelPropagation	0.043	0.174	0.459	0.596	0.616	0.643	0.621	0.706
	leadingEigenvector	0.588	0.629	0.532	0.583	0.629	0.656	0.630	0.663
	walktrap	0.643	0.628	0.642	0.705	0.667	0.720	0.677	0.743
Pascal	labelPropagation	0.059	0.091	0.460	0.615	0.572	0.591	0.609	0.607
	leadingEigenvector	0.432	0.467	0.437	0.474	0.459	0.503	0.549	0.567
	walktrap	0.528	0.535	0.627	0.634	0.571	0.574	0.569	0.575
IV04	labelPropagation	0.042	0.201	0.577	0.651	0.659	0.677	0.679	0.700
	leadingEigenvector	0.582	0.593	0.582	0.593	0.639	0.646	0.639	0.646
	walktrap	0.698	0.681	0.708	0.688	0.706	0.672	0.708	0.722

\mathbf{F}_1) との比較により, 変換後のネットワークから自己ループを除去することが性能向上に役立つことが分かる. さらに, 表 4 では上記を併用して用いた場合 ($\tilde{\mathbf{F}}$ および $\tilde{\mathbf{F}}_1$) に最良の結果が見られることが多く, 両者を併用することは効果的であるといえる.

5.4 考察

グラフやネットワークを扱う研究ではノード間の類似度をリンクの重みとして表現して活用することが多い [20], [23]. グラフ理論における接続行列を重み付き接続行列 $\tilde{\mathbf{B}}$ に拡張することにより, 変換後のグラフにおいてもオリジナルのネットワークにおける重みを活用することが可能となる. また, ネットワーク上の酔歩に基づく手法 [6] では変換後のグラフが自己ループを持つため変換前後のグラフ構造の対応がとれないという課題に対し, 式 (18) を通じて自己ループを除去することを提案した.

先行研究 [6] を提案した著者らも重み付きネットワークへの応用を提案している [7]. 文献 [7] では重み付き線グラフに対して下記の隣接行列が提案された.

$$\hat{\mathbf{C}}_{\alpha\beta} = \sum_i \tilde{\mathbf{B}}_{i\alpha} \mathbf{B}_{i\beta} (1 - \delta_{\alpha\beta})$$

$$\hat{\mathbf{E}}_{\alpha\beta} = \sum_{i, k_i > 1} \frac{\tilde{\mathbf{B}}_{i\alpha}}{s_i - w_\beta} \mathbf{B}_{i\beta} (1 - \delta_{\alpha\beta})$$

ここで $s_i = \sum_j \tilde{\mathbf{A}}_{ij}$, w_β はリンク β の重み, $\delta_{\alpha\beta}$ はリンク α, β に対するクロネッカーのデルタを表す. 本稿と同様に重み付き接続行列を用いるが, 重み付き接続行列, 従来の 0-1 接続行列, $(1 - \delta_{\alpha\beta})$ との積和で定義されるため, 上記

の $\hat{\mathbf{C}}$ や $\hat{\mathbf{E}}$ は自己ループを含まないが対称行列とはならない. このため, 重み付き無向グラフに対しても $\hat{\mathbf{C}}$ や $\hat{\mathbf{E}}$ で表現される重み付き線グラフは有向グラフとなり, グラフ理論における線グラフの素直な拡張とはなっていない.

ラベルありデータに基づく評価と異なり, コミュニティ発見では正解に対応するコミュニティラベルがないことが多い. モジュラリティはあくまでネットワークの構造に基づく評価指標にすぎず, 提案法は必ずしも「良い」重複コミュニティの発見を保証するものではない. しかし, ノード分割に対しては式 (24) のモジュラリティが (様々な拡張 [9], [13] は提案されているが) ほぼ標準的な指標として用いられることと比べて, 重複コミュニティ発見に対しては標準的な指標はまだ知られていない. 本稿では従来のモジュラリティをノードのソフトな割当てに拡張することを提案し, また, リンク分割に対して拡張した指示行列を提案した. 特に, 式 (29) はオリジナルのネットワークを別のネットワークに変換するか [6], [8], あるいはネットワークの辺を直接コミュニティに割り当てるか [2] に対して不変な定義であるため, 重複コミュニティ発見に対して広く適用可能であると考えられる. ただし, 式 (29) の Q_s の性質に関する詳細な解析は今後の課題である.

