

# コミュニティ抽出法とその展望

湯田 聰夫

ネットワーク構造からリンクが密な部分集合を抽出することを、複雑ネットワーク科学ではコミュニティ抽出と呼ぶ。密につながる集団は、同じ属性や傾向をもつ集団であることが多いため有用な情報となる。古くはグラフ理論や数理社会学から近年のWebサイエンスまで幅広くコミュニティ抽出手法として全体を俯瞰し、近年提案された高速なニューマン法とその派生法を概説する。オンライン上には既に数千万から億の単位でユーザがつながっており、集団間の情報流や効果的なマーケティング法の研究などORにおいてもコミュニティ抽出法は有効と考え、その可能性を解説し、今後を展望したい。

キーワード：ネットワーク分析、クリーク、コミュニティ抽出、ニューマン法、SNS

## 1. はじめに

「コミュニティ (community)」という用語は、地域や同好の集まりを指す場合が多い。ただし、ここではあくまでも構造的にサブグラフ内のノード間リンク密度が相対的にサブグラフ外へのリンク密度よりも高いという特徴により規定されるものをコミュニティと定義する。

近年提案された多くのコミュニティ抽出法は、解析的には古くからあるデータの階層クラスタリング手法の一種と言えるが、M. E. J. Newmanはネットワーク・クラスタリングの解としてコミュニティという用語を意識的に分けて使用[1]し、多くの研究者が彼に倣っている。

構造的なコミュニティ抽出が注目されるのは、関係が密な集団を具体的な属性に依存せず構造的に抽出できるからである。抽出されたコミュニティには「その集団を形成する関係性」が反映されていると期待できる。すでに確認されたものとして、友人のネットワークでは人種・年齢・性別・趣味や職場、論文共著者ネットワークでは研究領域という関係性が、構造的に抽出された[1]。

さまざまな応用のひとつとして、例えば生命科学において、作用が未知のタンパク質も相互作用ネットワークから既知のタンパク質のノード属性と、構造的な稠密さから予測する可能性も示されている[2]。

図1に抽出のイメージ図を示す。図1(a)が元のデータ、抽出されたコミュニティ単位で濃淡をつけたものが図1(b)、コミュニティ単位でノードとエッジを束ねてコミュニティ間のネットワーク構造にしたものが図1(c)である。

近年のIT技術やWeb技術の進展に伴い、大規模な関係性ネットワークが獲得できるようになった。それを意味ある集団単位で縮約できる手法は、ORにも資するところがあるだろう。コミュニティ抽出法は2004年まで5千ノード規模が計算の限界だった。台頭著しいSNS(Social Networking Site)においては、既に1億人を超えるユーザがネットワーク構造を成して交友している。今までに、大規模ネットワーク構造のコミュニティ抽出、そしてその応用研究が展開されようとしている。

本稿は以下、次のような構成である。2節においてグラフ理論、数理社会学、Webサイエンスにおけるコミュニティ抽出法を概観し、3節においてNewmanらの手法を解説する。4節で高速化について触れ、5節でSNSの実データの分析結果を示し、6節ではコミュニティ抽出法の可能性と、今後の展望を記してまとめとする。文中のより詳細な説明は、文献[3]5

