社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ICS－136　(19)
2004／8／5

帰無仮説としてのランダムネットワーク

藤原 義久[1], 湯田 聰夫, 下原 勝憲

ATR ネットワーク情報学研究所, 〒 619-0288 けいはんな学研都市

帰無仮説としてランダムネットワークを利用することについてレビューを行い，その応用とともに発表する．(1) ランダムネットワークのさまざまな量を計算する母関数法について述べ，二部グラフへの適用例をあげる．(2) betweenness というネットワークの中心性を測る量を使い，ネットワークのコミュニティ (クラスター) 構造を発見する際に，コミュニティ間の連結性をランダムな連結からのずれとして測る方法について述べる．ここまではレビューであり，紙面の都合上限られた形で述べる．講演では，これらの手法を応用したオリジナルな研究を，人的ネットワークと社会経済的な関係性ネットワークについて述べる．

Random network as a null hypothesis

Yoshi Fujiwara[1], Kikuo Yuta, Katsunori Shimohara

ATR Network Informatics Laboratories, Kyoto 619-0288

We review on utlization of random network as a null hypothesis, and talk about its applications. Reviewed are (1) generating-functional method for calculating several topological quantities of network with application to bipartite graph, (2) quantification of inter-communicties connectivity in finding communicty or cluster structure by centrality measure of betweenness. In the workshop, we will talk about our original work of application to human network and social-economic network.

1　はじめに

「ランダムなネットワーク＝次数 (degree) 分布が Poisson 分布である」と単純に誤解している人がいるが，そのようなランダムなネットワークは，(以下にも述べるように) 特殊なクラスである．何をもって「ランダム」とするかは，どのような仮説の下で何をランダムと考えるかによって当然異なってくる．2 つのノード間を結ぶエッジ があるかないかが，他のエッジ の存在とは独立に一定の確率で与えられるようにして作られた，ランダムなネットワークが，上記等号の右辺となる．

一方，次数分布というのはネットワークのトポロジーを特徴づける上で，ノードについて「1 次」の構造にすぎない．その意味は次のような操作を考えると分かる．簡単のために単純な (すなわち自己ループも多重エッジもない) 連結無向グラフで記述できるネットワークがあるとする．全エッジ それぞれをハサミで切り取り，異なるノードをつなぐという制限以外，まったくでたらめにエッジ をノリでひっつける．この操作で次数分布は完全に不変であるが，複数のノード間の連結性 (例えばクラスター係数など) に自明でない高次構造があったとしてもそれは壊される．

これを逆手にとれば，次数分布という 1 次構造だけを保存しそれ以外はでたらめであるという仮説の下で作られたランダムネットワークは，観測されたもとのネットワークの高次構造を調べるための帰無仮説として使えることを意味する．ごく最近，与えられた次数分布をもつランダムネットワークについて，さまざまなトポロジカルな量を求める手法が急速に発展してきた [2, 3].

本報告では，その手法 [3] について解説をして，特に二部グラフへの適用例と，betweenness を用いたネットワークのコミュニティ (クラスター) 解析への応用についてレビューを行う．研究会講演では，これらの手法を応用したオリジナルな研究を，人的ネットワークと社会経済的な関係性ネットワークについて発表する [10].

2　ランダムグラフ

任意の次数分布をもつ十分大きな無向グラフを考える．次数分布以外のトポロジカルな性質に関してランダムであるとする．すなわち，与えられた次数分布から全ノードの次数が独立同分布 (i.i.d.) として実現され，その次数系列 (degree sequence; [1]) をもつ可能なすべてのグラフから一様ランダムに一つグラフが選ばれる．この操作により生成されるすべてのグラフを考える．以下に述べる平均や統計量は，そのアンサンブルについて計算されるとする[2].

全ノードから一様ランダムにノードを選んだとき，

[1]Contact: yfujiwar@atr.jp .

