

確率的に生成されるネットワークの規模の評価Ⅱ

—特定の次数列を満たす大規模ランダムグラフについて—

浜 口 幸 弘

1. はじめに

本稿は浜口（2014）の続編に相当する。本稿では，Molloy and Reed（2000）で得られたいくつかの結果を評価するために，ランダムグラフが特定の部分グラフを含む可能性について，期待値の大きさの視点から独自の分析評価を行うものとする。そして本稿で得た結果を Molloy and Reed（2000）の結果と比較して，その類似性と相違性について考察する。

2. 概念と用語の定義

ここでは本稿に関係する Molloy and Reed（2000）による概念と用語の定義を概括する。頂点数 n のランダムグラフ G に対して，各頂点の次数は与えられているものとし，頂点 $i (1 \leq i \leq n)$ の次数を d_i とし，次数 j の頂点の個数を n_j （前回 $d_j(n)$ としていたが，ここでは便宜上改める）と記す。そして，次数列を $\mathbf{D} = n_0, \dots, n_j, \dots$ とする ($j \leq n-1$)。また， $n = \sum_i n_i$ は十分に大きな数とする。このとき，次数列 $\mathbf{D}$ の条件は次のように与えられる。：

次数列 $\mathbf{D}$ に対して， $\lim_{n \rightarrow \infty} \frac{n_i}{n} = \lambda_i$ となるような定数 $\lambda_i \geq 0$ が存在して， $\frac{\sum_{i \geq 1} i n_i}{n} = K + o(1)$ となる定数 K が存在する。

Molloy and Reed（2000）によれば，基準量 $Q(\mathbf{D}) = \sum_{i \geq 1} i(i-2)\lambda_i$ の正負に応じて連結成分の大きさが大きく変わってくる。すなわち，正の場合には大規模な連結成分が生成される結論を得ている。本稿では，1 より小さい正の定数 δ_1 と δ_t ($\delta_1 + \delta_t = 1$) に対して， $n_1 = \delta_1 n$ および $n_t = \delta_t n$ とする。このとき， $Q(\mathbf{D}) = -\delta_1 + t(t-2)\delta_t$ である。また，初期状態における頂点コピーの総数を $2m = n_1 + t n_t$ とする (m は構成されるグラフの辺の本数)。そして前回同様，各頂点の次数が予め 1 または $t (t \geq 3)$ として与えられていて，2 頂点がランダムに接続されるグラフについて議論する。なお，特にこのグラフを対象とする場合は G_t と記すものとする。

さて, Molloy and Reed (2000) によるコンフィグレーションモデルの構成アルゴリズムでは, 頂点コピーのランダムな選択に関して若干の規則を設けているが, 頂点コピーをまったくランダムに選択しても同じ結果となる。すなわち, $i-1$ 回目 ($i \geq 2$) にランダムに選択した頂点コピーと, i 回目にランダムに選択した頂点コピーを接続して辺を構成するという条件のもとでは, 上記 2 つの方法によって辺が構成される確率は明らかに同じである。簡単な例をあげると, 頂点数 6 で各頂点 1, ..., 6 の次数が 1 のグラフを考える。このとき, 辺 $\{1, 2\}$ が構成される確率は前者の場合と後者の場合は, ともに $\frac{1}{5}$ である。したがって, 以降では Molloy and Reed (2000) による構成アルゴリズムではなく, 後者の考え方に基づいて議論を行う。

3. 期待値による連結成分の大きさの評価

次数列が与えられた頂点数 n のランダムな単純グラフを G とし, 同じ次数列が与えられた頂点数 n のランダムな多重グラフを G^* とする。 Q を特定の性質を満たすグラフの集合とすれば, 両者のランダムグラフは次のように関係づけられる。

$$P\{G \in Q\} = \frac{P\{G^* \in Q \wedge G^* \text{ is simple}\}}{P\{G^* \text{ is simple}\}} = \frac{P\{G^* \in Q\} P\{G^* \text{ is simple} \mid G^* \in Q\}}{P\{G^* \text{ is simple}\}}$$

$P\{G^* \text{ is simple}\}$ は正の定数をとる (Bollobás (2001)) ので, $\lim_{n \rightarrow \infty} P\{G^* \in Q\} = 1$ ならば, $\lim_{n \rightarrow \infty} P\{G \in Q\} = 1$ となる。したがって, G の代わりに G^* を対象にすることでその性質を分析できる。よって以降では, G^* を考えることにする。そして, 次数が 1 または t からなるランダム多重グラフ G_t^* が特定の部分グラフを含む可能性について, 期待値の大きさの視点から考察する。すなわち, 前回のようにグラフ生成のプロセスを直接分析するのではなく, 期待値の評価という間接的方法を用いることにする。そのための基本となる考えが以下である。

