

推薦論文

# ノードの注目度に基づく機能コミュニティ抽出法

伏見 卓恭<sup>1,a)</sup> 齊藤 和巳<sup>1</sup> 池田 哲夫<sup>1</sup> 風間 一洋<sup>2</sup>

受付日 2013年3月20日, 採録日 2013年5月2日

**概要:** ネットワーク分析において、ノード全体より一部のノードまたはノード集合に焦点が向けられる場合がある。本稿では、分析者が注目したいネットワーク内の一部のノード集合に注目度を設定し、分析の解像度を制御する枠組みを提案する。ノード集合の選び方により得られる処理結果の特徴がどのように変化するかは自明ではないため、提案法により抽出される機能コミュニティの性質および有効性について評価する。複数のネットワークを用いた実験より、対象ノード集合がネットワークに分散している場合は、各対象ノードの周辺ノードを細かく分類し遠いノードを粗く分類するという結果が得られた。対象ノード集合のノードどうしが互いに近傍に存在する場合は、対象ノード集合内のノードを細かく分類し、その他のノードを粗く分類するという結果が得られ、分析の解像度を制御できることが示唆された。

**キーワード:** 機能コミュニティ, 注目度, パーソナライズ, ノード集合

## Functional Community Extraction Method Based on Degree of Attention to Nodes

TAKAYASU FUSHIMI<sup>1,a)</sup> KAZUMI SAITO<sup>1</sup> TETSUO IKEDA<sup>1</sup> KAZUHIRO KAZAMA<sup>2</sup>

Received: March 20, 2013, Accepted: May 2, 2013

**Abstract:** In network analysis, we sometimes focus on partial subsets of whole network. In this paper, we propose a framework that sets the degrees of attention and controls analytic resolution of target nodes set which we focus on. Since it is not obvious how characteristics of results depend on how to choose a set of nodes, we evaluate the characteristics and validity of functional communities extracted by our proposed method. From our experimental results using several types of synthetic and real networks, when the target nodes are distributed over the network, we obtained the results where the peripheral nodes of each target node were classified finely and far nodes were classified coarsely. When the target nodes are mutually close or adjacent, we obtained the results where target nodes were classified finely and the other nodes were classified coarsely. Therefore it is suggested that our method makes it possible to control analytic resolution in our experiments.

**Keywords:** functional community, degree of attention, personalize, node set

### 1. はじめに

複雑ネットワーク分析の分野では、ネットワーク内の密に結合する部分をコミュニティとしてとらえ、効率良くコミュニティを抽出する手法 [1] やネットワーク内での重要

ノードを抽出する手法 [2] など、さまざまな分析手法が提案されている。

既存のコミュニティ抽出手法として、Newman による Modularity というネットワーク分割指標を用いたコミュニティ抽出手法が高速で、大規模ネットワークに対しても有効であり注目を浴びている [3]。これは、ノードどうしの結合が疎な部分を切断し、いくつかのサブネットワーク

<sup>1</sup> 静岡県立大学大学院経営情報イノベーション研究科  
Graduate School of Management and Information of Innovation, University of Shizuoka, Shizuoka 422-8526, Japan

<sup>2</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

a) j11507@u-shizuoka-ken.ac.jp

本稿の内容は 2012 年 11 月の WebDB フォーラム 2012 にて発表され、同シンポジウムプログラム委員会により情報処理学会論文誌データベースへの掲載が推薦された論文である。

に分割する方法である。また、ネットワーク上でのノードどうしが密結合したサブネットワークをコミュニティと見なして、クリーク (clique) の条件を緩めたサブネットワークを見つける手法も提案されている [4], [5], [6]。これらの手法は、リンク構造の粗密に着目し、ネットワークをいくつかのサブネットワークに分割する。さらに、ネットワーク全体に対する大域的な役割が類似するノード群をコミュニティとして抽出する機能コミュニティ抽出法がある [7], [8]。機能コミュニティ抽出法は、ネットワーク上でのランダムウォークに着目し、PageRank スコアの収束過程における各ノードのステップごとのスコアを特徴ベクトルとし、その類似度に基づきノードを分類する。この手法では、全ノードを対等に扱い、ネットワーク全体を客観視した際の各ノードの機能に基づくといえる。たとえば、会社組織内の人間関係ネットワークでは、社長という階層上のトップノードから順に、部長、課長のように、Web ページのハイパーリンクネットワークでも、トップページという階層上のトップノードから順に、コンテンツの一覧を示すページ、コンテンツの詳細内容のページなどのように、それぞれのノードに階層や役割がある。これらの公式的な役割だけでなく、部門内に外部からの情報を伝達する役割の媒介度の高い社員や、小グループ内でのハブ的役割の社員など、非公式に定まる機能が類似するノードも同定する。

これらの既存手法では、ネットワーク内の全ノードを対等に扱い、各ノードの特徴量を計算しランキング、あるいはノード群を抽出する。しかしノード全体より、一部のノードまたはノード集合に焦点を当てて分析したい場合がある。本稿では、分析者が注目したいネットワーク内の一部のノード集合に注目度を設定することで対象ノード集合にバイアスをかける枠組みを提案する。つまり、全ノードを平等とするのではなく、注目したいノード集合の周辺の解像度を細かく、離れた部分は粗く分析できる枠組みである。

