2007年輪講用資料

氏名：奥村　浩徳

●Scale-Free Network -complex webs in nature and technology -

5. Graph generating models

5.1　Random graph model

最も単純なグラフモデルは、ErdösとRényi(1959)によって導入されたランダムグラフモデルである。頂点の数をnとし、それぞれの頂点が他の全ての頂点とつながっているとすると、辺の数はn(n-1)/2本ある。ここでもし1つの辺が確率pで生成するとすれば実際に存在する辺の数をE(n)として、

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

となる。そして、n個の頂点に対してm本の辺があるランダムグラフをG(n,m)とすれば、それが実現する確率P(G(n,m))は

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

となる。平均次数<k>は、1本の辺が2つの頂点につながっていることを考慮して、

$$\langle k \rangle = \frac{2\mathrm{m}}{n} = p(n-1) \approx pn$$

である。

ランダムグラフモデルにおいて次数がkである頂点が存在する確率 P_k は、1つの頂点と他の頂点との間にひける辺の数は最大n-1個であり、実際はk本の辺が生成しているので、

$$P_k = {}_{n-1}C_k\, p^k (1-p)^{n-1-k} = \frac{(n-1)!}{(n-1-k)!\,k!} p^k (1-p)^{n-1-k}$$

となる。この式でkを1からn-1まで変えたものを足し合わせると、

$$\sum_{k=1}^{n-1} P_k = {}_{n-1}C_1\, p^1 (1-p)^{n-2} + \ldots + {}_{n-1}C_{n-1}\, p^{n-1} (1-p)^{n-1-(n-1)} = (p+(1-p))^{n-1} = 1$$

となって正しいことが分かる。

P_k は二項分布であるが、nが限りなく大きくpが0に近いときにはポアソン分布で近似できて、

$$P_k = {}_nC_k\, p^k (1-p)^{n-k} \approx \frac{(np)^k e^{-pn}}{k!}$$

となる（nが限りなく大きいのでn-1はnとみなした）。平均次数＜k＞はpnと近似できたので P_k は、

$$P_k = \langle k \rangle^k \frac{e^{-\langle k \rangle}}{k!}$$

と書き直すことができる。

次にランダムグラフネットワークの直径Dを考える。頂点iに1本の辺で直接繋がっている頂点の数 N_i^1 は、iは<k>の辺を持つとして、

$$N_i^1 = \langle k \rangle = pn$$

である。iに繋がっているそれぞれの頂点も<k>の辺を持つとすれば、2つの辺でiに繋がっている頂点の総数 N_i^2 は、

$$N_i^2 \approx N_i^1 \langle k \rangle = \langle k \rangle^2$$

である。最初の近似は N_i^1 に1本の辺で繋がっている頂点がまた N_i^1 に含まれている可能性があるために起こる。このようにして頂点iにd本の辺で繋がっている頂点の数は大体 $\langle k \rangle^d$ である事が分かる。このdがネットワークの直径Dと等しいと仮定すると、

$$N_i^D \approx \langle k \rangle^D$$

となって、iに直接繋がっている頂点の総数nは

$$n = N_i^D + N_i^{D-1} + \ldots + N_i^1$$

となる。 N_i^D は N_i^{D-1} 以下の頂点数に比べてかなり大きいと考えられるので、

$$n \approx N_i^D \approx \langle k \rangle^D$$

のように近似できる。この式で両辺、対数をとれば

$$D \approx \frac{\ln(n)}{\ln(\langle k \rangle)}$$

が導かれる。

5.2 The small-world model

次に見るのはWattsとStrogatz(1998)によって提唱されたスモールワールドモデルである。このモデルは整列したn個の頂点を考えることからスタートする。頂点たちは、1次元の場合は円、2次元の場合は正方格子、3次元の場合は立方格子をつくるものとする。このとき、1つの頂点はz個先までの頂点と直接繋がっているとする(したがって1つの頂点に繋がっている頂点の総数は2zである)。

そして既存の1本の辺に対して、2つのランダムな頂点の間に新しくショートカットが生じる確率をpとする。するとショートカットの総数は、ネットワーク上にnz本の辺が存在していることに注意して、nzp本である。ショートカットは本数が少なくてもネットワークの平均頂点間距離や直径を劇的に小さくする効果がある。よってpが相当小さくても、平均頂点間距離が適度に小さく、クラスター係数はそれなりに大きいという現実に即したモデルができあがる。

