情報処理学会論文誌　数理モデル化と応用　Vol.6　No.2　137–146 (Aug. 2013)

リンクの向きに着目した機能コミュニティとモチーフの関係分析

伏見 卓恭[1,a)]　斉藤 和巳[1]　池田 哲夫[1]　風間 一洋[2]

受付日 2012年11月2日, 採録日 2013年2月5日

概要：本稿では，情報の発信や受信，片思いや両想いなどリンクの向きとして表れるノードの役割に着目し，有向ネットワークから機能的に類似するノード群で構成される機能コミュニティを抽出する手法を提案する．提案法の有効性を検証するとともに，無向化したネットワークに対する結果と比較し違いを分析する．また，ネットワークの局所的なリンク構造を分析するネットワークモチーフを用いた手法とも比較する．複数のネットワークを用いた評価実験から，無向化したネットワークに対する分析やモチーフを用いた手法では抽出できないノードの機能を，提案手法では抽出できることを示す．さらに，提案法における PageRank 計算時の大域ジャンプ確率の大小が処理結果を左右するため，実ネットワークを用いて本手法に最適な大域ジャンプ確率を検討する．評価実験より，大域ジャンプ確率を $\alpha \simeq 0$ にすると，有向ネットワーク内のノード群が有する多様な機能によるコミュニティ抽出結果が得られることも示す．

キーワード：機能コミュニティ，有向ネットワーク，PageRank，大域ジャンプ確率，ネットワークモチーフ

An Analysis of the Relation between Functional Community and Network Motif Based on Link Directions

Takayasu Fushimi[1,a)]　Kazumi Saito[1]　Tetsuo Ikeda[1]　Kazuhiro Kazama[2]

Received: November 2, 2012, Accepted: February 5, 2013

Abstract: In this paper, in order to detect nodes' functions such as sending/receiving information, one-way/bidirectional relationships and so forth, we propose a method for extracting communities each of which consists of functionally similar nodes from directed networks. We confirm effectiveness and usefulness of our proposed method in comparison with two methods, a standard functional community extraction method intended for undirected networks and a method based on network motif analysis which reveals local link structures. From our experimental results using artificial and real networks, we show that our method can extract some reasonable functional communities which can not be extracted by two comparison methods. We also analyze the values of global jump probabilities which affect the results of community extraction in the PageRank calculation step of our proposed method. We show that, when we set the values of global jump probabilities as $\alpha \simeq 0$, then, we can obtain reasonable communities by various function of nodes from our experiments.

Keywords: functional community, directed network, PageRank, global jump, network motif

1. はじめに

Web 技術の発展を契機に，Web 上でも多くの複雑ネットワークが見受けられるようになってきている．これらのネットワークにおいて，すべてのノードは均質ではなく，各ノードは固有の立場や役割，機能を有している．このよ

[1] 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
[2] 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

うな性質に基づき，多大なノード群をクラスタリングしたり，重要ノードを抽出したりするための手法が提案されている [1], [2], [3]．ネットワーク構造に関しても，全体が均質ではなく，リンクが密な部分もあれば疎な部分もあり，コミュニティ構造を有することが指摘されている [4]．

本稿では，ネットワークに対する各ノードの役割・機能・立場が類似したノードからなるコミュニティを抽出する手法を扱う．周辺ノードとのリンク関係の類似性，階層的地位，相対的位置などが類似するノードを抽出する方法として，著者らの機能コミュニティ抽出法がある [5], [6]．この方法は，無向ネットワークに対して，ネットワーク全体でのランダムウォークにより類似経路構造を探す方法であり，PageRank 反復計算時の PageRank スコア変化曲線の類似性を用いる手法である．この方法により，会社組織内のネットワークやウェブサイト内のハイパーリンクネットワークなどのような階層性を有するネットワークから，類似した立場にあるノード群を抽出できることが示されている．

一方，現実のネットワークには有向ネットワークも多く，これらを無向ネットワークに単純化して処理するとリンクの向きという情報が失われてしまうので，有向ネットワークのままで機能コミュニティを抽出することは重要である．本稿では，有向ネットワークを無向化するのではなく，有向ネットワークそのものから，機能が類似するノード群からなるコミュニティを抽出する手法を提案する．無向ネットワークを対象とした既存の機能コミュニティ抽出法では，ネットワーク内での相対的位置や階層的地位などの機能・役割に基づきノードを分類できる．一方，上司への相談や報告を密に行う社員や上司からの連絡を他の社員に知らせる社員など，同一階層の社員であっても役割は異なる場合がある．提案法では，このような無向化したネットワークを対象とした場合では分からない，片思いや両想いといったノード間関係の方向性や情報を送信/受信する役割など，リンクの向きとして表れる機能に基づきノードを分類することを目的としている．

既存の機能コミュニティ抽出法は，無向ネットワークにおいて直接隣接するノードへのランダムウォークのみを考慮しているが，出リンクを持たないノードを有したり，複数の強連結成分からなるような有向ネットワークを対象とする場合，PageRank による変化曲線計算時において，大域ジャンプ確率 α に適切な値を設定する必要がある．したがって，提案法における大域ジャンプ確率の値の設定についても考察をする．

また，有向ネットワークの局所的なリンク構造やリンクの向きに着目した分析指標としてネットワークモチーフがあげられる [7]．各ノードが，どのモチーフパターンにどれだけ含まれているかを要素としたベクトルの類似度は，局所的なリンク構造の類似性を反映できる．本稿で提案する有向機能コミュニティ抽出法では，ネットワークモチーフを用いた手法だけでは検出できないノードの機能が抽出可能であることも示す．

