

主査：松浦 博 教授
副査：中野良平 教授
副査：池田哲夫 教授
指導教員：斉藤和巳 教授

ネットワークにおける機能的類似ノード群の抽出

学籍番号 1124507

伏見卓恭

平成25年 12月 13日

論文要旨

近年，Web上でのユーザ間のコミュニケーションやWebのハイパーリンク構造など，現実のあらゆる場面でネットワーク構造が見受けられるようになったのを発端として，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．ネットワークを構成するノードはみな一様ではなく，それぞれ固有の機能や役割などを有し，各ノードは他のノードと相互に影響を及ぼし合っている．

大規模ネットワークから，ある種の性質が類似するノード群，すなわち，コミュニティを抽出する研究は国内外で広く行われおり，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．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り，機能コミュニティと呼ぶノード群の抽出を課題とする．機能コミュニティとは，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，次数などの性質が類似する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．この手法は，各ノードに対して，1ステップで迎れる隣接ノードにメールをランダムに渡す試行を繰り返した際（ランダムウォーク），各ノードがメールを受信する期待値の時間変化を各ノードの機能ベクトルとして定義する．機能ベクトル間の類似度を計算し，類似度に基づきいくつかのノード群に分類する．

さらに，ランダムウォークに基づく基本手法をベースにリンクの向きを考慮した有向機能コミュニティや各ノードの視点に基づくパーソナライズ機能コミュニティ，情報拡散現象の下での機能に着目した情報拡散機能コミュニティなどの概念を提案し，それらを抽出する手法を提案する．

本論文の各章では，手法による結果の違いや有効性を評価するために，人工ネットワークおよび現実ネットワークを複数用いた評価実験を行う．提案する各手法や既存手法により得られる実験結果の違いについて比較評価する．各章における評価実験では，ネットワーク可視化結果にコミュニティごとに定めた色でノードを着色し，定性的に評価する．可視化結果から機能的に類似するノード群を抽出できていないこと（有効性），既存の手法では得られない結果が得られること（有用性）を示す．

Abstract

In recent years, various kinds of networks such as communication network between users on a web service, hyperlink structures of web pages and so forth, can be found around our lives and a number of research topics on complex networks have been studied. Each of nodes is not homogeneous and has intrinsic functions or roles in a network.

Extracting some groups, each of which consists of nodes with similar characteristics has been widely studied and various kinds of methods have been proposed what is called community extraction method. In this paper, we address the problem of extracting the groups of functionally similar nodes from a network. As functional properties of nodes, we focus on hierarchical levels, relative locations and/or roles with respect to the other nodes. For this problem, we propose a novel method for extracting functional communities from a given network. This method consists of two steps: the method first assigns a feature vector to each node, which is assumed to be some functional properties, by using calculation steps of PageRank scores for nodes from an initial score vector. Then, in a case that the supposed number of functional communities is K , the method divides all the node into K groups by using the K -medians clustering method based on the cosine similarity between a pair of the feature vectors.

Furthermore, we extend the previous described method to directional version, personalized version, information diffusion version and so forth, and each chapter in this paper we compare these extended methods to the basic method and some existing methods.

In our experiments using several types of synthetic and real networks, we evaluate the characteristics of functional communities extracted by our proposed method. From our experimental results, we confirmed that our method can extract functional communities, each of which consists of nodes with functionally similar properties.

凡例

数式，図表などは情報処理学会の規定に準拠する．

目次

論文要旨		i
Abstract		ii
凡例		iii
第1章	序論	1
1.1	研究背景	1
1.2	研究目的	2
1.3	論文構成	3
第2章	機能コミュニティ	5
2.1	同値性	5
2.1.1	構造同値	5
2.1.2	正則同値	6
2.2	機能コミュニティ	7
2.3	関連研究	7
第3章	機能コミュニティ抽出法：FCE法	10
3.1	はじめに	10
3.2	リンク密度に基づくコミュニティ抽出	11
3.3	ノード間類似度に基づくコミュニティ抽出	11
3.3.1	Vertex Similarity法	12
3.3.2	SimRank法	13
3.4	機能コミュニティ抽出法：FCE法	13
3.4.1	PageRankと機能ベクトル	13
3.4.2	ノード間の類似度計算	14

3.4.3	<i>K</i> -medians法によるクラスタリング	15
3.4.4	時間計算量	16
3.5	評価実験	16
3.5.1	ネットワークデータ	16
3.5.2	可視化による定性的評価	20
3.5.3	時間計算量による定量的評価	26
3.6	考察	27
3.7	おわりに	27
第4章	有向ネットワークに対する機能コミュニティ抽出法：DFCE法	29
4.1	はじめに	29
4.2	有向機能コミュニティ抽出法：DFCE法	30
4.2.1	FCE法における機能ベクトル（参考）	30
4.2.2	DFCE法における機能ベクトル	31
4.3	評価実験	32
4.3.1	ネットワークデータ	32
4.3.2	モチーフパターンベクトルによる分類	33
4.3.3	実験設定	33
4.3.4	Treeネットワークの処理結果の評価	38
4.3.5	Hierarchicalネットワークの処理結果の評価	39
4.3.6	Hoseiネットワークの処理結果の評価	40
4.3.7	Yakuネットワークの処理結果の評価	41
4.3.8	変化曲線による大域ジャンプ確率の検討	42
4.4	次数不変リンク張替に対する頑健性	44
4.4.1	リンク張替法	44
4.4.2	評価結果	46
4.5	おわりに	47
第5章	個々のノードの視点に基づく機能コミュニティ抽出法：PFCE法	51
5.1	はじめに	51
5.2	FCE法	52
5.2.1	機能ベクトル構築法	53
5.3	PFCE法への拡張	53

目次	vi
5.4 評価実験	54
5.4.1 ネットワークデータ	55
5.4.2 実験結果と考察	56
5.4.3 変化曲線によるエゴジャンプ確率に関する考察	59
5.5 おわりに	60
第6章 情報拡散モデルに基づく機能コミュニティ抽出法：IDFCE法	68
6.1 はじめに	68
6.2 情報拡散機能コミュニティ抽出法：IDFCE法	69
6.2.1 情報拡散モデル	69
6.2.2 ボンドパーコレーション	70
6.2.3 情報受信期待値変化ベクトルの計算	71
6.3 評価実験	72
6.3.1 ネットワークデータ	72
6.3.2 実験設定	72
6.3.3 実験結果	73
6.3.4 コミュニティ抽出結果の類似性分析	77
6.3.5 機能ベクトルに関する考察	77
6.4 おわりに	78
第7章 時間減衰重みを導入した機能コミュニティ抽出法：TDFCE法	79
7.1 はじめに	79
7.2 時間減衰重み付き機能コミュニティ抽出法：TDFCE法	80
7.2.1 機能ベクトル構築法	80
7.2.2 時間減衰関数	81
7.3 評価実験	81
7.3.1 人工ネットワークに対する処理結果	82
7.3.2 実ネットワークに対する処理結果	83
7.4 おわりに	83
第8章 結論	86
謝辞	90
参考文献	90

本論文に関する原著論文

94

第1章

序論

本章では、本論文の研究背景および研究目的について言及する。

1.1 研究背景

近年様々な分野において、大規模・複雑な事象をネットワークとしてとらえ、ノード間の相互関係やネットワーク構造、ネットワーク上での現象を分析する研究が盛んに行われている。現実の事象の中にはさまざまな つながり が存在するためネットワークも多種多様である。友人関係や知人関係、相談関係、ブログサイト、Twitter、商品のレビューサイトなどにおけるユーザ間の関係などの社会的なつながり、道路網、電力網、インターネット網、空路などのインフラ、Webページ群のハイパーリンクや自然界の食物連鎖や遺伝子、タンパク質、神経細胞のつながりなどがあげられる。上述したつながりの例は、それぞれを構成する個体・オブジェクトをノードとし、ノード間のつながりをリンクとしたネットワークとして扱うことができる。

現実ネットワークの多くはスケールフリー性 [1, 2] やスモールワールド性 [3, 4] といった特徴的な構造を持つことが明らかとなり、近年ますます注目を集めている。これらの研究では、現実ネットワークの有する特徴や構造を模写した人工ネットワークのモデルが提案されている [5, 6, 7]。さらに、ネットワークを評価するための指標も多く提案されている。社会ネットワーク分析の分野では、中心性指標という、ネットワーク内のノードの性質をネットワーク構造から導出・定量化し、その値でランキングすることで重要ノードを抽出する手法が広く用いられている [8, 9, 10, 11, 12, 13]。また、Webのハイパーリンク構造から重要ページを検索するためのランキング手法として著名なPageRank アルゴリズム [14, 15] や書誌の引用関係から重要書籍を検索するためのHITSアルゴリズム [16]なども提案されている。

このように現実ネットワークにおいて、各ノードは均質ではなく、各ノードは固有の性質や立場、役割、機能を有しており、これらに基づき、多大なノード群をクラスタリングしたり、重要ノードを抽出するための手法が提案されている。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 [17]．

既存のコミュニティ抽出手法として，Newmanらによる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 [18, 19, 20, 21, 22]．さらに，スペクトラルグラフ分析の手法であるNormalized Cut法 [23]やRatio Cut法 [24]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，ネットワーク上でのノード同士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)やクリークの条件を緩めたサブネットワークをみつけるための様々な手法が提案されている [25, 26, 27, 28]．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いてリンク構造の粗密に着目し，全ノード集合をいくつかの部分集合に分割することに主眼を置いている．

また，ネットワークにおいて構造上類似した立場にあるノードの概念として同値性がある．同値なノード群を同定する代表手法であるREGE，CATREGEアルゴリズムは，計算量の点で大規模なネットワークには対応できない [29]．同値性は社会学において古典的な考え方であるため，文献 [29]は20年前の研究であるが，近年では同値性の概念を発展させ，近似的に同定する手法として Leichtら [30]やJehら [31]の手法がある．Leichtらは，“類似ノードの周囲のリンク関係は類似する”という仮定のもとで，ノード間の類似度を定義している．Jehらは“共通の隣接ノードを有するノードは互いに類似している”という直観に基づき，ノード間の類似度を定義している．隣接行列やラプラシアン行列によるカーネルを用いて，ノード間の類似度を計算する手法も提案されている [32, 33]．ソーシャル・ネットワーク上での情報拡散において，類似した立場や役割のノード，すなわち正則同値なノード同士は，同様の情報を保持する可能性があるという知見 [34]もある．ノードの同値性により，いくつかのグループにクラスタリングすることは，ネットワーク上でのノードの動向を調査する点で重要と考えられている [35]．

1.2 研究目的

本論文では，ネットワークにおける各ノードの機能・役割に着目し，類似の機能を有するノード群を抽出することを試みる．ある種の類似する性質を有するノード群をコミュニティと呼ぶことにならば，機能的に類似するノード群を機能コミュニティと呼ぶ．

機能コミュニティを抽出するために，ネットワーク内の各ノードの機能を表現した機能ベクトルを構築し，機能ベクトル間の類似度に基づいてノードを分類する．ノードの機能とは，ネットワーク内での役割や立場・地位，他のノードとの位置関係などを意味する．会社内のコミュニケーションネットワークの例では，部長や課長，一般社員などの階層的な地位や外部の情報に精通し内部に伝達するゲートキーパーなどの非公式な役割などを意味する．これは前述した同値性を拡張した概念ということもできる．

機能コミュニティを抽出することで、社内組織ネットワークにおける効果的な人材配置、コミュニケーションネットワークにおけるグループ内のコミュニケーションの効率化、共著関係ネットワークにおける各著者の役割・貢献度推定、道路ネットワークにおける地域特性を考慮した道路敷設などへの応用が期待できる。

1.3 論文構成

本論文は以下のような構成である。第1章の序論に続いて、第2章で本論文で最も重要な概念である機能コミュニティについて、関連概念と比較しながら定義、整理する。

第3章で基本となる機能コミュニティ抽出法（FCE法）に関して、提案法の内容および主要技術について述べ、関連手法と比較して評価する。FCE法はランダムウォークにおける各ノードへの到達確率の時間変化をベースにした手法であり、後述する拡張手法のベースになっている。同値性を近似的に同定する手法や従来の密結合するノード群を抽出する手法と比較し、FCE法の有用性などについて述べる。

第4章で機能コミュニティ抽出法を有向ネットワークへ適用するために拡張した方法（DFCE法）について述べ、FCE法や関連手法と比較して評価する。コミュニケーションやハイパーリンクなど現実ネットワークの多くは非対称であり、リンクには向きが付されている。有向ネットワークから機能コミュニティを抽出するDFCE法と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。

第5章で個々のノードの視点に基づいた機能によりノードを分類する方法（PFCE法）について述べ、FCE法と比較して評価する。FCE法により抽出される機能コミュニティは、ネットワーク全体を客観視した大域的な機能が類似するノードからなるコミュニティである。一方、ノードの役割・機能は視点を変えることによって変化すると想定される。そこで、個々のノード（エゴノード）から主観視した機能に基づくパーソナライズ機能コミュニティという概念及びその抽出手法を提案する。

第6章で情報拡散の拡散する期待値を機能ベクトルとして定義する手法（IDFCE法）について述べ、採用する情報拡散モデルの違いについて評価する。FCE法ではランダムウォーク（PageRank）モデルを仮定していたが、より直接的にコミュニケーションにおける機能を表現するために、機能ベクトル計算に情報拡散モデルを採用する。

第7章で機能ベクトル時間減衰重みを導入し、周辺ノードとの関係を強調させた手法（TDFCE法）について述べ、FCE法と比較して評価する。FCE法は、PageRankが収束するまでのスコア変化を要素とするベクトルを用いて、各ノードが有する機能ごとに分類する手法である。クラスタ係数の大きいサブネットワークでは、隣接するノードペアの次数の差が小さく、比較的早期にスコア変化が平坦化され、高次元特徴ベクトルから個々のノードの特徴を抽出するのが困難となる場合がある。また、PageRankのスコア変化は、早期では局所的な特徴を、末期では大域的な特徴を示す。この点を考慮し、PageRank

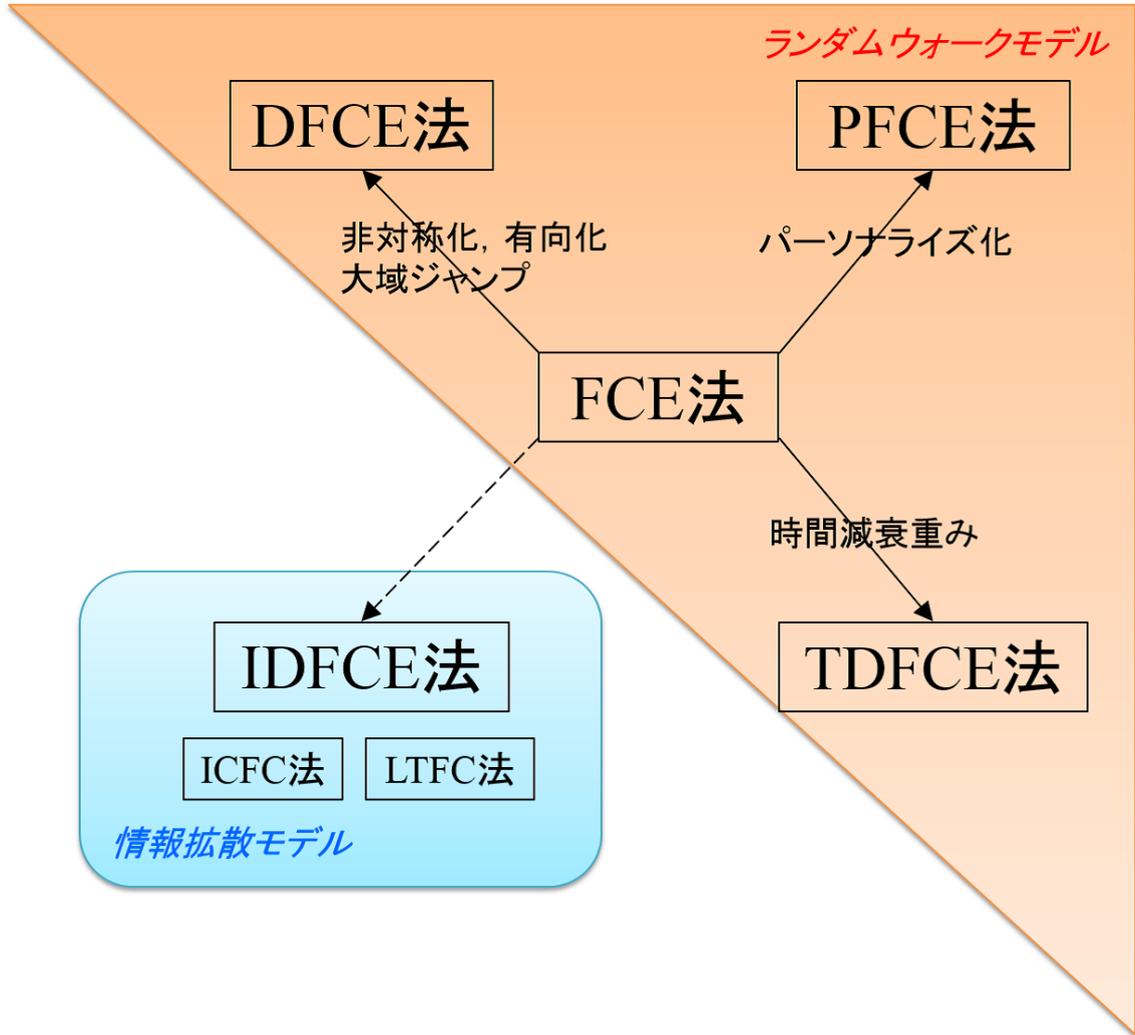


図1.1 機能コミュニティ抽出法 (FCE法) の全体像

スコアが局所的構造の影響をより強く受けている初期段階により着目できるように、高次元機能ベクトルに時間減衰関数による重みを導入する。

最後に本論文のまとめと今後の展望について述べる。

第2章

機能コミュニティ

この章では、本論文で重要となる同値性の概念および機能コミュニティについて説明する。

2.1 同値性

ネットワークにおける同値性とは、隣接ノードとの局所的なリンク傾向の類似性であり、定義により構造同値 (structural equivalence) や自己同型同値 (automorphic equivalence), 正則同値 (regular equivalence) に分類される (図 2.1)。中でも、広く知られている構造同値と正則同値について、ノード集合 V とリンク集合 E からなる有向ネットワーク $G(V, E)$ を用いて説明する。



図2.1 同値性概念の全体像

2.1.1 構造同値

ノード u と v が構造同値であるとは、 u, v が隣接する全てのノードとの関係が同一であることをいう [36, 37]。すなわち構造同値なノードを入れ替えても、ネットワーク構造は全く変わらない。形式的には、ノード u, v の出隣接ノード集合をそれぞれ $F(u), F(v)$ 、入隣接ノード集合をそれぞれ $B(u), B(v)$ としたとき、 $F(u) = F(v)$ かつ $B(u) = B(v)$ ならばノード u と v は構造同値である。

図 2.2に例を図示する．同一色のノードどうしは構造同値なノードである．構造同値を求めるアルゴリズムにはSTRUCTUREアルゴリズムやCONCORアルゴリズムがある．

2.1.2 正則同値

正則同値は構造同値の制約を緩和した，すなわち，一般化した概念である [35, 37]．ノード u と v が正則同値であるとは， u, v が正則同値な任意のノードペア x, y に対する接続関係が同一である場合である．形式的には，各ノード u を整数値へマッピングした像を $I(u)$ とし，ノード u の出隣接ノードの像の集合を $\mathcal{I}^{out}(u) = \{I(v)|v \in F(u)\}$ ，入隣接ノードの像の集合を $\mathcal{I}^{in}(u) = \{I(v)|v \in B(u)\}$ とする． $\mathcal{I}^{out}(u) = \mathcal{I}^{out}(v)$ かつ $\mathcal{I}^{in}(u) = \mathcal{I}^{in}(v)$ ならば ノード u と v は正則同値である．写像 f の像が等しいノードどうしは同値なノードである．

構造同値と異なり，リンク相手同士が同値なノードであれば，必ずしも相手ノードが同一ノードである必要はない．例えば，ある会社組織において，部長らは上司として社長と，部下として何人かの課長と一般社員とリンクしていると仮定する．どの部長も，1人の社長と何人かの異なる部下とリンクしているが，結合パターンは社長ノード，課長ノード，一般社員ノードとリンクしている点において，部長という正則同値なノード群が同定される．図 2.3に例を図示する．同一色のノードどうしは正則同値なノードである．正則同値を求めるアルゴリズムにはREGEアルゴリズムやCATREGEアルゴリズムなどがある．

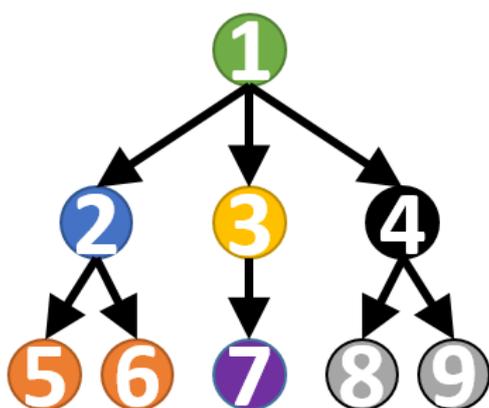


図2.2 構造同値の例

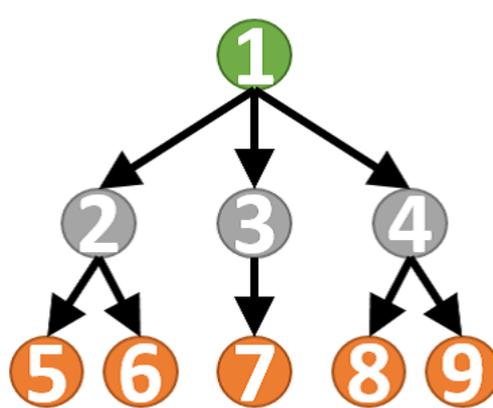


図2.3 正則同値の例

2.2 機能コミュニティ

本論文で提案・定義する概念である機能コミュニティとは、ネットワーク内での各ノードの機能が類似するノード群により構成されるコミュニティである。ネットワーク内での相対的位置や階層的地位、周辺ノードとの関係のパターンや次数などが類似すれば、ノードが提供する機能が類似するという考えに基づいた概念である。同値性のように局所的なリンク傾向に主眼を置くのではなく、ネットワーク全体に対する個々のノードの役割や立ち位置に着目したものである。例えば、社長、部長、課長、一般社員のような局所的なリンク構造から得られる形式的に定められた立場ではなく、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を同定することを目的としている（図 2.4）。

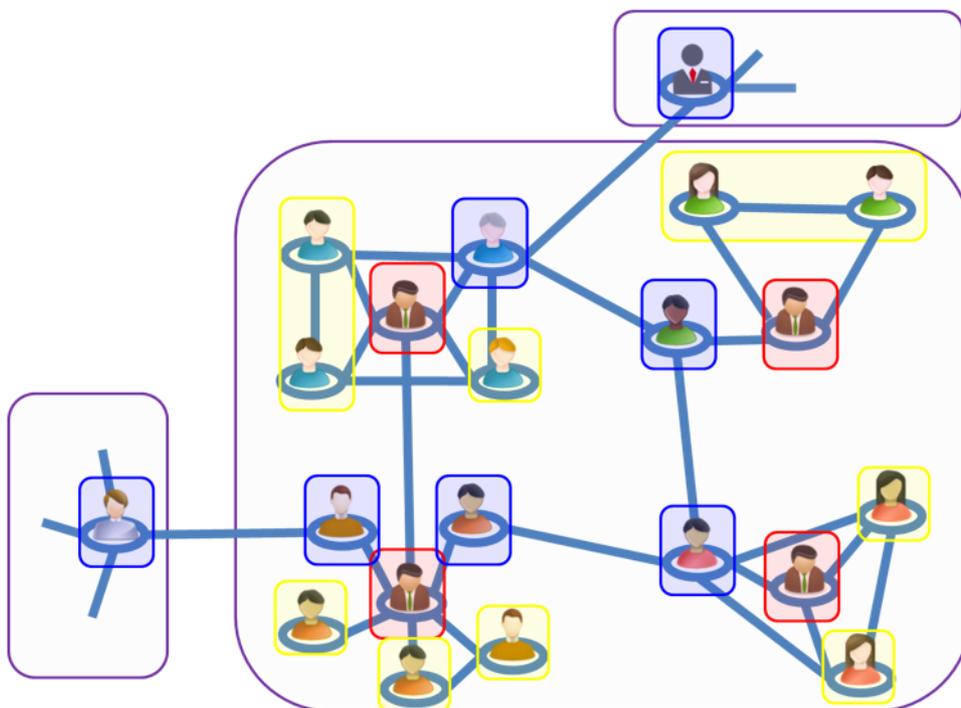


図2.4 機能コミュニティの例

2.3 関連研究

ネットワークにおけるノードの機能や役割に着目しノードどうしの類似性を判定する手法が多く研究されている。上述した構造同値なノードを同定する手法として、STRUCTUREアルゴリズム [38]やCONCORアルゴリズム [29]がある。同様に、正則同値なノードを同定する手法として、REGEアルゴ

リズムやCATREGEアルゴリズムがある [29]。これらの手法は組み合わせ最適化の手法を用いて厳密に正則同値性を計算するため、ノード数の5乗オーダー (REGE) や3乗オーダー (CATREGE) がかかる。そのため、これらの文献でも小さいモデルネットワークによる評価しかなされていない。特に、現実ネットワークの多くは幾何学的なグラフと比較して複雑なリンク構造を有しているため、同値なノード群を抽出する手法では、1つも抽出できない場合がある。

近年では同値性の概念を発展させ、近似的に同定する手法として Leichtら [30]やJehら [31]の手法がある。Leichtらは、“類似ノードの周囲のリンク関係は類似する” という仮定のもとで、ノード間の類似度を定義している。Jehらは“共通の隣接ノードを有するノードは互いに類似している” という直観に基づき、ノード間の類似度を定義している。隣接行列やラブラシアン行列によるカーネルを用いて、ノード間の類似度を計算する手法も提案されている [32, 33]。いずれの手法も、ノード数分の類似度行列を反復し計算するため、REGEアルゴリズムほどではないものの、大規模ネットワークへの適用が困難な場合がある。具体的な比較は次章以降を参照されたい。

社会ネットワーク分析の分野では、中心性指標という、ネットワーク内のノードの性質をネットワーク構造から導出・定量化する指標がある [8]。これらの指標により定量化したノードの性質をベクトル化しクラスタリングする手法も提案されている。

権威ノードとハブノードを抽出するHITSアルゴリズムにおける隣接行列の反復計算 (べき乗法) をベースに、2つの有向ネットワークの各ノード間の類似度を計算する手法も提案されており、辞書における単語をノードとし、共起単語をリンクでつないだグラフにおいて同意語 (シノニム) を抽出する応用例を示している [39]。同意語は単語ネットワークにおいて同様の役割を果たすという仮定に基づいた評価実験である。

ネットワーク内で機能が既知のノードがある状況において、機能が未知であるノードの機能を予測する手法も提案されている [40]。あらかじめ機能が既知のノード間の類似度を、有する機能集合のJaccard係数により定義し、未知のノードはネットワーク構造上最も類似するノード (機能が既知) の機能を割り当てることで予測する手法である。「各ノードからの他のノードの見え方が似ているノードどうしは類似している」という仮定のもとに、ネットワーク構造上での類似度を再帰的に計算している。類似度行列を反復させることで求めるこの手法も、大規模ネットワークへの適用が困難な場合がある。

リンク密度に基づくコミュニティの概念とコミュニティ内でのノードの役割という概念をベースに、ある1つのコミュニティへの依存度を表すCommunity Metricと、他のノードとの関係の多さを表す次数という2つの尺度から、全ノードを4つのカテゴリ (Big Fish, Ambassadors, Loners, Bridges) に分類する手法も提案されている [41]。Big Fishとは、次数は高いがコミュニティ内へのリンクが多い傾向にあるノード群であり、いわゆる、ある集団の中の大物的なノードである。Ambassadorsとは、次数は高いが1つのコミュニティに固執せず複数のコミュニティへリンクを多く持つノード群であり、いわゆ

る，大使的なノードである．Lonersとは，次数が低く他のコミュニティのノードともあまりリンクを持たないノード群であり，いわゆる，ひとりぼっち的なノードである．Bridgesとは，次数は低いが1つのコミュニティに固執せず複数のコミュニティとコネクションがあるノード群であり，いわゆる，弱い紐帯である．このように，コミュニティ内でのノードの役割によりノードを分類している．

次章以降で説明する提案法，機能コミュニティ抽出法は各ノードに対して，各ノードの機能を表現した機能ベクトルを定義する．そしてベクトル間の類似度によりノード間の類似度を計算しクラスタリングする手法である．この点が関連研究との大きな違いの一つであり，陽に機能ベクトルを定義するため，機能ベクトルを修正することで，派生した機能コミュニティを抽出する手法への拡張が容易である．また，中心性のように1つの観点による分類ではなく，多次元的な分類を実現する点で異なる．

第3章

機能コミュニティ抽出法：FCE法

3.1 はじめに

現実社会における人と人とのつながりや、Webサービスにおけるユーザ間のつながりなどのソーシャル・ネットワークに対する関心が高まっている。人間関係に留まらず、Webページのハイパーリンクネットワークや道路網など、あらゆるところで複雑ネットワークが見受けられるようになり、それに伴い多くの研究がなされている [42, 43]。

現実ネットワークにおいて、全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、機能を有しており、これらに基づき、多大なノード群をクラスタリングしたり、重要ノードを抽出するための手法が提案されている。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 [17]。

既存のコミュニティ抽出手法として、Newmanらによる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 [19]。さらに、スペクトラルグラフ分析の手法であるNormalized Cut法 [23]やRatio Cut法 [24]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、ネットワーク上でのノード同士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) やクリークの条件を緩めたサブネットワークを見つけるための様々な手法が提案されている [25, 26, 27, 28]。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いてリンク構造の粗密に着目し、全ノード集合をいくつかの部分集合に分割することに主眼を置いている。

また、ネットワークにおいて構造上類似した立場にあるノードの概念として同値性がある。同値なノード群を同定する代表手法であるREGE, CATREGEアルゴリズムは、計算量の点で大規模なネットワークには対応できない [29]。同値性は社会学において古典的な考え方であるため、文献 [29]は20年前の研究であるが、近年では同値性の概念を発展させ、近似的に同定する手法として Leichtら [30]やJeh

ら [31]の手法がある．Leichtらは，“類似ノードの周囲のリンク関係は類似する”という仮定のもとで，ノード間の類似度を定義している．Jehらは“共通の隣接ノードを有するノードは互いに類似している”という直観に基づき，ノード間の類似度を定義している．隣接行列やラプラシアン行列によるカーネルを用いて，ノード間の類似度を計算する手法も提案されている [32, 33]．ソーシャル・ネットワーク上での情報拡散において，類似した立場や役割のノード，すなわち正則同値なノード同士は，同様の情報を保持する可能性があるという知見 [34]もある．ノードの同値性により，いくつかのグループにクラスタリングすることは，ネットワーク上でのノードの動向を調査する点で重要と考えられている [35]．

この章では，“機能・役割が類似するノードであれば，PageRankスコアの収束パターンも類似する”と考え，機能コミュニティと呼ぶ類似した機能を持つノード集合を抽出する方法（以下，FCE法）を提案する．各ノード間に類似度を定義し，類似度に基づいてノードをクラスタリングし，近似的に正則同値なノード集合を抽出する手法および我々の機能コミュニティを抽出する手法について比較，考察する．具体的には，前述したLeichtらとJehらの手法および我々の手法により，ノード間類似度行列を計算する．大規模ネットワークへの適用を視野に入れ，計算量により定量的に，可視化により定性的に評価する．

3.2 リンク密度に基づくコミュニティ抽出

リンク密度に基づくコミュニティ抽出法の代表例である，Newmanクラスタリング法（以下 Newman法）について簡単に触れる．Newman法では，コミュニティ抽出の度合を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Modularity Q は，コミュニティ i と j 間のリンク数の総リンク数に対する割合 e_{ij} を要素とする $K \times K$ の対称行列 \mathbf{E} を定義し， $Q = \sum_{i=1}^K (e_{ii} - a_i^2) = \text{Tr}(\mathbf{E}) - \|\mathbf{E}^2\|$ 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．コミュニティの具体的な抽出法は，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，Modularityが最も増加するノード同士を結合するステップを繰り返す．Modularityが最も高くなるステップ数でコミュニティを出力する．

3.3 ノード間類似度に基づくコミュニティ抽出

この節では，既存のノード間類似度の各種計算手法について述べる．正則同値性は，周辺リンク構造の一致を調べるものであるが，現実ネットワークの構造のばらつきに適應できるように，構造的なノード間類似度として扱うことが多い．以下で述べる類似度行列に基づきクラスタリングし，コミュニティを抽出する．

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. 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. $|\Gamma(u)|$ をノード u の次数という. 自己リンク付きノード u は $u \in \Gamma(u)$ である.