このモデルにおいてnが限りなく大きいとすると次数分布 P_k は、少なくとも各

頂点が 2z の辺を持つとして、

$$P_k = {}_nC_{k-2z}\left(\frac{2zp}{n}\right)^{k-2z}\left(1-\frac{2zp}{n}\right)^{n-k+2z}$$

となる($k \geq 2z$ の場合。それ以外の k については $P_k = 0$)。この式で最初に n 個の頂点から k-2z 個の頂点を選び出しているのは、ある1つの頂点の辺の数 k が 2z 以上であれば k-2z 本の辺はショートカットであり、n 個の頂点のうちの k-2z 個の頂点と繋がっているからである。そしてある1つの頂点は 2z 本の整列した辺を持っているのでその頂点は 2zp 本のショートカットを持つことになる。そのショートカットの本数 2zp を頂点数 n(n が限りなく大きいので n-1 を n で近似)で割れば、他の頂点がショートカットの受け先となっている確率 $\frac{2zp}{n}$ が出てくる。2つの頂点のショートカットが互いに接続されている可能性もあるが n が限りなく大きいので、ここではその可能性は無視する。

Newman,Moore,Watts(2000)らはさらにこのモデルの解析を進めて、平均頂点間距離 l が、

$$l = \frac{n}{z'} f(npz') \qquad \left(z' = \frac{z}{2} \ , \quad f(x) = \frac{1}{2\sqrt{x^2+2x}} \tanh^{-1} \frac{x}{2\sqrt{x^2+2x}} \right)$$

となる事を示した。以下この式の導き方の概略を示す。

ランダムに選んだ頂点 i から各頂点までの距離の和をとって、その平均を求めれば平均頂点間距離が出てくる。簡単の為に、1次元の時の円のモデルを考えて、全ての値は実数であるとする。頂点 i を中心として、円に沿ったインクが両側に距離 r で拡がっていくとする。インクに含まれていない頂点の数を s(r)とし、インクに挟まれたギャップの数を t(r)とする(これはインクの部分の数に等しい)。

1つの頂点は z 個先までの頂点と直接つながっていて、インクは両側に拡がっていくので、r に対する s(r)の変化量は

$$\frac{ds}{dr} = -2zt \quad \leftarrow (1)$$

となる。このモデルでは最初 zn 本の辺が存在しているので、発生するショートカットの数は znp 本となる。よってランダムな1つの頂点がショートカットの端点となっている確率は $2znp/n = 2np$ である。r が r+1 に増加すると、1つのインクの部分は 2z で拡がっていく。インクの部分の数は t 個あるので、インクとショートカットが出会う確率は、

$$2zp \times (2z \times t) = 4z^2 pn$$

である。また、ショートカットの端がギャップの部分につながる確率は $\frac{s}{n}$ であるので、

新たにできるインクの部分の個数は $\dfrac{4z^2 pst}{n}$ となる。

インクが拡がるにつれてギャップが1つ消える（従ってインクの部分の数も1つ減る）という事象も起こる。r が r+ir に少しだけ変化するときにギャップの幅が 2zir より小さければ、ギャップは両側からインクに侵食されて消滅する。ギャップの幅が 2zir より小さい確率を求めるに、まずギャップの幅が x である確率 p(x)を求める。

今、インクに含まれていない頂点の数は s 個あるのでその部分に t-1 個の仕切り（これがインクの1個の部分に相当する）をランダムに入れるという動作を行う（最初の 1 つの仕切りは固定しておく）。ただし簡単の為に仕切りの幅は考えない。幅 x のギャップを確保するためには、まず t-1 個の仕切りの中から 1 個選び出し、それを最初の仕切りからちょうど x だけ離れた場所に設置する。そして残りの t-2 個の仕切りは、最初の仕切りから x より多く離れた場所にランダムに入れる。これを式で表すと次のようになる。

$$p(x) = {}_{t-1}C_1 \frac{1}{s}\left(\frac{s-x}{s}\right)^{t-2} = \frac{t-1}{s}\left(1-\frac{x}{s}\right)^{t-2}$$