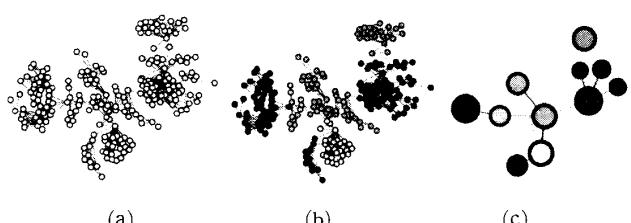


図1 コミュニティ抽出例

## 2. リンクが密な部分集合

### 2.1 クリーク関連

クリークとは、3個以上のノードの集合において、集合内のノード全てが互いにリンクし合っているような部分集合であり、コミュニティの考え方では最大のリンクの密度となる。クリーク自体は完全グラフとも呼ばれ、その部分集合は必ず全て完全グラフであるが、ここでは可能な部分完全グラフの中で最大のものをクリークと呼ぶ。ノード数  $k$  のクリークとなるには  $k(k-1)/2$  本のリンクがすべて結合している必要があるため、実際のグラフではサイズの大きなクリークが抽出されることは稀である。見た目に結びつきの強い集団があっても、その一部が小さなクリークとして検出されるにとどまる。よってこの厳しすぎる条件をいくぶん緩和したものが提案されている（解説[4]）。

クリークは全てのノードが必ず互いに距離1の範囲に含まれていることが明らかである。これを拡張し「互いに距離  $n$  以下の範囲にすべてのノードが含まれている集合」を半径  $n$  のクリークとして「 $n$ -クリーク」と呼ぶ。半径が1の場合は通常のクリークに他ならない。

クリークの自然な拡張に思われる  $n$ -クリークだが、すぐに違和感があるケースに気がつく。部分集合内のノードが全て距離  $n$  以下で到達できるとしても、その到達条件を満たす際に、部分集合の外を経由して距離  $n$  以下を実現している場合、取り出された部分集合だけでは、距離  $n$  以下で相互に到達できなくなる。図2では2-クリーク  $\{1, 2, 3, 4, 5\}$  の場合である。この不自然さを避けるため、 $n$ -クリークの定義にさらに条件を加えたものが提案されている。

「 $n$ -クラン ( $n$ -clan)」は、 $n$ -クリークのうち、部分グラフ内部を通った場合の距離では  $n$  を超えてしまうようなものを除いたものと定義される。図2では  $\{1, 2, 3, 4, 5\}$  は2-クリークだが2-クランではない。 $\{2, 3, 4, 5, 6\}$  は、2-クリークかつ2-クランである。「 $n$ -クラブ ( $n$ -club)」は、 $n$ -クリークからノードを削って、部分グラフ内での距離がすべて  $n$  以下とするものと定義される。図2では、 $\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{2, 3, 4, 5, 6\}$  が2-クラブとなる。 $n$ -クリーク、 $n$ -クラン、 $n$ -クラブは、ノード間距離で定義されているが、別なアプローチとしてリンク数を緩和する定義がある。

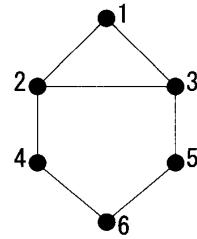


図2 クリーク関連

ノード数が  $N$  のクリークにおいては全てのノードが集合内に  $N-1$  本のリンクが必要であるのに対し、「 $k$ -プレックス ( $k$ -plex)」は集合内の全てのノードが少なくとも  $N-k$  本以上のリンクを有する部分集合として定義される。ここで、1-プレックスはクリークと等価である。

クリークに関する2005年にNature誌に掲載されたG. Pallaらの研究[2]に触れたい。彼らは、「クリーク・パーコレーション (clique percolation) 法」と後に呼ばれる手法を提案し、タンパク質の相互作用、語彙連想、共同研究者の3種の実データにおいて、クリーク同士がノードを共有して重なり合っている集合を抽出すると意味がある場合が多いことを示した。構成するノード数が  $k$  の2つのクリークが  $k-1$  のノードを共有している場合にこの2つは「接している」と定義する。互いに接しているクリーク全ての和を「 $k$ -クリーク-コミュニティ」と呼んでいる。 $k$ -クリーク-コミュニティは物理系の研究者による定義のため、先の  $n$ -クリークとは意味は異なることを補足しておく。