3.3.1 Vertex Similarity法

LeichtらのVertex Similarity法(以下 Versim法)は, 正則同値性の概念を拡張し, “類似ノードの周囲のリンク関係は類似する” という仮定の下で類似度行列を計算している. Versim法による類似度行列の計算法を以下に示す.

ノード u, v 間の類似度 $r(u, v)$ を再帰的に計算する:

$$r(u, v) = \phi \sum_{w \in V} a(u, w) \cdot r(w, v) + \delta(u, v). \quad (3.1)$$

ここで, $0 < \phi < 1$ は減衰係数で, $\delta(u, v)$ はクロネッカーのデルタである. 式 3.1を行列表記し整理すると,

$$\begin{aligned} \mathbf{R} &= \phi \mathbf{A} \mathbf{R} + \mathbf{I} \\ &= [\mathbf{I} - \phi \mathbf{A}]^{-1} \\ &\simeq \mathbf{I} + \phi \mathbf{A} + \phi^2 \mathbf{A}^2 + \dots \end{aligned} \quad (3.2)$$

となる. ここで \mathbf{I} は単位行列を表す. 隣接行列 \mathbf{A} の l 乗の u, v 要素 $a^l(u, v)$ は, ノード u からノード v への距離 l のパス数を表す. すなわち, ノード u からノード v へのパス数が多ければ多いほど, ノード u とノード v は類似度が高くなる. また減衰係数により, 距離が短いパスに大きな重みが付くことになる. 各項に対して, パス数の期待値で正規化し,

$$r(u, v) = \sum_{l=0}^{\infty} C_l^{uv} a^l(u, v) \quad (3.3)$$

とする. ここで, $C_l^{uv} = \frac{2|E|}{|\Gamma(u)| \cdot |\Gamma(v)|} \lambda^{-l+1}$ であり, λ は隣接行列 \mathbf{A} の最大固有値である. 各ノードの次数を対角要素にもつ行列 \mathbf{D} を用いて, 式 3.3を行列表記し, 式 3.2にならぬ整理すると,

$$\begin{aligned} \mathbf{R} &= 2\lambda|E|\mathbf{D}^{-1} \left(\mathbf{I} - \frac{\alpha}{\lambda} \mathbf{A} \right) \mathbf{D}^{-1} \\ \mathbf{DRD} &= \frac{\alpha}{\lambda} \mathbf{A}(\mathbf{DRD}) + \mathbf{I} \end{aligned} \quad (3.4)$$

ここで, $0 < \alpha < 1$ は減衰係数で, 初期値 $(\mathbf{DRD})_0 = \mathbf{I}$ とし, 式 3.4を $\|(\mathbf{DRD})_s - (\mathbf{DRD})_{s-1}\| < \varepsilon$ となるか, 所定の回数まで繰り返し計算することで類似度行列 \mathbf{R} を得る.

Versim法の主たる時間計算量は, 行列 \mathbf{DRD} の収束までの反復回数 S とし, 各ノードペアに対して, 隣接するノードの類似度を足し合わせるため, $O(S \times |V|^2 \times \bar{d})$ である. ここで, \bar{d} は平均次数を表す.

3.3.2 SimRank法

JehらのSimRank法（以下 Simrank法）は，“共通の隣接ノードを有するノードは互いに類似している”という仮定の下で類似度行列を計算している．Simrank法による類似度行列の計算法を以下に示す．

ノード u, v 間の類似度 $r(u, v)$ を再帰的に計算する：

$$r(u, v) = \frac{\phi}{|\Gamma(u)| \cdot |\Gamma(v)|} \sum_{i \in \Gamma(u)} \sum_{j \in \Gamma(v)} r(i, j). \quad (3.5)$$

ここで ϕ は減衰係数であり， $u = v$ の場合は $r(u, v) = 1$ とする．初期値 $\mathbf{R}_0 = \mathbf{I}$ とし，式 3.5 を $\|\mathbf{R}_s - \mathbf{R}_{s-1}\| < \varepsilon$ となるか，所定の回数まで繰り返し計算することで類似度行列 \mathbf{R} を得る．

Simrank法の主たる時間計算量は，行列 \mathbf{R} の収束までの反復回数 S とし，各ノードペアに対して，その隣接ノード数の積のペアの類似度を足し合わせるため， $O(S \times |V|^2 \times \bar{d}^2)$ である．Versim法より平均次数 \bar{d} 倍計算量がかかることがわかる．

3.4 機能コミュニティ抽出法：FCE法

FCE法は，対象とするネットワークとコミュニティ数 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. 各ノードの機能ベクトルとしてPageRankスコア変化曲線 x_v を構築;
3. 各ノードペアの機能ベクトル x_u と x_v のコサイン類似度 $\rho(u, v)$ を計算;
4. K -medians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;

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手法であり，階層的地位や相対的位置，周辺ノードとの隣接関係など，ネットワーク構造から窺い知ることのできるノードの機能により，ノードを分類する．各ステップの詳細な説明は参考文献 [44, 45]を参照されたい．

3.4.1 PageRankと機能ベクトル

FCE法は，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索する方法で，大域ジャンプを除いたPageRankの反復計算時のスコアの収束曲線を機能ベクトルとし，ベクトル間のコサイン類似度により類似度行列を定義する [44]．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．同様にランダムウォークも，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．

大域ジャンプなしのPageRankを用いた機能ベクトル構築法を以下に示す．行推移確率行列 \mathbf{P} は，各要素を $p(u, v) = a(u, v)/|\Gamma(u)|$ とする．各ノードのランダムウォークにおける到達期待値ベクトル \mathbf{y} は， $y(v) \geq 0$ で $\sum_{v \in V} y(v) = 1$ となる．繰り返しステップ数 s を用い，ランダムウォーク期待値ベクトル \mathbf{y} は以下の更新式の極限分布として定義される：

$$\mathbf{y}_s^T = \mathbf{y}_{s-1}^T \mathbf{P} \quad (3.6)$$

ここで \mathbf{b}^T は \mathbf{b} ベクトルの転置を表わす．単一コンポーネントの自己ループ付き無向ネットワークを対象とすれば，推移確率行列 \mathbf{P} は非周期かつ既約であるため，初期ベクトルによらない唯一の最大固有値を有し，極限分布が定常ベクトルに収束することがペロン・フロベニウスの定理により保証される．

また，ノード u に注目すると，

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} y_{s-1}(v) \cdot p(v, u) \\ &= \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \end{aligned} \quad (3.7)$$

で計算される．ノード u の値の極限值は，ノード u の次数 $|\Gamma(u)|$ により決定される [46]. *1

$$y_\infty(u) = \frac{|\Gamma(u)|}{\sum_{v \in V} |\Gamma(v)|}. \quad (3.8)$$

このことは，式 3.8を式 7.2の $y_{s-1}(u)$ に代入すると，

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} \left\{ \frac{1}{|\Gamma(v)|} \cdot \frac{|\Gamma(v)|}{\sum_{w \in V} |\Gamma(w)|} \right\} \\ &= \frac{|\Gamma(u)|}{\sum_{w \in V} |\Gamma(w)|} \end{aligned} \quad (3.9)$$

となり，明らかである．反復を繰り返し，各ノードの値は式 3.8に収束する． $\|\mathbf{y}_s - \mathbf{y}_{s-1}\| < \varepsilon$ となるか，所定の回数 S まで繰り返し，各反復回数でのノード u の値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 の機能ベクトル（収束曲線）と呼ぶ．各ノードの収束する値は，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．FCE法では，初期ベクトル $\mathbf{y}_0 = (1/|V|, \dots, 1/|V|)^T$ で収束曲線を計算する．

3.4.2 ノード間の類似度計算

各ノードの機能ベクトル間のコサイン類似度により，ノード間の類似度を計算する．ノード u と v の類似度 $\rho(u, v)$ は，

$$\rho(u, v) = \frac{\mathbf{x}_u^T \cdot \mathbf{x}_v^T}{\|\mathbf{x}_u\| \cdot \|\mathbf{x}_v\|} \quad (3.10)$$

*1 文献 [46]ではVoter Modelに関して述べられているが，無向ネットワークではPageRankも同様である．

で計算される．コサイン類似度は，ノルムが 1 となるよう正規化するため，最終的な収束した値の高低は影響せず，収束までの変化パターンの類似性による．以上のようにしてノード間類似度 $\rho(u, v)$ を要素とする類似度行列 $\mathbf{R} = [\rho(u, v)]$ を得る．

3.4.3 K -medians法によるクラスタリング

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ピボットを選定する．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）ピボットのクラスタに割り当てる．

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，既に選定したピボットを固定し，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ピボット集合に追加し， K 個のピボットを求める方法である．既に選定したピボット集合を \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，本論文では，以下の目的関数を考える．

$$f(\mathcal{P} \cup \{w\}) = \sum_{v \in \mathcal{V}} \max\{\mu(v; \mathcal{P}), \rho(v, w)\}. \quad (3.11)$$

ここで， $\mu(v; \mathcal{P})$ は既に選定されたピボッ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 3.11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，最も類似度の高いピボット $p_k \in \mathcal{P}$ のクラスタ C_k に割り当てる．

図 3.1で，具体例を用いて説明する．各オブジェクトは2次元のベクトルで表され，類似度が高いほど近傍に位置しているとする．

まず，全オブジェクトの中から最も他のオブジェクトとの類似度の和が最大のオブジェクトを選定し，1つ目のピボットとする（図 3.1(a)）．次に，1つ目のピボット，あるいは2つ目のピボット候補のいずれか類似度が大きい方のオブジェクトとの類似度の和が最大なピボット候補を2つ目のピボットとして選定する（図 3.1(b)）．以下同様にして，既に選定されたピボット，あるいは k 番目のピボット候補のいずれか類似度が大きい方のオブジェクトとの類似度の和が最大となるピボット候補を k 番目のピ

ピボットとして選定し、これを所望の $k = K$ になるまで繰り返す (図 3.1(c)(d)). K 個のピボット選定が終了し、各オブジェクトを最も類似度の高いピボットのクラスタに割り当てる (図 3.1(e)).

3.4.4 時間計算量

FCE法の時間計算量は、PageRankスコアの収束までの反復回数 S とし、PageRankスコア計算に $S \times |E|$ 、コサイン類似度計算に $S \times |V|^2$ であり、主たる計算量は $O(S \times |V|^2)$ である。Versim法より平均次数 \bar{d} 倍、Simrank法より平均次数の2乗 \bar{d}^2 倍計算量が少ないことがわかる。表 3.1に各手法の時間計算量をまとめる。

表3.1 各種法の時間計算量

手法	Versim法	Simrank法	FCE法
時間計算量	$O(S \times V ^2 \times \bar{d})$	$O(S \times V ^2 \times \bar{d}^2)$	$O(S \times V ^2)$

3.5 評価実験

コミュニティ抽出結果を可視化により定性的に、また実行時間により定量的に評価する。FCE法と同様に、Versim法、Simrank法により算出した類似度に基づき、 K -medians法により K 個のクラスタに分割する。

3.5.1 ネットワークデータ

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モデルにより生成した人工ネットワークである [6]。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。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、この章ではHierarchicalネットワークと呼ぶ。

2 つ目のネットワークは、Lattice ネットワークである。2次元平面上の正方格子であり、縦に10、横に10でノード数は100のネットワークを作る。この章ではLattice ネットワークと呼ぶ。このネットワークは、上下左右に同じ構造が連続しており、クラスタの判別が難しい事例である。

3 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。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

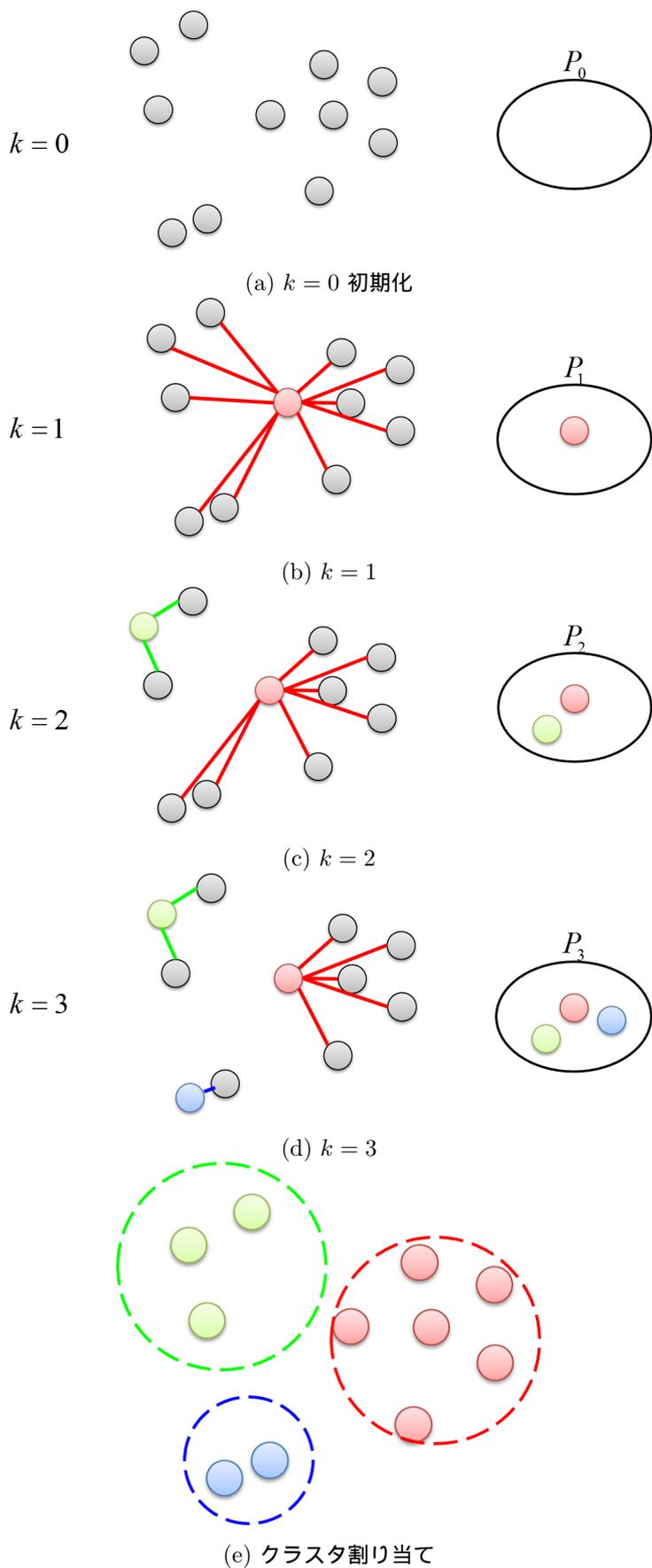


図3.1 K -medians法の説明 (貪欲法)

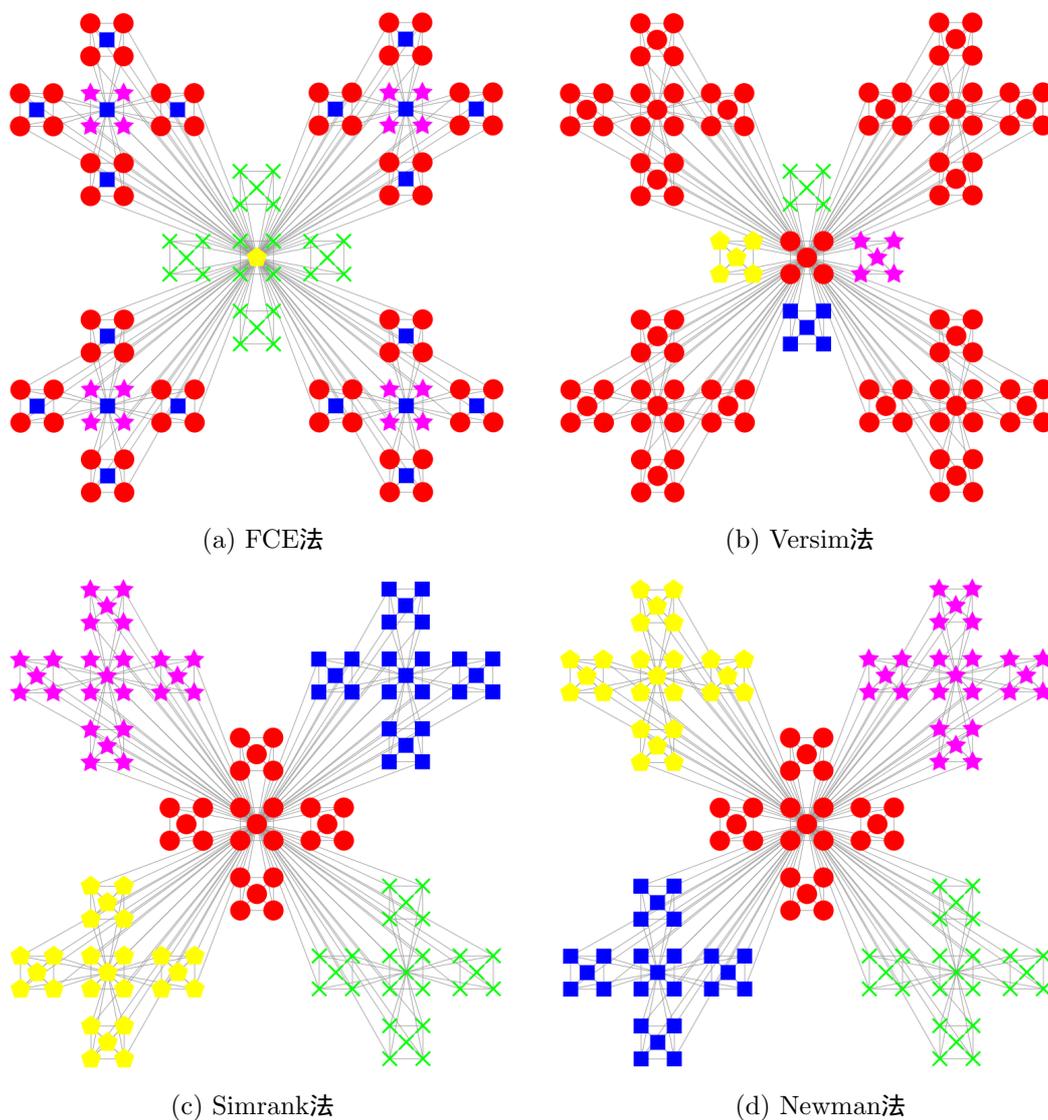


図3.2 Hierarchicalネットワーク ($K = 5$) 色とマーカーの種類は各クラスターを意味し、リンクの長さに本質的な意味はない。

ト性を有する [47]。この章ではKarateネットワークと呼ぶ。

4つ目のネットワークは、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構築したハイパーリンクネットワークである。本稿ではスペースの都合上、法政大学情報科学部のホームページ^{*2}のネットワーク（以下 Hoseiネットワーク）に対する結果を示す。この例はWebサイトの構造を分析するために用いる。

5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県版 [48]をもとに作成した。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。この章では、このネットワーク

^{*2} 法政大学情報科学部 <http://cis.k.hosei.ac.jp/>

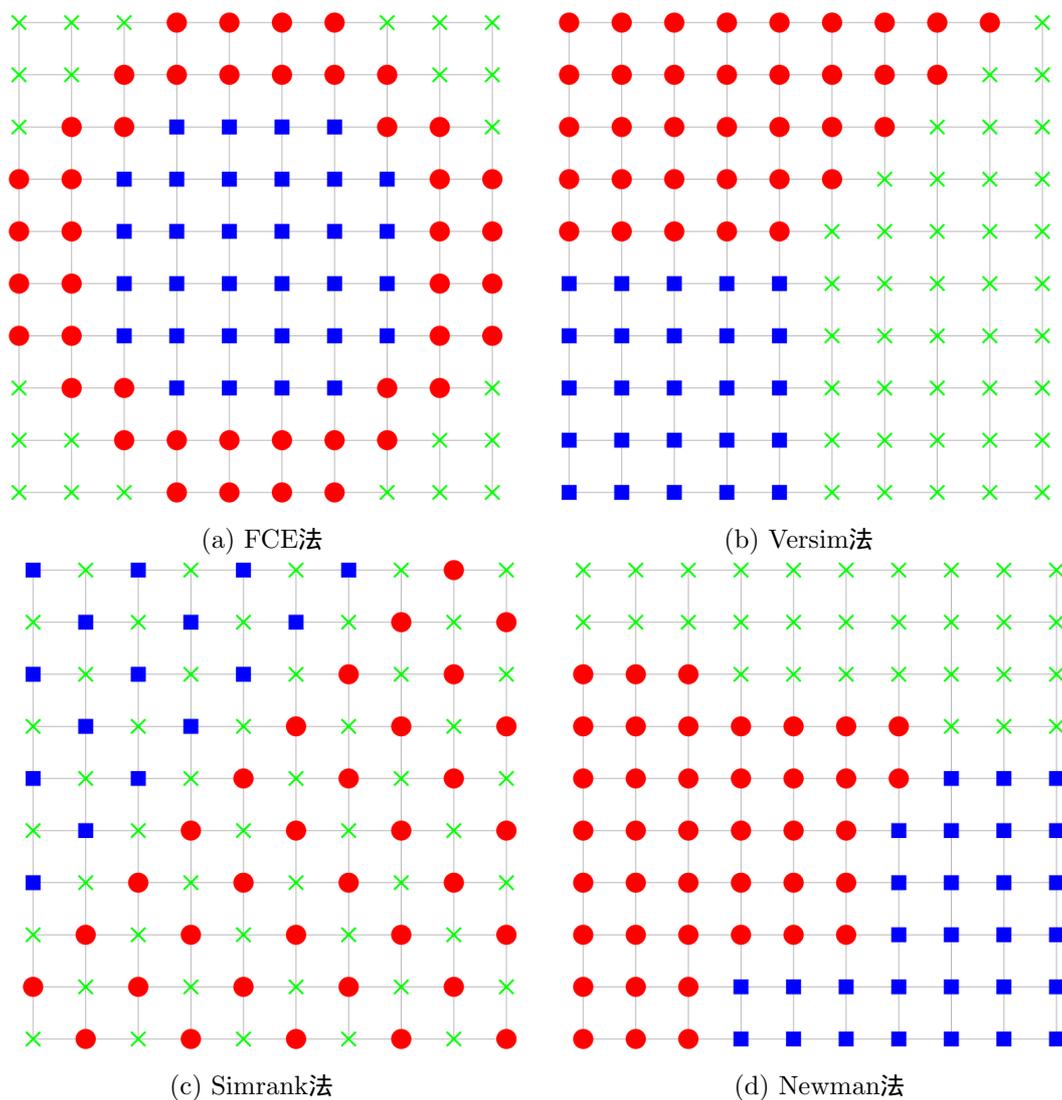


図3.3 Latticeネットワーク ($K = 3$) 色とマーカーの種類は各クラスタを意味し、リンクの長さに本質的な意味はない。

をShizuokaPrefネットワークと呼ぶ。ShizuokaPrefネットワークのノード数は104,526，リンク数は155,960である。

6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 [49]をもとに作成した。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。この章では，このネットワークをShizuokaCityネットワークと呼ぶ。ShizuokaCityネットワークのノード数は31,963，リンク数は44,467である。

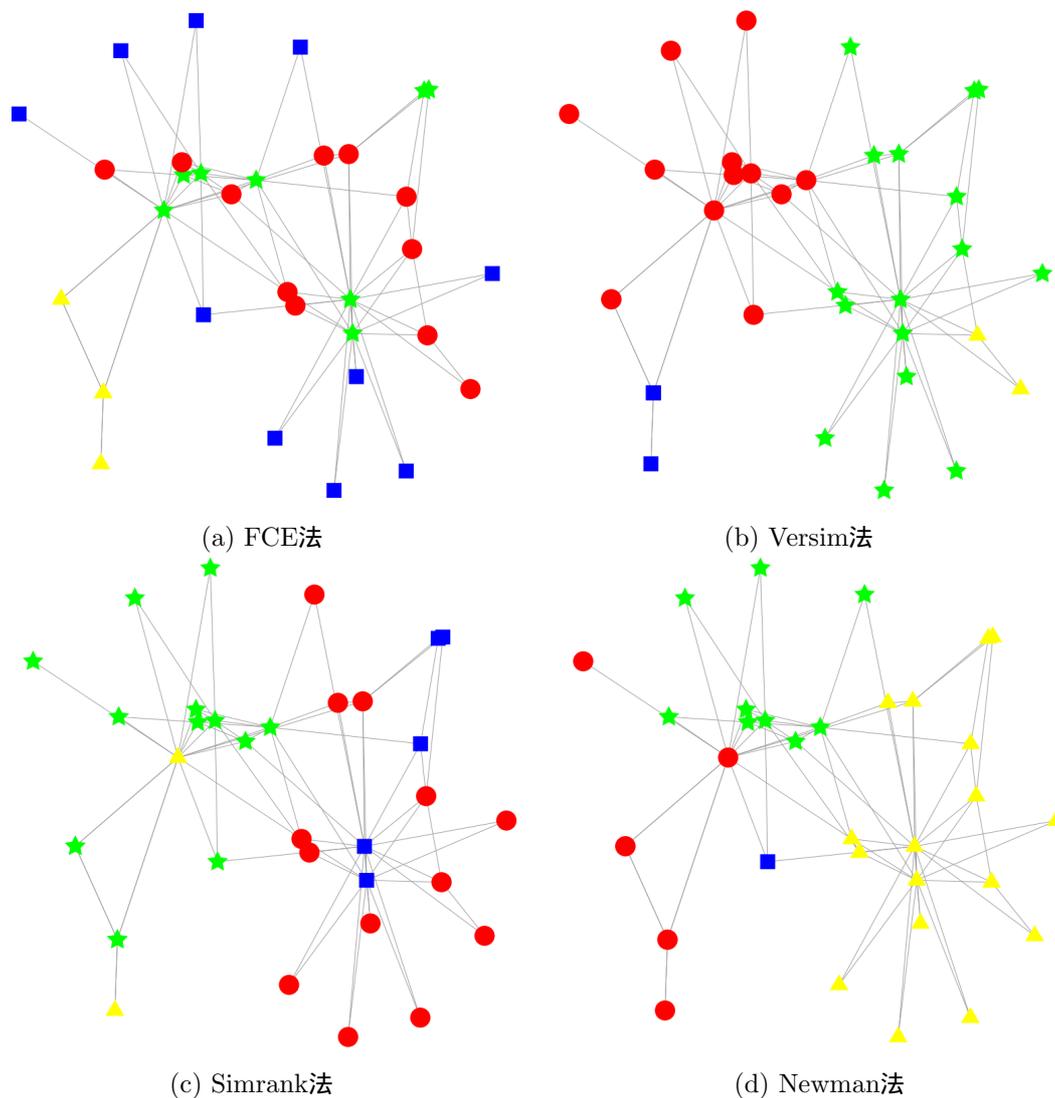


図3.4 Karateネットワーク ($K = 4$) 色とマーカーの種類は各クラスタを意味し、リンクの長さに本質的な意味はない。

3.5.2 可視化による定性的評価

上述したネットワークに対する結果を図 3.2 ~ 3.7 にそれぞれ示す。なお説明の便宜上、適切なクラスタ数 K を図示しているが、他の K の場合でも我々の実験の範囲では、ほぼ同様の結果が得られた。各手法の収束判定は $\varepsilon = 10^{-12}$ とした。Hierarchical ネットワークは文献 [6] に従い、Karate ネットワークおよび Hosei ネットワークはクロスエントロピー法により可視化した [50]。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は意味はないことに注意する。各可視化結果において、同一の色・マーカー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。

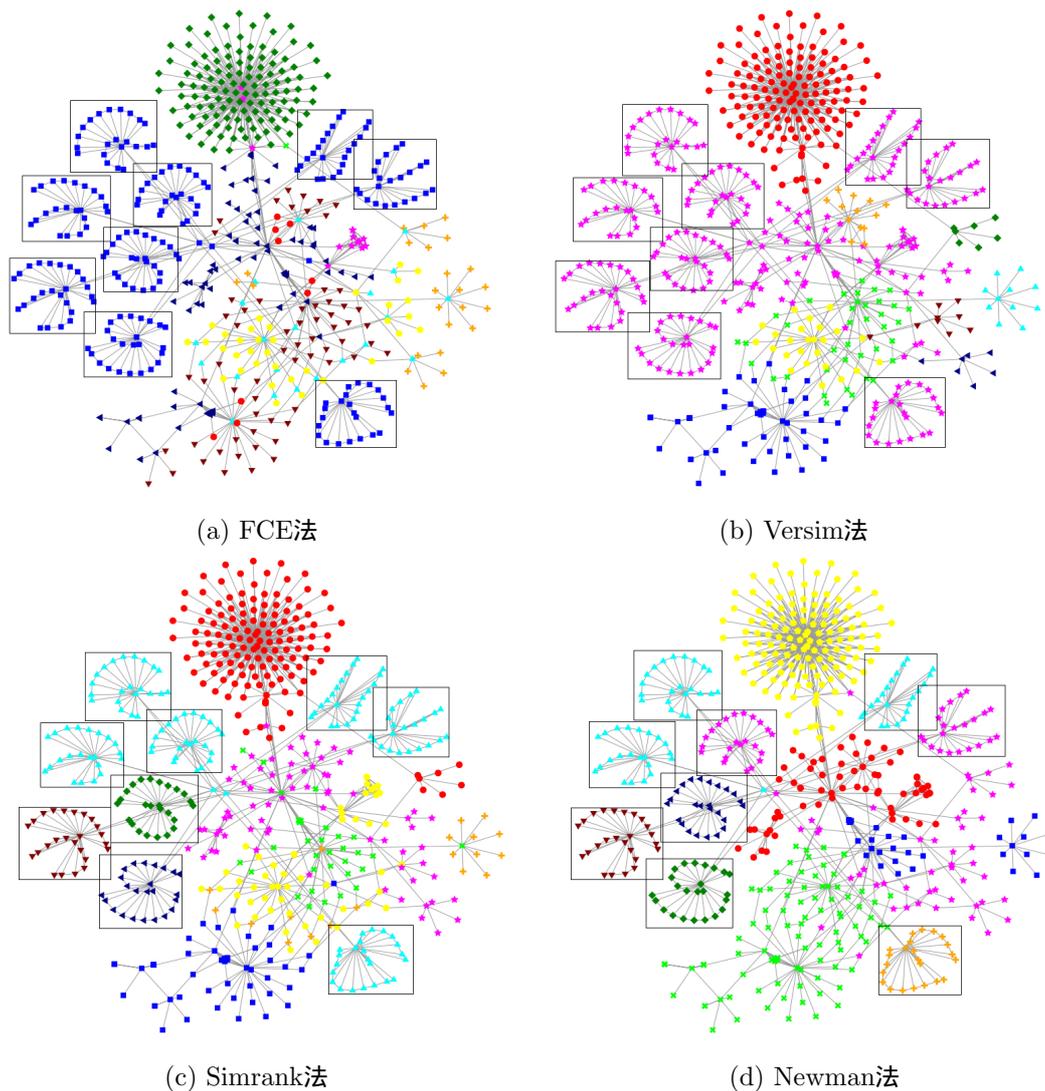


図3.5 Hosiネットワーク ($K = 10$) 年度ごとの教員成果報告ページ群を白抜きで四角で囲っている。色とマーカーの種類は各クラスターを意味し、リンクの長さに意味はない。