これを 0 から 2zir まで x で積分すればギャップの幅が 2zir 以下である確率 $p_{2\mathrm{zir}}(x)$ が次のように出てくる。

$$p_{2\mathrm{zir}}(x) = \int_0^{2\mathrm{zir}} p(x)\,dx = \left[-\left(1-\frac{x}{s}\right)^{t-1}\right]_0^{2\mathrm{zir}} = 1-\left(1-\frac{2\mathrm{zir}}{s}\right)^{t-1}$$

r が r+1 に変化する時は、この式で ir=1 となって、z の値が s に比べてかなり小さければ $\dfrac{2\mathrm{z}}{s} \ll 0$ で2項展開の近似式が使えるので、

$$1-\left(1-\frac{2\mathrm{zir}}{s}\right)^{t-1} = 1-\left(1-\frac{2\mathrm{z}}{s}\right)^{t-1} \simeq 1-\left(1-\frac{2\mathrm{z}(t-1)}{s}\right) = \frac{2\mathrm{z}(t-1)}{s}$$

となる。ギャップの数は t 個あるので、r が r+1 に変化するとき消滅するギャップの数は s は $\dfrac{2\mathrm{zt}(t-1)}{s}$ 個である。以上より r に対する t の変化量は、

$$\frac{dt}{dr} = \frac{4\mathrm{z}^2 pst}{n} - \frac{2\mathrm{zt}(t-1)}{s} \quad \leftarrow (2)$$

となる。（1）と（2）が解析の本質的な役割を果たし、テクニカルないくつかの計算の後に r が求まる。その r を用いて平均頂点間距離 l を求めることができる。

スモールワールドモデルの クラスター係数 C は、ショートカットが既存の辺を繋ぎ変えてできるのか、それとも単純にネットワークに追加されるかによって違ってくる。Barrat と Weigt(2000)によるとショートカットが既存の辺を確率 p で繋ぎ変えてできる

場合のクラスター係数Cは、各頂点が2z本の整列した辺を持つとして、

$$C=\frac{3(z-1)}{2(2z-1)}(1-p)^3$$

となる。ここで $\frac{3(z-1)}{2(2z-1)}=\frac{3z(z-1)}{2}/\frac{2z(2z-1)}{2}$ はpが0の時のクラスター係数で、$\frac{3z(z-1)}{2}=\frac{z(z-1)}{2}+\frac{z(z-1)}{2}+\{(z-1)+(z-2)+...+2+1\}$ は1つの頂点に直接繋がっている頂点2zの間に存在している辺の数、$\frac{2z(2z-1)}{2}$ は2zの間に存在することのできる辺の数である。そして $(1-p)^3$ は三角形がそのまま維持される確率である。

一方、Newman(2002)によるとクラスター係数の定義は、

C=3×(存在する3角形の個数)/(少なくとも1つの頂点が残りの2つの頂点と繋がっているような、3つの頂点の組の個数)

であり、ショートカットが新しく追加されるときのクラスター係数は、

$$C=\frac{3z(z-1)}{2z(2z-1)+4z^2p(p+2)}$$

となる。

5.3 The Barabási-Albert model

Barabási と Albert(1999)によって提唱されたモデル(以下BAモデルとする)は現実世界の成長するネットワーク(インターネットやWWWなど)を説明するのに最も適している。このモデルの特徴は時間毎に一定数の辺を持った頂点が追加されていくことと、追加される頂点が高い確率で次数の多い頂点に接続する(優先的選択)ことである。BAモデルにおいてグラフが成長していく具体的なアルゴリズムは次の通りである。

1. 接続されていない状態の n_0 個の頂点を用意する。

2. 時間毎に新たな頂点達を追加する。新たな頂点は m_0 本の辺を持つ。

3. 新しい頂点は確率 $\Pi(k_i)$ で既存の頂点iに接続されるものとする。その確率は、

$$\Pi(k_i)=\frac{k_i}{\sum_{j=1}^{n} k_j}$$

である。ただし最初に頂点が追加された時のみ、接続先の頂点はランダムに選ばれるものとする。

この3つのステップから次数分布がべき乗則に従うということを導く。簡単の為に各時間ごとに1つの頂点がネットワークに追加されるとする。このとき時刻 t におけるネットワークの頂点数 n と辺の数 m はそれぞれ、

$$n=n_0+t$$

$$m=\frac{1}{2}\sum_{j=1}^{n} k_j=m_o t$$

で表される。2つ目の式の $\frac{1}{2}$ は1つの辺を2つの頂点が共有しているという意味である。k を連続関数だとみなした時、頂点 i の辺の数 k_i の単位時間あたりの変化量は、

$$\frac{\partial k_i}{\partial t}=m_0\Pi(k_i)=m_0\frac{k_i}{\sum_{j=1}^{n} k_j}=m_0\frac{k_i}{2\mathrm{m}_0 t}=\frac{k_i}{2\mathrm{t}}$$

