

# ネットワーク機能コミュニティ抽出法

## Extracting Functional Communities in Networks

伏見 卓恭<sup>▼</sup> 齊藤 和巳<sup>◆</sup>  
風間 一洋<sup>▲</sup>

Takayasu FUSHIMI Kazumi SAITO  
Kazuhiro KAZAMA

本論文では、新たな視点でのネットワーク構造の把握を目的に、従来のリンク密度に基づくコミュニティとは異なる、機能コミュニティと呼ぶサブネットワークの抽出を課題とする。機能コミュニティは、連結性を仮定しない、ネットワークにおけるノードの相対的位置や役割、階層的地位、次数などの性質が同じノード群を意図する。この目的のため、各ノードの PageRank 収束過程のスコア推移に基づき、ノードをクラスタリングする方法を提案する。現実のハイパーリンクネットワーク及び代表的な人工ネットワークを用いた評価実験により、ノードの相対的位置や階層的地位などによりノードが分類された機能コミュニティを抽出できることを示す。

To uncover the functions and the latent structure of large-scale complex networks, the task of extracting community is an important research topic. In this paper, we focus on nodes' functions such as relative locations in network, roles, statuses, degree and so forth, and try to extract functional community which consists of nodes with the same function in order to uncover network structure with a fresh eye. The functional community is different from community based on link density, in terms of not supposing connectivity of nodes in the functional community. To extract the functional community, we propose a new nodes-clustering method using the PageRank score convergence curve. In our experiments using representative artificial networks and real web hyperlinks networks, we show how our proposed method behaves and that proposed method can extract meaningful functional communities.

### 1. はじめに

近年、大規模な複雑ネットワークの研究が盛んに行われている。人間関係ネットワークやWebのハイパーリンクネットワークにおいて、重要ノードを抽出するための手法は広く知られているが、実ネットワークは大規模かつ複雑な構造を有

するため、背後に隠れる機能や骨格となる構造を解明することは困難である。さらに、ネットワークを構成する各ノードはそれぞれ固有の機能や役割などを有し、他のノードと相互に影響を及ぼし合っている。しかし、つながり方にはある種の傾向、すなわち、似た性質を持つノード間にはリンクが生まれやすいという特徴を有すると考えられている[1]。さらにソーシャル・ネットワークは、コミュニティ構造という特徴を有する[2]。コミュニティ構造もある種のリンク傾向から生成されると考えられる。大規模で複雑なネットワークの骨格となる構造や機能を分析する際、ノード同士が結合し合うコミュニティを抽出するタスクは重要な研究課題であり、今までに様々なコミュニティ抽出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の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]。さらに、スペクトラルグラフ分析の手法であるNormalized Cut法[4]やRatio Cut法[5]なども挙げられる。これらは、クラスタ内リンクを多く、クラスタ間リンクを少なくする、すなわち、ノード同士の結合が疎な部分を切断し、いくつかのノード集合(サブネットワーク)に分割する方法である。一方、ネットワーク上のノード同士が密結合したような緻密な部分をコミュニティと見なして、完全部分グラフであるクリーク(clique)やその条件を緩和したサブネットワークを抽出するために様々な手法が提案されている[6][7][8]。

これらを代表とする既存のコミュニティ抽出手法の多くは、無向ネットワークにおけるリンク構造の疎密に注目している。本研究では、ネットワークを構成する各ノードの機能に着目する。ネットワーク内におけるノードの相対的位置、他のノードやネットワークに果たす役割などをノードの基本的な機能と考える。例えば、会社組織内の人間関係ネットワークでは、最終的な意思決定や大きな経営判断を行う社長や、トップからの命令を部下に伝達する部長、部下に具体的な命令を下し、部下からの提案を上司へ伝える課長、協力して任務を実行する社員など、ネットワーク内で果たす役割がある。同様にWebページのハイパーリンクネットワークでも、様々なコンテンツへの玄関口的役割を果たすトップページ、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な役割が存在する。

本論文では、このような各ノードの役割などをノードの機能としてとらえ、同質の機能を有するノード集合を機能コミュニティと定義し、機能コミュニティを抽出することを課題とする。たとえば、A部門の部長、B部門の部長など、部長という機能を持つノードで構成される部長コミュニティを抽出することを目的とする。従来の手法では、リンク密度に基づき連結性を仮定したようなノード集合をコミュニティとして考えているため、各部門をコミュニティとしてとらえ、部門コミュニティを抽出することは可能であるが、水平的な部長コミュニティを抽出することは困難である。このような視点からコミュニティを抽出することにより、従来のコミュニティ抽出とは異なるネットワーク構造の把握が可能になると考える。