本稿では線グラフへの変換を通じたリンク分割に基づく重複コミュニティ発見を扱ったが, リンクに対して最短距離法に基づく階層的クラスタリングを行う手法も提案されている [2]. リンク分割という観点からは本稿よりも直接的なアプローチであるが, 観測されるネットワークには表現されていないリンク間の類似度を別途定義する必要がある. また, コミュニティをノードの密な集まりととらえる

場合には、2章の a) で紹介した手法 [1] はクリークに基づいて密な部分グラフを直接的に抽出できるという利点があるが、計算時間のかかるクリークの同定が必要となる。

6. おわりに

本稿では、複数のコミュニティへの所属を許容する重複コミュニティの発見を実現するために、ネットワークの重みを反映する重み付き線グラフを提案した。ネットワーク(グラフ)の接続関係から定義される従来の線グラフを拡張し、オリジナルのネットワークにおけるリンクの重みを反映して線グラフ上での重みを活用する手法を提案するとともに、変換後のネットワークから自己ループを除去する手法を提案した。さらに、ノード分割に基づく従来のモジュラリティを拡張し、重複コミュニティ発見におけるソフトなノード分割に対するモジュラリティを提案した。

提案法を人工ネットワークおよび実ネットワークに適用して評価し、他手法との比較を通じてその有効性を確認した。特に、本稿で提案したオリジナルのネットワークにおける重みの活用と自己ループの削除がそれぞれ性能向上に役立つことを確認した。今後は発見したコミュニティにおける重複度などの解析を進めて提案した隣接行列をさらに拡張するとともに、隣接行列に対する酔歩からの解釈に取り組む予定である。

謝辞 本研究の一部は文部科学省科研費 (No.24300049)、カシオ科学振興財団、豊田理化学研究所の補助による。有益なご指摘を賜りました査読者の方々に深く謝意を表します。

参考文献

[1] Palla, G., Derényi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, Vol.435, pp.814–818 (2005).

[2] Ahn, Y.-Y., Bagrow, J.P. and Lehmann, S.: Link communities reveal multiscale complexity in networks, *Nature*, Vol.466, pp.761–764 (2010).

[3] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol.286, pp.509–512 (1999).

[4] Dempster, A., Laird, N. and Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*, Vol.39, No.2, pp.1–38 (1977).

[5] Diestel, R.: *Graph Theory*, Springer (2006).

[6] Evans, T. and Lambiotte, R.: Line graphs, link partitions, and overlapping communities, *Physical Review E*, Vol.80, No.1, pp.016105:1–8 (2009).

[7] Evans, T. and Lambiotte, R.: Line Graphs of Weighted Networks for Overlapping Communities, *European Physical Journal*, Vol.B 77, pp.265–272 (2010).

[8] Gregory, S.: Finding Overlapping Communities Using Disjoint Community Detection Algorithms, *Complex Networks*, Springer, pp.47–61 (2009).

[9] 原田恵雨, 鈴木育男, 山本雅人, 古川正志: コミュニティの対応関係を考慮した二部モジュラリティによるコミュ

ニティ分割, コンピュータソフトウェア, Vol.28, No.1, pp.127–134 (2011).

[10] Harville, D.A.: *Matrix Algebra From a Statistician's Perspective*, Springer (2008).

[11] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the Spread of Influence through a Social Network, *Proc. KDD'03*, pp.137–146 (2003).

[12] Mika, P.: *Social Networks and the Semantic Web*, Springer (2007).

[13] Murata, T.: A New Tripartite Modularity for Detecting Communities, コンピュータソフトウェア, Vol.28, No.1, pp.154–161 (2011).

[14] Clauset, A., Newman, M.E.J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol.70, No.6, p.066111 (2004).

[15] Newman, M.: Finding community structure using the eigenvectors of matrices, *Physical Review E*, Vol.76, No.3, p.036104 (2006).

[16] Newman, M.: *Networks: An Introduction*, Oxford University Press (2010).

[17] Pons, P. and Latapy, M.: Computing communities in large networks using random walks, *Journal of Graph Algorithms*, Vol.10, No.2, pp.191–218 (2006).

