情報処理学会論文誌 Vol.58 No.6 1246–1257 (June 2017)

ネットワーク上での特徴量分布を考慮したアノテーション付与法

伏見 卓恭[1,a)] 佐藤 哲司[2] 斉藤 和巳[3] 風間 一洋[4]

受付日 2016年10月2日, 採録日 2017年3月3日

概要：ある種のネットワークでは，隣接するノードどうしは互いに類似の特徴量を持つ傾向が発見されている．しかし一般に，ネットワーク上のすべてのノードが類似の特徴量を有するわけではなく，特徴量分布の変化点が存在すると想定できる．すなわち，ネットワーク上の一部のノード群を抽出したとき，その特徴量分布は他のノード群上の分布と大きく異なると仮定できる．さらに，特徴量分布は，分布の中心から距離が離れるに従って減衰していくと考えられる．そこで，各ノードの持つ特性や投稿内容などのコンテンツを特徴ベクトルで表現し，距離減衰重みを導入した合成ベクトルを考える．それらをクラスタリングすることで，特徴量が偏在する意味的凝集性と，ノード群が隣接する構造的凝集性のあるノード群を抽出し，各ノード群に偏在する特徴量を用いてアノテーションを付与する方法を提案する．また，すべてのノードの近隣に同数の類似ノードが存在するわけではなく，ノードによって減衰の程度を調整しなければならない．したがって，減衰重みを制御する指数減衰関数のパラメータをノードごとに推定する方法についても説明する．複数の実データを用いた評価実験により，推定したパラメータとノードの性質について考察する．さらに，クラスタリングして得られるノード群へのアノテーションの妥当性について，意味的凝集性と構造的凝集性を定量的に評価する．

キーワード：特徴量分布，距離減衰，アノテーション，凝集性

Annotation Method Considering Feature Distribution over Network Structure

Takayasu Fushimi[1,a)] Tetsuji Satoh[2] Kazumi Saito[3] Kazuhiro Kazama[4]

Received: October 2, 2016, Accepted: March 3, 2017

Abstract: In some types of real networks, the tendency for some nodes in a network to connect to other nodes that have similar features have been found. However generally, not all nodes on the network possess the similar features and there exist some change points of the feature distributions. That is to say, when we extract a certain node group, feature distribution over this node group differs in those over other nodes. We also assume that the density of feature distribution gradually or rapidly decays according to the distance from center of the distribution. In this paper, we deal with feature vectors of nodes which are obtained from activities or contents the nodes posted and their resultant vectors with introducing distance based decay weights. By clustering these resultant vectors, we attempt to extract semantically and structurally cohesive node groups, then annotate node groups with features which significantly occur in them. Moreover, since the number of similar neighbor nodes differs in each node, we should control the degree of decay with respect to each node. Therefore, we explain an estimation method of exponential decay function paramter which control distance based decay weights. From experimental evaluations using three real networks, we discuss the relation between estimated parameters and nature of nodes. Furthermore, we quantitatively evaluate annotation results of extracted node groups in terms of semantic and structural cohesiveness.

Keywords: feature distribution, distance-based decay, annotation, cohesiveness

© 2017 Information Processing Society of Japan

1. はじめに

TwitterやFacebook，CookpadなどのSNSや，レビューサイト，ブログサイトなどのソーシャルメディアでは，ユーザ間に多くのインタラクションが存在し，ネットワークとしてとらえ分析することにより，様々な知見が得られている [2], [3], [7], [8]．このようなネットワークにおいて隣接関係にあるユーザ間には，共通の特徴があると考えられる [6]．たとえば，化粧品に関するレビューサイトを利用するユーザを，商品に対するレビュー評点を要素とする商品次元のベクトルで表現する．このとき，隣接関係にあるユーザ間のベクトルは比較的類似する傾向にある．しかし，ネットワーク上のすべてのユーザ同士が類似ベクトルを有するとは，通常考えられない．すなわち，ネットワークのどこかで特徴量分布の変化点が存在すると考えられる．さらに，各ノードは隣接関係にあるノードと類似特徴を有する傾向にはあるが，その程度はノードによって様々である．

本研究では，ネットワーク構造上の特徴量分布を仮定して，連結する（構造的凝集性のある）ノード群に対して，その近辺に特徴量が偏在する度合い（意味的凝集性）を定量化する．すなわち，近傍に類似の特徴量を有するノードが多く存在するようなノード群を抽出することを目的とする．そして，ノード群上に偏在する特徴量を用いて，ノード群にアノテーションを付与する手法を提案する．ノード群が連結しておらず，ネットワーク上のいたるところに点在する場合，アノテーション付与が困難である．すなわち，アノテーション付与の観点から，対象ノード群にはある程度の意味的凝集性と構造的凝集性が求められる．ゆえに，提案手法では，意味的凝集性と構造的凝集性をあわせ持つノード群を抽出する．提案手法により，共通あるいは類似の特徴量を有するノード群が，ネットワークのどのあたりに位置するかなどを把握することができるようになる．構造のみに焦点を当てた従来のコミュニティ抽出では，密につながる隣接ノード群は抽出できるが，かならずしも類似の特徴量を有する意味的凝集性は保証されない．そこで，意味的凝集性と構造的凝集性を保証したノード群を抽出する．このようなノード群を抽出することで，ソーシャルメディアなどで同一の興味話題に関する議論がさかんなグループなのか，情報収集を目的として類似しないユーザとつながっているグループなのか，グループの特性評価にもつながる．

具体的には，各ノードの特徴量を特徴ベクトルで表現し，近隣ノードの特徴ベクトルを合成することで，ノード群上の特徴量分布を表現する．ただし，直接隣接するノードだけでなく，数ステップ離れた近隣ノードの特徴ベクトルも考慮する．この際，どの程度離れたノードまで合成すべきかは自明ではない．また一般に，離れたノードの影響も少なからず受けるが，その影響は遠いほど小さくなると考えるのが自然である．そこで，ノード間距離に応じて減衰する重みを乗じながら合成ベクトルを構築していく [16]．減衰度合いは各ノードによって異なるが，類似ノードが近隣に狭く存在するときは減衰を強くし，近隣に広く存在するときは減衰を弱くする．すなわち，各ノードに対して，適切な減衰を実現するパラメータを推定する必要がある [17]．また，減衰関数として指数的減衰やべき乗的減衰が考えられるが，いずれの減衰関数を用いても類似の結果が得られることを確認済みである [18]．本研究では，合成ベクトルの構築法 [16] と減衰パラメータの推定法 [17] をまとめ，各ノードの特徴ベクトルと合成ベクトルのコサイン類似度が最大になるようにノードごとの減衰パラメータを推定する．推定されたパラメータの値は，各ノードが特徴量を共有する近隣ノードの範囲を規定する意味もある．近隣ノードの定義として，CNM法 [1] などにより抽出されたコミュニティを用いることもできるが，コミュニティの境界が特徴量分布の境界である保証はない．また，特徴量分布の境界が厳密に存在するわけでもないため，減衰させることにより連続的に表現する．

本稿の構成は以下のとおりである．2章で関連する既存研究に言及し，3章で本研究の提案手法について詳しく説明する．4章で比較に用いる手法を解説し，5章で複数の実データを用いた評価実験により，推定したパラメータの値とノードの性質について考察し，抽出したノード群およびアノテーション結果について定量的に評価する．最後に，本稿のまとめと今後の課題について述べる．