本論文では、同質の機能を有するノード群を同一のコミュニティとして抽出する手法を提案する。提案法では無向ネットワークを対象として、Webページのランキングアルゴリズムとして有名なPageRank[9]の収束過程の変化曲線によりノードをクラスタリングし、機能コミュニティを抽出する。

<sup>▼</sup> 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 [j11507@u-shizuoka-ken.ac.jp](mailto:j11507@u-shizuoka-ken.ac.jp)

<sup>◆</sup> 正会員 静岡県立大学経営情報学部  
[k-saito@u-shizuoka-ken.ac.jp](mailto:k-saito@u-shizuoka-ken.ac.jp)

<sup>▲</sup> 非会員 日本電信電話株式会社未来ねっと研究所  
[kazama@ingrid.org](mailto:kazama@ingrid.org)

PageRankによる各ページ（ノード）のスコアは、推移確率行列の定常ベクトルの対応する要素の値によって定義される。大域ジャンプ確率  $\alpha = 0$  のとき、無向ネットワークでの各ノードのスコアは各ノードの次数のみに依存する。しかし、定常ベクトルへの収束の過程は各ノードによって異なり、周辺のノードとの関係やつながり方、ネットワーク構造が影響する。提案法では、各ノードの定常ベクトルへの収束の変化曲線の類似性を用いてノードをクラスタリングする。直観的には、時刻変化に伴う各ノードのスコアの変化傾向により、ノードをクラスタリングすることになる。

提案法により抽出された機能コミュニティがどのような特徴を有しているかを検証するために、リンク密度に基づくコミュニティ抽出法であるNewmanクラスタリング法と、抽出されたコミュニティの可視化結果を定性的に比較する。代表的なネットワークモデルにより生成した人工ネットワークおよび現実のネットワークデータを用いて評価する。

本論文は以下のような構成である。提案法のアルゴリズムおよびPageRankのスコア収束過程をクラスタリングするK-median法について2節で説明する。そして、提案法の可視化による評価実験の結果を3節で示す。最後に本研究のまとめと今後の展望を4節で述べる。

## 2. 機能コミュニティ抽出法

本節では1節で述べた機能コミュニティを抽出する提案法について詳細に説明する。

### 2.1 提案アルゴリズム

ノードの機能、地位、階層や役割は、周辺ノードとの隣接関係、周辺ノードの次数、ネットワーク内での相対的な位置などの影響を受ける。一方PageRankは、任意のノードからスタートし、各ステップでそのノードに到達する期待値を計算している。1節での例を用いると、任意の社員からメールを送り、各ステップでメールが到達する期待値を計算していることになる。提案法は、PageRankによる各ステップでの期待値（収束曲線）を特徴ベクトルとし、そのコサイン類似度により、ノードをクラスタリングすることでコミュニティを抽出する。ただし、同じ機能コミュニティに属しているノードであっても、収束曲線が類似しているだけで、最終的な収束結果のPageRankスコアが近いわけではなく、大きく異なることも多い。すなわち、PageRankスコアでクラスタリングしているわけではない。

また、本論文ではクラスタリングの方法として、K-median法を採用する。なぜなら、一般的に、平均（mean）より中央値（median）の方が外れ値に対して頑健であることが知られており、K-medianクラスタリングの貪欲解法は、解の一意性が保障されている。さらに、貪欲解法の目的関数のサブモジュラ性より、厳密解ではないものの、ある程度妥当な精度で最悪ケースの解品質が理論的に保証されている[10]。

提案法は、PageRankスコアの収束曲線を各ノードの特徴ベクトルとし、K-median法によりクラスタリングするため、特徴ベクトルの次元  $T$  およびクラスタ数  $K$  がパラメータとなる。従ってアルゴリズムの入力は、ネットワーク  $G=(V, E)$ 、特徴ベクトルの次元パラメータ  $T$  およびクラスタ数パラメータ  $K$  である。以下に提案法のアルゴリズムを示す。