[2]次数系列を一つ固定してアンサンブルを構成することもできる (例えば [2]). 統計力学でいう「ミクロカノニカル」アンサンブルに相当するが，グラフのサイズが無限大の極限では次数系列はすべて実現されつくされるので，どちらも同じ結果を与えると考えられる．

そのノードの次数が k である確率を p_k とする．分布 p_k に対して母関数を

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \tag{1}$$

で定義する．規格化は $G_0(1) = 1$ で与えられる．母関数が求まれば，逆に次数分布は

$$p_k = \frac{1}{k!} \left. \frac{d^k G_0(x)}{dx^k} \right|_{x=0} \tag{2}$$

と求まる．次数についての統計量は母関数から直接計算できる．例えば，次数のモーメントの平均は

$$\langle\, k^n \,\rangle = \sum_{k=0}^{\infty} k^n p_k = \left(x \frac{d}{dx} \right)^n G_0(x) \Big|_{x=1} \tag{3}$$

特に次数の平均は $\langle\, k \,\rangle = G_0'(1)$ である．

全エッジ から無差別にエッジ を選ぶ．その端点から出るエッジ 数 (選ばれたエッジも含んで) が k になる確率は kp_k に比例する．このこと (*) に注意すると，次の確率に対する母関数が導入できる．全ノードから一様ランダムに 1 つノードが選ばれたとする．この条件の下で，そのノードから出るエッジ をランダムに選んでたどり，最隣接ノードにいく．そのノードの outgoing edge(たどってきたエッジ を除く残りのエッジ) の数が k となる確率は，$(k+1)p_{k+1}$ に比例するので，この確率を生成する母関数は

$$\frac{\sum_{k=0}^{\infty}(k+1)p_{k+1}x^k}{\sum_{m=0}^{\infty} p_m} = \frac{G_0'(x)}{G_0'(1)} \equiv G_1(x) \tag{4}$$

となる (分母は規格化 $G_1(1) = 1$ のための因子である)[Fig. 1].

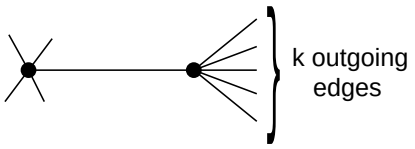

Figure 1: 最隣接ノードの outgoing edge 数.

閑話休題 (*) この確率が p_k ではなく，kp_k に比例するという事実は重要である．あなたが知合いのネットワークからでたらめに選ばれたとする．あなたの知合いの中から一人をランダムに取り出す操作は，実はランダムサンプリングではないことを意味する．なぜなら次数のより大きな人は，まさにその性質により，あなたの知合いである確率が高くなるからである．別の言い方をすると，孤高の人と超人気者は，人をたどるというこのやり方では異なる確率でバイアスを受けてサンプリングされる．

$G_1(x)$ が最隣接ノードからの outgoing edge 数の分布を生成する母関数を与えたが，それから 2 番目の最隣接ノードの総数の確率を生成する母関数をそれから構成できる．そのために母関数のべき乗を利用する．

例えば 2 つのノードを互いに独立に取り，そのエッジ 数の総数が k である確率を生成する母関数は，$[G_0(x)]^2 = \sum_{jk} p_j p_k x^{j+k}$ である[3]．なぜなら，それを展開した x^n の係数には，$p_j p_{n-j}$ $(j = 0, 1, \cdots, n)$ の組合せがすべて現れるからである．一般にある object のもつ特性量 k の分布を生成する母関数が与えられたとき，m 個の独立な object の実現に対応する，特性量の和の分布を生成する母関数はもとの母関数の m 乗で与えられる．

全ノードから一様ランダムに 1 つノードが選ばれたとする．このとき最隣接ノード数が k であるという条件の下で，それら最隣接ノードから出る outgoing edge の総数が取る値の分布を生成する母関数は $[G_1(x)]^k$ であるから，2 番目の最隣接ノードの総数の分布を生成する母関数は

$$\sum_k p_k [G_1(x)]^k = G_0(G_1(x)) \tag{5}$$

で与えられる Fig. 2[4].

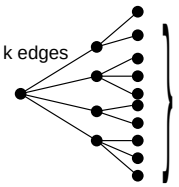

Figure 2: 2 番目の最隣接ノードの outgoing edge 数.