2. 関連研究

本研究と関連の深い意味的特徴と構造的特徴の両方を用いたコミュニティ抽出手法がいくつか提案されている．Yangら [13] は，各コミュニティに関係ない特徴量を隔離し損じたり，各コミュニティに影響を与えうる特徴量を考慮し損じるなどの生成モデルが持つ欠点を克服するため，PCL-DCモデルと呼ぶ識別モデルを提案した．このモデルでは，各コミュニティにおけるノードの人気度と各ノードのコミュニティ帰属度合いを用いて各ノードペア間のリンク存在確率を推定するPopularity-based Conditional Linkモデルと重み付きの特徴ベクトルを用いてコミュニティ帰

1 東京工科大学コンピュータサイエンス学部
School of Computer, Tokyo University of Technology, Hachioji, Tokyo 192–0982, Japan
2 筑波大学図書館情報メディア系
Faculty of Library, Information and Media Science, University of Tsukuba, Tsukuba, Ibaraki 305–8550, Japan
3 静岡県立大学経営情報学部
School of Management and Information, University of Shizuoka, Shizuoka 422–8526, Japan
4 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University, Wakayama 640–8510, Japan
a) takayasu.fushimi@gmail.com

© 2017 Information Processing Society of Japan

属確率を計算する Discriminative Content モデルを組み合わせている．それにより，有意に高い精度のコミュニティを抽出している．この手法では，各ノードの特徴量を考慮することで，より高い精度でコミュニティを抽出することを目指しているが，特徴量が偏在するノード群を抽出することを目的としていない点で，本研究とは異なる．実際に Yang らの評価実験でも，Modularity など構造的なコミュニティ分割の観点での評価しかなされていない．

Zhou ら [14] は，構造的特性と属性情報の両方を統合した類似度を計算し，SA-Cluster と呼ぶグラフクラスタリングアルゴリズムを提案した．この手法では，まずオリジナルネットワークに属性ノードを追加した augmented グラフを定義する．そして，augmented グラフ上のランダムウォークのパス数により，各ノード間の距離を計算する．この距離に基づき，K-medoids クラスタリングをし，一様な属性を有し密に結合するノード群を抽出する．本研究で対象としている各ノードの特徴量は，ノードの活動などから得られるコンテンツを素性としたものであり，特徴量数は非常に多い．augmented グラフ上のランダムウォークを想定している Zhou らの手法に，多数の特徴量を適用するには明らかに限界がある．実際に Zhou らの評価実験でも，ノードに付されたごく少数の属性情報のみを使用している．

Sun ら [9] は，Email などのコミュニケーションフローから得られる送信者，受信者，メッセージという 3 つ組からテンソルを構築し，テンソル分解を行う．そして，各モードごとに次元をクラスタリングし，相関のあるクラスタをモードをまたいで抽出する．最終的に，抽出したクラスタを階層的に可視化することでデータを分析する手法を提案している．

Kuramochi ら [4] は，与えられたグラフ構造から，極大クリークなどの密なノード集合をノード，クリーク間のリンクをリンクとした交グラフを構築する．交グラフにおけるノード間のリンクには，特徴量より算出する重みを付与する．この際に，交グラフのノード（密なノード集合に相当）内のノードの特徴量を併合し，TF・IDF をかけている．本研究の提案手法でも，周辺ノードの特徴ベクトルを合成するが，距離に従って減衰させながら合成する点，および，減衰の強弱をノードごとに推定する点で異なる．

Mcauley ら [5] は，ソーシャルメディアにおける各ユーザ（ノード）が関与するソーシャルサークルを抽出するために，各ノードのエゴネットワークを対象に，ユーザのプロファイル情報（所属や居住地域，学歴など）の一致性に基づく生成モデルを構築した．最尤推定により提案モデルを学習した結果，各ノードの属するソーシャルサークルをベースラインより高い精度で推定できたことを示している．文献 [5] では，エゴネットワークを対象とすることによりパラメータの次元を減少させているため，本研究が対象としている一般的なネットワークに適用するのは困難だと考えられる．

Wu らは，与えられたネットワークに対し，ノード間の類似度などを重みとした Conceptual ネットワークにおける重みの和が最大で，Physical ネットワーク（実際の接続関係）において連結となる Densest Connected Subgraph を抽出する手法を提案している [11]．この手法では，低次数ノードを枝刈りすることで効率的なアルゴリズムを実現しているが，本研究では，与えられたすべてのノードを構造的・意味的つながりのあるノード群に分割する点で異なる．いわば，コア抽出とコミュニティ分割の違いに相当する．

アノテーション法の関連研究として，小林らの可視化結果へのアノテーション法がある [15]．この手法では，2 次元上に可視化されたオブジェクト群に対して，可視化座標の近接性に基づいて最小全域木を構築する．そして，オブジェクト群が有する特徴量に基づく尤度関数を定義し，尤度関数が最大になるリンクを切断することにより，オブジェクト群を部分集合に分割していく．最終的に得られた部分集合群は，特徴的な特徴量分布を有する部分集合になっている．各部分集合に対し，Z スコアにより統計的有意に出現する特徴量を抽出し，アノテーションのラベルとして採用する．本研究では，ネットワーク構造を対象としている点で，最小全域木を切断する小林らの手法と関連する．しかし，ネットワーク構造は木構造と異なり，一般に 1 本のリンクを切断することによりノード集合を分割することができない．したがって，小林らの手法をそのまま適用することはできない．

また，ナイーブな手法として，各ノードの特徴ベクトルをそのままクラスタリングする方法が考えられるが，ノード間の隣接関係を考慮できない．逆に，Clauset らの CNM 法 [1] により，ノード群をコミュニティに分割する方法も考えられるが，コミュニティ内の異なる特徴の共存に対応できない．

3. 提案手法

提案手法は，ネットワーク構造 $G = (V, E)$，各ノード $u \in V$ の J 次元特徴ベクトル $\mathbf{x}_u$ およびクラスタ数 K を入力とし，以下の手順で全ノードを K 個のノード群に分割し，各ノード群にアノテーションを付与する．

CA1 各ノードを中心に，距離減衰合成ベクトルを構築する；

CA2 距離減衰合成ベクトルを K-medoids 法によりクラスタリングする；

CA3 各ノード群に統計的有意に出現する特徴量を Z スコアにより算出する；

CA4 Z スコア上位の特徴量により，各ノード群にアノテートする；

以下では各手順を詳しく説明する．

© 2017 Information Processing Society of Japan

3.1 距離減衰合成ベクトル

各ノード u に対して，他ノード v へのグラフ距離（最短パス長）を $d(u,v)$ とする．ただし，$d(u,v)=d(v,u)$ であり，$d(u,u)=0$ である．各ノード u を中心に，隣接するノード v の特徴ベクトル $\mathbf{x}_v$ を，指数的減衰関数 $\exp(-\lambda d)$ を用いて距離 d に従って減衰させながら合成ベクトルを構築する；

$$\begin{aligned}\mathbf{y}_u &= \sum_{d=0}^{D_u}\exp(-\lambda d)\sum_{v\in\Gamma_d(u)}\mathbf{x}_v \\ &= \sum_{v\in V}\exp(-\lambda d(u,v))\mathbf{x}_v. \qquad (1)\end{aligned}$$