(1) 特徴ベクトル計算：各ノード  $u$  の機能を表すPageRankスコアの  $T$  次元収束曲線  $\mathbf{x}_u$  を計算する；

(2) 類似度計算：各ノードの収束曲線間のコサイン類似度  $\rho(u, v)$  を計算する；

(3) クラスタリング：ノード間の類似度  $\rho(u, v)$  から、K-median法によりノードをクラスタリングする；

(4) 出力：クラスタ集合  $\mathcal{P}_K = \{p_1, \dots, p_K\}$  を出力する；

(1) のPageRankアルゴリズムによる収束曲線の計算法および、(3) のK-median法によるクラスタリング法の詳細について以降で説明する。

### 2.2 PageRank スコア収束過程

この節では、各ノードのPageRankスコアを計算する方法を述べる。PageRankは本来、Webページ群のような有向ネットワークとするが、本研究では無向ネットワークを対象として説明する。

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。自己リンクを持つノードもあり、その場合  $a(v, v) = 1$  となる。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。自己リンク付きノード  $u$  は  $u \in \Gamma(u)$  である。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  $\mathbf{y}(v) \geq 0$  で  $\sum_{v \in V} \mathbf{y}(v) = 1$  となる。繰り返しステップのパラメータ  $t$  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：

$$\begin{aligned} \mathbf{y}_t^T &= \mathbf{y}_{t-1}^T \left( (1-\alpha)\mathbf{P} + \alpha \mathbf{e} \mathbf{z}^T \right) \\ &= (1-\alpha)\mathbf{y}_{t-1}^T \mathbf{P} + \alpha \mathbf{z}^T \end{aligned} \quad (1)$$

ここで  $\mathbf{b}^T$  は  $\mathbf{b}$  ベクトルの転置を表わし、 $\mathbf{e} = (1, \dots, 1)^T$  である。このモデルは確率  $\alpha$  で、ユーザは確率分布  $\mathbf{z}$  に従って大域ジャンプすること（ランダムサーファージャンプ）を意味する。 $\mathbf{z}$  は  $z(v) > 0$  で  $\sum_{v \in V} z(v) = 1$  となるような確率分布である。行列  $(1-\alpha)\mathbf{P} + \alpha \mathbf{e} \mathbf{z}^T$  は Google 行列と呼ばれている。標準的な PageRank では、適切に初期化された  $\mathbf{y}_0$  を用いて式(1)の更新式により  $\mathbf{y}$  を更新する。また、ノード  $u$  のスコアに注目すると、

$$\begin{aligned} y_t(u) &= (1-\alpha) \sum_{v \in \Gamma(u)} \left\{ y_{t-1}(v) \cdot p(v, u) \right\} + \alpha \cdot z(u) \\ &= (1-\alpha) \sum_{v \in \Gamma(u)} \left\{ \frac{y_{t-1}(v)}{|\Gamma(v)|} \right\} + \alpha \cdot z(u) \end{aligned} \quad (2)$$

で計算される。

本研究では、ネットワーク上で直接隣接するノードとの関係に着目し、大域ジャンプ確率  $\alpha = 0$  とし説明するが、一般的な  $\alpha \neq 0$  のPageRankへの拡張は容易である。また、無向ネットワークを対象とした場合、 $\alpha = 0$  のPageRankスコアベクトルは、基本的な意見形成モデルとして知られるVoter Modelの影響度ベクトルと一致する[11]。大域ジャンプ確率  $\alpha = 0$  とし、各反復でのノード  $u$  のスコアは、

$$y_t(u) = \sum_{v \in \Gamma(u)} \frac{y_{t-1}(v)}{|\Gamma(v)|} \quad (3)$$

で計算される。無向ネットワークにおいて、 $\alpha = 0$  でのノード  $u$  のスコアの極限值は、ノード  $u$  の次数  $|\Gamma(u)|$  により決定される。

$$y_{\infty}(u) = \frac{|\Gamma(u)|}{\sum_{v \in \Gamma(u)} |\Gamma(v)|} \quad (4)$$

このことは、式(4)を式(3)の  $y_{t-1}(u)$  に代入すると、

$$y_t(u) = \sum_{v \in \Gamma(u)} \left\{ \frac{1}{|\Gamma(v)|} \cdot \frac{|\Gamma(v)|}{\sum_{w \in \Gamma(v)} |\Gamma(w)|} \right\} \\ = \frac{|\Gamma(u)|}{\sum_{w \in \Gamma(u)} |\Gamma(w)|} \quad (5)$$