### 2.2 $k$ -コア関連

ORでクリークというとNP困難として知られる「最大クリーク問題」やその近似解法がある。以上述べてきた全ての計算は、大きいサイズや高密度では近似解法無しでは計算できない。比較的少ない計算量で、最大クリークのサイズの下限を与える「 $k$ -コア ( $k$ -core)」[5]に触れたい。部分ネットワーク内の全てのノードの次数が  $k$  以上になっている場合、その部分ネットワークは  $k$ -コアと呼ばれる。低次数ノードから刈り込み、その結果低次数になったノードも刈り込んでいくことで求められる。さらに下限を絞る手法として「 $k$ -デンス ( $k$ -dense)」[6]がある。 $k$ -デンスとは、隣接する任意のノード対が  $k-2$  個以上のノードを共有する部分ネットワークである。

以上は、個々のノードのリンク数といった値を直接

評価し定義するアプローチである。他に、ネットワーク全体に対する相対的な密度の高さを評価するアプローチとして、ラムダセット、LS セット、Bock Husain コミュニティ（解説[4]）や、min-cut (max-flow) 法（関連[7][8]）や、後述する Newman コミュニティ[9][10]などもある。

### 2.3 ブロックモデル

密にリンクしている集団の抽出について、数理社会学の分野では過去にパラダイムシフトがあった（解説[11]）。集団間の関係性の研究の流れを汲むが、密な集団を抽出するのではなく、構造の同値性を元に解析する手法の登場である。集団内にリンクがなくても、集団間の関係が同じなら、同一ブロックとして抽出される手法である。例えば企業の取引ネットワークにおいて、原料→生産→卸→小売と考えた場合、このブロック間に有向グラフが張られるが、基本的にブロック内のリンクは疎であろう。このブロック化とブロック間の関係性抽出が可能な手法である。

事前に設定する集団間の関係に向けて近づけていくように行列演算していくこともできる。つまり同じブロック内のリンク密度が高くなるように設定したブロック化の特殊解が、複雑ネットワーク科学でいうところのコミュニティ抽出といえる。参考書[4]にアルゴリズムの詳細は掲載されている。ダウンロードできる体験版<sup>1</sup>により、社会ネットワーク分析ツールを実際に触ってみることもできる。

### 2.4 Web コミュニティ抽出

Web サイエンスの領域では、Web ページをノードとし、ハイパーリンクをリンクとして、WWW のネットワーク構造を解析することで、同じトピックを有する Web ページの集合（=Web コミュニティ）を抽出する手法が数多く研究されている。Flake らの max-flow 法の派生法[7]とともに、Kleinberg の HITS アルゴリズム[12]におけるハブ・オーソリティの評価法多くの手法の源流を成している。

HITS 法は、Web の構造の特徴を巧みに利用している例として概説したい。多くのページからリンクされているページをオーソリティと呼び、多くのページをリンクしているページをハブと呼ぶ。この再帰的定義により評価する。

具体的には、まずあるキーワードに関連するウェ

ブ・ページを事前に収集していることが前提になる。あるページのリンク先は容易に分かるが、リンク元を探すのは多くの場合困難で、大規模なクローリングが必要になることが多い。全ページのオーソリティ値に初期値を与える、リンク元のオーソリティ値をリンク先にてハブ値として合計し値が拡散しないようベクトル長が 1 になるように規格化する。

次にこのハブ値をさらにリンク先のページのオーソリティ値として合計して正規化し、この操作を繰り返していくことで、全ページのハブ値とオーソリティ値が求まる。

高いオーソリティ値やハブ値のノードの隣接先から部分集合を得ると関連する Web ページの集合を得ることができ、これを Web コミュニティと呼んでいる。

Web コミュニティは関連するトピックを有する集団を意図しており、必ずしもネットワーク構造上でリンクが高密度なわけではないが、ハブとオーソリティをセットにした二部グラフ上での高密度な集合を形成するが多く、二部グラフのコミュニティ抽出研究とは関連が深い。

