

ネットワーク構造解析に基づく重複クラスタ抽出法

高木 允[†] 田村 慶一^{††}
森 康真^{††} 北上 始^{††}

本研究では、Newman らの提案するネットワーク構造解析に基づくクリスピーなクラスタリングアルゴリズムを改良し、ひとつのノードが複数のクラスタに重複してクラスタリングされることを許したグラフのクラスタリング手法を提案する。提案する手法は、一度 Newman らの提案するアルゴリズムを用いてグラフのクラスタリングを行った後、それぞれのクラスタをひとつのノードとした縮約グラフの作成を行う。縮約グラフを再度 Newman らの提案するアルゴリズムを用いてクラスタリングし、重複してクラスタリングされるノードを識別する。重複を許したクラスタリングを行うことで、クリスピーなクラスタリングに比べ、柔軟なクラスタリングを行うことができる。ブログのトラックバックデータを用いた評価実験を行い、提案手法の有効性を示すことができた。

A Extraction Method of Overlapping Cluster based on Network Structure Analysis

MAKOTO TAKAKI,[†] KEIICHI TAMURA,^{††} YASUMA MORI^{††} and HAJIME KITAKAMI^{††}

In this paper, we propose a method of overlapping cluster based on network structure analysis which improves the clustering algorithm proposed by Newman et al. Newman's clustering algorithm is the crisp clustering algorithm. In the proposed technique, first, we cluster the nodes using the Newman's algorithm. Then, we make the contraction graph which is considered a cluster as a node. In addition, we cluster the created contraction graph by using the Newman's clustering algorithm again and identify the overlapping nodes. Overlapping clustering is more flexible than crisp clustering. The experimental results using the trackback data based on blog represented efficacy of proposed technique.

1. はじめに

近年、ノードと辺で構成されるグラフを分類するためのクラスタリング手法が様々研究されている。グラフのクラスタリングは、大量の複雑な情報を整理し、有用な知識を発見するための有効な手法である。

クラスタリング手法には、大きく分けて2種類の手法が存在する。ひとつは、ノードをノードの属性により分類するクラスタリング手法である。このクラスタリング手法を「内容参照型クラスタリング」と呼ぶことにする。もうひとつは、ノードが辺によって繋がっているという事象によりノードを分類するクラスタリング手法である。このクラスタリング手法を「ネットワーク構造解析に基づくクラスタリング」と呼ぶこと

にする。

内容参照型クラスタリングにおいては、グラフの種類により、それぞれのノード間の類似度などを算出する必要があり、多種多様なグラフを扱う場合、効率性が低下してしまう。ネットワーク構造解析に基づくクラスタリングにおいては、ノード間に辺が張られているという事象を基にクラスタリングを行うため、内容参照型のクラスタリング手法のようにノードの類似度や属性を考慮しなくてもよい。そこで、本研究では特にネットワーク構造解析に基づくクラスタリング手法に焦点を当てる。

ネットワーク構造解析に基づくクラスタリング手法として、Newman らが大規模なグラフの高速なクラスタリング手法を提案している。Newman らは Q という評価指標を用いたクラスタリング手法を提案している。 Q の値が高ければよいクラスタリング結果であると判断し、複数のクラスタを統合していくことでグラフ全体のクラスタリングを行う。Newman らが提案しているアルゴリズムでは、 Q の増減である ΔQ のみを計算するため、大規模なグラフにおいても高速にク

[†] 広島市立大学大学院情報科学研究科・日本学術振興会 DC
Graduate School of Information Sciences, Hiroshima City University・JSPS Research Fellow

^{††} 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

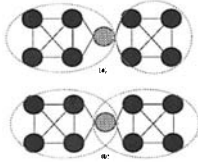


図1 クリस्पなクラスタリングと重複を許すクラスタリング

クラスタリングを行うことができる。この手法はネットワーク構造解析に基づくクラスタリング手法においては現在最も高速に動作するアルゴリズムであるといえる。しかしながら、Newman らの手法はクリस्पなクラスタリングを行うため、ひとつのノードが複数のクラスタに重複してクラスタリングされることを許していない。また、識別されるクラスタの境界面が曖昧になり、クラスタリング結果からより精度の高い知見を得ることができないという問題がある。

