解 説

複雑ネットワーク上の確率モデル

Stochastic Models on Complex Networks

竹 居 正 登^{*1} 今 野 紀 雄^{*2} ^{*1}大阪電気通信大学工学部 ^{*2}横浜国立大学大学院工学研究院

Masato Takei^{*1} and Norio Konno^{*2}

^{*1}Faculty of Engineering, Osaka Electro-Communication University ^{*2}Faculty of Engineering, Yokohama National University

1. は じ め に

本稿では、複雑ネットワークの数学モデルとして種々のランダムグラフを概観し、その上の確率モデルとしてランダムウォークと投票者モデルを考える。後で見るように、二つの確率モデルは密接に関係している。また伝染病のモデルであるコンタクトプロセスに関しては今野 [1] を参照のこと。文中で年号のみ示した結果については、Durrett によるテキスト [2] 等、参考文献中の文献表を見ていただきたい。

2. ランダムグラフ

2.1 グラフ

頂点と辺からなる図形をグラフと呼ぶ。グラフ $G = (V, E)$ は頂点の集合 V と辺の集合 E の組である。辺で結ばれた異なる 2 頂点は隣接しているという。二つの頂点をいくつかの辺をたどって結べるかどうかに着目すると、グラフを連結成分に分けることができる。

頂点 x から出ている辺の総数 $d(x)$ を x の次数という。グラフのすべての頂点の次数が d に等しいとき d -正則であるという。

2 点間の距離は、一方から他方に行くときに通る辺の最小本数と定義する。 n 個の頂点からなる連結グラフにおいて、 $\binom{n}{2}$ 組の頂点の組み合わせ全体にわたる 2 点間の距離の平均を平均頂点間距離と呼ぶ。グラフ中に三角形などの特定のつながり方がどの程度見いだされるかの尺度として、クラスター係数が定義される (文献 [3] を参照)。

2.2 古典的ランダムグラフ

n 個の頂点 $\{1, 2, \dots, n\}$ があり、異なる頂点を結ぶ辺が存在する確率が各組ごとに独立に p というランダムグラフを考える。このモデルを Erdős-Rényi ランダムグラフと呼

び、 $ER(n, p)$ と表す。任意の頂点について、その次数の分布は二項分布 $B(n-1, p)$ に従う。特に、平均次数は $(n-1)p$ である。そこで、平均次数 λ を一定にして総頂点数 n を大きくする極限をとるときに、ランダムグラフの従う確率法則がどのようなものになるか、またランダムグラフに高い確率で観測される特徴はどのようなものであるかを考察する： $ER(n, \lambda/n)$ において、 $n \rightarrow \infty$ のとき

- 次数の分布は、平均 λ の Poisson 分布に近づく。
- $\lambda < 1$ ならば、高い確率で、どの連結成分の大きさも $O(\log n)$ 。
- $\lambda > 1$ ならば、高い確率で、最大連結成分の大きさは $O(n)$ であり、2 番目以降の連結成分の大きさは $O(\log n)$ 。最大連結成分における 2 点間の平均距離は、高い確率で $O(\log n)$ 。

(ここで $f(n)$ が $O(g(n))$ であるとは、十分大きい n で $C \leq |f(n)/g(n)| \leq C'$ を満たす正の定数 C, C' が取れることを意味する)

2 頂点が辺でつながれる確率 p の頂点数 n に対する増大度を高めると、 n が大きいときに孤立した点ができる確率は低くなる： $c > 1$ のとき、 $ER(n, (c \log n)/n)$ において、孤立点がなく全体が連結である確率は $n \rightarrow \infty$ のとき 1 に収束する。

現実のネットワークについての調査結果と比較すると、 n が大きいとき次数分布が Poisson 分布に近い分分布の裾が薄く、三角形の密度 (クラスター係数) も低い点が異なっている。

Erdős-Rényi のランダムグラフでは次数分布が Poisson 分布に漸近するが、逆に、次数分布を先に与えてランダムグラフを構成することがある。古くから調べられている例に、ランダム d -正則グラフがある： n 個の頂点を持ち、各点の次数がすべて d であるグラフを一樣な確率で選ぶモデルを $\mathcal{G}_d(n)$ で表す。一般の次数分布に関しては Molloy-Reed (1995, 1998) がある。