一方、あるノード (エゴノードと呼ぶ) に注目し、そのノードの視点から見た機能・役割の類似するノード群を抽出する手法であるパーソナライズ機能コミュニティ抽出法が提案されている [9]。たとえば、会社組織内の人間関係ネットワークでは、社長は絶対的なトップノードであり、他のノードから見ても変化しない立場であると考えられる。しかし、各部門の部長や課長などは、自身の属する部門と他部門の部長とでは、自身から見た役割・機能は異なると考えられる。また、自身の直接隣接するノードやその隣接するノードなど、自身の近くのノードは直接的な相互関係が比較的大きいため、機能によって細分化できるが、自身の遠くのノードは相互関係が比較的小さく、同一視されやすいと想定できる。くだけた言い方をすると、“個人から見れば遠くのノードはみんな一緒に見える”ということになる。この手法では、上記の例のような考え方を基に、

PageRank 計算の初期ベクトルにおいてエゴノードのみに値を設定することで、エゴノード視点における他のノードの機能ベクトルを計算する。

本稿で提案する手法は、既存のパーソナライズ機能コミュニティ抽出法を拡張し、複数ノードに対して任意の確率分布で注目度を設定する。しかし、ノード集合の選び方により得られる処理結果の特徴がどのように変化するかは自明ではない。そこで本稿では、分析者が焦点を当てたい対象ノード集合に対して注目度を設定し、抽出される機能コミュニティの性質および有効性について評価する。

本稿は以下のような構成である。2章で通常の大域的機能コミュニティ抽出アルゴリズムおよび主要技術である PageRank スコア変化曲線と  $K$ -median 法について説明する。3章で本稿の提案法である注目度を導入した機能コミュニティ抽出法について説明する。そして、提案法の可視化による評価実験の結果を4章で示す。最後に本研究のまとめと今後の展望を5章で述べる。

## 2. 機能コミュニティ抽出法

この章では、ノードの機能に着目し、機能の類似するノード群から構成される機能コミュニティを抽出する方法について説明する。

機能コミュニティ抽出法は、ノード集合  $V$ 、リンク集合  $E$  からなるネットワーク  $G = (V, E)$  とコミュニティ数  $K$  を入力とし、以下のようなアルゴリズムにより機能コミュニティを抽出する。

- (1) 各ステップでの PageRank スコアベクトル  $\{y_1, \dots, y_s\}$  を計算;
- (2) 各ノードの特徴ベクトルとして PageRank スコア変化曲線  $x_v$  を構築;
- (3) 各ノードペアの特徴ベクトル  $x_u$  と  $x_v$  のコサイン類似度  $\rho(u, v)$  を計算;
- (4)  $K$ -median 法により全ノードを  $K$  個のグループに分割;
- (5) 機能コミュニティ  $\{C_1, \dots, C_K\}$  を出力;

以下に詳細を説明する。

### 2.1 PageRank スコア変化曲線

機能コミュニティ抽出法では、Web ページのランキングアルゴリズムとして用いられる PageRank [10] において、各ノードの PageRank スコアが収束するまでのステップごとのスコアを各ノードの特徴ベクトルとして採用する。

無向ネットワーク  $G = (V, E)$  の各ノードに 1 から  $|V|$  までの整数値を一意に割り振る。ここで、 $(u, v) \in E$  のとき  $a(u, v) = 1$ 、それ以外るとき  $a(u, v) = 0$  とし隣接行列  $\mathbf{A} \in \{0, 1\}^{|V| \times |V|}$  を定義する。各ノード  $u \in V$  に対して、 $\Gamma(u)$  をノード  $u$  の隣接ノード集合とする。すなわ

ち、 $\Gamma(u) = \{v \in V; (u, v) \in E\}$  となる。ここで、行推移確率行列  $\mathbf{P}$  の各要素は、 $p(u, v) = a(u, v)/|\Gamma(u)|$  である。通常、 $|\Gamma(u)|$  をノード  $u$  の次数という。各ノードの PageRank スコアを要素とするベクトル  $\mathbf{y}$  は、 $y(v) \geq 0$  で  $\sum_{v \in V} y(v) = 1$  となる。繰返しステップのパラメータ  $s$  を用い、PageRank スコアベクトル  $\mathbf{y}$  は以下の更新式の極限分布として定義される：

$$\mathbf{y}_s^T = \mathbf{y}_{s-1}^T \mathbf{P} \quad (1)$$

ここで  $\mathbf{b}^T$  はベクトル  $\mathbf{b}$  の転置を表す。ノード  $u$  に注目すると、

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} \{y_{s-1}(v) \cdot p(v, u)\} \\ &= \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \end{aligned} \quad (2)$$

で計算される。反復回数  $S$  まで反復を繰り返し、各ステップでのノード  $u$  の PageRank スコアを要素としたベクトルを  $\mathbf{x}_u = (y_1(u), y_2(u), \dots, y_S(u))^T$  と定義する。このベクトル  $\mathbf{x}_u$  をノード  $u$  の PageRank スコア変化曲線と呼ぶ。

## 2.2 $K$ -median クラスタリング

$K$ -median 法は、非階層クラスタリングで有名な  $K$ -means 法と同様に、 $N$  個のオブジェクト集合  $\mathcal{V}$  が与えられたとき、オブジェクト集合を  $K$  個のクラスに分割する手法である。任意のオブジェクトペア  $u, v \in \mathcal{V}$  間に、適切な類似度  $\rho(u, v)$  が定義されていると仮定し、オブジェクト集合の中から他のオブジェクトとの類似度の和が高い代表オブジェクトを選定する。他のオブジェクトは、自身と最も類似度の高い（距離の小さい）代表オブジェクトのクラスに割り当てる。