ここで，$D_u=\max_{v\in V}\{d(u,v)\}$ はノード u から全ノードへ到達するための最大グラフ距離を表し，$\Gamma_d(u)=\{v \mid d(u,v)=d\}$ はノード u から距離 d にあるノードの集合を表し，$\Gamma_0(u)=\{u\}$ とする．このベクトル $\mathbf{y}_u$ を，ノード u を中心とした**距離減衰合成ベクトル**と呼び，オリジナルの特徴ベクトル $\mathbf{x}_u$ と区別する．距離減衰合成ベクトルは，ノード u と直接隣接，あるいは，近くに存在するノードの特徴ベクトルを重みを強くして足し込む．逆に遠くに存在するノードの特徴ベクトルを重みを弱くして足し込んでいる．したがって，ベクトルとしては幾分か均されているが，隣接するノード間で類似のベクトルになりやすくなる．さらに，遠くに類似の特徴ベクトルを有するノードが存在しても，周辺ノードの特徴ベクトルが異れば，異なる距離減衰合成ベクトルになるため，クラスタリングした際に別のクラスタに割り当てられる．

3.2 K-medoids クラスタリング

距離減衰合成ベクトルに対して K-medoids 法 [10] を用いて，全ノード集合 V を K 個のノード群 $\{V_1,V_2,\ldots,V_K\}$ に分割する．

K-medoids 法は，オブジェクト集合 V とその要素 v, $w\in V$ 間の類似度 $\rho(v,w)$ が与えられたとき，以下の目的関数を最大にするような代表オブジェクト集合 P を求める．

$$\mathcal{J}(P)=\sum_{v\in V}\max_{w\in P}\{\rho(v,w)\}.$$

$K=|P|$ 個の代表オブジェクトを抽出し，残りのオブジェクト群を最も類似する代表オブジェクトのクラスタに割り当てることで，オブジェクト集合を K 個のクラスタに分割する．K-medoids 法の解法には反復法や貪欲法があるが，K-means 法と異なり解の一意性が保証される貪欲法を用いる．なお，提案手法において K-medoids 法を採用しているが，他のクラスタリング手法を用いることも可能である．提案手法では，類似度 $\rho(u,v)$ は，距離減衰合成ベクトル間のコサイン類似度とする：

$$\rho(u,v)=\frac{\mathbf{y}_u^T\mathbf{y}_v}{\|\mathbf{y}_u\|\|\mathbf{y}_v\|}.$$

3.3 Z スコア

各ノード群 V_k に有意に多く出現する特徴量を Z スコアを用いて抽出する．ここで，ネットワーク全体の特徴量の分布を以下のように定義する：

$$p_j=\frac{\sum_{u\in V}x_{u,j}}{M}.$$

ここで分母の M は，確率にするための正規化項であり，$M=\sum_{v\in V}\sum_{j=1}^{J}x_{v,j}$ である．また，ノード群 V_k に属するノードの特徴量を以下のように合算する：

$$q_j^{(k)}=\sum_{u\in V_k}x_{u,j}.$$

このとき，ノード群 V_k に対する特徴量 j の Z スコアは以下のように計算する：

$$z_j^{(k)}=\frac{q_j^{(k)}-M_kp_j}{\sqrt{M_kp_j(1-p_j)}}.$$

ここで，$M_k=\sum_{v\in V_k}\sum_{j=1}^{J}x_{v,j}$ である．ノード群 V_k に特徴量 j が出現する期待値 (M_kp_j) に対して有意に多いか少ないかにより，各ノード群の特徴的な特徴量を抽出する．提案手法では，各ノード群ごとに Z スコア上位 H 件の特徴量をアノテーション用のラベルとして採用する．

3.4 パラメータ推定法

上述した距離減衰合成ベクトル $\mathbf{y}_u$ は，近隣ノードの特徴ベクトルを減衰させながら合成させて構築する．これは，近隣ノードの特徴ベクトルは類似する傾向にあり，ノード間距離が離れるにつれ類似度は小さくなるという前提がある．しかし，すべてのノードのまわりに同数の類似ノードが存在するわけではなく，ノードによって減衰の程度を調整しなければならない．本稿では，各ノードの距離減衰合成ベクトルが特徴ベクトルとコサイン類似度の意味で最も類似するように各ノードのパラメータ λ_u を設定する．ノルムを 1 に正規化した特徴ベクトルを $\mathbf{x}_u$ とし，

$$F_u(\lambda_u)=\mathbf{x}_u^T\frac{\sum_{v\in V\setminus\{u\}}\exp(-\lambda_u d(u,v))\mathbf{x}_v}{\left\|\sum_{v\in V\setminus\{u\}}\exp(-\lambda_u d(u,v))\mathbf{x}_v\right\|} \qquad (2)$$

という目的関数を考える．ここでは，距離減衰合成ベクトルの計算に自身の値を合成していない点で式 (1) と異なることに注意されたい．

目的関数 (2) を最大化するようなパラメータ λ_u を求める手順を説明する．ノード u に対して，距離 d にあるノードの特徴ベクトルの合成ベクトルを

$$\mathbf{f}_{u,d}=\sum_{v\in\Gamma_d(u)}\mathbf{x}_v$$

とし，ノード u の特徴ベクトルとの内積を

$$g_{u,d}=\mathbf{x}_u^T\mathbf{f}_{u,d}$$

とする．そして，ノード u からの距離の和が d になる合成

© 2017 Information Processing Society of Japan

ベクトル $\mathbf{f}_{u,d_1}$ と $\mathbf{f}_{u,d_2}$ ペア間の内積を足し合わせ

$$h_{u,d} = \sum_{d_1+d_2=d} \mathbf{f}_{u,d_1}^T \mathbf{f}_{u,d_2}$$

とすると，式 (2) は以下のように書き換えられる：

$$F_u(\lambda_u) = \frac{\sum_{d=1}^{D_u} \exp(-\lambda_u d) g_{u,d}}{\sqrt{\sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}}}.$$

計算の便宜上，対数をとった以下の目的関数を最大にするようなパラメータ λ_u を求める：

$$\log F_u(\lambda_u) = \log \sum_{d=1}^{D_u} \exp(-\lambda_u d) g_{u,d} - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}. \quad (3)$$

ここで，補助的な変数を

$$r_{u,d} = \frac{\exp(-\lambda_u d) g_{u,d}}{\sum_{d'=1}^{D_u} \exp(-\lambda_u d') g_{u,d'}}$$

とすると，式 (3) は以下のように書き換えられる：

$$\log F_u(\lambda_u) = \sum_{d=1}^{D_u} \bar{r}_{u,d} \{(-\lambda_u d) + \log g_{u,d}\} - \sum_{d=1}^{D_u} \bar{r}_{u,d} \log r_{u,d} - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}.$$

パラメータ λ_u に関係のない項などを除くと

$$Q_u(\lambda_u) = -\lambda_u \sum_{d=1}^{D_u} \bar{r}_{u,d} \cdot d - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}$$

となり，1 階微分は，

$$\frac{dQ_u(\lambda_u)}{d\lambda_u} = -\sum_{d=1}^{D_u} \bar{r}_{u,d} \cdot d + \frac{\sum_{d=2}^{2D_u} \exp(-\lambda_u d) \cdot d \cdot h_{u,d}}{2 \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}}$$

となる．ここで，

$$s_{u,d} = \frac{\exp(-\lambda_u d) h_{u,d}}{\sum_{d'=2}^{2D_u} \exp(-\lambda_u d') h_{u,d'}}$$

とすると 2 階微分は，

$$\frac{d^2 Q_u(\lambda_u)}{d\lambda_u^2} = -\frac{1}{2} \left\{ \sum_{d=2}^{2D_u} s_{u,d} \cdot d^2 - \left(\sum_{d=2}^{2D_u} s_{u,d} \cdot d \right)^2 \right\}$$