本稿は以下のような構成である．最初に機能コミュニティ抽出法のアルゴリズムとして，無向ネットワークおよび有向ネットワークに対する PageRank スコアの変化曲線を計算する方法をそれぞれ説明し，次いで変化曲線をクラスタリングする K-median 法について 2 章で説明する．そして，有向機能コミュニティ抽出法の有効性および有用性を検証するために，3 章で可視化による機能コミュニティ抽出結果を示すとともに，大域ジャンプ確率に関する一考察を述べ，最後に本研究のまとめと今後の展望を 4 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．最初に文献 [5] のように無向ネットワークに対する PageRank スコア変化曲線の計算法について説明し，その後有向機能コミュニティ抽出法における変化曲線の計算法について説明する．

機能コミュニティ抽出法は，ネットワーク $G = (V, E)$ とクラスタ数 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 $\{\mathbf{y}_1, \cdots, \mathbf{y}_T\}$ 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 $\mathbf{x}_v$ を構築；
(3) 各ノードペアの特徴ベクトル $\mathbf{x}_u$ と $\mathbf{x}_v$ 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 $\{\mathcal{C}_1, \cdots, \mathcal{C}_K\}$ を出力；

以下に詳細を説明する．

2.1 無向ネットワークに対する変化曲線

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 $a(u, v) = 1$，それ以外のとき $a(u, v) = 0$ 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．この手法では，初期ベクトルを $\mathbf{y}_0 = (1/|V|, \ldots, 1/|V|)^T$ とし，繰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：

$$\mathbf{y}_t^T = \mathbf{y}_{t-1}^T \mathbf{P} \tag{1}$$

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，

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

で計算される．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 PageRank スコアを要素としたベクトルを $\mathbf{x}_u = (y_1(u), y_2(u), \ldots, y_T(u))^T$ と定義する．このベクトル $\mathbf{x}_u$ をノード u の変化曲線と呼ぶ．

2.2　有向ネットワークに対する変化曲線

有向ネットワーク $G = (V, E)$ に対して，上記の無向ネットワークと同様に隣接行列 $\mathbf{A} \in \{0,1\}^{|V| \times |V|}$ を定義する．各ノード $u \in V$ に対して，ノード u の子ノード集合を $F(u) = \{v \in V; (u,v) \in E\}$，ノード u の親ノード集合を $B(u) = \{v \in V; (v,u) \in E\}$ とする．ここで，行推移確率行列 $\mathbf{P}$ の各要素は，$p(u,v) = a(u,v)/|F(u)|$ である．この手法では，初期ベクトルを $\mathbf{y}_0 = (1/|V|, \ldots, 1/|V|)^T$ とし，繰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：

$$\begin{aligned} \mathbf{y}_t^T &= \mathbf{y}_{t-1}^T \left((1-\alpha)\mathbf{P} + \alpha \mathbf{e}\mathbf{z}^T\right) \\ &= (1-\alpha)\mathbf{y}_{t-1}^T \mathbf{P} + \alpha \mathbf{z}^T. \end{aligned} \tag{3}$$

ここで $\mathbf{e} = (1, \cdots, 1)^T$ である．このモデルは確率 α で，ユーザは確率分布（大域ジャンプベクトル） $\mathbf{z}$ に従って大域ジャンプすることを意味する（ランダムサーファージャンプ）．$\mathbf{z}$ 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 (3) の更新式により $\mathbf{y}$ を更新する．ノード u に注目すると，

$$\begin{aligned} y_t(u) &= (1-\alpha) \sum_{v \in B(u)} \{y_{t-1}(v) \cdot p(v,u)\} + \alpha \cdot z(u) \\ &= (1-\alpha) \sum_{v \in B(u)} \left\{ \frac{y_{t-1}(v)}{|F(v)|} \right\} + \alpha \cdot z(u) \end{aligned} \tag{4}$$

で計算される．

また，出リンクを持たないぶら下がりノード（dangling node）u に対して，パーソナライズベクトル $\mathbf{g}$ を導入する．$\mathbf{g}$ は，$g(v) > 0$ で $\sum_{v \in V} g(v) = 1$ となるような確率分布である．このモデルは，出リンクのないような（$|F(u)| = 0$）ぶら下がり（dangling）Web ページ u から，確率 $g(v)$ でページ v へジャンプすることを意味する．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 PageRank スコアを要素としたベクトルを $\mathbf{x}_u = (y_1(u), y_2(u), \ldots, y_T(u))^T$ と定義する．このベクトル $\mathbf{x}_u$ をノード u の変化曲線と呼ぶ．この手法において，大域ジャンプベクトルは $\mathbf{z} = (1/|V|, \ldots, 1/|V|)^T$，パーソナライズベクトルは $\mathbf{g} = (1/|V|, \ldots, 1/|V|)^T$ である．

2.3　K-median クラスタリング

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 $\mathcal{V}$ 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 $u, v \in \mathcal{V}$ 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．一般的に，平均（mean）より中央値（median）の方が頑健であることが知られている．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．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 [8]．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．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．

$$f(P \cup \{w\}) = \sum_{v \in V} \max\{\mu(v\,; P), \rho(v,w)\}. \tag{5}$$

ここで，$\mu(v\,; P)$ 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，$\mu(v\,; P) = \max_{w \in P}\{\rho(v,w)\}$ 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) $k \leftarrow 1$，$\mathcal{P}_0 \leftarrow \emptyset$，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 (5) で $\hat{p}_k = \arg\max_{w \in V \setminus \mathcal{P}_{k-1}}\{f(\mathcal{P}_{k-1} \cup \{w\})\}$ を求め，$\mathcal{P}_k \leftarrow \mathcal{P}_{k-1} \cup \{\hat{p}_k\}$ とする；
(3) $k = K$ ならば $\hat{\mathcal{P}}_K = \{\hat{p}_1, \ldots, \hat{p}_K\}$ 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 $p_k \in P$ のクラスタ $\mathcal{C}_k$ に割り当てる．