$K$ -median の解法には反復法や貪欲法があるが、機能コミュニティ抽出法では解の一意性が保証される貪欲法を採用する。貪欲法とは、すでに選定した代表オブジェクトを固定し、ある評価関数値を最大にするオブジェクトを求め、目的関数が増加するならば代表オブジェクト集合に追加し、 $K$  個の代表オブジェクトを求める方法である。すでに選定した代表オブジェクト集合を  $\mathcal{P}$  とし、新たに追加を試みるオブジェクトを  $w$  とするとき、本稿では、以下の目的関数を考える。

$$f(\mathcal{P} \cup \{w\}) = \sum_{v \in \mathcal{V}} \max\{\mu(v; \mathcal{P}), \rho(v, w)\}. \quad (3)$$

ここで、 $\mu(v; \mathcal{P})$  はすでに選定された代表オブジェクトとの類似度の最大値を表し、 $\mu(v; \mathcal{P}) = \max_{w \in \mathcal{P}} \{\rho(v, w)\}$  で定義される。以下に貪欲法による  $K$ -median 法のアルゴリズムを説明する。

(1)  $k \leftarrow 1$ ,  $\mathcal{P}_0 \leftarrow \emptyset$ , 各オブジェクト  $v \in \mathcal{V}$  に対し、

$\mu(v; \emptyset) \leftarrow 0$  と初期化する；

(2) 式 (3) で  $\hat{p}_k = \arg \max_{w \in \mathcal{V} \setminus \mathcal{P}_{k-1}} \{f(\mathcal{P}_{k-1} \cup \{w\})\}$  を求め、 $\mathcal{P}_k \leftarrow \mathcal{P}_{k-1} \cup \{\hat{p}_k\}$  とする；

(3)  $k = K$  ならば  $\hat{\mathcal{P}}_K = \{\hat{p}_1, \dots, \hat{p}_K\}$  を出力し終了する；

(4) 各オブジェクト  $v \in \mathcal{V}$  に対し、 $\mu(v; \mathcal{P}_k)$  を求め、 $k \leftarrow k + 1$  としステップ (2) へ戻る。

各オブジェクトを、最も類似度の高い代表オブジェクト  $p_k \in \mathcal{P}$  のクラス  $\mathcal{C}_k$  に割り当てる。

## 3. 注目度に基づく機能コミュニティ抽出法

この章では、提案法である注目度を考慮した機能コミュニティ抽出法について説明する。前述したように、大域的な機能コミュニティ抽出法は全ノードを対等に扱い、ネットワーク全体に対する各ノードの機能に基づき抽出する。この手法では、ノード全体の中でも特に注目したいノード集合が存在する場合など、注目度の違いを考慮できない。

本稿で導入する注目度の枠組みは、任意の注目度を付与した対象ノードを、注目度に応じてより精緻に、その他のノードを粗く分析することで、分析対象の解像度を制御することを目的とする。我々が扱うネットワーク構造は、友人関係やメールのやりとりなどの現実世界の写像であることが多い。そのような場合は、ネットワーク全体を俯瞰した視点から画一的に扱うだけでは不十分である。たとえば、現実世界においては、自分の周辺を詳細に知りたいとか、特に中間管理職達の周辺構造を明確にしたいなどの、ネットワーク中の特定の場所に注目した分析が必要となる。

本稿で導入する注目度を設定することで、対象ノード集合をその注目度に応じて不均一に扱うことを可能にする。すなわち、最も注目するノードに最も大きな値を設定し、次に注目するノードには幾分か小さな値を設定するというように、各ノードの注目度を要素に持つ注目度ベクトルを定義する。ノード  $v$  の注目度を  $c(v)$  とし、注目度ベクトル  $\mathbf{c}$  を以下のように設定する。

$$\mathbf{c} = \left( \dots, \frac{c(v)}{C}, \dots \right)^T \quad (4)$$

ここで、正規化定数は  $C = \sum_{v \in \mathcal{V}} c(v)$  である。提案法では、機能コミュニティ抽出法における PageRank スコア計算の初期ベクトルを  $\mathbf{y}_0 = \mathbf{c}$  とすることで、注目度に応じた機能コミュニティを抽出する。注目度ベクトルとして、ある対象ノードを中心とする正規分布や複数ノードを中心とする混合正規分布などのように任意の連続分布を設定できるため、離散的な初期ベクトルを想定していた既存のパーソナライズ機能コミュニティより汎用的な手法であるといえる。

注目度  $c(v)$  を全ノードに対し一様に設定した場合、通常の機能コミュニティと等価となり、すべてのノードを対等に扱った大域的な機能コミュニティが抽出される。注目

度  $c(v)$  をあるノードのみに設定した場合、エゴジャンプ確率  $\beta = 0$  としたパーソナライズ機能コミュニティと等価となる。

初期ベクトルを上記のように設定することで、反復計算の初期段階では、対象ノード集合およびその周辺ノードのみが値を持ち、遠くのノードは 0 のままとなる。周辺ノードは、対象ノード集合との関係（局所的なリンク構造）により、特徴ベクトルである PageRank スコア変化曲線が変動するため、対象ノード集合の近くでは特徴ベクトルの違いが強調され、より精緻に分類されることが期待できる。

## 4. 評価実験

本章では、現実の Web ネットワークデータなどを対象に、注目度導入により拡張した機能コミュニティ抽出法を評価する。対象ノード集合を変化させた際の処理結果について、妥当なコミュニティ抽出結果が得られるか、可視化により定性的に評価する。

### 4.1 ネットワークデータ

評価実験では、提案法の特徴および妥当性を評価するために、人工ネットワークおよび 3 つの実ネットワークを用いる。

1 つ目のネットワークは、Ravasz らによって提案された階層性のあるネットワークを生成する HN モデルにより生成した人工ネットワークである [11]。階層性のあるネットワークとは、企業内の社員のネットワークや Web サイトのハイパーリンクネットワークのようにトップノードと他のすべてのノード間にはリンクが張られているが、その他のノードどうしは限られた範囲でのみリンクが張られているような構造を持っている。すなわちトップノード（社長やトップページに相当するノード）は高い次数を有しているが、クラスタ係数が非常に小さいことになる。一方、その他のノード（一般社員や普通のページに相当するノード）は低い次数を有しているが、狭い範囲内で密につながっているためクラスタ係数が大きくなる。本稿では Hierarchical ネットワークと呼ぶ。