[18] Raghavan, U., Albert, R. and Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E*, Vol.76, p.036106 (2007).

[19] Roussopoulos, N.D.: A max $\{m, n\}$ algorithm for determining the graph H from its line graph G, *Information Processing Letters*, Vol.2, No.4, pp.108–112 (1973).

[20] von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol.17, No.4, pp.395–416 (2007).

[21] Watts, D.J.: *Small Worlds: The Dynamics of Networks Between Order and Randomness*, Princeton Univ Press (2003).

[22] Whitney, H.: Congruent Graphs and the Connectivity of Graphs, *American Journal of Mathematics*, Vol.54, pp.150–168 (1932).

[23] 吉田哲也: ネットワークのノード情報を考慮した正則化モジュラリティ固有空間法, 情報処理学会論文誌 数理モデル化と応用, Vol.5, No.1, pp.66–72 (2012).

付 録

A.1 定理 2 の証明

Proof.

$$\begin{aligned} \mathbf{1}_m^T \mathbf{C} &= \mathbf{1}_m^T (\mathbf{B}^T \mathbf{B} - 2\mathbf{I}_m) = \mathbf{k}^T \mathbf{B} - 2\mathbf{1}_m^T \\ &= \mathbf{k}^T \mathbf{B} - \mathbf{1}_n^T \mathbf{B} = (\mathbf{k} - \mathbf{1}_n)^T \mathbf{B} \end{aligned}$$

命題 1 の式 (6) より $\mathbf{1}_m^T \mathbf{B}^T = \mathbf{k}^T$ であり, また式 (7) より $2\mathbf{1}_m^T = \mathbf{1}_n^T \mathbf{B}$ である. 式 (11) の右側も, 命題 1 の式 (6), 式 (7) の右側を用いて同様に成り立つ. 他方, 命題 1 を用いて式 (12), 式 (13) は以下のように示せる.

$$\begin{aligned} \mathbf{1}_m^T \mathbf{E} &= \mathbf{1}_m^T \mathbf{B}^T \mathbf{D}^{-1} \mathbf{B} = \mathbf{k}^T \mathbf{D}^{-1} \mathbf{B} = \mathbf{1}_n^T \mathbf{B} = 2\mathbf{1}_m \\ \mathbf{1}_m^T \mathbf{E}_1 &= \mathbf{1}_m^T \mathbf{B}^T \mathbf{D}^{-1} \mathbf{A} \mathbf{D}^{-1} \mathbf{B} = \mathbf{k}^T \mathbf{D}^{-1} \mathbf{A} \mathbf{D}^{-1} \mathbf{B} \\ &= \mathbf{1}_n^T \mathbf{A} \mathbf{D}^{-1} \mathbf{B} = \mathbf{k}^T \mathbf{D}^{-1} \mathbf{B} = \mathbf{1}_n^T \mathbf{B} = 2\mathbf{1}_m \end{aligned}$$

同様に, 命題 1 の右側の式を用いて式 (12), 式 (13) の右

側の式が成り立つ。 □

A.2 定理 3 の証明

Proof. 式 (16) より行列 \mathbf{M}_{wo} の対角要素はすべて零である。 $\mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2}$ は対角行列であるため、行列 \mathbf{M}_{wo} に $\mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2}$ を左と右から (右からの場合にはその転置行列を) 掛けて行列 \mathbf{M}_{wo} の行と列をスケールしても対角要素は不変である。このため、 $\mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{M}_{wo} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2}$ の対角要素はすべて零であり、式 (19) が成り立つ。

\mathbf{M} は対称行列であるため、式 (16) の \mathbf{M}_{wo} も対称行列である。上記と同様に、行列 \mathbf{M}_{wo} に $\mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2}$ を左と右から (右からの場合にはその転置行列を) 掛けても対称性は不変である。このため、式 (18) の第 1 項と第 2 項はともに対称行列であるため、式 (20) が成り立つ。

