

ネットワークにおける 機能的類似ノード群の抽出

Extraction of functionally similar nodes in a network

伏見 卓恭

Takayasu Fushimi

本研究では、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、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。機能コミュニティ抽出のため、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ステップで辿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などの概念を提案し、それらを抽出する手法を提案する。

博士論文の各章では、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。

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 study, 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 and so forth, and each chapter in the thesis we compare these extended methods to the basic method and some existing methods.

In our experiments using several types of networks, we evaluate the characteristics of functional communities extracted by our proposed method.

1. はじめに

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。これらつながりの例は、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。

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。

既存のコミュニティ抽出手法として、Clauset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広く用いられている[1]。この手法は、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。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[2]。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[2]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[3]やJehら[4]の手法がある。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[5]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。

本研究では、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ない、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(図1参照)。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。

本稿では、博士論文の中で提案した手法の中で最も基本となる機能コミュニティ抽出法 (Functional Community

Extraction法: 以下, FCE法), に関して, 提案法の概念および主要技術について述べ, 従来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。

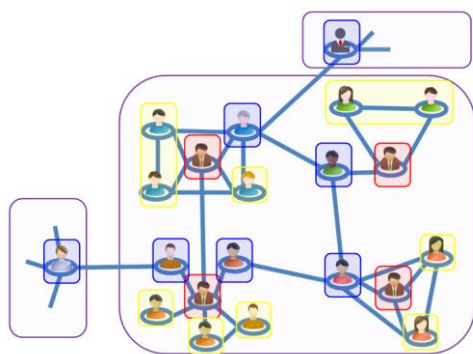


図1 機能的に類似するノード

2. 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 $\{C_1, \dots, C_k\}$ を抽出する。

1. 各ステップ s 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 \mathbf{x}_v を構築;
3. 各ノードペアの機能ベクトル \mathbf{x}_u と \mathbf{x}_v のコサイン類似度 $\rho(u, v)$ を計算;
4. K -medians法により全ノードを K 個のグループに分割;

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。手順1. と2. について詳細に説明する。

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。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する。

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 \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:

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

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,

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

この反復計算を, PageRank計算が収束するか, 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. 評価実験

現実のネットワークデータなどを対象に, FCE法を用いて機能コミュニティを抽出する。FCE法および比較のためのCNM法[1]による抽出結果を可視化により定性的に評価する。

3.1 ネットワークデータ

評価には以下のネットワークデータを用いる。

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[6]により生成した人工ネットワークである。以下, Hierarchicalネットワークと呼ぶ。