確率変数 X_k をランダムグラフ G_t^* に含まれる大きさ $k = k(n)$ の閉路の個数とする。 X_k がポアソン分布することを示せば, 前述の考え方に従い, 一定条件下にある G_t に対して, $\lim_{n \rightarrow \infty} P\{X_k = 0\} = 0$ すなわち, $\lim_{n \rightarrow \infty} P\{X_k \geq 1\} = 1$ が成り立つ。

そこで次の命題を提案するが, その中では不等式 $\left(\frac{n}{e}\right)^n \leq n!$ および $\left(\frac{n}{k}\right) \leq \left(\frac{en}{k}\right)^k$ による評価を適宜行うことにする。また, $(n)_r = n(n-1)\cdots(n-r+1)$ と定義する。

命題 4

X_k をランダムグラフ G_t^* に含まれる大きさ $k = k(n)$ の閉路の個数とする。また, 1 より小さい正の定数 δ_1 と δ_t に対して, $n_1 = \delta_1 n$ および $n_t = \delta_t n$ とする ($\delta_1 + \delta_t = 1$)。さらに, 任意の $0 < \delta < 1$ に対して, $k = (\log n)^\delta$ とする。ただし, $t(t-2)\delta_t > \delta_1$ である。

このとき, $\lambda = \frac{1}{2k} \left(\frac{t(t-1)}{t + \frac{\delta_1}{\delta_t}} \right)^k$ とすれば, X_k は次のようなポアソン分布をする。

$$P\{X_k=i\} = \frac{\lambda^i e^{-\lambda}}{i!}$$

証明

まず, $n = n_1 + n_t$ であり, $2m = n_1 + tn_t$ とおく。 G_t^* において, 大きさ k の閉路をなす頂点コピーのすべての組み合わせの個数を C_k とする。また, このような任意の閉路 1 つが存在する確率を p_k とすれば, 次のようになる。

$$C_k = \frac{(k-1)!}{2} \binom{n_t}{k} (t(t-1))^k, \quad p_k = \frac{1}{(2m-1)(2m-3)\cdots(2m-2k+1)}$$

よって,

$$\begin{aligned} E(X_k) &= C_k p_k = \frac{(t(t-1))^k}{2k} \frac{n_t(n_t-1)\cdots(n_t-k+1)}{(n_1+tn_t-1)(n_1+tn_t-3)\cdots(n_1+tn_t-2k+1)} \\ &= \frac{(t(t-1))^k}{2k} \frac{1}{\left(t + \frac{\delta_1}{\delta_t} - \frac{1}{\delta_t n}\right)} \frac{\left(1 - \frac{1}{\delta_t n}\right)}{\left(t + \frac{\delta_1}{\delta_t} - \frac{3}{\delta_t n}\right)} \cdots \frac{\left(1 - \frac{k-1}{\delta_t n}\right)}{\left(t + \frac{\delta_1}{\delta_t} - \frac{2k-1}{\delta_t n}\right)} \end{aligned}$$

よって,

$$\frac{t(t-1)^k}{2k} \left(\frac{1 - \frac{k-1}{\delta_t n}}{t + \frac{\delta_1}{\delta_t} - \frac{1}{\delta_t n}} \right)^k \leq E(X_k) \leq \frac{(t(t-1))^k}{2k} \left(\frac{1}{t + \frac{\delta_1}{\delta_t} - \frac{2k-1}{\delta_t n}} \right)^k$$

このことから, $E(X_k) \sim \frac{1}{2k} \left(\frac{t(t-1)}{t + \frac{\delta_1}{\delta_t}} \right)^k$ を得る。

次に, r 個の k -閉路から構成される順序つき r -tuples の個数の確率変数を $(X_k)_r$ とし, その期待値を $E((X_k)_r)$ とすると, 各閉路の頂点共有の有無から以下のように分解できる。

$$E((X_k)_r) = E_1((X_k)_r) + E_2((X_k)_r)$$

前者の項は各閉路が頂点を共有しない場合の $(X_k)_r$ の期待値であり, 後者はそうでない場合の $(X_k)_r$ の期待値である。

$$\begin{aligned} E_1((X_k)_r) &= \binom{n_t}{k} \binom{n_t-k}{k} \cdots \binom{n_t-(r-1)k}{k} \left(\frac{(k-1)!}{2} \right)^r \frac{(t(t-1))^{kr}}{(n_1+tn_t-1)(n_1+tn_t-3)\cdots(n_1+tn_t-2(r-1)k-2k+1)} \\ &= \left(\frac{(t(t-1))^k}{2k} \right)^r \frac{(n_t)_k (n_t-k)_k \cdots (n_t-(r-1)k)_k}{(n_1+tn_t-1)(n_1+tn_t-3)\cdots(n_1+tn_t-2(r-1)k-2k+1)} \end{aligned}$$