2.3 優先的選択モデル

Barabási-Albert (1999) は、次数の多い頂点に新たな頂点がつきやすい (優先的選択) という機構で生成されるランダムグラフのモデルを提案し、次数が k となる確率が

原稿受付 2007 年 9 月 27 日

キーワード: Complex Networks, Stochastic Models, Random Graphs, Random Walks, Voter Models

^{*1}〒572-8530 寝屋川市初町 18-8

^{*2}〒240-8501 横浜市保土ヶ谷区常盤台 79-5

^{*1}Neyagawa-shi, Osaka

^{*2}Hodogaya-ku, Yokohama-shi, Kanagawa

Ck^{-3} の形のべき乗則に従うことを示す議論を与えた。

彼らの論文においてはモデルの決め方の詳細が書かれていないこともあり、様々な修正案が提出されている。本稿では次の定義を採用する： $n = 1, 2, \dots$ に対して、 n 個の頂点を持つランダムグラフ G_n を次のように構成する。 G_1 を頂点 1 のみが孤立しているグラフとし、 G_2 を頂点 1 と頂点 2 が 1 本の辺で結ばれたグラフとする。 $n \geq 3$ に対して、 G_{n-1} に新たな頂点 n を付け加え、既存の頂点のどれか一つと辺で結んで G_n を構成する。この際、既存の頂点の選ばれる確率は G_{n-1} の段階での度数に比例するものとする。このモデルを $BA(n, 1)$ と呼ぶ。新しく付け加える頂点と（重複を許した） m 個の既存の点とが辺で結ばれるモデル $BA(n, m)$ は、 $BA(mn, 1)$ で頂点 $(k-1)m+1, (k-1)m+2, \dots, km$ を縮約し頂点 k とすることによって得られるものとする。

2.4 スモールワールドモデル

円周上に n 個の頂点があり、各頂点は左右 $k/2$ 個ずつの頂点と隣接していると考えて辺で結ぶ。次に、割合 p の辺をランダムに選んでつなぎかえる（つなぎかえにおいて細かいルールがいくつもあるが、それについては述べないことにする）。 $p = 0$ では円環状の格子のままであり、 $p = 1$ では Erdős-Rényi のモデルに似たランダムグラフになる。Watts-Strogatz (1998) は、 p が小さいときこの両極端の状況の両方の性質が得られて、クラスター係数が高く平均頂点間距離の小さいネットワークが実現することを示した。

つなぎかえのルールの煩雑さのために、Watts-Strogatz モデルの数学的に厳密な解析は難しい。より扱いやすいモデルがいくつか提案されているので、そのうちの二つを紹介する。

- NW スモールワールド：Newman-Watts (1999) による。円環状の格子から辺を取り除くことはせず、平均 $pkn/2$ の Poisson 分布に従う本数の弦（ショートカットとも呼ばれる）を追加する。このモデルを $NW(n, k, p)$ で表す。
- BC スモールワールド：Bollobás-Chung (1988) は、Watts-Strogatz 以前に、円環状の格子にランダムに弦を付け加える（より正確には、 n 個の頂点をランダムに組にして $n/2$ 本の弦を引く）ことで平均頂点間距離が短くなることを示している。文献 [2] に従って、このモデルを BC スモールワールドと呼び、 $BC(n, k)$ で表す（Bollobás-Chung が扱ったのは $k = 1$ の場合である）。

3. ランダムウォーク

3.1 離散時間 Markov 連鎖

$\{X_n\}_{n=0,1,2,\dots}$ がグラフ $G = (V, E)$ 上の離散時間 Markov 連鎖であるとは、

$$P(X_{n+1} = y | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x)$$

$$= P(X_{n+1} = y | X_n = x) =: K(x, y)$$

を満たすときにいう。 $K = \{K(x, y)\}_{x, y \in V}$ を推移確率行列と呼ぶ。頂点 x から出発して n ステップ後に頂点 y に到達する確率は行列 K の n 乗 K^n の (x, y) 成分 $K^n(x, y)$ で与えられる。