となり，ブレースの中は 2 次のモーメント同様に非負であるため，2 階微分自体はつねに 0 以下となる．1 階微分が λ_u に関して閉じた形で書けないため，本研究ではニュートン法により，補助変数 $r_{u,d}$ とパラメータ λ_u を反復して更新し求める．この推定されたパラメータは，値が大きいほど距離減衰の効果が強く，近隣ノードの値のみを大きな重みで，遠くのノードの値はほとんど無視する．逆に値が 0 に近いほど距離減衰の効果は弱く，近隣も遠方も同程度の重みで合成する．すなわち，近隣に類似の特徴ベクトルを有するノードが広く存在するか狭く存在するかにより値が異なる．

推定したパラメータによる距離減衰合成ベクトルを用いることで近隣に類似ノードが存在しないノードや近隣のみに類似ノードが存在するノードに関して，必要以上に特徴ベクトルを均すことを避けられるため，各ノードの環境に合わせて合成される．それにより，入力された特徴ベクトル $\mathbf{x}$ をより反映したクラスタリング結果が期待できる．

特徴ベクトルの次元を J，平均ノード間距離を $\bar{D}$ とすると，全ノードのパラメータ推定に要する時間計算量は，$h_{u,d}$ 計算に要する $O(|V| \times 2\bar{D} \times J)$ である．本提案手法では，様々な次元圧縮技術を用いて圧縮したベクトルを用いることにより，J を小さくでき高速化も可能である．

4. 比較に用いる手法

この章では，提案手法と比較するクラスタリング手法について説明する．

4.1 K-medoids クラスタリング

1 つ目は，特徴ベクトルをそのまま K-medoids 法によりクラスタリングする手法である．類似度は提案手法と同様，以下のようにコサイン類似度 $\rho(u,v) = \mathbf{x}_u^T \mathbf{x}_v / \|\mathbf{x}_u\| \|\mathbf{x}_v\|$ を用いる．各ノードの特徴ベクトル $\mathbf{x}$ をそのまま用いている点で，提案手法と異なることに注意されたい．この手法では，構造的凝集性はいっさい考慮せず，意味的凝集性のみを対象としている．以後，Kmed 法と呼ぶ．

4.2 CNM クラスタリング

2 つ目は，Clauset らによって提案された CNM 法である [1]．以下に示す，リンク構造に基づくモジュラリティ

$$\mathcal{Q} = \sum_{k=1}^{K} (e_{kk} - a_k^2)$$

を最大化するようにノードを分割し，コミュニティを抽出する．実際には，ΔQ を最大化することで高速化を図っている．ここで，e_{kk} は，全リンク数に対するコミュニティ k 内のリンク数の比率を表し，$a_k = \sum_{h=1}^{K} e_{kh}$ は，コミュニティ k のノードが持つリンク数の比率を表している．この手法では，意味的凝集性を陽に考慮せず，構造的凝集性を対象としている．以後，CNM 法と呼ぶ．

4.3 MST 分割

3 つ目は，小林らによって提案された MST 分割法である [15]．この手法では，2 次元上に可視化されたオブジェクト群に対して，可視化座標の近接性に基づいて最小全域木を構築する．そして，オブジェクト群が有する特徴量に

© 2017 Information Processing Society of Japan

基づく以下の尤度関数を定義し，尤度関数が最大になるようにリンクを $K-1$ 本切断することにより，オブジェクト群を K 個の部分集合に分割していく．我々の提案手法における Z スコア計算の記述 $q_j^{(k)}$ を利用すると，

$$\mathcal{L} = \sum_{k=1}^{K-1} \sum_{j=1}^{J} q_j^{(k)} \log q_j^{(k)} / q^{(k)}$$

となる．ここで，$q^{(k)} = \sum_{j=1}^{J} q_j^{(k)}$ である．最終的に得られた部分集合群は，特徴的な特徴量分布を有する部分集合になっている．本稿では，隣接関係を反映した可視化結果を出力できるクロスエントロピー法 [12] により可視化する．以後，MST 法と呼ぶ．

5. 評価実験

この章では，3 つの実データを用いて，クラスタリングして得られるノード群へのアノテーションの妥当性について，意味的凝集性と構造的凝集性を定量的に評価する．さらに，推定したパラメータとノードの性質について考察する．

5.1 ネットワークデータ

各ノードの特徴量が明確に分かる以下の 3 つのネットワークを評価実験に用いる．特に，以下の Web ネットワークと Wiki ネットワークは，各ノード（Web ページ/記事）のコンテンツが明快で，提案手法により得られた結果が妥当なのかが直観的に把握できるため，評価実験に採用する．

1 つ目のネットワークは，ある大学のウェブサイトにおけるハイパーリンク構造である[*1]．ウェブページをノード，ハイパーリンクを無向化しリンクとした．各ノードの特徴ベクトルは，ウェブページの内容を形態素解析して得られる名詞群の Bag of Words とした．ノード数は 600，リンク数は 1,299，特徴ベクトルの次元数は 4,412 である．本稿では Web ネットワークと呼ぶ．Web ネットワークのクラスタ係数は，0.54 であり，文献 [7] に示される一般的なソーシャルネットワークと同程度である．

2 つ目のネットワークは，日本語ウィキペディア[*2]の人名の共起ネットワークである．人物記事をノード，5 つ以上の記事において共起関係のある人物間にリンクを張った無向ネットワークである．各ノードの特徴ベクトルは，記事内に出現する名詞群の Bag of Words とした．ノード数は 9,481，リンク数は 122,522，特徴ベクトルの次元数は 20,411 である．本稿では Wiki ネットワークと呼ぶ．Wiki ネットワークのクラスタ係数は，0.55 であり，一般的なソーシャルネットワークと同程度である．

3 つ目のネットワークは，レシピ投稿サイト Cookpad におけるユーザのつくれぽ関係である[*3]．ユーザをノード，つくれぽ関係をリンクとした有向ネットワークを構築し，最大強連結成分を抽出，無向化した．さらにつくれぽ関係が 10 以上あるような関係のみを抽出した．各ノードの特徴ベクトルは，投稿したレシピに使用する食材の使用頻度とした．ノード数は 7,815，リンク数は 40,569，特徴ベクトルの次元数は 4,171 である．本稿では Cookpad ネットワークと呼ぶ．Cookpad ネットワークのクラスタ係数は，0.12 であり，一般的なソーシャルネットワークと比べると低い値である．

5.2 推定パラメータに関する考察

提案手法において距離減衰重みを制御するパラメータを推定した結果 $\hat{\lambda}$ について考察する．本稿では，パラメータの推定値によってランキングし，上位，下位のノードの特徴について定性的に述べる．

Web ネットワークにおいて，$\hat{\lambda} > 1$ となるようなランキング上位は，「教員紹介ページ」が多くを占めていた．これらのページには，教員の経歴や研究内容などが書かれており，隣接する他の教員ページも類似の単語（名詞）を使用した内容になっている．さらに，他の教員ページへは少ないステップでたどり着くことができ，類似の内容のページが近隣に集まり，逆に遠くには類似のページが存在しないため，パラメータの値が大きくなり，強い距離減衰重みを実現させたと考えられる．$\hat{\lambda} \simeq 0$ となるようなランキング下位は「ニュースページ」や「お知らせページ」が多くを占めていた．これらのページには，所属学生や教員の受賞ニュースなどが書かれており，近隣はおろかネットワーク内に類似するページが存在しないため，パラメータの値が小さくなり，弱い距離減衰重みを実現させたと考えられる．

