

講義ノート ネットワーク理論入門

増山 幸一
明治学院大学経済学部

2015年3月

目次

1	序	1
2	ネットワークの基礎概念と社会ネットワークの特徴	3
2.1	ネットワークの例	3
2.2	ネットワークの定義	6
2.3	中心化と群集化の諸概念	10
2.4	現実社会でのネットワークの特徴	13
3	ネットワークの基本モデル	17
3.1	Erdős and Rényi 型ランダムネットワーク	17
3.2	配列モデル	23
4	スモールワールド・モデルとスケールフリー・モデル	25
4.1	スモールワールド・モデル	25
4.2	スケールフリー・モデル	29
5	結び	38

1 序

会社や組織の管理・運営、新商品の研究開発、大学のクラスやサークル、あるいは、趣味やスポーツの同好会など、様々な社会的経済的な活動を結節点として、現実社会での人間関係の人的ネットワークは形成されていく。私たちは、日常的に、人と人の間に形成された社会的なネットワークの中で日々の生活を送っている。以下の図は、ビクトル・ユーゴの小説『Les Misérables』における登場人物をノードとし、彼らの間の人間関係をリンクで結んで描いたものである。人間関係の大きさに応じてリンク数が多くなる。当然のことながら、主人公のジャン・バルジャンとファンティエヌとその娘コゼットに接続するリンク数が最も多い。ノード番号

11 が Valjean、ノード 2 3 が Fantine、ノード 2 6 が Cosette に対応する*1。

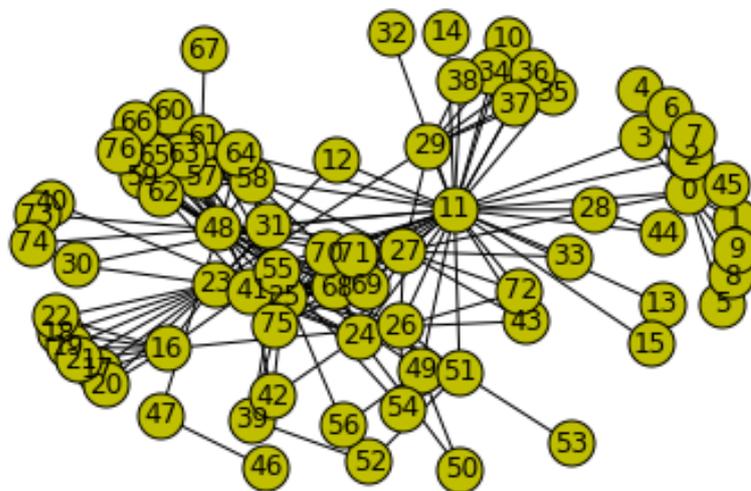


Fig.1.1 『Les Misérables』における人間関係

周知の通り、ネットワークの研究は、インターネット上での web のリンク接続、情報の交換と流通、イノベーションの伝搬、貿易取引と物流、金融資産の取引関係、友人関係、病気の伝染、生体内神経回路など、多種多領域にまたがる領域において急速に発展してきた。ネットワークは非常に多種多様な領域に広がっているが、大きく分けて4種類のネットワークに分類することができる*2。

一つは、人々の生活に見られる社会的経済的なネットワーク (social and economic networks) である。会社関係、あるいは取引関係、同級生同士の間関係、サークルを介した友人関係などから形成される人々のネットワークがこれに対応する。企業と企業間の取引関係のような、組織と組織間のネットワークもこれに入る。最近では、Facebook、LINE などの SNS を介した新しい人的ネットワークが形成されている。

次は、インターネット社会の基盤を形成する情報ネットワーク (information networks) である。Eメールの連鎖のネットワーク、学术论文の引用ネットワークやインターネット上において各 web ページのドキュメントがハイパーリンクされているなど、情報を対象物として連結されたネットワークである。

三つ目に、経済的活動のインフラを提供するネットワークとして、財やサービスの流通を行うために計画された技術的ネットワーク (technological networks) に分類されるものがある。代表的なものとして、各ドメイン間をつなぐインターネット回線、電力配電網、運輸交通網のようなインフラ的なネットワークである。

最後に、生物学的ネットワーク (biological networks) をあげることができる。生体内におけるニューロン神経のネットワーク、生態系における植物連鎖のネットワーク、病原菌の拡散過程を研究するときに採用される生体間の接触ネットワークなどが、これに当たる。

*1 このネットワークを描くための GML ファイルは、私のホームページ <http://www.geocities/mashyk3309/> の NetworkX のページにあります。

*2 これら4種類のネットワークの具体的な説明については、Newman(2010)を参照のこと。

近年における情報通信技術の驚異的な進歩は社会的ネットワークの性質に大きな影響を与えている。20世紀初頭以来、急速に、世界は小さくなっている。インターネットの進歩によりさらに世界は小さくなり続けている。多数の人々は新しい SNS を介して比較的少人数グループとの間で情報通信を行っている。政治的なブログに見られるように、新しい SNS の発達は各個人がより多様な意見を持つようになることを必ずしも保証しない。多分、SNS の進展は、他者の行動や意見を過剰に模倣する烏合の衆、群衆化を急速に作り出す可能性を持つ。Facebook、mixi、google+ に代表される新しい SNS は人々が入手する情報の性質を変化させたり、情報の処理方法を変化させるでしょう。

社会経済ネットワークはそれを構成するノード（結節点）が人間あるいは人的組織であり、人間あるいは人的組織の意思決定がネットワーク上で相互関係している点に特徴がある。だから、社会経済ネットワークの重要な特徴は、ネットワーク内のリンク接続の構造に依存するだけでなく、ネットワーク構造上で起きる行為の相互干渉の性格にも大きく依存する。ネットワーク内の各主体あるいは組織の意思決定はネットワーク上で相互依存するので、社会経済ネットワークの分析ではゲーム理論の枠組みが必要となってくる。ネットワーク内の各主体あるいは組織の意思決定が他の主体に影響を持たず現象を情報カスケード現象やネットワーク効果と呼んでいる。このような情報カスケード、群集化、ネットワーク効果がどのようなメカニズムで働き、どのような社会的現象を引き起こすのかを理解する必要がある。

この講義ノートでは、上記の課題に答えるために必要なネットワーク理論の基礎を解説する*3。とりわけ、経済分析において必須となるネットワークの考え方に基礎を提供することを目的にする。ネットワーク分析の数学的手法はグラフ理論と呼ばれる離散数学の1分野である。グラフ理論は、Euler(1735)が Königsberg 橋問題に対する数学的解答を準備する研究に出発点を持っている。

ネットワーク構造の概念を研究手法に導入した初期の研究は社会学においてである。Milgram(1967)によってスモール・ワールド (small worlds) という言葉が提起され、理論上の問題を抱えつつも、その後のネットワーク研究で重視されてきた。Watts(1999)は、WWW においてもスモール・ワールド性が成立することを指摘した。ネットワークの直径距離が短いこととクラスタリング係数値が相対的に大きいという2条件を満たすとき、スモール・ワールド性が成立するという。

次節では、社会経済ネットワークの分析で必要となるネットワーク理論の基礎的な概念と分析ツールについて説明し、現実社会で観察されるネットワークの典型的な事例を取り上げて、それらに共通する特徴について説明する。第3節では、基本的ランダムネットワーク・モデルの基本的考え方を説明する。第4節で、スモールワールド・ネットワーク、成長するネットワーク (スケールフリー・ネットワークのモデル) を導入し、ランダムグラフと対比しながら、それらの特徴を分析し、モデルの現実有効性について説明する。

2 ネットワークの基礎概念と社会ネットワークの特徴

2.1 ネットワークの例

最も基本的な意味で、ネットワークとは、ある対象物と他の対象物の組がリンクによって連結されているような対象物とリンクの組合せの集合である。ソーシャル・ネットワークの例として必ず登場するのが、Zachary による大学空手倶楽部を通じた学生間の友人関係のネットワークである。以下のグラフがそれを示し

*3 ネットワーク的思考を解説した入門書の代表格は、アルバート＝ラズロ・バラバシ著『新ネットワーク思考』NHK 出版 (2002 年)、ダンカン・ワッツ著『スモールワールド・ネットワーク』(2004 年) である。全く数式らしきものは出てきませんが、内容は相当に抽象的な所まで及んでいます。

ている。

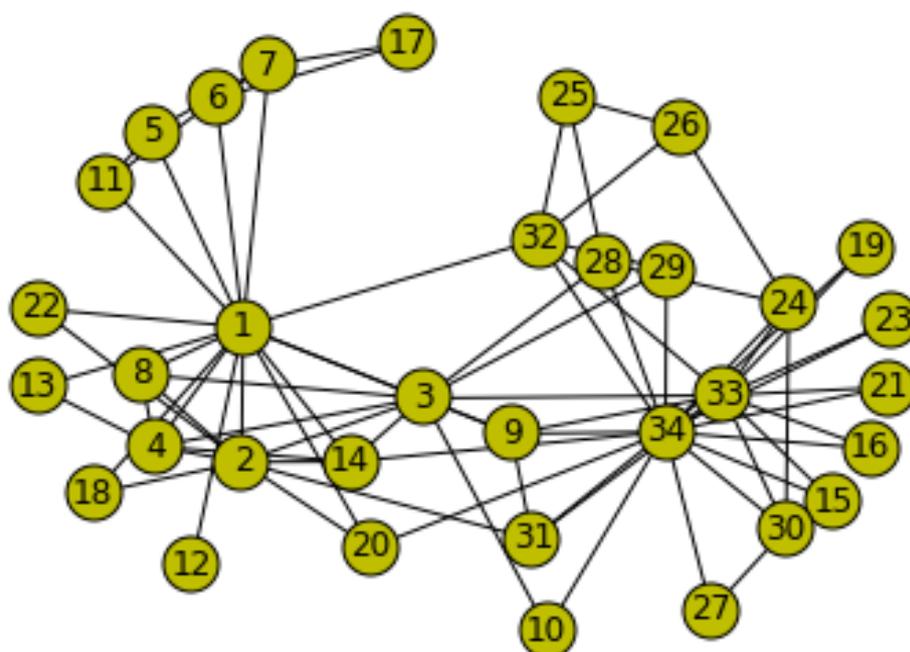


Fig.2.1 空手クラブを介した友人関係

34名からなる友人ネットワークの中で、ノード番号1と34が最も多いリンク（友人）数を持つ。

以下の図は、1970年12月時点でのARPANET(Advanced Research Projects Agency Network)と呼ばれるコンピュータ・インターネットのネットワーク構造を描いたものである。ARPANETは1969年に米国国防総省の国防高等研究計画局のもとに構築された研究・調査のための大型汎用コンピュータのネットワークである。最初はUCLAとスタンフォード大学との間で試験的に回線が引かれ、後に、UCLA、UCSB、SRI、スタンフォード大学、ユタ大学を結ぶ専用回線が敷設された。

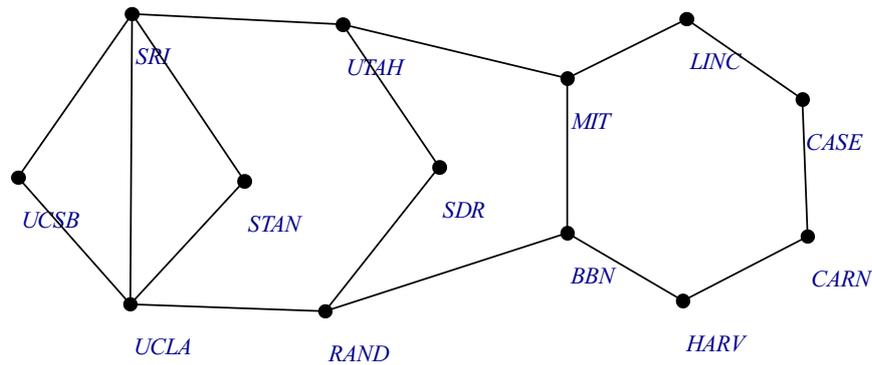


Fig.2.2 1970年12月でのARPANETネットワーク

各ノードはホスト・コンピュータが置かれている大学等を表現しており、リンクは各ホスト・コンピュータが専用回線で直接に接続していることを示す。ホスト・コンピュータを置く機関は、地理的には東海岸のMITやハーバード大学から西海岸のスタンフォード大学やUCLAなどの13大学等で、ネットワーク上では、リンクされた各ノード間の地理的距離はあまり関係がない。ちなみに、RAND研究所はカリフォルニア州サンタモニカに位置するが、BBN社はマサチューセッツ州ケンブリッジに位置する。このネットワークは情報通信ネットワークの典型例である。

学校生活、趣味やボランティア活動などの社会的な活動、あるいは、会社や組織を管理・運営する、商品販売の営業をする、共同研究開発に従事するなどという経済的な活動を結節点として、通常の社会的ネットワークは形成されていく。参加している社会的団体や経済的組織はネットワーク形成におけるフォーカル・ポイント（中心点）と呼ばれている。団体・組織の集合をフォーカル・ポイントとして形成されるネットワークは2部グラフ (bipartite graph) として表現できる。2部グラフでは、各ノードは、個人からなる集合と、団体・組織からなる集合のどちらかに2分類され、個人集合に属するノードと団体・組織に属するノードの間にリンクが形成される。下の図は米国有名企業の取締役役間の2項ネットワークの例である。

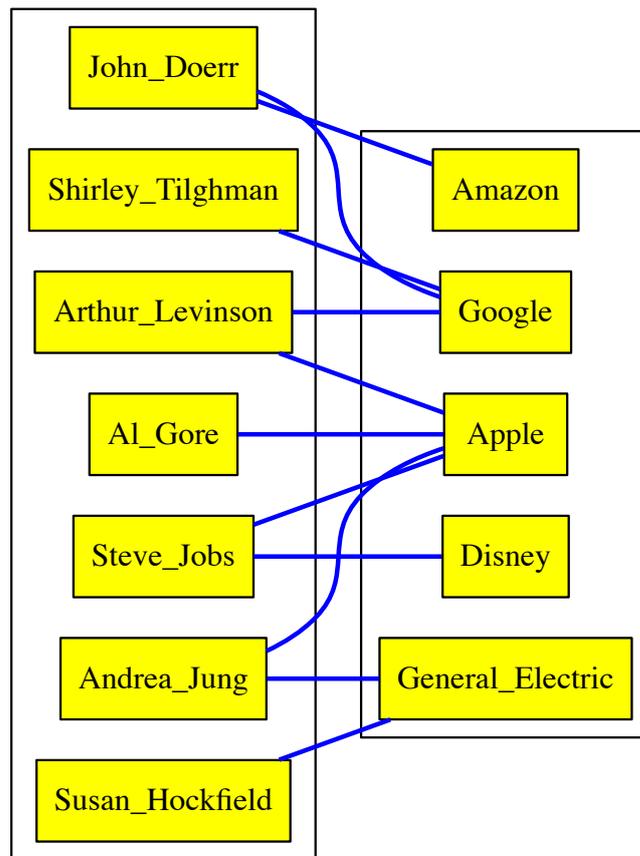


Fig.2.3 2部ネットワークの例 (2009年米国IT系会社の取締役会のネットワーク)