3. 評価実験

有向機能コミュニティ抽出法において，1) 機能の類似するノード群を抽出できているか（有効性），2) モチーフによる手法では抽出できない機能を抽出できているか（有用性），3) 無向化したネットワークを対象とした場合では抽出できない機能を抽出できているか（有用性）を評価する．

3.1 ネットワークデータ

実験では，有向機能コミュニティ抽出法により抽出したコミュニティの特徴をとらえるために，2 つの人工ネットワークを採用する．

1 つ目の人工ネットワークは，ツリー型のネットワークであり，トップノードから双方向リンクで子ノードとつながっている．さらにそれらの子ノードは，出リンク，入リンク，双方向リンクによりそれぞれ子ノードを有するような構造をしている．本稿では Tree ネットワークと呼ぶ．

2 つ目の人工ネットワークは，階層構造を持つネットワークである．Ravasz らによって提案された階層性のあるネットワークモデル [9] により生成した．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，本稿では Hierarchical ネットワークと呼ぶ．また，有向ネットワークでの有効性を検証するために，各リンクは一定の規則に従い，双方向リンク，出リンク，入リンクとした．

現実のネットワークに対して有効な結果が得られるかを実証するために 2 つの Web サイトから構築したハイパーリンクネットワークを採用する．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する．本稿では，紙面の都合上選択した 2 つの大学 Web サイトのハイパーリンクネットワーク（Hosei ネットワーク，Yaku ネットワーク）に対する結果を示す[*1]．

[*1] 法政大学情報科学部 http://cis.k.hosei.ac.jp/，静岡県立大学薬学部 http://w3pharm.u-shizuoka-ken.ac.jp/.

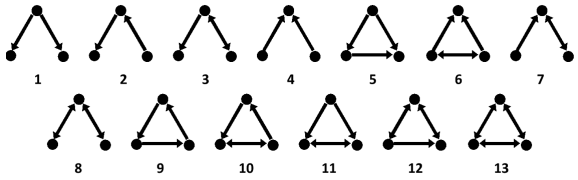


図 1　モチーフパターン
Fig. 1　Motif pattern.

3.2 モチーフパターンベクトルによる分類

有向機能コミュニティ抽出法の有用性を評価するために，ネットワークモチーフによる手法と比較する．

ネットワークモチーフは，有向ネットワーク内のリンクパターン（モチーフパターン）の出現頻度を数え上げることで，そのネットワークの特徴的なモチーフパターンを抽出する解析法である．本稿では，**図 1** に示す 13 種のモチーフパターンに対して，各ノードが各パターンにどの程度関与しているかを数え上げ，その値を要素とするベクトルを各ノードの特徴ベクトルとする．形式的には，ノード u がパターン i に $n_{u,i}$ 回関与していると，ノード u の特徴ベクトルは $\mathbf{m}_u = (n_{u,1}, \ldots n_{u,13})^T$ となる．

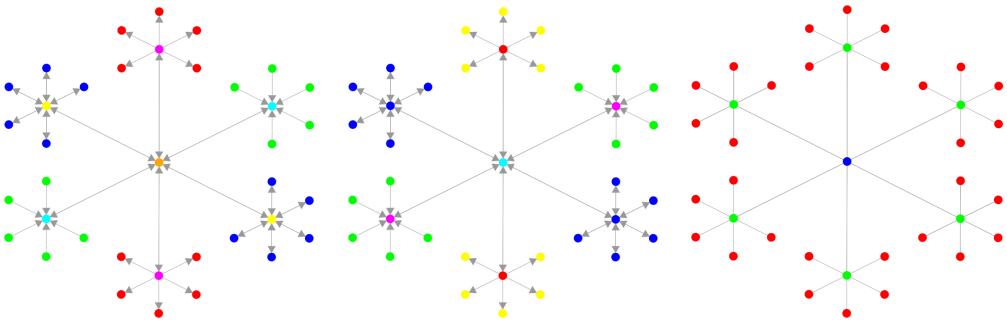
各ノード間の類似度およびクラスタリング法は，機能コミュニティ抽出法と同様に，コサイン類似度，K-median 法をそれぞれ適用する．本稿では，この手法をパターンベクトルによる分類と呼ぶ．

3.3 実験設定

4 つのネットワークに対し，反復終了ステップ数 $T = 500$，大域ジャンプ確率 $\alpha = 0.0001$ とし有向機能コミュニティを抽出する．**図 2**，**図 3**，**図 4**，**図 5** は，各ネットワークに対して，有向機能コミュニティ，パターンベクトルによる分類結果，および，無向化したネットワークに対して既存の機能コミュニティ抽出法により抽出した結果を示す．同じ色のノードは同一のコミュニティに属することを意味する．また，Hosei ネットワークと Yaku ネットワークのノード座標は，クロスエントロピー法により可視化した [10]．クロスエントロピー法は，ノード間の距離関係ではなく隣接関係によりノード座標を計算しており，可視化結果のリンクの長さに意味はないことに注意する．**図 6**，**図 7**，**図 8**，**図 9** は，K-median クラスタリングにより選定された代表ノードの変化曲線を表している．横軸はステップ数，縦軸は各ステップでの PageRank スコアを表している．変化曲線の色は図 2 (a) から図 5 (a) の可視化結果でのノードの色と対応している．凡例の上から順に選定されたノード順になっている．