2 つ目のネットワークは、ネットワーク分析のベンチマークとして広く用いられている、空手クラブ内の友人関係ネットワークである [12]。社会ネットワークの特徴であるスケールフリー性とスモールワールド性を有する。このネットワークには、インストラクタとアドミニストレータのそれぞれからなる 2 つの派閥が存在している。本稿では Karate ネットワークと呼ぶ。

3 つ目のネットワークは、Web のハイパーリンク・ネットワークである。大学のウェブサイト内のページを 2010 年 8 月に収集し、ウェブサイトのハイパーリンク構造からハイパーリンクネットワークを構築し無向化した。本稿で

は Hosei ネットワークと呼ぶ\*1。

4 つ目のネットワークは、全国デジタル道路地図データベース静岡市版 [13] をもとに作成した静岡市の道路網データであり、交差点をノード、交差点間の道路をリンクとしたネットワークである。パーソナライズ機能コミュニティにはエゴノードからの距離の遠近という概念があるので、平均距離が大きい道路ネットワークを用いる。本稿では ShizuokaR ネットワークと呼ぶ。

### 4.2 実験結果と考察

上述したネットワークに対し、収束判定基準を  $\|y_s - y_{s-1}\|_{L1} < 10^{-8}$ 、コミュニティ数  $2 \leq K \leq 10$  とし、提案法による機能コミュニティ抽出結果を図 3、図 4、図 5、図 6 に示す。なお説明の便宜上、最も適切と思われるコミュニティ数  $K$  を選択しているが、他の  $K$  の場合でも大きな矛盾のない結果が得られている。処理結果の比較のため、通常の機能コミュニティ、エゴノードを 1 つ選択した場合のパーソナライズ機能コミュニティの抽出結果も示す。本稿では、対象ノード集合内の各ノードの注目度は、簡単のため同一確率に設定する。

#### 4.2.1 Hierarchical ネットワークの処理結果

Hierarchical ネットワークの結果を図 3(a) および図 3(b) にそれぞれ示す。それぞれ対象ノード集合を黒い丸で囲っている。比較のため通常の機能コミュニティ抽出結果を図 1(a) に、パーソナライズ機能コミュニティ抽出結果を図 2(a) にそれぞれ示す。

このネットワークの特徴は、企業内の社員のネットワークや Web サイトのハイパーリンクネットワークのように階層性を有していることである。図 1(a) から、通常の機能コミュニティ抽出法では、階層的な地位のようなネットワーク内での大域的な機能・役割に基づいて、ノードを分類できていることが分かる。

図 2(a) において、黒い丸で囲っているエゴノードは、会社組織でいうとあるグループのリーダー的存在である。エゴノードの視点からの機能コミュニティ抽出結果を見ると、エゴノードの属するグループメンバ（桃色の星形）、周辺グループのグループリーダー（黄色の菱形）とそのメンバ（黄緑の三角形）、トップノード（赤色の丸）のように、エゴノードの直接関係するノードをその役割により細分化している。一方、エゴノードから遠くに存在するノード群は粗く 2 つのコミュニティに分類されている。このように、パーソナライズ機能コミュニティでは同じ機能のノードが存在しても、エゴノードとの距離が違えば同一コミュニティと判定されないという特徴があることが示唆された。

図 3(a) から、ネットワーク内に分散する類似の役割を果たすノードを対象ノード集合とした場合、対象ノードの周

\*1 法政大学情報科学部：<http://cis.k.hosei.ac.jp/>

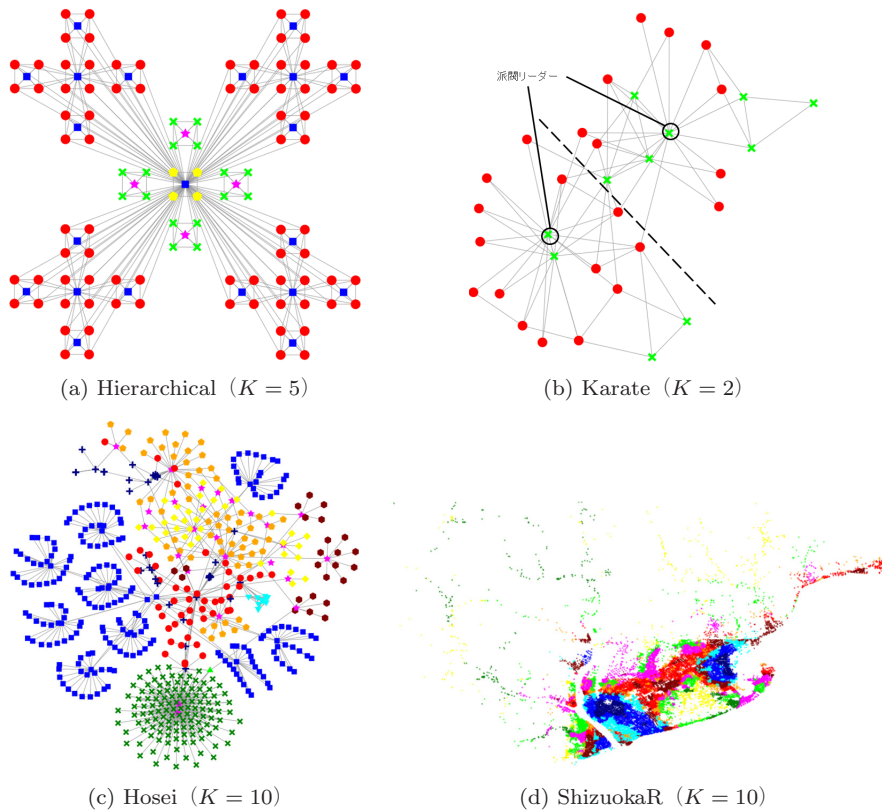


図 1 機能コミュニティ抽出結果

Fig. 1 Results of global functional communities.