さて、 K の定義から

$$P(X_{n+1} = y) = \sum_{x \in V} P(X_n = x) K(x, y)$$

である。そこで、 $\pi = \{\pi(x)\}_{x \in V}$ がこの Markov 連鎖の定常分布であるとは、すべての $y \in V$ で

$$\pi(y) = \sum_{x \in V} \pi(x) K(x, y)$$

を満たすことをいう（つまり、 π は行列 K の固有値 1 に対する固有ベクトルである）。例えば、 π がすべての $x, y \in V$ に対して

$$\pi(x) K(x, y) = \pi(y) K(y, x)$$

を満たすとき π は定常分布である。 K が周期性を持たずあらゆる点同士の行き来が正の確率で可能という条件のもとでは、唯一つの定常分布 π が存在し、任意の $x, y \in V$ に対して $n \rightarrow \infty$ のとき $K^n(x, y) \rightarrow \pi(y)$ となる（混合性：出発点 x を忘れる）。

二つの分布 $\pi = \{\pi(x)\}_{x \in V}$, $\tilde{\pi} = \{\tilde{\pi}(x)\}_{x \in V}$ の近さを測る尺度の一つに、全変動距離

$$d_{TV}(\pi, \tilde{\pi}) := \max_{A \subset V} |\pi(A) - \tilde{\pi}(A)|$$

がある。定常分布 π を持つ Markov 連鎖の混合時間 T_{mix} を

$$T_{\text{mix}} := \sup_{x \in V} \min\{n : d_{TV}(\{K^n(x, y)\}_{y \in V}, \pi) < e^{-1}\}$$

によって定義する。

3.2 連続時間 Markov 連鎖

連続時間の Markov 連鎖の特別な場合として、推移確率行列 K を持つ離散時間 Markov 連鎖を自然に連続時間にしたものを定義する：平均 1 の指数分布に従う時間だけ待ち、推移確率に従ってジャンプする。このとき、時刻 0 から時刻 t の間に起こるジャンプの回数は平均 t の Poisson 分布に従うので、時刻 0 で頂点 x から出発し時刻 t で頂点 y に到達する確率 $H_t(x, y)$ は

$$H_t(x, y) = \sum_{n=0}^{\infty} e^{-t} \frac{t^n}{n!} K^n(x, y)$$

となる。これは、行列の指数関数 $e^{-(I-K)t}$ (I は単位行列) の (x, y) 成分である。

3.3 格子上のランダムウォーク

単純ランダムウォークは、推移確率

$$K(x, y) = \begin{cases} 1/d(x) & (\{x, y\} \in E), \\ 0 & (\text{その他}) \end{cases}$$

を持つ Markov 連鎖である。\$d\$ 次元の正方格子 \$\mathbb{Z}^d\$ の上の単純ランダムウォークでは、空間の次元が \$d = 1, 2\$ の場合は再帰的（出発点に戻ってこない確率は 0）、\$d \geq 3\$ の場合は非再帰的（出発点に戻ってこない確率が正）であることが知られている。

3.4 複雑ネットワーク上のランダムウォーク：被覆時間

ネットワーク全体を探索するのに要する時間について調べる。連結なグラフ \$G = (V, E)\$ 上の単純ランダムウォーク \$X_n\$ を考える。頂点 \$v \in V\$ から出発したときに、\$X_n\$ が \$G\$ のすべての頂点を訪問するまでにかかる時間の平均値を \$C_v\$ で表す。\$G\$ の被覆時間 \$C_G\$ を

$$C_G := \max_{v \in V} C_v$$

によって定義する。頂点集合 \$V\$ の要素の個数が \$n\$ のとき、

$$(1 - o(1))n \log n \leq C_G \leq (1 + o(1))\frac{4}{27}n^3$$

が成り立つことが知られている (Feige, 1995)。この不等式の下限はすべての頂点間が辺でつながれた \$n\$ 頂点完全グラフによって達成される。

複雑ネットワーク \$G\$ 上の単純ランダムウォークの被覆時間 \$C_G\$ については、Cooper-Frieze による一連の研究がある (文献 [4] を参照)。 $f(n) \sim g(n)$ **whp** とは、任意の \$\varepsilon > 0\$ に対して \$1 - \varepsilon \leq f(n)/g(n) \leq 1 + \varepsilon\$ となる確率が \$n \rightarrow \infty\$ で 1 に収束することを意味する（‘with high probability’ の略）。