3.4 Tree ネットワークの処理結果の評価

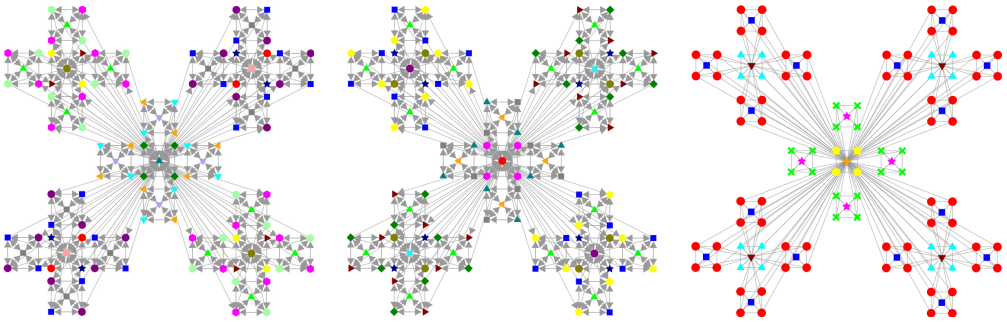
図 2 に Tree ネットワークのコミュニティ抽出結果を示す．有向機能コミュニティ抽出法の結果を見ると，各ノー



(a) 有向機能コミュニティ　(b) パターンベクトルによる分類　(c) 無向化機能コミュニティ

図 2　Tree ネットワーク

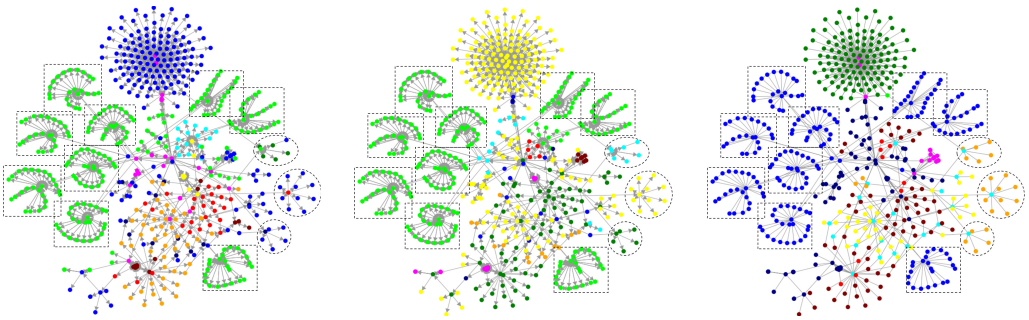
Fig. 2　Tree network.



(a) 有向機能コミュニティ　(b) パターンベクトルによる分類　(c) 無向化機能コミュニティ

図 3　Hierarchical ネットワーク

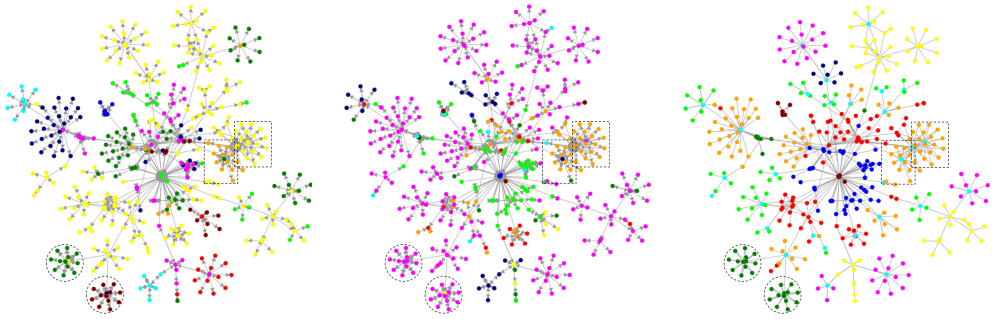
Fig. 3　Hierarchical network.



(a) 有向機能コミュニティ　(b) パターンベクトルによる分類　(c) 無向化機能コミュニティ

図 4　Hosei ネットワーク（$K = 10$）

Fig. 4　Hosei network ($K = 10$).



(a) 有向機能コミュニティ　(b) パターンベクトルによる分類　(c) 無向化機能コミュニティ

図 5　Yaku ネットワーク（$K = 10$）

Fig. 5　Yaku network ($K = 10$).

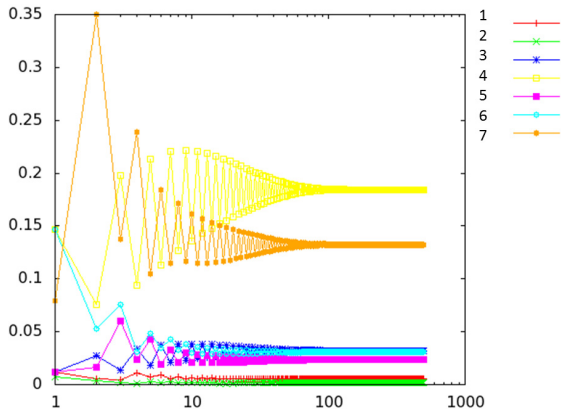


図 6　Tree ネットワーク 代表ノード変化曲線
Fig. 6　Tree network changes curves of representative nodes.