となる。始まりの時間を t_i として初期条件 $k_i(t_i)=m_0$ の元でこの式を解くと、

$$k_i(t)=m_0\left(\frac{t}{t_i}\right)^{\frac{1}{2}}$$

となる。この式は頂点 i の辺の数は $\sqrt{t}$ に従って増えていくという事を示している。次に頂点 i の辺の数 k_i が k より小さい確率は上の式を利用して、

$$P(k_i<k)=P(t_i>\frac{m_o^2 t}{k^2})$$

である。頂点は 1 ステップで一つずつネットワークに追加されていくので、時刻 t における総頂点数 n_0+t である。これらの頂点の中で $t_i>\frac{m_0^2 t}{k^2}$ をみたすものは時刻 $\frac{m_0^2 t}{k^2}+1, \frac{m_0^2 t}{k^2}+2, \ldots, t$ に加えられた頂点達で合計 $t-\frac{m_0^2 t}{k^2}$ 個ある。よって時刻 t である頂点が条件 $t_i>\frac{m_0^2 t}{k^2}$ を満たしている確率は、

$$P(k_i<k)=P(t_i>\frac{m_0^2 t}{k^2})=\frac{t}{n_0+t}-\frac{m_0^2 t}{n_0+t}\frac{1}{k^2}$$

となり、これを k で偏微分すれば

$$P(k)=\frac{\partial P(k_i<k)}{\partial k}=\frac{2\mathrm{m}_0^2 t}{n_0+t}\frac{1}{k^3}$$

が導かれて、 $P(k)\sim k^{-3}$ となって確かに次数分布はべき乗則に従っていることが分かる。

5.4 Modifications to the Barabáshi-Albert model

5.4.1 Fitness and preferential attachment model

BAモデルの拡張版としてBianconiとBarabáshi(2001)によって提唱された適合度モデルがある。これは個々の頂点が適合度を持ち、新しい頂点と接続しやすく（または接続しにくく）なっているモデルである。具体的には次のようになる。

1．BAモデルをベースとして、最初に n_0 個の異なった頂点を用意する。その際に頂点iは η_i という適合度を持つものとする。この η_i は確率分布 $\rho(\eta)$ によって決まる。

2．適合度 η_i はネットワークが成長する毎に変化する。

3．頂点iの優先的選択の確率 $\Pi(k_i,\eta_i)$ は、

$$\Pi(k_i,\eta_i)=\frac{\eta_i k_i}{\sum_{j=1}^{n}\eta_j k_j}$$

となる。

これによって辺の数が多い頂点がどんどん成長するのを抑えることができ、また新しい頂点も多くの辺を獲得できるという状況ができる。このモデルは個々の頂点の差別化をはかることによって現実世界のネットワークをうまく説明したモデルであるといえる。

5.4.2 Edges growth model

ネットワーク上で成長するのは頂点だけでなく、辺もまた成長すると考えられる。Krapivski, Rodgers, Redner(2001)達はネットワークに頂点を追加するだけでなく、既存の頂点の間に辺を追加するモデルを提唱した。このモデルでは辺に向きがあるとして、頂点aに入ってくる辺の数を k_a^{in} 、出て行く辺の数を k_a^{out} とする。最初の頂点数を0とし、時刻tの時の頂点の総数をN(t)、各頂点に入る辺の総数をI(t)、各

頂点から出て行く辺の総数をJ(t)とする。このときネットワークでは、

1．確率pで1本の辺を持つ新しい頂点が追加される。優先的選択の確率はターゲットの頂点をaとして $(k_a^{in}+\lambda)$ を用いて決まる。ただし $\lambda>0$ である。

2．確率 $q=(1-p)$ で新しい辺が既存のネットワーク上に発生する。このとき新しい辺が頂点bからaに向かうものである確率は $(k_a^{in}+\lambda)(k_b^{out}+\mu)$ を用いて決まる。ただし、 $\mu>0$ である。