- 連結な Erdős-Rényi ランダムグラフ：\$c > 1\$ のとき、\$G = \text{ER}(n, (c \log n)/n)\$ に対して

$$C_G \sim c \log \left(\frac{c}{c-1} \right) n \log n \quad \text{whp.}$$

- ランダム \$d\$-正則グラフ：\$d \geq 3\$ のとき、\$G = \mathcal{G}_d(n)\$ に対して

$$C_G \sim \frac{d-1}{d-2} n \log n \quad \text{whp.}$$

- 優先的選択モデル：\$m \geq 2\$ のとき、\$G = \text{BA}(n, m)\$ に対して

$$C_G \sim \frac{2m}{m-1} n \log n \quad \text{whp.}$$

3.5 複雑ネットワーク上のランダムウォーク：混合時間

定常状態への収束に関する混合時間について調べる。これは被覆時間の評価においても基本的な役割を果たしている。

周期性の問題を避けるために、単純ランダムウォークを少し変形しよう：ウォーカーが次数 \$d(x)\$ の頂点 \$x\$ にいるとき、確率 \$1/2\$ で次の時刻においても点 \$x\$ にとどまり、確率 \$1/2d(x)\$ ずつで次の時刻に \$x\$ の隣接点のいずれかに移動することにする。これは遷移確率行列 \$K = \{K(x, y)\}_{x, y \in V}\$ が

$$K(x, y) = \begin{cases} 1/2, \\ 1/2d(x) & (\{x, y\} \in E), \\ 0 & (\text{その他}) \end{cases}$$

で与えられる Markov 連鎖であり、「ものぐさランダムウォーク」(lazy random walk) とも呼ばれる。こう変形しても、長時間の挙動に本質的な違いは生じない。

定常分布 \$\pi\$ において \$S \subset V\$ に滞在する確率は \$\pi(S) := \sum_{x \in S} \pi(x)\$ である。定常分布に急速に収束するには、\$\pi(S)\$ の小さい \$S\$ に「つかまってしまう」可能性が小さければよいと考えられる。

$$\Phi(S) := \frac{\sum_{x \in S, y \in S^c} \pi(x) K(x, y)}{\sum_{x \in S} \pi(x)}$$

を \$S\$ のコンダクタンスと呼ぶ (\$S^c = V \setminus S\$)。この量を確率論的に解釈すると、初期分布として定常分布 \$\pi\$ をとったとき、\$S\$ から出発して、1 歩で \$S\$ から \$S^c\$ へと脱出する条件付確率である。Sinclair-Jerrum (1989) は、滞在確率がある程度小さい頂点集合にわたる最小コンダクタンス

$$\min_{S \subset V: S, S^c \neq \emptyset, \pi(S) \leq 1/2} \Phi(S)$$

を用いた混合時間の評価を与えた。\$\Phi\$ において最小値をとる \$S\$ の範囲を狭めるなどして、実用性を高め良い評価を得る研究も進んでいる (例えば文献 [5] を参照)。

複雑ネットワーク上のものぐさランダムウォークの混合時間 \$T_{\text{mix}}\$ について、次が知られている。

- 連結な Erdős-Rényi ランダムグラフ (Cooper-Frieze, 2003): \$c > 1\$ の \$\text{ER}(n, (c \log n)/n)\$ において、\$T_{\text{mix}} = O(\log n)\$.
- 与えられた次数分布を持つランダムグラフ (文献 [2] 6.3 節): ランダム \$d\$-正則グラフにおいては定常分布が \$\pi(x) = 1/d\$ となり、\$S\$ のコンダクタンスは、\$S\$ の体積 (頂点数) に対する \$S\$ の境界の辺 (\$S\$ と \$S^c\$ とを結ぶ辺) の本数の比という幾何学的な量に比例する。その最小値であるエッジ等周定数は Bollobás (1988) により詳しく評価されている。最小次数が 3 以上の任意の次数分布を持つランダムグラフに対しても、\$T_{\text{mix}} = O(\log n)\$ であることを示すコンダクタンスの下からの評価がある (文献 [2] 6.3 節)。ランダム 3-正則グラフに関しては文献 [6] も参照されたい。

