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)]

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ひと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，本稿ではリンクをコミュニティに割り当て，隣接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で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（グラフ）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，リンク分割などを通じた重複コミュニティ発見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)]

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 networks, and the results indicate that the proposed weighted line graphs 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 [13]．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，各頂点をひとつのコミュニティに割り当てられることが多かった [8], [9], [10], [11], [15]．しかし，実世界のネットワークでは一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等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 [4], [5] を

[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

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) 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することでリンク分割を行う．隣接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 [4] や先行研究 [5] 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を拡張し，リンク分割などを通じた重複コミュニティ発見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 [5], [6]，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．

提案法を重みを持つ人工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提案する重み付き線グラフの有効性を確認した．特に，提案するオリジナルのネットワークにおける重みの活用と重み付き線グラフにおける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 節で関連研究を紹介し，3 節で提案法の詳細を説明する．4 節で他手法との評価実験を報告し，提案法の有効性を議論する．5 節でまとめと今後の展望を述べる．

2. ネットワークからのコミュニティ発見

2.1 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．要素が全て 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$\boldsymbol{a}$から生成される対角行列を $diag(\boldsymbol{a})$ と表記する [*1]．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 $diag(\mathbf{A})$ 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 [*2]．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，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 $\boldsymbol{k} = \mathbf{A}\mathbf{1}_n$ 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 [4].

[*1] $diag(\boldsymbol{a})$ の第 i 対角要素は$\boldsymbol{a}$の第 i 要素である．
[*2] 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．

2.2 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 [8], [9], [10], [11]．また，発見したコミュニティの評価には 3.4 節で述べる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．

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 [*3]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 [6]．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], [5]．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．

2.3 線グラフに基づくリンク分割

上記の **c)** におけるリンク分割を実現する手法として，単

[*3] k-クリークコミュニティは k-1 個のノードを共有して連接する k-クリークの和集合と定義される．

純グラフに対する線グラフの活用が提案された [5]．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 [4])．**単純グラフ** G=(V, E) **の線グラフ** $L(G)$ **とは，**G **の仲の隣接する辺の組** $x, y \in E$ **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 [5]．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．

$$\mathbf{C} = \mathbf{B}^T\mathbf{B} - 2\mathbf{I}_m \tag{1}$$

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）隣接行列として推奨した．

$$\mathbf{E} = \mathbf{B}^T\mathbf{D}^{-1}\mathbf{B} \tag{2}$$

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

ここで $\mathbf{D}$ はネットワークの次数ベクトル$\boldsymbol{k}$ から生成される対角行列である．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．

3. 重み付き線グラフに基づく重複コミュニティ発見

3.1 先行研究での課題

2.3 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り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をリンクの重みとして表現することが多いが [12], [15]，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．

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 [14] があり，G が三角形や 4 頂点から成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むためこの性質は成り立たないことになる．

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り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，単純に比較することができなかった．

本稿では上記の課題に対するアプローチを提案する．3.2 節は **i)**，3.3 節は **ii)**，3.4 節は **iii)** に対応する．

3.2 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 $\boldsymbol{w} \in \mathbb{R}^m$ 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 [12], [15]．

3.2.1 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．

$$\tilde{\boldsymbol{k}} = \tilde{\mathbf{A}}\mathbf{1}_n \tag{4}$$

$$\tilde{\mathbf{D}} = diag(\tilde{\boldsymbol{k}}) \tag{5}$$

式 (4) のベクトル $\tilde{\boldsymbol{k}}$ 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$\boldsymbol{k}$ 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 α=(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．**

$$\mathbf{B}\mathbf{1}_m = \boldsymbol{k} \qquad \tilde{\mathbf{B}}\mathbf{1}_m = \tilde{\boldsymbol{k}} \tag{6}$$

$$\mathbf{1}_n^T\mathbf{B} = 2\mathbf{1}_m^T \qquad \mathbf{1}_n^T\tilde{\mathbf{B}} = 2\boldsymbol{w}^T \tag{7}$$

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

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わかる．

3.2.2 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．

$$\tilde{\mathbf{C}} = \tilde{\mathbf{B}}^T\tilde{\mathbf{B}} - 2diag(\boldsymbol{w}_2) \tag{8}$$
$$\tilde{\mathbf{E}} = \tilde{\mathbf{B}}^T\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{B}} \tag{9}$$
$$\tilde{\mathbf{E}}_1 = \tilde{\mathbf{B}}^T\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{B}} \tag{10}$$

ここで $\boldsymbol{w}_2 = \boldsymbol{w} \odot \boldsymbol{w}$ であり，$\odot$ は行列の要素積を表す [7].