Wiki ネットワークにおいて，$\hat{\lambda} > 1$ となるようなランキング上位は，「タレント芸能人ページ」が多くを占めていた．これらはバンドやコンビ，チームなどに所属している割合が高く，共起関係にある（同じチームに所属）ノードどうしは，類似の単語を使用する傾向にあるため，このような結果になったと考えられる．一方，「俳優や女優などの芸能人ページ」は，タレント芸能人と比べると，パラメータの値は低めに推定された．ウィキペディアの芸能人ページ内には，来歴・人物やエピソードなどが書かれているが，共起関係にある俳優同士（同じドラマに出演など）だからといって，これらに用いられる単語が類似するとは限らないため，それを反映した結果と考えられる．すなわち，チームやバンドの構成員たるタレント芸能人は，非常に類似度の高いノードが近隣に狭く存在する；様々な背景を持つ俳優芸能人は，中程度に類似するノードが近隣にやや広く存在する；という直感に合致した結果となった．$\hat{\lambda} \simeq 0$

*1 法政大学情報科学部（2010 年 8 月時点）http://cis.k.hosei.ac.jp/
*2 https://ja.wikipedia.org/
*3 http://cookpad.com/

© 2017 Information Processing Society of Japan

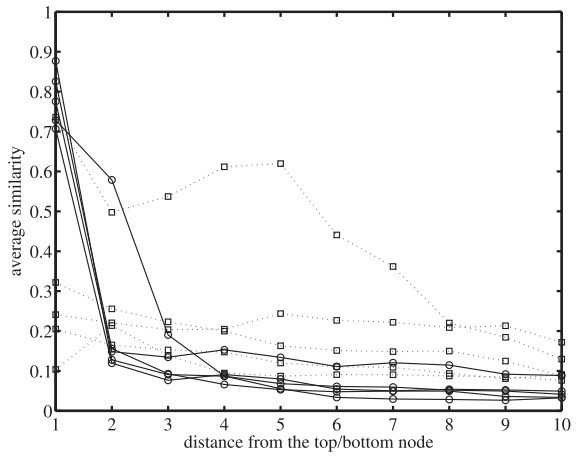


図 1　ノードからの距離に対する平均類似度の変化
Fig. 1　Average similarities w.r.t. the distance from a node.

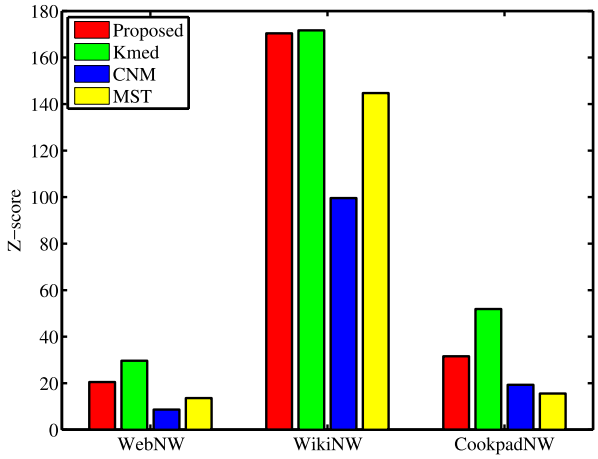


図 2　意味的凝集性の定量評価
Fig. 2　Quantitative evaluation of semantic cohesiveness.

となるようなランキング下位は，「特異な固有名詞を使用するページ」や「内容が少ないページ」などが見られた．これらのページは，ネットワーク内に類似するページが存在しないため，距離に関係なく様々なページの特徴ベクトルを合成することで，自身の特徴ベクトルとのコサイン類似度を高くしようとする．そのためパラメータの値が小さくなり，弱い距離減衰重みを実現させたと考えられる．

Cookpad ネットワークにおいて，$\hat{\lambda} > 1$ となるようなランキング上位は，「比較的小規模なカテゴリコミュニティに属するユーザ」が多くを占めていた．具体的には，「タレ」，「幼児食」，「ドリンク」などがあげられる．これらはカテゴリに投稿するユーザは，比較的小規模ながらコミュニティを形成しており，コミュニティ内では類似の食材を利用するユーザが多く存在する．一方で，コミュニティの外には類似するユーザがあまり存在しないため，距離減衰重みを制御するパラメータの値が大きく推定されたと考えられる．$\hat{\lambda} \simeq 0$ となるようなランキング下位は，「様々なカテゴリのレシピを投稿するユーザ」が多くを占めていた．様々な食材を使用するため，近隣だけでなく遠方のノードの特徴ベクトルをも合成しなければ高いコサイン類似度を得ることができない．いわば，コミュニティにあまり染まらないノードである．そのため，距離減衰重みを制御するパラメータの値が小さく推定されたと考えられる．これらの結果から，推定パラメータ $\hat{\lambda}$ の値は，コミュニティのサイズや所属するノードの特徴ベクトルのばらつき度に依存することが示唆された．

図 1 は，横軸があるノード u からの距離 d，縦軸がノード u から距離 d にあるノード群 $v \in \Gamma_d(u)$ との平均類似度 $\mathrm{avgsim}(u, d)$ を示すグラフである．具体的には，パラメータが高い値，あるいは，低い値に推定されたノードをそれぞれ 5 つずつ選定し，各ノード u と距離 d にあるノードの特徴ベクトル間の類似度を平均した値である：

$$\mathrm{avgsim}(u, d) = \frac{1}{|\Gamma_d(u)|} \sum_{v \in \Gamma_d(u)} \frac{\mathbf{x}_u^T \mathbf{x}_v}{\|\mathbf{x}_u\| \|\mathbf{x}_v\|}.$$

図 1 より，上述の考察のようにパラメータの値が高く推定されたノードの平均類似度（○実線）は，近傍のみに類似ノードが多いため，距離の増大にともない急激に減少する．反対に，パラメータの値が低く推定されたノードの平均類似度（□点線）は，近傍にも類似ノードが少なく，減少は少ない．本稿では，Wiki ネットワークの結果のみを示したが，Web ネットワークでも，Cookpad ネットワークでも同様の傾向が確認できた．各ネットワークの各ノードに対して推定されたパラメータを用いて，距離減衰合成ベクトルを計算し，ノード間の類似度を計算する．

5.3　アノテーション付与結果

比較手法と提案手法について，意味的凝集性と構造的凝集性の観点から定量的に比較する．あるクラスタに有意に出現する特徴量があれば，そのクラスタに意味的凝集性があるといえる．抽出したクラスタに意味的凝集性があるか，3.3 節で説明した Z スコアを用いて定量的に評価した結果を**図 2** に示す．実際には，各クラスタにおける Z スコアの降順ランキング i 位の特徴量 Z スコアを $z_{(i)}^{(k)}$ とし，10 クラスタ（$K = 10$）それぞれにおける上位 10 件（$I = 10$）の特徴量 Z スコアの平均をプロットした：

$$Z = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{I} \sum_{i=1}^{I} z_{(i)}^{(k)}$$

なお，本稿では $K = 10$ の結果のみを示しているが，クラスタ数を変えても，意味的凝集性と構造的凝集性の両方を反映できる提案手法の有効性に矛盾のない結果が得られている．図 2 を見ると，どのネットワークにおいても，Kmed 法が最も高い値を示している．これは，特徴ベクトルを直接クラスタリングしているので当然の結果であるが，提案手法も次いで高い値を示している．提案手法では，距

© 2017 Information Processing Society of Japan

離減衰合成ベクトルをクラスタリングしているため，このような結果が得られることは自明ではない．反対に，対象ネットワークの構造に依存するが，CNM 法では意味的凝集性は見られない傾向にある．CNM 法と同様に，ネットワークの隣接関係が影響する MST 法でも，意味的凝集性は見られない傾向にある．