- 優先的選択モデル (文献[7]): $m \geq 2$ の $BA(n, m)$ において, $T_{\text{mix}} = O(\log n)$.
- NW スモールワールド (文献[2] 6.6 節): $NW(n, k, p)$ において, T_{mix} は短くとも $O(\log^2 n)$, 長くとも $O(\log^3 n)$ と評価される. 関連するモデルについて, より詳しいことが分かっている: ランダム 3-正則グラフの各辺を, 独立な幾何分布 (パラメータは上記の p と関連付けてうまく選ぶ) に従う本数だけ辺を直列につないだものに置き換えて得られるランダムグラフでは $T_{\text{mix}} = O(\log^2 n)$ (Reed の結果, 文献[2] 6.7 節を参照).
- BC スモールワールド (Mossel の結果, 文献[2] 6.3 節を参照): $BC(n, 1)$ において, $T_{\text{mix}} = O(\log n)$.

4. 投票者モデル

投票者モデルは Clifford-Sudbury (1973) と Holley-Liggett (1975) により独立に導入された. 詳しくは文献[8]を参照.

4.1 投票者モデルの定義

グラフ $G = (V, E)$ 上の投票者モデルを次のように定義する. 各頂点 $x \in V$ は時刻 $t \geq 0$ で意見 $\xi_t(x) \in \{0, 1\}$ を持つとする (x にいる人が, ある問題に対して反対または賛成していることを表す). 各 $x \in V$ ごとに独立に, 平均 1 の指数分布に従う時間が経過するごとに周囲の意見を取り入れようとし, G 上で x に隣接する点の一つを一樣な確率で選んでその人の意見を採用する. 周囲の賛成・反対の比率に応じて意見を変化させると言い換えてもよい. ある時刻 t においてすべての $x \in V$ で $\xi_t(x) \equiv 1$ または $\xi_t(x) \equiv 0$ となったとすると, それ以降の時刻では変化が起らない. この状態を「コンセンサスが成立している」という.

4.2 ランダムウォークとの関係

頂点 $x \in V$ が時刻 t で意見 $\xi_t(x)$ を持っているというとき, 時刻 t までにどのような影響を受けて現在の意見を持つに至ったかを逆に追跡することを考える. この追跡の様子を「双対プロセス」と呼ぶ. 時刻 t での x の意見のもとをたどるとき, この双対プロセスは時刻 0 で x から出発すると定義する. 時刻 t 以前で, x が隣人の意見を取り入れた最後の時刻を s とするとき, この隣人が y であったとすると, 双対プロセスは時間区間 $[0, t-s]$ では x にとどまり, 時刻 $t-s$ に y にジャンプすると定義する. 以下同様に, 時刻 $t-s$ 以前で, y が隣人の意見を取り入れた最後の時刻までさかのぼり, その隣人にジャンプすることを繰り返しれば双対プロセスが構成できる. これは平均 1 の指数分布に従う時間ごとに隣接する点に移動するランダムウォークになっている. 各頂点から一斉に双対プロセスを走らせたとき, ある時刻で二つのプロセスが衝突すると, その時刻以後は二つのプロセスが合体して同じ動きをするように

なることが分かる. 投票者モデルは, このような「合体的ランダムウォーク」(coalescing random walk) の問題に帰着することができる.

4.3 無限格子上の投票者モデル

密度 p の Bernoulli 測度を初期分布とする投票者モデルの時刻 t での状態を ξ_t^p で表す. Holley-Liggett (1975) は $\mathbb{Z}^d$ 上の投票者モデルの定常状態について次のことを証明した.

- 次元 $d = 1, 2$ のとき, 初期の意見の配置によらずに, 完全なコンセンサスに到達する: 任意の $x, y \in \mathbb{Z}^d$ に対して, $t \rightarrow \infty$ のとき $P\{\xi_t(x) = \xi_t(y)\} \rightarrow 1$.
- 次元 $d \geq 3$ のとき, ξ_t^p は $t \rightarrow \infty$ で意見 1 のサイトの密度が p である定常状態 ξ_∞^p に分布の意味で収束する. これは, $d = 1, 2$ と $d \geq 3$ でランダムウォークの再帰性に違いが出ることに対応している.

