

講義ノート ネットワーク理論:母関数アプローチ

増山 幸一

明治学院大学経済学部

2015年3月: in progress version

目次

1	序	1
2	ランダムグラフ	1
2.1	ネットワークの定義	1
2.2	ランダムグラフ	4
3	確率母関数アプローチ	6
3.1	母関数の導入	6
3.2	巨大コンポーネントのサイズ分布	9
4	病原菌および新技術の拡散過程	12
4.1	病原菌の拡散過程	12
4.2	イノベーションの拡散過程	14

1 序

2 ランダムグラフ

2.1 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き、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。 $g_{ij} = 0$ 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 j からノード i へのリンク接続が存在しているときは、 $g_{ij} \neq 0$ である。ここでは、 $g_{ij} = 1$ を仮定する。リンク接続に方向性があるとき、例えば、 i から j へのリンク接続はないが、 j から i へのリンク接続があるとき、 $g_{ji} = 0, g_{ij} = 1$ である。このようなネットワークをリンク接続が方向付けられたネットワーク、有向グラフ (directed graph) という。

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） $g = (N, G)$ と標記する。 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）を表現しているため、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る。ノード集合を $N = \{A, B, C, D, E, F, G\}$ 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. 2.1 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

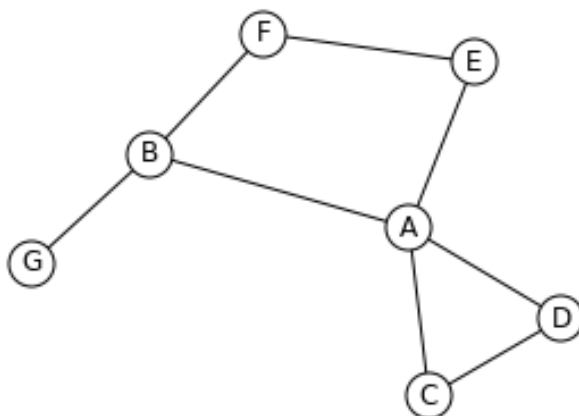


Fig.2.1 パスと連結したネットワーク

ネットワーク上で、相異なる二つのノード間をつなげるリンク接続があ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。すべてのリンクが異なっているウォークをトレイル (trail) という。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

発するノードと到達するノードが同じであるトレイルをサイクル (cycle) という。

Fig2.1 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 $N_i(g) = \{j \in N | g_{ij} = 1\}$ 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 $k_i(g) = |N_i(g)|$ はノード i の次数である。言い換えると、 $k_i(g) = |N_i(g)|$ 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、 $N_A = \{B, C, D, E\}$, $k_A = 4$ となっている。同様に、 $N_B = \{A, F, G\}$, $k_B = 3$ 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。

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。 $k_i = n - 1$ 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。