本研究では、Newman らが提案しているアルゴリズムを改良し、ひとつのノードが複数のクラスタに重複してクラスタリングされる手法を提案する。重複を許すクラスタリングを行うことで、クリस्पなクラスタリング手法では得られなかった新たな知見を見い出せる可能性が増加する。

提案手法では、一度、グラフ全体を Newman らによって提案されている手法でクラスタリングし、識別されたクラスタをそれぞれひとつのノードとみなした縮約グラフを作成する。作成された縮約グラフを再度 Newman らによって提案されている手法でクラスタリングし、複数のクラスタに重複してクラスタリングされるノードを発見する。

本論文の構成は以下の通りである。2 で問題定義し、3. で関連研究について述べ、4. で Newman らが提案するアルゴリズムについて説明する。5. で提案手法について説明し、6. で評価実験を行った結果を示す。7. でまとめる。

2. 問題定義

本章では、本研究で扱う課題の問題定義を示す。

ネットワークを無向グラフ $G = (V, E)$ (V : 頂点集合, E : 辺集合) で表すものとする。ネットワーク上のノードがグラフの頂点になり、リンクが辺となる。また、無向グラフは多重辺、自己ループのない単純無向グラフのみを考える。各頂点は一意にラベリング付けされており、頂点番号 i の頂点を v_i と表現する。また、 v_i と v_j 間の辺を e_{ij} として表すものとする。

ここで、クラスタ i に所属する頂点集合を $C_i \in \mathcal{V}$ と表すと、クラスタリング C は以下のようにすべて

のクラスタからなる集合として表現される。ここで、 $1 \leq i \leq n$ である。

$$C = \{C_1, C_2, \dots, C_n\} \quad (C_i \subset V)$$

$$V = \bigcup_{i=1}^n C_i \quad C_i \cap C_j \subset V \quad (i \neq j)$$

クラスタ i に所属する頂点 j を行列 X で表現することを考える。この行列を所属行列と呼ぶこととする。

所属行列 X の要素 x_{ij} の値が 1 のとき、頂点 j がクラスタ i にクラスタリングされているとし、値が 0 のとき、頂点 j はクラスタ i にはクラスタリングされていないものとする。頂点数が n のとき、クラスタリングの結果、取り出されるクラスタの最大数は n であるため、 $1 \leq i \leq n, 1 \leq j \leq n$ が成り立っている。

ここで、 \mathcal{F} をクラスタリングの良さを示す評価関数と定義し、 $\mathcal{F}(G, X)$ が返す値を評価値とする。本研究で扱う問題は、以下のように $\mathcal{F}(G, X)$ の値を最大にする X を求める組合せ最適化問題として定義できる。

$$\begin{aligned} & \text{maximize} \quad \mathcal{F}(G, X) \\ & \text{subject to} \quad \sum_j x_{ij} \geq 0 \\ & \quad \quad \quad \sum_i x_{ij} \geq 1 \end{aligned} \quad (1)$$

ある頂点 v_i は n 個のクラスタにそれぞれクラスタリングされる可能性があるため、1 つの頂点のクラスタリングとしては $2^n - 1$ 個考えられる。このクラスタリングの組合せが頂点数 n 分組み合わせる必要があるため、考えられる解の組合せは $(2^n - 1)^n$ となる。

3. 関連研究

インターネットのリンク構造をはじめ、ソーシャルネットワーク、人的ネットワーク、文献参照関係やタンパク質の相互作用ネットワークなど様々なネットワークから、クラスタと呼ばれる構造を抽出するアルゴリズムの研究が盛んにおこなわれている。

近年、盛んに研究がおこなわれているのが、ネットワーク構造解析に基づくクラスタリング手法である。ネットワーク構造解析に基づくクラスタリング手法は、ネットワークが表すグラフ上の頂点をクラスタリングすることにより、全体の構造、特に、クラスタとクラスタのつながり方を取り出せるという利点を持っている。

ネットワーク構造解析に基づくクラスタリング手法

は、データクラスタリングにおけるデータを頂点に置き換えたものであるといえる。データクラスタリングでは、fuzzy c-means 法^{1),2)}など、fuzzy クラスタリングにおいて重複を許したクラスタリング手法が盛んに研究されている^{3),4)}。一方、ネットワーク構造解析に基づくクラスタリング手法については、重複を許した手法に関する研究は少ない。