Hierarchicalネットワーク

Hierarchicalネットワークの結果(図 3.2)を比較すると、FCE法(a)では、階層上の同質(同一階層)のノード、すなわち、同質の機能・役割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。会社組織でたとえるならば、部長、課長のような役割別にコミュニティが抽出されている。一方Newman法(d)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。会社組織でたとえるならば、営業部、人事部のような部門別にコミュニティが抽出されている。Simrank法(c)でも同様に、密結合するノード群が同一のコミュニティとして抽出されている。またVersim法(b)では、隣接度により結合するノード群()と外側の()と

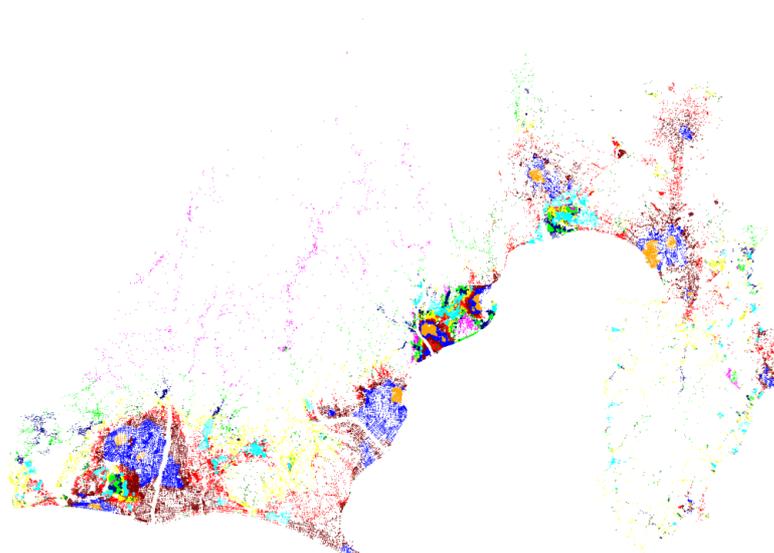
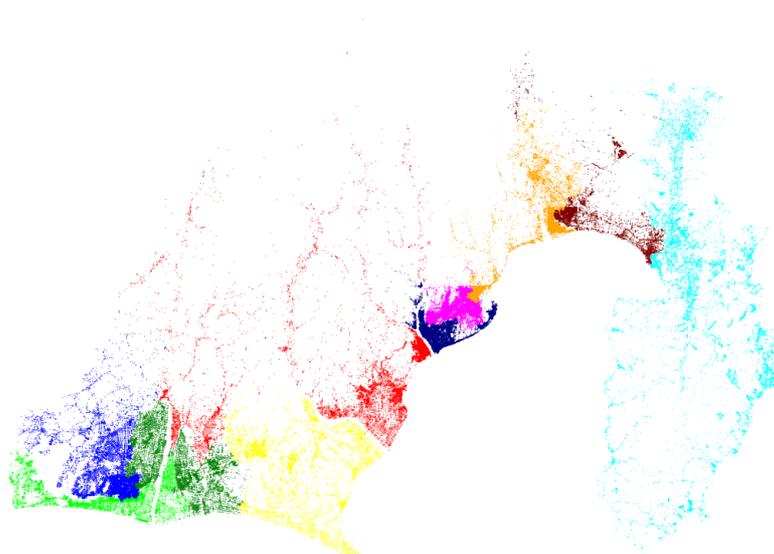
(a) FCE法 ($K = 10$)(b) Newman法 ($K = 10$)

図3.6 ShizuokaPrefネットワーク

直接結合しないノード群（以外）に分割されている。

Latticeネットワーク

Latticeネットワークの結果（図 3.3）を比較すると，FCE法では，ネットワークの全体に対する相対的な位置関係の違いで，中心部（），末端部（×），中間部（）に分割されている．一方Newman法

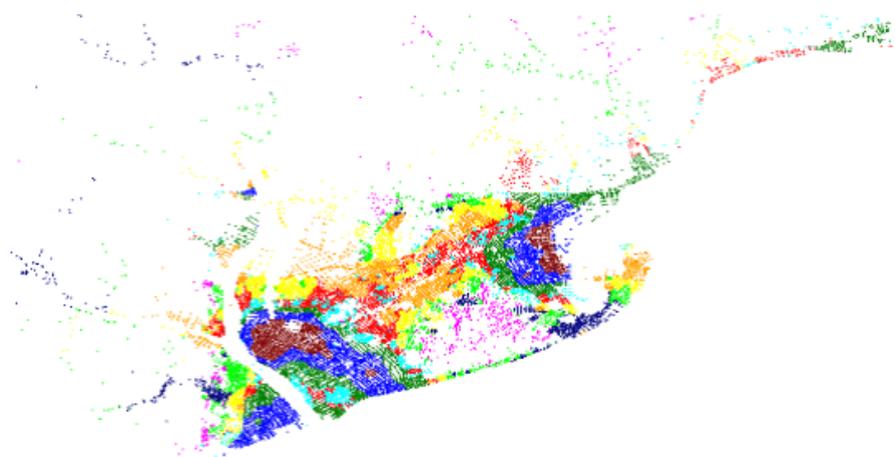
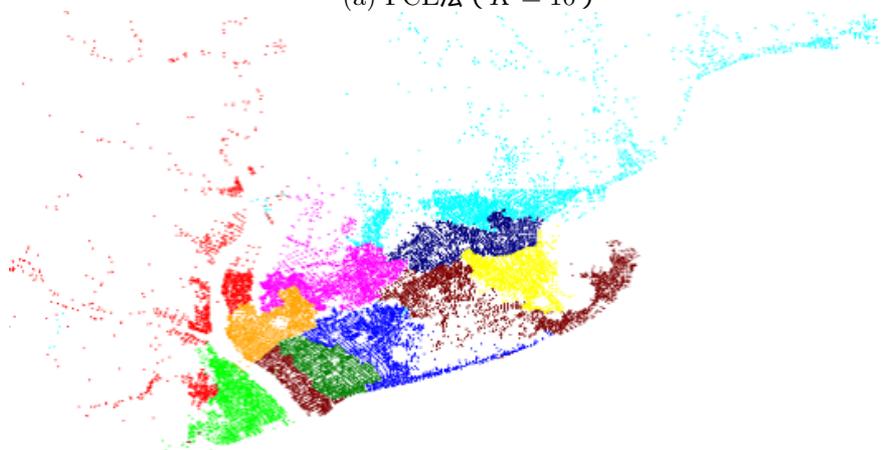
(a) FCE法 ($K = 10$)(b) Newman法 ($K = 10$)

図3.7 ShizuokaCityネットワーク

およびVersim法の結果は、リンク密度や隣接度により近傍ノードを群をまとめて分割している様子が見える。Simrank法の結果は、共通隣接ノードを有するノードペアの類似度が高くなるため、2部グラフのような強い周期性（ \times 、 \times ）がみられる結果となった。Latticeネットワークは正方形格子で規則正しい構造をしているため、局所的に見るとどの部分も等しい構造をしており、局所的な構造しか考えないVersim法やSimrank法は、例えば同じ機能を果たしているはずの四隅が同一のコミュニティに分類されないように、ノードを立場・役割で分類することができない。FCE法はネットワーク全体を見ているので、全体における立場・役割でノードを分類できている。

Karateネットワーク

Karateネットワークの結果（図 3.4）を比較すると、FCE法では、ハブ的な存在のノード（ \times ）、ハブ間の橋渡しの役割かつ互いに結合するノード（ \times ）、ハブとだけ友人関係にあるノード（ \times ）、ハブと

繋がりがあつかつ小グループを形成するノード()に分類されている。抽出されたコミュニティ内のノード同士は隣接していない場合もあるが、同様な立場にあり役割・機能が類似したノードを同一のコミュニティとして抽出している。一方Versim法やNewman法では、隣接性を強く考慮しているため、密結合するノード群を同一のコミュニティとして抽出している。Simrank法では、共通隣接ノードを有するノードペアの類似度が高くなるため周期性()がみられる(2部グラフのように完全に分かれていないため、完全な周期性ではない)。

Hoseiネットワーク

Hosei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迎れるが、年度間のリンクは存在しない。Hoseiネットワークの結果(図 3.5)を比較すると、FCE法では、可視化結果の左側部分の6つのノード群や右上2つ、右下1つのノード群のノード()は同じコミュニティに分割されている。このノード群は、上述した対象大学の各年度の教員の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。Versim法でも同様に、研究成果ページ群()を同一コミュニティとして抽出しているが、それらの間にあるページも同一のコミュニティとして抽出している点でFCE法と異なる。これは、定義式から隣接性を強く考慮しているためと考えられる。一方Newman法は、各年度の教員研究成果ページ間に直接リンクが存在しないため、異なるコミュニティとして抽出しているSimrank法では、教員研究成果ページ群を同一コミュニティとして抽出しているが、抽出漏れがあることがわかる。

ShizuokaPrefネットワーク

静岡県道路網に対する結果を図 3.6(a)に示す。本研究で用いる道路ネットワークは大規模であり、VerSim法ならびにSimRank法では時間で適切な結果が得られないため、FCE法による結果のみを示す。図 3.6(a)を見ると、西(図中左)から高塚駅、焼津駅、静岡駅、清水駅、富士宮駅、沼津駅、三島駅など主要な市町村の東海道線の駅周辺地域が抽出できているのがわかる(橙、青、茶)。これら抽出された駅周辺の繁華街の規模そのものは異なるものの、類似の道路構造をしていることから同一コミュニティとして抽出できていると考えられる。一方、浜松市も主要な市町村の一つであるが、第二次世界大戦で静岡県で最も大きな被害を受け、戦後に都市や道路が新たに再構築されたために、上記の都市と同一のコミュニティにならなかったと考えられる。完全に一致はしていないが、駅周辺の繁華街以外にも住宅街・集落地域(赤)や山岳地帯(桃)、工業用地を含む同一サイズの格子状に造られた地域(水)などが抽出された。

一方Newman法では、対象データにおける地域コミュニティに類する結果が得られた(図 3.6(b))。

ShizuokaCityネットワーク

静岡市道路網に対する結果を図 3.7(a)に示す。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二ヶ所にそれぞれの市の中心部が存在する。図 3.7(a)を見ると、概ね駅周辺（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わかる。また、コミュニティが同心円状に広がっていることも見て取れる。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。静岡市道路網での機能コミュニティの抽出結果は、静岡市の土地利用方針図^{*3}における土地利用区分と類似しており、抽出結果にある程度の妥当性があると言える。さらに、静岡市道路網は静岡県道路網の中央部分にも含まれているが、静岡県全体での処理結果と静岡市のみでの処理結果は大きな矛盾のない結果となっており、静岡県全体の処理結果の細分類になっていることもわかる。

一方Newman法では、対象データにおける地域コミュニティに類する結果が得られた（図 3.7(b)）。

定性的評価のまとめ

これら異なる構造のネットワークに対する実験結果より、FCE法では、トップノードからの深さやネットワーク内での相対的位置、周辺リンク構造の類似性など、ネットワークに対する機能が類似する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。Versim法やSimrank法は、FCE法と近い結果が得られる場合があった。Versim法は隣接性を強く意識しているため、FCE法で同一と判定されたコミュニティ間を繋ぐ部分（異なるコミュニティ）も一緒に抽出される傾向があった。Simrank法では、直接隣接することより共通隣接ノードを有するかに焦点を当てているため、コミュニティ抽出結果に周期性が見られた。Versim法やSimrank法は、正則同値なノードを近似的に同定するための手法であるが、局所的なリンク構造の類似性に直接着目することから、ノードの機能を考慮しないNewman法と近い結果が得られる場合もあった。

また、ある種の類似機能を有するノード群があった場合、それらのノードはある特有の構造を有することが観測できる。しかし、現実ネットワークのように表出する構造にバラつきがある場合、構造だけで機能の類似性を判定することは困難である場合がある。一方、ネットワーク全体でのランダムウォークにおいて、各ノードへの到達確率という期待値の収束曲線の類似性を扱うFCE法では、ある程度のバラつきがある場合においても正しく同定できたと考えられる。

^{*3} http://www.city.shizuoka.jp/deps/tosikeikaku/koshin_toshi_kei_toshimasu_shizuoka_2syo_kakudai_2_3-b.htm

3.5.3 時間計算量による定量的評価

ノード間類似度に基づく3手法を時間計算量の点から定量的に評価する．各手法の収束判定は $\varepsilon = 10^{-12}$ とした．評価実験には，ノード数・リンク数の異なる6つのランダムなネットワークを用いる（表 3.2）．理論的には，Versim法はFCE法の平均次数 \bar{d} 倍，Simrank法は平均次数の2乗 \bar{d}^2 倍の計

表3.2 ネットワークの統計量

	NW1	NW2	NW3	NW4	NW5	NW6
$ V $	100	500	1000	5000	10000	50000
$ E $	500	2500	5000	25000	50000	250000

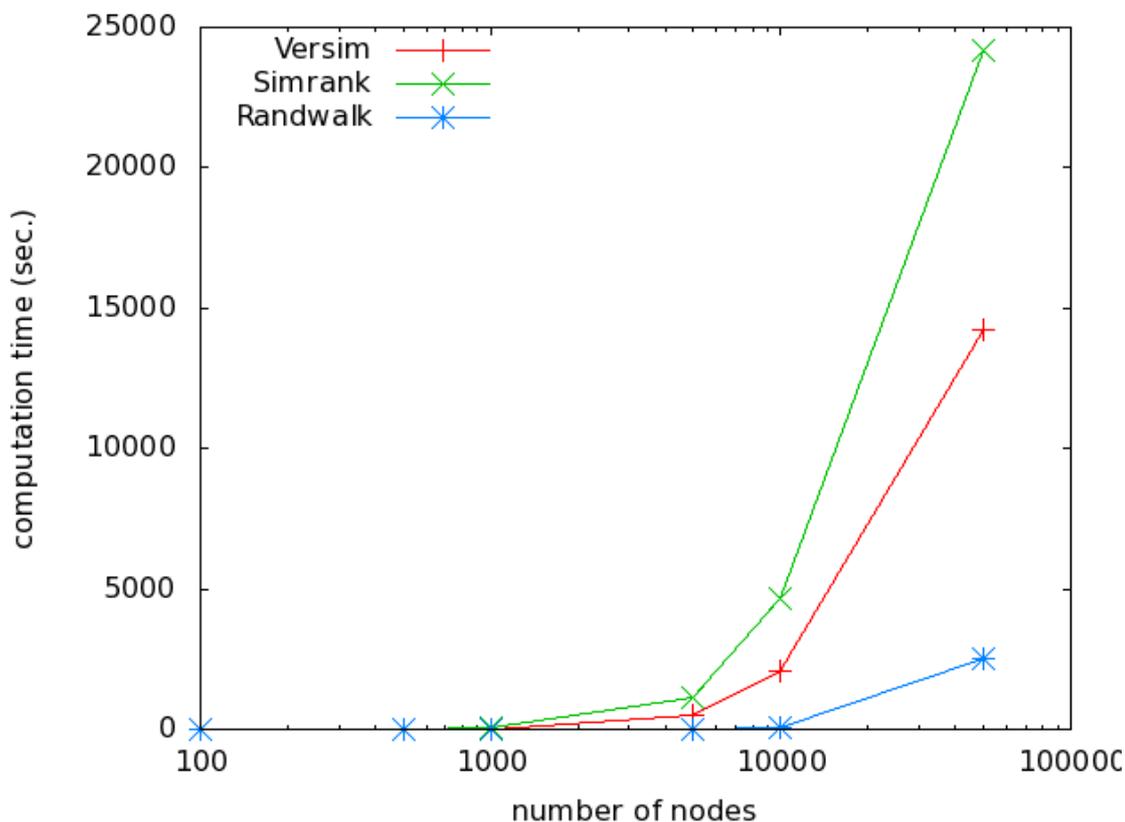


図3.8 実行時間

算量を必要とする．Versim法とSimrank法は収束するまで類似度行列を反復させるが，FCE法は機能ベクトルが収束するまで反復させるだけなため，反復回数においてFCE法が有利となる．また，Versim法とSimrank法は局所的なノード間の関係を見ている一方で，FCE法はネットワーク全体を考慮しているが，図 3.8から実際の計算時間を比較すると，FCE法は他の2手法より高速に計算できており，大規模な

ネットワークに対しても有効な手法であると言える。

3.6 考察

FCE法は、比較した他の手法と異なり、機能コミュニティをより明示的に抽出できることがわかった。評価実験より、あるノードと類似した機能のノードを有する部分ネットワークを抽出可能であることが示唆された。FCE法における機能ベクトルの要素は、ランダムウォークの各ステップにおける期待値を意味する。会社組織でたとえるならば、任意の社員からランダムに隣接社員へE-mailを送信する試行を繰り返したとき、各ステップにおけるE-mailを受け取る確率ということができる。周辺社員とのリンクパターンが類似する社員同士は、各ステップでE-mailを受け取る確率の推移が類似すると自然に想定できる。このようにして、FCE法では、機能（性質・役割）の類似するノードを同定できたと考えられる。

ノードの性質として、ソーシャル・ネットワーク分析の分野で有名な中心性指標 [8]やPageRank [15], HITS [16]などが挙げられる。これらの指標は単一尺度であるため、次数や近接度、媒介度、ハブ度といった特定の性質を判定することしかできない。一方、FCE法による機能コミュニティは、機能ごとにクラスタリングすることから、ハブやオーソリティ、ゲートキーパなどのグループを抽出可能であり、多次元的な分析が可能となる。実際に、図 3.4(a)におけるハブノード()とハブ間の橋渡しノード()はともに、HITSランキング(ハブ度)上位となり分類できない。クラスタ係数に関していうと、ハブノード()は極めて小さく、ハブ間の橋渡しノード()は相対的に大きいという違いでFCE法は分類できている。このように、中心性概念にはない性質のノード群も分類することができた。

FCE法は、初期ベクトル $y_0 = (1/|V|, \dots, 1/|V|)^T$ により収束曲線を計算している [44]。しかし、ノードの役割や機能は、視点となるノードを変えれば異なるものになるというのが直感的である。すなわち、ノード u に対するノード w の役割は、ノード v に対するノード w の役割とは異なるということである。FCE法は、初期ベクトル $y_0 = (0, \dots, 1, \dots, 0)^T$ のようにノード u に対応する要素のみ 1 でその他の要素は 0 としたベクトルとすることにより、ノード u の視点から、他のノードの機能・役割をクラスタリングすることができる。我々はパーソナライズ機能コミュニティと呼ぶ。収束する値は初期ベクトルに依存しないが、収束曲線のパターンは変化するため、通常FCE法とは異なる結果が得られる。個々のノードからの視点によるコミュニティ抽出法は、今までにない新たなパラダイムであり、今後の発展性に大いに期待できると考えられる。

3.7 おわりに

この章では、従来のリンク密度に基づくコミュニティとは異なり、ノード間の類似性に着目したコミュニティ抽出法に焦点を当てた。ノード間の同値性を近似的に計算する手法としてVersim法とSimrank

法，類似機能を有するノードを同定するFCE法を取り上げ，3手法に基づいたコミュニティ抽出の結果を可視化により定性的に，時間計算量により定量的に評価した．Versim法，Simrank法ともに近似的に正則同値を同定するための手法であるため，FCE法と近い結果が得られる場合があった．しかしVersim法は，隣接性が強く考慮されているため，FCE法で異なると判定されたコミュニティも同一コミュニティとして抽出する傾向があった．またVersim法，Simrank法は，局所的なリンク構造の類似性に直接着目することから，ノードの機能を考慮しないNewman法と近い結果が得られる場合もあった．FCE法は，機能ベクトルとしてPageRankスコアの収束曲線を用いて，大域的な構造上における現象の類似性を見ており，かつスケーラビリティのある方法である．その結果，局所構造をみるVersim法やSimrank法と比較して，全体構造をみるFCE法は同質の機能・役割を有するノード群を要素とするコミュニティをより適切に抽出できることが示された．また，中心性指標とは異なり，一義的な性質の判別ではなく，性質ごとのコミュニティを抽出可能であることを示した．さらに，個々のノードからの視点によるコミュニティ抽出といった，発展性のある手法であることも示唆された．今後は，有向ネットワークや多重ネットワーク，2部グラフなどの一般的なネットワークを対象としたコミュニティ抽出などへの拡張を検討していくつもりである．

第4章

有向ネットワークに対する 機能コミュニティ抽出法：DFCE法

4.1 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、全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、重要ノードを抽出するための手法が提案されている [8, 14, 15]。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 [17]。

この章では、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FCE法がある [44, 45]。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、これらが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機能コミュニティを抽出することは重要である。この章では、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（DFCE法）を提案する。無向ネットワークを対象とした標準のFCE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おこな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わからない、片思いや両思いと

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。

標準のFCE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。従って、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一考察をする。

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 [51]。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。この章で提案するDFCE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。

4.2 有向機能コミュニティ抽出法：DFCE法

この節では、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。DFCE法における変化曲線の計算法について説明する。

DFCE法は、有向ネットワーク $G = (V, E)$ 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 $\{y_1, \dots, y_S\}$ を計算;
2. 各ノードの機能ベクトルとしてPageRankスコア変化曲線 x_v を構築;
3. 各ノードペアの機能ベクトル x_u と x_v のコサイン類似度 $\rho(u, v)$ を計算;
4. K -medians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;

以下に機能ベクトル構築法について詳細を説明する。

4.2.1 FCE法における機能ベクトル（参考）

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 $a(u, v) = 1$ 、それ以外の場合 $a(u, v) = 0$ とし隣接行列 $A \in \{0, 1\}^{|V| \times |V|}$ 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 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 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 $y_0 = (1/|V|, \dots, 1/|V|)^T$

とし、繰り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:

$$\mathbf{y}_s^T = \mathbf{y}_{s-1}^T \mathbf{P} \quad (4.1)$$

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

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} \{y_{s-1}(v) \cdot p(v, u)\} \\ &= \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \end{aligned} \quad (4.2)$$

で計算される。反復回数 S 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 $\mathbf{x}_u = (y_1(u), y_2(u), \dots, y_S(u))^T$ と定義する。このベクトル \mathbf{x}_u をノード u の変化曲線（機能ベクトル）と呼ぶ。

4.2.2 DFCE法における機能ベクトル

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 $\mathbf{y}_0 = (1/|V|, \dots, 1/|V|)^T$ とし、繰り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:

$$\begin{aligned} \mathbf{y}_s^T &= \mathbf{y}_{s-1}^T ((1 - \alpha)\mathbf{P} + \alpha\mathbf{e}\mathbf{z}^T) \\ &= (1 - \alpha)\mathbf{y}_{s-1}^T \mathbf{P} + \alpha\mathbf{z}^T. \end{aligned} \quad (4.3)$$

ここで $\mathbf{e} = (1, \dots, 1)^T$ 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 (5.3) の更新式により \mathbf{y} を更新する。ノード u に注目すると、

$$\begin{aligned} y_s(u) &= (1 - \alpha) \sum_{v \in B(u)} \{y_{s-1}(v) \cdot p(v, u)\} + \alpha \cdot z(u) \\ &= (1 - \alpha) \sum_{v \in B(u)} \left\{ \frac{y_{s-1}(v)}{|\Gamma(v)|} \right\} + \alpha \cdot z(u) \end{aligned} \quad (4.4)$$

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 S 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 $\mathbf{x}_u = (y_1(u), y_2(u), \dots, y_S(u))^T$ と定義する。このベクトル \mathbf{x}_u をノード u の変化曲線(機能ベクトル)と呼ぶ。この手法において、大域ジャンプベクトルは $\mathbf{z} = (1/|V|, \dots, 1/|V|)^T$, パーソナライズベクトルは $\mathbf{g} = (1/|V|, \dots, 1/|V|)^T$ である。

4.3 評価実験

DFCE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。

4.3.1 ネットワークデータ

実験では、DFCE法により抽出したコミュニティの特徴を捉えるために、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。この章では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 [6]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、この章では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したがい、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。この章では、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

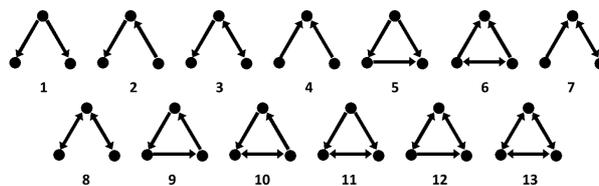


図4.1 モチーフパターン

ク, Yaku ネットワーク) に対する結果を示す*1.

4.3.2 モチーフパターンベクトルによる分類

DFCE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.

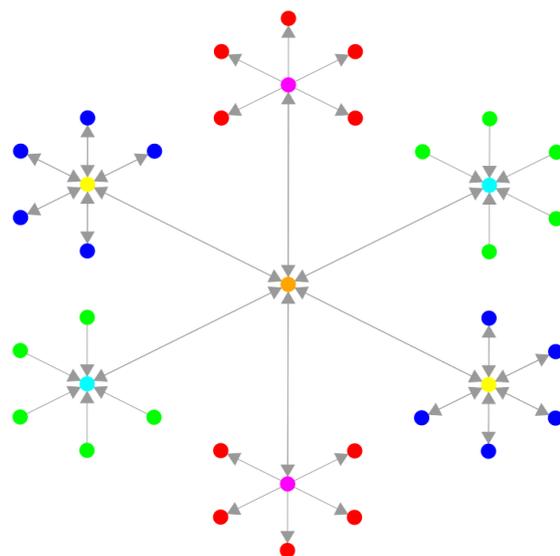
ネットワークモチーフは, 有向ネットワーク内のリンクパターン(モチーフパターン)の出現頻度を数え上げることで, そのネットワークの特徴的なモチーフパターンを抽出する解析法である. この章では, 図 4.1 に示す13種のモチーフパターンに対して, 各ノードが各パターンにどの程度関与しているかを数え上げ, その値を要素とするベクトルを各ノードの特徴ベクトルとする. 形式的には, ノード u がパターン i に $n_{u,i}$ 回関与していると, ノード u の特徴ベクトルは $\mathbf{m}_u = (n_{u,1}, \dots, n_{u,13})^T$ となる.

各ノード間の類似度およびクラスタリング法は, FCE法と同様に, コサイン類似度, K -medians法をそれぞれ適用する. この章では, この手法をパターンベクトルによる分類と呼ぶ.

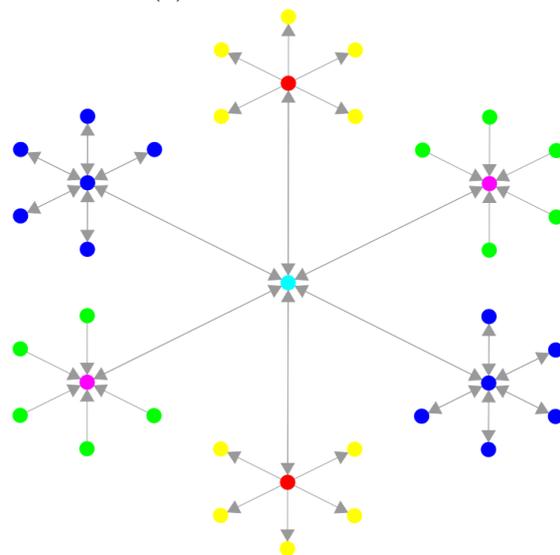
4.3.3 実験設定

4つのネットワークに対し, 反復終了ステップ数 $S = 500$, 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 4.2から図 4.5までは, 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標準のFCE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 [50]. 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 4.6から図 4.9は, K -medians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 4.2(a)から図 4.5(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.

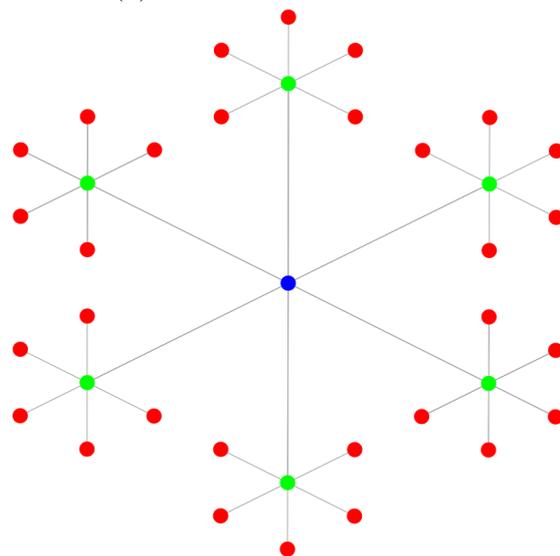
*1 法政大学情報科学部<http://cis.k.hosei.ac.jp/>, 静岡県立大学薬学部<http://w3pharm.u-shizuoka-ken.ac.jp/>.