同様にして，3 番目の最隣接ノードの総数に対する母関数は，$G_0(G_1(G_1(x)))$ となる．一般に m 番目の最隣接ノードに対しての母関数を $G^{(m)}(x)$ と表すと，$G^{(1)}(x) = G_0(x)$ かつ，$m > 1$ に対して $G^{(m)}(x) = G^{(m-1)}(G_1(x))$ と再帰的な関係式から決定できる．m 番目の最隣接ノードの総数の平均 z_m は，それら母関数から $z_m = (d/dx)G^{(m)}(x)|_{x=1}$ により求まり，$z_m = z_1(z_2/z_1)^{m-1}$ で与えられる．z_1 と z_2 だけで決まるのは，独立性を仮定したからであり，実際には m が大きくなるにつれ，グラフの有限サイズを占めるようになるので，それは悪い近似になると考えられる．

しかし仮に，ランダムに選んだノード pair 間の最短経路の典型的な長さ ℓ を，思い切って $1 + \sum_{m=1}^{\ell} z_m = N$ (N は全ノード数) として評価すると，$N \gg z_1, z_2 \gg z_1$ の場合に，$\ell = [\frac{\ln(N/z_1)}{\ln(z_2/z_1)}] + 1$ を得る．

[3] ランダムグラフは十分に大きく，それらのノードがエッジ を共有する確率はノード数の逆数に比例して十分に小さく無視できると仮定している．これは以下の，m 番目の最隣接ノードの話でも同様である．

[4] $[G_1(x)]^k$ を展開した係数は，すべて条件付き確率を表しており，それぞれに p_k という確率が乗算されると考えると分かりやすい．

母関数を用いた手法は，連結成分のサイズ分布・その統計量・巨大成分の出現と相転移・クラスタ係数などを含む，ランダムグラフの性質を調べるのに応用できる [3].

閑話休題 ランダムに選んだノードの最隣接の最隣接，すなわち知合いの知合い (“2-link”) のサイズの平均は

$$z_2 = G_0''(1) = \langle\, k^2 \,\rangle - \langle\, k \,\rangle^2 = \langle\, k \,\rangle^2 - \langle\, k \,\rangle + \sigma_k^2 \tag{6}$$

で与えられる．ここで $\sigma_k^2 = \langle\, k^2 \,\rangle - \langle\, k \,\rangle^2$ は次数分布の分散である．この表式は，導出の際に述べたように，知合いの知合い間で重複がないことを前提にした近似的なものであるが，2-link のサイズに関して重要である．平均次数 $z_1 = \langle\, k \,\rangle$，すなわち平均的な知合いの数が 100 人であるとしよう．知合いの知合い全体の総数をざっと評価するには，${z_1}^2$ すなわち 10,000 人とできる．もちろんこれはオーバーラップがあるので，実際にはこれよりも少ないと考えられる．重複のために実際には少ないという点は正しいが，べき分布のように次数分布の裾野が長い場合には，最初の評価で「少なく」見積もってしまっており，その場合，重複を考慮しても，10,000 人よりも大きなサイズになることが多いのである．これは (6) において，$\sigma_k^2 \gg \langle\, k \,\rangle$ であることによる．2-link のサイズは naive に考えるよりも大きい ([4])!.

3 例

3.1 二項分布

$$p_k = \binom{n}{k} p^k (1-p)^{N-k} \tag{7}$$

の場合

$$\begin{aligned} G_0(x) &= \sum_{k=0}^{N} \binom{n}{k} p^k (1-p)^{N-k} x^k \\ &= (1+px-p)^N \to e^{z(x-1)} \end{aligned} \tag{8}$$

ここで最後に $N \to \infty$ かつ $Np \to z =$ 定数の極限を取った．このとき次数分布は，$p_k = (1/k!) z^k e^{-z}$ なる Poisson 分布で与えられる．

$$G_1(x) = e^{z(x-1)} = G_0(x) \tag{9}$$

Poisson 分布を次数分布にもつランダムネットワークは，$G_1(x) = G_0(x)$ なる性質をもつ[5]．すなわちランダムにノードを選んでも，ランダムに選んだエッジ をたどってノードを選んでも，同じ次数分布が得られるという特殊な性質がある．