Palla らは、文献 5) において、クリークパーコレーション法と呼ばれるネットワーク構造解析に基づくクラスタリング手法を提案している。クリークパーコレーション法では、最初に、頂点の数 k ごとに k -クリークすべてを列挙する。次に、 k -クリークについて、頂点が $k-1$ 個重なり合っている k -クリーク同士を 1 つのクラスタとして抽出していく。クリークパーコレーション法では、すべてのクリークを列挙しているため、結果的に、複数のクラスタにクラスタリングされる頂点が存在する。

クリークパーコレーション法は、重複を許すネットワーク構造解析に基づくクラスタリング手法となっているが、以下の 2 つの課題が存在する。 k の値をどのように設定するか、また、どの k が最適なクラスタを抽出することができているかはユーザが得られた結果を分析する必要がある。 k の数が小さいと全体が 1 つのクラスタとなり、 k の数が大きいと、少数のクリークが抽出されるだけである。また、クリークしか取り出されないため、全体を分類するには不向きであるという欠点が存在する。

重複を許したクラスタリング手法として、文献 6) で提案されている MODES が挙げられる。MODES は、タンパク質の相互作用ネットワークデータを対象とした、頻出かつ密な部分グラフ抽出法である。この手法は、文献 7), 8) で提案されているカット手法を組み合わせ、改良した手法である。密の判定基準としては、文献 7) で示されている基準である $min_cut > n/2$ を使用している。ここで、 min_cut はグラフの最小カットの値であり、 n はグラフ全体のノード数である。

この手法は、重複してクラスタリングされるノードの存在を許しており、パラメータとして重複の度合い、抽出される頻出部分グラフの最小サイズなどを最初に指定する必要がある。MODES は複数のパラメータ設定が必要であり、カット手法を用いるため、最終的に抽出された部分グラフからクラスタ間の繋がりを把握することが難しく、パラメータの設定により結果が変化するため、グラフ全体を分類して解析する場合には適さない。

さらに、Zhang らは、Graph Spectral Cut と、fuzzy

c-means 法を利用した、ネットワーク構造解析に基づくクラスタリング手法を提案している⁹⁾。提案されている手法では、Graph Spectral Cut の途中段階で最適なカットを見つけるために使用されている k -means 法の部分を fuzzy c-means 法に置き換えたものであり、重複を許すクラスタリング手法となっている。

提案されている手法では、頂点数が 30 程度のネットワークにおいて、重複を取り出すことに成功している。しかしながら、fuzzy c-means 法の実行後、クラスタへの所属比率が λ 以上の頂点を、複数のクラスタにクラスタリングするという閾値 λ の設定が必要である。 λ の設定は、ネットワークの規模や構造に依存するため、最適な λ の値を何らかの手法で見つける必要があり、最適な λ を見つけるための指導原理的な指標が存在しない。

4. Newman らのアルゴリズム

4.1 モジュール性

無向グラフ $G(V, E)$ が与えられ、隣接行列 A の要素 A_{ij} が以下のように与えられているとする。

$$A_{ij} = \begin{cases} 1 & \text{if vertex } v \text{ and } w \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases}$$

頂点 v の次数 k_v は以下の式で表される。

$$k_v = \sum_{w \in V} A_{vw}$$

頂点 v が所属するクラスタ番号を c_v とし、頂点 v と頂点 w とが同じクラスタに所属するかどうかを示す関数を以下のように、

$$\delta(c_v, c_w) = \begin{cases} 1 & \text{if } c_v = c_w \\ 0 & \text{otherwise} \end{cases}$$

と定義する。

クラスタ内部に存在する辺の割合が多いクラスタリング結果ほど、良いクラスタリングといえる。ここで、クラスタ内部に存在する辺の割合を式で表すと、

$$\frac{\sum_{v,w \in V} A_{vw} \delta(c_v, c_w)}{\sum_{v,w \in V} A_{vw}} = \frac{1}{2m} \sum_{v,w \in V} A_{vw} \delta(c_v, c_w) \quad (2)$$