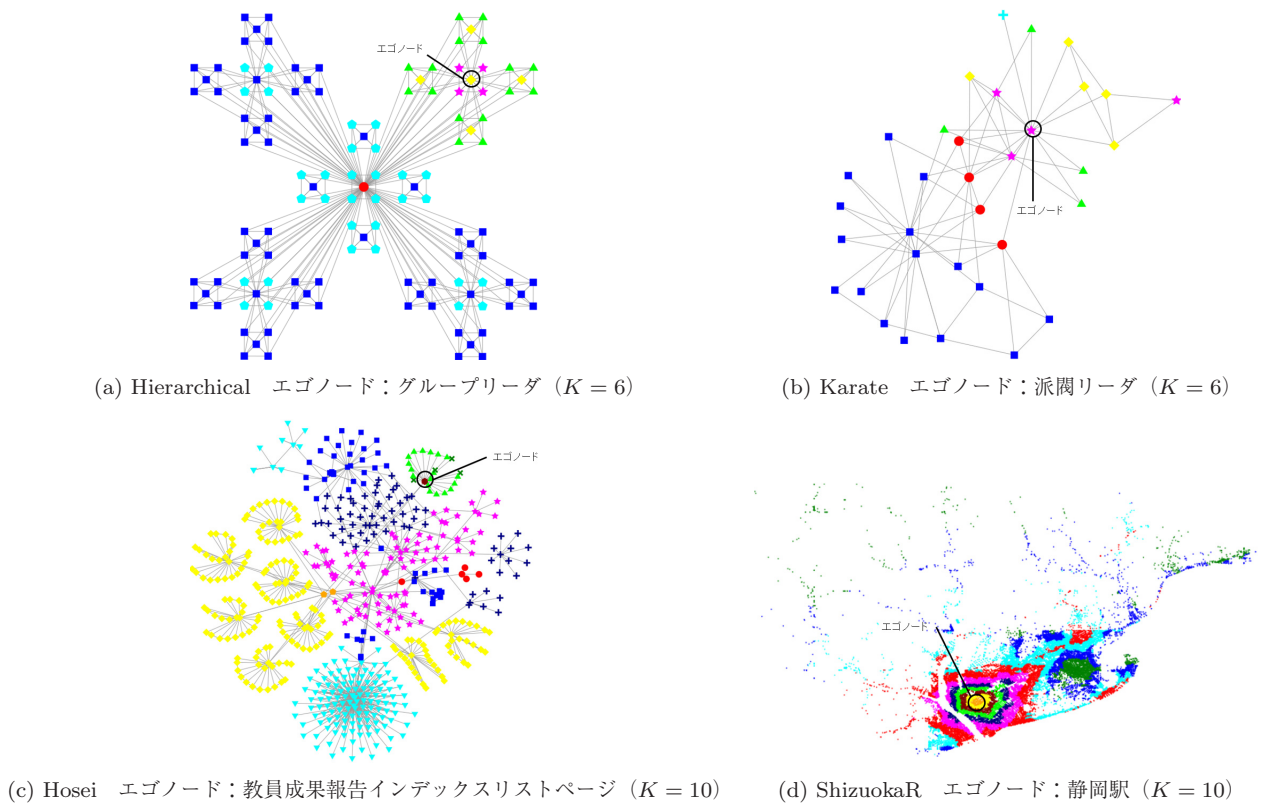
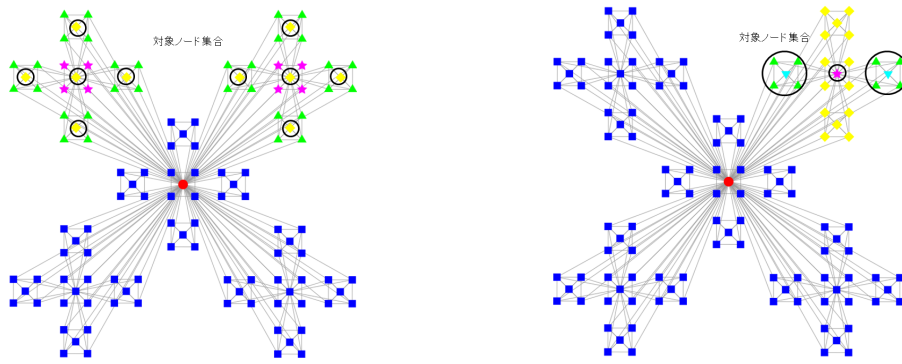


図 2 パーソナライズ機能コミュニティ抽出結果

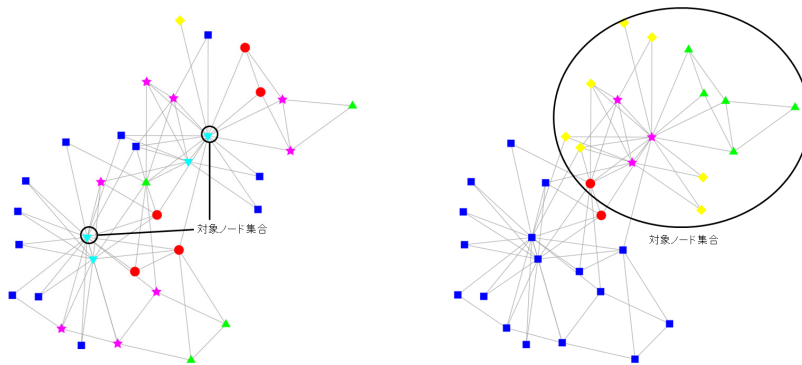
Fig. 2 Results of personalized functional communities.