4.4 有限トーラス上の投票者モデル

有限グラフ上の投票者モデルでは, いずれ必ずコンセンサスが成立する. したがって, コンセンサスが成立するまでの時間とグラフの大きさとの関係に興味を持たれる. Cox (1989) は有限トーラス $(\mathbb{Z} \bmod N)^d$ 上の投票者モデルにおいてコンセンサスに到達するまでの時間 τ_N について次のことを証明した:

$$\begin{aligned} d = 1 \text{ のとき} & \quad s_N = N^2, \\ d = 2 \text{ のとき} & \quad s_N = N^2 \log N, \\ d \geq 3 \text{ のとき} & \quad s_N = N^d \end{aligned}$$

とおくと, 次元 d に依存する定数 C_d がとれて

$$\begin{aligned} \tau_N &= O(s_N), \\ E\tau_N &\sim C_d \{-p \log p - (1-p) \log(1-p)\} s_N. \end{aligned}$$

$d \geq 3$ のとき, 十分大きい $o(s_N)$ の時刻に観察すると, この有限系の状態は無限系の定常分布に近いことも分かる.

$(\mathbb{Z} \bmod N)^d$ の頂点数は $O(N^d)$ だから, $d = 1, 2$ のとき τ_N はネットワークの体積より大きいオーダーになり, $d \geq 3$ では体積と同じオーダーになる.

4.5 複雑ネットワーク上の投票者モデル

次の結果は連結な Erdős-Rényi ランダムグラフ, 優先的選択モデル, BC スモールワールド, NW スモールワールドに適用可能である.

- n 個の頂点を持つランダムグラフ上でランダムに二つの頂点を選び, そこから二つの独立な連続時間ランダムウォークを出発させる. このとき, 二つのランダムウォークが会うまでの時間の平均値 $\sim Cn$ である (文献[2] 6.8 節).
- n 個の頂点を持つ種々のランダムグラフ上の投票者モデル $\{\xi_t^p\}_{t \geq 0}$ においてコンセンサスが成立するまでの

時間は $C_p n$ 以上である (文献 [2] 6.9 節). 前節で述べた Cox (1989) の結果を踏まえると, $d \geq 3$ のタイプの挙動が示唆される.

Durrett は, 一般にランダムグラフ G 上の投票者モデルに対して G に依存する定数 C がとれて頂点数 $n \rightarrow \infty$ のときコンセンサスに到達するまでの時間 $\sim Cn$ となると予想している. こうしたことを証明するためには, すべてのサイトにウォーカーが存在するという状態から出発した合体的ランダムウォークの挙動の解析が必要である. その第 1 段階の結果として, 文献 [2] 定理 6.9.4 を参照されたい.

参考文献

[1] 今野紀雄: “複雑ネットワーク上のコンタクト・プロセスへ向けて”,

数理科学, no.524, pp.61–68, 2007.

[2] R. Durrett: Random graph dynamics. Cambridge University Press, 2006.

[3] 増田直紀, 今野紀雄: 複雑ネットワークの科学. 産業図書, 2005.

[4] C. Cooper and A. Frieze: “The cover time of the preferential attachment graph,” J. Combin. Theory Ser. B, vol.97, no.2, pp.269–290, 2007.

[5] N. Fountoulakis and B. Reed: “Faster mixing and small bottlenecks,” Probab. Theory Relat. Fields, vol.137, no.3–4, pp.475–486, 2007.

[6] N. Berestycki and R. Durrett: “Limiting behavior for the distance of a random walk,” preprint, 2007.

[7] M. Mihail, C. Papadimitriou and A. Saberi: “On certain connectivity properties of the internet topology,” J. Comput. System Sci., vol.72, no.2, pp.239–251, 2006.

[8] T.M. Liggett: Stochastic interacting systems: Contact, voter and exclusion processes. Springer-Verlag, 1999.

竹居正登 (Masato Takei)

2005 年神戸大学大学院自然科学研究科博士課程後期課程修了. 博士 (理学). 慶應義塾大学理工学部特別研究助手, 横浜国立大学工学部非常勤講師等を経て, 現在大阪電気通信大学工学部専任講師. 専門は確率論. 日本数学会会員.

今野紀雄 (Norio Konno)

1982 年東京大学理学部数学科卒業, 1987 年東京工業大学大学院理工学研究科博士課程単位取得. 現在横浜国立大学大学院工学研究院教授. 博士 (理学). 研究テーマは, 無限粒子系, 量子ウォーク, 複雑ネットワーク.