となる。この値が大きなクラスタリングほど内部に存在する辺の割合が多くなり、評価の高いクラスタリングであるといえる。

しかしながら、この式で表される割合だけでは、グラフ全体を 1 つのクラスタとするとき最大値 1 となっているために、評価値としては使用できない。そこで、辺をランダムに張り替えたとき、頂点 v と頂点 w の

間に辺が張られる確率 $k_v k_w / 2m$ を A_{vw} から引いた値を A_{vw} と置き換えたものをモジュール性の度合い Q として、

$$e_{ij} = \frac{1}{2m} \sum_{v,w \in V} A_{ij} \delta(c_v, i) \delta(c_w, j), \quad a_i = \frac{1}{2m} \sum_{v,w \in V} k_v k_w \delta(c_v, i)$$

$$Q = \frac{1}{2m} \sum_{v,w \in V} \left[A_{ij} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w) \quad (3)$$

定義する。ここで、とおくと、 Q は以下のように、

$$Q = \sum_i (e_{ii} - a_i^2) \quad (4)$$

となる。

Algorithm 1 Newman 入力: $G(V,E)$ 出力: C

```

1:  $C := \phi$ ;
2: for all  $v \in V$  do
3:    $C := C \cup \{v\}$ ; /* 各頂点を1つのクラスタとする */
4: end for
5:  $\Delta Q := \text{MAKE\_DELTA\_Q}(G, C)$ ;
   /* 関数 MAKE\_DELTA\_Q( $G, C$ ) は  $\Delta Q_{ij}$  を作成する関数 */
6: while (1) do
7:    $\{max\_dq, i, j\} := \text{GET\_MAX\_DELTA\_Q}(\Delta Q)$ ;
   /* 関数 GET\_MAX\_DELTA\_Q は  $\Delta Q_{ij}$  の最大値  $max\_dq$  とその添字  $\{i, j\}$  を返す関数 */
8:   if  $max\_dq > 0$  then
9:      $C := \text{JOIN}(C, i, j)$ ;
     /* 関数 JOIN( $C, i, j$ ) は、 $C_i$  と  $C_j$  を結合して1つのクラスタとする関数 */
10:     $\text{UPDATE\_DELTA\_Q}(Q, i, j)$ ; /*  $\Delta Q$  を更新 */
11:   else
12:     return  $C$ ; /* 得られたクラスタリング結果を返す */
13:   end if
14: end while

```

4.2 アルゴリズム

Newman のクラスタリングアルゴリズムは、 Q の値を最大にするようなクラスタリング結果を求める組合せ最適化問題となる。組合せの数は頂点数の指数オーダー存在するため、厳密解を求めるのではなく、貪欲アルゴリズムにより、近似最適解を求めている。

最初に1つの頂点を1つのクラスタとし、階層的に

クラスタ同士を結合する。どのクラスタ同士を結合するかは、2つのクラスタを結合することにより、 Q の値がどれだけ増加するかで判断する。

クラスタ i とクラスタ j とを結合したときに、増加する Q の値である ΔQ_{ij} は以下の式により求めることができる。

$$\Delta Q_{ij} = 2(e_{ij} - a_i a_j) \quad (5)$$

Algorithm1 に、Newman のクラスタリングアルゴリズムを示す。最初に、グラフの各頂点を1つのクラスタとする。次に、関数 **MAKE_DELTA_Q** を用いて2つのクラスタ i とクラスタ j を結合したときのモジュール性の増加値 ΔQ_{ij} を計算する。そして、関数 **GET_MAX_DELTA_Q** を用いて増加値 ΔQ_{ij} の最大値を求める。最後に、増加値 ΔQ_{ij} が最大値であるクラスタ同士を結合して1つのクラスタとし、関数 **UPDATE_DELTA_Q** により、 ΔQ を更新する。以下、6行目に戻り、同じ処理を繰り返し、クラスタを結合していく。増加値 ΔQ_{ij} がすべて負の値となれば、これ以上結合を続けたとしても Q の値が減少するため、処理を終了する。