よって,

$$\left(\frac{(t(t-1))^k}{2k} \right)^r \frac{(n_t-rk+1)^{kr}}{(n_1+tn_t-1)^{kr}} \leq E_1((X_k)_r) \leq \left(\frac{(t(t-1))^k}{2k} \right)^r \frac{n_t^{kr}}{(n_1+tn_t-2rk+1)^{kr}}$$

このことから、 $E_1((X_k)_r) \sim (E(X_k))^r$ となる。一方、 $E_2((X_k)_r)$ については以下のように評価できる。順序つき r -tuples に現れる異なる頂点数を i とすると、 $1 \leq i \leq kr-1$ である。よって、これらの頂点の選び方は $\binom{n_t}{i}$ となる。後は、計算を簡単にするために、複数の閉路において j 回共有される頂点 v を j 個の頂点 $v_1, \dots, v_j$ とみなし、選んだ i 個の頂点を kr 個として r 個の閉路に割り当てる。起こり得る順序つき r -tuples の個数を s とすれば、以下を得る。

$$s \leq \sum_{i=1}^{kr-1} \binom{n_t}{i} \binom{kr}{k} \binom{kr-k}{k} \cdots \binom{k}{k} \left(\frac{(k-1)!}{2} \right)^r (t(t-1))^{kr}$$

また、各順序つき r -tuple が G_t に含まれる確率 p_r は次のようになる。

$$p_r = \frac{1}{(2m-1)(2m-3)\cdots(2m-2kr+1)} \leq \frac{1}{(n-2kr)^{kr}}$$

よって、

$$\begin{aligned} E_2((X_k)_r) &= s p_r \leq \frac{1}{(2m-2kr)^{kr}} \left(\frac{t^{2k}}{2} \right)^r (kr)! \sum_{i=1}^{kr-1} \binom{n_t}{i} \\ &\leq \left(\frac{t^{2k}}{2} \right)^r (kr)! \frac{kr-1}{(2m-2kr)^{kr}} \left(\frac{en_t}{kr-1} \right)^{kr-1} \leq \frac{e(kr)^2}{n_t} \left(\frac{t^2}{2^{\frac{1}{k}} \left(t + \frac{n_1}{n_t} \right)} \right)^{kr} = o(1) \end{aligned}$$

したがって、 $E((X_k)_r) \sim (E(X_k))^r$ となる。すなわち、Bollobás (2001) 定理 1.22 によれば、主張が示されたことになる。□

この命題から、 $P\{X_k=0\} = e^{-\lambda} \rightarrow 0$ となり、ほぼすべての G_t は大きさ $k=k(n)$ の閉路を少なくとも 1 つ含むことを導ける。これは、Molloy and Reed (2000) の結果を改善したものではないが、 X_k がポアソン分布することがわかるという点では、若干意味ある情報が得られると言える。

次に、 G_t に含まれる最大の連結成分の大きさを評価するために、大きさ $k=k(n)$ の部分木 T_k の出現確率を評価したい。ところが、 X_k をランダムグラフ G_t^* に含まれる T_k の個数とすると、 $E((X_k)_r)$ を求めることは前述の閉路の場合よりかなり複雑になるので、 X_k がポアソン分布するか否かを評価することは容易ではない。よって今回は、以下のような手順に従い、 $E(X_k)$ を導出し、マルコフの不等式を利用して X_k の大きさを確率的に評価する。