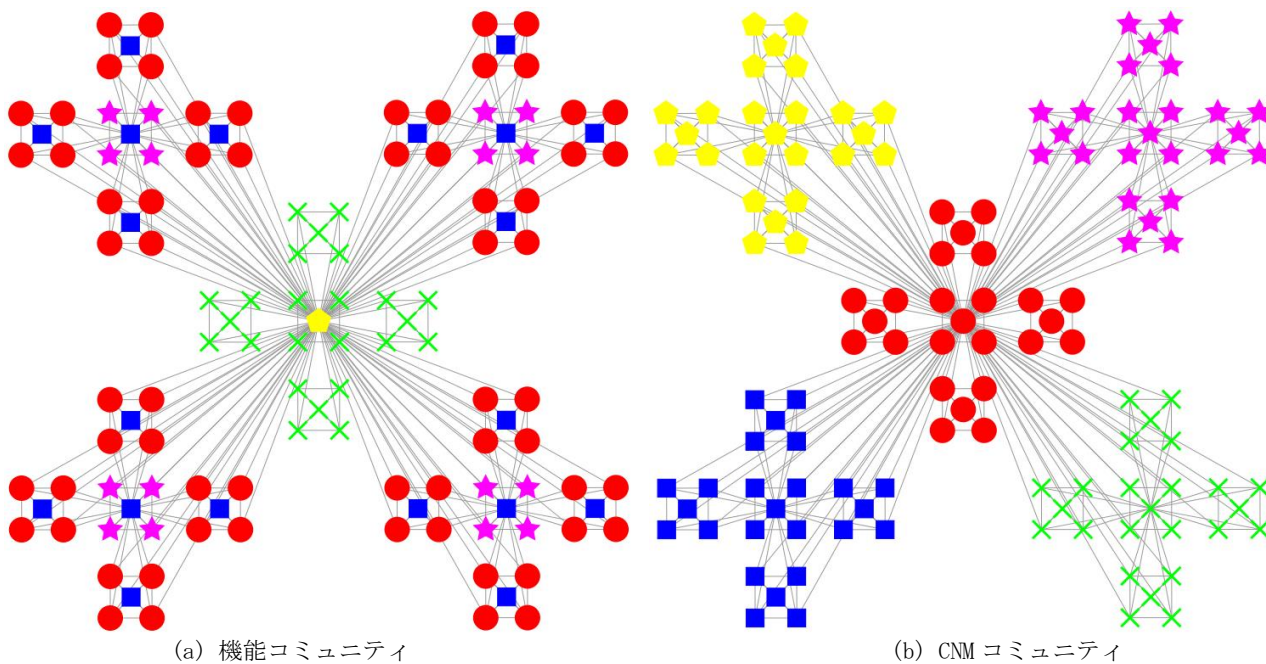


図2 Hierarchicalネットワークに対する処理結果

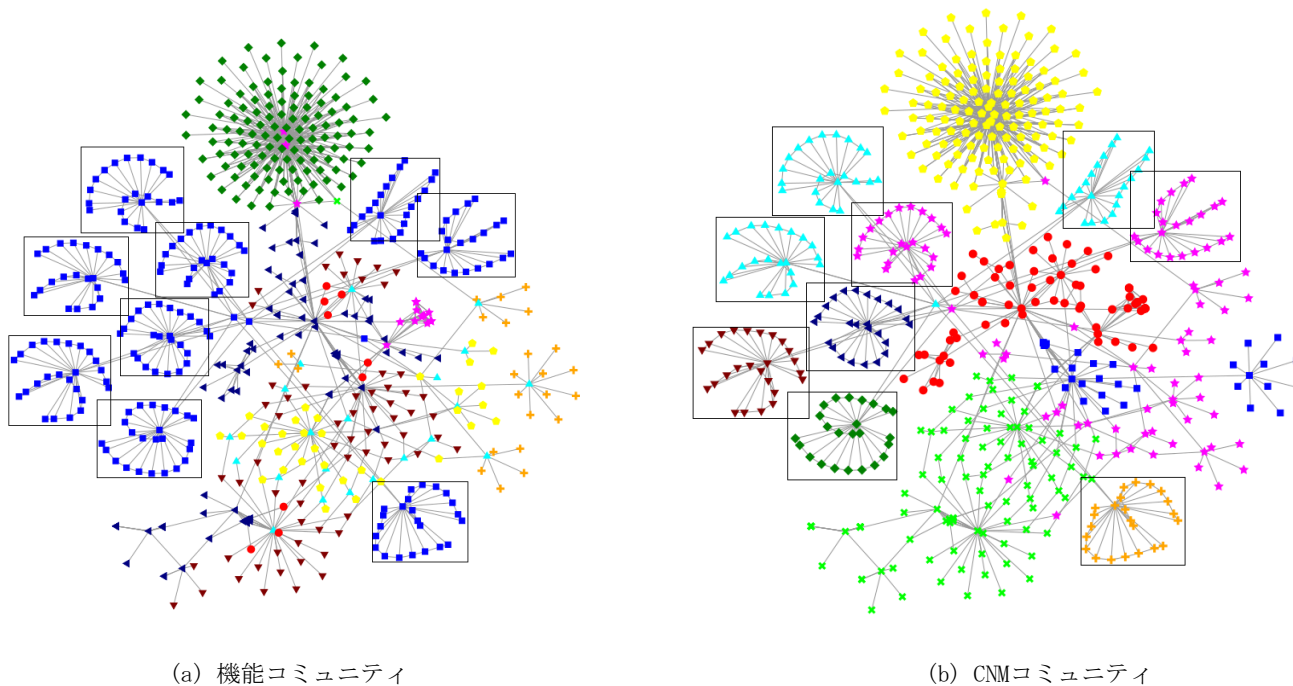


図3 Hoseiネットワークに対する処理結果

2 つ目のネットワークは、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、無向化して構築したハイパーリンクネットワークである。以下、Hoseiネットワークと呼ぶ。