1．か2．のどちらかの事象が起こる。以下の考察はこの条件の下で行う。

最初の頂点数は0なので、 $N(t)=pt$ である。また事象1．2．のどちらが起こっても辺の数は1本追加されるので $I(t)=J(t)=t$ である。時刻tのときの、入ってくる辺の数（in-degree）がi、出て行く辺の数（out-degree）がjである頂点の総数を $N_{ij}(t)$ とする。この時 $N_{ij}(t)$ のtに対する変化量は

$$\frac{dN_{ij}}{dt}=(p+q)\left[\frac{(i-1+\lambda)N_{i-1,j}-(i+\lambda)N_{ij}}{I+\lambda N}\right] +q\left[\frac{(j-1+\mu)N_{i,j-1}-(j+\mu)N_{ij}}{J+\mu N}\right]+p\,\delta_{i0}\delta_{j1} \quad \leftarrow(3)$$

となる。(3)の右辺の最初の項は入ってくる辺の変化量に注目した時の、 $N_{ij}(t)$ の変化量である。時刻がtからt+1に変わった時にネットワークには必ず1つの辺が追加されるが、この時in-degreeがi-1である頂点は、

$$(p+q)\frac{(i-1+\lambda)N_{i-1,j}}{\sum_{i,j}(i+\lambda)N_{ij}}=(p+q)\frac{(i-1+\lambda)N_{i-1,j}}{I+\lambda N} \quad \leftarrow(4)$$

の割合でin-degreeがiである頂点に変化するので、これが N_{ij} の増加分である。同様にin-degreeがiである頂点は $(p+q)\frac{(i+\lambda)N_{ij}}{I+\lambda}$ ←(5)の割合で、in-degreeがi+1の頂点に変化するのでこれが減少分となり、(4)と(5)を組み合わせて最初の項が出てくる。事象2．が起こった場合はさらにout-degreeの変化に注目して、上と同様の考察をすれば(3)の右辺の第2項が出てくる。最後の項は1．が起こったときにネットワークにin-degreeが0、out-degreeが1の頂点が発生する確率である（ただし δ_{ij} は $i=j$ のとき1でそれ以外の時は0となる関数である）。

$N_{ij}(t)$ はtに対して線形的に変化するのでin-degreeがi、out-degreeがjの頂点の発生確率を n_{ij} とすれば $N_{ij}(t)=t\,n_{ij}$ となる。これを(3)に代入して、さらに

$$a = q\frac{1+p\lambda}{1+p\mu} \;,\; b = 1+(1+p)\lambda$$

という置き換えをすれば、

$$n_{ij} = \frac{(i-1+\lambda)n_{i-1,j} + a(j-1+\mu)n_{i,j-1} + p(1+p\lambda)\delta_{i0}\delta_{j1}}{i+a(j+\mu)+b} \quad \leftarrow (6)$$

が得られる。時刻 t での in-degree が i である頂点の数を $A_i(t)$ 、out-degree が j である頂点の数を $B_j(t)$ とすれば、 $A_i(t) = \Sigma_j N_{ij}(t)$、 $B_j(t) = \Sigma_i N_{ij}(t)$ となる。さらに in-degree が i の頂点の分布を I_i 、out-degree が j の頂点の分布を O_j とすれば $A_i(t), B_j(t)$ ともに t に関して線形的に変化していくので、 $A_i(t) = t\,I_i$, $B_{\mathrm{j}}(t) = t\,O_{\mathrm{j}}$ という関係式が得られて、

$$I_i = \sum_j n_{ij}, O_j = \sum_i n_{ij} \quad \leftarrow (7)$$

となる。(6)を(7)に代入すれば、

$$(i+b)I_i = (i-1+\lambda)I_{i-1} + p(1+p\lambda)\delta_{i0} \quad \leftarrow (8)$$

$$\left(j+\frac{1}{q}+\frac{\mu}{q}\right)O_j = (j-1+\mu)O_{j-1} + \frac{p(1+p\mu)}{q}\delta_{j1} \quad \leftarrow (9)$$

が出てくる。ここでガンマ関数

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$$

を用いると、自然数 n に対して $\Gamma(n) = (n-1)!$ となるので適当な自然数 λ を選んで(8)は

$$I_i = \frac{(i-1+\lambda)\times\ldots\times\lambda}{(i+b)\times\ldots\times(1+b)}(I_0 + p(1+p\lambda)) = \frac{\Gamma(i+\lambda)\Gamma(b+1)}{\Gamma(i+b+1)\Gamma(\lambda)}(I_0 + p(1+p\lambda))$$

$$\leftarrow (10) \qquad \left(I_0 = p\frac{1+p\lambda}{b}\right)$$

と書き直すことができる。同様にして(9)も、適当な自然数 μ を選んで

$$O_j = \frac{\Gamma(j+\mu)\Gamma(2+q^{-1}+\mu q^{-1})}{\Gamma(j+1+q^{-1}+\mu q^{-1})\Gamma(1+\mu)}\left(O_1 + \frac{p(1+p\mu)}{q}\right) \qquad \leftarrow (11)$$