[5]逆に $G_1(x) = G_0(x)$ を $G_0(x)$ に対する微分方程式とみて，条件 $G_0(1) = 1$ の下で解くと，Poisson 分布が確率分布となる．

3.2 指数分布

$p_k = (1 - e^{-1/\kappa}) e^{-k/\kappa}$ の場合

$$G_0(x) = \frac{1 - e^{-1/\kappa}}{1 - e^{-1/\kappa} x} \tag{10}$$

$$G_1(x) = \left[\frac{1 - e^{-1/\kappa}}{1 - e^{-1/\kappa} x} \right]^2 \tag{11}$$

3.3 べき分布 (cut-off を含む)

$p_k = C\, k^{-\tau}\, e^{-k/k_0}$ $(k \geq 1)$．ここで k_0 は cut-off で，C は規格化定数である．polylogarithm という特殊関数

$$\mathrm{Li}_\tau(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^\tau} \tag{12}$$

を用いると，$C^{-1} = \mathrm{Li}_\tau(e^{-1/k_0})$ と書ける．母関数は

$$G_0(x) = \frac{\mathrm{Li}_\tau(e^{-1/k_0} x)}{\mathrm{Li}_\tau(e^{-1/k_0})} \tag{13}$$

また，$(d/dx)\mathrm{Li}_\tau(x) = x^{-1}\mathrm{Li}_{\tau-1}(x)$ に注意すると

$$G_1(x) = \frac{\mathrm{Li}_{\tau-1}(e^{-1/k_0} x)}{x \mathrm{Li}_\tau(e^{-1/k_0})} \tag{14}$$

3.4 経験分布

実際に観測されるネットワークでは，まず次数分布の経験分布が与えられるのが普通である．つまり次数が k であるようなノードの数 n_k が観測される．このように任意の分布に対しては，経験分布を用いて

$$G_0(x) = \frac{\sum_{k=0}^{\infty} n_k x^k}{\sum_{k=0}^{\infty} n_k} \tag{15}$$

と (4) で母関数を構成できる．

4 二部グラフの縮約への応用

映画共演，論文の共著や複数企業の取締役会など，二部グラフとして表現できるネットワークは多い Fig. 3 (a)．母関数の応用例として，二部グラフを縮約したグラフの次数分布を求める．

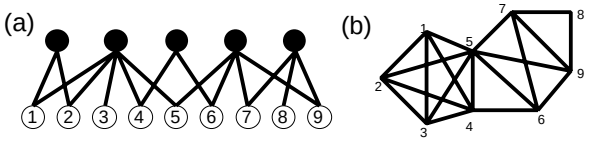

Figure 3: (a) 二部グラフ: 黒丸が「映画」, 白丸が「俳優」を表す．(b) 同じ「映画」を共有する「俳優」にリンクをはることで縮約したグラフ.

説明を分かりやすくするため，Fig. 3 (a) にあるように，二部グラフのノードの一方の種類を「映画」,

他方を「俳優」と表現することにする．他の例の場合には，論文や取締役会であり，共著者や役員である．二部グラフは，Fig. 3 (b) に示したように一部グラフに縮約できる．同じ映画・論文・取締役会に参加した人の間に互いにリンクをはるという操作を行えばよい[6]．

俳優 i が出ている映画の数を $d_i^{(a)}$，俳優の総数を N とする．また映画 i に共演している俳優の数を $d_i^{(f)}$，映画の総数を M とする．俳優一人当りの平均出演映画数を μ, 映画一つ当りの平均共演者数を ν とする．このとき定義から，$\mu = (1/N)\sum_{i=1}^{N} d_i^{(a)}$, $\nu = (1/M)\sum_{i=1}^{M} d_i^{(f)}$ であり，俳優と映画の間のエッジ総数はどちら側からみても同じだから，$\mu/M = \nu/N$ である．