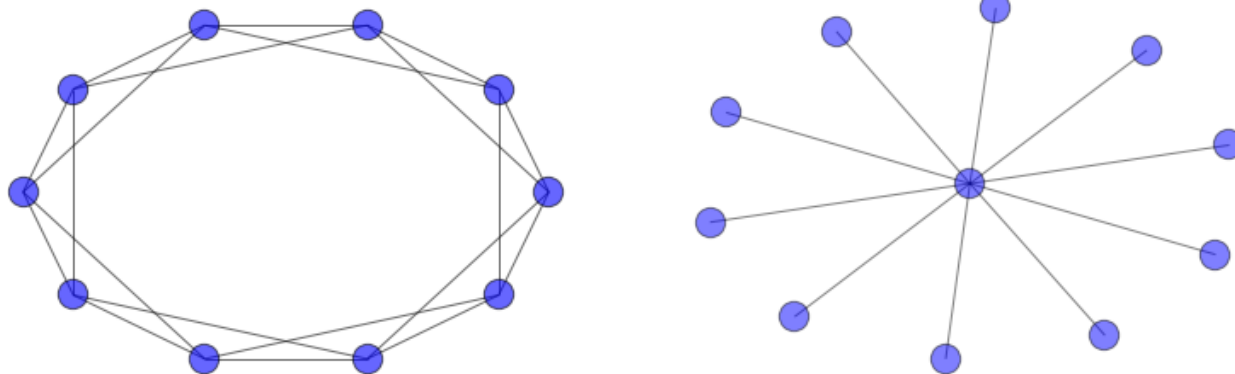


Fig.2.1 正則グラフとスター型グラフ

次数 k を持つノードの割合 (次数分布) $\text{Pr}(k)$ は

$$\text{Pr}(k) = \frac{|N_k(g)|}{n}$$

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。

2.2 ランダムグラフ

$N = \{1, 2, \dots, n\}$ 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 k 個のリンクが接続している確率は、

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

となる。任意に選ばれたノードが次数 d となる確率は

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

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 $\langle k \rangle$ の計算をすると、

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

平均次数は $\langle k \rangle = (n-1)p$ である。各ノードのクラスタリング係数が $\langle k \rangle / (n-1) = p$ 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 $\mu = (n-1)p$ としてもポアソン分布になる。

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

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・ネットワークという。

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

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

である。両辺の対数を取ると、

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

となる。 n が非常に大きければ、近似式

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

が成立する。よって、

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

となり、

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

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、

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

が成立する。右辺を S で微分すると、

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

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、

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

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。

3 確率母関数アプローチ

3.1 母関数の導入

ネットワークの次数分布が p_k , $k = 0, 1, 2, \dots$ で与えられているとき、対応する母関数を $G_0(x)$ と表記すると、

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \quad (2)$$

と定義される。 p_k は各ノードの次数が k となる確率を示す。当然、 $G_0(1) = \sum p_k = 1$ と正規化されている。また、 $|x| \leq 1$ と仮定する。

母関数は元の確率分布のすべての情報を持っている。次数が k となる確率 p_k は、母関数の k 次微係数を計算すれば復元できる。正確には、

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

ネットワーク全体での次数の平均値は

$$z = \langle k \rangle = \sum_k k p_k = G'_0(1)$$

と計算できる。より一般的に、

$$\langle k^n \rangle = \sum_k k^n p_k = \left[\left(x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}$$

が成立つ。ノード間の次数分布が互いに独立である場合、2つのノードの次数の和の確率分布の母関数は、各ノードの母関数の積 $G_0(x)^2$ となる。一般的に、 m 個のノードの次数の和の確率分布に対応する母関数は $G_0(x)^m$ である。

ネットワークの統計量で重要な概念の一つは、任意に一つのリンクを選択したとき、このリンクに連結されているノードの次数分布にかかわる統計量である。このノードの次数分布を ξ_k , $k = 0, 1, 2, \dots$ とする。あるリンクを辿るとき、次数 k のノードに連結されている確率は k に比例する。このリンクに沿って到達したノードの次数は $k p_k$ に比例する。だから、その次数分布は $\xi_k = k p_k / \sum_k k p_k$ となるの。対応する母関数は

$$\sum_{k=0}^{\infty} \xi_k x^k = \frac{\sum_k k p_k x^k}{\sum_k k p_k} = x \frac{G'_0(x)}{G'_0(1)}$$

となる。今、ランダムに一つのノード i を選び、そのノード i から出ているリンク先のノード j を考える。このノード j の次数 k_j から元のリンク $\{ij\}$ を差し引いたときの、 j の次数分布 $\hat{\xi}_k$ は

$$\hat{\xi}_k = \xi_{k+1}, \quad k = 0, 1, 2, \dots$$

となるので、隣人ノードの次数分布に対応する母関数は

$$G_1(x) = \sum_{k=0}^{\infty} \hat{\xi}_k x^k = \sum_{k=1}^{\infty} \xi_k x^{k-1} = \frac{G'_0(x)}{G'_0(1)}$$