3.2 実験結果と考察

上述したネットワークに対し、 $S = 1,000$ としFCE法により機能コミュニティを抽出した結果を示す。なお説明の便宜

上、ある特定のコミュニティ数 K を図示しているが、他の K の場合でもほぼ同様の結果が得られた。

Hierarchicalネットワークの結果を図2に示す。これらの図を比較すると、FCE法では階層上の同質（同一階層）のノード同士が同じ色に塗り分けられており、同質の機能を持つノード群が同一のコミュニティとして抽出されていることがわかる。会社組織で例えるなら、部長コミュニティが抽出できている。一方CNM法の結果は、リンク密度によるコミュニテ

ィ抽出のため、密に隣接するノード同士が同一のコミュニティとして抽出されていることがわかる。会社組織で例えるなら、部署コミュニティが抽出できている。

Hoseiネットワークの結果を図3に示す。これらの図を比較すると、FCE法では可視化結果の四角で囲まれた部分は同じ色で塗り分けられている。このノード群は、対象大学の各年度の教員の研究成果報告ページであり、ノードの機能としては同質であると考えられ、同一のコミュニティとして抽出できている。一方CNM法の結果は、年度ごとに異なるコミュニティとして抽出されていることがわかる。これは、異なる年度のノード同士が直接リンクされていないために、異なるコミュニティとして抽出されている。

4. おわりに

本研究では、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であるFCE法を提案した。従来のコミュニティ抽出手法であるCNM法との比較し、FCE法が抽出するノード群の性質を定性的に評価した。また、本論文では、FCE法を土台として、いくつかの拡張手法を提案している(図4参照)。

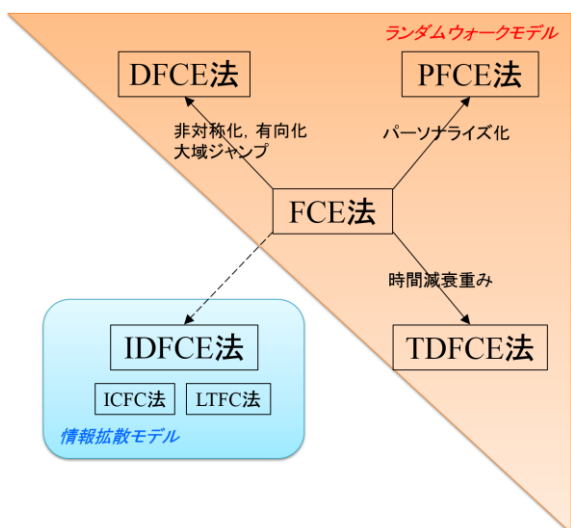


図4 FCE法とその拡張

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。評価実験により、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。

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。

FCE法による結果との比較により、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。

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LIFC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。

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ーを当てはめることができ、識別できた。

本研究では、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異なり、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、博士論文各章で提案したように様々な拡張を可能にした。今後は、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。

【文献】

[1] Clauset, A., et.al.: Finding community structure in very large networks, Physical Review E, Vol.~70, No.~6, pp. 066111+ (2004).
 [2] Borgatti, S.: Two algorithms for computing regular equivalence, Social Networks, Vol.~15, No.~4, pp. 361--376 (1993).
 [3] Leicht, E.~A., et.al.: Vertex similarity in networks, Physical Review E, Vol.~73, No.~2, pp. 1--10 (2005).
 [4] Jeh, G., et.al.: SimRank: a measure of structural-context similarity, Proc. of the 8th ACM SIGKDD international conf. on Knowledge discovery and data mining, pp. 538--543 (2002).
 [5] Christakis, N.~A., et.al.: Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives, Little, Brown and Company (2009).
 [6] Ravasz, E., et.al.: Hierarchical organization in complex networks, Physical Review E, Vol.~67, No.~2, pp. 026112+ (2003).