ランダムに俳優を選んだとき，その俳優の出演している映画数が j である確率分布を p_j とする．また，ランダムに映画を選んだとき，その映画の共演者数が k である確率分布を k とする．それぞれに対する母関数を

$$f_0(x) = \sum_{j=0}^{\infty} p_j x^j \tag{16}$$

$$g_0(x) = \sum_{k=0}^{\infty} q_k x^k \tag{17}$$

と表す．確率の規格化から $f_0(1) = 1 = g_0(1)$ であり，平均の定義から $f_0'(1) = \mu$, $g_0'(1) = \nu$ である．

いま，全俳優からランダムに一人を選び，出演している映画からランダムに一つを選んだとき，その映画の他の共演者の総数 (すなわち outgoing edge) の分布を生成する母関数は，(4) を導出した同じやり方を用いて

$$g_1(x) \equiv \frac{1}{\nu} g_0'(x) \tag{18}$$

で与えられる (Fig. 4 (a))．同様に，全映画からランダムに一つ選び，その中の一人の俳優をランダムに選んだとき，その俳優の他の出演映画の総数 (outgoing edge) の分布を生成する母関数は $f_1(x) \equiv (1/\mu) f_0'(x)$ である．

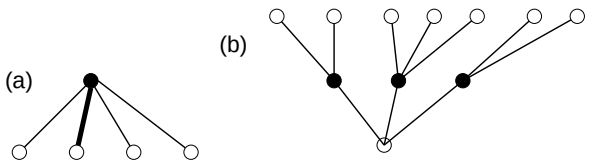

Figure 4: (a) ランダムに選んだ俳優の出演した映画からの outgoing edge (b) ランダムに選んだ俳優の共演者総数

次に，Fig. 4 (b) にあるように，ランダムに選んだ俳優の 2 番目の最隣接ノードの総数，つまりランダムに選んだ俳優の共演者の総数を与える分布を生成する母関数は，(5) を導出した同じやり方を用いて

$$f_0(g_1(x)) \equiv G_0(x) \tag{19}$$

となる．すなわち，Fig. 3 (b) にある，縮約した一部グラフの次数分布がランダムグラフという帰無仮説の下で計算できる．(映画の方についても同様な母関数が求められる)．

簡単な例として，p_j も q_k も Poisson 分布であるような場合を考える．このとき例 (1) で示したように，$f_0(x) = \exp[\mu(x-1)]$, $g_0(x) = \exp[\nu(x-1)]$, かつ $f_1(x) = f_0(x)$, $g_1(x) = g_0(x)$ である．縮約したグラフの次数分布を生成する母関数は

$$G_0(x) = \exp[\mu(e^{\nu(x-1)} - 1)] \tag{20}$$

であり，次数分布 $r_k = (1/k!)(d/dx)^k G_0(x)|_{x=0}$ を求めると

$$r_k = \frac{\nu^k}{k!} \exp[\mu(e^{-\nu}-1)] \sum_{m=1}^{k} \left\{ \begin{array}{c} k \\ m \end{array} \right\} [\mu e^{-\nu}]^m \tag{21}$$

ここで

$$\left\{ \begin{array}{c} n \\ k \end{array} \right\} = \sum_{s=1}^{k} (-1)^{k-s} \frac{s^n}{s!\,(k-s)!} \tag{22}$$

は n 個の番号付きボールを k 個 ($k \le n$) の箱に入れる分割 (partition) の組合せの数を表す，いわゆる第 2 種の Stirling 数である．

上記の例に対応する二部グラフをシミュレーションで生成し，観測された次数分布と比較した結果を Fig. 5 に示す．ここでは，$M = 10^4$, $N = 10 - 5$, $\mu = 1.5$, $\nu = 15$ に対して，グラフの realization 一つだけに対する結果を比較してみた．(cf. [3, Fig. 7])．グラフのサイズが小さくても，かなりの程度よい結果が得られる．