2.2 ネットワークの定義

ネットワークは、抽象的には、ノード (nodes, vertices) と各ノードを接続するリンク (edges, links) の集合からなる。各ノードに番号 i をつける。 n 個のノードからなるネットワークを考えると、ノードの集合を $N = \{1, 2, 3, \dots, n\}$ とする。ノード i がこのネットワークに属していれば、 $i \in N$ である。ノード i とノード j との間のリンク接続を変数 g_{ij} で表現する。 $g_{ij} = g_{ji} = 0$ であるとき、ノード i とノード j の間には何の関係もない。ノード i の行動とノード j の行動の間に関係性があるときは、 $g_{ij} \neq 0$ または $g_{ji} \neq 0$ である。ノード間の関係に方向性がないとき、つまり、リンク接続が両方向であるとき、 $g_{ij} = g_{ji} = 1$ となる。無方向グラフ (undirected graph) と言う。依存関係に方向性があるとき、例えば、 i から j への依存性はないが、 j から i への依存関係があるとき、 $g_{ji} = 0, g_{ij} = 1$ などと表現する。このようなネットワークをリンクが方向づけられているので、有向グラフ (directed graph) という。以下での議論では、断りがない限り、無方向グラフを対象とする。

ノードの集合 N と各ノード間のリンク構造 G が与えられるとき、一つのグラフが記述できる。これを一般的にグラフ (ネットワーク) $g = (N, G)$ と表記する。 G は (i, j) 要素を g_{ij} とする $n \times n$ 行列である。行列

G は各ノード間の直接的な連結関係（隣人関係）を表現しているので、隣接行列 (adjacency matrix) と呼ばれる。

以下のネットワーク例を取上げます。 $N = \{A, B, C, D, E, F, G\}$ とする。各ノード間のリンク構造、隣接行列 G が以下の表で与えられている。

例 2.1

隣接行列の例

	A	B	C	D	E	F	G
A	0	1	1	1	1	0	0
B	1	0	0	0	0	1	1
C	1	0	0	1	0	0	0
D	1	0	1	0	0	0	0
E	1	0	0	0	0	1	0
F	0	1	0	0	1	0	0
G	0	1	0	0	0	0	0

このグラフを視覚化して描いてみると、Fig. 2.4 に描かれたネットワークとなる。このネットワークの特徴を見てみよう。ノード A はノード B、C、D、E とリンクで接続している。つまり、ノード B、C、D、E はノード A の隣人 (neighbors) である。隣人の数を次数 (degree) と言う。ノード A は 4 人の隣人、つまり、次数 4 を持つ。ノード B は 3 人の隣人を持つ。

an example of connected networks

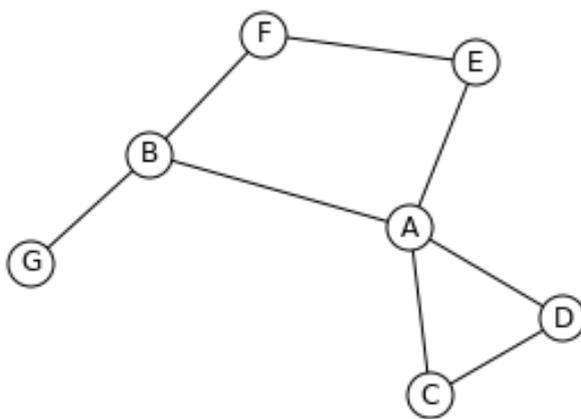


Fig.2.4 パスと連結されたネットワークの例

ネットワーク上で、リンクによって連結しているノードの点列をウォーク (walk) という。同一のノードが一つのウォークに複数回登場することもあり得る。ウォークの中のノードがすべて異なっているとき、パス (path) という。出発するノードと終端ノードが同じようになっているノードの点列をサイクル (cycle) という。

Fig2.4の例でいうと、ノード A からノード G に到達するパスは 2 個あり、A-E-F-B-G と A-B-G である。このうち、A-B-G の方が最短距離で到達するパスになっている。A-C-D-A はサイクルになっている経路である。A-E-F-B-A もサイクルである。

正確に定義する。ノード i に連結するリンクをもつすべてのノードの集りを $N_i(g) = \{j \in N | g_{ji} = 1\}$ で表記する。 N_i がノード i の隣人集合である。ノード i に直接的にリンクされた隣人の数 $\eta_i(g) = |N_i(g)|$ はノード i の次数である。言い換えると、 $\eta_i(g) = |N_i(g)|$ はネットワーク g に属するノード i の隣人の数を表す。ここで、 $|X|$ は集合 X の要素数を表現している。上の例では、 $N_A = \{B, C, D, E\}$, $\eta_A = 4$ となっている。同様に、 $N_B = \{A, F, G\}$, $\eta_B = 3$ である。

k 個の隣人を持つノードの集合を $N_k(g)$ とすると、グラフ内のすべてのノードは、互いに異なる $n - 1$ 個の集合

$$N_1, N_2, \dots, N_{n-1}$$

に分割できる。上の例では、

$$N_1 = \{G\}, N_2 = \{C, D, E, F\}, N_3 = \{B\}, N_4 = \{A\}$$

となっている。

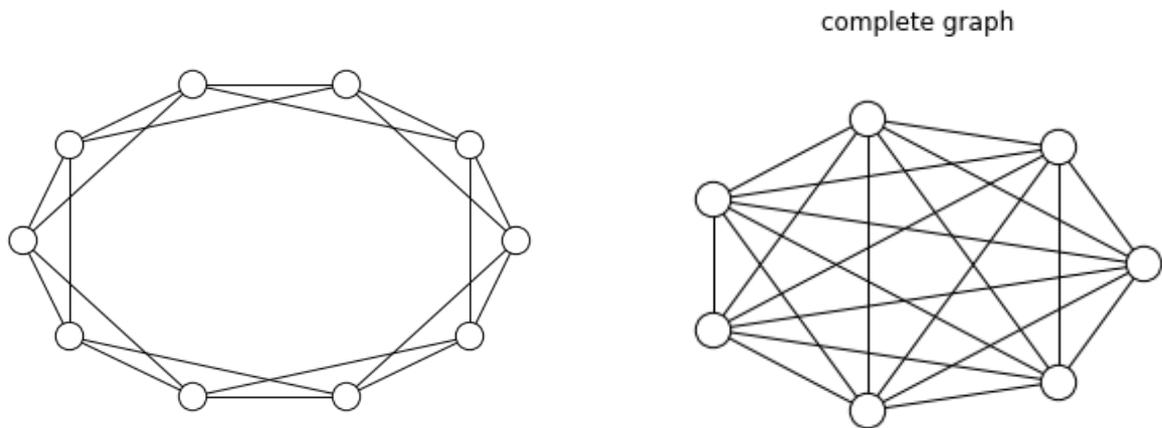


Fig.2.5 正則グラフと完備グラフ

すべてのノードが同一数の次数を持つようなネットワークを正則グラフ (regular graph) という。Fig. 2.5 に正則グラフの例が描かれている。この例では各ノードは同一の次数 4 を持っている。 $\eta_i = n - 1$ であるとき、つまり、各ノードがネットワーク内の他のすべてのノードと連結しているとき、完備ネットワーク (complete network) という。一つのノードを中心として、他のすべてのノードがこの中心ノードとだけ連結しているネットワークをスター型ネットワークと言う。

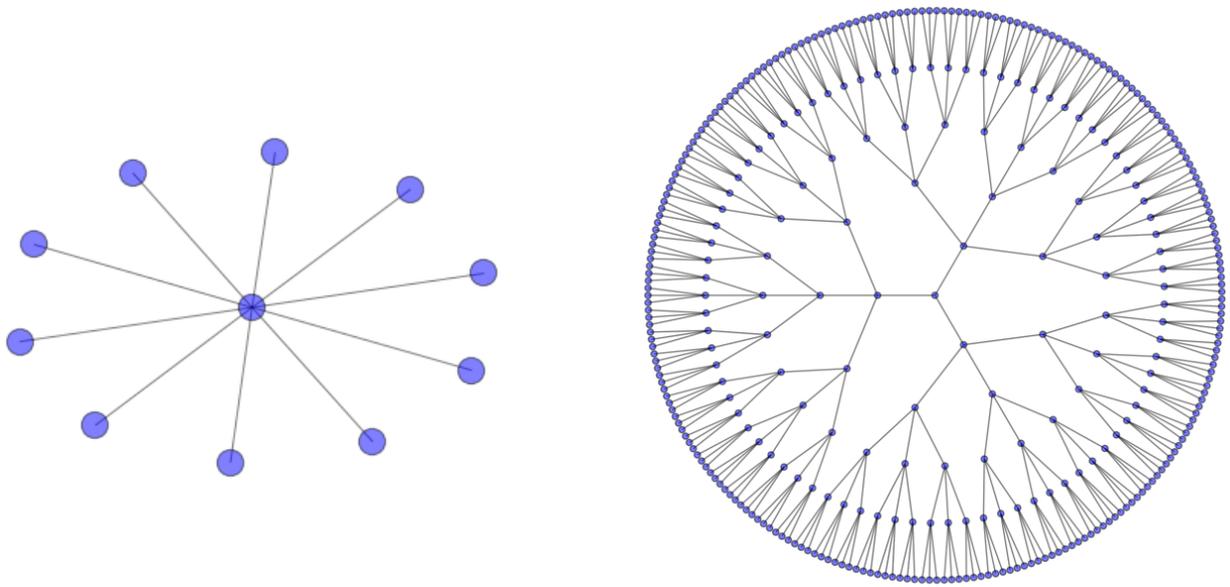


Fig.2.6 スター型グラフとツリー状になったスター型グラフ

次数 k を持つノードの割合 (次数分布) $\Pr(k)$ は

$$\Pr(k) = \frac{|\mathbb{N}_k(g)|}{n}$$

と定義される。明らかに、 $\sum_{k=0}^{n-1} \Pr(k) = 1$ である。再び上の例で計算すると、 $N = 7$ なので

$$\Pr(1) = \frac{1}{7}, \Pr(2) = \frac{4}{7}, \Pr(3) = \frac{1}{7}, \Pr(4) = \frac{1}{7}$$

となる。次数分布の分散は、スター型ネットワークでは、 n とともに増加し、正則ネットワークではゼロとなる。

ここで、ネットワークの特徴を表現する上で重要な言葉を定義する。ネットワーク内のある部分集合を取り出し、この部分集合に属する異なる二つのノードの間にパスが存在するとき、これらのノードは同一のコンポーネント (*component*) に属するという。コンポーネントはネットワークの部分ネットワークで、その中のすべてのノードが連結している最大の部分集合である。ネットワークは通常、複数のコンポーネントを含んでいる。ネットワーク上のすべてのノードの組がそれらを接続するパスを持っているとき、ネットワークは連結している (*connected*) という。連結したネットワークはただ一つのコンポーネントをもつ。例えば、Fig 2.4 のネットワークは連結しており、ただ一つのコンポーネントから形成されている。

以下の例では、コンポーネントが複数個存在するケースを描いている。ネットワーク全体のノード数は 15 である。ネットワーク全体としては連結していないが、各コンポーネント内の各ノードは連結している。

an example of components

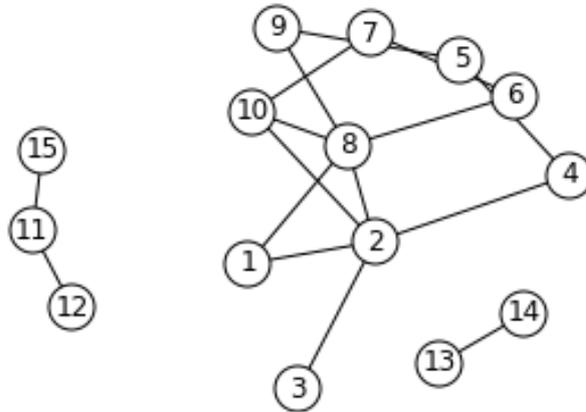


Fig.2.7 サブネットワークとコンポーネント

二つのノードを結ぶ最短パスの距離を測地距離（最短距離） (*geodesic distance*) という。ネットワーク上にある二つのノード間の距離は通常、この最短距離のことをさす。ネットワーク上の任意の二つのノード間の最短距離の最大値をネットワークの直径 (*diameter*) という。各ノード間の最短距離の平均を取った値をパスの平均距離という。

2.3 中心化と群集化の諸概念

社会・経済ネットワークの分析で最初に登場する代表的な例の一つは15世紀初期イタリア・フィレンツェにおける *Medici* 家の興隆を巡る家系間のネットワーク分析である。政略的結婚を通じた血縁ネットワークの拡大をテコに富と政治的権力を獲得していった過程を分析するために、以下で定義する中心化度 (*centrality*) の概念が活用された。

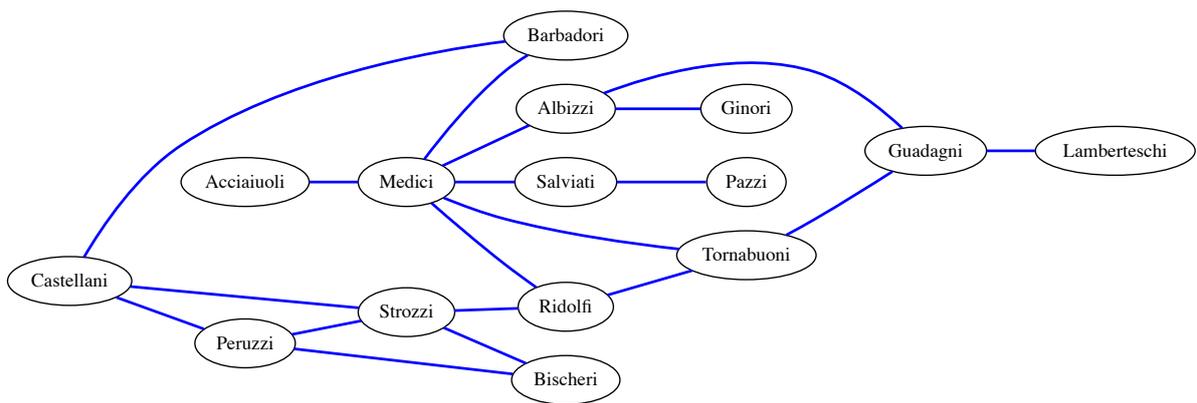


Fig.2.8 15世紀フィレンツェにおける *Medici* 家と主要家系のネットワーク

Padgett and Ansell(1993)の研究によれば、15世紀フィレンツェにおける主要な血縁門閥のネットワークは Fig.2.6 にある通りである。この例を通して、中心化度とりわけ中継中心化度の概念の重要性を理解する。

各ノードがネットワーク上でどれくらいの重要性を占めているかを測る計数として中心化度 (centrality) を定義する。中心化度は4種類の統計量に従って計測することが出来るので、異なる4種類の定義が存在する。

1. 次数 (degree) に基づく中心化度の計測：ノードがどのように接続されているかを計測
2. 近接性 (closeness) に基づく中心化度の計測：各ノードが互いにかに容易に到達し得るのかを計測
3. 中継性 (betweenness) に基づく計測：他のノードに到達する上でいかに重要な役割を果たすかを計測
4. 隣人 (neighbors) の特性に基づく計測：ノードの隣人がいかに重要なあるいはいかに強い影響力を持つかを計測

次数中心化度 (degree centrality) とは、ノード i の次数中心化度 $C_d(i; g)$ とは、このノードの次数を最大可能な次数で割った値をいう。

$$C_d(i; g) = \frac{\eta_i(g)}{n-1}$$

中継中心化度 (betweenness centrality) は以下のように定義される。ノード i とノード j を結びつける最短距離のパスの総数を $P(i, j)$ と表記し、これらのパス経路の中にノード k が含まれているパス経路数を $P_k(i, j)$ と表記する。ノード k の中継中心化度 (betweenness centrality) は、直感的は、

$$C_b(k; g) = \sum_{i \neq j, k \notin \{i, j\}} \frac{P_k(i, j)}{P(i, j)}$$

である。しかし、この定義だと、ノード数 n が大きくなるにつれて、この中心化度の値も大きくなる。値を正規化するためには、つまり、区間 $[0, 1]$ に収めるためには、

$$C_b(k; g) = \sum_{i \neq j, k \notin \{i, j\}} \frac{P_k(i, j)/P(i, j)}{(n-1)(n-2)/2}$$