$$\left(O_1 = p\frac{1+p\mu}{1+q+\mu}\right)$$

と書き直すことができる。(10)や(11)で 0 よりかなり大きな i について、ガンマ関数の漸近的振る舞いを考えれば

$$I_i \sim i^{-(2+p\lambda)}, O_j \sim j^{-(1+p^{-1}+\mu pq^{-1})}$$

という次数分布のべき乗則が得られる。

5.4.3 Aging process model

BA モデルの最後の拡張版は頂点が歳をとっていくというものである。俳優も辞めたり死んだりして映画には出演しなくなるし、あまり引用されないような論文はやがて忘れられる。Klemn と Eguíluz(2002)はそのような事象をうまく反映させたモデルをつくった。そのアルゴリズムは次のようである。

1. 最初に n 個の頂点が辺を持たずにあるとして、その中の m 個の頂点がアクティブな状態となっている。そこに1つの頂点をアクティブな状態で追加する。

2. 追加された1つの頂点はアクティブな状態の m 個の頂点全てと接続される。

3. アクティブな状態の m+1 個の頂点から1つの頂点を選んで、ノンアクティブな状態にする。アクティブな状態の頂点 i がノンアクティブになる確率 P_i^{deact} は

$$P_i^{deact}=\frac{1}{N}\frac{1}{(k_i+a)}=\frac{1}{\Sigma_{j\in A}(k_j+a)^{-1}}\frac{1}{(k_i+a)}$$

である。ここで $\frac{1}{N}=\frac{1}{\Sigma_{j\in A}(k_j+a)^{-1}}$ （A はアクティブな頂点の集合）であり、a は 0 より大きい定数である。

このモデルにおいても次数分布はべき乗則に従っている。以下にそれを示す。
まず次数 k の頂点の分布 $p(k)$ は連続であるとみなす。時刻 t+1 において辺の数が k+1 である頂点の分布 $p_{t+1}(k+1)$ は、時刻 t のときに辺の数が k である頂点の分布 $p_t(k)$ に頂点がアクティブのまま維持される確率をかけたものである（なぜなら頂点がアクティブであれば次の時刻には次数が1増えるから）。よって、

$$p_{t+1}(k+1)=[1-P_i^{deact}]p_t(k)=\left[1-\frac{1}{N}\frac{1}{(k+a)}\right]p_t(k)$$

となる。ここで時間発展がないものとすれば $p_{t+1}(k)=p_t(k)=p(k)$ となり、次数 k についての $p(k)$ の変化量は

$$\frac{dp(k)}{dk}=p(k+1)-p(k)=-\left[\frac{1}{N}\frac{1}{(k+a)}\right]p(k)$$

となる。この微分方程式を解くと $p(k)=C(k+a)^{-1/N}$ (C は任意定数)となって $p(k)\sim(k+a)^{-1/N}$ が示される。

5.5 Copying model

遺伝子のネットワークでは DNA が自己複写するので、頂点が自己複写すると

いうネットワークモデルも考えられる。そのようなモデルは Aiello,Chung,Lu(2000)らによって提唱された。ここでは最初に n_0 個の頂点を用意して毎ステップごとにランダムに1つの頂点が複製されるものとする。グラフは有効グラフであるとする。その時、

1．確率 α で、新しく複製された頂点の m_0 本の出て行く辺（out-degree）が繋ぎ変えられる。

2．確率 $1-\alpha$ で、新しく複製された頂点の out-degree はそのまま保たれる。

のどちらかの事象が起こる。よって頂点 i の時間 t に対する入ってくる辺（in-degree）の変化量は、総頂点数 n が十分多い状態では

$$\frac{\partial k_i^{in}(t)}{\partial t}=(1-\alpha)\frac{k_i^{in}(t)}{n}+m_0\frac{\alpha}{n}$$

となる。ここで右辺の第1項は、out-degree が頂点 i に繋がっている頂点が n 個の頂点の中から選ばれて複写され、その out-degree が保たれる状態を表す。右辺の第2項はランダムに選ばれて複写された頂点の m_0 本の out-degree が繋ぎ変えられて、それが頂点 i に接続される状態を表す。ネットワークに追加された時刻を t_i として初期条件 $k_i^{in}(t_i)=0$ の下で上の式を解くと、

$$k_i^{in}(t)=\frac{m_0\alpha}{1-\alpha}\left[\left(\frac{t}{t_i}\right)^{1-\alpha}-1\right]$$