対角行列 $\mathbf{D} \in \mathbb{R}^{\ell \times \ell}$ に対し、 $\mathbf{D}\mathbf{1}_\ell = \mathbf{d} = \mathbf{1}_\ell \odot \mathbf{d}$ が成り立つ。ここで、ベクトル \mathbf{d} は \mathbf{D} の行和であり、 \odot は要素積である [10]。 $\text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2}$ は対角行列であるため、以下が成り立つ。

$$\begin{aligned} & \mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{M}_{wo} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2} \mathbf{1}_\ell \\ = & \mathbf{D}_M^{1/2} \text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{M}_{wo} \mathbf{1}_\ell \odot (\text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2}) \end{aligned} \quad (\text{A.1})$$

$$= \mathbf{D}_M^{1/2} \mathbf{m}_{wo}^{1/2} \odot (\text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{D}_M^{1/2}) \quad (\text{A.2})$$

$$= \mathbf{D}_M^{1/2} \mathbf{1}_\ell \odot \mathbf{D}_M^{1/2} \quad (\text{A.3})$$

$$= \mathbf{D}_M^{1/2} \mathbf{D}_M^{1/2} \mathbf{1}_\ell \quad (\text{A.4})$$

$$= \mathbf{d}_{D_M} \quad (\text{A.5})$$

ここでベクトル \mathbf{d}_{D_M} は式 (15) の \mathbf{D}_M の行和である。式 (A.1) は上記のように成り立つ。ベクトル \mathbf{m}_{wo} は式 (17) の \mathbf{M}_{wo} の行和であるため、式 (A.2) で $\text{diag}(\mathbf{m}_{wo})^{-1/2} \mathbf{m}_{wo} = \mathbf{m}_{wo}^{1/2}$ となり、また要素積の定義から式 (A.3) が成り立つ。最後に、上記の性質と要素積の性質から式 (A.4) が成り立つ。

さらに、式 (18) の第 1 項に対して $\mathbf{M}_{wo} \mathbf{1}_\ell = \mathbf{M} \mathbf{1}_\ell - \mathbf{D}_M \mathbf{1}_\ell = \mathbf{M} \mathbf{1}_\ell - \mathbf{d}_{D_M}$ となる。このため、 $\mathbf{M} \mathbf{1}_\ell - \mathbf{d}_{D_M}$ と \mathbf{d}_{D_M} の和により式 (21) が成り立つ。 □

A.3 人工ネットワークの生成手順

コミュニティ数 c 、コミュニティあたりのノード数 n_c 、全体ネットワークでのリンク重み w_u 、コミュニティ内の重み比率 $r_m > 1$ に対し、Barabási-Albert (BA) モデル [3] を用いて人工ネットワークを以下の手順で生成した。

1. $n_c \times c$ 個のノードを持つ比較的密なネットワークを BA モデルで生成し*7, リンクの重みを w_u とする。
2. コミュニティに対応する、 n_c 個のノードを持つ比較的

*7 リンクが密になるように、BA モデルでの初期次数は 20 とした。

疎なを BA モデルで c 個生成し*8し、リンクの重みを $w_u \times r_m$ とする。

3. 2. で生成したコミュニティを、相互に重ならないよう、1. で生成したネットワークの隣接行列の対角ブロックに埋め込む。
 4. 各ノードごとに、そのノードが所属しない他のコミュニティを 1 つランダムに選択する。次に、2. で生成時のノード度数分を上限として選択したコミュニティに属するノードを選択し、そのノードとのリンクの重みを 2. と同様に $w_u \times r_m$ とする。
1. で生成した比較的小きな重みで連結するネットワークの中に、2. で生成した大きな重みを持つ疎なコミュニティを 3. で埋め込む。また、4. により各ノードは他のコミュニティとも強く結び付くことになり、生成したネットワークは重複コミュニティの構造を持つことになる。実験では $w_u = 1$, $r_m = 100$ としてコミュニティ内のノードどうしがより強く結び付くようにした。



吉田 哲也 (正会員)

1968 年生。1991 年東京大学工学部航空工学科卒業。1997 年東京大学大学院博士課程修了。工学博士。同年大阪大学大学院基礎工学研究科助手。2001 年大阪大学産業科学研究所助手。2004 年北海道大学大学院情報科学研究科助教授。現在、同大学准教授。主に機械学習、知識獲得、データマイニング等の研究に興味を持つ。人工知能学会会員。

*8 疎な部分ネットワークとなるように 1. での 1/10 として BA モデルでの初期次数を 2 とした。