## 3. Newman コミュニティ

本節では、2008 年時点で、特に大規模ネットワークに対して有効性が高い抽出法と位置づけられる Newman 法とその派生法を少し詳しく紹介したい。優れた特質として、圧倒的に計算量のオーダーが小さい点、パラメータ調整が不要な点が挙げられる。ネットワーク構造データを与えるだけで分割結果を求めることができる。そこでまず、Newman 法がどのような経緯で提案され発展してきたのか、アルゴリズムとその変遷を概説する。

### 3.1 Girvan-Newman 法

2002 年、当時サンタフェ研究所に所属していた 2 人の物理学者 M. Girvan と M. E. J. Newman によって、エッジ媒介中心性 (edge betweenness (解説[3]2 章)) による新しいコミュニティ抽出法が提案[8]された。エッジ媒介中心性は、社会学者の L. Freeman により 1977 年にノードに対して提案された指標を拡張している。ネットワーク内の任意のノード間の最短経路を考えるとき、より多くのノード間の経路に選ばれるリンク（エッジと同義）ほど高くなるように定義されている。つまりコミュニティ間をより独占的につないでいるリンクほど高い値を得られる。かれらの提案手法は、切断法と呼ぶことができ、ネットワーク全体

<sup>1</sup> UCINET6

<http://www.analytictech.com/ucinet/ucinet.htm>

から最大の媒介中心性をもつリンクを除去し、残るネットワークでの媒介中心性の再計算と除去を繰り返し、最終的にコミュニティを浮かび上がらせようとする手法である。先行研究のように高密度な部分を直接発見しようとする点が新しく、クリアな結果を得ていた。

媒介中心性の計算は、2001年にNewmanが提案した高速なアルゴリズムにより計算時間のオーダーが $O(mn^2) \rightarrow O(mn)$ と劇的に改善していた。それを用いることで提案手法は実現でき、全体で $O(m^2n)$ の計算量であった。 $m, n$ はネットワーク全体のリンク数、ノード数に相当する。

全てのリンクの媒介中心性が算出されると、各ノード間結合ウェイトとして階層クラスタ化できる。得られるデンドログラム(*dendrogram*)のどの深さで分割しても、その深さにおけるコミュニティ分割が得られる(解説[3]2章)。

### 3.2 モジュラリティ

どの段階でもコミュニティが抽出できるということは、どこが最良の分割条件か分からぬともいえる。この分割終了条件を明確にするために、2004年 NewmanとGirvanは、モジュラリティ(*Modularity*) $Q$ という指標を提案した。

ネットワーク全体が $V_1, V_2, \dots, V_L$ と $L$ 個の重複しないコミュニティに分割された際、 $Q$ は次のように定義される。

$$Q = \sum_{l=1 \dots L} Q_l = \sum_{l=1 \dots L} (e_{ll} - a_l^2) \quad (1)$$

$e_{ll}$ は $V_l$ 内部のリンクの存在確率を意味し、 $a_l$ は、無向グラフであっても敢えて出・入リンクとして「リンク端」を分けて考えたとき、 $V_l$ 内にあるリンク端総数のネットワーク全体に対する存在確率を意味し、それぞれ次の式で得られる。

$$e_{ll} = \frac{1}{2m} \sum_{i \in V_l} \sum_{j \in V_l} A(i, j) \quad (2)$$

$$a_l = \frac{1}{2m} \sum_{i \in V_l} \sum_{j \in V} A(i, j) \quad (3)$$

$A(i, j)$ は、ネットワークの隣接行列で、ノード $i, j$ 間にリンクがあると1、なければ0を返し、ネットワークのリンク端の総計は $2m$ である。

$Q$ は極めて強力な分割指標である。 $e_{ll}$ は各コミュニティ内部の密度がそれぞれ高いことを求め、 $a_l$ は全体をひとつのコミュニティにしている場合や、ランダムな分割に対して $Q$ を下げる補正項として導入されている。 $a_l \times a_l$ とはネットワーク全体からランダムに1本リンクを取り出した際に、 $V_l$ から出てくる