(a) 有向機能コミュニティ

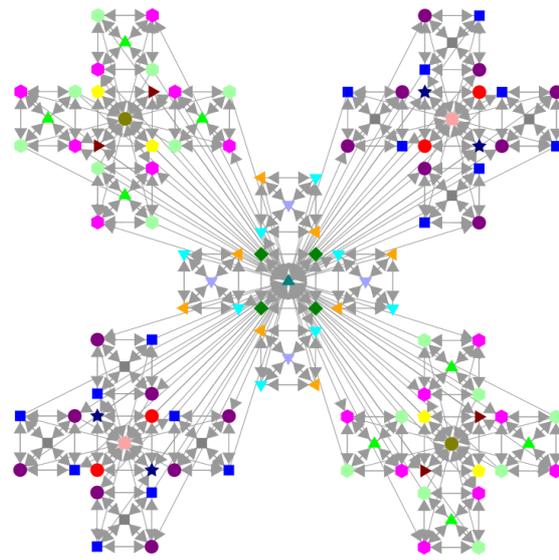


(b) パターンベクトルによる分類

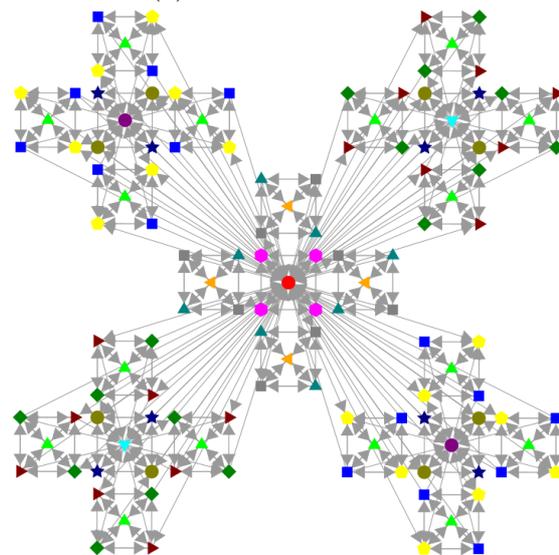


(c) 無向化機能コミュニティ

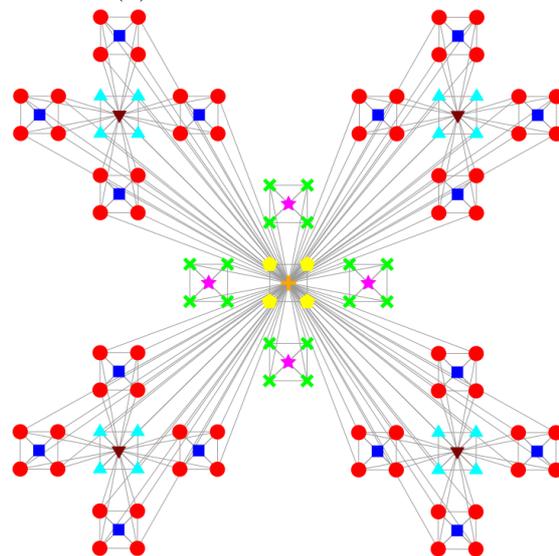
図4.2 Treeネットワーク



(a) 有向機能コミュニティ

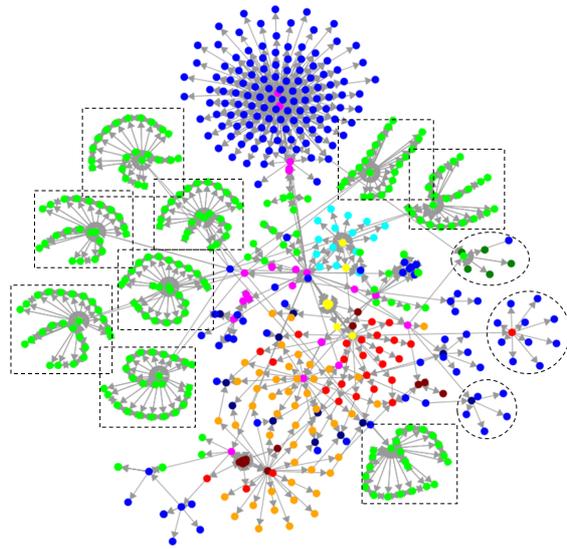


(b) パターンベクトルによる分類

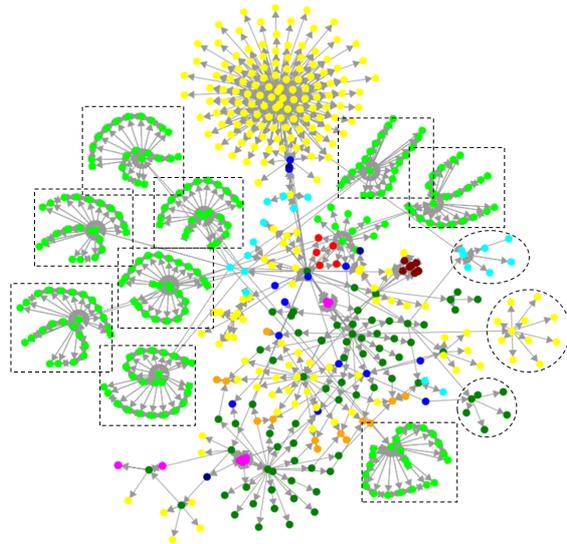


(c) 無向化機能コミュニティ

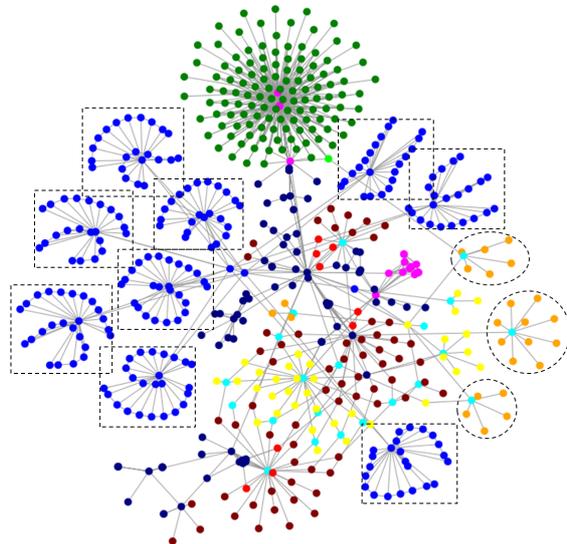
図4.3 Hierarchicalネットワーク



(a) 有向機能コミュニティ

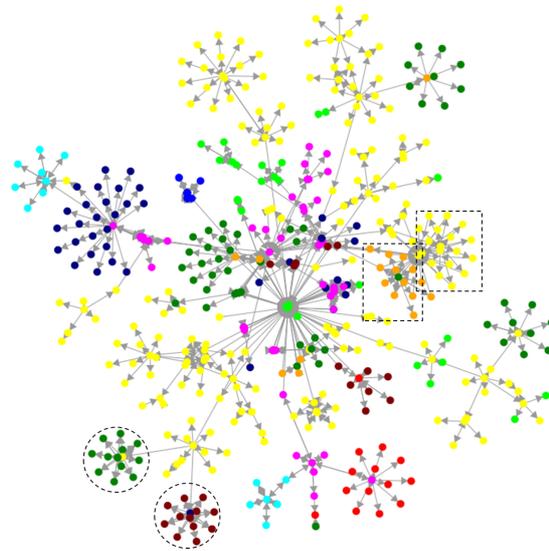


(b) パターンベクトルによる分類

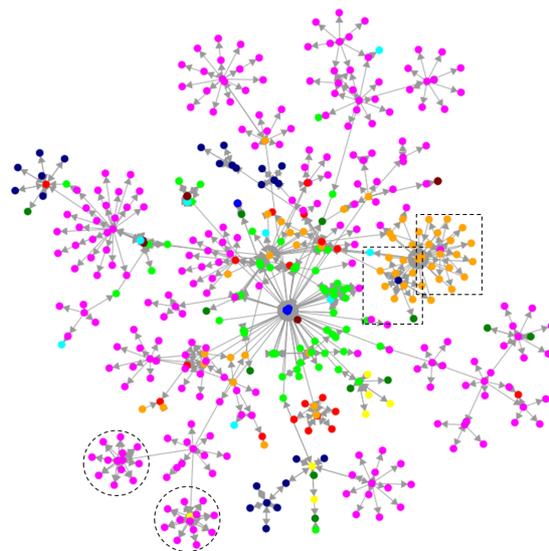


(c) 無向化機能コミュニティ

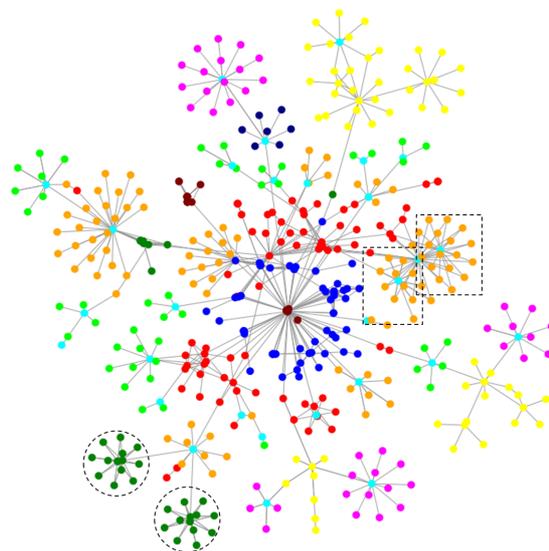
図4.4 Hoseiネットワーク ($K = 10$)



(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4.5 Yakuネットワーク ($K = 10$)

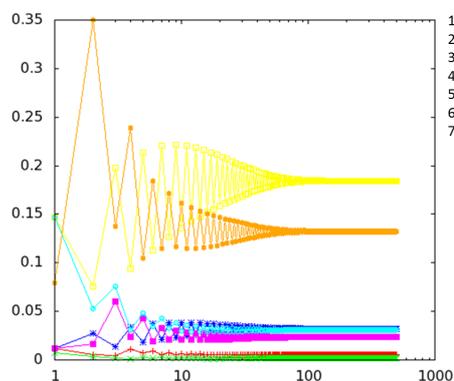


図4.6 Treeネットワーク 代表ノード変化曲線

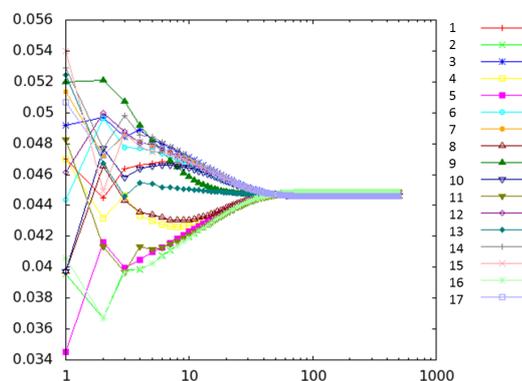


図4.7 Hierarchicalネットワーク 代表ノード変化曲線

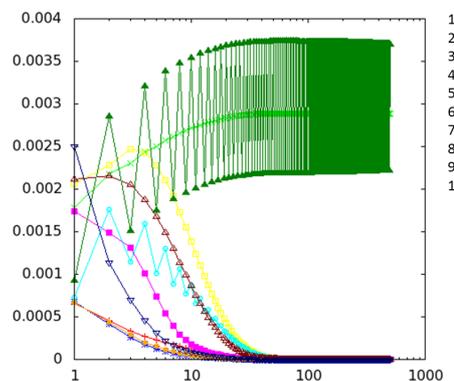


図4.8 Hoseiネットワーク 代表ノード変化曲線

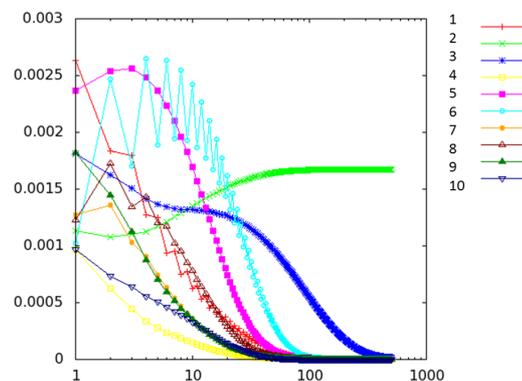


図4.9 Yakuネットワーク 代表ノード変化曲線

4.3.4 Treeネットワークの処理結果の評価

図 4.2にTreeネットワークのコミュニティ抽出結果を示す．DFCE法の結果を見ると，各ノー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わかる（図 4.2(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繋がれた中心部分と隣接ノードにそれぞれ分類されている．双方向リンクで繋がっている部分は，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もち原始的でないため，極限值を持たなく周期性が表れることが知られている [52]．無向二部グラフ的構造の中心ノードである図 4.6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

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

しており、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わかる。

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わかる(図 4.2(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標準のFCE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 4.2(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わかり、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わかる。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わかり、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しないで、相補的である。

4.3.5 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、各ノードは自身の属するセグメント内のノードと双方向リンクで繋がっているが、最上位ノードと、あるいは一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同士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わかる(図 4.3(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 4.3(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わかる。

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わかる(図 4.3(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、標準のFCE法の結果を見ると、階層

^{*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。

^{*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。

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 4.3(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。

4.3.6 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 4.4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辿れるが、年度間のリンクは存在しない。また、図 4.4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 4.4(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、標準のFCE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 4.4(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言うと、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わかる(図 4.4(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繋がっており、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 4.8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なく同一視されてしまっている(図 4.4(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 4.4(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。

*⁷ $K = 8$ 以上には分割できないため、 $K = 8$ の結果を示す。

4.3.7 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 4.5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 4.5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。

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」の関係が多くみ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わかる。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。

4.3.8 変化曲線による大域ジャンプ確率の検討

DFCE法のパラメータとして、大域ジャンプ確率 α 、反復数 T などが挙げられる。反復数は、PageRankスコアの収束する $\|y_t - y_{t-1}\|_{L1} < \varepsilon$ 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。

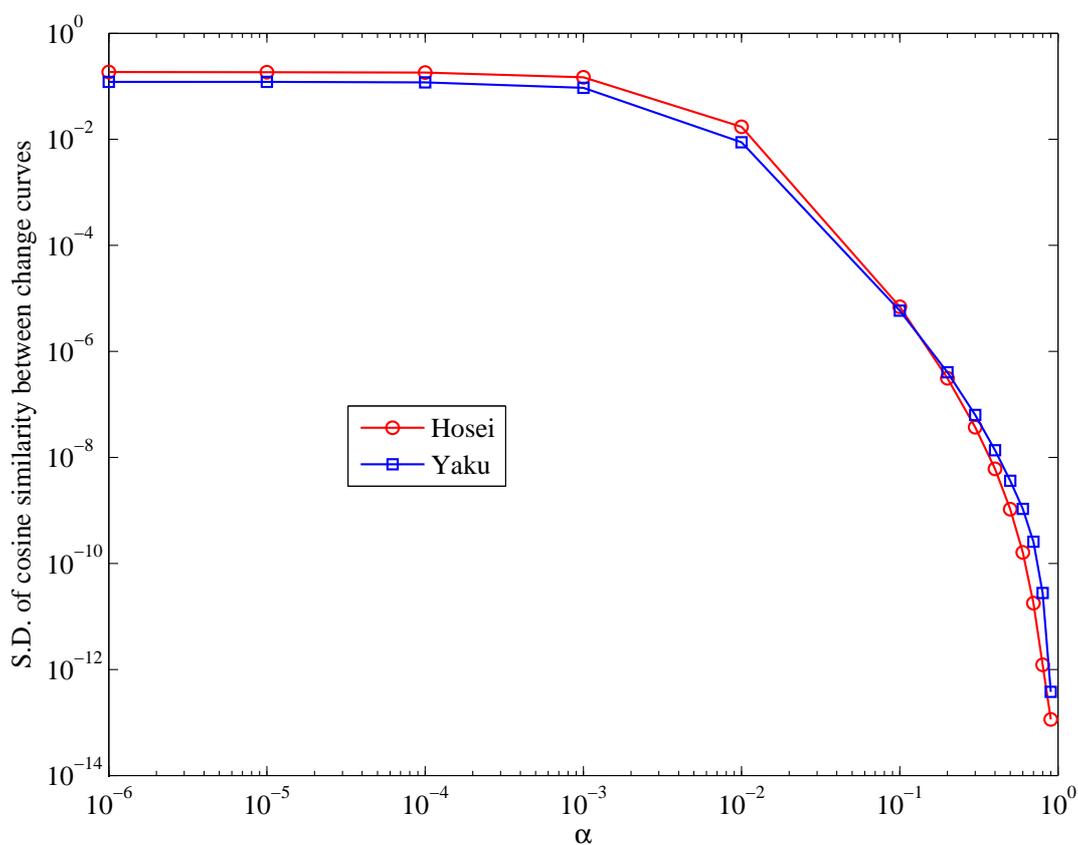


図4.10 大域ジャンプ確率と変化曲線

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 4.10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 4.10より、大域ジャンプ確率が大きいほど、コサイ

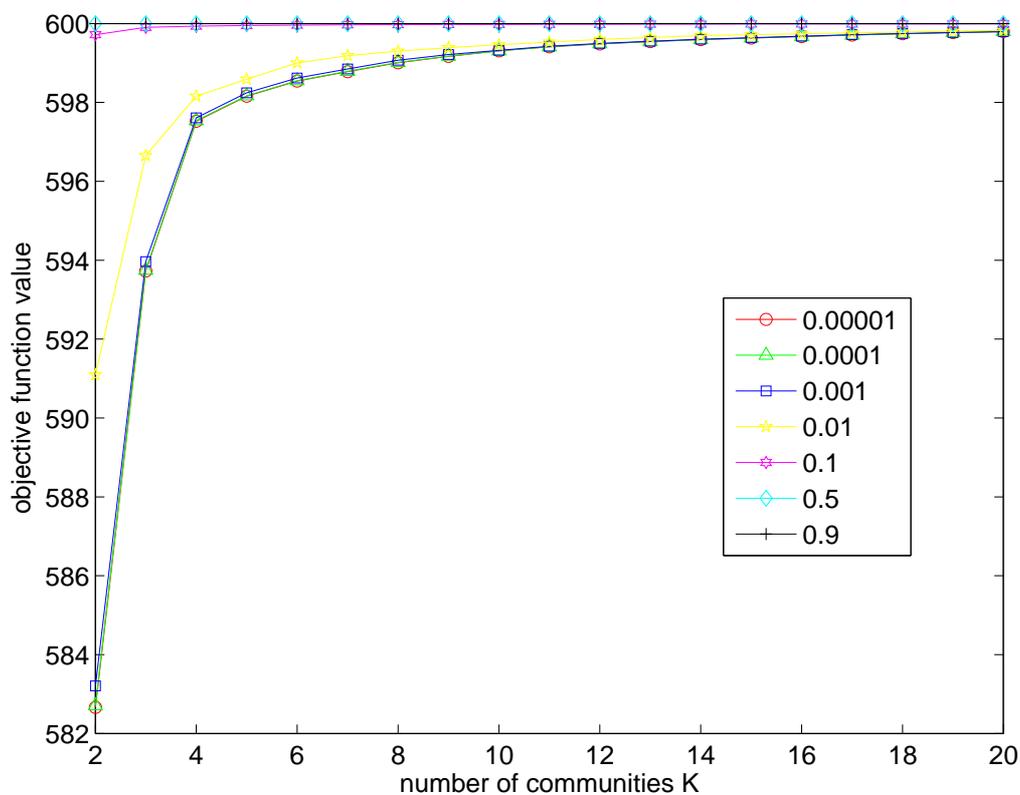


図4.11 Hoseiネットワーク 目的関数値

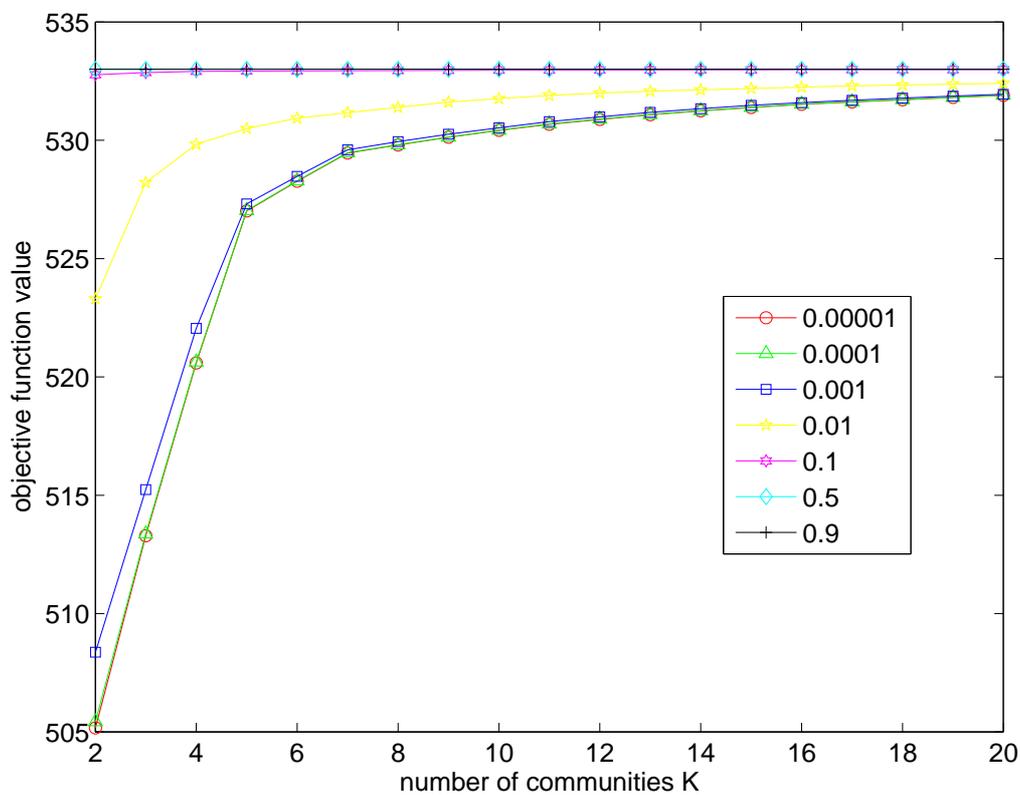


図4.12 Yakuネットワーク 目的関数値

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窺える。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、この章では $\alpha = 0.0001$ での実験結果を掲載した。

つぎに、大域ジャンプ確率を変化させた時の目的関数値の変化を検証する。この章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言うと、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。FCE法やDFCE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 -medians法の目的関数値がある。そこで、大域ジャンプ確率を変化させ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 4.11, 図 4.12）。横軸にコミュニティ数、縦軸に目的関数値、各マーカーは異なる大域ジャンプ確率を表す。

図 4.11 と図 4.12 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わかる。

PageRank を検索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 [15]。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の値は異なる。

4.4 次数不変リンク張替に対する頑健性

各ノードの出次数・入次数を不変でリンクを張替えるノイズをネットワークに与え、機能コミュニティの頑健性を評価する。

4.4.1 リンク張替法

張替確率 $p = 10^{-k}$ ($k = 0, \dots, 3$) で任意のリンクを張替える。

張替ネットワークはオリジナルネットワークのリンクを、各ノードの出次数・入次数を変化させずに、

張替確率 p に従ってリンクをランダムに張り替えることで構築する[?]. 詳細は、下記に示す.

1. ネットワークの有向リンクリスト $L_E = (e_1, \dots, e_{|E|})$ を用意する
 e_1 はリンクの両端のノードからなる順序対 $(L_{from}(1), L_{to}(1))$ であり, 第1要素から第2要素へのリンクを示す.
2. 有向リンクリスト L_E から2つのノードリスト L_{from} と L_{to} を生成する
3. ノードリスト L_{to} の2つの要素を張替確率 p でスワップし, 部分的に順序の変わったノードリスト L'_{to} を得る
4. ノードリスト L_{from} と ノードリスト L'_{to} とを結合し L'_E を得る

得られたリンクリスト $L'_E = ((L_{from}(1), L'_{to}(1)), \dots, (L_{from}(|E|), L'_{to}(|E|))) = (e'_1, \dots, e'_{|E|})$ は張替ネットワークの有向リンクを表す. 但し, リンク張替の際には, 自己リンクと多重リンクにならないようにする. 張替モデルでのリンクのスワップを図示したものが図4.14である. 図4.14では, 赤と青の実線リンクがスワップ対象のリンクで, それぞれ赤と青の点線リンクに張り替えている. スワップをした結果, 真のネットワークである図4.14 (1) と張替ネットワークである図4.14 (3) での, 各ノードの出次数・入次数は変化していない. 図4.15は, 多重リンクのためリンクを張り替えられない例である.

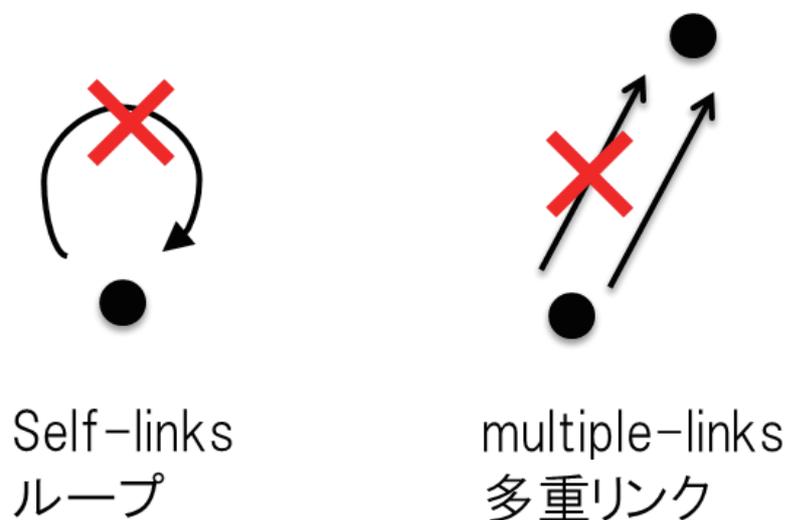


図4.13 禁止事項

張替確率を大きくするにつれ, 推移確率行列のEigen-gapが大きくなり, 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. またランダムネットワークでは, ネットワーク内でのノードの役割に差がなくなり, 変化曲線の分散が小さくなると考えられる. これらの理由が

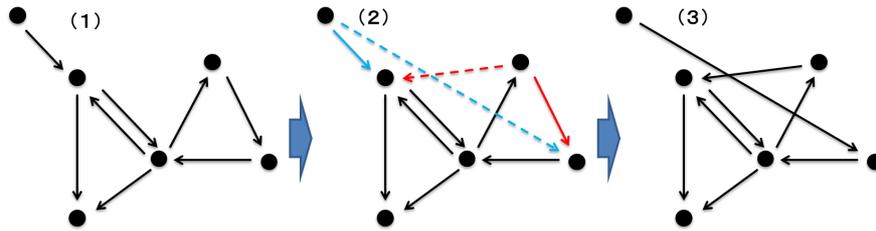


図4.14 リンク張替例

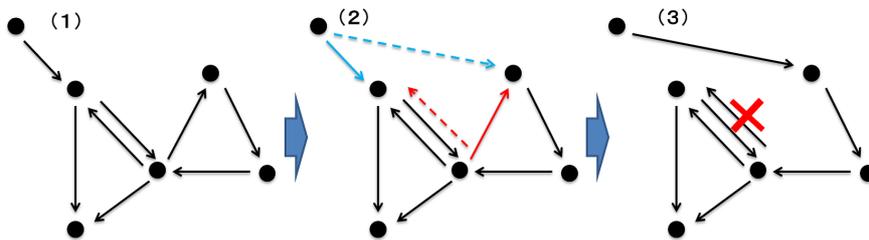


図4.15 リンク張替できない例

ら，機能コミュニティ抽出法の変化曲線は，各ノードの局所的なリンク構造のみを反映した特徴ベクトルを用いることになる．従って，張替前のネットワーク（オリジナルネットワーク）とは異なるコミュニティ抽出結果となると考えられる．

4.4.2 評価結果

Hoseiネットワークに対してリンク張替を行い，可視化した結果を図 4.16に示す．オリジナルネットワークとの比較のために，オリジナルネットワークのノード座標を用い，ノードの色（機能コミュニティ）およびリンクは張替後の結果を示している．また，オリジナルネットワークと各張替ネットワークのコミュニティ抽出結果のConfusion Matrixを図 4.17に示す．縦軸はオリジナルネットワーク，横軸は張替ネットワークでの機能コミュニティIDをそれぞれ意味し，Confusion Matrixの i 行 j 列目は，オリジナルネットワークで i 番目，張替ネットワークで j 番目のコミュニティに属しているノード数を表す．行列の白い要素ほど多くのノードが属していることを表す．

図 4.16(a)や(b)を見ると，張替確率 $p = 10^{-3}, 10^{-2}$ では張替えられるリンク数も少なく，オリジナルネットワークの構造も維持されている．図 4.17(a)や(b)を見ても，コミュニティIDは異なるが多くのノードがオリジナルでの同一コミュニティノードと同一のコミュニティに属している（濃い白い要素が各行に1つある）ことがわかる．

図 4.16(c)や(d)を見ると，張替確率 $p = 10^{-1}, 10^{-0}$ では張替えられるリンク数が多くなり，オリジナルネットワークの構造が崩れ始めている．図 4.17(c)や(d)を見ても，オリジナルでの同一コミュニティ

ノードが分割したり、異なるコミュニティノードが統治をし始め（濃い白い要素が少なくなり、灰色が多くある）、オリジナルネットワークとは異なる機能コミュニティ抽出結果となっていることがわかる。

Confusion Matrixの各行/列のエントロピーにより定量的に評価した結果を図 4.18に示す。エントロピーの値を反転し（評価値）、各行/列の平均値を張替確率ごとでプロットしたものである。各行のエントロピーの平均値が低いほど、オリジナルネットワークでのある機能コミュニティが分割されたことを意味する。各列のエントロピーの平均値が低いほど、オリジナルネットワークでの異なる機能コミュニティが統治されたことを意味する。エントロピーの値は、5回張替ネットワークを生成し、それらの平均値である。

図 4.18(a), (b)より、予想通り張替確率を大きくするにしたがい、オリジナルネットワークでの機能コミュニティが分割・統治され、オリジナルネットワークとは異なるコミュニティが抽出されていることがわかる。

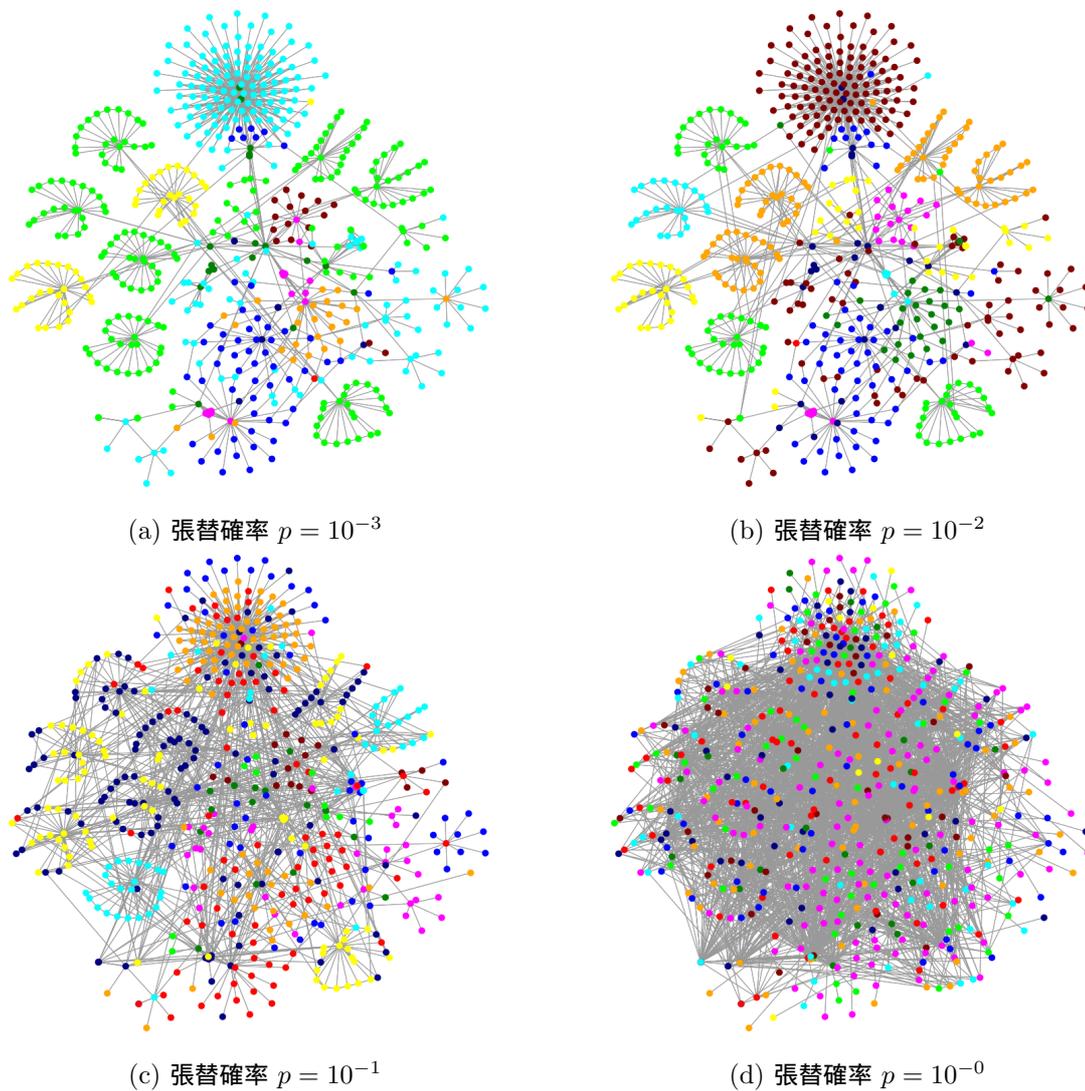
図示はしていないが、Yakuネットワークでも同様の傾向が見られた。

4.5 おわりに

この章では、有向ネットワークを対象に、無向化した場合にはわ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。DFCE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。