となり、明らかである。

反復を繰り返し、各ノードのスコアは式(4)に収束する。反復回数  $T$  まで反復を繰り返し、各反復回数でのノード  $u$  のスコアを要素としたベクトルを  $\mathbf{x}_u = (y_1(u), y_2(u), \dots, y_T(u))^T$  と定義する。このベクトル  $\mathbf{x}_u$  をノード  $u$  の収束曲線と呼びノード  $u$  の特徴ベクトルとする。各ノードの収束するスコアは、各ノードの次数のみで決まるが、一般に収束曲線は次数のみでは決まらない。周辺ノードの影響や周辺ノードとの相対的な位置関係、ネットワーク構造の影響を受ける。提案法では、初期ベクトル  $\mathbf{y}_0 = (1/|V|, \dots, 1/|V|)^T$  とする。

### 2.3 K-median 法

$K$ -median ( $K$ -medoidとも呼ばれる)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  $V$  が与えられたとき、オブジェクト集合を  $K$ 個のクラスタに分割する手法である。任意のオブジェクトペア  $u, v \in V$  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し、類似度の高い(距離の小さい)オブジェクトペアは同じクラスタに、類似度の低い(距離の大きい)オブジェクトペアは異なるクラスタに属するように分割する。 $K$ -medianの解法には反復法や貪欲法があるが、本研究では代表オブジェクトが計算ごとに不変である貪欲法を使用する。貪欲法とは、既に選定した代表オブジェクトを固定し、ある目的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することで、結果の代表オブジェクト集合を求める方法である。各オブジェクトは、最も類似度の高い代表オブジェクトと同じクラスタに割り当てられる。既に選定した代表オブジェクト集合を  $P$  とし、新たに追加を試みるオブジェクトを  $w$  とするとき、本論文では、以下の目的関数  $f(\cdot)$  を考える。

$$f(P \cup \{w\}) = \sum_{v \in V} \max\{\mu(v, P), \rho(v, w)\}. \quad (6)$$

ここで、 $\mu(v, P)$  は既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、 $\mu(v, P) = \max_{w \in P} \{\rho(v, w)\}$  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)  $k \leftarrow 1, P_0 \leftarrow \emptyset$ , 各オブジェクト  $v \in V$  に対し、 $\mu(v; \emptyset) \leftarrow 0$  と初期化する；
- (2) 式(6)で  $\hat{p}_k = \operatorname{argmax}_{w \in V \setminus P_{k-1}} \{f(P_{k-1} \cup \{w\})\}$  を求め、 $P_k \leftarrow P_{k-1} \cup \{\hat{p}_k\}$  とする；
- (3)  $k = K$  ならば  $\hat{P}_K = \{\hat{p}_1, \dots, \hat{p}_K\}$  を出力し終了する；
- (4) 各オブジェクト  $v \in V$  に対し、 $\mu(v; P_k)$  を求め、 $k \leftarrow k + 1$  とステップ(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  $p_k$  のクラスタに割り当てる。

### 3. 評価実験

本節では、現実の Web ネットワークデータなどを対象に、提案法を用いて機能コミュニティを抽出する。提案法および比較のための Newman クラスタリング法による抽出結果を可視化により定性的に評価する。

#### 3.1 ネットワークデータ

実験では、機能コミュニティ抽出法の特性を評価するために、2つの代表的な人工ネットワークおよび2つの実ネットワークを用いる。

1つ目の人工ネットワークは、Hierarchical ネットワークである。HN モデルは、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モデルである[12]。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、その他のノード同士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ほか)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ほか)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。このような性質を有するネットワークを HN モデルにより生成し、本論文では Hierarchical ネットワークと呼ぶ。

2つ目の人工ネットワークは、Karlsruhe ネットワークである。Karlsruhe はドイツの都市の名で、宮殿を中心に道路網が放射状に伸びる計画都市で中心から同心円状に道路網が広がり、中心からいくつかのスポークが放射状に伸びており、この都市の道路網の交差点をノード、交差点間の道路をリンクとしてモデル化したネットワークを Karlsruhe ネットワークと呼ぶ。本論文では、提案法の性質を評価するために、同心円状のリング数を 10、スポーク数を 10 とし、中心ノードを含めた 101 ノードのネットワークを 2つ作り、最外部のノードの 1つ同士を結んだネットワークを用いる。