(a) 対象：各グループの中間階層ノード ( $K = 5$ ) (b) 対象：隣接する 2 つのグループ内のノード集合 ( $K = 6$ )

図 3 Hierarchical 提案法による処理結果

Fig. 3 Results of Hierarchical network by proposed method.



(a) 対象：各派閥のリーダー ( $K = 6$ ) (b) 対象：インストラクタ派閥ノード集合 ( $K = 5$ )

図 4 Karate 提案法による処理結果

Fig. 4 Results of Karate network by proposed method.

辺ノードは階層の違いなどにより細かく分類されていることが分かる。対象ノードから遠くに存在するノード群はほとんど1つのコミュニティとして抽出されている。図 3(b) から、互いに近傍に存在するノードを対象ノード集合とした場合においては、対象ノード集合内をその階層により細分化し、その他のノードに関しては特異なトップノードを除いてほとんど1つのコミュニティとして抽出していることが分かる。

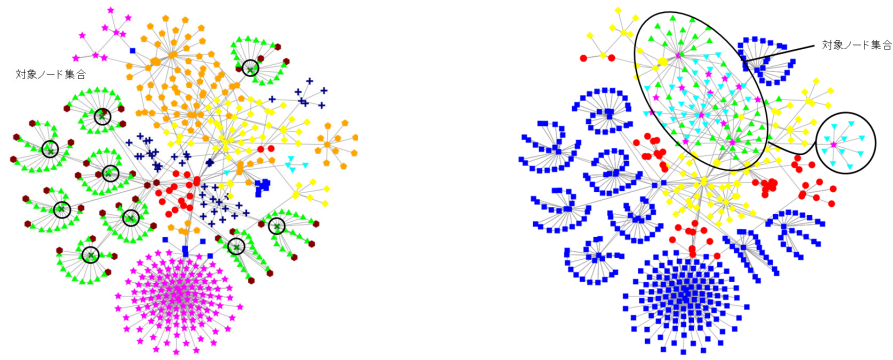
#### 4.2.2 Karate ネットワークの処理結果

Karate ネットワークの結果を図 4(a) と図 4(b) にそれぞれ示す。それぞれ対象ノード集合を黒い丸で囲っている。比較のため通常の機能コミュニティ抽出結果を図 1(b) に、パーソナライズ機能コミュニティ抽出結果を図 2(b) にそれぞれ示す。Karate ネットワークは、階層性が視認できるよう、ばねモデルにより可視化している [14]。

このネットワークの特徴は、図 1(b) のように2つの派閥リーダーが存在し、他のノードはどちらかの派閥に所属している。図中の点線は派閥の境界を表している。図 1(b) から、通常の機能コミュニティ抽出法では、ハブノード（派閥のリーダー）とその他ノードのようなネットワーク内での大域的な機能・役割に基づいて、ノードを分類できていることが分かる。

図 2(b) において、黒い丸で囲っているエゴノードは派閥リーダーである。エゴノードの視点からの機能コミュニティ抽出結果を見ると、エゴノードの周辺ノード、特にエゴノードの派閥に属するノードを細分化している。どちらの派閥にも強いコネクションを持つノード（赤色の丸）、ハブ的存在のノード（桃色の星形）、ハブノードとのみ隣接しているノード（黄緑の三角形）、エゴノードとだけ隣接しているノード（水色の十字形）などに分類されている。遠方のノード、すなわち、他の派閥のノードの多くは同一のコミュニティとして抽出されている。

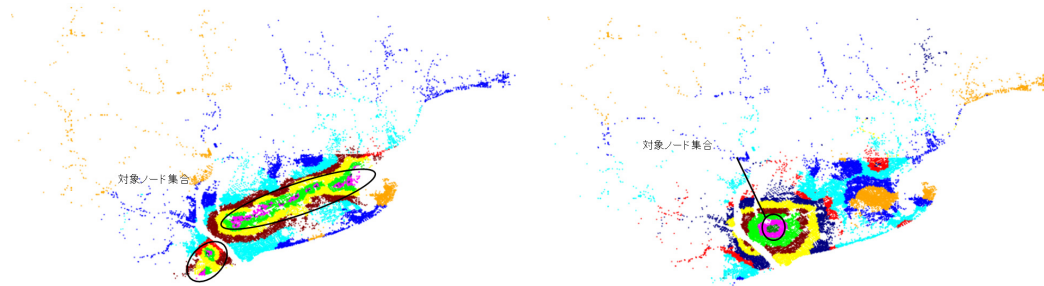
図 4(a) から、各派閥のリーダーを対象ノード集合とした場合、対象ノードの周辺ノードがその役割や隣接関係などにより細かく分類されていることが分かる。図 2(b) の処理結果を複数ノードへ拡張した結果が得られたといえる。図 4(b) から、ある派閥内のノード全体のように、互いに近傍に存在するノードを対象ノード集合とした場合には、対象ノード内の役割により細かく分類されていることが分かる。具体的には、派閥内の中心人物（桃色の星形）、中心人物とのみつながる人物（黄色の菱形）、派閥内の小グループ（黄緑の三角形）というように細分化されている。対象以外のノード群は、ほとんど1つのコミュニティとして抽出され、粗く分類されている。



(a) 対象：各年度の成果報告ページへのインデックス ( $K = 10$ ) (b) 対象：教員紹介ページ集合 ( $K = 6$ )

図 5 Hosei 提案法による処理結果

Fig. 5 Results of Hosei network by proposed method.