が得られる。後は 5．3 の後半部分と同じように $P(k^{in})$ を求めれば、

$$P(k^{in})\sim\left[k^{in}+m\frac{\alpha}{1-\alpha}\right]^{-\frac{2-\alpha}{1-\alpha}}$$

となって、次数分布のべき乗則が出てくる。

5.6 Fitness based model

現実の世界では頂点どうしの組み合わせによって、そこにリンクが生まれやすかったり生まれにくかったりするようなネットワークがある。Servedio,Caldarelli,Buttá (2004)達はネットワークの頂点どうしの関係に注目したモデルを提唱した。その特性は次のようなものである。

・最初 n 個の頂点が存在するものとする。ある頂点 i にはその頂点を特徴づける実数 x_i が確率分布 $\rho(x)$ に応じて与えられているものとする。この実数 x_i は現実世界でのその頂点の重要度や地位に対応している。

・頂点 i と j の間には相性度 $f(x_i, x_j)$ に応じてリンクが張られるものとする。グラフに向きが無ければ $f(x_i, x_j) = f(x_j, x_i)$ となる。

このとき実数 x を与えられた頂点の次数は、

$$k(x) = \int_0^\infty f(x, y) n\rho(y) dy$$

となる。ここで $n\rho(y)$ は実数 y を与えられた頂点の数である。

もし k(x)の値が x についての単調増加（減少）関数ならば、k の値を決めれば x の値もまた決まる。そうすると、k の次数分布 P(k)は

$$P(k) = \rho(x(k)) \frac{dx(k)}{dk}$$

となる。 $\rho(x(k))$ は k を決めた時の x の確率分布を表していて、 $\frac{dx(k)}{dk}$ は k の変化量に対する x(k)の変化量を表している。

例として x_M を頂点の最高値として $f(x, y) = \frac{xy}{x_M^2}$ と定義すれば

$$k(x) = \frac{nx}{x_M^2} \int_0^\infty y\rho(y) dy = n\frac{\langle x \rangle x}{x_M^2}$$ (<x>は頂点の平均の値)

となって、 $x = \frac{x_M^2}{\langle x \rangle x} k$ が出てきて

$$P(k) = \frac{x_M^2}{n\langle x \rangle} \rho\left(\frac{x_M^2}{n\langle x \rangle} k\right)$$

が求まる。ここで $\frac{x_M^2}{n\langle x \rangle}$ は一定の値なので、P(k)というのは結局 k によって決められるという事が分かるが、ここでは k は x によって一意に決められるので、P(k)というのは x の分布 $\rho(x)$ を元にして決まる事が分かる。よって $\rho(x)$ がべき乗則に従っていれば、P(k)もべき乗則に従う。

5.7 Graph from optimization principles

これまでに挙げたものとは違うグラフの作り方として、頂点や辺にある値を与えたときに、任意の測度（例えば平均頂点間距離など）の値が最も小さくなるように頂点や辺に変化を加えていくという方法がある。

代表的なものとしてメトロポリス法があり、これを使えば十分多い回数のシミュレーションを行ってグラフを最適化状態にすることができる。具体的にはある測度 E を用意して、グラフ A からグラフ B に状態を仮に変化させたときに、 $E(B) \leq E(A)$ ならばそのまま変化を認め、 $E(B) > E(A)$ ならば確率 $e^{-(E(B)-E(A))}$ で変化を認める。こ

の試行を十分多い回数行えば測度の最小値 E_{min} が得られるようなグラフに行き着く。

5.7.1　Optimized graphs: cost function and transport

グラフの最適化を利用した例として次のようなものがある。生き物たちは個々に、一定の時間ごとに、生存していく為のエネルギー量B(Metabolic late)を必要とする。立体的な生き物が必要とするBの値は、その生物の質量(総エネルギー数)をMとして

$$B \propto M^{3/4} \quad \leftarrow (12)$$

という関係式(Kleiber relation)を持つ。今立体的な生物を、細胞間の距離を1だとして、一辺の長さがLの正立方体 L^3 に見立てる。その中に各細胞が格子状に存在しているとすれば、細胞の数は L^3 個ある(正確には $(L+1)^3$ 個だが、 $L \gg 1$ なので L^3 と近似する)。ここで各細胞が一定時間ごとに必要とするエネルギーの平均をaとすると