実ネットワークとして、複数の国公立大学のウェブサイト内のページを 2010 年 8 月に収集し、各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た。Newman クラスタリング法との比較のため、ハイパーリンクネットワークを無向化した。本論文では、スペースの都合上 Hosei ネットワーク(法政大学情報科学部 <http://cis.k.hosei.ac.jp/>)および Keijo ネットワーク(静岡県立大学経営情報学部 <http://ai.u-shizuoka-ken.ac.jp/>)の 2つの大学 Web サイトのハイパーリンクネットワークに対する結果を示す。

#### 3.2 Newman クラスタリング法

評価実験では、提案法の性質を検証するため、リンク密度に基づく代表的なコミュニティ抽出法である Newman クラスタリング法[13]と比較する。以下に Newman クラスタリング法のアイデアについて簡単に触れる。

Newman クラスタリング法では、コミュニティ抽出の度を  $\text{Modularity}$  という定量的指標により評価している。 $K$ 個のコミュニティに対する  $\text{Modularity}$  は、コミュニティ  $i$  と  $j$  間のリンク数の総リンクに対する割合  $e_{ij}$  を要素とする  $K \times K$  の対称行列  $\mathbf{E}$  を定義し、

$$Q = \sum_{i=1}^K (e_{ii} - a_i^2) = \operatorname{Tr}(\mathbf{E}) - \|\mathbf{E}\|^2 \quad (7)$$

で計算される.

ここで,  $a_i = \sum_{j=1}^K e_{ij}$  であり,  $\|\mathbf{B}\|$  は行列  $\mathbf{B}$  の要素の和 (L1 ノルム) である. この値が高ければ, 同一コミュニティ内のノード間にリンクが相対的に多いことになる.

具体的なコミュニティの抽出方法は, 階層的クラスタリングと同様にデンドログラムを用いて, 最初は各ノードが単一でコミュニティを形成し, Modularity が最も増加するノード同士を結合しコミュニティを形成するステップを繰り返して, Modularity が最も高くなるようなステップ数およびコミュニティ数を出力する.

### 3.3 実験結果と考察

上述したネットワークに対し,  $T = 10,000$  とし提案法により機能コミュニティを抽出した結果を図 1 および図 3 に, Newman クラスタリング法による結果を図 2 および図 4 にそれぞれ示す. なお説明の便宜上, ある特定のクラスタ数  $K$  を図示しているが, 他の  $K$  の場合でもほぼ同様の結果が得られている.

Hierarchical ネットワークの結果 (図 1 左および図 2 左) を比較すると, 提案法では階層上の同質 (同一階層) のノード同士が同じ色に塗り分けられており, 同質の機能を持つノード群が同一のコミュニティとして抽出されていることがわかる. 1 節で例示したような, 部長コミュニティを抽出できている. 一方 Newman クラスタリング法の結果は, リンク密度によるコミュニティ抽出のため, 密に隣接するノード同士が同一のコミュニティとして抽出されていることがわかる. 1 節で例示したような, 部門コミュニティを抽出している.

Karlsruhe ネットワークの結果 (図 1 右および図 2 右) を比較すると, 提案法ではネットワークにおける相対的位置の等しいノード同士が同じ色に塗り分けられており, 同質の機能を持つノード群が同一のコミュニティとして抽出されている. 一方 Newman クラスタリング法の結果では, 2 つの円間の疎なリンクで分断され, 概ね左下の円と右上の円の 2 つのコミュニティに分けられている.