(a) 対象：市内の各駅最近隣交差点 ( $K = 10$ )

(b) 対象：静岡駅周辺交差点集合 ( $K = 10$ )

図 6 ShizuokaR 提案法による処理結果

Fig. 6 Results of ShizuokaR network by proposed method.

#### 4.2.3 Hosei ネットワークの処理結果

Hosei ネットワークの結果を図 5(a) と図 5(b) にそれぞれ示す。それぞれ対象ノード集合を黒い丸で囲っている。比較のため通常の機能コミュニティ抽出結果を図 1(c) に、パーソナライズ機能コミュニティ抽出結果を図 2(c) にそれぞれ示す。Hosei ネットワークは、隣接性が視認できるよう、クロスエントロピー法により可視化している [15]。

このネットワークの特徴は、教員の成果報告ページが年度ごとに別のディレクトリにまとめて整理されて公開されていることである。なお、インデックスページからどの年度にもたどれるが、年度間のリンクは存在しない。図 1(c) から、通常の機能コミュニティ抽出法では、直接隣接していない年度ごとの成果報告ページが同一コミュニティとして抽出されるなど、ネットワーク内での大域的な機能・役割に基づいて、ノードを分類できていることが分かる。

図 2(c) において、黒い丸で囲っているエゴノードは対象年度の教員成果報告ページ群へのインデックスページである。エゴノードの視点からの機能コミュニティ抽出結果を見ると、エゴノードの周辺ノード、すなわち同年度の成果報告ページは、初めの教員と最後の教員ページのような末端に位置するページと中間に位置するページのように、細かなリンク構造の違いにより細分化されている。一方、エゴノードから遠くに存在するノード群は、通常の機能コミュニティ抽出結果と比較して、細分化されない傾向に

ある。

図 5(a) から、各年度のインデックスページを対象ノード集合とした場合、対象ノードの周辺ノードは末端に位置するか中間に位置するかなどのリンク構造の違いにより細かく分類されていることが分かる。対象ノードから遠くに存在するノード群は、周辺ノードと比較して粗く分類されている。図 5(b) から、教員紹介ページ群のように、互いに近傍に存在するノードを対象ノード集合とした場合には、対象ノード内の役割により細かく分類されていることが分かる。具体的には、学部の教員ページ (黄緑の三角形)、大学院の教員ページ (水色の逆三角形)、高校生向けのコンテンツを含むページ (黄色の菱形)、インデックスページや各研究室のトップページなどのハブノード (桃色の星形) というように細分化されている。対象以外のノード群は、対象ノード集合と比較して粗く分類されている。

#### 4.2.4 ShizuokaR ネットワークの処理結果

ShizuokaR ネットワークの結果を図 6(a) と図 6(b) にそれぞれ示す。比較のため通常の機能コミュニティ抽出結果を図 1(d) に、パーソナライズ機能コミュニティ抽出結果を図 2(d) にそれぞれ示す。ShizuokaR ネットワークは、交差点の実座標により可視化している。

現在の静岡市は、2003 年の旧静岡市と旧清水市の新設合併により発足した。このために、このネットワークは、西の静岡駅周辺と東の清水駅周辺の 2 カ所にそれぞれの市

の中心部が存在する。図 1(d) から、通常の機能コミュニティ抽出法では、おおむね駅周辺（紺、青、水）、学園地域・民家が多く並ぶ地域（黄）、山岳地帯（桃）、農村地帯（橙）のようにネットワーク内での大域的な機能・役割に基づいて、ノードを分類できていることが分かる。

図 2(d) において、黒い丸で囲っているエゴノードは静岡駅周辺の交差点の 1 つである。エゴノードの視点からの機能コミュニティ抽出結果を見ると、エゴノードの周辺ノードは、エゴノードからの距離に従い細分化されている。これは、歴史的な経緯などにより、駅や城下町からの距離により地域の機能が変化し、それを反映した道路構造になっているからと考えられる。一方、エゴノードから遠くに存在するノード群は、通常の機能コミュニティ抽出結果と比較して、細分化されない傾向にある。注目している駅の周辺においては静岡市の土地利用方針図<sup>\*2</sup>における土地利用区分と類似した結果が得られた。またこのネットワークは、現実世界の物理的制約を受けたことで平面グラフに近い構造<sup>\*3</sup>をしており、距離が離れたノードとのショートカットが存在しないので、距離がより支配的になる結果となったと考えられる。

図 6(a) から、静岡市内にある JR ならびに私鉄の各駅周辺の交差点を対象ノード集合とした場合、対象ノードの周辺ノードは駅からの距離や道路構造の違いにより細かく分類されていることが分かる。対象ノードから遠くに存在するノード群は、周辺ノードと比較して粗く分類されている。図 6(b) から、静岡駅周辺の交差点群のように、互いに近傍に存在するノードを対象ノード集合とした場合においても、対象ノードからの距離が強く影響する結果が得られた。

これらの結果をまとめると、どのネットワークに対する処理結果も、対象ノードがネットワーク内に分散する場合、各対象ノードの周辺ノードは細分化、遠くに存在するノードは粗く分類された。対象ノードがネットワーク内で互いに近傍に存在するグループとした場合、対象ノード集合内をその役割により細分化、対象以外のノードは粗く分類された。以上より、注目度を付したノード集合について、相対的に詳細に分析ができることが示唆された。