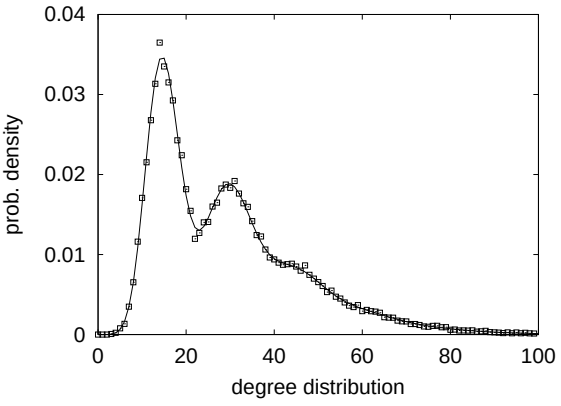

Figure 5: Poisson 分布を双方向の次数分布とするランダムな二部グラフを縮約し，次数分布を求めたもの．(白丸: a realization, 実線: 解析解)．

実際に観測されるデータでは，二部グラフ双方の次数分布 p_j, q_k は経験分布として与えられる．その

[6]縮約の過程で当然情報は落ちる．すなわち同じ一部グラフを与える異なる二部グラフが一般には存在する．

場合に，例 (4) に述べた母関数の多項式表式を用いて，帰無仮説の下での縮約グラフの次数分布を計算することができ，それを実データと比較できる．

5　コミュニティ解析への応用

社会ネットワークの研究で，ネットワークの中心性 (centrality) という指標がしばしば用いられる [5]．エージェント (行為者) が，どの程度に中心的なのか末端的なのかという，ネットワーク内での位置の重要性を測る指標である．何をもって重要とよぶかにより，いくつかの基準が考えられる．よく使用される 3 つの中心性指標をあげると

1. ノードのもつ紐帯 (tie) の数
2. ノード間の距離
3. ノードのもつ媒介性

がある．

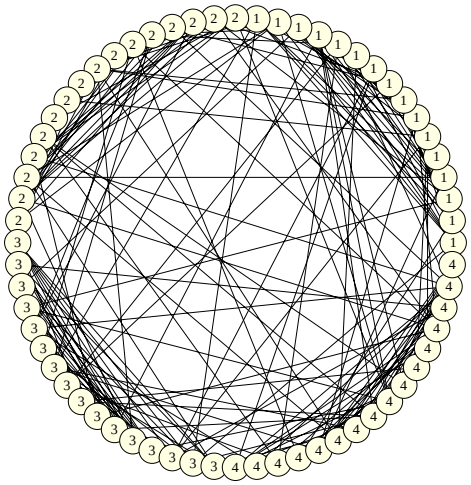

Figure 6: 4 つのコミュニティ構造をもつネットワーク

1. は次数が大きなノードほど重要であると考える指標である．2. は特定のノードから，ネットワーク内のすべての他のノードへの最短経路の長さを測り，その合計から，そのノードからどれくらいの距離で他のノードにつながっているかを指標とする．この場合，ネットワーク内のどの人にもできるだけ短い経路で情報を伝達できるノードほど中心的なノードであると解釈していることになる．

3. は，特定のノードが，他のノード間の関係性をどのように媒介するかその媒介性 (between とよばれる; [6]) により，中心性を定義するものである．1. がノードについて 1 次構造，2. がノードのペアにもとづく 2 次構造であるのに比べ，3. は 3 次構造をみていることになる．この場合，ネットワーク内で，このノードがなければ情報が伝わらない，もしくは伝わりにくいというノードほど中心的なノードであると解釈していることになる．

情報媒介の担い手という意味では，ノードよりもむしろエッジに中心性があるとみなす方が自然かもしれない．情報は相手との関係性がなければ媒介のしようがなく，ノードのペアに紐帯が存在することの方が中心性にとって重要と考えられるからである．したがって，エッジの betweenness ということが考えられる [7, 8].