と定義する方が便利である。ここで、分母の値 $(n-1)(n-2)/2$ はノード k 以外の $n-1$ 個のノードから二つのノードを取り出す組合せ (リンク) の総数である。

この定義に沿って、Fig 2.8 の例における家系 i と家系 j を結びつける最短距離の経路の総数を $P(i, j)$ と表記し、これらの経路の中に家系 k が含まれている経路数を $P_k(i, j)$ と表記する。Barbadori 家と Guadagni 家を結ぶ最短距離の経路は、3つのリンク (枝) を経由する Barbadori - Medici - Tornabuoni - Guadagni という経路と Barbadori - Medici - Albizzi - Guadagni の経路の2経路である。だから、 $i = \text{Barbadori}$ 、 $j = \text{Guadagni}$ としたとき、 $P(i, j) = 2$ である。 $k = \text{Medici}$ とすると、 $P_k(i, j) = 2$ である。 $k = \text{Strozzi}$ とおくと $P_k(i, j) = 0$ 、 $k = \text{Albizzi}$ とおくと $P_k(i, j) = 1$ となる。Medici 家は Barbadori 家と Guadagni 家を結びつける上で中心的役割を果たす家系となっている。ネットワークにおけるこうした中心的役割を数量的に表現するのが、中継中心化度である。この計量化を活用すると、Medici 家の係数は 0.5219 となる。つまり、Medici 家以外の2家系を結ぶ最短経路総数のうち 52% の経路は Medici 家を經由していることがわかる。Strozzi 家の中継集中度指数を計算すると、0.103 となる。Medici 家の次にこの値が大きい家系は Guadagni

家で、その値は 0.2545 である。このことから、Medici 家の政略結婚がネットワークの構築にいかに重要な役割をはたしたかを理解できる。

この例に対して Networkx を用いて計算した詳細な結果を以下の表に示す。

例 2.2

フィレンツェにおける家系ネットワークの例

	clustering	degree sequence	degree centrality	betweenness centrality
'Castellani'	0.6666	3	0.2142	0.0549
'Peruzzi'	0.6666	3	0.2142	0.0219
'Strozzi'	0.3333	3	0.2142	0.1025
'Ridolfi'	0.3333	3	0.2142	0.1135
'Barbadori'	0.3333	3	0.2142	0.0934
'Medici'	0.0666	6	0.4285	0.5219
'Tornabuoni'	0.3333	3	0.2142	0.0915
'Bischeri'	0.0	1	0.0714	0.1043
'Guadagni'	0.0	4	0.2857	0.2545
'Lamberteschi'	0.0	1	0.0714	0.0
'Acciaiuoli'	0.0	1	0.0714	0.0
'Salviati'	0.0	2	0.1428	0.1428
'Pazzi'	0.0	1	0.0714	0.0
'Albizzi'	0.0	3	0.2142	0.2124
'Ginori'	0.0	1	0.0714	0.0

群集化 (clustering) の概念を理解するために、再び Fig.2.4 の例に戻ろう。ノード A の隣人は 4 人 (C,D,E,B) いる。この 4 人の中で知人同士になっている組は C-D の一組だけである。4 人の中から二人を取り出す組合せの数は $4!/2!2! = 6$ である。従って、隣人同士の組合せ (6 組) の中で知人同士である割合は $1/6$ となっている。これをクラスタリング係数という。つまり、自分の隣人が互いに知合いになっている確率である。

Fig.2.8 で、Medici 家にリンクを持つ家系ノードは 6 であり、そのうち一組だけ、Tornabuoni 家と Ridolfi 家のみが直接的にリンク接続している。なので、Medici 家のクラスタリング係数は

$$1/\frac{6!}{2!4!} = 1/15 = 0.0666$$

と計算できる。上記の結果とも符合する。

ノード i のクラスタリング係数は

$$Cl_i = \frac{\text{ノード } i \text{ の二人の隣人が直接的にリンクされている組合せの数}}{\text{ノード } i \text{ の隣人から二人を取り出す組合せの数}}$$

となる。より、正確に言うと、ノード i の群集化係数 (clustering coefficient)、 $Cl_i(g)$ 、は

$$Cl_i(g) = \frac{\sum_{j \in N_i(g)} \sum_{k \in N_i(g)} g_{jk}}{\eta_i(\eta_i - 1)} = \frac{\sum_{j < k, j, k \in N_i(g)} g_{jk}}{\eta_i(\eta_i - 1)/2}$$

と定義される。

ノード i がノード j と知合いで、ノード j がノード k と知合いであるとき、2本のリンク ij と jk からなるパス (path) $i-j-k$ が存在する。 i が k と知合いであれば、このパスは閉じていると言う。つまり、3辺 ij 、 jk 、 ki からなる3角形の閉じたパスができる。クラスタリング係数は、長さが2のパスの中で、3角形の閉じたパスになっているパスの割合を計測している。

$$\text{クラスタリング係数} = \frac{\text{長さが2の閉じているパスの数}}{\text{長さが2のパスの数}}$$

ネットワーク全体での平均的なクラスタリング係数 Cl を求める方法は

$$Cl = \frac{\text{閉じた3角形になっているパスの数} \times 3}{\text{長さ2のパスで連結された3ノードの組合せの数}}$$

で計算することになる。分子に $\times 3$ が登場するのは、3角形の頂点が3個あり、パスの出発するノードとしては3種類あり得るので。

Fig 2.6 から理解できる通り、スター型ネットワークでは、中心ノード以外のノード間にはリンクがないので、クラスタリング係数の値は0となる。完備ネットワークではすべてのノードが互いに接続しているのでクラスタリング係数は1となる。クラスタリング係数のもう一つの定義は各ノードのクラスタリング係数の加重平均を採用することである。

互いにリンクしているノード同士から構成される集合の最大の部分集合を派閥 (clique) という。コンポーネントとの相違に注意して下さい。 $k \geq 3$ として、 $I = \{i_1, i_2, \dots, i_k\} \subset N$ が以下の条件を満たすとき派閥という: I のすべての組合せ、 $i_x, i_y \in I$ に対して $g_{i_x i_y} = 1$ が成立し、この性質を持ち、かつ $I \subset I'$ となる集合 I' が存在しない。完備ネットワークでは、ネットワーク全体が派閥になる。

2.4 現実社会でのネットワークの特徴

現実社会でのネットワークの特徴と理論モデルの特徴とを対比するために、最もシンプルで典型的なランダム・ネットワークの理論モデルである Erdős and Rényi 型ネットワークの特徴を明らかにしておこう。基本的な考え方は、以下の通りである。 n 個のノードから任意に2つのノードを選び、例えば、サイコロを振って6の目が出たら、このノード間にリンクを張る。それ以外の目が出たとき何もしない。再び、任意に2つのノードを選び、サイコロを振って6の目が出たら、このノード間にリンクを張る; それ以外の目が出たとき何もしない。これを何回も続ける。この結果、ある種のネットワークが形成される。この手続きに従って形成されるネットワークをランダムネットワークという。これを数学的に定式化する。

ノードの集合 $N = \{1, 2, \dots, n\}$ の中から、任意に選び出した二つのノード間にリンクが張られる確率 (リンク形成確率) を、各ノードの位置に依存せず、 p とする。ただし、 $0 < p < 1$ である。各ノードはネットワーク上の $n-1$ 個のノードとリンクされる可能性があるが、 $n-1$ 個のノードのうち d 個のノードとリンク接続している確率は

$$p^d (1-p)^{n-1-d}$$

である。 $n - 1$ 個の中から d 個のノードを取り出す組合せの数は

$$\frac{(n-1)!}{d!(n-d-1)!} = \frac{n \dots (n-d-2)}{d!}$$

である。従って、任意に選ばれたノードの次数が d となる確率は

$$\Pr(d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}$$

である。平均次数は

$$\sum_{d=0}^{n-1} \Pr(d) d = (n-1)p$$

である。形成されるネットワーク上のリンク総数は確率変数であるが、大数の法則により、 n が非常に大きくなると、リンク総数平均値は

$$\frac{n(n-1)}{2} p$$

となることが知られている。各ノードの集中化（クラスタリング）係数が p になることは自明である。さらに、 n が非常に大きくなるときには、パラメター $\mu = (n-1)p$ を平均値としてもつポアソン分布になる。

$$\Pr(d) = \frac{e^{-\mu} \mu^d}{d!}, \quad d = 0, 1, 2, \dots, n-1.$$

この次数分布で表現されるネットワークをポアソン型ランダム・ネットワーク (Poisson random graph) ともいう。ポアソン分布をしたランダム・ネットワークの例を以下の図に示す。この数値例は、 $n = 1000$, $p = 0.01$ とおいて、描いている。

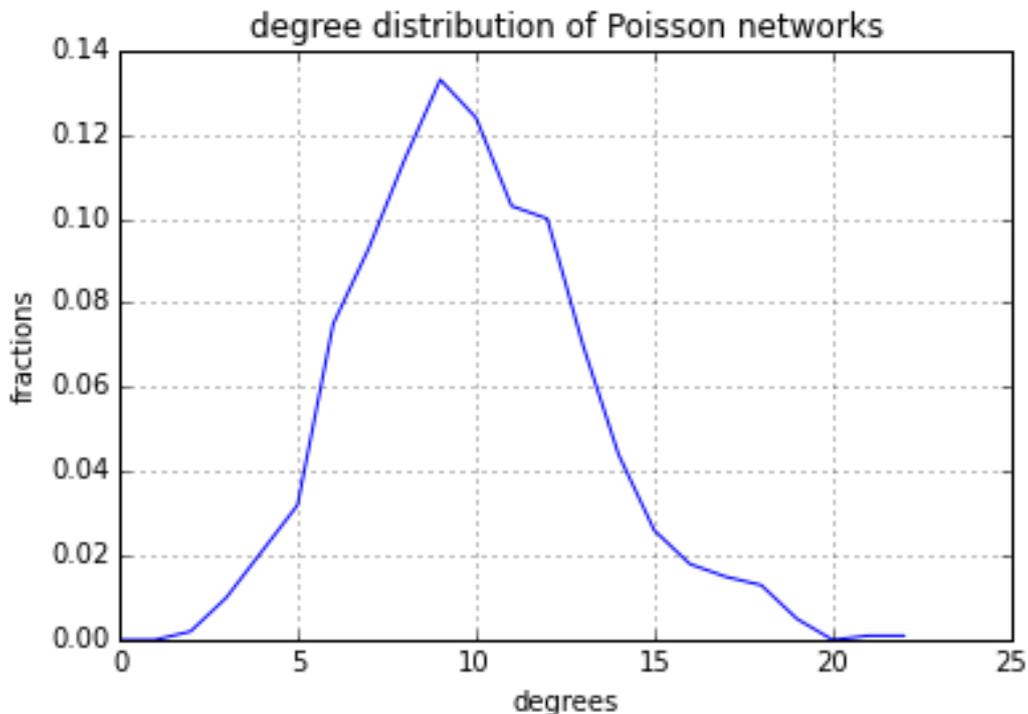


Fig. 2.9 ポアソン型ランダム・ネットワークの次数分布： $n = 1000$, $p = 0.01$

総ノード数は 1000 なので、平均次数は $np = 10.0$ である。次数分布は釣り鐘形状をしている。大多数のノードの次数（隣人数）は次数分布 5 から 15 の範囲に収まっている。この範囲以外の次数を持つノード数は全体の 5% 以下となっている。クラスタリング係数は理論とおり、 $p = 0.01$ である。パスの平均距離は 3.2 である。

社会ネットワークの研究で最もよく知られている例はスモール・ワールド (small worlds) という言葉で表現されているモデルであろう。スモール・ワールドという言葉が意味することは、現実の社会ネットワークではネットワークの直径が想像されるよりもはるかに小さく、パスの平均距離が短いという事実である。Milgram(1967) は、現実社会におけるネットワークのパス距離がいかほどかを調べるために実験を行った。ネブラスカに住む住人 96 人を電話帳から任意に選び、彼らに封書を送った。封書の中には、ハーバード大学名入りの”パスポート”が封入されていて、これを Milgram の (ボストンに住む) 友人 (A 氏とする) に送り返すような願い書が入っていた。この友人 A 氏に関する情報は住所と職業だけが開示されていた。封書を受け取った人々は、A 氏を直接・間接に知っているであろうと予想される友人に、あるいは同じ職業に就いている友達に、”パスポート”を送ることを要請された。これを受け取った人々も同じように行動することが要請されていた。この結果、18 個の”パスポート”が A 氏のもとに届いた。平均的に 5.9 回の転送を経由して届いたことが判明した。人々は世界中の誰とでも 6 次の隔たり (six degrees) でつながっている。世界は見たよりもはるかに小さい。

一般的に、人々が平均的に 100 人の友人・知人を持っているとすると、友人の友人数は 10000 名となる。従って、友人の友人の友人の数は 100 万人となる。人々が平均的に d 個のリンク (友人) を持っている、距離 6 のパスによって到達可能なノード数は $x = d^6$ である。両辺の対数を取ると、 $\ln x = 6 \times \ln d$ から $x = 200,000,000$ (人口が 2 億人) とおくと、米国の人々の平均的友人数 d は約 24.1 人となる。

その後、スモール・ワールド・モデルは Watts and Strogatz(1998) によってネットワークモデルとして定式化され、スモール・ワールド性が WWW ネットワークにおいても観測できることが Watts(1999) によって確認されている。

以下の表は生物学、経済学、数学、物理学における論文の共著者のネットワークを調べた結果である*4。

項目	生物学	経済学	数学	物理学
ノード総数	1,520,521	81,217	253,339	52,909
平均次数	15.5	1.7	3.9	9.3
パスの平均距離	4.9	9.5	7.6	6.2
最大コンポーネントの直径	24	29	27	20
平均クラスタリング係数	0.09	0.16	0.15	0.45
最大コンポーネントのシェア	0.92	0.41	0.82	0.85

これら 4 領域の中で、経済学での論文著者のネットワークは、平均次数が最も小さく、パスの平均距離が最も長く、最大コンポーネントの占めるシェアが最低になっている。生物学のネットワークは、平均次数が最大であり、平均クラスタリング係数が最も低く、最大コンポーネントのシェアが最大となっている。生物学の研究は多数の研究者による共同研究が非常に多いといえる。それに比較して、経済学の研究は一人あるいは少数の

*4 表の値は Jackson(2008) の 58 ページから引用。

研究者による研究が多い。平均クラスタリング係数が最も大きいのは物理学であり、その値は 0.45 である。この値は、ランダムネットワークが生み出すクラスタリング係数の大きさに比較すると非常に大きな値だと言える。物理学のネットワークでは、ノード数が 52909、平均次数が 9.3 だから、ランダム・ネットワークでこの平均次数を生み出すためには、各リンクが生成される均一確率は $9.3/52908 = 0.00018$ でなければならない。この値は 0.45 と比べると桁違いに小さい。

以下の表は、1970年代から1990年代にかけて *EconLit* に収録された経済学論文の著者のネットワークの特徴を分析した *Goyal, Leij and Moraga-González(2006)* の研究結果を示したものである。