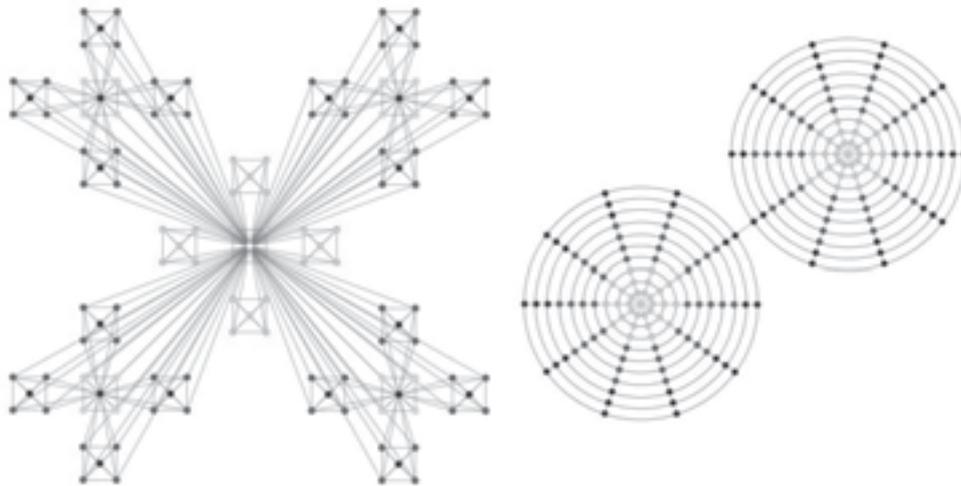


図 1 人工ネットワーク 機能コミュニティ抽出結果

Fig.1 Results of Functional Communities of Artificial Networks

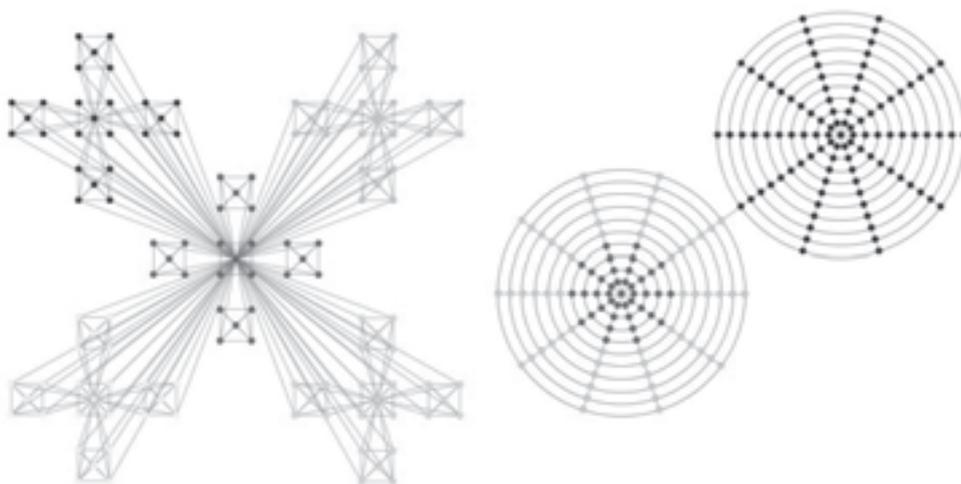


図 2 人工ネットワーク Newman クラスタリング結果

Fig.2 Results of Newman Clustering for Artificial Networks

上記の人工ネットワークに対する実験結果より、提案法はノードの階層的地位や相対的位置といった同一の機能を持つノード群を同一の機能コミュニティとして抽出できることがわかり、2つのコミュニティ抽出手法の性質の顕著な違いを見ることができる。

実ハイパーリンクネットワークに対する実験結果を図3および図4に示す。各ネットワークはばねモデルにより可視化した[14]。

Hosei ネットワークの結果(図3左および図4左)を比較すると、提案法では可視化結果の四角で囲まれた部分は同じ色で塗り分けられている。このノード群は、対象大学の各年度の教員の研究成果報告ページであり、ノードの機能としては同質であると考えられ、同一のコミュニティとして抽出できている。一方 Newman クラスタリング法は、年度ごとに異なるコミュニティとして抽出している。これは、異なる年度のノード同士が直接リンクされていないために、異なるコミュニティとして抽出されている。

Keijo ネットワークの結果(図3右および図4右)を比較すると、提案法では可視化結果の中央の塊部分は同じ色で塗り分けられている。このノード群(Web ページ群)は、対象大学の News ページであり、相互にはリンクを張りあっていないが、リンク構造はほぼ同質であり、トップページからの深さは等しい。提案法は、このようなノード群を同

一のコミュニティとして抽出している。一方 Newman クラスタリング法は、直接つながりっていない News ページのノード群の一部を、異なるコミュニティとして抽出している。

実ハイパーリンクネットワークに対する実験結果より、提案法はトップノードからの深さやリンク構造の類似性など、同質の機能を持つノード群をコミュニティとして抽出できることが示された。