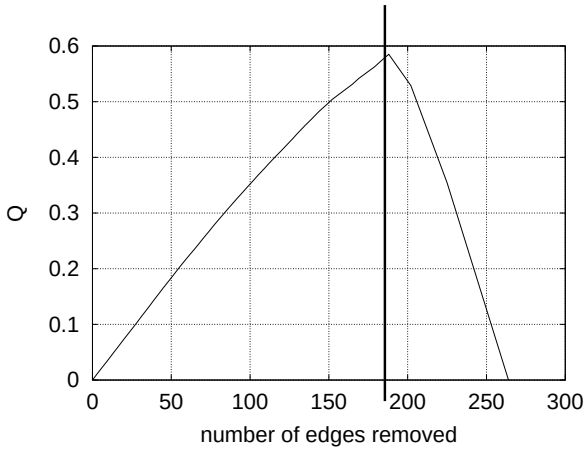

Figure 7: 縦軸: Q, 横軸: 除去して残ったエッジの数なので，アルゴリズムの進行に伴い，右から左へ Q が推移する．たての実線はそこで最大値を取ることを意味する．

さて，普通の社会的な組織には，部署に代表されるクラスタ (あるいはコミュニティ) 構造が存在している．その場合，同じコミュニティ内部に存在するエッジと，異なるコミュニティを結ぶエッジは，上で述べた中心性に差が生じていることが多い．そこでエッジの betweenness を用いれば，情報媒介という観点からコミュニティ構造がどのようなものなのかを解析することができる．

グラフからエッジを除去していく，いわゆる divisive な方法で，コミュニティを同定することができる．このアルゴリズムは，最初に全エッジを除去してから「仲間」として近いノードをつないでいく，通常のクラスタ解析のような aggolomerative なアルゴリズムと異なり，

1. もとのグラフからスタート
2. betweenness の最も大きなエッジを除去する
3. 停止条件が満たされるか，完全にノードがバラバラになるまで，残ったグラフを対象に，2 を走らせる

で与えられる．3 の停止条件をつけなければ，通常のクラスタ解析で得られるような樹形図 (dendrogram) が得られる．このアルゴリズムの途中で得られたクラスタ分けは，何をもってよいとすべきであろうか．

3 の停止条件として，[8] で提案されているのが，帰無仮説としてのランダムグラフからのずれである (もともとのアイデアは assortative mixing [9]).

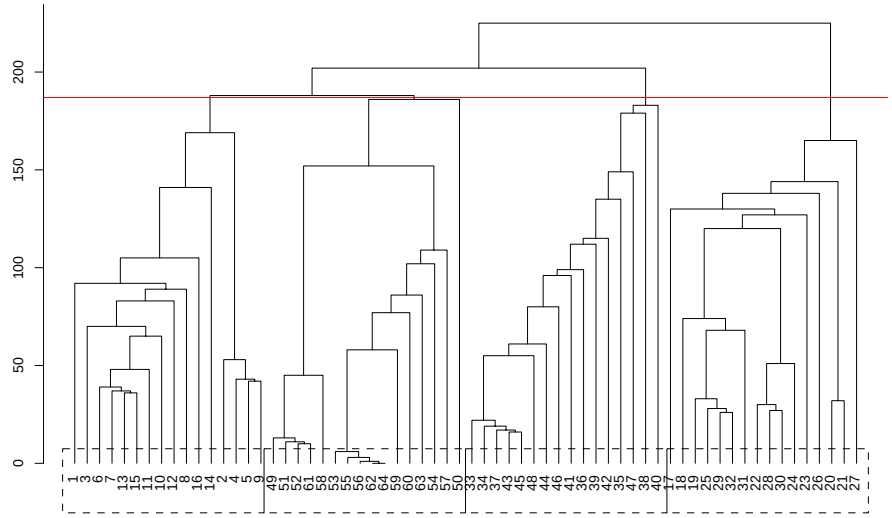

Figure 8: 樹形図 (dendrogram)．枝にある番号が 16 ごとに設定したクラスタになっている．よこの実線は Q の最大値に相当し，そこで 4 つのコミュニティに正しく分かれる．縦軸: 樹形図のそのレベルで，グラフに残っているエッジ数.