次に，抽出したクラスタにノード間の隣接性を表す構造的凝集性があるかを定量的に評価した結果を**図 3** に示す．実際には，割り当てたクラスタ番号を各ノードの属性値として，以下のように Assortative 係数 [6] を計算することにより評価した：

$$R = \frac{\sum_{k=1}^{K} e_{kk} - \sum_{k=1}^{K} a_k^2}{1 - \sum_{k=1}^{K} a_k^2}.$$

ここで，e_{kh} は，全リンクに対するクラスタ k に属するノードと h に属するノードがつながっている割合であり，$a_k = \sum_{h=1}^{K} e_{kh}$ は，クラスタ k に属するノードの割合を表す．図 3 を見ると，ノードの隣接関係のみを考慮している CNM 法が最も高い値を示しているが，この結果は自明である．次いで提案手法も高い値を示しており，同一クラスタのノードどうしが隣接関係にあることが，可視化結果からだけでなく定量的にも示された．反対に，隣接関係をいっさい考慮していない Kmed 法は最も低い値を示している．これらの結果から，提案手法は本稿の目的である，隣接関係を考慮した特徴的な意味を有するノード群を抽出できていることが分かった．

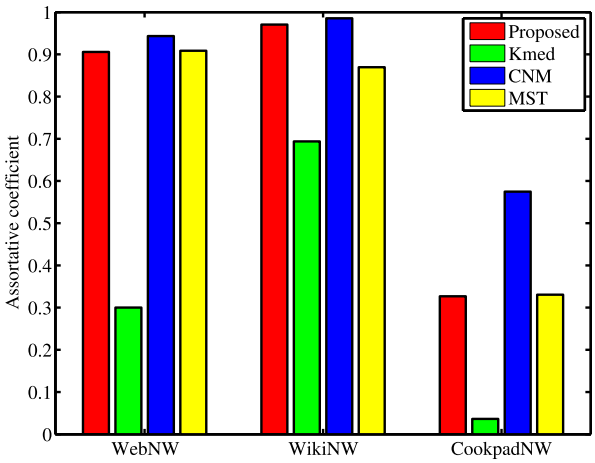


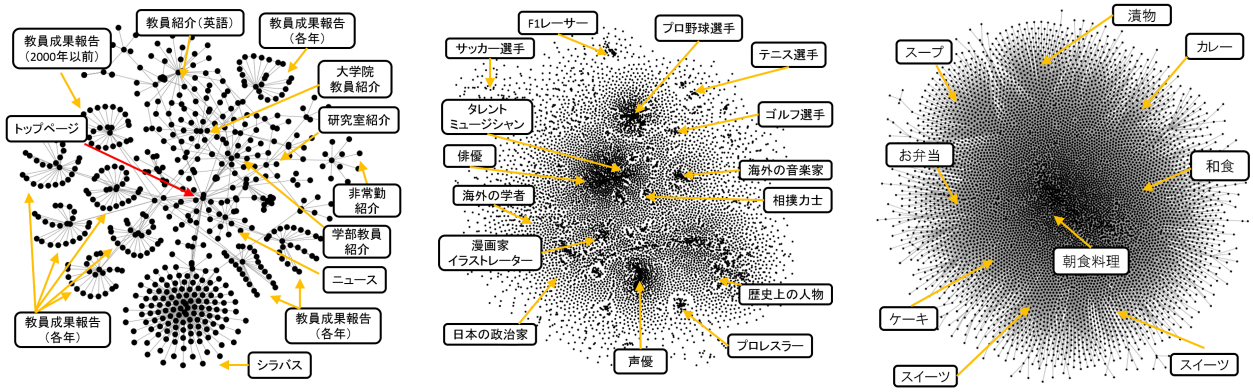
図 3　構造的凝集性の定量評価
Fig. 3　Quantitative evaluation of structural cohesiveness.

次に，提案手法によるクラスタリングおよびアノテーションを定性的に評価する．**図 4** に，可視化結果に人手でアノテーションを付与したものを示す．また可視化には，以下に示すクロスエントロピー $E_{u,v}$ の和 $\sum_{u \in V} \sum_{v \in V \setminus \{u\}} E_{u,v}$ を最小化させることにより，ノードの布置座標を計算するクロスエントロピー法 [12] を用いた：

$$E_{u,v} = -a_{u,v} \ln \rho(u,v) - (1 - a_{u,v}) \ln(1 - \rho(u,v)).$$

ここで，$a_{u,v}$ は隣接行列の u 行 v 列目の値，すなわち，ノード u と v がつながっていれば 1，つながっていなければ 0 である．また，$\rho(u,v) = \exp(-\|\mathbf{x}_u - \mathbf{x}_v\|^2/2)$ は，ノード u と v の布置座標ベクトル $\mathbf{x}_u$，$\mathbf{x}_v$ 間のユークリッド距離の自乗 $\|\mathbf{x}_u - \mathbf{x}_v\|^2$ が小さいほど値が大きくなるような減衰関数である．この可視化手法はバネモデル法などと異なり，クラスタ係数がある程度高いような密な構造を持つネットワーク可視化の際，構造がつぶれにくいという利点があり，隣接関係にあるノードどうしは平面上で近くに布置される．

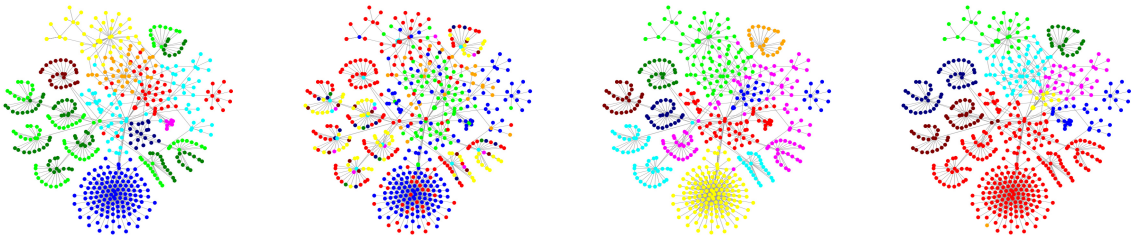
図 5 に，$K = 10$ とした際の Web ネットワークに対するノード分割結果を示す．同一色のノードは，同一ノード群に分割されたことを意味する．以下，分割されたノード群をクラスタと呼ぶ．可視化結果から分かることとして，(b) では，隣接するノードでも異なるクラスタ（色）が割り当てられており，本稿の目的である構造的凝集性を考慮できていない．(c) では，ネットワーク構造上綺麗に分割できているが，各クラスタに意味的凝集性があるという保証はない．(d) では，可視化座標の近接性に依存している



(a) Web ネットワーク　(b) Wiki ネットワーク　(c) Cookpad ネットワーク

図 4　人手による可視化結果へのアノテーション
Fig. 4　Manual annotation for visualization results.