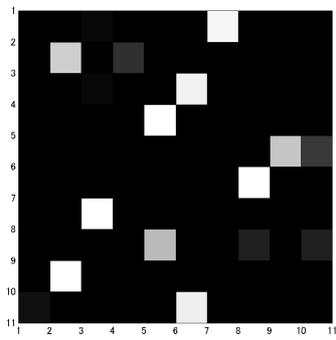
また、有向リンクを有する有向ネットワークでは、出リンクを持たないぶら下がりノードや、複数の強連結成分の存在などにより、PageRankスコアの反復計算時に、ランクシンクなどの問題が発生する場合がある。そのため、 $\alpha = 0$ では有向ネットワークに適用できない。この章では、大域ジャンプ確率を $\alpha \neq 0$ とし、パラメータ α に対する一考察をした。 $\alpha \simeq 1$ とすると、PageRankスコアの収束が速くなるため、特徴ベクトルの実質的な次元数が減少する。得られるコミュニティ数 K も少なくなることが実験から明らかになった。さらに、どのノー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

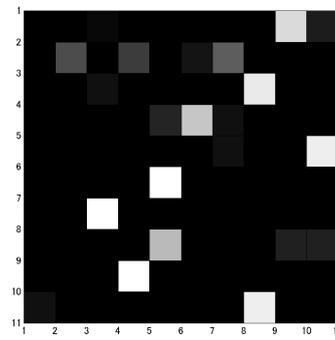
図4.16 張替後Hoseiネットワーク ($K = 10$)

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。

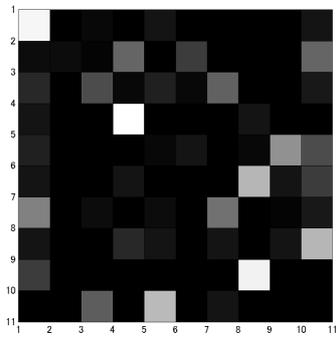
今後は、さらに多様なネットワークでDFCE法の有効性を検証するとともに、大域ジャンプ確率 α やクラスター数 K などの設定方法を詳細に分析していくつもりである。また、どのようなリンク構造なら同一あるいは異なるコミュニティとして抽出できるかを詳細に検証し、機能コミュニティの必要十分性を評価していくつもりである。



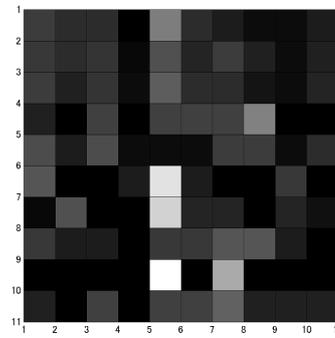
(a) 張替確率 $p = 10^{-3}$



(b) 張替確率 $p = 10^{-2}$

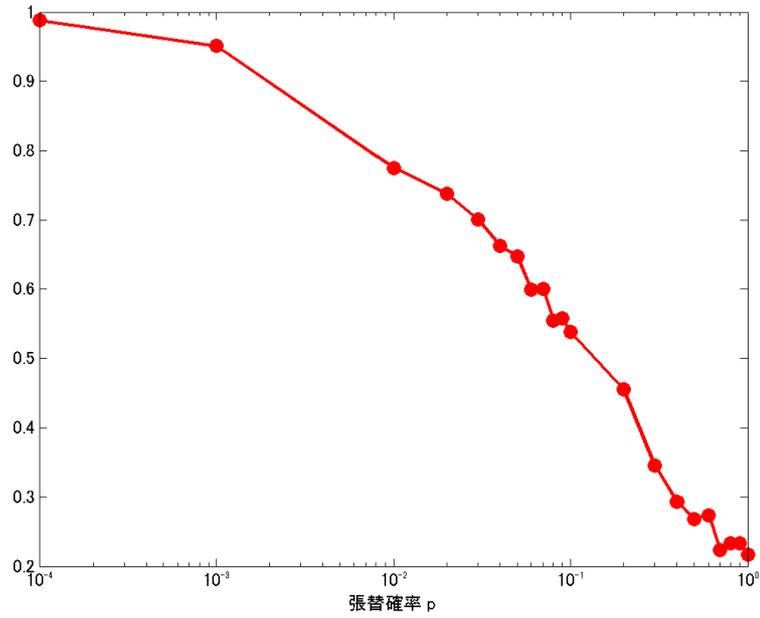


(c) 張替確率 $p = 10^{-1}$

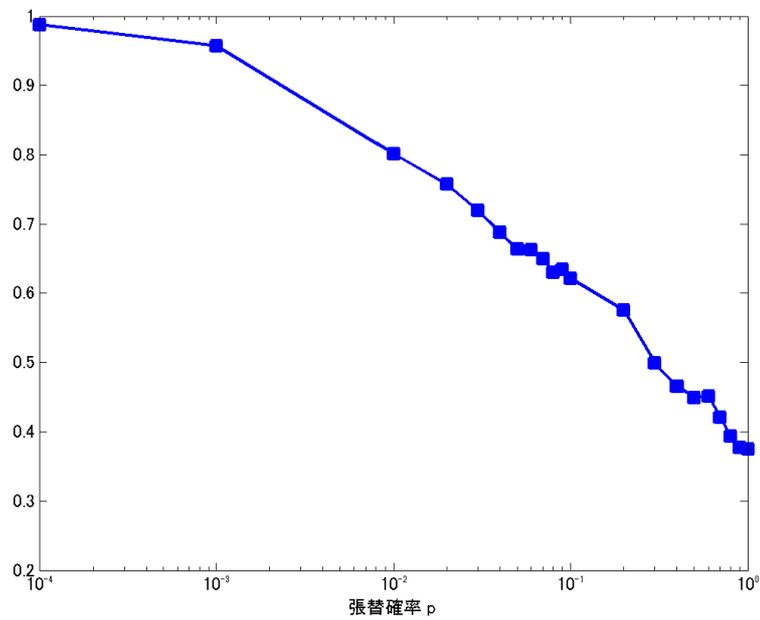


(d) 張替確率 $p = 10^0$

図4.17 Confusion Matrix ($K = 10$)



(a) 行エントロピー



(b) 列エントロピー

図4.18 張替確率とエントロピーの推移 ($K = 10$)

第5章

個々のノードの視点に基づく 機能コミュニティ抽出法：PFCE法

5.1 はじめに

現実社会における人間関係や、TwitterやmixiといったWebサービスのソーシャル・ネットワークに対する関心が高まっている。人間関係に留まらず、Webページのハイパーリンクネットワークや道路網など、あらゆるところで複雑ネットワークが見受けられるようになっている。これらに代表される現実ネットワークにおいて、全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、機能を有していると想定される。各ノードの特徴に基づき、多大なノード群の類似構造を分析したり、重要ノードを抽出するための手法 [8] が提案されている。ネットワーク構造に関しても全体が均等ではなく、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 [17]。

既存のコミュニティ抽出手法として、Newmanらによる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 [19]。この手法は、ノード同士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に分割する方法である。また、ネットワーク上でのノード同士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ための手法も提案されている [25, 26, 27]。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。

また、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た機能コミュニティがある [44, 45]。著者らのFCE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手法である。文献 [44]では、人工ネットワークおよびハイパーリンクネットワークを対象に、Newmanらの方法によるコミュニティ抽出結果と提案手法による機能コミュニティの違いを定性的に評価し、従来のコミュニティ抽出結果とは異なる、妥当な機能コミュニティが抽出できることを

示した。文献 [45]では、局所的なリンク構造の類似性の概念として構造同値や正則同値を取り上げ、それらを近似的に抽出する手法とFCE法の違いを評価した。FCE法は、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手法であると言える。例えば、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、課長のようにそれぞれのノードに階層的地位や役割がある。同様にWebページのハイパーリンクネットワークでも、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。

ただし、ノードの役割・機能は視点を変えることによって変化することがある。例えば、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。

この章ではこれらの点に着目し、個々のノードから主観視した機能に基づくパーソナライズ機能コミュニティ抽出法（PFCE法）を提案する。具体的には、FCE法におけるPageRank計算の初期ベクトルにバイアスを加え、さらにエゴジャンプと呼ぶ機構を導入する。人工ネットワークおよび実ネットワークを用いた評価実験より、選択されたノード（エゴノード）の視点に基づく機能の類似するコミュニティを抽出できることを示す。

5.2 FCE法

この節では、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。

FCE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. 各ノードの機能ベクトルとしてPageRankスコア変化曲線 x_v を構築;
3. 各ノードペアの機能ベクトル 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;

以下に機能ベクトル構築法について詳細を説明する.

5.2.1 機能ベクトル構築法

FCE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 [15]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 $a(u, v) = 1$, それ以外の場合 $a(u, v) = 0$ 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. 繰り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:

$$\mathbf{y}_s^T = \mathbf{y}_{s-1}^T \mathbf{P} \quad (5.1)$$

ここで \mathbf{b}^T はベクトル \mathbf{b} の転置を表わす. ノード u に注目すると,

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} \{y_{s-1}(v) \cdot p(v, u)\} \\ &= \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \end{aligned} \quad (5.2)$$

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 (機能ベクトル) と呼ぶ.

5.3 PFCE法への拡張

通常の機能コミュニティは, すべての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全域的な機能の類似度により抽出される. この章では, ノード個人の視点から見た他のノードの機能・役割の類似度によりコミュニティを抽出する方法を, Personalized PageRank [53] を基にして提案する.

視点の原点となる着目するノードをエゴノードと呼び, エゴノード集合を $R \subset V$ と表す. PageRankスコア計算の初期ベクトルを $\mathbf{y}_0 \in \{0, |R|^{-1}\}^{|V|}$ で定める. 但し,

$$y_0(v) = \begin{cases} \frac{1}{|R|} & (\text{if } v \in R) \\ 0 & (\text{otherwise}) \end{cases}$$

とする．これにより，反復計算の初期段階では，エゴノード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，エゴノードとの関係（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，エゴノードから見た各ノードの機能・役割を表すことができると考えられる．反復を繰り返すに従い，遠くのノードも値を持つことになり，同様にエゴノードとの関係などによりPageRankスコア変化曲線が変動する．

この拡張に加え，PageRankスコア計算にエゴジャンプと呼ぶ機構を提案・導入する．エゴジャンプとは，マルコフ連鎖の過程において，任意のノードからエゴノードへエゴジャンプ確率 β でジャンプするものである．繰り返しステップのパラメータ s を用い，PFCE法でのPageRankスコアベクトル \mathbf{y} の更新式を以下のように定義する：

$$\begin{aligned}\mathbf{y}_s^T &= \mathbf{y}_{s-1}^T ((1-\beta)\mathbf{P} + \beta\mathbf{e}\mathbf{z}^T) \\ &= (1-\beta)\mathbf{y}_{s-1}^T \mathbf{P} + \beta\mathbf{z}^T.\end{aligned}\quad (5.3)$$

ここで $\mathbf{z} \in \{0, |R|^{-1}\}^{|V|}$ である． \mathbf{z} は $z(v) \geq 0$ で $\sum_{v \in V} z(v) = 1$ を満たす．

また，ノード u のスコアに注目すると，

$$\begin{aligned}y_s(u) &= (1-\beta) \sum_{v \in \Gamma(u)} \{y_{s-1}(v) \cdot p(v, u)\} + \beta \cdot z(u) \\ &= (1-\beta) \sum_{v \in \Gamma(u)} \left\{ \frac{y_{s-1}(v)}{|\Gamma(v)|} \right\} + \beta \cdot z(u)\end{aligned}\quad (5.4)$$

で計算される．

通常のFCE法と同様に，反復回数 S 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 $\mathbf{x}_u = (y_1(u), y_2(u), \dots, y_S(u))^T$ 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（機能ベクトル）と呼びノード u の特徴ベクトルとして扱う．

提案法における初期ベクトルの調整および，エゴジャンプの導入により，選択されたエゴノードの近くでは，局所的なリンク構造が強く影響するため細分化される．一方エゴノードの遠くでは，エゴノードからの直接影響が少なく，分類の粒度が粗くなる．

この章で拡張したPageRankスコア変化曲線の計算法は，エゴノードとしてネットワークの全ノードを選択した場合 ($R = V$)，初期ベクトルは全ノードで対等となり，エゴジャンプは確率 β でネットワーク全域にジャンプすることになる（ランダムサーファージャンプ）．したがって，Brinらにより提案されたオリジナルのPageRankと等価となり，自然な拡張になっている．

5.4 評価実験

この節では，現実のWebネットワークデータなどを対象に，提案法を用いてパーソナライズ機能コミュニティを抽出する．いくつかのノードをエゴノードとして選択し，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．

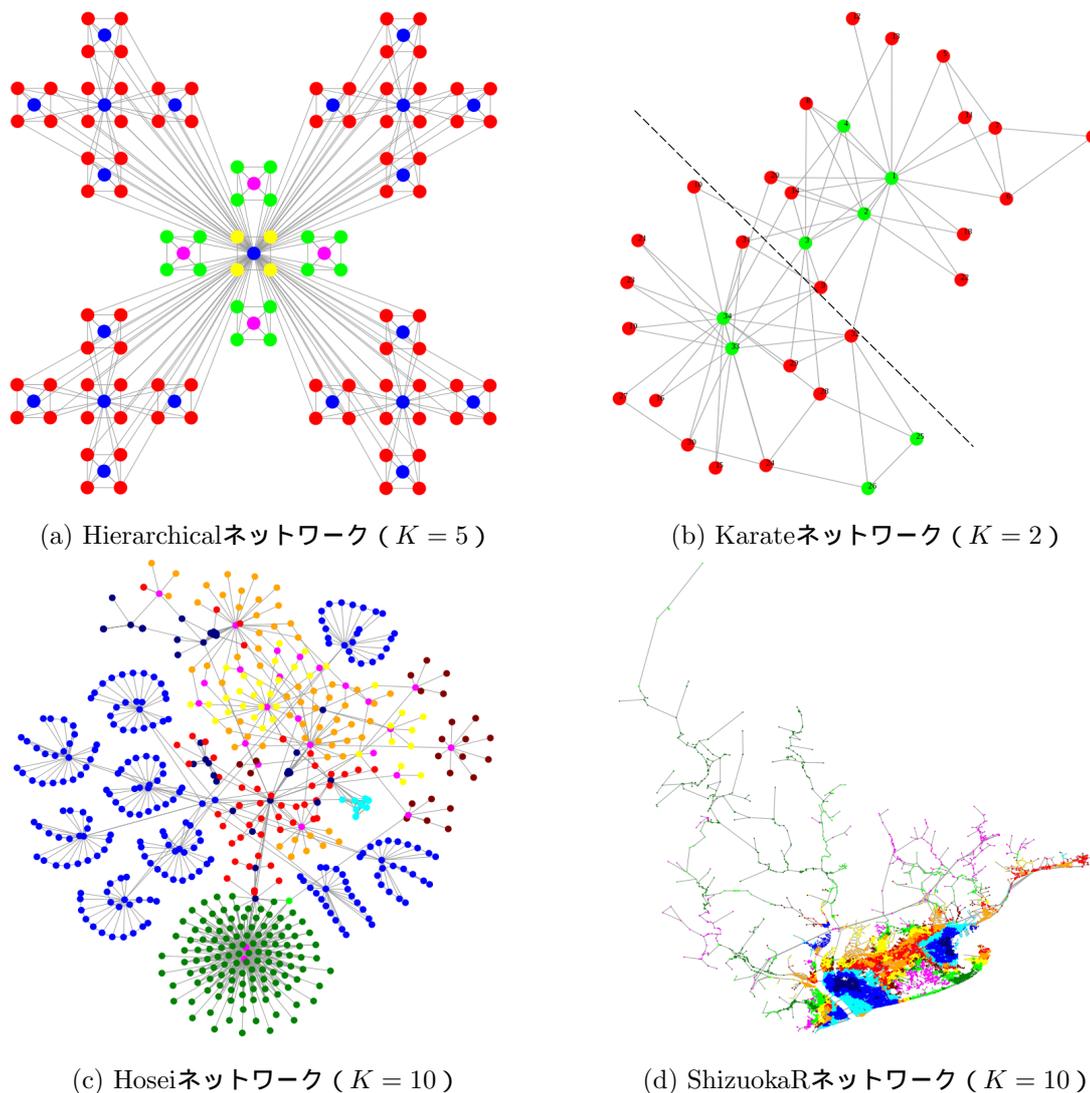


図5.1 機能コミュニティ抽出結果

5.4.1 ネットワークデータ

評価実験では、PFCE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モデルであるHNモデルにより生成した人工ネットワークである [6]。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。この章では、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 [47]。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、インストラクター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。この章では、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。この章では、Hoseiネットワークと呼ぶネットワークを用いる*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 [49] 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。この章では、ShizuokaR ネットワークと呼ぶ。

5.4.2 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 $\|y_s - y_{s-1}\|_{L1} < 10^{-8}$ 、エゴジャンプ確率を $\beta = 0.0001$ 、コミュニティ数 $2 \leq K \leq 10$ とし、提案法によるパーソナライズ機能コミュニティ抽出結果を図 5.2から図 5.5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 の場合でも矛盾のない結果が得られている。この章では、エゴノードを1つ選択した結果を示す。

Hierarchicalネットワークの処理結果

Hierarchicalネットワークの結果を図 5.2(a)および図 5.2(b)にそれぞれ示す。それぞれエゴノード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 5.1(a)に示す。

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 5.1(a)から、通常のFCE法では、階層的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わかる。

ノード番号26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。ノード番号26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ー（桃）、周辺グループのグループリーダー（黄）とそのメンバー（緑）、トップノード（赤）のように、エゴノードの直接関係

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

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2つのコミュニティ（青と水）に分類されている。

ノード番号63は、会社組織でいうと、一般社員に相当するノードである。ノード番号63の視点からの機能コミュニティ抽出結果を見ると、同一グループ内の他の一般社員（緑）とグループリーダー（橙）、一階層上のグループリーダー（桃）、同一階層の他グループのリーダー（茶）と一般社員（青）のように詳細に分類できる。一方、エゴノードから遠くに存在するノード群の多くは1つのコミュニティ（水）として抽出される。

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。

Karateネットワークの処理結果

Karateネットワークの結果を図 5.3(a)と図 5.3(b)にそれぞれ示す。それぞれエゴノード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 5.1(b)に示す。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 [54]。

このネットワークの特徴は、ノード番号1と34のノードが派閥のリーダーであり、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 5.1(b)から、通常のFCE法では、ハブノード（派閥のリーダー）とその他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わかる。

ノード番号1は、前述したように派閥のリーダー的存在である。ノード番号1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）、ハブ的存在のノード（桃）、ハブノードとのみ隣接しているノード（緑）、エゴノードとだけ隣接しているノード（水）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。

ノード番号17は、ネットワークのはずれに位置するノードである。ノード番号17の視点からの機能コミュニティ抽出結果を見ると、エゴノードの隣接ノード（青）とそのグループ（緑）、派閥内の他のノード（赤）、他の派閥のノード（黄）に分類されている。この結果から想定できることとして、このエゴノードが派閥内で端に存在しているため、親しい友人（直接隣接するノード）以外の派閥内のノードは同一機能しか有していないということである。これらの結果を踏まえると、ネットワークの機能と同時に、特にエゴノードから遠い部分で派閥のようなコミュニティ概念も考慮されることが示唆された。

機能コミュニティは、ネットワーク構造から窺い知ることのできるノードの役割（階層性ほか）により分類できる手法である。Karateネットワークのように階層性が明確に表れないネットワークでは、派閥のリーダーやはずれノードなどの役割によりコミュニティが抽出される。それらをエゴノードとした

結果，エゴノードとの関係性や他の派閥のノードとの距離感などにより分類された結果が得られた．

Hoseiネットワークの処理結果

Hoseiネットワークの結果を図 5.4(a)と図 5.4(b)にそれぞれ示す．それぞれエゴノード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 5.1(c)に示す．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 [50]．

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辿れるが，年度間のリンクは存在しない．図 5.1(c)から，通常のFCE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わかる．

ノード番号303は，シラバスページ群の1ページであり，対象大学内のある講義に関するページである．今年度開講講義全てのシラバスに対してリンクをするページからのみリンクされている．ノード番号303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，特にシラバス群は本年度開講シラバス群（茶）と未開講シラバス群（深緑），学科横断受講指定講義のシラバス（黄緑）に細分化されている．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．

ノード番号350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．ノード番号350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，最初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．また，通常の機能コミュニティ抽出結果と同様に，ノードの大域的な機能に基づく分類という全体的な傾向は保持している．

ShizuokaRネットワークの処理結果

ShizuokaRネットワークの結果を図 5.5(a)と図 5.5(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 5.1(d)に示す．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二ヶ所にそれぞれの市の中心部が存在する．図 5.1(d)から，通常のFCE法では，概ね駅周辺（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わかる．

ノード番号6329は、静岡駅周辺の交差点の1つである。ノード番号6329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。

ノード番号21557は、清水駅周辺の交差点の1つである。ノード番号21557の視点からの機能コミュニティ抽出結果を見ると、上記の結果と類似した結果が得られた。特に、両駅から遠く離れた山間部（図中上の方）は色は違うものの配色パターンが酷似していることがみてとれる。近隣地域はエゴノードから同心円状に分類される傾向、遠隔地域は同様のパターンに分類されていることから、通常の機能コミュニティのように、各ノードの大域的な役割と、エゴノードからの距離が反映できていると言える。さらに、どちらの駅をエゴノードとして選択した場合においても、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^{*2}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^{*3}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。

5.4.3 変化曲線によるエゴジャンプ確率に関する考察

提案法のパラメータとして、任意のノードからエゴノードへジャンプする確率であるエゴジャンプ確率 β があげられる。エゴジャンプ確率の取りうる値の範囲は、 $0 \leq \beta \leq 1$ であるが、適切なパーソナライズ機能コミュニティを抽出するため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 P の第1固有値と第2固有値のEigen-gapおよび、エゴジャンプ確率 β の値に依存する。 $\beta \simeq 1$ では、PageRankスコアの収束が速くなり、特徴ベクトルの実質的な次元数が減少する。さらに、エゴノードの影響が支配的になり、エゴノード以外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の抽出が困難になる。図 5.6に、Karateネットワーク、Hoseiネットワークに対して、横軸にエゴジャンプ確率、縦軸に平均反復回数をプロットした。平均反復回数は、収束判定基準を $\|y_s - y_{s-1}\|_{L1} < 10^{-8}$ とし、エゴノードをランダムに選び、PageRankスコアが収束するまでのステップ数の10回の平均である。図 5.6から、 10^{-4} あたりから平均反復回数が安定し、充実した特徴ベクトルを得られることがわかる。

図 5.7に、Karateネットワーク、Hoseiネットワークに対して、横軸にエゴジャンプ確率、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 5.7から、Karateネットワー

^{*2} http://www.city.shizuoka.jp/deps/tosikeikaku/koshin_toshi_kei_toshimasu_shizuoka_2syo_kakudai_2.3_b.htm

^{*3} 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、殆どの道路は他と立体交差しない。

クでは 10^{-1} , Hoseiネットワークでは 10^{-4} あたりから標準偏差が安定し, ノード固有の多様な特徴ベクトルを得られることがわかる. そのためこの章では $\beta = 0.0001$ の実験結果を掲載した.

実際のKarateネットワークにおける代表ノードの変化曲線(正規化後)を図5.8に示す. 横軸は反復のステップ数 s , 縦軸は各ステップにおけるPageRankスコアを意味する. 代表ノードとは, 提案法における K -medianクラスタリングで代表ベクトルとして選ばれたノードを意味する. 図5.8から, エゴジャンプ確率 $\beta \simeq 0$ ならば, 上述したようにPageRankスコアの収束が比較的遅く, 変化曲線はノード固有の多様な形状をしていることがわかる. 一方, エゴジャンプ確率 $\beta \simeq 1$ ならば, 上述したようにPageRankスコアの収束が速く変化曲線の変動が少ないことが見て取れる, エゴノード以外は初期値が0で, エゴノードからの距離に従いPageRankスコアを親ノードから授かるため, 曲線間の違いとしては収束するまでのわずかな期間でありそれはエゴノードからの距離を意味している. したがって, パーソナライズ機能コミュニティにおいてエゴジャンプ確率 $\beta \simeq 1$ ならば, エゴノードからの距離が支配的になることが示唆される.

5.5 おわりに

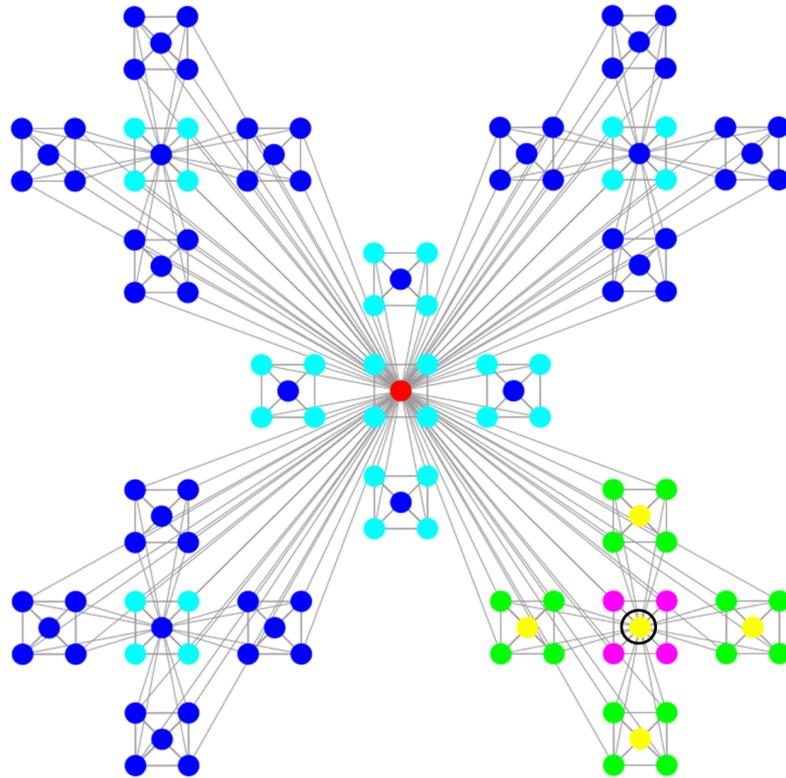
本研究では, 従来とは異なり, 各ノードの視点に基づくコミュニティの概念としてパーソナライズ機能コミュニティおよび, その抽出手法を提案した. パーソナライズ機能コミュニティは, エゴノードの視点からの他のノードの機能・役割が類似するノードをコミュニティとしたもので, コミュニティ抽出では新しいパラダイムであるといえる. 提案法は, 各ノードの機能を表す特徴ベクトルであるPageRankスコア変化曲線の計算において, Personalized PageRank に基づき, 初期ベクトルのエゴノードに対応する要素を1, その他のノードの対応する要素を0とする調整を加えた(和が1になるように正規化する). さらに, 任意のノードから確率 β でエゴノードにジャンプする, エゴジャンプという機構を導入した.

人工ネットワーク, ソーシャル・ネットワーク, ハイパーリンクネットワーク, 道路網といった構造の異なる複数種のネットワークを用いた評価実験により, エゴノードの視点から見て同質な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. すなわち, 客観的に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. しかし, エゴノードからの距離で分類しているわけではない. また通常の機能コミュニティとの比較により, エゴノード近くを細分化し, 遠くを大まかに分類する結果が得られた. これは, 自身の周辺は自身との関係が強く, 直接的に影響するため, 分類の粒度が細くなるが, 自身から離れた位置に存在するものは関係が弱いいため, 分類の粒度が粗くなるという直観に対応している.