項目	1970s	1980s	1990s
著者総数	33,770	48,608	81,217
平均次数	0.894	1.244	1.672
(標準偏差)	1.358	1.765	2.303
最大コンポーネントのサイズ	5253	13808	33027
最大コンポーネントのシェア	15.6%	28.4%	40.7%
準最大コンポーネントのサイズ	122	30	30
単独著者の総数	16,735	19,315	24,578
単独著者のシェア	49.6%	39.7%	30.3%
クラスタリング係数	0.193	0.182	0.157
最大コンポーネントにおける平均距離	12.86	11.07	9.47
(標準偏差)	4.03	3.03	2.23

1970年代に発表された論文の著者（ノード）数は 33,770 で、1990年代の著者数は 81,217 となり、2倍以上に拡大している。平均次数は 0.89, 1.244, 1.672 というように非常に小さい。巨大コンポーネントの占めるシェアが 15.6% から 40.7% へと拡大し、その平均距離は縮小傾向にある。つまり、スモール・ワールド化している。クラスタリング係数は一貫して低い値をとっている。1970年代では、平均次数が 0.894、クラスタリング係数は 0.193 であった。同じ平均次数を持つランダム・ネットワークでクラスタリング係数を計算すれば、 $0.84/33769=0.000026$ となる。現実に観察されたクラスタリング係数は 0.193 であるから、ランダム・ネットワークでの計算値に比べて 1 万倍となっている。

以下の図は Newman が使用したネットワーク研究者間の友人関係ネットワークの次数分布を示している。

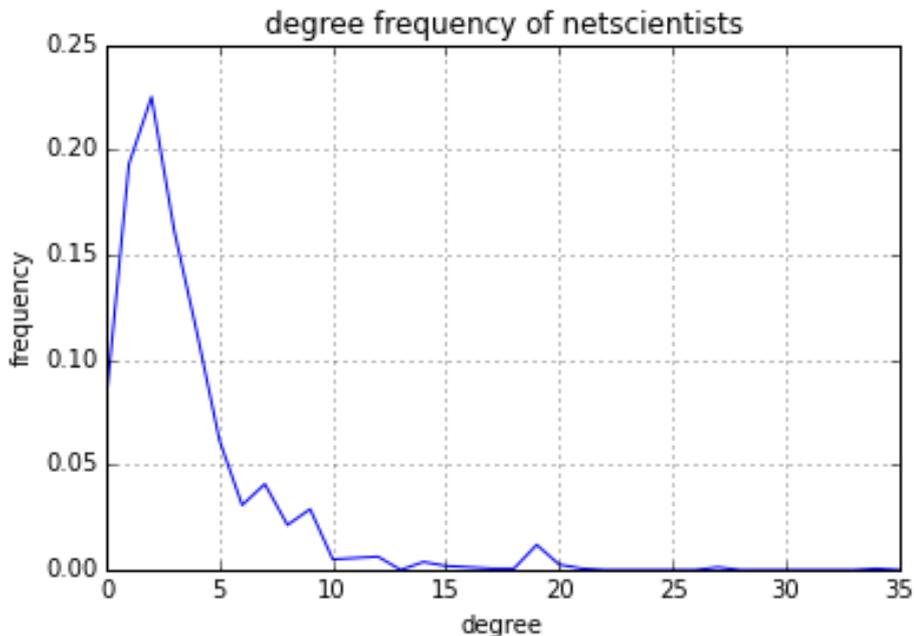


Fig2.10 ネットワーク研究者の友人間ネットワークの次数分布

この次数分布の図と Fig. 2.9 のグラフとを比較して明確になることは、人的なネットワークでは次数分布はポアソン分布とは大きく異なるという事実である。また、クラスタリング係数は 0.63 となっている。

これらのデータは、現実のネットワークのリンクは合理的な意思のもとで、あるいは戦略的、相互依存関係を背景に形成されていることを示している。ネットワークの形成が必ずしもすべてランダムには形成されていないことを意味する。Goyal, et al.(2006) によれば、観測された次数分布はランダム・ネットワークが生成するポアソン分布に比較して、かなり大きな分散を持つ。Jackson and Rogers(2007) がより鮮明に描き出したように、現実の社会ネットワークにおける次数分布の多くはべき乗則分布 (power law) に近い分布をしていると言える。これらの研究は、社会経済ネットワークにおけるリンクの形成が、ポアソン分布に従うように全く不規則にランダムに形成されている訳ではないが、ある種の戦略的な意思のもとで、ある程度の不規則性を持ちつつも形成されている事実を明らかにしている。

3 ネットワークの基本モデル

3.1 Erdős and Rényi 型ランダムネットワーク

最も簡単なランダムグラフはノード総数 n とリンク (エッジ) 数 m が固定されているネットワークである。このネットワークはランダムに選び出した二つのノード間にリンクを張る操作を m 回繰り返すと作成できる。このランダムグラフを $G(n, m)$ と表記する。ノードの選択がランダムであるから、全く同じリンク構造を持つランダムグラフは得られない。ネットワークのサンプルが作成できるだけである。このランダムグラフ $G(n, m)$ の平均リンク数は m で、ノード総数は n なので、各ノード当たり平均次数は $m/(n/2)$ である。何故なら、一つのリンクには両端に総計で二つのノードが接続されているからである。

通常、分析の便宜上から、ランダムグラフはリンク数を与えないで、リンク接続確率 p を与えた形式、

$G(n, p)$ で表現される。このランダムグラフで、総リンク数が m になる確率は

$$\Pr(G(n, m)) = p^m (1 - p)^{n(n-1)/2 - m}$$

となる。この形式で理論展開した初期の研究が *Erdős and Rényi* によって行われた。

$N = \{1, 2, \dots, n\}$ から、任意に選び出した二つのノード間にリンクが張られる確率（リンク形成確率）が、各ノードの位置に依存せず、同一の p であるとする。ただし、 $0 < p < 1$ である。ある一つのノードを選択すると、残りのノードは $n - 1$ 個の中から選択されるので、あるノードに d 個のリンクが接続している確率は、

$$p^d (1 - p)^{n-1-d}$$

となる。任意に選ばれたノードが次数 d となる確率は

$$Pr(d) = \binom{n-1}{d} p^d (1 - p)^{n-1-d}$$

である。これは 2 項分布 (*binomial distribution*) である。ランダムグラフの平均リンク総数は、 $M = n(n-1)/2$ とおくと、

$$\langle m \rangle = \sum_{m=0}^M m \Pr(G(n, m)) = pM = \frac{n(n-1)}{2} p$$

と計算できる。この結果は直観的にも正しい。何故なら、形成可能なノードの組合せ数とリンク形成確率の積になっているから。

リンク総数が m であるとき、ネットワークの平均次数は $2m/n$ になることは分かっている。リンク形成確率を用いて平均次数 $\langle d \rangle$ の計算をすると、

$$\langle d \rangle = \sum_{d=0}^{n-1} Pr(d) d = (n-1)p$$

平均次数は $\langle d \rangle = (n-1)p$ である。各ノードのクラスターリング係数が $\langle d \rangle / (n-1) = p$ に、集中化係数も p になることは自明である。

確率論の知識から、 n が非常に大きくなるときには、平均次数をパラメータ $\mu = (n-1)p$ としてもつポアソン分布になる。

$$\Pr(d) = \frac{e^{-\mu} \mu^d}{d!}.$$

この次数分布で表現されるネットワークをポアソン型ランダムネットワーク (*Poisson random graph*) という。下のグラフはランダムネットワークの例である。

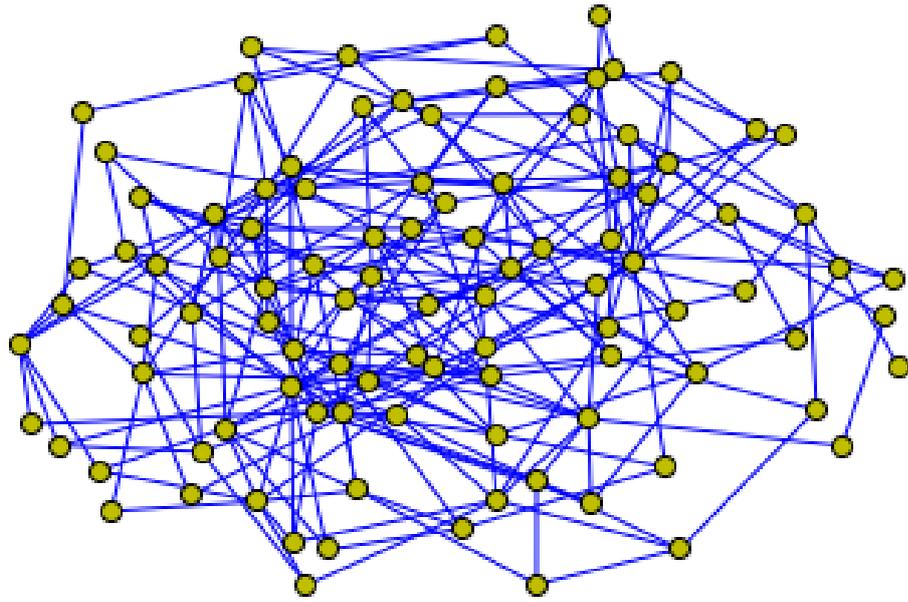


Fig3.1 Erdos Renyi 型ランダムネットワークの例 : $n = 100, p = 0.05$

これはノード数 $n = 100$ 、 $p = 0.05$ と与えたとき、Python/Netwrokx を用いて描いたものである。次数分布は以下ようになる。理論上は、平均次数は $\langle d \rangle = 4.95$ である。

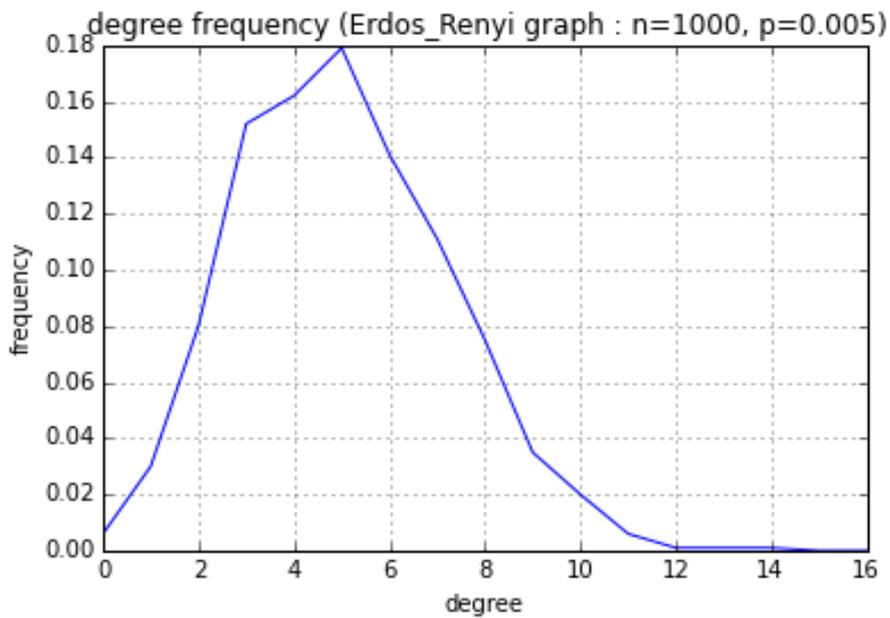


Fig3. 2 Erdos Renyi 型ランダムネットワークの次数分布 : $n = 1000, p = 0.005$

ランダムネットワーク理論の前提は徹底した平等主義である。つまり、リンクは完全にランダムに連結さ

れ、どのノードも同じ確率でリンクを獲得する。ラスベガスで大当たりを取るチャンスは誰にも平等にあるのと同様に。しかし、運の悪いノードと運の良いノードが出てくる。運の良いノードは多数のリンクを獲得するが、運の悪いノードはリンクを一つも獲得できない(ある者は貧しく、ある者は豊かになる)。しかし、エルデシュとレーニイの予測によれば、ネットワークにおけるノード数が十分大きくなり、リンクが完全にランダムに連結されるとすると、ほとんどすべてのノードは近似的に同数のリンクを持つようになる。つまり、各ノードが持つリンク数の確率分布はポアソン分布に従う*⁵。ノードの大部分は平均的ノードと同数のリンク数を持ち、平均からのずれが大きいリンク数のノードはごく少数にすぎない。われわれの大多数はほぼ同数の友人や知人を持つ。平均を著しく逸脱する人たちはごく少数である。

任意に2つのノードを選び、サイコロを振って、リンクを連結することを繰り返すと、多くのノードからなるネットワークができる(最初に、ノードの総数は所与とする)。リンク連結を繰り返すことは、リンク形成確率 p を増大させることに対応する。 p が大きくなって、すべてのノードが少なくとも一つのリンクをもつ程度にリンクが形成された段階で、奇跡が起こる。巨大クラスター(「巨大コンポーネント」)が出現する。大半のノードが一つの巨大なクラスターに属して、その中のどのノードから出発してもそのクラスター内のどのノードへも到達できる。物理学では、「パーコレーション」(相転移 *phase transition*) とよぶ。水が凍る瞬間のような現象に対応する。リンク数が臨界数に達するとき、ネットワークに根本的な変化が生じる。

この段階で、私達は、グローバルな社会的ネットワーク(巨大クラスター)から誰一人逃げられないことになる。地球上のすべての人と面識が無いが、任意に選ばれた二人をつなぐ経路は必ず存在する。それは、なぜか?エルデシュとレーニイの答えはこうである。巨大クラスターが発生するための条件が一人一人が持つ友人数が1以上であることになるから。ネットワークにおけるリンク数の閾値が1を越えると、巨大クラスターが生まれる。反対に、各ノードのリンク数が平均して1未満であるとき、ネットワークは相互に孤立した小さなクラスターに分裂している。

$n = 50$ のとき、 p の値を0から少しずつ大きくすると、ネットワーク内のノード間接続の状態がどのように変化するかを *NetworkX* などのソフトを用いてシミュレーション実験することができる。 $p = 1/n^2 = 0.0004$ で一つのリンクが形成され、 $p = n^{-3/2} = 0.003$ で二つ以上のノードが連結される。 $p = 0.01$ に到達すると、3つのノードからなるコンポーネントが登場し、 $p = 1/n = 0.02$ でサイクルが登場する。 p がこれ以上の値では、巨大コンポーネントが生まれる。さらに、 $p = \log n/n = 0.08$ の値を超えると、ネットワーク内のすべてのノードが連結される。このように、 $p(n)$ の値がある閾値を超えて増大するとき、ネットワークの連結性に構造的な変化をもたらす。これをネットワークの相転移現象 (*phase transition*) という。下のグラフがこの様子を描いている。