#### 4. おわりに

本研究では、従来のコミュニティの概念とは異なる機能コミュニティに着目し、ネットワークにおける同質の機能を持つノード群を抽出する方法を提案した。

提案法は、各ノードの機能を表す特徴ベクトルを PageRank スコアの収束曲線により定義し、特徴ベクトル間のコサイン類似度によりノードをクラスタリングする方法である。評価実験により、提案法は同質の機能を持つノード群をコミュニティとして抽出できることを示した。また既存手法との比較により、提案法により抽出されるコミュニティ性質の違いを定性的に検証した。提案法は、連結性を仮定しないため、同一階層のノード群やネットワーク内の相対的な位置が類似したノード群をコミュニティと

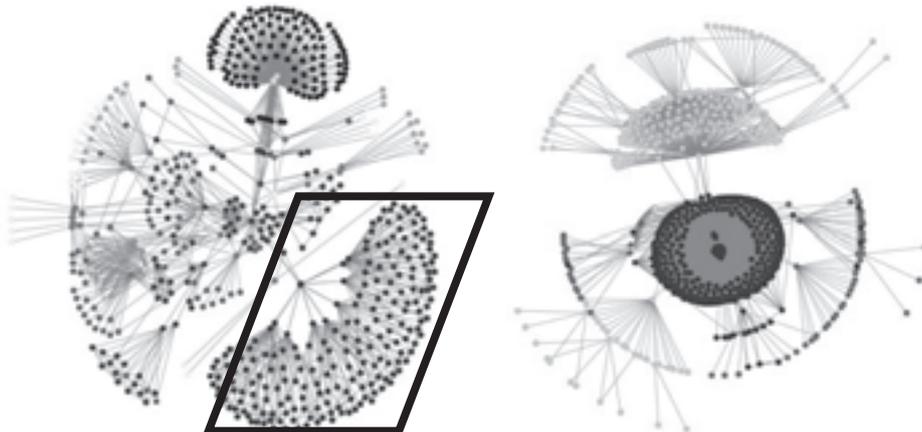


図3 ハイパーリンクネットワーク 機能コミュニティ抽出結果

Fig.3 Results of Functional Communities of Real Web-sites Networks

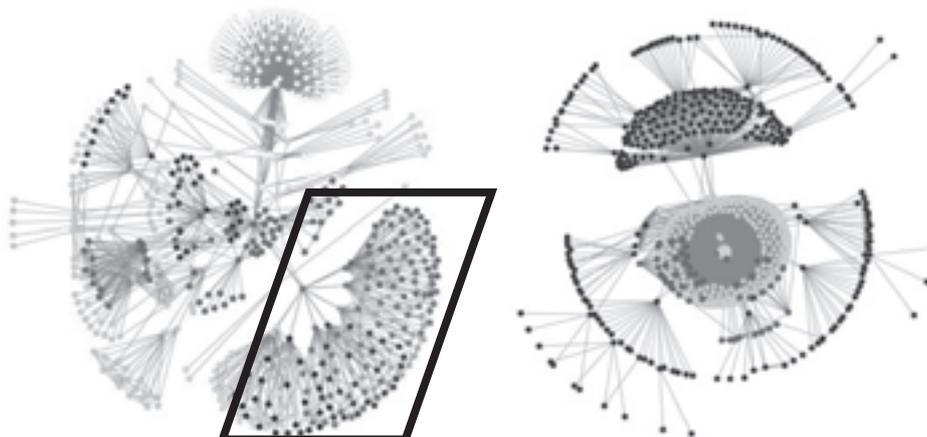


図4 ハイパーリンクネットワーク Newman クラスタリング結果

Fig.4 Results of Newman Clustering for Real Web-sites Networks

して抽出でき、水平的コミュニティを抽出可能にした。

さらに、PageRankによりノードの収束曲線を計算しているため、当該分野ではあまりない、一般的な有向ネットワークや重み付きのネットワークなどへの自然な拡張が期待できる。また、 $K$ -median クラスタリング法の  $K$  の値を変化させることで、コミュニティを細分化できるという利点を有する手法である。

提案法は、注目するノードがあった時に、そのノードと類似した機能・役割のノードを有する部分ネットワークの抽出などが可能である。例として、研究組織内のゲートキーパ(組織外部の情報源に精通しており、外部情報源から得た情報を効果的に組織のメンバーに伝えるという重要な役割を有するノード)に注目する。あるゲートキーパが属する機能コミュニティには、他の部門のゲートキーパも属している。ゲートキーパを発見して支援するような体制を整え、各部門内のコミュニケーションが円滑に行えるようになり、組織全体の効率向上に貢献できる。