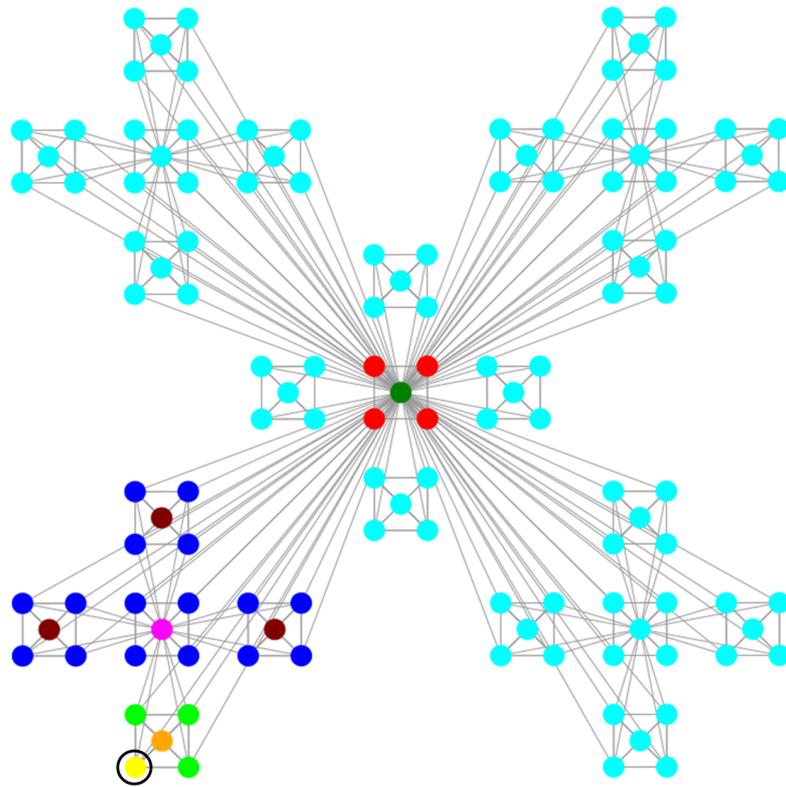
本手法の結果は, ドライバーがカーナビなどを閲覧するに際して, ドライバーのいる交差点付近は興味があるためその情報を細分化し, 遠方の地域は比較的興味が薄いいため細分化する必要がない, といっ

た直観に対応した応用が考えられる。

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。また、エゴノードとして複数ノードを選択した場合での検証やエゴジャンプ確率の有用性を深く追求していくつもりである。

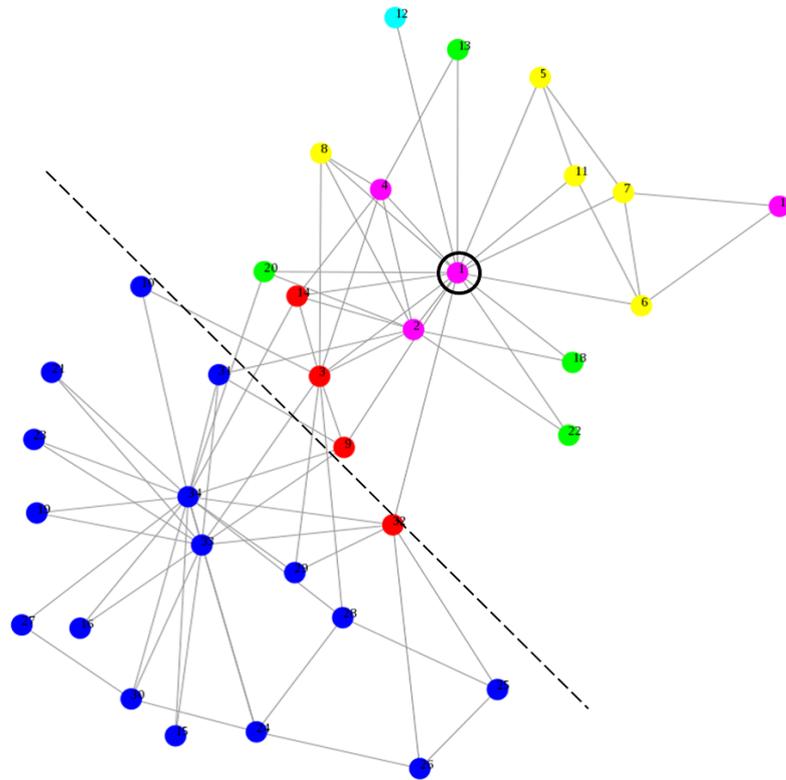


(a) エゴノード 26 グループリーダー ($K = 6$)

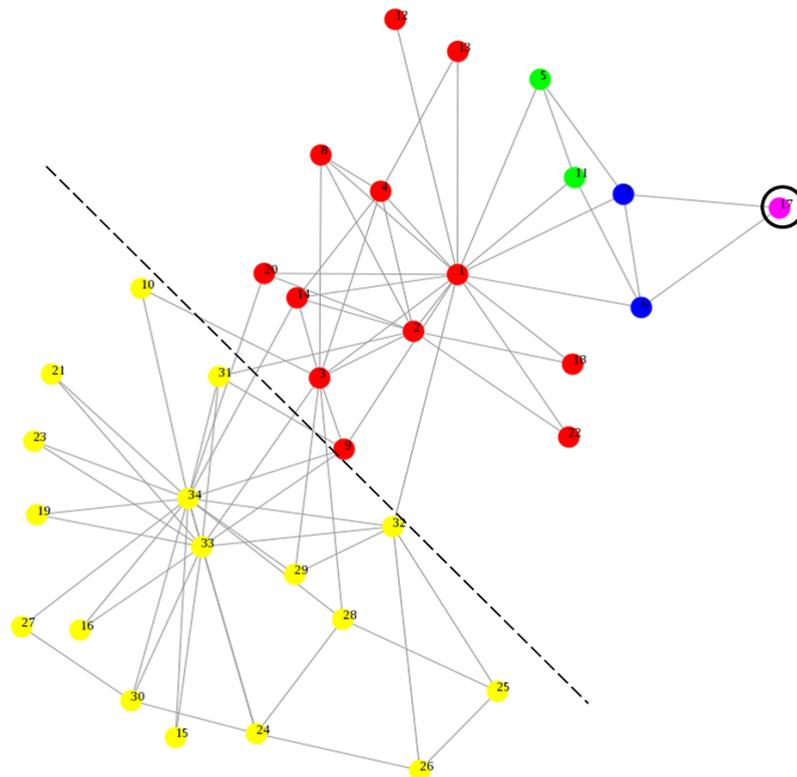


(b) エゴノード 63 最下層ノード ($K = 9$)

図5.2 Hierarchicalネットワーク パーソナライズ機能コミュニティ抽出結果

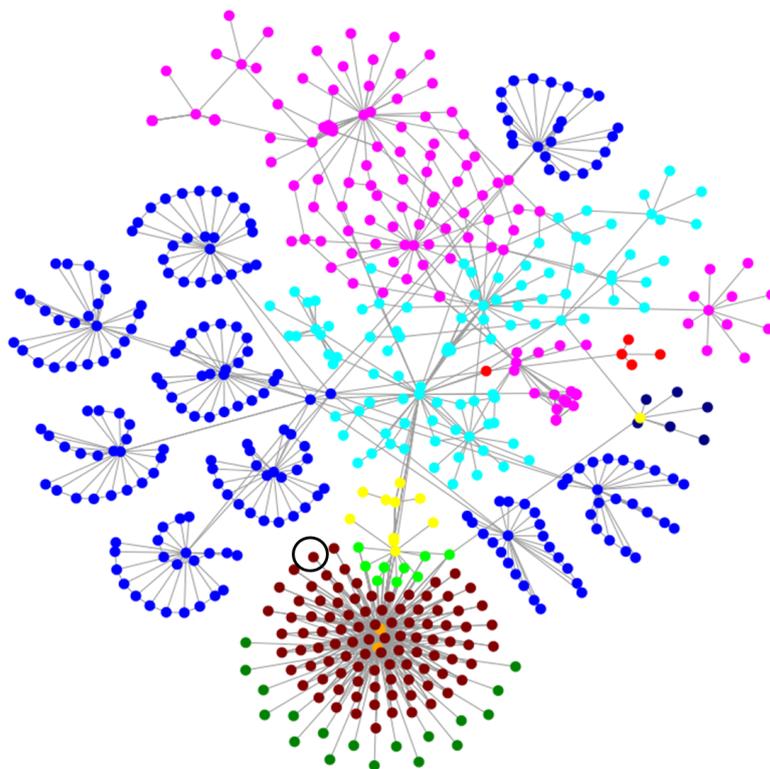


(a) エゴノード 1 派閥のリーダー ($K = 6$)

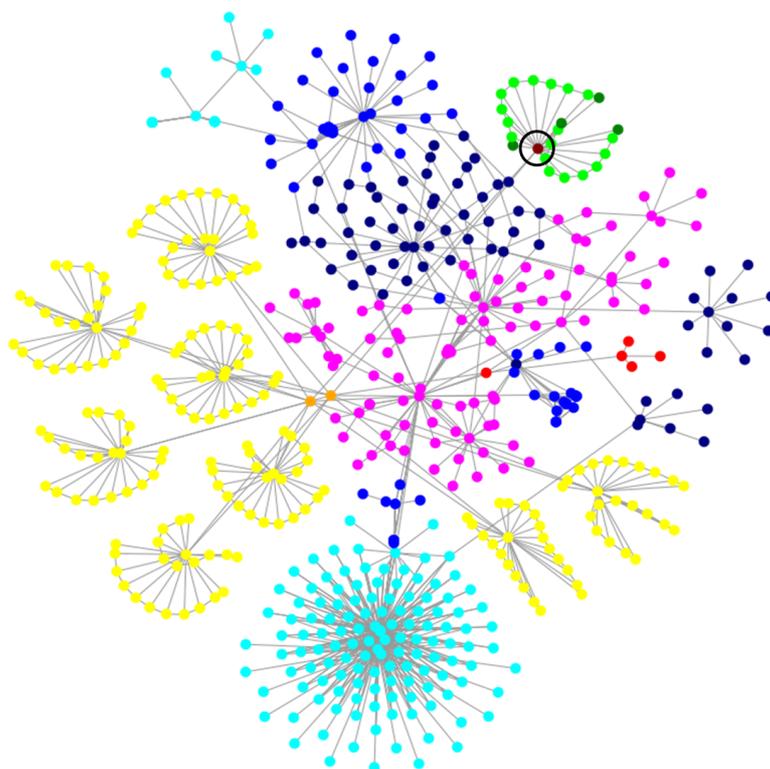


(b) エゴノード 17 はずれノード ($K = 5$)

図5.3 Karateネットワーク パーソナライズ機能コミュニティ抽出結果

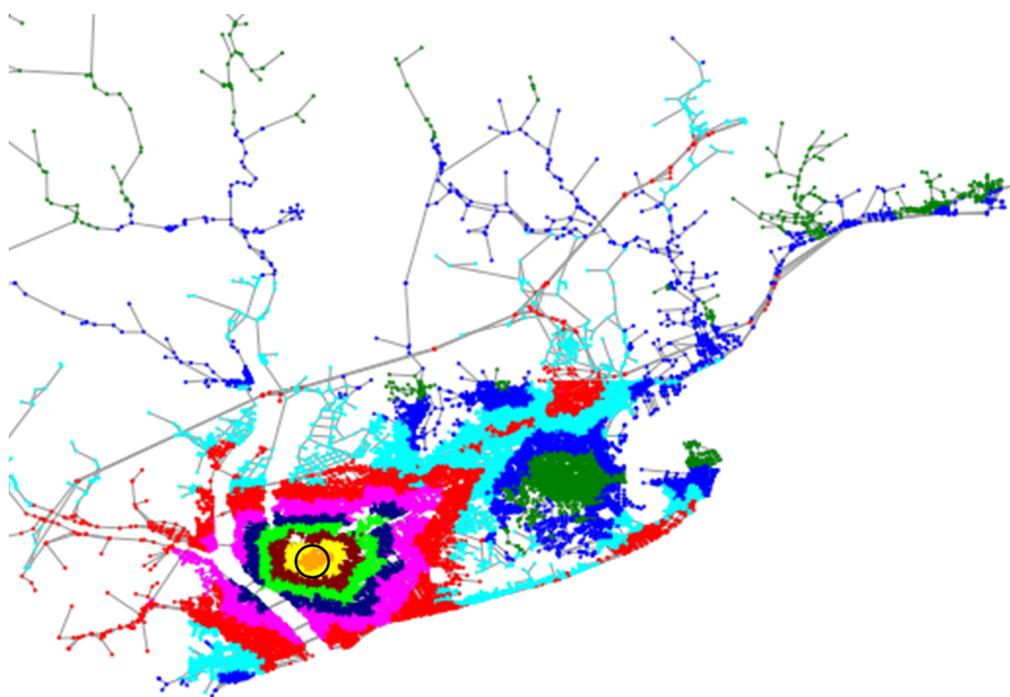


(a) エゴノード 303 シラバス ($K = 10$)

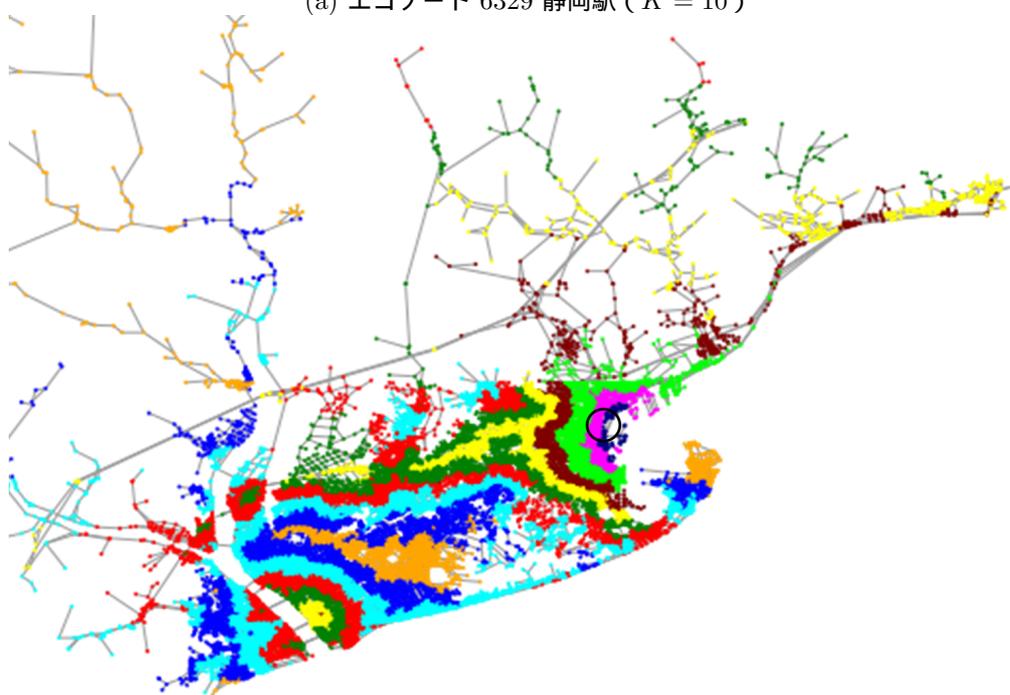


(b) エゴノード 350 教員成果報告ページ ($K = 10$)

図5.4 Hoseiネットワーク パーソナライズ機能コミュニティ抽出結果



(a) エゴノード 6329 静岡駅 ($K = 10$)



(b) エゴノード 21557 清水駅 ($K = 10$)

図5.5 ShizuokaRネットワーク パーソナライズ機能コミュニティ抽出結果

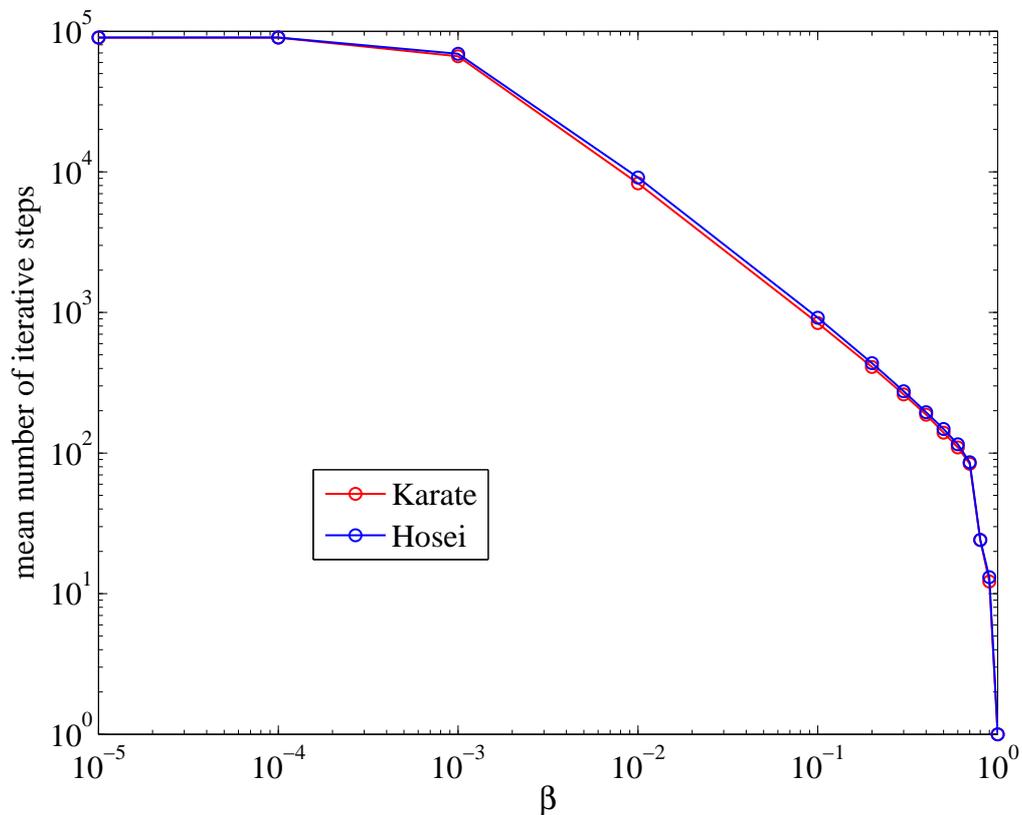


図5.6 エゴジャンプ確率と変化曲線の次元

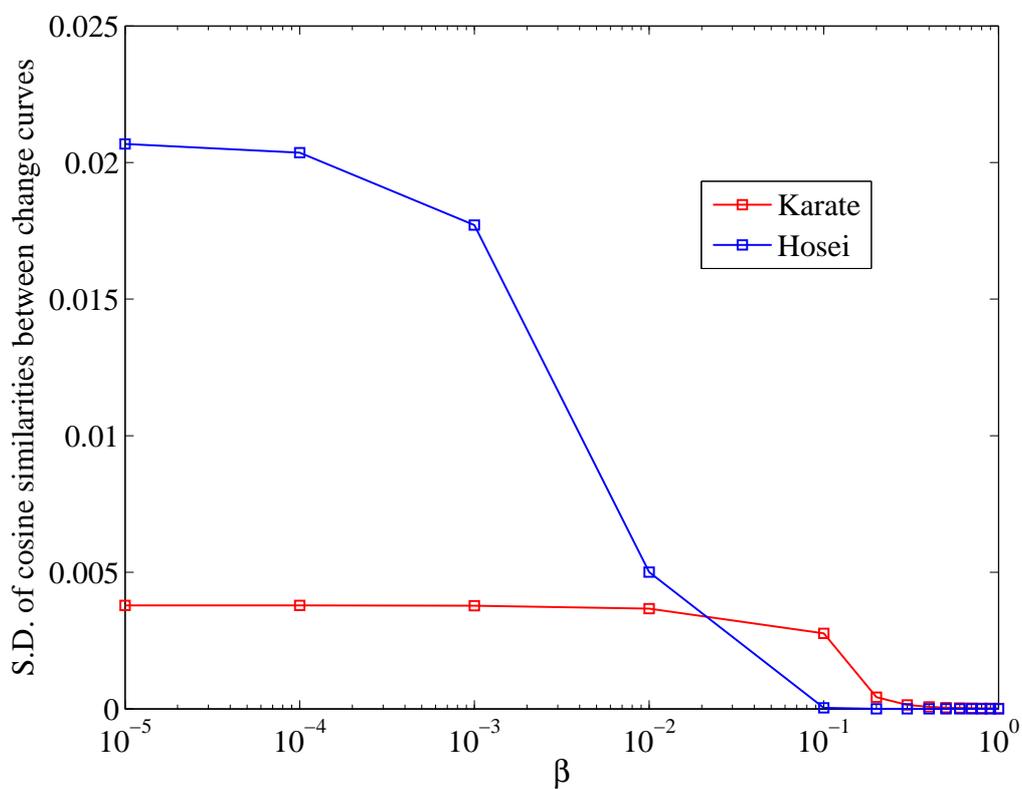
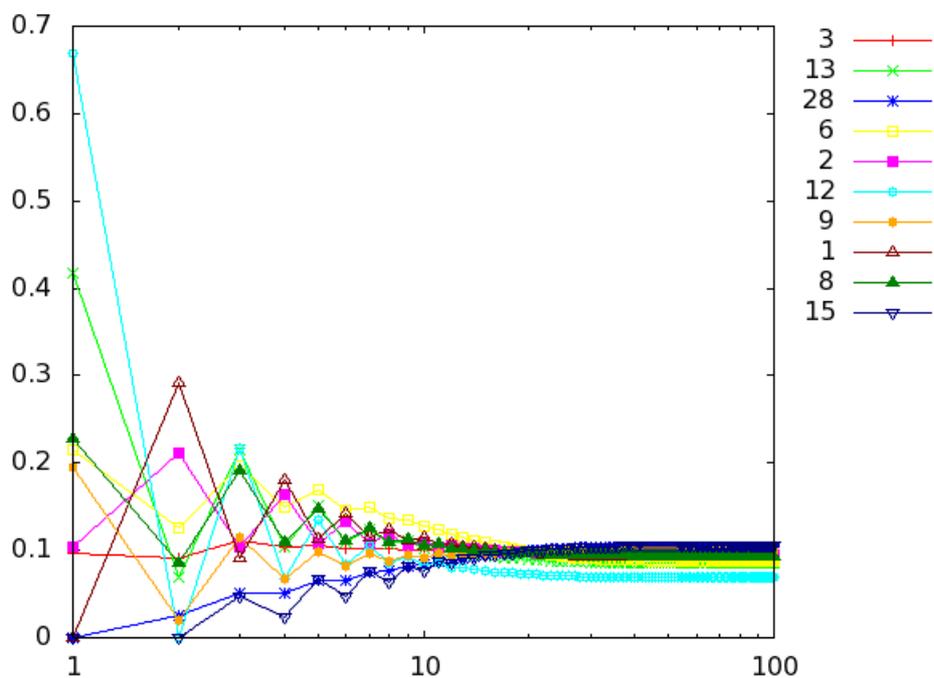
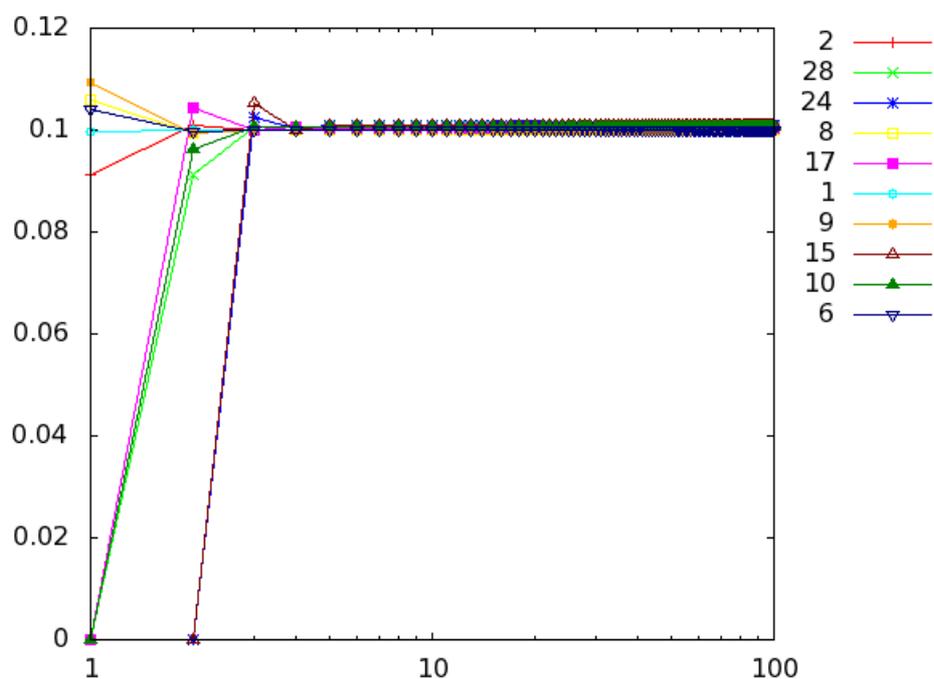


図5.7 エゴジャンプ確率と変化曲線の多様性



(a) $\beta = 0.0001$



(b) $\beta = 0.9$

図5.8 Karateネットワーク 代表ノードの変化曲線

第6章

情報拡散モデルに基づく 機能コミュニティ抽出法：IDFCE法

6.1 はじめに

SNSやレビューサイト内でのユーザ間のつながりなど、複雑ネットワークがあらゆる場面で見受けられるようになり、それらを対象とした分析が盛んに行われている。また、近年ではソーシャルグラフ上でのクチコミ（情報拡散）を利用した情報推薦や広告などの技術が多くみられる。このように、身近に存在するネットワーク上での情報拡散現象を分析・モデル化・予測することで、Webマーケティングなどへの貢献が期待できる。

ネットワーク上での情報拡散過程において、発信された情報をすぐに受信できるノードもいれば、拡散過程の終わりまで情報を受信できないノードもいる。さらに、一般に情報拡散の経路は複数考えられ、多くノードとつながるノードは情報が集まる可能性も高くなる。このように、ネットワーク構造などの様々な要因が重なり合い、情報受信の期待値は各ノードにより異なる。情報受信の期待値が類似するノードは、ネットワーク内で類似の機能・役割を有すると考えられる [34]。

類似機能を有するノードを抽出する手法として、FCE法がある [44, 45]。この手法は、ネットワーク上でのランダムウォークにおける各タイムステップの期待値の変化の類似性によりノードを分類する。各タイムステップでの期待値変化が類似するノードはネットワーク内で類似の立場にあると仮定し、相対的位置や階層的地位などネットワーク内での大域的機能によりノードを分類する。

しかし、現実のソーシャルネットワークを対象とする場合、ランダムウォークモデルに基づき抽出された機能コミュニティは、抽出されたノードがもつさまざまな属性を考慮しないと具体的な機能性が判明しないことから、その意味付けが困難な場合がある。この章では、情報拡散モデルに基づき、情報受信の期待値変化の類似性によりノードを分類する手法（IDFCE法）を提案する。複数のネットワークを用いた評価実験より、FCE法では区別が困難なノードを抽出可能なことなどを示す。

6.2 情報拡散機能コミュニティ抽出法：IDFCE法

代表的な情報拡散モデルを用い、情報受信の期待値変化の類似性に基づきノードを分類するIDFCE法について説明する。

IDFCE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情報拡散機能コミュニティを抽出する。

1. 情報拡散モデルにより、各タイムステップでの情報受信期待値ベクトル $\{y_1, \dots, y_S\}$ を計算;
2. 各ノードの機能ベクトルとして期待値変化ベクトル x_v を構築;
3. 各ノードペアの機能ベクトル x_u と x_v のコサイン類似度 $\rho(u, v)$ を計算;
4. K -medians法により全ノードを K 個のグループに分割;
5. 情報拡散機能コミュニティ $\{C_1, \dots, C_K\}$ を出力;

以下に、アルゴリズムの主要技術に関する詳細を説明する。

6.2.1 情報拡散モデル

情報拡散過程において、情報を受信した状態をアクティブな状態、それ以外の状態を非アクティブな状態と定義する、この章では離散同期型かつSIR型のモデルを用いる。すなわち、情報拡散過程は、離散時刻 $t \geq 0$ で展開し、一度アクティブになったノードは二度と非アクティブにはならない。IDFCE法で用いる代表的な情報拡散モデルである、独立カスケード (Independent Cascade, 以下ICモデル) モデルと線形閾値 (Linear Threshold, 以下LTモデル) モデルについて説明する [55, 56]。

ICモデルでは、各リンク (u, v) に対して、実数値 $p_{u,v} \in [0, 1]$ を前もって指定する。ここで $p_{u,v}$ はリンク (u, v) を通しての拡散確率である。ICモデルの情報拡散過程は、アクティブノードの初期集合 A が与えられたとき、次のように進んでいく。ノード u が時刻 s でアクティブになったとき、 u は非アクティブな子ノード v をアクティブにする唯一の機会が与えられ、その試行は確率 $p_{u,v}$ で成功する。成功した場合、ノード v は時刻 $s + 1$ でアクティブになる。この成功・失敗に関わらず、ノード u がノード v をアクティブにする機会はこの時のみである。ネットワーク中で試行対象となるノードがなくなったとき、情報拡散過程は終了する (図 6.1参照)。

LTモデルでは、任意のノード $v \in V$ に対して、その親ノード u からの重み $w_{u,v} (> 0)$ を、 $\sum_{u \in \Gamma(v)} w_{u,v} \leq 1$ となるように、前もって指定する。各ノードの v の閾値 θ_v を区間 $[0, 1]$ から一様ランダムに設定する。LTモデルの情報拡散過程は、アクティブノードの初期集合 A が与えられたとき、次のように決定論的に進んでいく。時刻 s で非アクティブなノード v は、時刻 s でアクティブな

親ノード u から重み $w_{u,v}$ の影響を受ける．アクティブな親ノードからの重みの和が，自身の閾値を超えたときノード v はアクティブになる．すなわち，時刻 s でアクティブなノード v の親ノード集合を $\Gamma_t(v)$ とすると， $\sum_{u \in \Gamma_t(v)} w_{u,v} \geq \theta_v$ であれば，ノード v は時刻 $s+1$ でアクティブになる．ネットワーク中で試行対象となるノードがなくなったとき，情報拡散過程は終了する（図 6.2参照）．

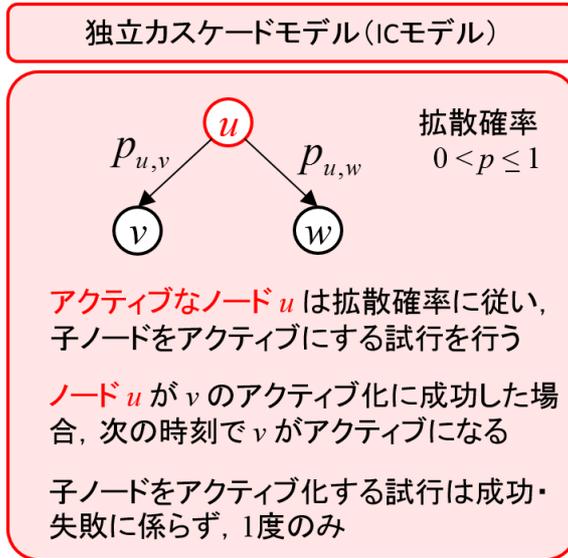


図6.1 ICモデル

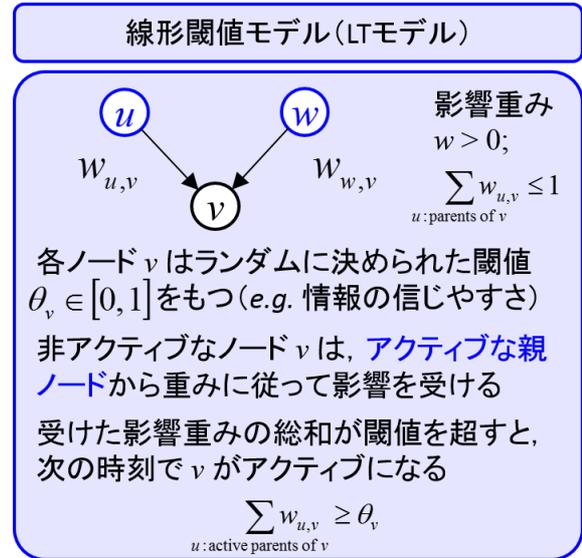


図6.2 LTモデル

6.2.2 ボンドパーコレーション

上述したICモデルとLTモデルによるシミュレーションを，ボンドパーコレーションにより効率的に実行する [57]．ネットワーク G 上でのボンドパーコレーションプロセスは， G の各リンクが任意の確率でランダムに「有効」か「無効」かを割り当てられる．形式的には，リンク数次元 $|E|$ のベクトル r の集合 R_G を以下のように定める．

$$R_G = \left\{ r = (r_{u,v})_{(u,v) \in E} \in \{0, 1\}^{|E|} \right\} \tag{6.1}$$

各リンク $(u, v) \in E$ はランダムベクトルの要素が $r_{u,v} = 1$ で「有効」， $r_{u,v} = 0$ で「無効」を割り当てられる．1つのランダムベクトル r により「有効」と割り当てられたリンク集合を E_r と表記し，「無効」リンクを除いたグラフをBPグラフ $G_r = (V, E_r)$ と呼ぶことにする．

情報拡散の文脈でいえば，シミュレーションごとに確率的に生成されるランダムベクトル r が異なるため，BPグラフ G_r が異なる．しかし，1回のシミュレーションでは1つのBPグラフを用いることができ，決定論的な情報拡散現象として扱うことができる．換言すると，1つのBPグラフにおけるリンクは，情報拡散によって情報が伝わったリンク，すなわち，有効なリンクを意味するため，1つのBPグラフは情報が伝わったリンクのみのグラフを意味する．