となる。 $G_1(x)$ は隣人ノードの超過次数分布 (excess degree distribution) の母関数と理解される。最初のノード i から出ているリンク数 k は確率 p_k に従い、リンク先隣人の超過次数分布は $\hat{\xi}_k$ に従う。だから、ノード i に連結する隣人ノードの隣人総数の確率分布に対応する母関数は

$$\sum_k p_k [G_1(x)]^k = G_0(G_1(x))$$

となる。ここで、 k 個の独立な確率分布 $\hat{\xi}$ の和の母関数は各確率分布の母関数 $G_1(x)$ の積 $[G_1(x)]^k$ になる性質を用いた。 $G_0(G_1(x))$ は隣人ノードの友人数に対応する母関数となっている。これは 2 ステップ先の友人数の母関数と理解されるので、

$$G_2(x) = G_0(G_1(x))$$

と表現する。同様の手続きから、3 ステップ先の友人数の母関数は

$$G_3(x) = G_0(G_1(G_1(x)))$$

となる。一般的に、 d ステップ先の友人数の母関数は

$$G_d(x) = G_0(\underbrace{G_1(\dots G_1(x))}_{d-1})$$

となる。

これらの関係式から、隣人ノードの友人数 (2 ステップ先の隣人ノード数) の平均値 z_2 が計算できる。

$$z_2 = \left[\frac{d}{dx} G_0(G_1(x)) \right]_{x=1} = G'_0(1) G'_1(1) = G''_0(1) = \langle k^2 \rangle - \langle k \rangle$$

である。3 ステップ先の隣人ノード数の平均値 z_3 は

$$z_3 = \left[\frac{d}{dx} G_2(G_1(x)) \right]_{x=1} = [G'_2(G_1(x))]_{x=1} G'_1(1) = z_2 \times G'_1(1)$$

と計算できる。 $G'_1(1) = z_2 / G'_0(1) = z_2 / z_1$ であるので、

$$z_3 = \left(\frac{z_2}{z_1} \right)^2 z_1$$

となる。一般的に

$$z_d = \left[\frac{z_2}{z_1} \right]^{d-1} z_1, \quad d = 3, 4, \dots$$

が成立する。

従って、通常、 $z_1 > 0$ であるので、あるノードから出発したリンク接続がネットワーク全体のノードのほとんどに到達できる条件は $z_2 / z_1 > 1$ である。巨大コンポーネントが生まれる条件は

$$\langle k^2 \rangle > 2 \langle k \rangle \tag{3}$$

で与えられる。

例 3.1 (Poisson distribution)

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$$

ここで、パラメータ $\lambda > 0$ は平均値となる。

$$\langle k \rangle = \lambda, \text{Var}(k) = \lambda$$

各ノードの隣人分布の母関数 $G_0(x)$ は

$$G_0(x) = e^{\lambda(x-1)}$$

となる。隣人ノードの隣人の確率母関数 $G_1(x)$ は $G_0(x)$ に等しい。

例 3.2 (exponential distribution)

$$p_k = (1 - e^{-\lambda})e^{-\lambda k}, k = 0, 1, 2, \dots$$

ここで、 λ は正のパラメータである。各ノードの次数分布の母関数 (generating function) $G_0(x)$ は

$$G_0(x) = \frac{e^\lambda - 1}{e^\lambda - x}$$

となる。隣人ノードの隣人数の確率母関数 $G_1(x)$ は

$$G_1(x) = \left[\frac{e^\lambda - 1}{e^\lambda - x} \right]^2$$

例 3.3 (power law distribution)

$$p_k = \frac{k^{-\alpha} e^{-\lambda k}}{Li_\alpha(e^{-\lambda})}, k = 1, 2, 3, \dots$$