© 2017 Information Processing Society of Japan

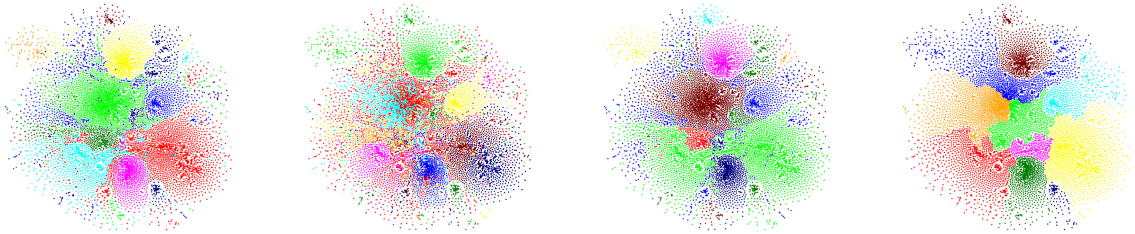


(a) 提案手法　(b) Kmed 法　(c) CNM 法　(d) MST 法

図 5　Web ネットワークに対するクラスタリング結果：$K = 10$
Fig. 5　Clustering results of Web network: $K = 10$.

表 1　Web ネットワークのアノテーション特徴量
Table 1　Annotation characteristics of Web network.

	ノードの色	1 位	2 位	3 位	4 位	5 位
クラスタ 1	#ff0000	科学	情報	研究	コンピュータ	学科
クラスタ 2	#00ff00	node	algorithm	image	virtual	convert
クラスタ 3	#0000ff	科目	授業	理解度	春	秋
クラスタ 4	#ffff00	research	year	student	advisor	English
クラスタ 5	#ff00ff	課程	セミナー	指導	単位	博士
クラスタ 6	#00ffff	画像	映像	描画	認識	動画
クラスタ 7	#ffa500	領域	研究	開発	プロジェクト	非常勤
クラスタ 8	#800000	page	proceeding	transaction	press	edition
クラスタ 9	#008000	model	browser	object	agent	function
クラスタ 10	#000080	掲載	受賞	時間割	更新	開催



(a) 提案手法　(b) Kmed 法　(c) CNM 法　(d) MST 法

図 6　Wiki ネットワークに対するクラスタリング結果：$K = 10$
Fig. 6　Clustering results of Wiki network: $K = 10$.

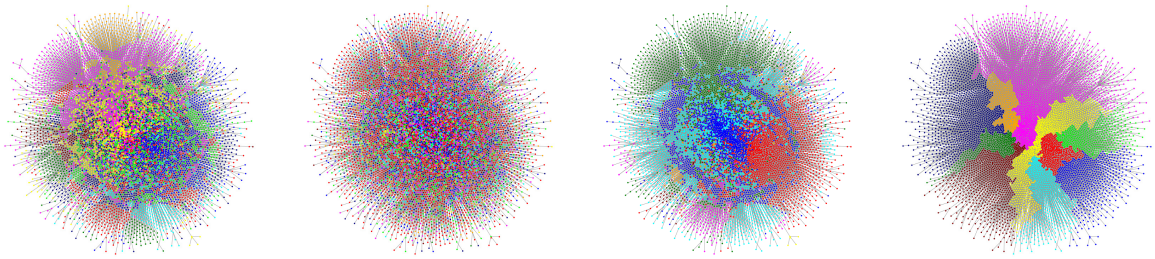
ため，(c) のコミュニティ抽出とも異なる結果が得られた．また，(c) と (d) に対する Z スコア上位の特徴量には共通点がなく，アノテーションには不向きな特徴量であった．これらと比較して推定パラメータを用いた (a) の提案手法では，隣接関係を考慮しているため，連結したノード群単位で同一のクラスタに割り当てられており，かつ，意味的に類似するノードが同一のクラスタに割り当てられている．提案手法により抽出された特徴量を**表 1** に示す．正解ラベルと図 5 (a) のノードの色を念頭において見ると，どのクラスタに付されたアノテーション特徴量も，ある程度クラスタに属するノードの特色を表すものが抽出されている．特に，CNM 法では「教員成果報告ページ」として同一クラスタに分けられていたノード群が，提案手法では第 2 クラスタのような画像処理系の教員成果報告ページと，第 9 クラスタのような Web 系の教員成果報告ページに分けられている．提案手法は，意味的凝集性を考慮するため，近隣に存在していても特徴ベクトルが大きく異なれば分離することが可能である．

図 6 に，$K = 10$ とした際の Wiki ネットワークに対するノード分割結果を示す．可視化結果から分かることとして，(b) では，隣接するノードでも異なるクラスタ（色）が割り当てられており，本稿の目的である構造的凝集性を考慮できていない．(c)，(d) では，ネットワーク構造上綺麗に分割できているが，意味的な部分で分割されている保証はない．しかし，共起ネットワークの性質上，ある程度の意味的凝集性のあるノード群が抽出されたように見受けられる．これらと比較して推定パラメータを用いた (a) の提案手法では，(b) の意味的凝集性と (c) の構造的凝集性の

© 2017 Information Processing Society of Japan

表 2　Wiki ネットワークのアノテーション特徴量
Table 2　Annotation characteristics of Wiki network.

	ノードの色	1 位	2 位	3 位	4 位	5 位
クラスタ 1	#ff0000	旧暦	幕府	徳川	藤原	元年
クラスタ 2	#00ff00	テレビ	出演	フジテレビ	ドラマ	番組
クラスタ 3	#0000ff	交響	ピアノ	音楽	フランス	ローマ
クラスタ 4	#ffff00	野球	選手	プロ	投手	本塁打
クラスタ 5	#ff00ff	アニメ	声優	ガンダム	戦士	ロボット
クラスタ 6	#00ffff	内閣	議員	選挙	大臣	大統領
クラスタ 7	#ffa500	サッカー	ワールドカップ	得点	代表	リーグ
クラスタ 8	#800000	グランプリ	ドライバー	レース	モナコ	フェラーリ
クラスタ 9	#008000	漫画	連載	文庫	作品	手塚
クラスタ 10	#000080	場所	優勝	王座	オープン	プロレス



(a) 提案手法　(b) Kmed 法　(c) CNM 法　(d) MST 法

図 7　Cookpad ネットワークに対するクラスタリング結果：$K = 10$
Fig. 7　Clustering results of Cookpad network: $K = 10$.

表 3　Cookpad ネットワークのアノテーション特徴量
Table 3　Annotation characteristics of Cookpad network.

	ノードの色	1 位	2 位	3 位	4 位	5 位
クラスタ 1	#ff0000	薄力粉	砂糖	牛乳	強力粉	マーガリン
クラスタ 2	#00ff00	オリーブオイル	塩	ニンニク	白ワイン	植物油
クラスタ 3	#0000ff	酒	コショウ	醤油	だし汁	ごま油
クラスタ 4	#ffff00	食パン	E マフィン	マヨネーズ	ベーコン	卵
クラスタ 5	#ff00ff	醤油	みりん	麺つゆ	酒	だし汁
クラスタ 6	#00ffff	バニラオイル	無塩バター	全粒粉	上白糖	牛乳
クラスタ 7	#ffa500	塩麹	きゅうり	ナス	大根	甘酢
クラスタ 8	#800000	海苔	ご飯	チーズ	ハム	ウィンナー
クラスタ 9	#008000	グラニュー糖	生クリーム	卵黄	薄力粉	無塩バター
クラスタ 10	#000080	じゃがいも	ウスターソース	ルー	玉ねぎ	カレー粉

両方を考慮できているように見える．実際に抽出されたアノテーション特徴量を**表 2** に示す．正解ラベルと図 6 (a) のノードの色を念頭において見ると，どのクラスタに付されたアノテーション特徴量も，ある程度クラスタに属するノードの特色を表すものが抽出されている．特に，CNM 法では「歴史上の人物」と「政治家」が同一クラスタに分けられていたノード群が，提案手法では第 1 クラスタの「歴史上の人物」と第 6 クラスタの「政治家」に分けられている．