定理 2. 重み付き線グラフに対し以下の性質が成り立つ.

$$\mathbf{1}_m^T\mathbf{C} = (\boldsymbol{k} - \mathbf{1}_n)^T\mathbf{B} \qquad \mathbf{1}_m^T\tilde{\mathbf{C}} = (\tilde{\boldsymbol{k}} - \mathbf{1}_n)^T\tilde{\mathbf{B}} \tag{11}$$
$$\mathbf{1}_m^T\mathbf{E} = 2\mathbf{1}_m^T \qquad \mathbf{1}_m^T\tilde{\mathbf{E}} = 2\boldsymbol{w}^T \tag{12}$$
$$\mathbf{1}_m^T\mathbf{E}_1 = 2\mathbf{1}_m^T \qquad \mathbf{1}_m^T\tilde{\mathbf{E}}_1 = 2\boldsymbol{w}^T \tag{13}$$

証明は省略する．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 3.4 節で述べる．

3.3 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．

$$\boldsymbol{m} = diag(\mathbf{M}) \tag{14}$$
$$\mathbf{D}_M = diag(\boldsymbol{m}) \tag{15}$$
$$\mathbf{M}_{wo} = \mathbf{M} - \mathbf{D}_M \tag{16}$$
$$\boldsymbol{m}_{wo} = \mathbf{M}_{wo}\mathbf{1}_\ell \tag{17}$$

$\boldsymbol{m}$ 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 $\boldsymbol{m}$ 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（対角要素は全て零），$\boldsymbol{m}_{wo}$ 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．

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

ここで $diag(\boldsymbol{m}_{wo})$ は $\boldsymbol{m}_{wo}$ 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 M に対し，式 *(18)* で定義する行列 N に対して以下の性質が成り立つ.

$$diag(\mathbf{N}) = \mathbf{0}_\ell \tag{19}$$
$$\mathbf{N}^T = \mathbf{N} \tag{20}$$
$$\mathbf{N}\mathbf{1}_\ell = \mathbf{M}\mathbf{1}_\ell \tag{21}$$

証明は省略する．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．

2.3 節での行列 $\mathbf{E}$ と $\mathbf{E}_1$，および 3.2.1 節で定義した $\tilde{\mathbf{B}}$ 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.

$$\mathbf{1}_m^T\mathbf{F} = \mathbf{1}_m^T\mathbf{E} = 2\mathbf{1}_m^T \qquad \mathbf{1}_m^T\tilde{\mathbf{F}} = \mathbf{1}_m^T\tilde{\mathbf{E}} = 2\boldsymbol{w}^T \tag{22}$$
$$\mathbf{1}_m^T\mathbf{F}_1 = \mathbf{1}_m^T\mathbf{E}_1 = 2\mathbf{1}_m^T \qquad \mathbf{1}_m^T\tilde{\mathbf{F}}_1 = \mathbf{1}_m^T\tilde{\mathbf{E}}_1 = 2\boldsymbol{w}^T \tag{23}$$

上記の性質の意義は 3.4 節で述べる．

3.4 重複コミュニティ発見に対するモジュラリティ

3.4.1 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 $\mathbf{A}$ に基づいて表現できる [8].

$$Q = \frac{1}{\mathbf{1}_n^T\mathbf{A}\mathbf{1}_n}\,\mathrm{tr}(\mathbf{S}^T(\mathbf{A} - \mathbf{P})\mathbf{S}) \tag{24}$$

ここで $\mathbf{P}_{ij} = k_ik_j/2m$ 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り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 $\tilde{\mathbf{A}}$ を用いて同様に定義される．その場合，正規化係数は重みの総和（$\sum_{i,j}\tilde{\mathbf{A}}_{ij}$）となる．

3.4.2 ソフトなノード分割に対するモジュラリティ

式（??）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り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り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り当てを表現するため，式 (24) 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 $\tilde{\mathbf{S}}$ の各行がノードの確率的な割り当てを表すという意味でソフトな割り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．

$$\tilde{\mathbf{S}}_{ij} \geq 0, \quad \forall i,j \tag{25}$$

$$\tilde{\mathbf{S}}\mathbf{1}_c = \mathbf{1}_n \tag{26}$$

系 4 の式 (22) と (23) は行列 $\mathbf{E}$ と $\mathbf{E}_1$ [5]，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．

$$Q_s = \frac{1}{\mathbf{1}_n^T\mathbf{A}\mathbf{1}_n}\mathrm{tr}(\tilde{\mathbf{S}}^T(\mathbf{A}-\mathbf{P})\tilde{\mathbf{S}}) \tag{27}$$