リンク端密度と $V_l$ に入るリンク端密度の積、つまり $V_l$ の内部のリンク密度を意味する。よって、有意にコミュニティとして内部に密度の高い $V_l$ が分割されていないと高い $Q_l$ は得られず、またどこかの分割だけ高くして他が低い場合も、全体の $Q$ を高くはできない。

モジュラリティはあくまで「分割されたネットワーク」の評価指標として働き、どのような分割手法であっても、分割結果の程度を比較できる。一般のネットワークにおいて、有意なコミュニティが抽出された際、モジュラリティは $0.3 < Q < 0.7$ 程度になることが報告されている[13]。

### 3.3 Newman法

Newmanは、モジュラリティを提案した直後に、斬新なコミュニティ抽出法を提案した[9]。提案された手法は、媒介中心性を使わず、モジュラリティを最大にするようにリンクを追加していく手法だった。開始時点で各ノードを内包ノード数が1の仮コミュニティとして存在させ、まず $Q_{initial}$ を求める。次に、全リンクに対して結合した場合の $Q$ の増加分 $\Delta Q_{ij}$ を求め、最大の $\Delta Q$ を与えるリンクを介している仮コミュニティ同士を結合する。結合による変化を反映して $\Delta Q$ を再計算していく山登り法(*hill climbing algorithm*)であった。計算が進むにつれ仮コミュニティの結合が進み、内包ノード数が大きくなっていく。そして最大の $Q$ の時点で、残る全ての $\Delta Q_{ij}$ が負となって計算が終わる。 $\Delta Q_{ij}$ の計算は、

$$\Delta Q_{ij} = 2(e_{ij} - a_i a_j) \quad (4)$$

となる。媒介中心性の手法の計算量のオーダーが $O(m^2n)$ であったが、Newman法は $O((m+n)n)$ まで効率化された。

ここで、モジュラリティを分割過程に用いて、モジュラリティ最大化を目指すアルゴリズムが抽出するコミュニティを、Newmanコミュニティ(NC)と呼ぶことにしたい。

## 4. 抽出の高速化

### 4.1 Clauset-Newman-Moore(CNM)法

A. Clauset, M. E. J. Newman, C. Mooreにより提案されたCNM法[10]は、Newman法と計算結果は全く同等であるが、計算上の改善が施されている。先のNewman法では $\Delta Q_{ij}$ の更新ごと、 $e_{ij}, a_i, a_j$ を更新する必要があったが、計算量の少ない $\Delta Q_{ij}$ と $a_i, a_j$ の更新だけで計算できることを発見し、結果とし

て計算量のオーダーが実質的に  $O(n \log^2 n)$  になることが期待された。事実、それまで 5,000 ノードを超えたネットワークからのコミュニティ抽出はツールが存在しなかったが、10 万ノード規模での解析ができるようになった。

#### 4.2 Danon-Diaz-Arenas (DDA) 法

L. Danon, A. Diaz-Guilera, A. Arenas は、仮コミュニティの結合時に、リンクの含有量が大きいものに小さいものが取り込まれていくプロセスを確認した。より高い  $Q$  を得るために、 $\Delta Q_{ij}$  更新の際、結合の両端のリンク含有量が同程度である結合が優先されるようにして、わずかにより高い  $Q$  を得た。速度は CNM 法と変わらない程度と報告された[14]。より高い  $Q$  を求める研究は他にも盛んで、サイズは数万までの適用となるが、Newman による新しい手法が高性能として報告されている[15]。

#### 4.3 DDA の改良法

筆者らは、DDA 法は原理的に高速化にも向いていることを発見し、異なるアルゴリズムの性能を比較検討するためのプラットフォームを考案し、新しいデータ構造と DDA 法の改良を提案した[16]。結果として 100 万ノードでオリジナル CNM の 192 倍の高速化を達成し、同時に  $Q$  の向上も得た。この時点で 2,000 万ノード、1 億リンクまでは計算可能となった。現在も数億ノード、数百億リンク規模の解析を目指す高速化研究が展開されている。