図 7 に，$K = 10$ とした際の Cookpad ネットワークに対するノード分割結果を示す．可視化結果から分かることとして，おおむね上記 2 つのネットワークの結果と矛盾のない結果が得られた．アノテーション特徴量を**表 3** に示す．こちらの結果も，「スイーツ」の食材が多い第 1 クラスタ，「朝食料理」の食材が多い第 4 クラスタ，「つけもの」の食材が多い第 7 クラスタ，「お弁当」の食材が多い第 8 クラスタ，「ケーキ」の食材が多い第 9 クラスタ，「カレー」の食材が多い第 10 クラスタなど，アノテーションとして有用な特徴量が抽出されている．

6.　おわりに

本稿では，距離減衰重みを導入した合成ベクトルを考え，ノード間の隣接関係を考慮した，ノード群へのアノテーション法を提案した．距離減衰を実現するに際し，各ノー

© 2017 Information Processing Society of Japan

ドごとに異なる周辺ノードとの親和性や異端児度を意味するパラメータを数理的枠組みのもとで推定した．実ネットワークを用いた評価実験より，推定パラメータは，ある程度妥当にノードの異端児度を示していることを確認した．さらに，推定パラメータを用いて距離減衰重み付き合成ベクトルを構築し，アノテーション付与を実現した．アノテーション法としての評価では，Kmed 法で抽出できる意味的凝集性と CNM 法で抽出できる構造的凝集性の両側面をあわせ持っており，隣接関係にあるノード群に対し適切なアノテーションを付与できることが示唆された．今後は，人工データを用いて提案手法の精度などを定量的に評価するとともに，大規模ネットワークでも耐えられるようにアルゴリズムの高速化にも着手する予定である．

謝辞 本研究は，JSPS 科研費（No.16H02904）の助成を受けたものである．本研究の評価に際し，クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した．ここに記して謝意を示す．

参考文献

[1] Clauset, A., Newman, M.E.J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol.70, No.6, 066111 (online), DOI: 10.1103/PhysRevE.70.066111 (2004).

[2] Domingos, P.: Mining Social Networks for Viral Marketing, *IEEE Intelligent Systems*, Vol.20, No.1, pp.80–82 (2005).

[3] Gruhl, D., Guha, R., Liben-Nowell, D. and Tomkins, A.: Information Diffusion Through Blogspace, *Proc. 13th International Conference on World Wide Web*, New York, NY, USA, ACM, pp.491–501 (2004).

[4] Kuramochi, T., Okada, N., Tanikawa, K., Hijikata, Y. and Nishida, S.: Community Extracting Using Intersection Graph and Content Analysis in Complex Network, *Proc. 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, WI-IAT '12*, Washington, DC, USA, IEEE Computer Society, Vol.1, pp.222–229 (2012).

[5] Mcauley, J. and Leskovec, J.: Learning to Discover Social Circles in Ego Networks, *Advances in Neural Information Processing Systems 25*, Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q. (Eds.), pp.539–547, Curran Associates, Inc. (2012).

[6] Newman, M.E.J.: Assortative mixing in networks, *Structure*, Vol.2, No.4, p.5 (online), DOI: 10.1103/PhysRevLett.89.208701 (2002).

[7] Newman, M.E.J.: The Structure and Function of Complex Networks, *SIAM Review*, Vol.45, pp.167–256 (2003).

[8] Newman, M.E.J., Forrest, S. and Balthrop, J.: Email networks and the spread of computer viruses, *Physical Review E*, Vol.66, 035101 (2002).

[9] Sun, J., Papadimitriou, S., Lin, C.-Y., Cao, N., Liu, S. and Qian, W.: MultiVis: Content-Based Social Network Exploration through Multi-way Visual Analysis, *SIAM International Conference on Data Mining*, SIAM, pp.1064–1075 (2009).

[10] Vinod, H.: *Integer Programming and The Theory of Grouping*, Vol.64, An Official Journal of the American Statistical Association (1969).

[11] Wu, Y., Jin, R., Zhu, X. and Zhang, X.: Finding Dense and Connected Subgraphs in Dual Networks, *Proc. IEEE 31st International Conference on Data Engineering* (*ICDE2015*), pp.915–926 (2015).

[12] Yamada, T., Saito, K. and Ueda, N.: Cross-entropy directed embedding of network data, *Proc. 20th International Conference on Machine Learning* (*ICML03*), pp.832–839 (2003).

[13] Yang, T., Jin, R., Chi, Y. and Zhu, S.: Combining Link and Content for Community Detection: A Discriminative Approach, *Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, New York, NY, USA, ACM, pp.927–936 (2009).

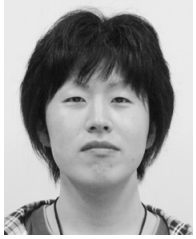
[14] Zhou, Y., Cheng, H. and Yu, J.X.: Graph Clustering Based on Structural/Attribute Similarities, *Proc. VLDB Endow.*, Vol.2, No.1, pp.718–729 (2009).

[15] 小林えり，斉藤和巳，池田哲夫，大久保誠也：L1 埋め込みによるアノテーション付き可視化法，第 7 回 Web とデータベースに関するフォーラム（WebDB Forum2014）(2014).

[16] 伏見卓恭，佐藤哲司，斉藤和巳，風間一洋：ノード間の隣接関係を考慮したアノテーション法の提案，ネットワークが創発する知能研究会（JWEIN2015）(2015).

[17] 伏見卓恭，佐藤哲司，斉藤和巳，風間一洋：距離減衰重みを導入したノード群へのアノテーション付与法，第 8 回 Web とデータベースに関するフォーラム（WebDB Forum2015）(2015).

[18] 伏見卓恭，佐藤哲司，斉藤和巳，風間一洋：構造特性と意味特性を考慮した中心性指標の提案，第 8 回データ工学と情報マネジメントに関するフォーラム（DEIM2016）(2016).



伏見 卓恭 （正会員）

2014 年静岡県立大学大学院経営情報イノベーション研究科博士後期課程修了．同年静岡県立大学大学院経営情報学部客員研究員．2015 年筑波大学図書館情報メディア系特別研究員 (PD)．2017 年より東京工科大学コンピュータサイエンス学部助教．複雑ネットワーク，可視化の研究に従事．博士（学術）．人工知能学会，日本データベース学会各会員．

佐藤 哲司 （正会員）

1980 年山梨大学工学部電子工学科卒業．同年日本電信電話公社 (現, NTT) 武蔵野電気通信研究所に入所．1994 年工学博士（大阪大学）．2007 年より筑波大学図書館情報メディア系教授．情報検索・知識発見，社会ネットワーク分析，社会インタラクションに興味を持つ．電子情報通信学会，日本データベース学会各会員．

© 2017 Information Processing Society of Japan

斉藤 和巳　（正会員）

1985 年慶應義塾大学理工学部数理科学科数学専攻卒業．1998 年東京大学博士（工学）．2007 年より静岡県立大学経営情報学部教授．複雑ネットワークの研究に従事．電子情報通信学会，人工知能学会，日本神経回路学会，日本応用数理学会，日本行動計量学会，日本データベース学会各会員．

風間 一洋　（正会員）

1988 年京都大学大学院工学研究科精密工学専攻修士課程修了．同年日本電信電話（株）入社．2012 年より和歌山大学システム工学部教授．Web 情報検索，Web マイニングの研究に従事．博士（情報学）．人工知能学会，日本ソフトウェア科学会，日本データベース学会，ACM 各会員．

© 2017 Information Processing Society of Japan