式 (**??**) と同様，式 (27) 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 (27) で用いる．

3.5 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 (27) 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無しネットワークもリンクの重みを 1 とみ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．

$$\tilde{\mathbf{H}}_{\alpha k} = \begin{cases} w_\alpha & \text{if link } \alpha \text{ (with weight } w_\alpha) \\ & \text{is assigned to community } k \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

式 (28) の $\tilde{\mathbf{H}}$ から指示行列 $\tilde{\mathbf{S}}$ を以下のように定義する．

$$\tilde{\mathbf{S}}_{ik} = \frac{\sum_{\alpha \text{ adjacent to } i}\tilde{\mathbf{H}}_{\alpha k}}{\sum_k\sum_{\alpha \text{ adjacent to } i}\tilde{\mathbf{H}}_{\alpha k}} \tag{29}$$

最尤法などと同様に式 (29) 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 (29) で定義する $\tilde{\mathbf{S}}$ は式 (25) を満たし，同様に式 (26) も満たす．このため，式 (29) の $\tilde{\mathbf{S}}$ はリンク分割に対する指示行列とみなすことができる．

4. 評価

4.1 実験設定

4.1.1 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] を用いて生成した．生成したネットワークでは，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 1 つ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る．

4.1.2 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[11]
2) leadingEigenvector [8]
3) walktrap [10]

LabelPropagation[11] 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 [10] 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 [8] は，グラフカット [12] と同様に式（24）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．

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上記の手法を適用してオリジナルのネットワークのリンク分割を行った．

4.1.3 評価指標

線グラフ上でのノード分割から式 (29) の指示行列を計算し，リンク分割に対して式 (27) で提案した Q_s を評価した [*4]．なお，4.1.2 節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．

4.1.4 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 [5]：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 $\tilde{\mathbf{F}}_1$

[*4] Q_s の値が大きいほど良いリンク分割に対応する．

表 1 人工ネットワークに対する結果
Table 1 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**

4.2 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 **??** 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 4.1.2 節の手法を 10 回適用したため，計 100 回の平均を報告する．

結果を表 1 に示す．表 1 の各列は 4.1.4 節の隣接行列，行は 4.1.2 節の手法に対応し，行ごとに Q_s の最大値を太字で示す．ほぼ全ての人工ネットワークと手法の組み合わせで（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．

生成した人工ネットワークでは，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 $\mathbf{E}$，$\mathbf{E}_1$ [5] 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. 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，リンク分割などを通じた重複コミュニティ発見に対するモジュラリティを提案した．

提案法を重みを持つ人工ネットワークに適用して評価し，他手法との比較を行い，提案する重み付き線グラフの有効性を確認した．特に，オリジナルのネットワークにおける重みの活用と重み付き線グラフにおける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．

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．

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] Diestel, R.: *Graph Theory*, Springer (2006).
[5] Evans, T. and Lambiotte, R.: Line graphs, link partitions, and overlapping communities, *Physical Review E*, Vol. 80, No. 1, pp. 016105:1–8 (2009).
[6] Gregory, S.: Finding Overlapping Communities Using Disjoint Community Detection Algorithms, *Complex Networks*, Springer, pp. 47–61 (2009).
[7] Harville, D. A.: *Matrix Algebra From a Statistican's Perspective*, Splinger (2008).
[8] Newman, M.: Finding community structure using the eigenvectors of matrices, *Physical Review E*, Vol. 76, No. 3, p. 036104 (2006).
[9] Newman, M.: *Networks: An Introduction*, Oxford University Press (2010).
[10] 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).
[11] 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).
[12] von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol. 17, No. 4, pp. 395–416 (2007).
[13] Watts, D. J.: *Small Worlds: The Dynamics of Networks Between Order and Randomness*, Princeton Univ Press (2003).
[14] Whitney, H.: Congruent Graphs and the Connectivity of Graphs, *American Journal of Mathematics*, Vol. 54, pp. 150–168 (1932).
[15] 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).