ゲートキーパは媒介中心性が高いため、媒介中心性が高いノード集合を抽出すればゲートキーパ群を近似的に抽出できる。しかし、媒介度が高いノードと低いノードには分類できるが、他の機能を有するノードを考える場合、他の属性値を用いなければならない。すなわち、中心性概念のような単一尺度では、特定の機能を判定することしかできない。提案法は特定機能ノード群を抽出するのではなく、機能ごとにクラスタリングするため、その点で効率的な手段と言える。

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。

#### [謝辞]

本研究は、NTT 未来ねっとと研究所との共同研究、及び、科研費(23500128)の支援を受けて行ったものである。

#### [文献]

- [1] M.E.J. Newman.: "Mixing patterns in networks." *Physical Review E*, Vol.67, No.2, pp. 026126+ (2003).
- [2] M.E.J. Newman and Juyong Park.: "Why social networks are different from other types of networks." *Phys. Rev. E*, Vol.68, No.3, p. 036122 (2003).
- [3] M.E. J. Newman.: "Detecting community structure in networks." *The European Physical Journal B - Condensed Matter and Complex Systems*, Vol.38, No.2, pp. 321-330 (2004).
- [4] J. Shi and J. Malik.: "Normalized Cuts and Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.8, pp. 888-905 (2000).
- [5] L.Hagen and A.B. Kahng.: "New spectral methods for ratio cut partitioning and clustering." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol.11, No.9, pp. 1074-1085 (1992).
- [6] G.Palla, I.Derenyi, I.Farkas, and T.Vicsek.: "Uncovering the overlapping community structure of complex networks in nature and society." *Nature*, Vol. 435, pp. 814-818 (2005).
- [7] K. Saito, T. Yamada, and K. Kazama.: "The k-dense method to extract communities from complex networks." In Djamel Zighed, Shusaku Tsumoto, Zbigniew Ras,

- and Hakim Hacid, editors, *Mining Complex Data*, Vol. 165 of *Studies in Computational Intelligence*, pp. 243-257. Springer Berlin / Heidelberg (2009).
- [8] S. B. Seidman.: "Network structure and minimum degree." *Social Networks*, Vol.5, No.3, pp. 269 -287 (1983).
- [9] A. N. Langville and C. D. Meyer.: "Deeper inside pagerank." *Internet Mathematics*, Vol.1, p. 2004 (2004).
- [10] G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher.: "An analysis of approximations for maximizing submodular set functions." *Mathematical Programming*, Vol.14, pp. 265-294 (1978).
- [11] T. Fushimi, K. Saito, M. Kimura, H. Motoda, and K. Ohara.: "Finding relation between pagerank and voter model." In *Proceedings of the 11th international conference on Knowledge management and acquisition for smart systems and services*, PKAW'10, pp. 208-222, Berlin, Heidelberg, Springer-Verlag (2010).
- [12] E. Ravasz and A. L. Barabasi.: "Hierarchical organization in complex networks." *Physical Review E*, Vol.67, No.2, pp. 026112+ (2003).
- [13] A. Clauset, M.E.J. Newman, and C. Moore.: "Finding community structure in very large networks." *Physical Review E*, Vol.70, No.6, pp. 066111+ (2004).
- [14] T.Kamada and S.Kawai.: "An algorithm for drawing general undirected graphs." *Inf. Process. Lett.*, Vol.31, pp. 7-15 (1989).

#### 伏見 卓恭 Takayasu FUSHIMI

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 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。電子情報通信学会学生会員。日本データベース学会学生会員。

#### 齊藤 和巳 Kazumi SAITO

静岡県立大学経営情報学部教授。1985 慶応義塾大学工学部数理科学科数学専攻卒業、1998 東京大学 博士(工学)。複雑ネットワークの研究に従事。情報処理学会、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)。

#### 風間 一洋 Kazuhiro KAZAMA

NTT 未来ねっとと研究所主任研究員。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。博士(情報学)。Web 情報検索、Web マイニングの研究に従事。情報処理学会、人工知能学会、日本ソフトウェア科学会、ACM 各会員。