## 5. おわりに

本稿では、ネットワーク内のノードを平等に扱い分析する従来の手法とは一線を画し、焦点を当てたいノードに注目度を設定することで分析の解像度を制御する枠組みを提案した。具体的には、機能コミュニティ抽出法に対して初期ベクトルを注目度ベクトルとすることで、対象ノード集

合をより詳細に分類し、他のノードを粗く分類する手法を提案した。この手法において、ノード集合の選び方により得られる処理結果の特徴がどのように変化するかは自明ではない。

本稿では、既存研究で用いているネットワークデータを対象に、提案法を適用し、抽出される機能コミュニティの性質および有効性について評価した。実験より、対象ノード集合がネットワークに分散している場合は、各対象ノードの周辺ノードを細かく分類し遠いノードを粗く分類するという結果が得られた。対象ノード集合のノードどうしが互いに近傍に存在する場合は、対象ノード集合内のノードを細かく分類し、その他のノードを粗く分類するという結果が得られ、分析の解像度を制御できることが示唆された。

今後は、実問題への応用を視野に入れて、多様なネットワークで提案法の有効性を検証していくつもりである。さらに、対象とするノード集合の選び方や、注目度ベクトルとして設定する確率分布による処理結果の違いについて、より詳細に分析していくつもりである。

謝辞 本研究は、科学研究費補助金基盤研究 (C) (23500128) の支援を受けて行ったものである。

## 参考文献

- [1] Newman, M.E.J. and Park, J.: Why social networks are different from other types of networks, *Phys. Rev. E*, Vol.68, No.3, p.036122 (online), DOI: 10.1103/PhysRevE.68.036122 (2003).
- [2] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol.1, No.3, pp.215–239 (online), DOI: 10.1016/0378-8733(78)90021-7 (1979).
- [3] Newman, M.E.J.: Detecting community structure in networks, *The European Physical Journal B – Condensed Matter and Complex Systems*, Vol.38, No.2, pp.321–330 (online), DOI: 10.1140/epjb/e2004-00124-y (2004).
- [4] 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).
- [5] Saito, K., Yamada, T. and Kazama, K.: The k-Dense Method to Extract Communities from Complex Networks, *Mining Complex Data*, Zighed, D., Tsumoto, S., Ras, Z. and Hacid, H. (Eds.), *Studies in Computational Intelligence*, Vol.165, pp.243–257, Springer Berlin/Heidelberg (2009).
- [6] Seidman, S.B.: Network structure and minimum degree, *Social Networks*, Vol.5, No.3, pp.269–287 (online), DOI: DOI:10.1016/0378-8733(83)90028-X (1983).
- [7] 伏見卓恭, 斉藤和巳, 風間一洋: ネットワーク機能コミュニティ抽出法, 日本データベース学会論文誌, Vol.10, No.3, pp.13–18 (2012-02).
- [8] 伏見卓恭, 斉藤和巳, 風間一洋: 機能性に基づくコミュニティ抽出法の比較, 情報処理学会論文誌 データベース, Vol.5, No.3, pp.26–35 (2012).
- [9] 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋: 異なる視点からのノード機能に基づくコミュニティ抽出法, 日本データベース学会論文誌, Vol.11, No.3, pp.27–32 (2013).
- [10] Langville, A.N. and Meyer, C.D.: Deeper inside pagerank, *Internet Mathematics*, Vol.1, p.2004 (2004).

<sup>\*2</sup> [http://www.city.shizuoka.jp/deps/tosikeikaku/koshin\\_toshikei\\_toshimasu\\_shizuoka\\_2syo\\_kakudai\\_2.3.b.htm](http://www.city.shizuoka.jp/deps/tosikeikaku/koshin_toshikei_toshimasu_shizuoka_2syo_kakudai_2.3.b.htm)

<sup>\*3</sup> ごく一部の道路（東名高速道路やバイパス）は他の道路と立体交差するが、ほとんどの道路は他と立体交差ししない。



- [11] Ravasz, E. and Barabási, A.L.: Hierarchical organization in complex networks, *Physical Review E*, Vol.67, No.2, p.026112+ (online), DOI: 10.1103/PhysRevE.67.026112 (2003).
- [12] Zachary, W.: An information flow model for conflict and fission in small groups, *Journal of Anthropological Research*, Vol.33, pp.452-473 (1977).
- [13] 全国デジタル道路地図データベース静岡市版 (2011).
- [14] Kamada, T. and Kawai, S.: An algorithm for drawing general undirected graphs, *Inf. Process. Lett.*, Vol.31, pp.7-15 (1989).
- [15] 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).



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。2005年京都大学大学院情報学研究科システム科学専攻博士課程修了。2012年和歌山大学システム工学部教授、現在に至る。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。

(担当編集委員 大野 成義)



伏見 卓恭 (学生会員)

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在学中。2011年静岡県立大学大学院経営情報学研究科修士課程修了。複雑ネットワークの研究に従事。電子情報通信学会、日本データベース学会、人工知能

学会各学生会員。



斉藤 和巳 (正会員)

静岡県立大学経営情報学部教授。1985年慶応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)。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。著書に「ウェブサイエンス入門—インターネットの構造を解き明かす」(NTT出版)。



池田 哲夫 (正会員)

1979年東京大学理学部情報科学科卒業。1981年東京大学大学院理学系研究科情報科学専攻修士課程修了。同年日本電信電話公社(現、NTT)電気通信研究所入所。2002年岩手県立大学ソフトウェア情報学部教授。2006年静岡県立大学経営情報学部教授。専門は、データベース工学、情報検索、社会情報システム等。博士(工学)(東京大学)。電子情報通信学会、日本データベース学会、言語処理学会、ACM各会員。