*⁵ 簡単な証明は、確率論のテキストや Newman(2010)chapt.12 を参照して下さい。

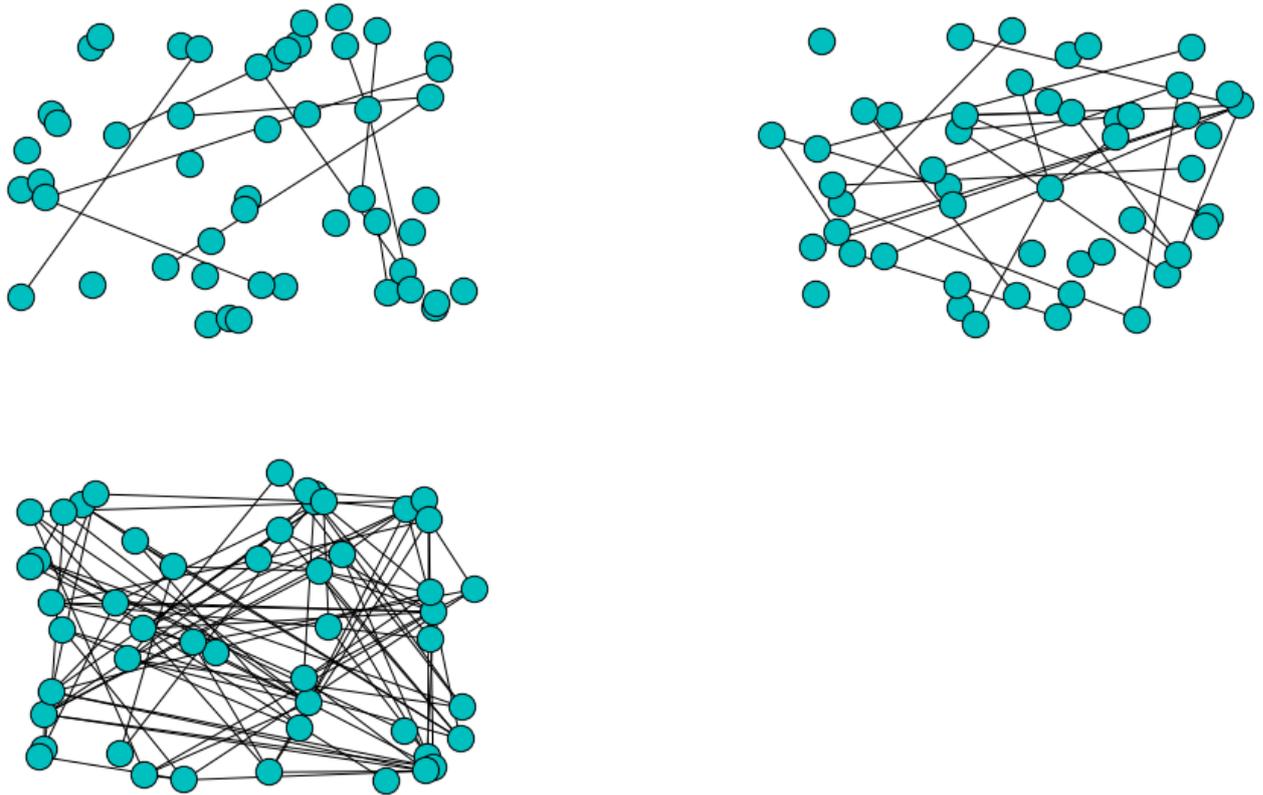


Fig3.3 Erdos Renyi 型ランダムネットワークの相転移現象: $p = 0.01$, $p = 0.02$, $p = 0.08$ の順に描かれている

社会学によれば、多くの方は 200 人から 5000 人の名前を上げることができる。平均的なニューロンは、数十から数千のニューロンとリンクしている。小さな会社でも、数百のサプライヤーや顧客とリンクしている。(大企業は数百万のリンクをもつ。) 現実のネットワークは 1 という閾値をはるかに超えたリンク数を持つ。ランダム・ネットワーク理論によれば、平均リンク数が 1 を越えると、巨大クラスターに含まれないノードの個数は指数関数的に減少する。結果として、われわれは、何者も逃れられない詰んだ網の目のようなネットワークに属する。

以上の直観的なストーリーを理論的に裏付けることにしよう。 n が非常に大きくなりネットワークが大規模になるとき、各ノードの平均次数、次数分布、ネットワークの直径や平均パスの長さの大きさがどのようになるのかを分析する。 n が無限大になるとき、ネットワークの各種統計量が有限の値に漸近 (収束) する確率を

計算しなければならない。

リンク形成確率をネットワークの規模 n に依存する形で表現するので、 $p = p(n)$ と表現する。上の例からも理解できる通り、ネットワーク内のすべてのノードが連結している (*connected*) ためには、リンク形成確率 $p(n)$ が閾値を超えなければならない。この閾値を表現する関数を閾値関数 (*threshold function*) といい、 $t(n)$ と表記する。上の例では、閾値関数 $t(n)$ は、

$$t(n) = \frac{\log n}{n}$$

となっている*6。このとき、ネットワークが連結しているためには

$$p(n) \geq \frac{\log n}{n}$$

が成立する必要がある。この条件式は、 n が非常に大きいならば平均次数が $\langle d \rangle = np(n)$ となるので、

$$\langle d \rangle > \log n$$

を意味する。この条件は、ネットワークのサイズが大きくなるに伴って、平均次数も増大しなければならないことを含意する。 $\langle d \rangle < \log n$ が成立するとき、ネットワークは、すべてのノードがリンク連結されていないので、互いにリンク接続していないサブネットワークやクラスターから構成されている。

現実のネットワークでは、 n が非常に大きくなっても、平均次数はそれほど大きくならないか、ほとんど変化しない。だから、 $\langle d \rangle < \log n$ が成立していると想定される。ネットワークの規模が拡大しても、平均次数がそれほど変化しないのであれば、リンク接続が失われるノードが存在し、ネットワークが異なる部分コンポーネントに分解することになる。これらのコンポーネントのサイズはどのような特徴を持つのだろうか？巨大なコンポーネントができるのだろうか、それとも非常に小さな規模のコンポーネントになるのだろうか？

Erdős and Rényi の証明を援用して、この疑問に答える。ネットワークのサイズ n が大きくなるにつれて、その規模に比例してコンポーネントのサイズも拡大するようなコンポーネントを巨大コンポーネントと呼ぶことにする。巨大コンポーネントに属さないノードの割合を u とする。この巨大コンポーネントに属するノードの比率は $1 - u$ である。巨大コンポーネントが存在するとき、 $1 - u > 0$ である。あるノード i がこの巨大コンポーネントに属さないのであれば、他のノード j へのリンクを持たないか、リンクした先のノード j が巨大コンポーネントに属さない。前者の確率は $1 - p(n)$ である。後者の確率は、 $p(n)u$ となる。よって、ノード i がノード j を経由して巨大コンポーネントに属さない確率は

$$1 - p(n) + p(n)u = 1 - p(n)(1 - u) = 1 - \frac{\langle d \rangle}{n - 1}(1 - u)$$

と与えられる。あるノード i が他のすべてのノードを介しても巨大コンポーネントに属さない確率は

$$u = \left[1 - \frac{\langle d \rangle}{n - 1}(1 - u)\right]^{n-1}$$

である。両辺の対数を取ると、

$$\ln u = (n - 1) \ln \left[1 - \frac{\langle d \rangle}{n - 1}(1 - u)\right]$$

*6 このような閾値関数が存在することを初めて証明したのが、Erdős and Rényi(1959) に他ならない。より詳しい説明は、Jackson(2008) や Vega-Redondo(2007) を参照のこと。

となる。 n が非常に大きければ、近似式

$$\ln\left[1 - \frac{\langle d \rangle}{n-1}(1-u)\right] \approx -\frac{\langle d \rangle}{n-1}(1-u)$$

が成立する。よって、

$$\ln u = -\langle d \rangle(1-u)$$

となり、

$$u = e^{-\langle d \rangle(1-u)}$$

が得られる。巨大コンポーネントに含まれている確率 S は $S = 1 - u$ だから、

$$S = 1 - e^{-\langle d \rangle S} \tag{1}$$

が成立する。右辺を S で微分すると、

$$\langle d \rangle e^{-\langle d \rangle S}$$

が得られる。方程式 (1) の解 S が正の値を取る条件は、右辺のグラフの $S = 0$ での傾きが 1 以上となることである。(右辺は S の凹関数である。) すなわち、

$$\langle d \rangle = (n-1)p(n) > 1$$

が成立するならば、巨大コンポーネントが存在し、そのサイズは方程式 (1) の解 S で与えられる。 n が十分大きいときには、

$$p(n) > 1/n$$

ならば、一つの巨大なコンポーネントが生まれる。反対に、 $p(n) < 1/n$ ならば、ネットワーク内の各コンポーネントのサイズは相対的に無視できるほど小さくなり、ほとんどのノードはツリー状のネットワークを構成する。

ポアソン・ランダムネットワーク・モデルでは、各ノードのリンク接続は互いに独立に、確率 $p(n)$ で一様に形成される。平均次数を $\langle d \rangle$ として、各ノードが他のノードとリンクを持つ確率は $p(n) = \langle d \rangle/n$ である。クラスタリング係数は $\langle d \rangle/n$ である。平均次数 $\langle d \rangle$ が n の増加より早く増大しない限り、 n が非常に大きくなるにつれて、クラスタリング係数 $\langle d \rangle$ はゼロに収束する。この性質は現実の社会ネットワークのクラスタリング係数と比較して小さすぎるので、現実のネットワークを説明するためには、モデルの拡張が必要となる。

3.2 配列モデル

ノードの次数分布を具体的に与えて、この次数分布と同じ分布を生み出すように形成されたランダムネットワークを配列モデル (*configuration model*) という。この配列モデルの性質について分析する。ネットワークのノード総計 n を与え、各ノード i の次数 $\eta_i, i \in N$ を与える。各 $k = 0, 1, \dots, n-1$ に対して

$$Pr(k) = \frac{|\{i \in N : \eta_i = k\}|}{n}$$

と次数分布を定める。各ノードの次数の数だけ各ノードのインデックスのコピーを作成し、それらを要素とする集合を W とする。集合 W はノード名のインデックスを要素とする集合で、総計で $Z = \sum_{k=0}^{n-1} \eta_k$ 個の要素を持つ。配列モデルは、この集合の中から要素をランダムに 2 個取り出し、取り出された二つのノード間にリンクを形成する。これを Z 回行う。このようなモデル構成では、同一のノード同士に複数のリンクが形成されてしまう可能性を排除できない。ある的確な条件が満たされる場合、統計計算上は、こうした病理現象を避けることができるので、ネットワークモデルとして有効であるといえる。

配列モデルでも、ランダムネットワークと同じように、各ノードの次数間には相関関係がない。しかし、次数 d を持つノードの隣人ノードの次数分布が $Pr(d)$ になるとは限らない。次数 d のノードの隣人数は d 個で、各隣人ノードが次数 d を持つ確率は $Pr(d)$ である。ノード j が次数 k を持っているとき、ノード i がノード j にリンクする確率は、ノード j が次数 1 を持つ場合に比較して、 k 倍大きい。従って、隣人ノードの次数分布 $\xi(\cdot)$ は $Pr(d)d$ に比例する。次数分布を確率分布になるように正規化すれば、隣人の次数分布は

$$\xi(d) = \frac{Pr(d)d}{\sum_{d=0}^{\infty} Pr(d)d} \quad (2)$$

となる。この分布関数を用いると、あるノード i がリンク接続した先の隣人 j がリンクするノード i 以外の隣人数の平均値は

$$\sum_{d=0}^{\infty} (d-1)\xi(d) = \frac{\sum_{d=0}^{\infty} d(d-1)Pr(d)}{\sum_{d=0}^{\infty} Pr(d)d} = \frac{\langle d^2 \rangle}{\langle d \rangle} - 1$$

である。ここで、 $\langle d^2 \rangle$ は次数分布 $Pr(x)$ のもとでの d^2 の期待値を表す。当然、 $d_1 = \langle d \rangle$ とおく。あるノード i は平均的に隣人数 $\langle d \rangle$ をもつ。隣人の隣人総数の期待値を d_2 と表記すると、 d_2 は

$$d_2 = \left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1 \right) \langle d \rangle = \langle d^2 \rangle - \langle d \rangle \quad (3)$$

と計算される。同じようにして、 r 個のリンクを経由した先のノードの隣人総数の期待値 d_r は

$$d_r = \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} d_{r-1} = \left(\frac{d_2}{d_1} \right)^{r-1} d_1$$

となることが知られている*7。 $\langle d \rangle > 0$ であれば、 $d_2 > d_1$ である限り、各ノード i はネットワーク全体のノードに到達可能であることを意味する。言い換えると、あるノードで新しい情報が生まれるとき、その情報がほぼネットワーク全体に広がって伝搬する条件が $d_2 > d_1$ となることを含意する。すなわち、

$$\langle d^2 \rangle - 2\langle d \rangle > 0 \quad (4)$$

が成立するならば、あるノードで起こったイノベーションはほぼネットワーク全体に拡散して伝染する。言い換えると、上記の不等式が成立するとき、巨大コンポーネントが存在し、ネットワークの大部分を覆っている。このとき、ネットワークの直径を l とすると、

$$\left(\frac{d_2}{d_1} \right)^{l-1} d_1 \approx n$$

が成立つ。変形すると

$$l \approx \frac{\log(n/d_1)}{\log(d_2/d_1)} + 1 \quad (5)$$

*7 Vega-Redondo(2007) を参照のこと。

となる。ネットワーク上の任意の二つのノード間の距離は最大で直径なので、ノード間のパスの平均距離もこの大きさに近いと想定される。

ポアソン・ランダムネットワークでは、 $\langle d^2 \rangle = \langle d \rangle + \langle d \rangle^2$ なので、上の不等式 (4) は $\langle d \rangle^2 > \langle d \rangle$ と簡単化できる。つまり、 $\langle d \rangle > 1$ が成立するとき、巨大コンポーネントが存在する。言い換えると、平均次数が 1 以上になるとき、巨大コンポーネントが発生する。上で指摘した通り、ポアソン・ランダムネットワークで巨大コンポーネントが存在する条件は $p(n) > 1/n$ 、つまり、 $\langle d \rangle = np(n) > 1$ が成立するときである。ポアソン・ランダムネットワークでは、 $d_2/d_1 = d_1 = \langle d \rangle$ なので、

$$l \approx \frac{\log(n/d_1)}{\log(d_2/d_1)} + 1 = \frac{\log n}{\log d_1}$$

となる。このようにして、配列モデルは Erdős and Rényi 型ランダムネットワークを含む一般的なランダムネットワークを拡張したモデルとなっている。

配列モデルはポアソン・ランダムモデルをより一般化したものであるが、ランダム・モデルとして共通の弱点を持つ。問題点は、隣人同士の間に関連関係が見られないことから生じる。ノード数が増加するにつれて、群集化係数がゼロになってしまい、現実の社会的なネットワークの事例に矛盾する。

4 スモールワールド・モデルとスケールフリー・モデル

4.1 スモールワールド・モデル

ランダムネットワーク理論では、複雑さ＝ランダムさで見なして定式化してきた。複雑過ぎるネットワーク＝ランダムなネットワークと想定した。現実にはランダムに形成されているのでしょうか？たいていの人はこの世界はランダムだとは思わない。複雑なシステムの背後には、何らかの秩序が存在するに違いない、と思っている。現実世界のネットワークはランダムネットワークの世界とは異なる組織原理をもつ、と思っている。ランダムネットワークでは、私の友人の二人が知り合いである可能性は、ベニスのゴンドラ乗りとエスキモーの漁師が知り合いである可能性と同じになっている。社会は密なクラスター同士が弱く連結したネットワークと予想されるので、この可能性は現実では起こりえない。ネットワークの新しいモデルへの一歩が必要となってくる。

マーク・グラノヴェッター (1973) は、強い絆 (*strong tie*) と弱い絆 (*weak tie*) という概念を提案して、「人はいかにして職を得るのか」の研究の中で、この課題に挑戦した。アンケートからの結論は、いまの職を得るときに「弱い社会的絆」が重要な役割を果たしている事実を明らかにした。社会は高度に連結されたクラスター構造にあってっており、これらのクラスターが外部の数本のリンクにより互いに関係している。高度に連結したクラスター内では、緊密な人間関係が構造化されている、一方で、親しい友人は共通の情報を持っているので、職探しには役に立たない。弱い絆の友達の方が知らない情報を持っているので、職探しには有効な情報をもたらしてくれる。