ここで、 α, λ は定数で、 $Li_\alpha(x)$ は

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

と定義される (polylogarithm) 関数である。各ノードの次数分布の母関数は

$$G_0(x) = \frac{Li_\alpha(xe^{-\lambda})}{Li_\alpha(e^{-\lambda})}$$

となる。 $\alpha \rightarrow 0$ のとき、つまり、単純なスケールフリー分布 $p_k = Ce^{-\alpha k}$ になるとき

$$G_0(x) = \frac{Li_\alpha(x)}{\zeta(\alpha)}$$

となっている。ここで、 $\zeta(\alpha)$ はリーマン・ゼータ (Rieman-zeta) 関数と呼ばれ、

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

隣人ノードの次数分布の母関数 $G_1(x)$ は

$$G_1(x) = \frac{Li_{\alpha-1}(xe^{-\lambda})}{xLi_{\alpha-1}(e^{-\lambda})}$$

となる。平均次数は

$$\langle k \rangle = G'_0(0) = \frac{Li_{\alpha-1}(e^{-\lambda})}{Li_{\alpha}(e^{-\lambda})}$$

と計算される。隣人ノードの平均次数 (the average number of second neighbors) は

$$z_2 = G''_0(1) = \frac{Li_{\alpha-2}(e^{-\lambda}) - Li_{\alpha-1}(e^{-\lambda})}{Li_{\alpha}(e^{-\lambda})}$$

となる。

純粹ベキ法則のケース、つまり、 $\lambda = 0$ の場合を計算する。

$$\langle k \rangle = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)} \tag{4}$$

$$\langle k^2 \rangle = \frac{\zeta(\alpha - 2)}{\zeta(\alpha)} \tag{5}$$

なので、巨大コンポーネント発生の閾値条件 (3) は

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

となる。数値計算によると、この不等式は $\alpha < 3.4788\dots$ の範囲で成立することが知られている。

3.2 巨大コンポーネントのサイズ分布

巨大コンポーネントが発生せず、有限サイズのコンポーネントが複数生まれるとき、各コンポーネントの平均サイズを計算することが必要となる。この計算は母関数を用いると容易に計算できることが知られている。ここで、ネットワーク内の任意のノード i を取り出し、このノード i はノード j にリンクしてるとしよう。隣人ノードの次数分布は ξ_k である。従って、以下のような可能性が考えられる。

1. 確率 ξ_1 で、ノード j はノード i のみと連結しており、それ以外のノードとはリンクしていない。
2. 確率 ξ_2 で、ノード j はノード i およびノード k と連結している。事前的には、リンク jk はリンク ij と同一の性質、同一の次数分布を持つ。
3. 確率 ξ_3 で、ノード j はノード i 、およびノード k 、 l と連結している。事前の情報では、リンク jk と jl はリンク ij と同一の性質、同一の次数分布を持つ。
4. 以下同じように、確率 ξ_4, ξ_5, \dots で、 \dots 。

ノード j から出発するリンクに連結するノード数の分布は、事前の情報では、ノード i から出発するリンクに接続するノード数の分布と同じである。(これを自己準拠性という。) ノード i の隣人であるノード j から出発

するリンクに連結されているノード数分布に対する母関数を $H_1(x)$ とすると、自己準拠性は

$$H(x) = x\xi_1[H(x)]^0 + x\xi_2[H_1(x)]^1 + x\xi_3[H_1(x)]^2 + \cdots = x \sum_{k=1}^{\infty} \xi_k (H_1(x))^{k-1}$$

を満たす必要がある*1。右辺の各項に x がかかっている理由は、隣人ノードに到達した最初のリンクの分を表している。この関係式を用いると、上の式は

$$H_1(x) = xG_1(H_1(x)) \quad (6)$$

と整理できる。

さて、ここで、ネットワーク上の任意のノードを取り出し、そこから出発する連結経路にリンクするノード数分布に対応する母関数 $H_0(x)$ を考えよう。上での議論を援用すると、母関数 $H_0(x)$ は

$$H_0(x) = x \sum_{k=0}^{\infty} p_k (H_1(x))^k$$

を満たす。 $H_0(x)$ は任意のノードを選んだときの、このノードがを含むコンポーネントのサイズ分布を与える母関数となっている。任意のノードの次数分布に対応する母関数 $G_0(x)$ を用いて上の式を書き換えると、

$$H_0(x) = xG_0(H_1(x)) \quad (7)$$

が得られる。以上の関係式を活用すると、コンポーネントの平均サイズが計算できる。

平均コンポーネントサイズを $\langle s \rangle$ と表記する。母関数の性質から

$$\langle s \rangle = H'_0(1)$$

なので、式 (12) を微分して、 $x = 1$ とおくと

$$\langle s \rangle = G_0(H_1(1)) + G'_0(H_1(1))H'_1(1) = 1 + G'_0(H_1(1))H'_1(1)$$