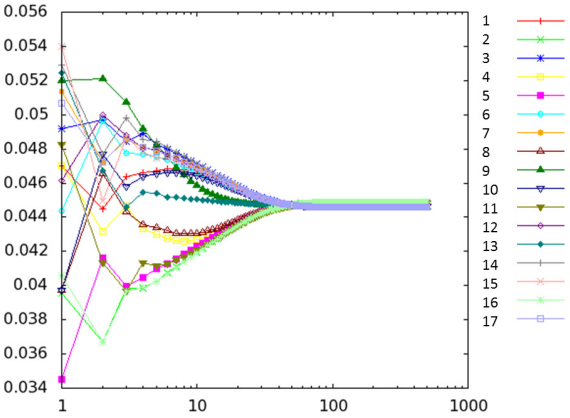


図 7　Hierarchical ネットワーク 代表ノード変化曲線
Fig. 7　Hierarchical network changes curves of representative nodes.

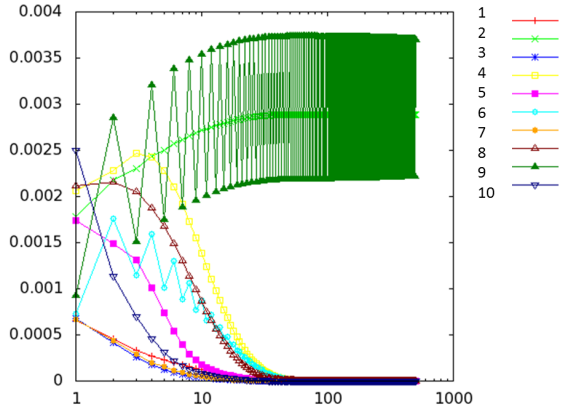


図 8　Hosei ネットワーク 代表ノード変化曲線
Fig. 8　Hosei network changes curves of representative nodes.

ドがその周辺ノードとのリンク関係により 7 つのコミュニティに分割されていることが分かる（図 2 (a)）[*2]．具体的には，最上部分と最下部分の中心ノード（ハブノード）とそれらの隣接ノード，左下部分と右上部分の中心ノード（権威ノード）とそれらの隣接ノード，左上部分と右下部分の双方向リンクによりつながれた中心部分と隣接ノードに

[*2]　$K = 7$ 以上には分割できないため，$K = 7$ の結果を示す．

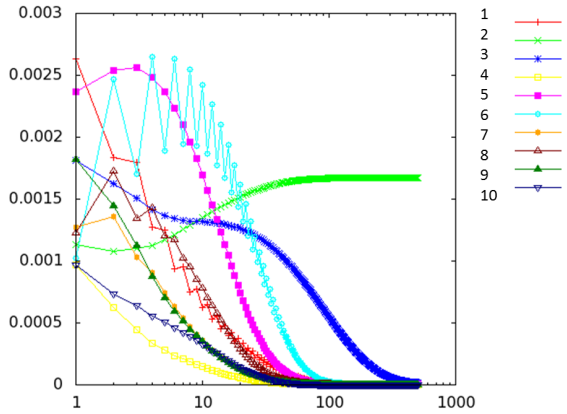


図 9　Yaku ネットワーク 代表ノード変化曲線
Fig. 9　Yaku network changes curves of representative nodes.

それぞれ分類されている．双方向リンクでつながっている部分は，無向二部グラフ的な構造を有している．二部グラフの隣接行列は，スペクトル円上に複数の固有値を持ち原始的でないため，極限値を持たず周期性が表れることが知られている [11]．無向二部グラフ的構造の中心ノードである図 6 の 4 つ目の代表ノードの変化曲線を見ると，二部グラフの特徴である周期性が表れており，他の曲線とうまく識別されている．このように，有向ネットワークに対しても，ノードの機能によって適切に分類できていることが分かる．

パターンベクトルによる分類の結果を見ると，有向機能コミュニティ同様に周辺ノードとのリンク傾向によりノードを分類できていることが分かる（図 2 (b)）[*3]．しかし，無向二部グラフ的構造の中心ノードが周辺ノードと同じクラスタに分類されてしまっている．これは，該当部分の中心ノードも周辺ノードもモチーフパターン 8 にしか出現しないことから，違いが判別できないからである．

一方，このネットワークを無向化したネットワークに対する既存の機能コミュニティ抽出法の結果を見ると，リンクが有する情報量が減り，有向機能コミュニティでは識別されていたハブノードと周辺ノード，権威ノードと周辺ノード，二部グラフ的構造の関係が同一視されてしまっている（図 2 (c)）[*4]．無向ネットワークに対する結果では，このような大域的な特徴が分かり，パターンベクトルによる分類の結果では，局所的なリンク構造の類似性が分かる．有向ネットワークに対する結果では，ノードの大域的な役割およびリンクの方向による局所的な特徴が分かり，両手法のハイブリッド的な結果が得られた．3 手法の処理結果は矛盾せず，相補的である．

3.5　Hierarchical ネットワークの処理結果の評価

Hierarchical ネットワークにおいて，各ノードは自身の

[*3]　$K = 6$ 以上には分割できないため，$K = 6$ の結果を示す．
[*4]　$K = 3$ 以上には分割できないため，$K = 3$ の結果を示す．

属するセグメント内のノードと双方向リンクでつながっているが，最上位ノードと，あるいは 1 階層上のノードとのリンク方向が異なるという特徴を有する．

有向機能コミュニティの結果を見ると，階層的地位の同じノードどうしは同じコミュニティに割り振られる傾向にあるが，隣接ノードとのリンク関係の違いによりさらに細分化されていることが分かる（図 3 (a)）[*5]．会社組織内では，上司への報告を主にする社員や上司からの連絡を受ける社員，セグメント内の社員と密に相談や話し合いをする社員など，それぞれ異なる役割を有している．その違いが図 3 (a) に表れていると考えられる．このように，有向ネットワークに対しても，ノードの機能によって適切に分類できていることが分かる．