$$B = aL^3$$

となる。一方質量Mは、各細胞間に張り巡らされている血管(これは格子の辺と見立てられる)の内に存在するエネルギーの総和であるとする。エネルギーは L^3 の中の任意の1点から供給されるものとして、一定時間に血液のよって距離1だけ運ばれるものとする。もし出発点からある頂点までの距離が d ならば、血管内には ad のエネルギーが存在していなくてはならない。出発点から他のすべての頂点までの距離の平均を $\langle d \rangle$ とすれば、

$$M = a\langle d \rangle L^3$$

となる。ここで L^3 の細胞に対して任意の頂点を出発点に選んで、メトロポリス法を使って $\langle d \rangle$ が最小となるようなグラフを求める。ただし血液は一方方向に流れてしかも交わることがないので、グラフは有効グラフとし一つの頂点には2つ以上の辺が入ってこないものとする。そうするとグラフは $\langle d \rangle = L$ となるものに行き着く(これに対する厳密な証明は存在するが、かなり難解なのでメトロポリス法を使った方が便利である)。そして驚くべき事に生物達はそのような最適化されたグラフを選んでいるのである。よって、

$$M = aL^4$$

となって、 $B = M^{3/4}$ ←(13)という関係式が出てくる。

以上の理論は、あくまで生物を完全な正立方体と考えたときのものである。実際には生物は完全な正立方体ではないので、個々の正立方体の部分に分割してそれらの総和として考える必要がある。よって完全に(13)の式が満たされるわけではないが、沢山の生物達の間に共通して(12)の関係が見られることの十分な説明にはなる。ただ、あまりに平面に近い生物だと $B \propto M^{2/3}$ という関係になる。この式は上の理論で正立方体を正方形と考えれば同様に導かれる。

5.7.2 The constant flux

自然界には川のように、大きな流れから段々といくつかの小さな流れに変化していくというモデルもよく見られる。大きな川は最終的には沢山の小さな川となって海に流れ込むのか湖を形成するのであるが、そのような小さな川の水量は実はあまり大差ない。よってそれらの小さな川の水量を一定値 a とすることができ、またその小さな川の個数を n とすれば、一番最初の川の水量は na となる。ここでもしある川の水量が na だった場合には、その場所の地理性質にもよるが、自然界では大体似たような川の水量の確率分布 P(na)が観察される（最小の川の個数もまた n 個になる）。これは川だけでなくその他の多くの流れについて共通して見られることである。次の章からはその具体例を見ていく。

●References

・Collective dynmics of “small-world” networks
(Watts D.J. and Strogatz S.H.;Nature,393.(1998))

・Extracting large-scale knowledge bases from the web
(Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan and Andrew Tomkins; Proceedings of the 25th International Conference on Very Large Data Bases.(1999))

・Size and forms in efficient transportation networks
(Jayanth R.Banavar , Amos Maritan and Andrea Rinaldo; Nature,399.(1999))

・The fourth dimention of life: Fractal geometry and allometric scaling of organisms
(Geoffrey B.W. , James H.B. and Brian J.E.; Science,284.(1999))

・A random graph model for massive graphs
(William Aiello, Fan Chung and Linyuan Lu; Proceedings of the 32nd ACM Symposium on Theory of Computing.(2000))

・Mean-field solution of the small-world network model
(Newman M.E.J. , Moore C. and Watts D.J.; Physical Review Letters,84.(2000))

・On the properties of small-world networks
(Barrat A., Weigt M.;The European Physical Journal B,13.(2000))

・Competition and multiscaling in evolving networks
(Ginestra Bianconi, Albert-László Barabási; Europhysics Letters,54.(2001))

・Degree distributions of growing networks
(Krapivski P.L., Rodgers G.L. and Render S.;Physical Review Letters,86.(2001))

・Highly clustered scale-free networks
(Konstantin Klemm and Victor M. Eguíluz; Rhysical Review E,65.(2002))

・The structure and function of networks
(Newman N.E.J.; Computer Physics Communications,147.(2002))

・Vertex intrinsic fitness: How to produce arbitrary scale-free net works
(Vito D.P. Servedio, Guido Caldarelli and Paolo Buttà; Rhysical Review E,70.(2004))

・「複雑ネットワークの科学」(増田直紀, 今野紀雄;産業図書(2005))