が得られる。ここで、 $H_1(1) = 1$ および $G_0(1) = 1$ を用いた。式 (11) を微分して $x = 1$ とおくと、

$$H'_1(1) = G_1(H_1(1)) + G'_1(H_1(1))H'_1(1) = 1 + G'_1(1)H'_1(1)$$

が得られるので、これを解いて

$$H'_1(1) = \frac{1}{1 - G'_1(1)}$$

となる。従って、平均コンポーネントサイズは

$$\langle s \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)} \quad (8)$$

と算できる。

$$G'_1(1) = \sum_{k=1}^{\infty} (k-1)\xi(k) = \frac{G''_0(1)}{G'_0(1)}$$

*1 母関数の自己準拠性を用いた計算方法および母関数を用いたコンポーネントサイズの計算方法は Newman, Strogatz and Watts(2001) によってランダムグラフの分析に導入された。コンパクトな解説については、Vega-Redondo(2007) の第2章を参照のこと。

および、 $G'_0(1) = \langle k \rangle = z_1$ を用いると、平均コンポーネントサイズは

$$\langle s \rangle = 1 + \frac{z_1}{1 - z_2/z_1} \quad (9)$$

と簡単に計算できることになる。ここで、 z_2 は隣人の隣人数の平均値である。 z_2/z_1 が限りなく 1 に近づいて行くと、平均コンポーネントサイズは無限大となる。 $G'_1(1) = 1$ が巨大コンポーネントが突然生まれる相転移の臨界値である。 z_2/z_1 が限りなく 1 に近づいていくとき、巨大コンポーネントが出現する。 $z_2 > z_1$ である限り、巨大コンポーネントがネットワークを埋め尽くす。

巨大コンポーネントが存在するとき、コンポーネントのサイズ分布の母関数値 $H_0(1)$ は 1 にならない。巨大コンポーネントの占める比率を $S > 0$ とすると、 $1 - S = H_0(1) < 1$ である。このとき、

$$S = 1 - H_0(1) = 1 - G_0(H_1(1))$$

が成立つ。ここで、 $u = H_1(1)$ とおくと、

$$S = 1 - G_0(u), \quad u = G_1(u)$$

が得られる。巨大コンポーネントを除いた平均コンポーネントサイズは

$$\langle s \rangle = \frac{H'_0(1)}{H_0(1)} = \frac{G_0(H_1(1)) + G'_0(H_1(1))G_1(H_1(1))}{H_0(1)}$$

と計算される。これを変形すると、

$$\langle s \rangle = 1 + \frac{zu^2}{(1-S)(1-G'_1(u))}$$

が得られる。 $S = 0(u = 1)$ のとき、(8) で与えられる上の結論と同じになる。

例 3.4 (power law distribution)

純粋スケールフリーネットワークの場合、 $\lambda \rightarrow 0$ のケースを考える。このとき、

$$G_0(x) = \frac{Li_\alpha(x)}{\zeta(\alpha)}$$

となっている。隣人ノードの次数分布の母関数 $G_1(x)$ は

$$G_1(x) = \frac{Li_{\alpha-1}(x)}{x\zeta(\alpha-1)}$$

となる。巨大コンポーネント以外のコンポーネントの平均サイズの計算式の u の値は

$$u = \frac{Li_{\alpha-1}(u)}{u\zeta(\alpha-1)} = \frac{\sum_{k=1}^{\infty} k^{-\alpha+1} u^k}{\zeta(\alpha-1)}$$