5. 提案手法

本章では、提案手法である、Newman らのアルゴリズムを改良した、重複を許すクラスタリング手法の説明を行う。提案するクラスタリングアルゴリズムは、Newman らのアルゴリズムでクラスタリングした後、縮約グラフの作成を行い、再度 Newman らのアルゴリズムでクラスタリングを行い、重複してクラスタリングされるノードを見つけ出す。以下に、アルゴリズムの説明、具体的な例を用いた説明を行う。

5.1 アルゴリズム

提案するアルゴリズムは4.2で説明したNewmanのアルゴリズムを用いる。以下、4.2で説明したNewmanのアルゴリズムを関数 $Newman(G)$ と表記する。Algorithm2 に、提案するアルゴリズムを示す。

まず、 $Newman(G)$ を用い、入力されたグラフ G のクラスタリングを行う。得られたそれぞれのクラスタを変数 NC に代入する。 $Newman(G)$ により得られたクラスタの数が n 個あれば、 $NC = \{NC_1, \dots, NC_n\}$ となる。そして、それぞれの NC_i について、関数 $MAKE_C_GRAPH(G, NC_i)$ を用いて、 NC_i をひとつのノードとした縮約グラフを作成する。このとき、ひとつのノードに縮約されたクラスタに繋がっている辺の数の情報はそのまま保存しておく。 $MAKE_C_GRAPH(G, NC_i)$ の出力として、縮約グラフ

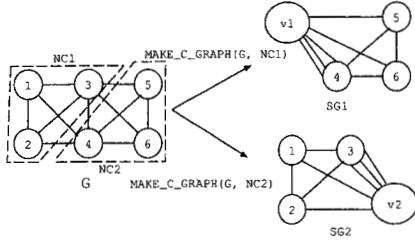


図 2 提案手法の具体例 1

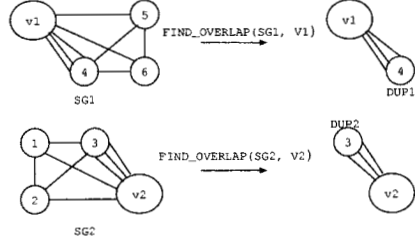


図 3 提案手法の具体例 2

SG_i と縮約されたクラスタであるノード v_i が得られる。

さらに、関数 $FIND_OVERLAP(SG_i, v_i)$ を用いて、重複してクラスタリングされるノードを見つけ出す。 $FIND_OVERLAP(SG_i, v_i)$ の出力として、重複してクラスタリングされるノード集合 DUP_i が得られる。得られたノード集合を NC_i に付加し、最終的な出力である C に挿入する。この手順を全ての NC_i について行い、重複してクラスタリングされる全てのノードを見つけ出す。

関数 $FIND_OVERLAP(SG_i, v_i)$ の動作は、次の通りである。縮約したノード v_i に隣接するノードをそれぞれ v_i と結合したときの ΔQ を計算する。そして、結合後に ΔQ が負となる v_i に隣接しているノードを取り除く。 v_i と結合しても ΔQ が負とならないノード集合を結果として出力する。

Algorithm 2 Overlapping_Newman 入力: $G(V,E)$ 出力: C

- ```

1: $C := \phi$; /*最終的な解を保存する変数*/
2: $NC := \phi$; /*Newman(G) によって得られるクラスタリング結果*/
3: $NC := Newman(G)$; /*Newman(G) の実行*/
4: for all $NC_i \in NC$ do
5: $\{SG_i, v_i\} := MAKE_C_GRAPH(G, NC_i)$ /*縮約グラフの作成*/
6: $DUP_i := FIND_OVERLAP(SG_i, v_i)$ /*重複するノードを見つける*/
7: $C := C \cup \{NC_i \cup DUP_i\}$ /*重複するノードを NC_i に追加*/
8: end for
9: return C; /*得られたクラスタリング結果を返す*/

```

## 5.2 例

本節では、具体的な図を用いて提案手法の説明を行う。図 2 から図 4 に提案手法の具体例を示す。6 つのノードで構成されるグラフ  $G$  を  $Newman(G)$