### 5. SNS のコミュニティ分析

大規模なネットワーク構造データのコミュニティ分析例として筆者らが行った SNS の分析[17]を紹介する。

対象は国内最大規模の SNS における交友関係ネットワークで、初期の安定成長しているときのデータであった。分析した最大連結成分は、 $n=360,802$  で  $m=1,904,641$  であった。次数分布にスケールフリー性があり、局所凝集性を知るための平均クラスタ係数が  $C=0.330$  と極めて局所に凝集していることが分かった。ネットワーク全体の平均経路長は  $L=5.53$  とサイズに比べ極端に小さく、局所の凝集性を合わせてスマートワールド性があることが分かる。

このデータに CNM 法により Newman コミュニティ抽出を行った。コミュニティ単位の関係を可視化したもののが図 3(a)である。コミュニティのサイズは内包しているノード数の対数に比例し、濃淡は視認性向上

のために付けている。描画そのものは全コミュニティ間に斥力、リンクにはバネを設定し物理モデル計算を行ってレイアウトしている。200 万本近いリンクを全て描画していたら、このような認識しやすい構造に可視化はできない。コミュニティ抽出により粗視化できることの例である。

図 3(a)が示すように、中心にある極めてサイズの大きいコミュニティと、左上にある中規模のもの、そして数多く存在する小規模のものと、サイズに偏りがあることが分かる。サイズとその順位を両対数でプロットすることで、より詳細に示したのが図 3(b)である。このギャップは、ノード内包数が 80~200 となるコミュニティがわずかしか抽出されなかったことを意味する。また大規模が 4 つ右下方に位置し、ギャップの左側に小規模なものが位置している。

NC サイズ分布におけるギャップは、筆者らの SNS 研究においてはじめて確認された。Amazon の購買の共起ネットワーク[10]や、他のコミュニティ抽出結果には存在しないギャップの正体を突き止めるべく、構造特性の分かっているネットワークモデルを用いて検証していった。

表 1 には、SNS とネットワークモデルのネットワーク指標比較結果を示した。以下ではその結果の主要な部分を解説する。まず、SNS が有しているスケールフリー性やスマートワールド性を実現しても、NC サイズ分布にギャップは確認されなかった。ここで改めて図 3(b)を見ると、両対数で全体が傾き -1 の直線に概ね従っていることが分かる。この分布に従うことを Zipf 則といい、都市のサイズと順位の分布など、様々なところに潜んでいる分布である。CNN の NC 分布はギャップこそ無いが Zipf 則に完璧に従っていた。A. Vazquez により提案された CNN モデル[18]は成長型のモデルであり、確率  $1-\mu$  でノードの追加、確率  $\mu$  で凝集化する。新しくノードを追加する際には、既存のノードから一様ランダムに 1 つ選んで結合し、その既存ノードの隣接先に「潜在リンク (Potential Edge)」を張る。確率  $\mu$  で潜在リンクを 1 本選んで顕在化させる。このとき、必ずリンクの三角形を 1 つ生成し凝集性を上げていく。

では、SNS にはあって CNN モデルには存在しないものは何であろうか。実際の SNS を観察すると、趣味や同好の集まりであるオンライン・グループや、検索機能を介して普段接点の無い昔の友人との交流や新しい出逢いを楽しむユーザが多いことに気がつく。