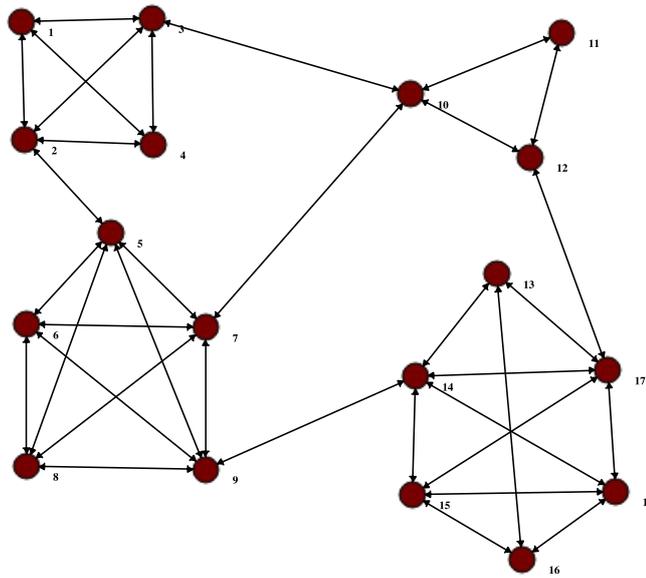


Fig.4.1 弱い絆と強い絆

社会的ネットワークは、強い絆を持つクラスターの集まりであり、このクラスター同士が弱い絆によって互いにリンクしているネットワークになっている。

私の親友が4人いるとき、この4人が全員友達ならば、4人の間に6本のリンクが形成されている。4人全員が友達でないときは、リンク数は6以下となる。4本のリンクがあったとすると、クラスタリング係数 $= 4/6 = 0.666 < 1$ となる。クラスタリング係数が1のとき、私の友人同士はすべて知り合いである。クラスタリング係数がゼロに近いとき、私の友人は互いに知り合いでない。多数の緊密なクラスターが弱い絆で結ばれている社会では、クラスタリング係数は高くなる。

Watts and Strogatz(1998)は、ネットワーク内各ノード間のパスの距離が短い、群集化係数がゼロよりも相当大きい値を持つという2条件を満たすネットワークをスモール・ワールドと呼んでいる。彼らは、群集化と弱い結びつきという2種類の社会ネットワークの特徴を格子モデルを用いてモデル化した。モデルは地理的近接性あるいは密接な人間関係を2次元格子上のノードで表現し、近接するノードに連結された一つのリンクを取り除き、そのリンク先をより遠い位置にあるノードをランダムに選びそれに連結する。

Watts and Strogatz(1998)のモデルを説明する前に、比較のためネットワークの格子モデルを取上げる。最も単純な格子モデルは1次元のケースである。1次元格子モデルの典型は、円周上に等間隔で n 個のノードが配置されている場合である。各ノードから距離 $k(k < n - 1)$ 以内にあるノードにリンクを連結する。すると、各ノードの次数は同一で $\langle d \rangle = c = 2k$ で、リンクの総計は $\langle d \rangle n = cn$ となる。3ノードを連結する閉じたパスの総数を計算する。各ノードからのリンク数は前後(左右)同数になっている。パスが閉じているためには、3ステップ目にいるノードまでの距離は $c/2$ 以下でなければならない。そうでないと、出発ノードに戻れない。最初の2ステップは同一の方向にあるノードに行かなければ行けないので、閉じたパスの組合せ数は

$$\binom{c/2}{2} = \frac{c}{4} \left(\frac{c}{2} - 1 \right)$$

である。二人の隣人を取り出す組合せの数は

$$\binom{c}{2} = \frac{c}{2}(c-1)$$

であるので、クラスタリング係数は

$$Cl = \frac{3\frac{c}{4}(\frac{c}{2}-1)}{\frac{c}{2}(c-1)} = \frac{3(c-2)}{4(c-1)}$$

と計算される。クラスタリング係数は $c=2$ のとき最小値 0 を取り、 c の値が無限大になるとき、 $3/4$ となる。ネットワークのサイズが大きくなっても、クラスタリング係数は小さな値にならない。この性質はランダムグラフの特徴と際立った差異となっている。

しかし、各ノードの次数分布は同一なので、いわゆる、正規グラフなので、現実のネットワークに特徴的な次数分布とは大きく異なる。言い換えると、スモールワールドではなく、ラージワールドになっている。ちなみに、このネットワークの平均距離を計算すると、 $n/2c$ となる*⁸。Watts and Strogatz(1998) は次数分布のランダムさ、つまり、ショートカット・パスを格子モデルに導入した。

円周上に n 個のノードを配置し、各ノードからの距離が k 以内に位置するノードに一本づつリンクを連結する。全体で $2nk$ 本のリンクが存在する。その後、それらのリンクの中から、確率 p で一つのリンクを任意に選択して、そのリンク先を円周上に配置されたノードからランダムに選び、このノードを新しいリンク先に変更して、そこに再連結する。各リンクが再連結される確率は p である。再連結されるパス（ショートカット）の総計は平均的に $2nkp$ 個となる。 $p=0$ のとき、完全な正規格子モデルで、 $p=1$ のケースは完全なランダムネットワークに対応する。再連結の確率 p の値を 0 から 1 に変化させることにより、正規格子モデルと完全ランダムネットワークの中間のネットワークの特徴を持つネットワークを作成できる。

このモデルにおける平均パスやクラスタリング係数などの各統計量の解析的な表現は難しい。Watts and Strogatz(1998) が示した通り、 $p \rightarrow 0$ になるとき、パスの平均距離は $n/2c$ 、群集化係数は $3/4$ となる。反対に、 $p \rightarrow 1$ になるとき、パスの平均長さは $\ln(n)/\ln(c)$ 、群集化係数は c/n となる。数値計算のシミュレーションにより、ショートカット作成確率 p の値が $0 < p < 1$ のある範囲にある場合、スモールワールドが出現することがわかる。

*⁸ この計算は Newman(2010) chapt.15 にある。

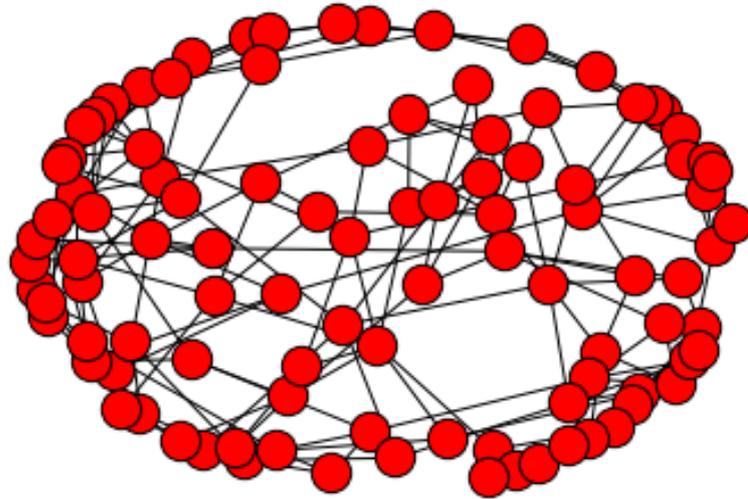


Fig.4.2a スモールワールド・ネットワークの例： $n = 100, m = 4, p = 0.2$

ここでの例は、ノード総数 100 で、各ノードが 4 人の隣人に連結している初期状態から、確率 0.2 でこのリンクを外してランダムに選んだ他のノードに連結される。Watts and Strogatz 型のモデルである。Networkx を用いて描いたものである。以下のグラフはこのスモールワールド・ネットワークに対応する次数分布を描いたものである。

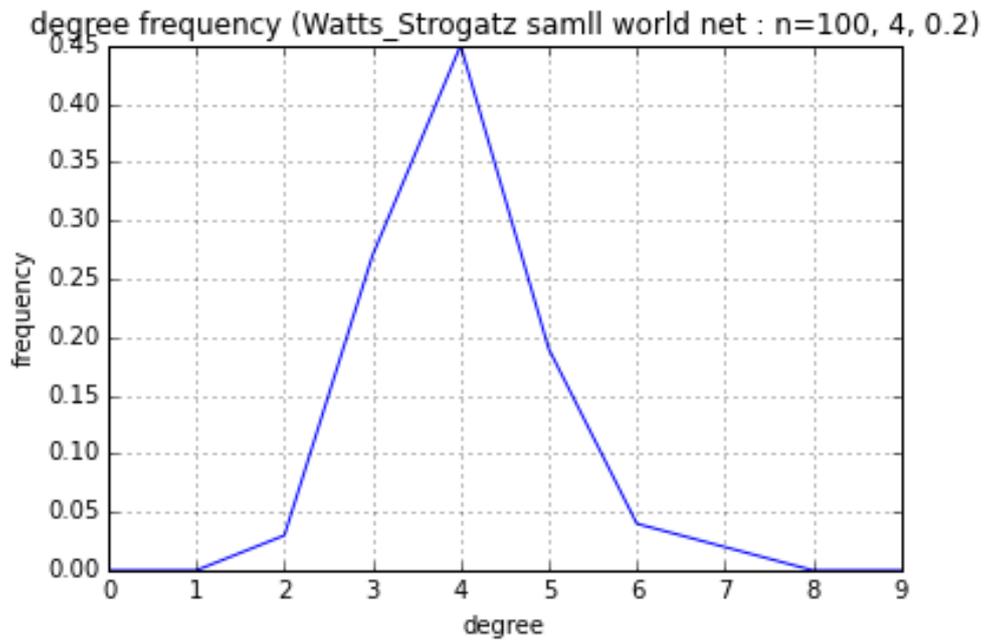


Fig.4.2b スモールワールド・ネットワークの次数分布： $n = 100, m = 4, p = 0.2$

WSモデルは現実のネットワークで観測される群集化とパスの距離に関する特徴を定式化することに成功しているが、あるノードから任意に指定されたノードへのパスの長さが長すぎるという欠点をもつ。Milgramの実験に見られるように、現実の社会では6ステップ以内で任意の人物同士を結ぶパスが存在している。WSモデルにおけるリンクは、近傍のノード同士を結ぶ強い結びつき（近距離リンク）と、遠い距離に位置するノードとを結ぶ弱い結びつき（長距離リンク）という2種類のリンクから構成されているが、長距離リンクの連結先がランダムに選択される。WSモデルの欠点は弱い結びつきと呼ばれる長距離リンクの接続先がランダムに選択されている設定に起因する。

4.2 スケールフリー・モデル

ワッツ＝ストロガツツのモデルは現実のネットワークで観測される大きい群集化と短いパス距離に関する特徴を定式化することに成功しているが、平均数よりもはるかに多いリンク数を持つノードの存在は認められない。平均数よりもはるかに多いリンク数を持つハブの存在を説明するモデルの提案が必要である。1999年、Barabási and Albert(1999)はスケールフリー・ネットワーク・モデルを提案した。

イタリアの経済学者V・パレートは『政治経済学の原理 (Cours d'economie politique)』(1897年)を発表し、その中で、パレート分布(80対20の法則)を提案した*9。エンドウ豆の収穫量の80%は20%のさやからもたらされる。イタリア国土の80%は人口の20%に所有されている。経済学におけるマーフィーの法則とも言える。収益の80%は従業員の20%があげている、顧客トラブルの80%は顧客の20%が持ち込んでいる、決定事項の80%は会議時間の20%内で処理されている、等々。

自然界に現れる多くの統計量の分布は釣り鐘型の形状を持つ。統計学の言葉でいえば、正規分布(normal distribution)に近い分布形をしている。身長分布は150センチから190センチ内にほとんど収まり、その中にピークがある。しかし、ウェブ・ページが持つリンク数を調べたら、釣り鐘の形状をしていなかった。

Albert, Jeong and Barabási(1999)は、world wide web上のウェブ・ページのリンク数分布は冪乗則(power law distribution)に従うことを発見した。冪乗則(ベキ法則)の分布はリンク数 k をもつウェブ・ページの数を $N(k)$ とすると、 $N(k) \sim k^{-\gamma}$ と表現される。両対数グラフに表現すると、右下がりの直線になる。指数の値 γ はその傾きの大きさになる。インカミング(in-coming)リンク数では、 $k \approx 2.1$ 、アウトゴーイング(out-going)リンク数では、 $k \approx 2.5$ となっている事実が発見された。

Barabási and Albert(1999)でも、ハリウッド俳優のリンク数もベキ法則に従うことも発見された。細胞内神経系も、 k の分子と相互作用する分子の数もベキ法則になっている。前節で見た通り、論文引用件数もベキ法則に従う。個々のネットワークは独自のベキ(冪)指数を持ち、その値は2から3になる。Soramaäki et al(2006)や稲岡・他(2003)の研究によれば、金融ネットワークの次数分布も冪乗則に従うことが知られている。ランダムネットワークとベキ法則に従うネットワークは、見た目も実際の構造も衝撃的に異なる。高速道路網にはハブは存在しないが、航空会社のルートマップを見れば明白になるように非常に多数の航空会社が乗り入れている空港がある。十数本の高速道路が合流する都市は存在しても、50本を超える高速道路を連結したハブの役割をする都市は存在しない。他方、航空会社のルートマップには、シカゴ、ダラス、デンバー、アトランタなどのハブ空港が存在する。

ベキ法則の特徴は小さな事象が無数に存在すると同時に、一握りの極めて大きな事象が共存することである。

*9 統計量 k のPareto分布は、累積分布形で $F(k) = 1 - k^{-\gamma+1}$ と表現され、密度分布形では $f(k) = (\gamma - 1)k^{-\gamma}$ と表現される。

る。鐘型の分布では、小さな事象は起こるが、非常に大きな事象は絶対に起こりえない。裾野は指数関数的に減少する。また、ベキ分布の特徴は、ピークが存在しない、平均的ノードがないことでもある。ネットワークを特徴づける平均などのスケール（尺度）が存在しない。だから、スケールフリー・ネットワークとも呼ばれる。ベキ法則に従うスケールフリー・ネットワークでは、少数のハブが存在し、そのハブがネットワークのトポロジーを基本的に決定する。ハブの登場と存在が、ネットワークの進化を支配する重要な組織原理になっている。

Barabási and Albert(1999) は、ネットワーク形成の動的な過程を明示的にモデルに導入することによって、次数分布がベキ乗則に従うことを示した。動的なリンク形成に関する第 1 の仮定は、新しいノードが生まれ、この新しいノードと既存のあるノードと間に新しいリンクが形成されるというネットワーク成長についての仮定である。第 2 の仮定は、新しいノードのリンク先の選択が、各既存ノードのそのときの影響力の大きさに偏向するようなメカニズムが働くという仮定である。この仮定を「優先的付着」(preferential attachment) の仮定という。豊かな人はより豊かになるというメカニズムを組み込むことに他ならない。

始めに、新しいノードのリンク先がランダムに選ばれるケースについて考える。時刻 $t = 1$ で二つのノードが存在し、これらは連結されているとする。これらのノードのインデックス（名前）を $0, 1$ とする。時刻が進み $t = 2$ になると、新しいノードが生まれる。この新しいノードは既存ノードの中から一様にランダムに選択された一つのノードにリンクされる。時間が進み、 $t = 3$ になると、また新しいノードが生まれ、この新しいノードは既存ノードの中から一様にランダムに選択された一つのノードにリンクされる。このようにして、時間の進行とともに、ノード数が増加し、リンク数も増えて行く。時間の進行とともに、ネットワークの規模は拡大して行く。ネットワークの規模が非常に大きくなる時、リンクの次数の確率分布は定常状態になって行く。この状態になると、次数分布が満たすべき関係式は

$$2p(k) = p(k-1) + \delta_{k,1}$$

となることが証明できる。これを解くと

$$p(k) = 2^{-k}, k = 1, 2, \dots$$

となる。これは指数的ネットワーク (exponential networks) の特殊ケースとなっている。この次数分布は、定性的な特徴としては、ポアソン・ネットワークから得られる分布と大きく異なるものではない。