6.2.3 情報受信期待値変化ベクトルの計算

上述した情報拡散モデルを用いたシミュレーションにより、各離散時刻で各ノード v がアクティブになる期待値を計算する。情報源ノードを u とした情報拡散過程の時刻 s でアクティブになったノード集合を $A_s(u)$ としたとき、

$$f_s^u(v) = \begin{cases} 1 & \text{if } v \in A_s(u) \\ 0 & \text{otherwise} \end{cases}$$

のように f_s^u を構築する。そして、各ノードを情報源とした際の期待値を計算する：

$$y_s(v) = \frac{1}{|V|} \sum_{u \in V} f_s^u(v). \tag{6.2}$$

ここで、これらを要素とするベクトル y_s を時刻 s での情報受信期待値ベクトルと呼ぶ。

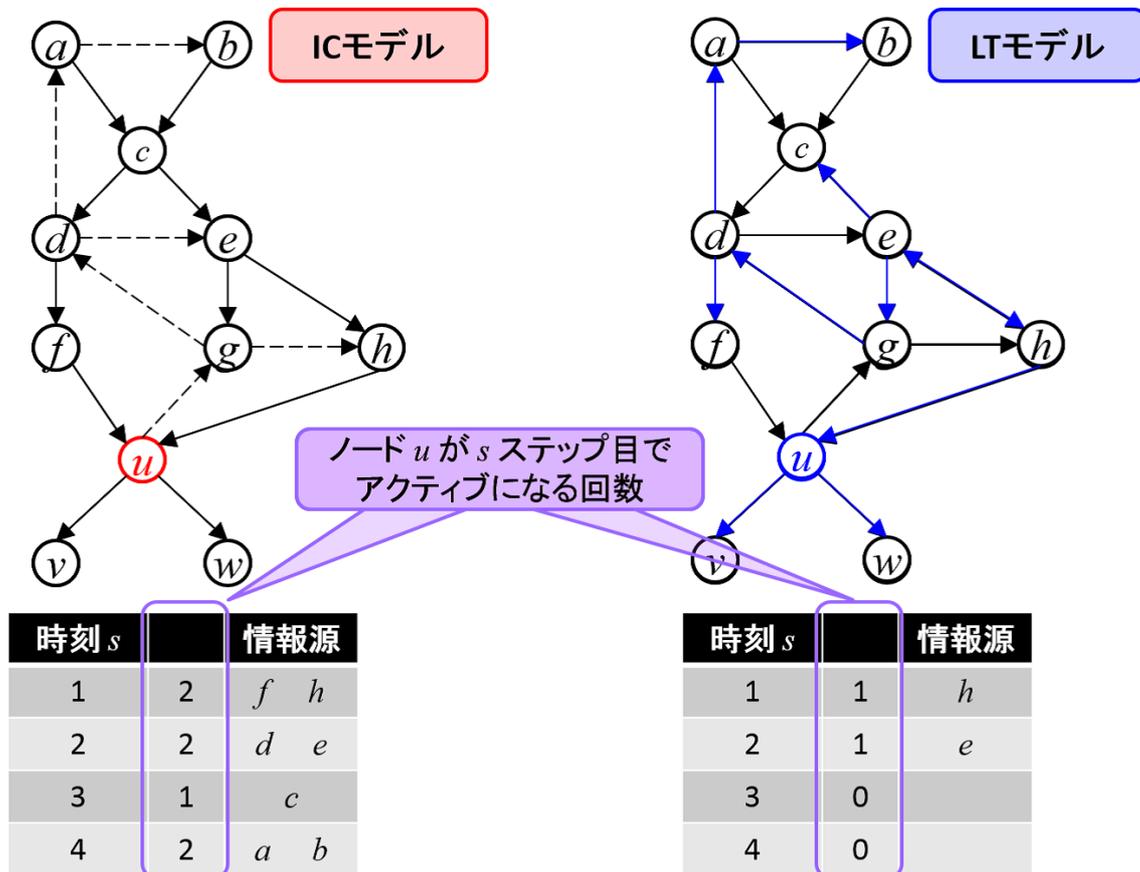


図6.3 1回のシミュレーションでのBPグラフと情報受信期待値ベクトル

図 6.3において、すべての線はネットワークのリンク、実線リンクはボンドパーコレーションによる有効リンク、点線リンクは無効リンクを意味する。1回のシミュレーションにおいて、時刻 s での情報受信期待値は、BPグラフにおける s ステップ離れた親ノードの数に等しくなる（図 6.3参照）。

シミュレーションを M 回実施した場合、各シミュレーションで得られる情報受信期待値ベクトルを平均して期待値を計算する。

次に、各ノードの機能を表す期待値変化ベクトルを以下のように構築する。ノード v の機能ベクトルである期待値変化ベクトルは、

$$\mathbf{x}_v = \{y_1(v), \dots, y_S(v)\} \quad (6.3)$$

と定義する。各ノードの期待値変化ベクトル間の類似度に基づき、ノードをクラスタリングする。

6.3 評価実験

IDFCE法により抽出されたコミュニティの性質を評価するために、前述したランダムウォークに基づく機能コミュニティ抽出法 (FCE法) と比較する。以下ICモデル、LTモデルに基づく機能コミュニティ抽出法を、それぞれICFC法、LTFC法と呼ぶ。

6.3.1 ネットワークデータ

1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。この章ではHoseiネットワークと呼ぶ^{*1}。

2つ目のネットワークは、Hoseiネットワークと同様に、大学のウェブサイト内のページを2012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ものである。この章ではYamaguchiネットワークと呼ぶ^{*2}。

3つ目のネットワークは、ブログのトラックバックネットワークである。2005年5月に“goo”^{*3}というサイトの「JR福知山線脱線事故」というテーマからトラックバックを10段辿ることにより収集し無向化した。この章ではTrackbackネットワークと呼ぶ。

6.3.2 実験設定

IDFCE法において、シミュレーション回数を $M = 10,000$ 、任意のリンク (u, v) における拡散確率を各ネットワークの平均次数の逆数 $p_{u,v} = p \simeq 0.5 \cdot |V|/|E|$ と設定する。各ノードが平均して1ノード程度をアクティブ化できるように設定した。コミュニティ数 K を $2 \leq K \leq 10$ の範囲で最も適切な値を選んだ時の抽出結果を図 6.4から図 6.6に示す。なお他の K の場合でもこの章の主張と大きな矛盾のない結果が得られている。可視化にはノード間の隣接関係が視認できるように、クロスエントロピー

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

^{*2} 山口大学 <http://www.yamaguchi-u.ac.jp/>

^{*3} <http://blog.goo.ne.jp/usertheme/>

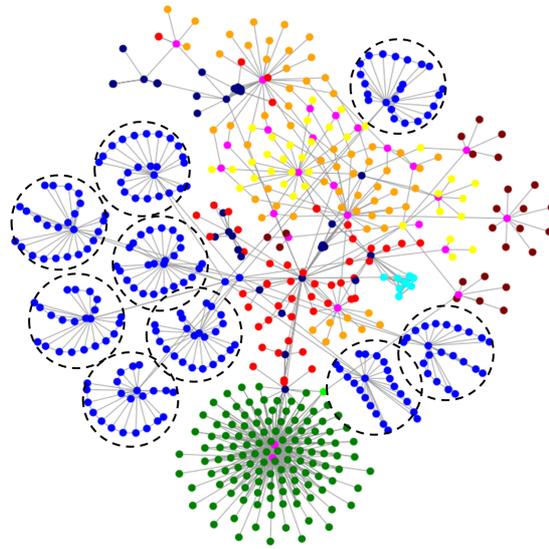
法を用いて[50]，コミュニティごとにノードを異なる色で着色した．

6.3.3 実験結果

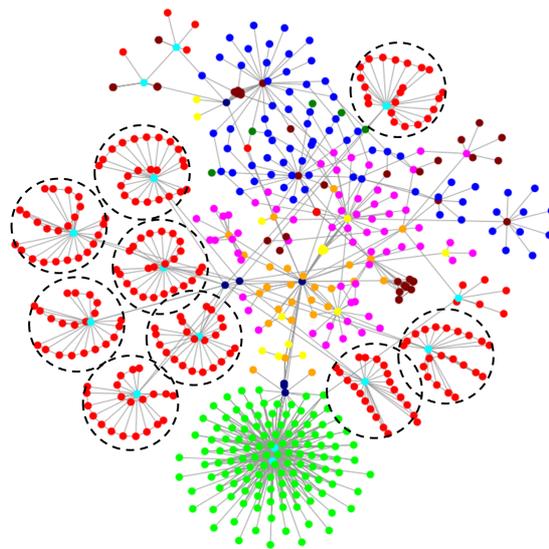
図 6.4はHoseiネットワークの処理結果である．FCE法の結果を見ると，図中下部のシラバスページ群や，点線で囲っている各年度の成果報告ページなど，同一の機能を有するノードが同一のコミュニティとして抽出されていることがわかる．ICFC法，LTFC法の結果を見ると，全体の傾向はFCE法の結果と類似している．しかし，点線で囲っている各年度の成果報告ページに関しては，FCE法では中心のインデックスページとその周辺の各教員の成果ページの区別ができていないが，ICFC法，LTFC法では情報拡散の点で重要ノードとなるハブノードとその他のノードを区別できている．ランダムウォークモデルにおいては，共通隣接ノードを有するクラスタ（三角形）を構成するノードを識別できないため [58]，インデックスページと連続する2つの教員成果ページが作り出すクラスタにおいて中心となるハブノード（インデックスページ）が同一のコミュニティと判定されると考えられる．情報拡散モデルにおいて各ノードがアクティブになるタイミングは，多くのノードへ情報を拡散させるハブノードからの距離が大きな要因であり，クラスタ係数は大きな影響はないと考えられる．

図 6.5はYamaguchiネットワークの処理結果である．FCE法の結果を見ると，中心部分のプレスリリースページや，点線で丸く囲っている月報の広報ページ，点線で四角く囲っている大学歴史紹介ページなど，同一の機能を有するノードが同一のコミュニティとして抽出されている．ICFC法，LTFC法の結果を見ると，全体の傾向はFCE法の結果と類似している．しかし，点線で丸く囲っている月報広報ページに関しては，FCE法では中心のインデックスページとその周辺のコンテンツページの区別ができていないが，ICFC法，LTFC法では情報拡散の点で重要ノードとなるハブノードとその他のノードを区別できている．一方，同じハブノードでもクラスタを形成していない四角く囲っている大学歴史紹介ページは，どの手法でもハブノードとリーフノードを識別できている．

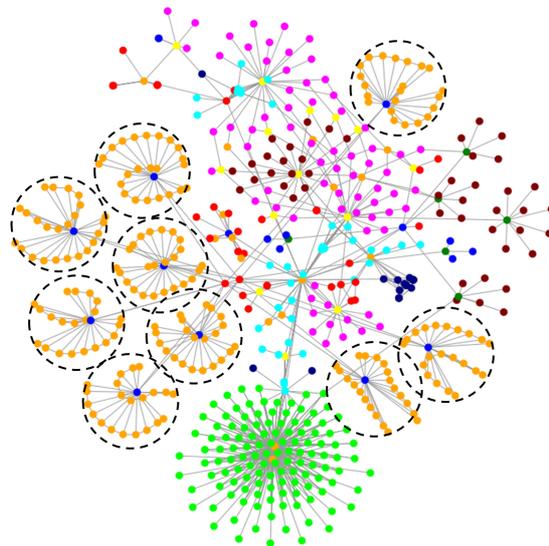
図 6.6はTrackbackネットワークの処理結果である．可視化に用いたクロスエントロピー法では，密に結合するコアな部分は可視化結果のより中心に集まるようにノードの座標を決定する．FCE法の結果を見ると，ネットワーク内のコアな部分を抽出できているが，近傍に位置するノードを同一のクラスタに分類する傾向がある．すなわち，比較的密につながるノード集合（いわゆるCNM的コミュニティ）内のノードを同一のコミュニティとして抽出している．これは上記のクラスタの議論と同様に，ランダムウォークモデルにおいては，クラスタを構成するノードの識別が困難であることが起因していると考えられる．ICFC法とLTFC法の結果を見ると，可視化結果で随所に見られる各ノード集合において，コアな部分とその周辺部分，さらに末端部分というように，コアなノード群からの距離や，近傍に位置するノードの数に基づいて層状にクラスタリングされていることが窺える．図示はしていないが，クラスタリング結果の機能ベクトルを見ると，ハブノードなどの影響度の大きいノード群からの距離によってク



(a) FCE法

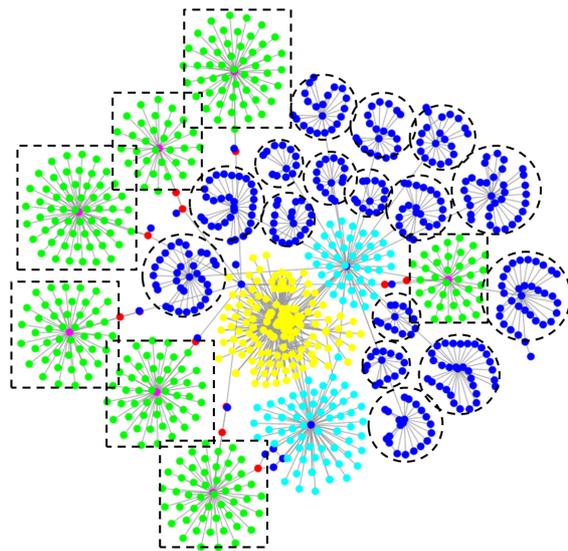


(b) ICFC法 ($p = 0.3$)

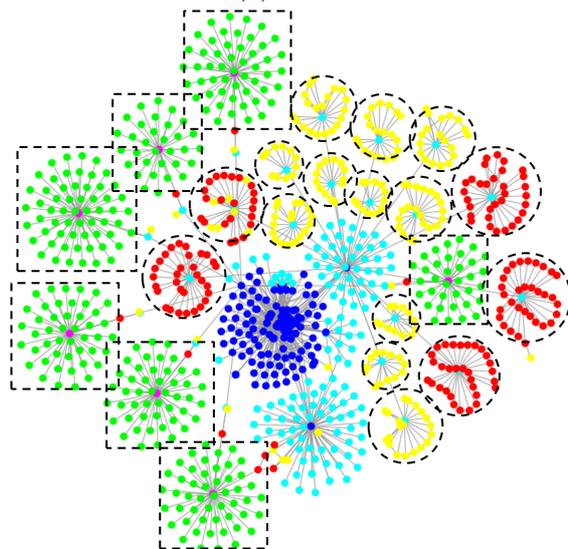


(c) LTFC法

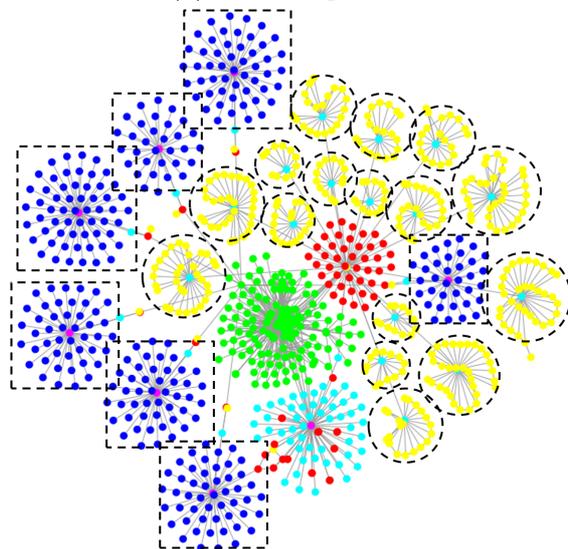
図6.4 Hosiネットワーク ($K = 10$)



(a) FCE法

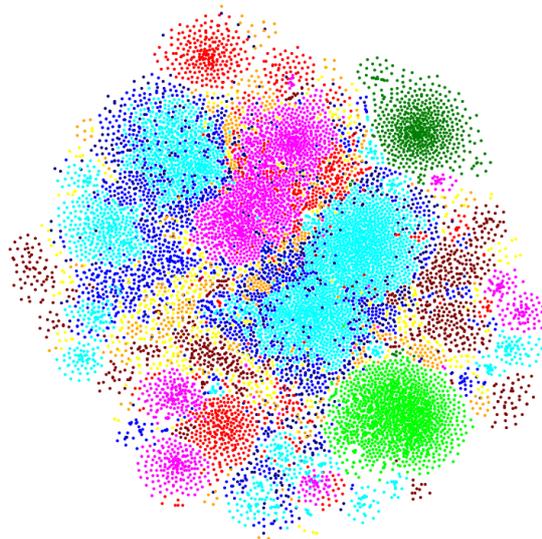


(b) ICFC法 ($p = 0.2$)

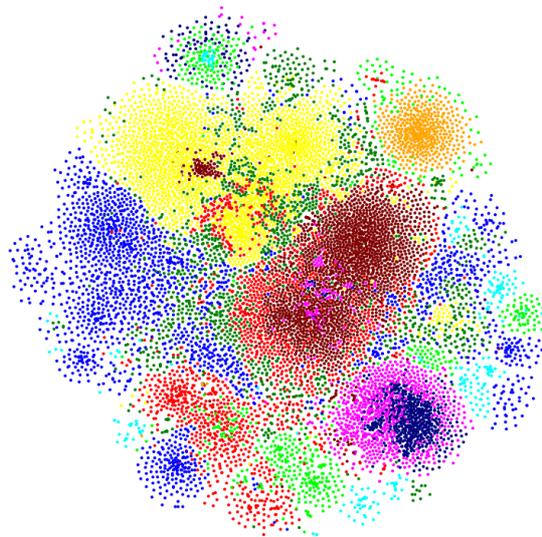


(c) LTFC法

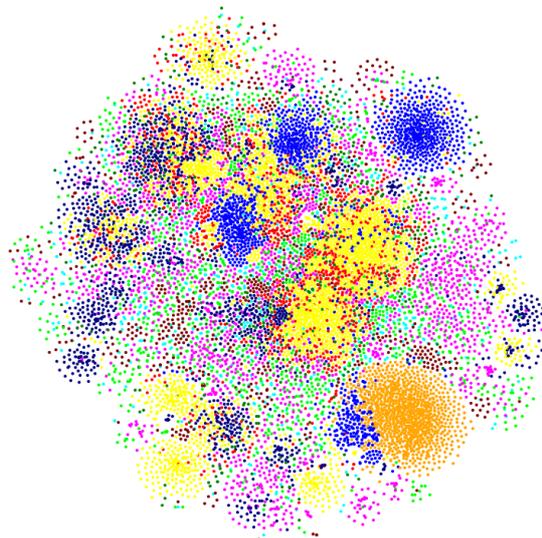
図6.5 Yamaguchiネットワーク ($K = 6$)



(a) FCE法



(b) ICFC法 ($p = 0.2$)



(c) LTFC法

図6.6 Trackbackネットワーク ($K = 10$)

ラスタリングしている傾向にあった。

以上の3つのネットワークに対する結果より，IDFCE法ではFCE法が比較的苦手としていたクラスタ係数の高いハブノードを抽出できることが示唆された。

6.3.4 コミュニティ抽出結果の類似性分析

各ネットワークのコミュニティ抽出結果の類似性評価として，正規化相互情報量 [59]を表 6.1に示す。正規化相互情報量は，0 から 1 の値をとり，値が大きいほどクラスタ集合が類似していることを示す。表中IC-RWの列はICFC法とFCE法によるコミュニティ抽出結果の類似性，LT-RWの列はLTFC法とFCE法によるコミュニティ抽出結果の類似性を示している。最終列はIDFCE法であるICFC法とLTFC法それぞれによる抽出結果の類似性を示している。表 6.1より，LTFC法の方がFCE法と類似する結果が得られていることがわかる。

表6.1 正規化相互情報量

ネットワーク	IC-RW	LT-RW	IC-LT
Hosei	0.63368	0.76294	0.61365
Yamaguchi	0.71809	0.85440	0.75614
Trackback	0.33061	0.42887	0.23388

ICFC法とLTFC法は仮定しているモデルが違うため，両手法の類似度は比較的高いが異なる抽出結果が得られた。両手法の大きな違いは，ノードの機能を表す機能ベクトルが意味するところにある。ICモデルによるシミュレーションでは，各リンクが与えられた拡散確率でそのリンクにおける子ノードをアクティブにするため，各ノードは複数の親ノードから情報を受け取れる可能性がある。すなわち，自分に情報が到達する経路は多く存在する。一方LTモデルによるシミュレーションでは，各ノードが持つ閾値を超した親ノードのみが子ノードをアクティブにするため，各ノードは唯一の親ノードから情報を受け取る。すなわち，自分に情報が到達する経路は限られる。したがって，距離 d 離れたノードを d 近傍ノードとすると，ICFC法では情報拡散において有効な d 近傍ノード数により，LTFC法では自分に情報が到達する経路長の期待値により分類していると考えられる。機能ベクトルのクラスタリング結果を見ると，近傍ノード数あるいはハブノードからの距離などに基づいて分類される傾向にある。

6.3.5 機能ベクトルに関する考察

上述したように，1回のシミュレーションにおいて，ICモデルでは，情報源が異なるたびに親ノードが異なるため，各時刻でアクティブになる期待値に単調性がないLTモデルでは，情報源ノードが異なっても，実効的な親ノードは1人であるため，各時刻でアクティブになる期待値は増えることがなく，単

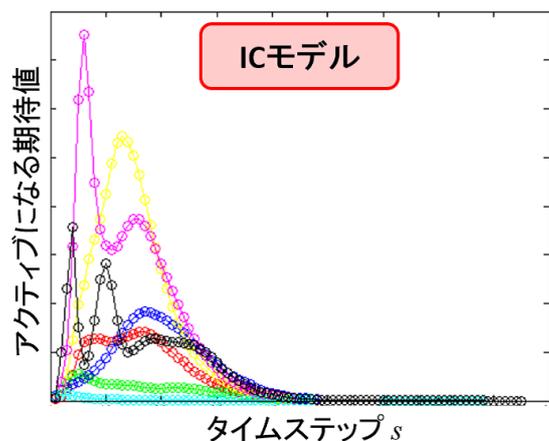


図6.7 ICFC法における情報受信期待値ベクトル

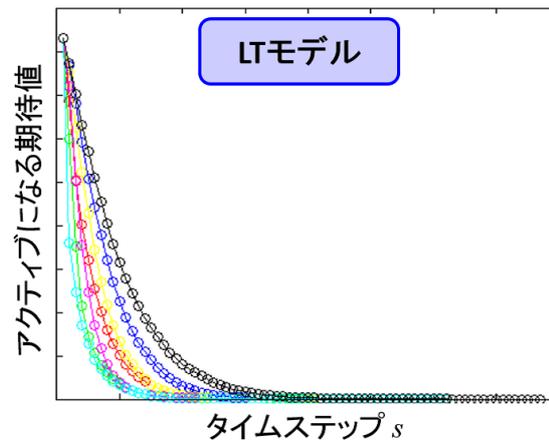


図6.8 LTFC法における情報受信期待値ベクトル

調に減少するこのような違いから，両手法で異なる結果が得られ，ICFC法では情報拡散において有効な d 近傍ノード数により，LTFC法では自分に情報が到達する経路長の期待値により分類していると考えられる．

6.4 おわりに

この章では，ランダムウォークモデルを仮定した通常のFCE法とは異なり，ネットワーク上での情報拡散モデルを仮定した機能コミュニティ抽出法（IDFCE法）を提案した．提案したICFC法とLTFC法による結果の基本的な分析をし，IDFCE法の有効性及びFCE法との相違点を明らかにした．今回用いた情報拡散モデルは，ノードが他のノードをアクティブにするタイミングが同期的であり，かつ，一度アクティブになったノードが再び非アクティブさらにアクティブになることができないSIR型である．今後は，情報拡散のタイミングに時間遅れを導入した非同期型や何度でもアクティブになることができるSIS型の情報拡散モデルによる検証を進めていくつもりである．

第7章

時間減衰重みを導入した 機能コミュニティ抽出法：TDFCE法

7.1 はじめに

ソーシャルネットワーク分析の分野では、個々のノードの性質に着目し、ノードをランキングすることで膨大な数のノードから重要ノードを抽出するための指標がいくつか提案されている [8, 10]。情報拡散や意見形成などネットワーク上での現象においても、隣接ノードを多く有するハブノードが重要な役割を果たすことが知られている。

単にハブノードといっても、属するサブネットワークによってハブ周辺ノードの隣接関係が異なる場合がある。周辺ノード同士にリンクがない場合や密結合する場合も想定できる。たとえば、共著ネットワークでは、指導教員がハブノードとなり、学生や研究員など周辺ノードとなる。研究室の方針によって、周辺ノード間のリンク密度やクラスタ係数などが異なると想定できる。Webのハイパーリンクネットワークでは、インデックスページがハブノード、コンテンツページが周辺ノードとなり、コンテンツページ間の関係によってハブノードないしはサブネットワークの性質が異なる。これらの例のように、ハブノードはその周辺ノードの関係によって立場が幾分か異なり、これらを詳細に分類することで、情報拡散や意見形成などのネットワーク上での現象における影響を詳細に分析可能になると期待できる。

著者らは、ネットワーク上でのノードの機能・役割を同定し、機能が類似するノード群からなる機能コミュニティを抽出する手法（FCE法）を提案した [44, 45]。

この章では、さらに、周辺ノードの隣接関係や周辺ノードとの関係からハブノードを詳細分類することを試みる。FCE法における機能ベクトルは、PageRankスコアがある程度収束するまで反復させた際の収束曲線である。しかし、密結合するサブネットワークでは、隣接ペアの次数の差が小さく、比較的早期に曲線が平坦化されるため、全ノードの機能ベクトルが収束するまでの高次元ベクトルから個々のノードの特徴を抽出するのが困難な場合がある。そこで、局所的構造の影響をより強く受けている初期段階に着目し、機能ベクトルに減衰関数を導入する。具体的には、機能ベクトルに対し、時間減衰関数

による重みを掛けることで，収束曲線の初期段階を強調する．そして，FCE法と同様に，ノードをクラスタリングする．

構造の異なる複数のネットワークを用いた評価実験より，減衰関数を導入することにより，通常のFCE法では困難であったクラスタ係数の高いサブネットワークにおけるハブノードも抽出可能であることを示す．さらに，周辺ノードの隣接関係（次数・クラスタ係数）の違いによりハブノードを詳細分類できることも示す．

7.2 時間減衰重み付き機能コミュニティ抽出法：TDFCE法

FCE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. 各ノードの機能ベクトルとしてPageRankスコア収束曲線 x_v を構築;
3. 各ノードペアの機能ベクトル間のコサイン類似度 $\rho(u, v)$ を計算;
4. K -medians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(2)機能ベクトル構築法および，時間減衰関数による重みとその導入法について説明する．

7.2.1 機能ベクトル構築法

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 [15]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機能を表す機能ベクトルとして採用する．

$(u, v) \in E$ のとき $a(u, v) = 1$ ，それ以外の場合 $a(u, v) = 0$ とし隣接行列 $A \in \{0, 1\}^{|V| \times |V|}$ 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．ここで，行推移確率行列 P の要素を $p(u, v) = a(u, v)/|\Gamma(u)|$ と定義する．各ノードのPageRankスコアを要素とするベクトル y は， $y(v) \geq 0$ で $\sum_{v \in V} y(v) = 1$ となる．繰り返しステップのパラメータ s を用い，ステップ s でのPageRankスコアベクトル y_s は $y_0 = (1/|V|, \dots, 1/|V|)^T$ から始める以下の更新式の極限分布として定義される:

$$y_s^T = y_{s-1}^T P \quad (7.1)$$

ここで b^T はベクトル b の転置を表わす．ノード u に注目すると，

$$y_s(u) = \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \quad (7.2)$$

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 の機能ベクトルとする．特にこの章では，ハブノードの詳細な分類を可能にするために，局所的な情報を含めるべく初期ベクトルの値も用い， $\mathbf{x}_u = (y_0(u), y_2(u), \dots, y_S(u))^T$ とする．

7.2.2 時間減衰関数

PageRankスコアの収束曲線は，各ノードから隣接ノードへのジャンプ，さらにその隣接ノードへ次々とランダムウォークして行く際の各ノードへの滞在確率の変化を表している．従って，収束曲線の初期段階では比較的局所的なリンク構造を反映していると考えられる．この点に注目し，初期段階をより強調させるために時間減衰関数による重み付けを導入する．時間減衰関数は，PageRankの離散時間の進行につれ小さな重みを付ける関数であり，指数的減衰関数とべき乗的減衰関数を用いることを検討する．

指数的減衰関数は， $w_{\text{exp}}(s; \lambda) = \exp(-\lambda s)$ で定義する．ここに， $\lambda \geq 0$ は減衰の強さを表すパラメータである．同様にべき乗的減衰関数は， $w_{\text{pow}}(s; \lambda) = s^{-\lambda} = \exp(-\lambda \log s)$ で定義する．

これらの減衰関数による重みを以下のように導入する（図 7.1参照）．

$$\mathbf{x}_u = (w(0; \lambda) \cdot y_0(u), \dots, w(S; \lambda) \cdot y_S(u))^T \quad (7.3)$$

特に， $\lambda = 0$ とした場合は通常の機能ベクトルと等価となり，自然な拡張である．

7.3 評価実験

評価実験には，ハブと，その周辺ノードの次数・クラスタ係数を変化させた複数のサブネットワークを含む人工ネットワークを用いた．

さらに，実データとして大学のウェブサイト内のページを2012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Yamaguchiネットワークを用いた^{*1}．

PageRankの反復回数 S は $S = 10000$ とした．機能コミュニティ数 K は， $2 \leq K \leq 10$ の範囲で実験し，最も適切だと思われる K の値を用いた．本文中に図示していない他の K においても，この章の主張と概ね矛盾しない結果が得られている．また，指数的減衰関数とべき乗的減衰関数を導入した結果に大きな違いはなかったため，指数的減衰関数による結果のみを示す．