パターンベクトルによる分類の結果を見ると，有向機能コミュニティ同様に周辺ノードとのリンク傾向によりノードを分類できていることが分かる（図 3 (b)）[*6]．しかし，モチーフパターンは 1 ステップあるいは 2 ステップ先のノードとのリンク構造しか考慮できないため，局所的なリンク構造の類似するノードしか分類できない．

一方，このネットワークを無向化したネットワークに対する，既存の機能コミュニティ抽出法の結果を見ると，階層的地位の高さによって分類されているが，同一階層内は無向化すると区別できないので同一視される（図 3 (c)）[*7]．このように，有向機能コミュニティでは最大で 17 コミュニティに分割できるのに対し，パターンベクトルによる分類では 14 コミュニティ，無向化すると 8 コミュニティしか抽出できない．有向機能コミュニティは両手法と比較して，明らかに有用な情報を含んだ役割の類似するノードを抽出できている．

3.6　Hosei ネットワークの処理結果の評価

Hosei ネットワークの特徴は，図 4 に点線で四角く囲っている部分のように，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．また，図 4 に点線で丸く囲っている部分のように，研究室一覧のページから各研究室のインデックスページ，コンテンツページなどが存在するという特徴もある．

有向機能コミュニティの結果を見ると，上述した各年度の成果報告ページが同一のコミュニティとして抽出されている（図 4 (a)）．各年度において成果報告ページ数は異なっているが，隣接ノードとのリンク関係が類似するため，同一のコミュニティとして抽出できていると考えられる．無向化したネットワークに対する，既存の機能コミュニティ抽出法の結果を見ても，同様に各年度の成果報告ページは，ノードの機能としては同質と考えられ，同一コミュニティとして抽出されている（図 4 (c)）．一方，点線で丸く囲っている研究室ごとのページに関していうと，有向機能コミュニティでは，一番上の該当コミュニティに注目すると，濃い緑色のコミュニティとして抽出されており，他の該当コミュニティとは識別されていることが分かる（図 4 (a)）．この研究室のページ群では，各ノードが双方向リンクでつながっており，上述した Tree ネットワーク同様二部グラフ的な構造になっている．図 8 の 9 つ目の代表ノードの変化曲線を見ても，二部グラフ特有の周期性が見て取れる．無向化した場合では，これらの違いが識別できず同一視されてしまっている（図 4 (c)）．

パターンベクトルによる分類の結果を見ると，有向機能コミュニティ同様に各年度の成果報告ページが同一のコミュニティとして抽出されている（図 4 (b)）．各年度において，隣り合う教員のページ間は双方向リンクが存在し，すべてのページは年度ごとのインデックスページとの間に双方向リンクが存在するため，モチーフパターン 13 が顕著に出現することが起因すると考えられる．一方，点線で丸く囲っている研究室ごとのページでは，Tree ネットワーク同様に無向二部グラフ的構造の中心ノードと周辺ノードの違いが見られない．

3.7　Yaku ネットワークの処理結果の評価

Yaku ネットワークの特徴も，図 5 に点線で囲っている部分のように，研究室一覧のページから各研究室のインデックスページ，コンテンツページなどが存在する点である．右上の点線で四角く囲っている部分は，入り口となるインデックスページが 2 つ存在しており，対象部分左側の緑色ノードは「高分子生物化学」という学問に関するインデックスページであり，対象の研究室内の学問的コンテンツ（タンパク質や酵素の説明）へのみリンクしている．対象部分右側の黄色ノードは研究室のトップページであり，上記のコンテンツだけでなくゼミ生紹介や研究内容紹介，教授の業績一覧などのゼミ的コンテンツにもリンクしている．

図 5 の Yaku ネットワークのコミュニティ抽出結果を比較すると，有向機能コミュニティの結果では，このようなリンク構造の違いを考慮し，学問的コンテンツとゼミ的コンテンツを異なるコミュニティとして抽出されているが，無向化した場合は同じ研究室のページ群は同一コミュニティとして抽出される．左下の点線で丸く囲っている部分は，対象部分の研究室の学会発表リストページであり，日本語版と英語版の 2 種類が存在している．日本語版と英語版ではハイパーリンクの構造が異なっており，有向ネットワークでは言語ごとに別の機能コミュニティとして区別されるが，無向化すると同一の機能コミュニティと見なされる．

*5　$K = 17$ 以上には分割できないため，$K = 17$ の結果を示す．
*6　$K = 14$ 以上には分割できないため，$K = 14$ の結果を示す．
*7　$K = 8$ 以上には分割できないため，$K = 8$ の結果を示す．

パターンベクトルによる分類の結果を見ると，局所的に類似したリンク構造を有する多くのノードが，同一のクラスタとして抽出されている（桃）．これは，Yaku ネットワークが全体的にツリーに近い構造をしており，「インデックスページとコンテンツページ」の関係が多く見られるからだと考えられる．中でも，双方向リンクを有するパターンを多く持つ（黄緑）と片方向リンクを有するパターン 4 を多く持つ（橙），パターン 1，2，3 を多く持つ（桃）に大別されている．パターンベクトルによる分類結果では，ネットワーク内の相対的な位置に関係なく，局所リンク構造の類似するノード群を抽出できている．しかし，上述した無向化した場合と同様に，有向ネットワークで区別できていたノードの機能が同一視されている点も見られる．