指数的分布を持つネットワークの次数分布は一般的に

$$p(k) = Ce^{-\lambda k}, \lambda > 0$$

と表現される。定数 C は確率分布であるための条件 $\sum_k p(k) = 1$ から $C = 1 - e^{-\lambda}$ となる。だから、指数ネットワークの次数分布は一般的に

$$p(k) = (1 - e^{-\lambda})e^{-\lambda k}, k = 1, 2, 3, \dots$$

と記述されている。

次に、Barabási and Albert(1999) が提案したリンク形成に優先的な付着性を導入したモデルを考える。新しいノードが生まれるとき、このノードのリンク先が各既存ノードの人気度（次数）の大小に依存するような確率分布に従うとする。時刻 t が進んで行くとき、任意に取り出されたノードが次数 k を持つ確率 $p(k)$ が従うべき方程式を導出することができる。時刻が無限大になり、 $t \rightarrow \infty$ になるとき、つまり、定常状態での次数分布 $p(k)$ は

$$p(k) = \frac{1}{2}[(k-1)p(k-1) - kp(k)] + \delta_{k,1}$$

に従うことが知られている。ここで、 $\delta_{k,1}$ はクロネッカーの関数である。

ネットワークが複雑になり巨大化するとき、次数 k を実数変数と見なし、連続系モデルで近似する方法が採用される。上記の差分方程式を微分方程式に変換すると、

$$p(k) = -\frac{1}{2} \frac{d[kp(k)]}{dk}$$

となる。これを解くと、

$$p(k) = Ak^{-3} \tag{6}$$

が得られる。ただし、 A は定数で、

$$\int_1^{\infty} Ak^{-3} dk = 1$$

を満たすように決定される。 $A = 2$ である。こうして、最終的に、次数分布がべき乗則になるネットワーク・モデルに到達する。こうして、リンク形成が優先的付着の原理に従う場合には、次数分布がべき乗則に

$$p(k) \propto k^{-\alpha}, \alpha = 3$$

に従うことを理論的に導出できる。

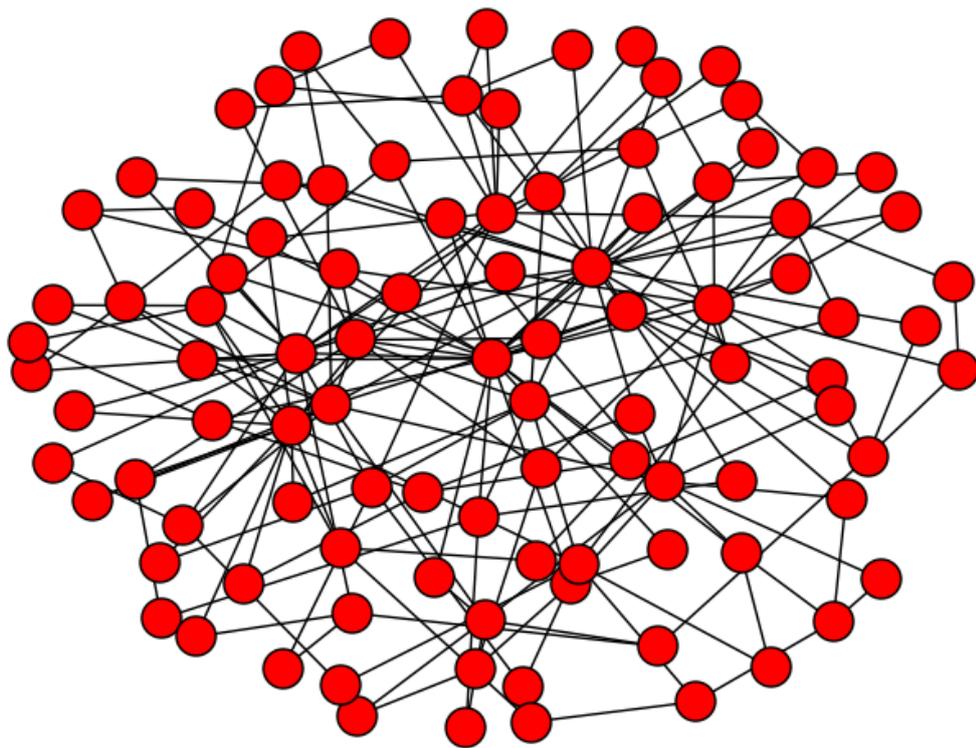


Fig 4.3a Barabási and Albert 型ネットワークの例 : $n = 100, m = 2$

これは、ノード数 $n = 100$ 、新しいノードがリンクできるノード数が $m = 2$ と与えたときのケースで、Python/Netwrokx を用いて描いたものである。このネットワークに対応する次数分布は下のように描かれる。

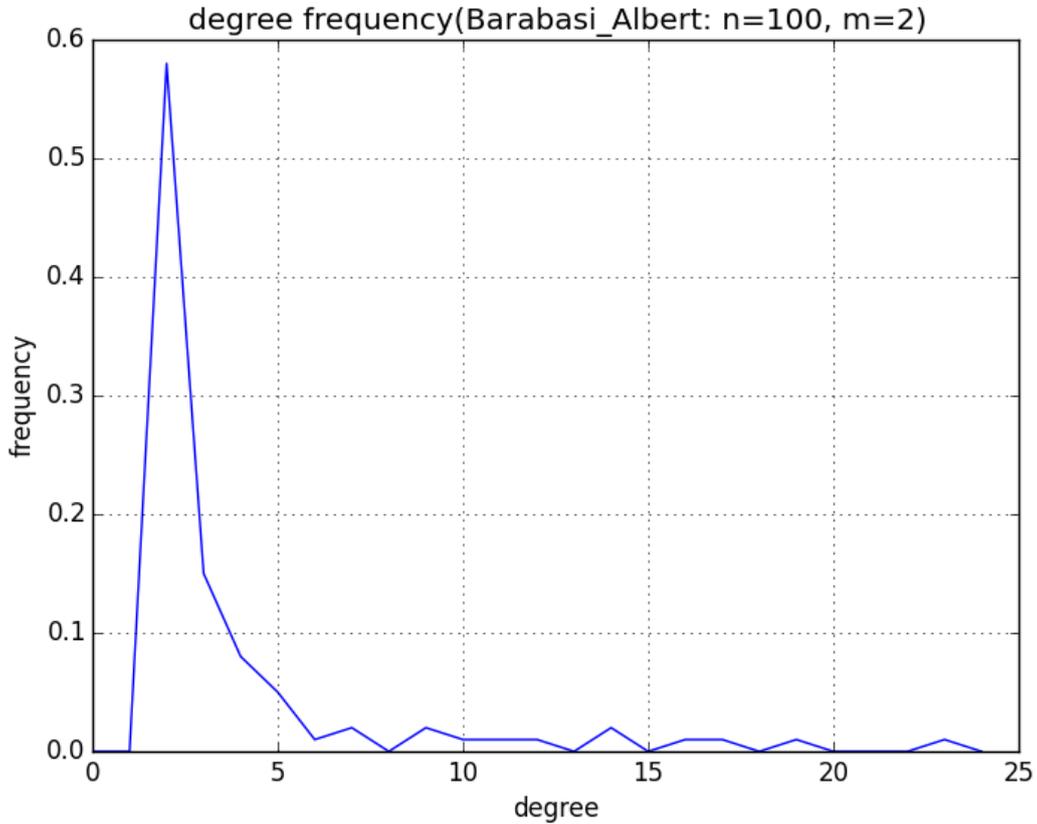


Fig 4.3b Barabási and Albert 型ネットワークの次数分布 : $n = 100, m = 2$

現実社会でのネットワークでは、冪乗則の指数は $2 < \alpha \leq 3$ となっている。この次数分布はスケールフリーであることを意味する。つまり、任意の正数 β に対して、

$$p(k) \propto p(\beta k)$$

が成立する。言い換えると、すべての次数 k において、 $p(k)/p(\beta k)$ は一定値となる。 $p(k)$ と $p(\beta k)$ のグラフを対数軸の平面に描くと、同じ傾きを持つ直線となる。

冪乗則ネットワークの純粹形における次数分布は

$$p(k) = Ck^{-\alpha}$$

と書かれる。 $k = 0$ のとき、この式は適用できない。通常、 $p(0) = 0$ とする。定数 C は、確率分布であるための条件から、

$$C \sum_{k=1}^{\infty} k^{-\alpha} = 1$$

を満たさなければならない。ここで、リーマン・ゼータ関数 (Rieman zeta function) $\zeta(\alpha)$ の定義

$$\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$$

を使用すると、 $C = 1/\zeta(\alpha)$ と書ける。この表現を用いると、冪乗則ネットワークの次数分布は

$$p(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}, \quad k = 1, 2, \dots \quad (7)$$

と表現されることになる。

この純粋冪乗則分布の際立った特徴の一つは、パレートの法則、ジッフの法則あるいはローレンツ曲線と呼ばれる性質が成立つことである*10。ネットワーク上で最高次数を持つノードの割合を P とし、この最高次数を持つノードに連結するノード数のネットワークに占める割合を W とすると、

$$W = P^{(\alpha-2)/(\alpha-1)}$$

なる関係が成立する。 $k_{min} = 20$ で、指数 $\alpha = 2.2$ を持つ WWW のベキ分布では、 $P = 0.5$ とおくと、 $W = 0.89$ が得られる。つまり、次数分布の上位 50 % を占めるノードのウェブページに連結されているリンク数の占める割合が全体の 89 % を占める。平均の半分以下の次数を持つノードへのリンク数は全体の 11 % にすぎないことが分かる。逆に、 $W = 0.5$ とおくと、 $P = 0.015$ となる。ネットワーク上の 50 % のリンク先は上位 1.5 % のノードによって占められている。

前節で演繹した通り、ネットワークに巨大コンポーネントが発生する閾値条件は不等式 (4) で与えられる。スケールフリー・ネットワークで巨大コンポーネントが生まれる条件をここで再掲すると、

$$\langle d^2 \rangle > 2\langle d \rangle$$

となる。

$$\begin{aligned} \langle d \rangle &= \sum_{k=0}^{\infty} k p(k) = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+1} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}, \\ \langle d^2 \rangle &= \sum_{k=0}^{\infty} k^2 p(k) = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+2} = \frac{\zeta(\alpha-2)}{\zeta(\alpha)} \end{aligned}$$

と計算できるので、スケールフリー・ネットワークにおける巨大コンポーネント存在の閾値条件は

$$\zeta(\alpha-2) > 2\zeta(\alpha-1)$$

となる。この式を解析的に解くことはできない。数値計算によれば、 $\alpha < 3.4788\dots$ となることが知られている。 $\alpha < 3.4788$ を満たす限り、冪乗則分布を持ついかなるネットワークにも巨大コンポーネントが存在する。

実は、冪乗則分布を持つどんなネットワークでも、 $\alpha < 3$ である限り、 $\zeta(\alpha-2)$ が無限大に発散する。 $\alpha > 2$

*10 Newman et al.(2005) を参照のこと。

である限り、 $\zeta(\alpha - 1)$ は有限値に留まる。つまり、 $2 < \alpha < 3$ である限り、冪乗則分布を持つどんなネットワークでも巨大コンポーネントが生まれる。 $\alpha \leq 2$ のときも、巨大コンポーネントが存在することは確かである。

純粋型の冪乗則分布の両辺の対数を取ると、

$$\ln p_k = \ln C - \alpha \ln k$$

が得られる。この式から、両軸を対数値とする平面上で、冪乗則のグラフは右下がりの直線となることが分かる。傾きの大きさは α である。以下の図は、Broder *et al.* (2000) が提供したデータを用いて描かれた、インターネットにおける WWW 間のハイパーリンク接続の次数分布である。横軸及び縦軸は対数値で取られている。

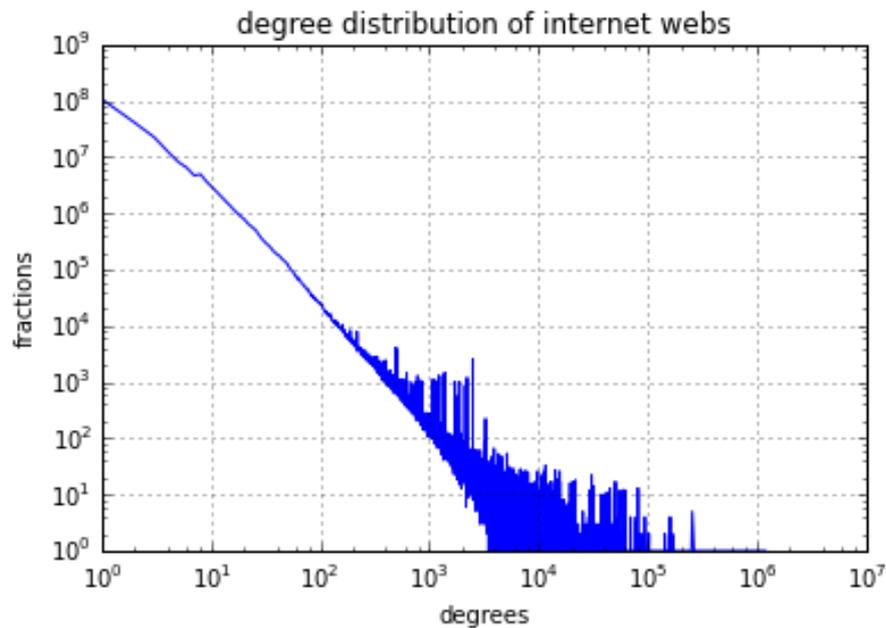


Fig.4.4 インターネットにおけるウェブページ・リンクの次数分布：対数グラフ

このインターネット WWW の次数分布は対数値を取った平面上で直線となっているように見る。Broder 達はこのネットワークは冪乗則分布に従っていると指摘した。

実際のデータを対数軸のグラフに描き、冪乗則分布に当てはまるか否かを検証するとき、いくつかの問題に直面する。次数が非常に大きなテールの領域では、ヒストグラムを描くためのデータ数は非常に少なくなる。つまり、データ数が非常に少ないならば、ノイズ・データの方が大きくなる可能性が生じる。上の図でもこの現象が見られる。このことの故に、描かれた分布が直線に従わない可能性が出てくる。この問題を避けるために取られる方法は、累積分布を使用することである。以下のような逆累積分布関数 (*the complement of the cumulative distribution function*)

$$P_k = \sum_{r=k}^{\infty} p_r$$

を定義する。 P_k は k 以上の次数を持つノードが占める比率である。つまり、任意に選ばれたノードが次数 k 以上を持つ確率となっている。

次数分布がテールの領域で冪乗則に従っていると。言い換えると、ある大きな次数 k_{min} よりも右側で冪乗則が成立つとすると、 $k \geq k_{min}$ に対して

$$P_k = C \sum_{r=k}^{\infty} r^{-\alpha}$$

が成立する。 k を実数として取り扱おうと、

$$P_k \approx \int_k^{\infty} r^{-\alpha} dr = \frac{C}{\alpha-1} k^{-(\alpha-1)}$$

なる関係式が得られる。従って、逆累積分布は指数を $-(\alpha-1)$ とする冪乗則分布に従う。この逆累積分布関数の概念を活用すれば、上記のデータ数不足に伴うエラー問題は解決する*11。多くの実証研究ではこのアプローチが採用されている。以下の図は、上で取上げたウェブリンクの次数分布の累積分布形状を描いている。