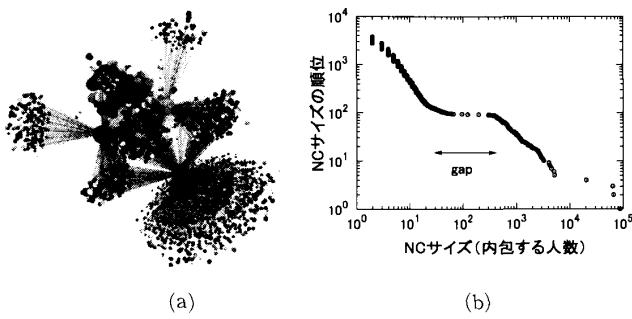


図3 (a) SNS-NC 粗視化と(b) NC サイズ分布

表1 SNS とモデルのネットワーク指標比較

network	r	C	Nnc	Q	SF	SW	Zipf	Gap
SNS	-0.121	0.346	5,032	0.591	+	+	+	+
BA	-0.009	0.000	24	0.257	+	-	-	-
WS	0.222	0.373	68	0.685	-	+	-	-
CNN	0.100	0.398	1,062	0.694	+	+	+	-
CNNR ( $r=8\%$ )	0.124	0.346	5,032	0.591	+	+	+	+

モデルは全て SNS のネットワークサイズと同等。

BA: Barabási-Albert モデル, WS: Watts-Strogatz モデル, CNN: Connecting-Nearset-Neighbor モデル, r: 次数相関係数, C: 平均局所クラスタ係数, Nnc: 抽出 NC 数, SF: Scal-Free 性, SW: Small-World 性, Zipf, Gap: NC サイズ分布における Zipf 則, ギャップの有無, 詳細は文献[3][17]

これはネットワーク構造的には、先の CNN を保ちながら経路上離れたノード同士が、ランダムにつながり凝集していくことを意味する。その振る舞いをモデル化し、CNNR (*CNN with Random linkage*) モデルと名づけた。このモデルの NC 分析結果を図4(b)に示している。図3(a), 図3(b)と比較しても、よく再現できていることが分かる。表1で確認できるように、各種の指標も再現できている。CNNR モデルはネットワークサイズが固定されると  $\mu$  が固定されるため、ランダム度  $r$  だけが可変となる。図4(b)に示すように、この  $r$  を大きくするほどギャップが大きくなることも確認した。

ではこの  $r$  の意味は何であろうか？ SNS 上での検索機能やオンライン・グループ機能が、まさにこの  $r$  を大きくしている。SNS誕生前とは比較にならない少ない労力で旧友と交流し、ニッチな同好の仲間と出会う機会を創出し、物理的な場所や時間から比較的自由に交流できるようになってきている。このギャップと、モデルで把握できる  $r$  は、オンライン上で新しく興っているコミュニケーションの本質的な一側面を

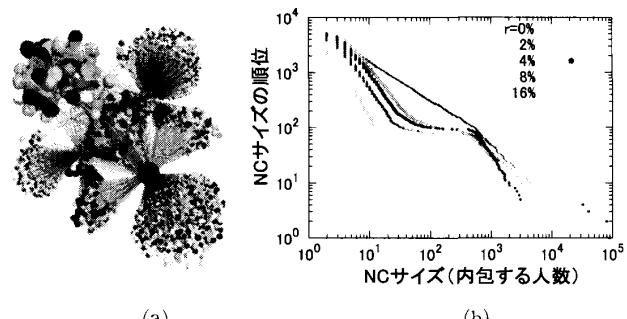


図4 (a) CNNR-NC 粗視化と(b) NC サイズ分布

捉えていると筆者らは考えている。

## 6. 可能性と将来展望

大規模ネットワークのコミュニティ抽出法の研究はその端緒についたにすぎない。モジュラリティを向上させるコミュニティ抽出が、その指標のセンスのよさと高速な応用アルゴリズムから多く用いられているが、万能ではないことも明らかになってきている。S. Fortunato らは、モジュラリティ自体が原理的にもつコミュニティの解像度があり、相対的に小さいコミュニティが抽出されないと報告した[19]。1 本のリンクで結合した 2 つのクリークが内部にあると、合わせて 1 つのコミュニティとして抽出される。