を満たす最小の正の値 u となる。この式の右辺の分子は明らかに正の数値で、分母は $\alpha \leq 2$ の範囲で発散する。よって、とき、 $\alpha \leq 2$ のとき $u = 0$ となる。つまり、巨大コンポーネントがネットワーク全体を覆う。 $\alpha > 2$ のとき、 $1 > u > 0$ である。従って、 $\alpha > 2$ の場合、ネットワークは巨大コンポーネントと小規模のコンポーネントから構成されている。

4 病原菌および新技術の拡散過程

4.1 病原菌の拡散過程

Callaway, et al.(2000)、Newman(2002)、および、Watts(2002) 達の研究において活用されてきたウイルスの相転移モデルを取上げる。ウイルスに感染して発病したノードからなるクラスターの性質を調べるために、任意に選んだノードが m 個の感染したクラスターに連結している確率を q_m として、 r_m を任意に選んだリンク先のノードが m 個の感染したクラスターに連結している確率とする*2。これらの確率分布に対応する母関数を

$$H_0(x) = \sum_m q_m x^m, \quad H_1(x) = \sum_m r_m x^m$$

とする。いま、次数 k のノードが病原菌に感染する確率を ρ_k とする。感染するメカニズムの詳細については、ここでは説明しないで感染率のみを仮定する。ランダムに選ばれたノードの次数が k で感染している確率は $p_k \rho_k$ である。この確率分布に対応する母関数は

$$F_0(x) = \sum_{k=0}^{\infty} p_k \rho_k x^k \quad (10)$$

と定義される。 $\rho = F_0(1)$ がネットワーク全体での感染確率の平均値である。任意に選んだリンクに連結するノードの次数は確率分布は ξ に従うので、隣人ノードの感染確率に対応する母関数は

$$F_1(x) = \sum_{k=1}^{\infty} \xi_k \rho_k x^{k-1} = \frac{\sum_k k p_k \rho_k x^{k-1}}{\sum_k k p_k} = \frac{F_0'(x)}{z_1}$$

である。 $F_1(1)$ は隣人ノードの感染確率の平均である。

一つのリンクをランダムに選び、そのリンクに連結しているノードの中で病原菌に感染（発病）したノード数（感染者のクラスターのノード数）の確率分布に対応する母関数は $H_1(x)$ である。このクラスターが空の場合もある。選ばれたリンクに連結したノード j が感染していない確率は

$$1 - \sum_{k=0}^{\infty} \xi_k \rho_k = 1 - F_1(1)$$

に等しい。感染しているノード j に連結しているとき、この感染ノード j の隣人の感染確率に対応する母関数は $F_1(x)$ でなければならない。 j から出るリンクに連結している感染クラスターのサイズ分布は $H_1(x)$ に従う。 $H_1(x)$ は自己準拠性を持たなければならないので、

$$H_1(x) = 1 - F_1(1) + x \left[\sum_{k=1}^{\infty} \xi_k \rho_k (H_1)^{k-1} \right] = 1 - F_1(1) + x F_1(H_1(x)) \quad (11)$$

が成立する。ランダムに選ばれたノードにリンクされている感染者クラスターのノード数（サイズ）分布に対応する母関数は $H_0(x)$ である。選ばれたノードが感染していない確率は $1 - F_0(1)$ で、次数 k で感染してい

*2 物理学のパーコレーション理論では、巨大コンポーネントのことを巨大クラスターと読んでいるので、ここでもコンポーネントの代わりにクラスターという呼び方をする。

る確率は ρ_k だから、この母関数は上と同様な理由から

$$H_0(x) = 1 - F_0(1) + x \left[\sum_{k=0}^{\infty} p_k \rho_k (H_1(x))^k \right] = 1 - F_0(1) + x F_0(H_1(x)) \quad (12)$$

となる。 $F_0(x)$ はネットワーク上で感染したノードからなるクラスターのサイズ分布を与えている。この式から感染ノードの平均クラスターサイズや相転移の閾値を計算することができる。