^{*1} 山口大学 <http://www.yamaguchi-u.ac.jp/>

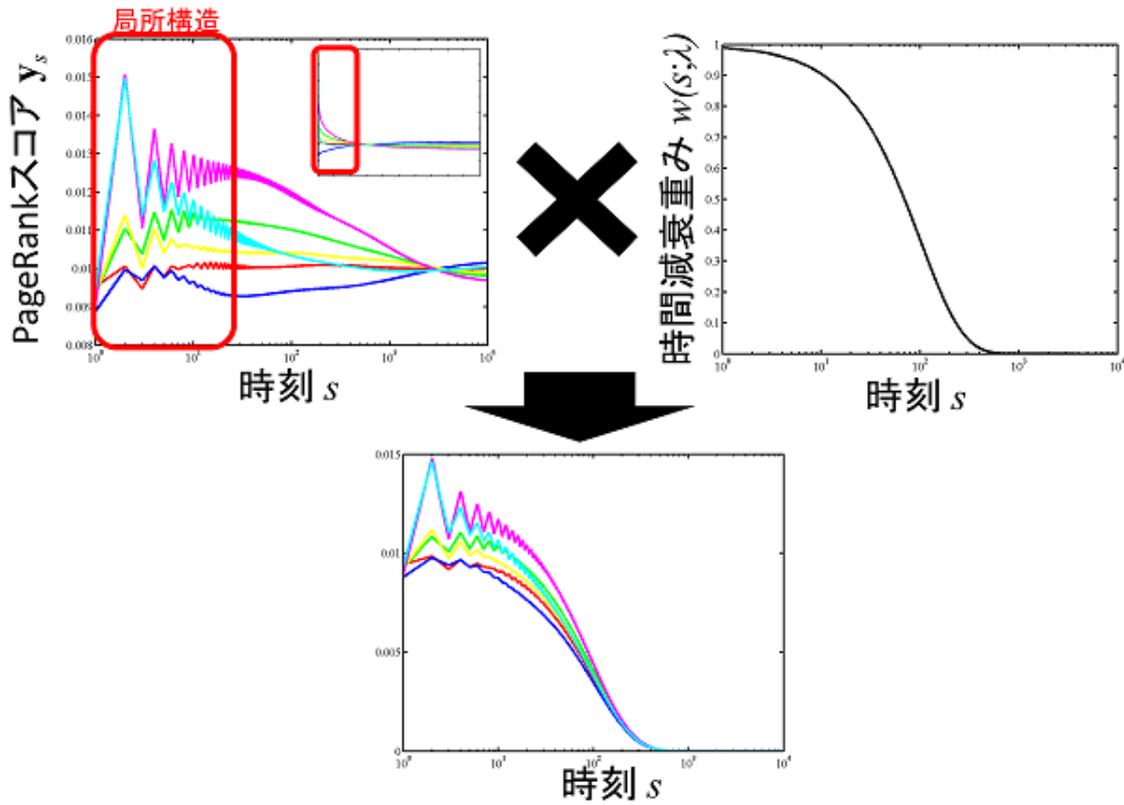


図7.1 時間減衰関数による重みと機能ベクトル

7.3.1 人工ネットワークに対する処理結果

人工ネットワークに対する結果を図 7.2 に示す。この人工ネットワークは、1つのハブノードと12個の周辺ノードからなるサブネットワークが8つ隣接するネットワークである。各サブネットワークにおける周辺ノード同士は、それぞれ 0近傍、1近傍、2近傍、3近傍ノードと隣接する構造である。図の中心に対して対称に位置するサブネットワークは同一の構造である。

図 7.2(a) は減衰の強さを表すパラメータ $\lambda = 0$ とした結果である。すなわち、通常の機能コミュニティ抽出結果である。図中の対象に位置するサブネットワークが同一の配色パターンとなっていることがわかる。すなわち、類似構造をもつサブネットワークにおける類似の立場にあるノードが抽出できている。しかしながら、周辺ノードのクラスタ係数が高いサブネットワークでは、ハブノードと周辺ノードの識別がうまくできていない。サブネットワーク内の次数相関が正に近づくほど、PageRankスコアの平坦化が比較的早まり、高次元の機能ベクトルに特徴が出にくくなるのが理由の1つと考えられる。図示はしていないが、当然のことながら機能コミュニティ数 K を多くすればハブノードが識別できるが、周辺ノードの次数が低いサブネットワークからハブノードが識別されていく。

図 7.2(b)は $\lambda = 1$ とした結果である。通常のFCE法では困難であった、クラスタ係数の高いサブネットワークのハブノードが識別できる。さらに、周辺ノードの関係性によりハブノードが異なる色で配色され、ハブノードの詳細な分類ができることが示唆された。

7.3.2 実ネットワークに対する処理結果

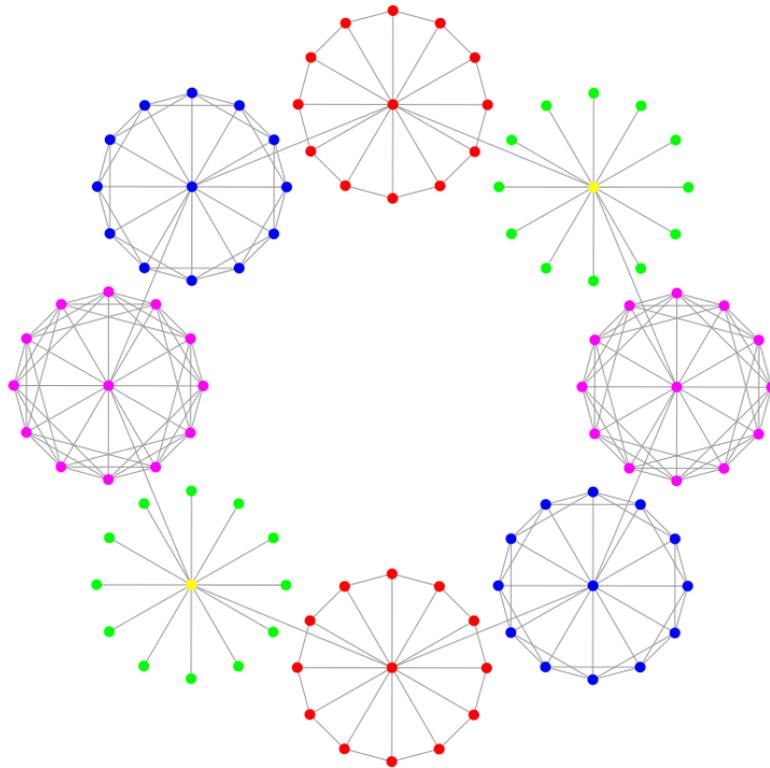
Yamaguchiネットワークに対する結果を図 7.3に示す。このネットワークでは、「各年度の学内イベントの特集ページ」サブネットワークや、「世代ごとの大学の歴史ページ」サブネットワークなど、ノード数は異なれど同一のトポロジ構造をしたサブネットワークを複数有している。

図 7.3(a)は $\lambda = 0$ とした結果である。左側部分にある「世代ごとの大学の歴史ページ群」や、「年度ごとのイベント特集ページ群」など、同一の機能を有すると考えられるノードが同一の機能コミュニティとして抽出されている。しかし、「年度ごとのイベント特集ページ群」に関しては、中心のインデックスページとその周辺のコンテンツページの区別ができない。

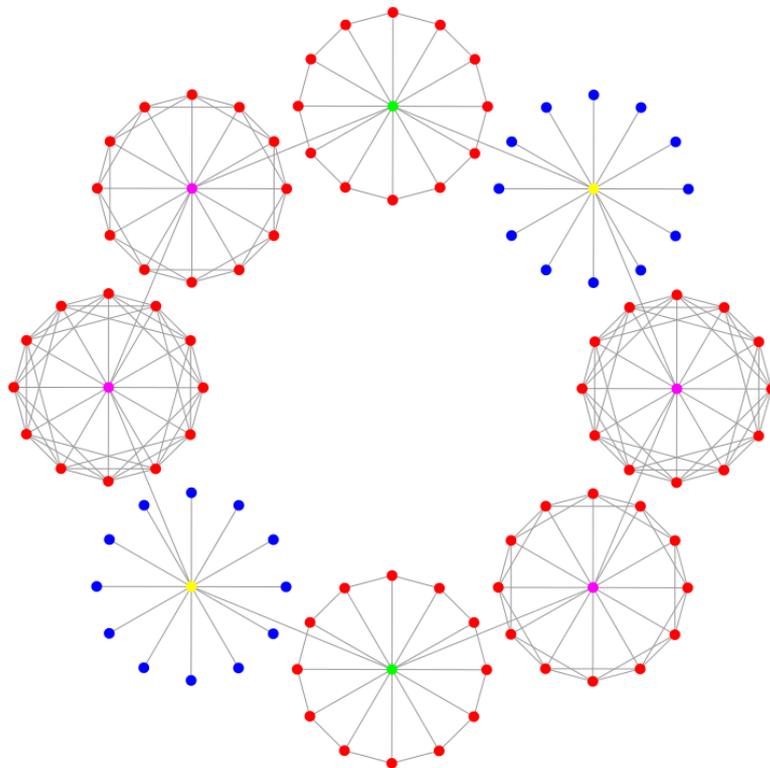
図 7.3(b)は $\lambda = 1$ とした結果である。類似の機能を有するノードは同一色で配色されていることが見て取れる。一方、通常のFCE法では困難であった、クラスタ係数が比較的高い「年度ごとのイベント特集ページ群」においても、ハブノードを識別できている。さらに、「大学の歴史ページ群」におけるハブノードとは異なる機能を持つノードとして識別されている。各サブネットワークでは周辺ノードの数が異なるため、ハブノードのクラスタ係数は異なる値となることを注意しておく。すなわち、単純なクラスタ係数によってハブノードを分類しているのではない。

7.4 おわりに

この章では、機能的に類似するノード群を抽出するFCE法に、時間減衰関数による重みを導入することで、周辺ノードの隣接構造の違いによるハブノードの詳細分類を試みた。実験の結果、この章で用いたネットワークに対してはハブの周辺ノードの隣接関係の違いからハブノードに複数のクラスタを当てはめることができ、識別できた。今後は、多様なネットワークに対し適用し、本手法の有効性を検証していきたい。

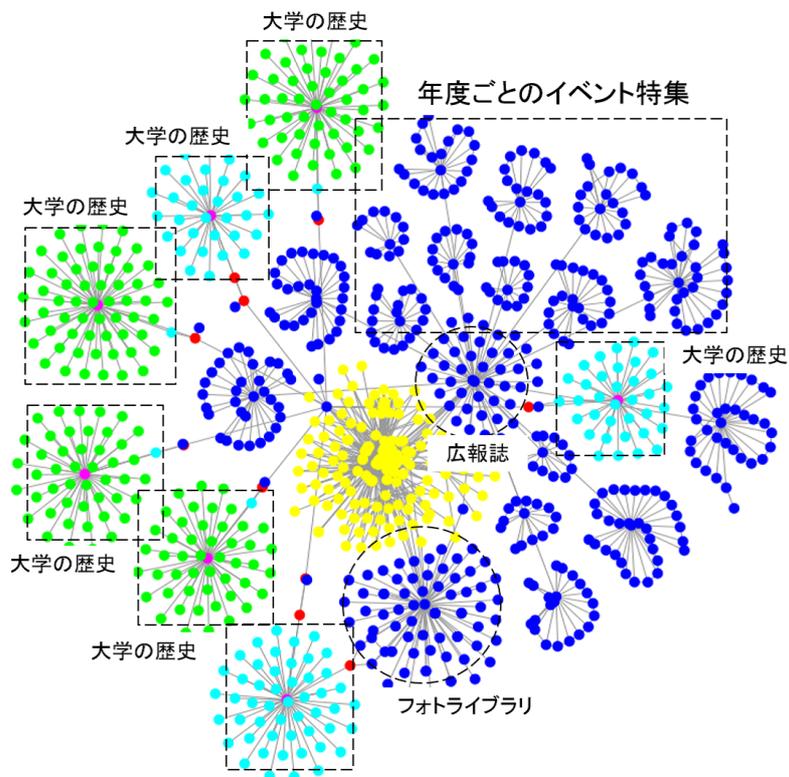


(a) $\lambda = 0$

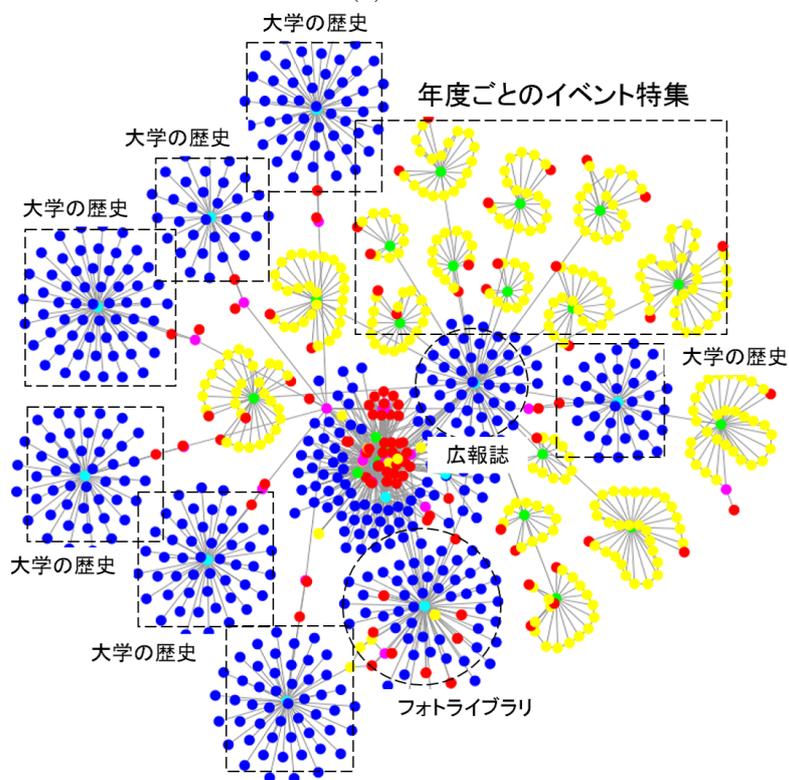


(b) $\lambda = 1$

図7.2 人工ネットワークの処理結果（指数的減衰） $K = 5$



(a) $\lambda = 0$



(b) $\lambda = 1$

図7.3 Yamaguchiネットワークの処理結果（指数的減衰） $K = 6$

第8章

結論

本論文では、ネットワークにおける各ノードの機能・役割に着目し、類似の機能を有するノード群、機能コミュニティを抽出する手法を提案した。

第3章では、従来のリンク密度に基づくコミュニティとは異なり、ノード間の類似性に着目したコミュニティ抽出法に焦点を当てた。ノード間の同値性を近似的に計算する手法としてVersim法とSimrank法、類似機能を有するノードを同定する提案手法(FCE法)を取り上げ、3手法に基づいたコミュニティ抽出の結果を可視化により定性的に、時間計算量により定量的に評価した。Versim法、Simrank法ともに近似的に正則同値を同定するための手法であるため、FCE法と近い結果が得られる場合があった。しかしVersim法は、隣接性が強く考慮されているため、FCE法で異なると判定されたコミュニティも同一コミュニティとして抽出する傾向があった。またVersim法、Simrank法は、局所的なリンク構造の類似性に直接着目することから、ノードの機能を考慮しないNewman法と近い結果が得られる場合もあった。FCE法は、機能ベクトルとしてPageRankスコアの収束曲線を用いて、大域的な構造上における現象の類似性を見ており、かつスケラビリティのある方法である。その結果、局所構造をみるVersim法やSimrank法と比較して、全体構造をみるFCE法は同質の機能・役割を有するノード群を要素とするコミュニティをより適切に抽出できることが示された。また、中心性指標とは異なり、一義的な性質の判別ではなく、性質ごとのコミュニティを抽出可能であることを示した。さらに、個々のノードからの視点によるコミュニティ抽出といった、発展性のある手法であることも示唆された。今後は、有向ネットワークや多重ネットワーク、2部グラフなどの一般的なネットワークを対象としたコミュニティ抽出などへの拡張を検討していくつもりである。

第4章では、有向ネットワークを対象に、無向化した場合にはわ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(DFCE法)を提案した。DFCE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可能性がある。そのため、 $\alpha = 0$ では有向ネットワークに適用できない。この章では、大域ジャンプ確率を $\alpha \neq 0$ とし、パラメータ α に対する一考察をした。 $\alpha \simeq 1$ 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。

DFCE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DFCE法の有効性を検証するとともに、大域ジャンプ確率 α やクラスター数 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。

第5章では、従来とは異なり、各ノードの視点に基づくコミュニティの概念としてパーソナライズ機能コミュニティおよび、その抽出手法（PFCE法）を提案した。パーソナライズ機能コミュニティは、エゴノードの視点からの他のノードの機能・役割が類似するノードをコミュニティとしたもので、コミュニティ抽出では新しいパラダイムであるといえる。PFCE法は、各ノードの機能を表す特徴ベクトルであるPageRankスコア変化曲線の計算において、Personalized PageRank に基づき、初期ベクトルのエゴノードに対応する要素を1、その他のノードの対応する要素を0とする調整を加えた（和が1になるように正規化する）。さらに、任意のノードから確率 β でエゴノードにジャンプする、エゴジャンプという機構を導入した。

人工ネットワーク、ソーシャル・ネットワーク、ハイパーリンクネットワーク、道路網といった構造の異なる複数種のネットワークを用いた評価実験により、エゴノードの視点から見て同質な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。すなわち、客観的に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。しかし、エゴノードからの距離で分類しているわけではない。また通常の機能コミュニティとの比較により、エゴ

ノード近くを細分化し、遠くを大まかに分類する結果が得られた。これは、自身の周辺は自身との関係が強く、直接的に影響するため、分類の粒度が細くなるが、自身から離れた位置に存在するものは関係が弱いので、分類の粒度が粗くなるという直観に対応している。

本手法の結果は、ドライバーがカーナビなどを閲覧するに際して、ドライバーのいる交差点付近は興味があるためその情報を細分化し、遠方の地域は比較的興味が薄いので細分化する必要がない、といった直観に対応した応用が考えられる。

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。また、エゴノードとして複数ノードを選択した場合での検証やエゴジャンプ確率の有用性を深く追求していくつもりである。

第6章では、ランダムウォークモデルを仮定した通常のFCE法とは異なり、ネットワーク上での情報拡散モデルを仮定した機能コミュニティ抽出法（IDFCE法）を提案した。提案したICFC法とLTFC法による結果の基本的な分析をし、IDFCE法の有効性及びFCE法との相違点を明らかにした。今回用いた情報拡散モデルは、ノードが他のノードをアクティブにするタイミングが同期的であり、かつ、一度アクティブになったノードが再び非アクティブさらにアクティブになることができないSIR型である。今後は、情報拡散のタイミングに時間遅れを導入した非同期型や何度でもアクティブになることができるSIS型の情報拡散モデルによる検証を進めていくつもりである。

第7章では、機能的に類似するノード群を抽出するFCE法に、時間減衰関数による重みを導入することで、周辺ノードの隣接構造の違いによるハブノードの詳細分類を試みた。実験の結果、この章で用いたネットワークに対してはハブの周辺ノードの隣接関係の違いからハブノードに複数のクラスタを当てはめることができ、識別できた。今後は、多様なネットワークに対し適用し、本手法の有効性を検証していきたい。

本論文では、従来のコミュニティ概念とは異なる、連結性を仮定しない、機能や役割が類似するノード群からなるコミュニティ概念、ならびに、抽出する手法を提案した。関連概念である同値性を近似的に抽出する手法などとなり、各ノードの機能をPageRankの収束曲線を用いてベクトルとして表現するところに新規性がある。ベクトルとして定義することで、各章で提案したように様々な拡張を可能にした。基本となるFCE法、ならびに、各種拡張手法に対して、さらに多様なデータを用いて有用性を検証していきたい。また、PageRankの収束曲線を用いるアイデアをベースとし、コミュニティ抽出以外の応用も考慮に入れ、研究を進めていきたい。

謝辞

本論文は筆者が静岡県立大学大学院経営情報イノベーション研究科経営情報イノベーション専攻博士後期課程に在籍中の研究成果をまとめたものである。同専攻教授 斉藤和巳 先生には指導教員として本研究の実施の機会を与えて戴き、その遂行にあたって終始、ご指導を戴いた。ここに深謝の意を表す。同専攻教授 池田哲夫 先生並びに同専攻准教授 武藤伸明 先生には副指導教員としてご助言を戴くとともに本論文の細部にわたりご指導を戴いた。ここに深謝の意を表す。また本研究の全般にわたり和歌山大学 システム工学部教授 風間一洋 先生には、研究を遂行するにあたり多くのご助言と激励のお言葉を戴くとともに本論文の細部にわたりご指導を戴いた。本論文の主査になっていただいた同専攻教授 松浦博 先生、副査になっていただいた中部大学工学部教授 中野良平 先生にも、中間審査会において貴重なご意見を賜ることができた。ここに深謝の意を表す。本学部 斉藤研究室の各位には研究遂行にあたり日頃より有益なご討論ご助言を戴いた。ここに感謝の意を表す。

参考文献

- [1] Barabási, A. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol. 286, No. 5439, pp. 509–512 (1999).
- [2] Albert, R., Jeong, H. and Barabási, A.-L.: Error and attack tolerance of complex networks, *Nature*, Vol. 406, pp. 378–382 (2000).
- [3] Watts, D. J. and Strogatz, S. H.: Collective dynamics of ‘small-world’ networks, *Nature*, Vol. 393, No. 6684, pp. 440–442 (1998).
- [4] Kleinberg, J.: The Small-World Phenomenon: An Algorithmic Perspective, *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pp. 163–170 (2000).
- [5] Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E*, Vol. 67, No. 5, pp. 056104+ (2003).
- [6] Ravasz, E. and Barabási, A. L.: Hierarchical organization in complex networks, *Physical Review E*, Vol. 67, No. 2, pp. 026112+ (2003).
- [7] Leskovec, J., Chakrabarti, D., Kleinberg, J. M., Faloutsos, C. and Ghahramani, Z.: Kronecker Graphs: An Approach to Modeling Networks, *Journal of Machine Learning Research*, Vol. 11, pp. 985–1042 (2010).
- [8] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, No. 3, pp. 215–239 (1979).
- [9] Sabidussi, G.: The centrality index of a graph, *Psychometrika*, Vol. 31, No. 4, pp. 581–603 (1966).
- [10] Newman, M. E. J.: A measure of betweenness centrality based on random walks, *Social Networks* (2005).
- [11] Brandes, U.: A Faster Algorithm for Betweenness Centrality, *Journal of Mathematical Sociology*, Vol. 25, pp. 163–177 (2001).
- [12] Katz, L.: A new status index derived from sociometric analysis, *Psychometrika*, Vol. 18, pp. 39–43 (1953).

- [13] Bonacich, P.: Power and Centrality: A Family of Measures, *The American Journal of Sociology*, Vol. 92, No. 5, pp. 1170–1182 (1987).
- [14] Brin, S. and L. Page: The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems*, Vol. 30, pp. 107–117 (1998).
- [15] Langville, A. N. and Meyer, C. D.: Deeper inside pagerank, *Internet Mathematics*, Vol. 1, p. 2004 (2004).
- [16] Kleinberg, J. M.: Authoritative sources in a hyperlinked environment, *J. ACM*, Vol. 46, pp. 604–632 (1999).
- [17] 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 (2003).
- [18] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, p. 026113 (2004).
- [19] Newman, M. E. J.: Detecting community structure in networks, *The European Physical Journal B - Condensed Matter and Complex Systems*, Vol. 38, No. 2, pp. 321–330–330 (2004).
- [20] Newman, M. E. J.: Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, Vol. 69, p. 066133 (2004).
- [21] Clauset, A., Newman, M. E. J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol. 70, No. 6, pp. 066111+ (2004).
- [22] Newman, M. E. J.: Modularity and community structure in networks, *Proceedings of the National Academy of Sciences*, Vol. 103, No. 23, pp. 8577–8582 (2006).
- [23] Shi, J. and Malik, J.: Normalized Cuts and Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888–905 (2000).
- [24] Hagen, L. and Kahng, A. B.: 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).
- [25] 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).
- [26] 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, Springer Berlin / Heidelberg, pp. 243–257 (2009).
- [27] Seidman, S. B.: Network structure and minimum degree, *Social Networks*, Vol. 5, No. 3, pp.

- 269 – 287 (1983).
- [28] 風間一洋, 佐藤進也, 斉藤和巳, 山田武士 : 人間関係の重なりを持つコミュニティ構造の抽出 (特集 ネットワークが創発する知能), コンピュータソフトウェア, Vol. 24, No. 1, pp. 81–90 (2007-01-26).
- [29] Borgatti, S.: Two algorithms for computing regular equivalence, *Social Networks*, Vol. 15, No. 4, pp. 361–376 (1993).
- [30] Leicht, E. A., Holme, P. and Newman, M. E. J.: Vertex similarity in networks, *Physical Review E*, Vol. 73, No. 2, pp. 1–10 (2005).
- [31] Jeh, G. and Widom, J.: SimRank: a measure of structural-context similarity, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, ACM, pp. 538–543 (2002).
- [32] Smola, A. J. and Kondor, R.: Kernels and Regularization on Graphs, *Machine Learning*, Vol. 2777, No. 212938, pp. 1–15 (2003).
- [33] Higham, N. J.: The scaling and squaring method for the matrix exponential revisited, *SIAM J. Matrix Anal. Appl.*, Vol. 26, p. 2005 (2005).
- [34] Christakis, N. A. and Fowler, J. H.: *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, Little, Brown and Company (2009).
- [35] Borgatti, S. P. and Everett, M. G.: Notions of Position in Social Network Analysis, *Sociological Methodology*, Vol. 22, No. 1992, pp. 1–35 (1992).
- [36] Lorrain, F. and White, H. C.: Structural equivalence of individuals in social networks, *The Journal of Mathematical Sociology*, Vol. 1, No. 1, pp. 49–80 (1971).
- [37] Wasserman, S. and Faust, K.: *Social Network Analysis: Methods and Applications*, Cambridge University Press (1994).
- [38] Burt, R. S.: Positions in Networks, *Social Forces*, Vol. 55, pp. 93–122 (1976).
- [39] Paul, Dooren, P. V., Blondel, V. D., Blondel, V. D. and Dooren, P. V.: A Measure Of Similarity Between Graph Vertices. With Applications To Synonym Extraction And Web Searching, *SIAM Review*, No. 46, p. 2004 (2004).
- [40] Holme, P. and Huss, M.: Role-similarity based functional prediction in networked systems: application to the yeast proteome., *J R Soc Interface*, Vol. 2, No. 4, pp. 327–333 (2005).
- [41] Scripps, J., Tan, P.-N. and Esfahanian, A.-H.: Node Roles and Community Structure in Networks, *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, WebKDD/SNA-KDD '07, New York, NY, USA, ACM, pp. 26–35

- (2007).
- [42] 池田哲夫, 斉藤和巳, 武藤伸明, 伏見卓恭 : 複雑ネットワーク分析手法による道路網の分析, 第20回地理情報システム学会研究発表大会(GISA2011) (2011).
- [43] Crucitti, P., Latora, V. and Porta, S.: Centrality measures in spatial networks of urban streets, *Physical Review E*, Vol. 73, No. 3, pp. 036125+ (2006).
- [44] 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).
- [45] 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).
- [46] Even-Dar, E. and Shapira, A.: A Note on Maximizing the Spread of Influence in Social Networks, *Internet and Network Economics* (Deng, X. and Graham, F., eds.), Lecture Notes in Computer Science, Vol. 4858, Springer Berlin / Heidelberg, pp. 281–286 (2007).
- [47] Zachary, W.: An information flow model for conflict and fission in small groups, *Journal of Anthropological Research*, Vol. 33, pp. 452–473 (1977).
- [48] : 全国デジタル道路地図データベース静岡県版 (2011).
- [49] : 全国デジタル道路地図データベース静岡市版 (2011).
- [50] Yamada, T., Saito, K. and Ueda, N.: Cross-entropy directed embedding of network data, *Proceedings of the 20th International Conference on Machine Learning (ICML03)*, pp. 832–839 (2003).
- [51] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: simple building blocks of complex networks., *Science (New York, N.Y.)*, Vol. 298, No. 5594, pp. 824–827 (2002).
- [52] Meyer, C.: *Matrix Analysis and Applied Linear Algebra*, SIAM: Society for Industrial and Applied Mathematics (2001).
- [53] Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web., Technical Report 1999-66, Stanford InfoLab (1999).
- [54] Kamada, T. and Kawai, S.: An algorithm for drawing general undirected graphs, *Inf. Process. Lett.*, Vol. 31, pp. 7–15 (1989).
- [55] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, New York, NY, USA, ACM, pp. 137–146 (2003).
- [56] Kimura, M., Saito, K., Nakano, R. and Motoda, H.: Extracting influential nodes on a social

- network for information diffusion, *Data Min. Knowl. Discov.*, Vol. 20, No. 1, pp. 70–97 (2010).
- [57] Kimura, M., Saito, K. and Nakano, R.: Extracting influential nodes for information diffusion on a social network, *Proceedings of the 22nd national conference on Artificial intelligence - Volume 2*, AAAI'07, AAAI Press, pp. 1371–1376 (2007).
- [58] Gfeller, D. and Paolo: Spectral Coarse Graining of Complex Networks, *Physical Review Letters*, Vol. 99, No. 3 (2007).
- [59] Cheng, H., Hua, K. A. and Vu, K.: Constrained locally weighted clustering, *PVLDB*, Vol. 1, No. 1, pp. 90–101 (2008).

本論文に関する原著論文

学術論文

1. 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, 2012.2 .
2. 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 .
3. 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, 2013.2 .
4. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“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”,
情報処理学会論文誌, 数理モデル化と応用 -ネットワークが創発する知能論文特集-, Vol.6, No.2,
pp.137–146, 2013.8 .
5. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“ノードの注目度に基づく機能コミュニティ抽出法”,
情報処理学会論文誌, データベース, Vol.6, No.4, pp.104–112, 2013.9 .

国際会議

1. Takayasu Fushimi, Kazumi Saito and Kazuhiro Kazama,
“Extracting Communities in Networks based on Functional Properties of Nodes”,
The 12th International Workshop on Knowledge Management and Acquisition for Intelligent Systems, pp.328–334, 2012.

研究会・大会

1. 伏見卓恭, 斉藤和巳, 風間一洋,
“PageRank影響度曲線によるコミュニティ抽出法の提案”,
ネットワークが創発する知能研究会 (JWEIN'11), 2011.
2. 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の提案”,
第4回Webとデータベースに関するフォーラム, 2011.
3. 伏見卓恭, 斉藤和巳, 風間一洋, 池田哲夫,
“有向ネットワークにおける機能コミュニティ抽出法の検証”,
ネットワークが創発する知能研究会 (JWEIN'12), 2012.
4. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“有向ネットワークにおける機能コミュニティと頑健性”,
第97回知識ベースシステム研究会 (SIG-KBS), 2012.
5. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋, 松尾真人,
“個々のノードの視点に基づく機能コミュニティ抽出法”,
第5回Webとデータベースに関するフォーラム, 2012.
6. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“道路ネットワークでの機能コミュニティ抽出による地域特性分析”,
第9回ネットワーク生態学シンポジウム, 2012.
7. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“情報拡散影響度に基づく機能コミュニティ抽出法”,
2013年度人工知能学会全国大会 (第27回), 2013.
8. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“ネットワークコミュニティの特性を用いたメタクラスタリング”,
ネットワークが創発する知能研究会 (JWEIN'13), 2013.
9. 伏見卓恭, 斉藤和巳, 池田哲夫, 風間一洋,
“減衰付き機能コミュニティ抽出法によるハブノードの詳細分類”,
第10回ネットワーク生態学シンポジウム, 2013.

受賞など

1. “ネットワーク機能コミュニティ抽出法の提案”，
第4回Webとデータベースに関するフォーラム 優秀論文賞
2. “ネットワーク機能コミュニティ抽出法の提案”，
第4回Webとデータベースに関するフォーラム 企業賞（ネクスト賞）
3. “ネットワーク機能コミュニティ抽出法”，
2011年度日本データベース学会 論文賞
4. “個々のノードの視点に基づく機能コミュニティ抽出法”，
第5回Webとデータベースに関するフォーラム 優秀論文賞
5. “個々のノードの視点に基づく機能コミュニティ抽出法”，
第5回Webとデータベースに関するフォーラム 企業賞（ミクシィ賞）
6. “道路ネットワークでの機能コミュニティ抽出による地域特性分析”，
第9回ネットワーク生態学シンポジウム ポスター優秀賞
7. “個々のノードの視点に基づく機能コミュニティ抽出法”，
2013年度山下記念研究賞（情報処理学会）