しかしクリークなど密な集団を「必要以上に分割」することは原理的に無いので、再帰的に抽出を繰り返することで、必ず全体に対して相対的に密な集団を抽出できる手法であると筆者らは考えている。大規模な人のネットワークから集団が抽出されると、その集団の属性や履歴を反映した推薦や検索などが行える可能性があり、個人やその身近な集団に焦点を当てたサービスにとっては、特に重要な技術的基盤技術となっていくと考えている。

Web 上には有向グラフの情報が多く、またその向きには重要な意味が含まれているため、有向グラフへの拡張も熱心に研究されているが、高速なものはいまだ十分ではない。

10 年後、20 年後を展望するとき、有向グラフ、二部グラフ、時系列解析、属性も含めた多元な解析、特に規模の制約が入ると、アルゴリズムとしての高速化が必須となり、いまだめども立たない挑戦的な研究課題は多い。

OR には、数理もアルゴリズムも実装も強く、さらに実践的な応用に強い関心のある研究者が多いと聞く。

ORの知識と経験が大規模ネットワークの構造解析に深く入ってくること、また実践的なORの分野へ、より貢献できる日が来ることを、心から楽しみにしている。

### 参考文献

- [1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, **45**, 167–256, 2003.
- [2] G. Palla, I. Derényi, I. Farkas and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, **435**, 814–818, 2005.
- [3] 林幸雄(編著), 大久保潤, 藤原義久, 上林憲行, 小野直亮, 湯田聰夫, 相馬亘, 佐藤一憲(著)「ネットワーク科学の道具箱」, 近代科学社, 2007.
- [4] S. Wasserman and K. Faust, "Social Network Analysis," *Cambridge University Press*, 1994.
- [5] S. B. Seidman, "Network structure and minimum degree," *Social Networks*, **5**, 269–287, 1983.
- [6] 齊藤和巳, 山田武士, 風間一洋, "k-dense法によるネットワークのコア部分抽出," 第2回ネットワーク生態学シンポジウム, 2006.
- [7] G. Flake, S. Lawrence, C. L. Giles and F. M. Coetzee, "Self-organization and identification of web communities," *Computer*, **35**, 66–70, 2002.
- [8] M. Girvan and M. Newman, "Community structure in social and biological networks," *PNAS*, **99**(12), 7821–7826, 2002.
- [9] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, **69**, 066133, 2004.
- [10] A. Clauset, M. E. J. Newman and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, **70**, 066111, 2004.
- [11] L. C. Freeman, "The Development of Social Network Analysis: A Study in the Sociology of Science," *Empirical Press*, 2004. (邦題: 社会ネットワーク分析の発展 邦訳: 辻竜平 出版: NTT出版, 2007.)
- [12] J. M. Kleinberg, "Authoritative sources in a hyper-linked environment," *J. ACM*, **46**(5), 604–632, 1999.
- [13] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, **69**, 026113, 2004.
- [14] L. Danon, A. Diaz-Guilera and A. Arenas, "Effect of size heterogeneity on community identification in complex networks," arXiv: physics/0601144, 2006.
- [15] M. E. J. Newman, "Modularity and community structure in networks," *PNAS*, **103**, 8577–8582, 2006.
- [16] Y. I. Leon-Suematsu and K. Yuta, "A Framework for Fast Community Extraction of Large-Scale Networks," *17 th International WWW Conference*: www2008.org/papers/pp253.html, 2008.
- [17] K. Yuta, N. Ono and Y. Fujiwara, "A gap in the community-size distribution of a large-scale social networking site," arXiv.org: physics/0701168, 2007.
- [18] A. Vazquez, "Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations," *Phys. Rev. E*, **67**, 056104, 2003.
- [19] S. Fortunato and M. Barthélémy, "Resolution limit in community detection," *PNAS*, **104**, 36, 2007.