以上のように，有向ネットワークにおいても妥当な機能コミュニティが抽出できていることが分かる．さらに，無向化したネットワークを対象とした場合のようなノードの大域的な役割の類似性とパターンベクトルによる分類のような局所的なリンク構造の類似性の両方の特徴を持つハイブリッド的な結果が得られることが示唆された．無向化した場合に比べ，リンクの方向により局所的な特徴を考慮することができるため，より精緻にノードの機能を分類できることも示唆された．

3.8　変化曲線による大域ジャンプ確率の検討

有向機能コミュニティ抽出法のパラメータとして，大域ジャンプ確率 α，反復数 T などがあげられる．反復数は，PageRank スコアの収束する $\|\mathbf{y}_t - \mathbf{y}_{t-1}\|_{L1} < \varepsilon$ となるステップ数 $T = t$ とすることで，$T > t$ とした場合と同質のコミュニティ抽出結果が得られることが示されている．すなわち，収束後の変化曲線は実質的な効果はない．

この節では，導入した大域ジャンプ確率の設定範囲について考察する．PageRank スコアの収束速度は，推移確率行列の第 1 固有値と第 2 固有値の Eigen-gap および，大域ジャンプ確率 α の値に依存する．$\alpha \simeq 1$ とすると，PageRank スコアの収束が速くなるため，特徴ベクトルの実質的な次元数が減少する．さらに，大域ジャンプのランダム性の影響が支配的になるため，どの変化曲線も形状が類似し，適切な機能コミュニティを抽出することが困難になる．**図 10** に，Hosei ネットワーク，Yaku ネットワークに対して，横軸に大域ジャンプ確率，縦軸に全ノードペアの変化曲線間のコサイン類似度の標準偏差をプロットした．図 10 より，大域ジャンプ確率が大きいほど，コサイン類似度の散らばりが小さくなり，全ノードの変化曲線の形状が類似していることがうかがえる．一方，大域ジャンプ確率が小さいほど，コサイン類似度のバラつきが大きくなり，変化曲線が類似するノードと類似しないノードが存在し，変化曲線の形状により有向機能コミュニティが抽出可能であることが示唆される．ただし，コミュニティにうまく分割できるかどうかを標準偏差だけでは語れないことを注意しておく．これらのネットワークに関していうと，$\alpha = 0.0001$ で安定しているため，本稿では $\alpha = 0.0001$ での実験結果を掲載した．

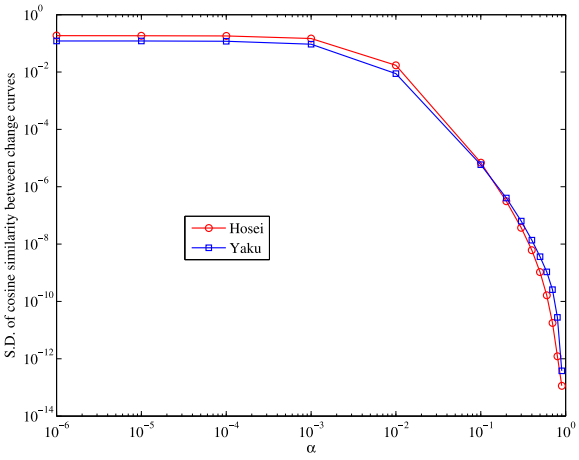


図 10　大域ジャンプ確率と変化曲線
Fig. 10　Global jump probability and changes curves.

次に，大域ジャンプ確率を変化させたときの目的関数値の変化を検証する．本稿で取り上げた人工ネットワークは，幾何学的かつ規則的な構造を有しているため，最適なコミュニティ数（コミュニティ分割の上限）が実験の範囲内で得られている．一般のネットワークに関していうと，各ノードがそれぞれわずかに異なる役割を有し，わずかにリンク構造が異なる場合，すべてのノードが異なる個別のコミュニティとして抽出される（分類の粒度が最小になる）可能性もある．また，これはクラスタリング問題において，適切なクラスタ数を決定するという難しい問題と等価である．機能コミュニティ抽出法においてコミュニティ数の決定に貢献するものとして，K-median 法の目的関数値がある．そこで，大域ジャンプ確率が変化した際の適切なコミュニティ数の代替として，各コミュニティ数に分割した際の目的関数値の値をプロットする（**図 11**，**図 12**）．横軸にコミュニティ数，縦軸に目的関数値，各マーカは異なる大域ジャンプ確率を表す．

図 11 と図 12 を見ると，上述した変化曲線の標準偏差に関する考察と同様に，コミュニティ数が少ない場合でも目的関数値が上限に達し，大域ジャンプ確率を大きくするにつれて得られるコミュニティ数が減少することが分かる．

PageRank をサーチエンジンの検索結果のランキングに用いる場合には，膨大な Web ページに対して上位のごくわずかの検索結果しか見られないこと，スパム的な Web ページを除外するための副次的な要素であることから，計算時間短縮を優先して値が選択されることが多い [3]．一方本手法では，変化曲線が各ノードの機能を表現し，大域ジャンプによる変化曲線の均質化を回避して，変化曲線の多様化が求められるため，両者の適切な大域ジャンプ確率

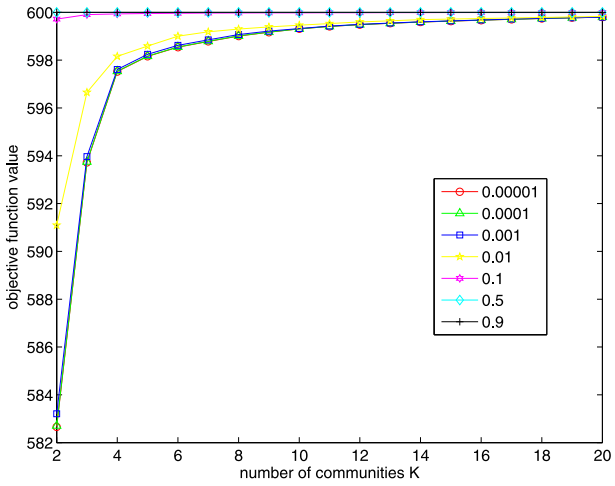


図 11　Hosei ネットワーク 目的関数値
Fig. 11　Hosei network objective function value.