感染したノードのクラスターサイズの平均 s は

$$\langle s \rangle = H'_0(1) = F_0(H_1(1)) + F'_0(H_1(1))H'_1(1) = F_0(1) + F'_0(1)H'_1(1)$$

と計算できる。ここで、 $H_1(1) = 1$ と正規化した。簡単な計算で、平均クラスターサイズが

$$\langle s \rangle = H'_0(1) = F_0(1) + \frac{F'_0(1)F'_1(1)}{1 - F'_1(1)} \quad (13)$$

となることが分かる。 $F'_1(1) = 1$ が相転移を引き起こす、巨大クラスターを生み出す臨界点である。

各ノードの感染確率が次数に依存せず同一ならば、つまり、すべての次数 k において $\rho_k = \rho$ ならば、

$$F_0(x) = \rho G_0(x), F_1(x) = \rho G_1(x)$$

が成立するので、(11), (12) 式は

$$\begin{aligned} H_1(x) &= 1 - F_1(1) + \rho x F_1(H_1(x)) \\ H_0(x) &= 1 - F_0(1) + \rho x F_0(H_1(x)) \end{aligned}$$

と簡単化される。感染したクラスターの平均サイズは

$$\langle s \rangle = \rho \left[1 + \frac{\rho G'_0(1)}{1 - \rho G'_1(1)} \right] = \rho \left[1 + \frac{\rho z_1}{1 - \rho z_2/z_1} \right]$$

と計算できる。相転移を引き起こす感染確率の臨界値 ρ_c は

$$\rho_c = \frac{1}{G'_1(1)} = \frac{z_1}{z_2}$$

で与えられる。感染確率が $\rho \geq \rho_c$ を満たすとき、いくつかのノードが病原菌に感染するならば、これらのノードから始まる感染過程はネットワークのほとんどすべてのノードを覆い尽くして拡散する。直観的には、ネットワークの次数分布の分散が平均値に比較して大きければ大きいほど、巨大クラスターに感染し易い。従って、ポアソン・ランダムネットワークに比較して、スケールフリー・ネットワークにおいて、より容易に巨大クラスターに感染する。単純なスケールフリー・ネットワークの分散は無限大なので、病原菌の感染に無防備である。

巨大クラスターのサイズは以下のように計算できる。前節における導出から、巨大クラスターの占有比率を S とすると、

$$H_0(1) = 1 - S$$

が成立する。だから、

$$S = 1 - G(u)$$

となる。ただし、 $u = G_1(u)$ である。これらの関係式から巨大クラスターのサイズを計算できる。

4.2 イノベーションの拡散過程

病原菌やウイルスの感染過程の研究で発展してきた *SIR* ランダムネットワーク・モデルや *SIS* ランダムネットワーク・モデルに模倣の閾値あるいは追従の閾値概念を適用して、社会的ネットワークにおけるイノベーションの浸透過程を分析する。

新技術を採用するときや新製品を購入しようとするとき、人々は社会的なネットワークを介して友人や知合いの行動や意見に大きな影響を受ける。ネットワーク上に配置された各主体は 2 種類の選択、*A* または *B* という行為を選択できるとする。行為 *B* がデフォルトの状態とする。各主体 *i* が社会的ネットワーク上で k_i 人の隣人を持つとき、彼らの行動を観察して、行為 *A* を採用している人数を r_i とする。主体 *i* の新技術の採用から得られるネットワーク効果の大きさを $f(k_i, r_i)$ とするとき、

$$\theta_{k_i} \leq f(k_i, r_i)$$

ならば、主体 *i* は行為 *A* を模倣する。そうでなければ、行為 *B* を続けるとする。各主体に対して確率密度分布 $h(\theta)$ からランダムに選ばれたある閾値 θ_k が一様に与えられる。 $h(\theta)$ は単位区間 $[0, 1]$ の上に定義された密度関数で、 $\int_0^1 h(\theta) d\theta = 1$ と正規化されている。 θ_k を次数 k を持つノードの適用の閾値 (*threshold to adopt or imitate*) と呼ぶ。