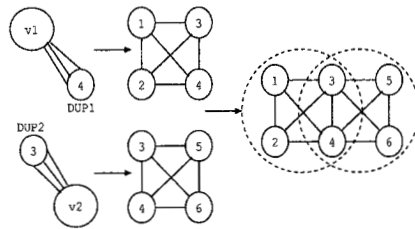


図 4 提案手法の具体例 3

によりクラスタリングすると、図 2 に示すように、 $NC_1$  と  $NC_2$  にクラスタリングされる。次に、 $MAKE\_C\_GRAPH(G, NC_i)$  を用いて縮約グラフを作成する。すると、図 2 の右側に示す 2 つの縮約グラフ  $SG_1$  と  $SG_2$  が作成される。

次に、 $FIND\_OVERLAP(SG_i, v_i)$  を用いて、作成された縮約グラフから重複してクラスタリングされるノードを発見する。図 3 に示すように、 $v_1$  とノード 4、 $v_2$  ノード 3 を結合すると  $\Delta Q$  の値が増加するため、それぞれ重複してクラスタリングされるノード  $DUP_1$ 、 $DUP_2$  として識別される。

最後に、 $v_1$  と  $v_2$  を元のグラフに復元し、重複を許したクラスタリングを終了する。最終的に、図 4 に示すようにノード 3、4 が 2 つのクラスタに重複してクラスタリングされるノードとなる。

## 6. 評価実験

本章では、まず、使用したネットワークデータについて説明を行い、Newman のアルゴリズムを用いてクラスタリングした結果、提案手法によりクラスタリングした結果を示す。さらに、クラスタリング結果をより詳細に解析した結果を示し、提案手法の有効性を確認する。

### 6.1 ネットワークデータ

評価には、文献 10) で収集した、プログラムのバックによる繋がりを含み、1 人のプログラマーをノードとした、単純無向グラフであるネットワークデータを用

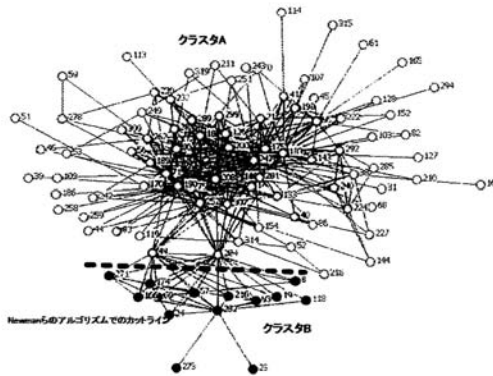


図5 Newmanらのアルゴリズムを適用した結果

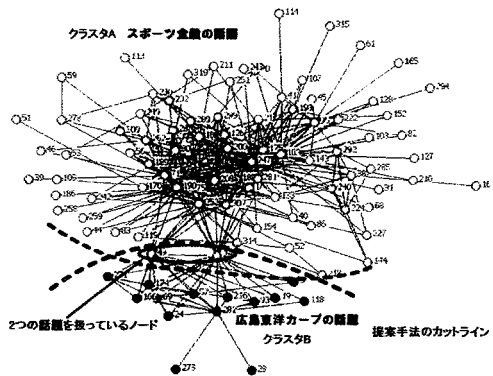


図6 提案手法を適用した結果

いた。用いたネットワークデータは、一人のプロガーがひとつのノードとなっており、プロガーが多趣味である場合など、複数のクラスタに属す可能性が高い。

## 6.2 評価

実際に提案した手法を実装し、評価実験を行った。図5にNewmanのアルゴリズムを適用した結果の一部を可視化したものを示す。図5には2つのクラスタ、クラスタAとクラスタB、クラスタの境界線が示されている。グラフデータとしてプロガーのグラフを扱っているため、2つのクラスタがどのような話題を主に扱っているのかを調べるために、クラスタ内に存在しているブログ記事の解析を *tf-idf* と手作業で行った。結果として、クラスタAはスポーツ全般の記事を扱っているプロガー集団であり、クラスタBはプロ野球球団である広島東洋カーブの話題を主に扱っているプロガー集団であった。

次に、図6に提案手法を適用した結果の一部を可視化したものを示す。結果として、2つのクラスタの境界に存在しているノード番号28と49のノードが2つのクラスタに重複してクラスタリングされた。ノード28と49は、図5ではクラスタAのみにクラスタリングされていたが、提案手法ではクラスタBにもクラスタリングされている。