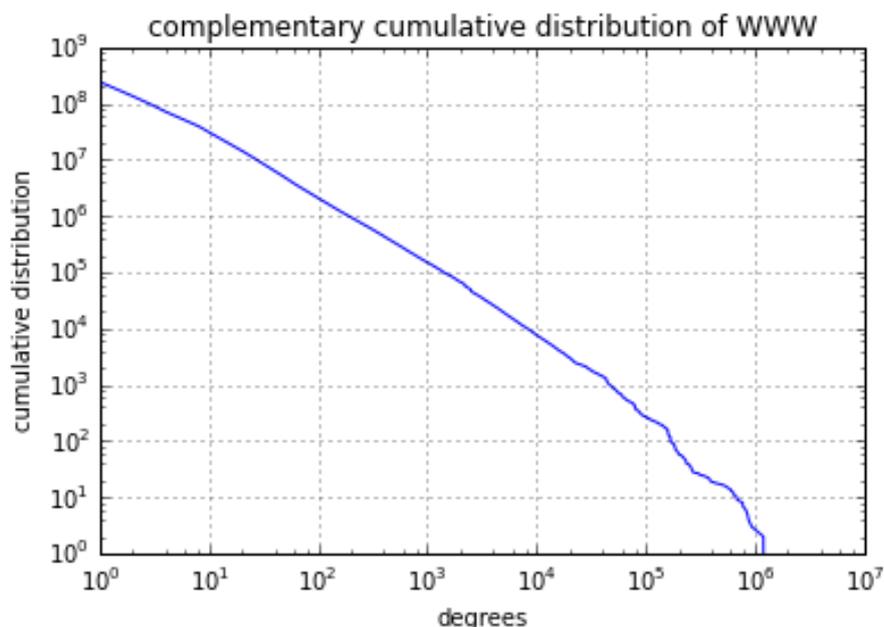


Fig.4.5 インターネットにおけるウェブページ・リンクの累積次数分布：対数グラフ

通常の次数分布のグラフに比べれば、曲線の形状が明確に見て取れる。一瞥すると、直線ではなく、曲線の形状をしていると思われる。全次数分布の範囲に対して一本の直線で回帰分析をして、その傾きを検証することは難しい。次数が小さい領域での傾きと次数が多い領域での傾きが異なる。複数の直線の組合せとして理解した方が合理的かも知れない。

純粋型のスケールフリー・ネットワークは次数が非常に大きくなる範囲ではデータによくフィットするが、次数が小さい範囲では現実のデータとの適合性は悪く、データとスケールフリーネットワークの次数分布の間には大きな差異が起きると指摘されてきた。ある水準以下の次数分布をカットオフする必要があると言われている。次数分布をある次数水準以下でカットオフしない形式で、よく使用されるスケールフリー・ネット

*11 この場合でも、別の新たな問題点が存在するが、この件については省略する。

ワークの次数分布は

$$p(k) = \frac{k^{-\tau} e^{-k/\kappa}}{Li_{\tau}(-1/\kappa)}, \quad k = 1, 2, 3, \dots$$

と与えられることが多い。ここで、 τ, κ は定数であり、関数 Li は *polylogarithm* と呼ばれている関数で

$$Li_{\tau}(z) = \sum_{k=1}^{\infty} k^{-\tau} z^k$$

と定義される。次数の平均値は

$$\langle d \rangle = \frac{Li_{\tau-1}(-1/\kappa)}{Li_{\tau}(-1/\kappa)}$$

となる。

また、*Jackson and Rogers(2007)* が提案した拡張版べき乗則分布では、累積分布関数が

$$F(k) = 1 - (rm)^{1+r} (k + rm)^{-(1+r)}$$

で与えられる。ここで、 $r > 0$ は定数で、平均次数 $\langle d \rangle = m$ となっている。 r の値はネットワークのランダム性の程度を表す定数と理解される。 r がゼロに近いとき、スケールフリー・ネットワークに近づいて行く。 r の値が無限大に近づいていくと、負の指数分布に漸近する。この次数分布のネットワークを拡張型スケールフリー・ネットワークと呼ぶことにする。このネットワークでは、

$$\langle d^2 \rangle = \frac{2rm^2}{r-1}, \quad r > 1 \text{ のとき}$$

で、 $r = 1$ のとき $E[d^2] = \infty$ となる。これは、スケールフリー・ネットワークの次数分布の分散が無限大になる性質を反映している。拡張型スケールフリー・ネットワークの次数分布を逆累積分布形式で描いたグラフは下の図である。

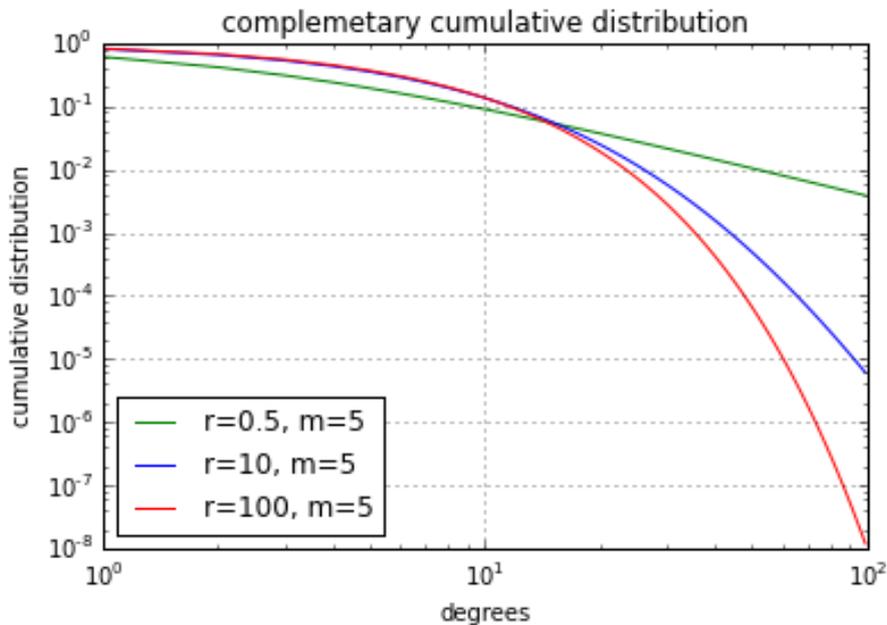


Fig.4.6 拡張型スケールフリー・ネットワークの累積次数分布：対数グラフ

明らかに、グラフは累積次数分布は直線ではなく、曲線の形状をしている。 r がゼロに近づくに連れて、直線の形状に接近する。*Jackson and Rogers* は、 $r = 0.57$ のケースが現実のネットワークの次数分布に最も良くフィットすると指摘した。このモデルは、上で見た WWW のウェブリンクの次数分布をある程度説明していると判断できる。

最後に、テールの領域で次数分布が冪乗則分布に従うケースを想定し、 k_{min} 以下の分布を切り捨てることにする。このとき、テールでの次数分布を

$$p_k = Ck^{-\alpha}, \quad \sum_{k=0}^{\infty} k^{-\alpha} = 1,$$

と与える。よって、 $k \geq k_{min}$ での次数分布は

$$p_k = \frac{\alpha - 1}{k_{min}} \left(\frac{k}{k_{min}} \right)^{-\alpha}$$

となる。

いずれにしても、現実の社会経済ネットワークの大多数をスケールフリー・ネットワークのみで説明することは難しいし、ランダムネットワークで説明することもできない。現実のネットワークは多くの現象はスケールフリー・ネットワークの特徴に合致しており、ある現象はスモールワールド・ネットワークによくフィットする。また、あるいくつかのケースでは、ランダムネットワークの性質を有する現象も見受けられる。

ネットワークの特徴を理解するためには、次数分布のモーメントを計算する必要がある。1次モーメントは

$$\langle d \rangle = \sum_{k=0}^{\infty} kp_k$$

であり、2次モーメントは

$$\langle d^2 \rangle = \sum_{k=0}^{\infty} k^2 p_k$$

と計算できる。一般的に、 m 次モーメントは

$$\langle d^m \rangle = \sum_{k=0}^{\infty} k^m p_k$$

である。インターネットのリンク連結のように、次数 $k \geq k_{min}$ の範囲において冪乗則が成立つと想定すると、 m 次モーメントは一般的に

$$\langle d^m \rangle = \sum_{k=0}^{k_{min}} k^m p_k + C \sum_{k=k_{min}}^{\infty} k^{m-\alpha}, \quad (8)$$

と計算される。次数 k が大きいときは、 k を実数として取り扱うことにすると、

$$\langle d^m \rangle = \sum_{k=0}^{k_{min}} k^m p_k + C \int_{k=k_{min}}^{\infty} k^{m-\alpha} dk = \sum_{k=0}^{k_{min}} k^m p_k + \frac{C}{m-\alpha-1} \left[k^{m-\alpha+1} \right]_{k_{min}}^{\infty}$$

と近似できる。右辺の第1項目はある次数分布（冪乗則分布でなくても良い）に基づいて計算した値で有限である。第2項の値はパラメータ m と α に大きく依存する。 $m - \alpha + 1 < 0$ であれば、この値は有限値となる。

しかし、 $m - \alpha + 1 \geq 0$ になるときは、右辺第 2 項は無限大に発散してしまう。言い換えると、 $\alpha \leq m + 1$ であれば、 m 次モーメントは発散する。ネットワークの特徴では、次数分布の分散が重要であるので、 $m = 2$ のケースがとりわけ重視される。この関係式から、 $\alpha \leq 3$ であるとき、分散は無限大になることが分かる。この結論は、次数分布として (7) を用いた計算と同じ結論をもたらす。

現実社会での通常のネットワークでは総ノード数 n は無限大でもなく、各ノードの最大次数はネットワーク上の総ノード数よりも小さい。つまり、 $k < n$ となっている。よって、次数分布の m 次モーメントは、 n が非常に大きいとき、そして、 $\alpha < m + 1$ であるならば、関係式

$$\langle d^m \rangle = \sum_{k=0}^{k_{min}} k^m p_k + \frac{C}{m - \alpha - 1} \left[k^{m-\alpha+1} \right]_{k_{min}}^n \sim n^{m-\alpha+1}$$

を満たす。2 次モーメントについては、

$$\langle d^2 \rangle \sim n^{3-\alpha}$$

となる。現実のインターネットの研究では、大まかに、 $n \approx 20000$ 、 $\langle d^2 \rangle \approx 1200$ の前後の値が使われている。これらの数値は理論上では非常に大きな値として取り扱われ、無限大に近い数値ともいえる。

5 結び

大規模なネットワーク理論の出発点は *Erdős and Rényi* によって発展してきたランダムグラフの理論である。近年急速に進んだコンピュータ技術の発展は、このランダムグラフ理論の現実有効性を検証する機会を提供してきた。大規模なネットワークからのデータが蓄積するにつれて、ランダムグラフから演繹される各種統計量の現実有効性に疑問が出されてきた。とりわけ、クラスタリング係数の値とネットワークの平均パス距離値が現実には観察されるネットワークから得られたデータから大きくかけ離れている事実が明らかになってきた。こうした理論と事実の乖離を解決するために、新しいネットワーク・モデルが提案されてきた。一つは、*Watts and Strogatz* を始めとする研究者達によるスモールワールド・モデルであり、もう一つは *Barabási and Albert* に代表される研究者達によって推進されてきたスケールフリー・モデルに他ならない。

現実の大規模ネットワークでは、スモールワールド性（大きなクラスタリング係数と短いパス距離）を持つこと、次数分布のある範囲では冪乗則分布が成立していること、これらの論点に関しては共通の理解となっていると思われる。分析する対象や分析の目的によって、どちらのモデルが有効であるかについての見解の相違は研究者間に存在している。また、感染学などでは、ランダムグラフの理論解析上での取り扱いが非常に容易であるので、ランダムネットワークのモデルが病原菌の感染過程を分析するために多方面で活用されている。

以上までの議論から理解できるように、ネットワーク理論では、暗黙裏にネットワークにおける各ノードの次数分布は確率分布として取り扱われる。言い換えると、ネットワーク理論では、ネットワークのノード数は非常に多く、無限大に近いと想定されている。従って、ネットワークの特徴を分析するとき、ネットワークに存在するノード数を無限大に大きくするというアプローチが採用される。しかし、現実のネットワークでノード数が無限大になることはあり得ない。この場合、例えば、次数分布の分散が無限大になるという結論を現実のネットワークに適用する際には、若干の留意が必要である。理論上で無限大になる性質は、現実的には、相当大きな値になることを含意すると理解すべきである。具体的にどの程度の大きさが無限大と理解されるべきかは分析対象となっているネットワークの総ノード数との相対的な対比で判断すべきだと思われる。

参考文献

- [1] Albert-László Barabási and Réka Albert(1999), Emergence of Scaling in Random Networks, *Science*, 286, 509-512.
- [2] Réka Albert, Hawoong Jeong, and Albert-László Barabási(1999), Diameter of the World-Wide Web *Nature*, 401, 130-131.
- [3] Béla Bollobás(1998), *Modern Graph Theory*, Springer.
- [4] A. Broder, R. Kumar, F. Maghoul, et al. (2000), Graph structure in the web, *Computer Networks*, 33, 309-320.
- [5] Antoni Calvó-Armengol and Matthew Jackson(2004), The Effects of Social Networks on Employment and Inequality, *American Economic Review*, 94(3), 426-454.
- [6] Reinhard Diestel(2010), *Graph Theory, Fourth Edition*, Springer.
- [7] David Easley and Jon Kleinberg(2010), *Networks, Crowds and Markets: Reasoning about a High Connected Worlds*, Cambridge University Press.
- [8] P. Erdős and A. Rényi(1959), On Random Graphs, *Publicationes Mathematicae Debrecen*, 6, 290-297.
- [9] Sanjeev Goyal(2007), *Connections: An Introduction to the Economic of Networks*, Princeton University Press.
- [10] Sanjeev Goyal, Marco J. van der Leij, and José Luis Moraga-González(2006), Economics: An Emerging Small Worlds, *Journal of Political Economy*, 114(2), 403-412.
- [11] Mark, Granovetter, (1973), The strength of weak ties, *American Journal of Sociology*, 78, 1360-1380.
- [12] Matthew O. Jackson(2008), *Social and Economic Networks*, Princeton University Press.
- [13] Matthew O. Jackson and B. W. Rogers(2007), Meeting Strangers and Friends: How Random are Social Network?, *American Economic Review*, 97(3), 890-915.
- [14] Mark E. J. Newman(2010), *Networks: An Introduction*, Oxford University Press.
- [15] Mark E. J. Newman(2005), Power laws, Pareto distributions, and Zipf's law, *Contemporary Physics*, 46, 323-351.
- [16] J. F. Padgett and C. Ansell(1993), Robust Action and the Rise of the Medici: 1400-1434, *American Journal of Sociology*, 98, 1259-1319.
- [17] K. Soramäki, M. L. Bechm J. Arnold, R. J. Glass, and W. E. beyeler(2006), The topology of interbank payment flows, *Federal Reserve Bank of New York, Staff Report no.243*.
- [18] Duncan J. Watts(1999), *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press.
- [19] Duncan J. Watts and Steven H. Strogatz(1998), Collective Dynamics of 'small world' Networks, *Nature*, 393, 440-442.
- [20] Fernando Vega-Redondo(2007), *Complex Social Networks*, Cambridge University Press.
- [21] 稲岡創、二宮拓人、谷口健、高安秀樹 (2003)、「金融機関の資金取引ネットワーク」、日本銀行金融市場局、ワーキングペーパー。