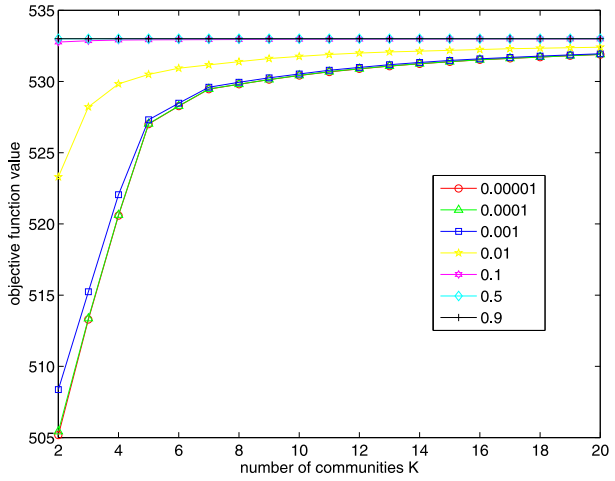


図 12　Yaku ネットワーク 目的関数値
Fig. 12　Yaku network objective function value.

の値は異なる．

4.　おわりに

本稿では，有向ネットワークを対象に，無向化した場合には分からないノード間関係の方向性や情報の流れなど，リンクの向きにより表出する機能に基づきノードを分類する手法を提案した．有向機能コミュニティ抽出法の有効性と有用性を検証するために，複数の人工ネットワークおよび Web ハイパーリンクネットワークを用いて評価した．可視化による定性的評価により，有向ネットワークに対しても類似の機能を有するノードを同一コミュニティとして抽出可能であり，本手法の有効性が示唆された．本手法は，無向化したネットワークを対象とした場合のようなノードの大域的な役割の類似性とパターンベクトルによる分類のような局所的なリンク構造の類似性の両方の特徴を持つハイブリッド的な結果が得られることが示唆された．無向化した場合に比べ，リンクの方向により局所的な特徴を考慮することができるため，より精緻にノードの機能を分類できることも示唆された．

また，有向リンクを有する有向ネットワークでは，出リンクを持たないぶら下がりノードや，複数の強連結成分の存在などにより，PageRank スコアの反復計算時に，ランクシンクなどの問題が発生する場合がある．そのため，$\alpha = 0$ では有向ネットワークに適用できない．本稿では，大域ジャンプ確率を $\alpha \neq 0$ とし，パラメータ α に対する考察をした．$\alpha \simeq 1$ とすると，PageRank スコアの収束が速くなるため，特徴ベクトルの実質的な次元数が減少する．得られるコミュニティ数 K も少なくなることが実験から明らかになった．さらに，どのノードの変化曲線も形状が類似し，適切な機能コミュニティを抽出することが困難になる．一方，$\alpha \simeq 0$ とすると，変化曲線間の類似度構造に散らばりが増え，変化曲線の形状により有向機能コミュニティが抽出可能であることが示唆された．

本手法は，多重ネットワーク，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．企業間取引ネットワークなどにおいては，階層だけでなく，取引量などの流量を重みとして加えることで，よりインフォーマティブな機能を抽出できることも期待できる．

今後は，さらに多様なネットワークで機能コミュニティ抽出法の有効性を検証するとともに，大域ジャンプ確率 α やクラスタ数 K などの設定方法を詳細に分析していくつもりである．また，どのようなリンク構造なら同一あるいは異なるコミュニティとして抽出できるかを詳細に検証し，機能コミュニティの必要十分性を評価していくつもりである．

謝辞　本研究は，NTT 未来ねっと研究所との共同研究，および，科研費（23500128）の支援を受けて行ったものである．

参考文献

[1] 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).
[2] Brin, S. and Page, L.: The Anatomy of a Large-scale Hypertextual Web Search Engine, *Computer Networks and ISDN Systems*, Vol.30, pp.107–117 (1998).
[3] Langville, A.N. and Meyer, C.D.: Deeper Inside Pagerank, *Internet Mathematics*, Vol.1, p.2004 (2004).
[4] 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).
[5] 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).
[6] 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.2, pp.1–10 (2012).
[7] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network Motifs: Simple Building Blocks of Complex Networks., *Science*,

Vol.298, No.5594, pp.824–827 (online), DOI: 10.1126/science.298.5594.824 (2002).
[8] Nemhauser, G.L., Wolsey, L.A. and Fisher, M.L.: An Analysis of Approximations for Maximizing Submodular Set Functions, *Mathematical Programming*, Vol.14, pp.265–294 (1978).
[9] 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).
[10] 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).
[11] Meyer, C.: Matrix Analysis and Applied Linear Algebra, *SIAM: Society for Industrial and Applied Mathematics* (2001).

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 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 慶應義塾大学理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．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．

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．