ノード28と49が扱っている記事の内容を調べるために *tf-idf* を用いて記事の解析を行った。結果を表1と表2に示す。それぞれの表は *tf-idf* の値上位5件のキーワードを示している。表1、表2のキーワードより、それぞれのノードにおいてプロ野球球団の広島東洋カーブに関するキーワードが上位にランキングされている。また、実際に2つのノードのブログ記事を手作業で調べたところ、両者とも広島東洋カーブのファンであり、主に広島東洋カーブのブログ記事

表1 28の *tf-idf* の結果

| 順位 | キーワード |
|----|-------|
| 1  | 広島    |
| 2  | カーブ   |
| 3  | ブログ   |
| 4  | 野球    |
| 5  | 対     |

表2 49の *tf-idf* の結果

| 順位 | キーワード |
|----|-------|
| 1  | カーブ   |
| 2  | 鯉     |
| 3  | 広島    |
| 4  | 夏     |
| 5  | 対     |

を書いていることが分かった。つまり、ノード28と49は、図5において、クラスタBにもクラスタリングされるべきノードであることが分かる。しかしながら、提案手法においては、2つのノードはクラスタAとクラスタBの両方にクラスタリングされており、2つのノードが「スポーツ全般」を扱うプロガーと交流があり、主に「広島東洋カーブ」についてのブログ記事を書いていることが分かる。

Newmanのアルゴリズムでは、クラスタ間の境界面のノードがどちらのクラスタにも属す可能性があるという曖昧さが残っていたが、重複を許す提案手法によりクラスタ間の境界面の曖昧性を取り除くことができた。本論文では、ブログデータの一部分のみの評価実験を示したが、他のクラスタ間においても同様な重複を取り出すことができた。

## 7. まとめ

本研究では、Newmanらが提案しているクラスタリングアルゴリズムを改良し、重複を許したクラスタリング手法を提案した。提案したクラスタリング手法は、縮約グラフの作成と縮約グラフの再クラスタリングにより、重複してクラスタリングされるノードを発見した。ブログのトラックバックデータを用いて評価実験を行ったところ、Newmanらが提案しているクラスタリングアルゴリズムでは見い出せなかった、複数のク

ラスタに重複してクラスタリングされるノードを発見できた。実際に重複してクラスタリングされたノードの特徴を調べてみたところ、複数の話題の記事を扱っているノードであることが分かり、提案手法の有効性を示すことができた。

今後は、ブログのトラックバックデータだけではなく、ソーシャルネットワークなどの大規模なネットワークデータを対象とした評価実験を行う予定である。

**謝辞** 本研究の一部は、日本学術振興会・特別研究員奨励費（課題番号：18・0205）、日本学術振興会・科学研究費補助金（基盤研究（C）（一般）、課題番号：17500097）の支援により行われた。

### 参 考 文 献

- 1) J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Cybernetics*, Vol. 3, pp. 32–57, 1973.
- 2) James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- 3) Joerg Reichardt and Stefan Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Physical Review Letters*, Vol. 93, p. 218701, 2004.
- 4) Joerg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, Vol. 74, p. 016110, 2006.
- 5) Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, Vol. 435, No. 7043, pp. 814–818.
- 6) Haiyan Hu, Xifeng Yan, Yu Huang, Jiawei Han, and Xianghong Jasmine Zhou. Mining Coherent Dense Subgraphs Across Massive Biological Networks for Functional Discovery. In *ISMB (Supplement of Bioinformatics)*, pp. 213–221, 2005.
- 7) Erez Hartuv and Ron Shamir. A Clustering Algorithm Based on Graph Connectivity. *Information Processing Letters*, Vol. 76, No. 4–6, pp. 175–181, 2000.
- 8) Jianbo Shi and Jitendra Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888–905, 2000.
- 9) S. Zhang, R.-S. Wang, and X.-S. Zhang. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A Statistical Mechanics and its Applications*, Vol. 374, pp. 483–490, January 2007.
- 10) 高木允, 森康真, 田村慶一, 黒木進, 北上始. ブログユーザ空間からの頻出なコミュニティ抽出法. 電子情報通信学会第 18 回データ工学ワークショップ

ブ, 2007.