まず、 T_k を固定して考え、 T_k を含む連結成分を C_k とする。そして、 $C_k \supsetneq T_k$ のとき、 $e \in C_k$ かつ $e \notin T_k$ を満たす辺 e の両端点が T_k に含まれるならば、 e を分離して 2 つの辺にし、各端点を次数 1 のダミーの頂点とする。あるいは、辺 e の 1 つの端点だけが T_k に含まれるならば、含まれない方の端点を次数 1 のダミーの頂点とする。こうして、 i 個のダミーの頂点 $v_1, \dots, v_i$ が得られるとする。そして、 T_k および $v_1, \dots, v_i$ とそれらに接続する辺から得られる孤立した部分木を T'_k とする。 T'_k は次数 1 と t の頂点から構成され、辺の本数が $k+i-1$ である。このとき、 i の範囲は $0 \leq i \leq 2(t-1) + (t-2)(k-2) = (t-2)k+2$ となる。また、 T'_k において、次数 1 の頂点と次数 t の頂点の個数をそれぞれ x_i および y_i とすれば、以下の式が成り立つ。

$$x_i + y_i = k \quad \text{かつ} \quad x_i + i + t y_i = 2(k+i-1) \quad (*)$$

よって, $x_i = \frac{k(t-2) - i + 2}{t-1}$, $y_i = \frac{k+i-2}{t-1}$ となる。このとき, x_i および y_i は整数であると仮定

して議論する (どちらか整数でない場合, T'_k を構成できない)。

次に, T'_k における頂点コピーの組み合わせを考える。まず, T'_k の 1 つの頂点を最上位レベルの根として位置づけ, その隣接点を次の下位レベルに位置づけ, それらの各頂点に隣接する点を次の下位レベルに位置づけ, この手順を繰り返して T'_k を平面上に位置づける。そして, 上位レベルの頂点から順に以下の手続きを行う。そのレベルで未手続きの次数 t の頂点があれば, 上位の 1 個の隣接点と下位の $t-1$ 個の頂点に割り当てる頂点コピーを選択する。その選択方法は $t!$ 通りある。よって, T'_k における頂点コピーの組み合わせは $(t!)^{y_i}$ となる。

次に, T'_k の総数を求める。(*) を満たす次数 1 の頂点が $x_i + i$ 個, 次数 t の頂点が y_i 個与えられているとき, Matousek and Nesetril (2008) によればこれらの頂点から構成される T'_k の総数は次のように与えられる。

$$\frac{(k+i-2)!}{((t-1)!)^{y_i}}$$

ここで, T'_k からダミーの頂点を除けば T_k を得るが, $i=0$ の場合を除き, T_k は重複して数えられるので, a_i で割り引いて補正する。したがって, T_k の頂点を選択して組み合わせるようにすれば, 以下を得る。

$$E(X_k) = \sum_{i=0}^{(t-2)k+2} \binom{n_1}{x_i} \binom{n_t}{y_i} (t!)^{y_i} \frac{(k+i-2)!}{((t-1)!)^{y_i} a_i} \frac{1}{(2m-1)(2m-3)\cdots(2m-2k+3)}$$

今回は上記の理由により, 最も簡単な $i=0$ の場合 ($a_i=1$), すなわち T_k が孤立木となる場合を対象にする。記号の定義は先の命題に従う。

命題 5

$\delta_1 > et(t-1)\delta_t$ とし, $k = \varepsilon n$ ($0 < \varepsilon < \delta_t$) とする。このとき, ほとんどすべての G_t は T_k を含まない。

証明

X_k を G_t^* に含まれる大きさ $k = k(n)$ の孤立木 T_k の個数とする。前述の議論から以下を得る。

$$\begin{aligned} E(X_k) &= \binom{n_1}{x_0} \binom{n_t}{y_0} (t!)^{y_0} \frac{(k-2)!}{((t-1)!)^{y_0}} \frac{1}{(2m-1)(2m-3)\cdots(2m-2k+3)} \\ &\leq \binom{n_1}{x_0} \binom{n_t}{y_0} t^{y_0} k! \frac{1}{(2m-1)(2m-3)\cdots(2m-2k+3)} = o(1) \end{aligned}$$

よって, マルコフの不等式から主張を得る。 □

この結果は, Molloy and Reed (2000) の結果に比べて弱い結果となっている。この場合, 期待値による評価方法に若干の問題があると思われる。

4. 課題

引き続き考察を行って、次回にまとめを行う予定である。

参考文献

- B. Bollobás, *Random Graphs* (2nd edition) (2001). Cambridge Univ. Press
- J. Matousek and J. Nešetřil, *Invitation to Discrete Mathematics* (2008), Oxford Univ. Press
- M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence", *citeseerx.ist.psu.edu* (2000)
- M. Molloy and B. Reed, "The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence", *Combinatorics, Probability and Computing* 7, 295-306 (1998)
- M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence", *Random Structures and Algorithms* 6, 161-180 (1995)
- 浜口幸弘, 確率的に生成されるネットワークの規模の評価 I —特定の次数列を満たす大規模ランダムグラフについて—, 明治学院大学経済研究 147 (2014)