もとのグラフの全ノード数を N として，それぞれ $1, 2, \cdots, N$ という番号が振られているとする．いま，このアルゴリズムの途中で，エッジ除去の結果，もとのグラフが k 個の非連結成分に分かれたとする．ノード番号の集合 $[N] \equiv \{1, 2, \cdots, N\}$ は，空でない互いに素な k 個に集合に分割 (partition) される．分割におけるそれぞれの集合に番号 $i = 1, \cdots, k$ をつける．

このとき上記アルゴリズムのステップ 1 でスタートしたもとのグラフにおいて，クラスタ i とクラスタ j を結ぶエッジが全エッジ数のどれくらいの割合を占めているかを e_{ij} で表し，e_{ij} を $k \times k$ の対称正方行列の要素とする．この行列のトレース $\sum_i e_{ii}$ は，同じクラスタ内を結んでいるエッジが全エッジの内どれくらいあるかを表しているので，クラスタがうまく分けられれば大きな値を取る．しかし，全ノードを完全にバラバラにするような自明なクラスタ分けでは，$\sum_i e_{ii}$ は最大値 1 を取るので，これ自体はよい指標ではない．

いま，$a_i \equiv \sum_j e_{ij} = \sum_j e_{ji}$ が，クラスタ i に端点をもつようなすべてのエッジ (自分自身のクラスタにループするものも他のクラスタにリンクするものも含まれる) の割合であることに注意する．$a_1, a_2, \ldots, a_k$ が与えられたとき，それら系列は固定するという制限以外，まったくでたらめにコミュニティ内およびコミュニティ間にエッジをはるようなランダムグラフを仮説のモデルとして考えることができる．

そのようなランダムグラフでは，エッジの端点の一方がクラスタ i に属していることが，多端がどのクラスタに属しているかということと統計的に独立であるから，e_{ij}=$a_i\, a_j$ という帰無仮説がすべての i, j で成り立つことになる．そこで指標

$$Q = \sum_i (e_{ii} - {a_i}^2) \qquad (23)$$

をもって，作成されたクラスタ分けがよいかどうかの指標とすることができる [8]．$0 \le Q \le 1$ であり，Q が最大となるところが全体のコミュニティ構造が (もしあれば) 存在しているところとなっている可能性が最も高い．この指標は [8] では modularity とよばれている．

簡単な例として，Fig. 6 にあるような 4 つのコミュニティ構造が平坦に存在して，互いにゆるく結びついているようなネットワークでアルゴリズムを走らせてみる．$N = 64$ を等しく 4 つの「コミュニティ」に分け，任意のノードが，コミュニティ内の他のノードと平均 6 の次数で，コミュニティ外の他のノードと平均 2 の次数で，リンクされている．

Fig. 7 に modularity Q，Fig. 8 に樹形図を示す．ただしグラフに残ったエッジ数を，図の横軸と縦軸にそれぞれ取った (cf. [8, Fig.6])．

6 応用

これらの手法を応用したオリジナルな研究として，人的ネットワークと社会経済的な関係性ネットワークについて解析を行った [10]．紙面の制限上，これらの実データを用いた解析結果は研究会講演で発表する．

References

[1] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).

[2] M. Molloy, B. Reed, Combinatorics, Probability and Computing **7** (1998) 295–306.

[3] M. E. J. Newman, S. H. Strogatz, D. J. Watts, Phy. Rev. E **64** (2001) 026118.

[4] M. E. J. Newman, Social Networks **25** (2003) 83–95.

[5] J. Scott, *Social Network Analysis: A Handbook*, (Sage Publications, London, 2000, 2nd ed).

[6] L. Freeman, Sociometry **40** (1977) 35–41.

[7] M. Girvan, M. E. J. Newman, Proc. Natl. Acad. Sci. **99** (2002) 7821–7826.

[8] M. E. J. Newman, M. Girvan, Phy. Rev. E **69** (2004) 026113.

[9] M. E. J. Newman, Phy. Rev. E **67** (2003) 026126.

[10] Y. Fujiwara, K. Yuta, K. Shimohara, in preparation.