

スケールフリーランダムグラフの直径に関する考察 I

浜 口 幸 弘

1. はじめに

近年、複雑な現実世界のネットワーク (complex real-world networks) は自然科学のみならず各分野で注目されておられ、とりわけスモールワールド (small-world) 現象は多くの研究が行われてきている。例えば最初の頃では、Albert, Jeong and Barabási (1999, 2000) による研究が有名である。この現象では、観測データやコンピュータシミュレーションによって、頂点数 n のグラフの直径はほぼ $\log n$ に近くなることが指摘されている。Barabási and Albert (1999) は最初にその現象のモデル化を行っているが、さらに Bollobás and Riordan (2004) は、それにランダムグラフの理論を取り入れてモデル化し、その直径を理論的に与えている。また、この文献はいろいろな分野で数多く引用されている。本稿では最初の段階として、Bollobás and Riordan (2004) の各定理における疑問点を指摘しつつ議論を進めていくことにする。

2. ランダムグラフの直径

Bollobás and Riordan (2004) によるスモールワールドのランダムモデル化を議論する前に、ランダムグラフにおける直径について述べる。なお、グラフの直径はグラフにおける最長の道 (path) の長さとして定義される。また、ここで対象にするランダムグラフのモデルは、確率空間 $\mathbf{g}(n, M)$ および $\mathbf{g}(n, p)$ のそれぞれに属するグラフ $G(n, M)$ および $G(n, p)$ である (M は辺の本数, p は任意の 2 頂点間における辺の接続確率を表す)。

さて、Bollobás (2001) はランダムグラフの直径の下限と上限を以下の定理として導いている。

[Bollobás (2001) の定理]

ほとんどすべてのランダムグラフのプロセスは以下の不等式を満たす。

$$\left\lceil \frac{\log n - \log 2}{\log \log n} \right\rceil + 1 \leq \text{diam} G_{\tau_0} \leq \left\lfloor \frac{\log n + 6}{\log \log n} \right\rfloor + 3$$

ここで、 τ_0 はランダムグラフのプロセスにおいて、グラフが連結される最小の辺の本数を意味する (hitting time)。

この定理の証明 (上限についての証明) では、省略されている部分もあるので以下に補足して検討していく。なお、参照する記号、定理と補題は、断りのない限り Bollobás (2001) 中に記載されているものである。

まず、 Q を特定の性質 (property) を満たすグラフの集合とすると、定理 2.2 を参照して、不等式 $P(G_M \text{ has } Q) \leq 3P(G_p \text{ has } Q)$ を証明に用いている。ところが、定理 2.2 ではこの不等式について直接は述べていない。この不等式は Q が単調増加 (例えば、グラフの直径がある上限 D で抑えられるということは、単調増加の性質) であるときに、定理 2.2 (iii) の特別な形であることを意味している。その証明に関しては、例えば、Frieze and Karoński (2015) に記されている。

次に、P. 269 の記述にある、「青色グラフにおいて与えられた m 個の頂点は、確率 $1 - o(n^{-1})$ で少なくとも $m \log n/2$ 個かつ高々 $3 \log n/2$ 個の隣接点をもつ」ことを検討してみる。補題 7.5 の証明中の記述から $G \in \mathbf{g}(n, M)$ における孤立点の個数は高々 $\log n$ であるから (これは定理 5.3 および定理 7.2 の $k_0=1$ の場合から導出できる)、巨大連結成分 (giant component) である青色グラフ $G'_M \subset G$ の頂点数を $n-r$ ($0 \leq r \leq \log n$) とする。すなわち、青色グラフ G'_M は確率空間 $\mathbf{g}(n-r, M)$ に属す任意のグラフとみなせる。ここで、 v を青色グラフにおける任意の頂点とし、 v の隣接点の個数を $d_1(v)$ とする。性質 $d_1(v) \leq \log n/2$ は単調減少であり、 $d_1(v) \geq 3 \log n/2$ は単調増加であるから、定理 2.2 (ii) により、 $p' = M / \binom{n-r}{2}$ として、 G'_M の代わりに $G'_{p'}$ で考えてもよい。 v の隣接点の平均個数は $(n-r-1)p' \sim \log n$ であり、隣接点の個数は n が十分大きいとき正規分布するから以下を得る。

$$P\left(d_1(v) \leq \frac{\log n}{2}\right) \sim P\left(d_1(v) \geq \frac{3 \log n}{2}\right) = \sum_{i \geq \frac{3 \log n}{2}} b(i; n-r-1, p) \leq \frac{C}{(\log n)^{\frac{1}{2}}} \frac{1}{n^{0.1}} = o(n^{-0.1})$$

(C : 定数)

ここで、 $\sum_{i \geq \frac{3 \log n}{2}} b(i; n-r-1, p)$ の近似に当たっては、本書において最も厳密な近似式を与える定理

1.1 を用いている。このことから、青色グラフにおいて与えられた任意の頂点は、確率 $1 - o(n^{-0.1})$ で少なくとも $\log n/2$ 個かつ高々 $3 \log n/2$ 個の隣接点をもつ。よって、青色グラフの頂点 v_i ($1 \leq i \leq m$) は少なくとも $\log n/2$ 個かつ高々 $3 \log n/2$ 個の隣接点をもつという事象を A_i とすると、以下が成り立つ。

$$P(A_1 \wedge \cdots \wedge A_m) \geq P(A_1) + \cdots + P(A_m) - (m-1)P(A_1 \vee \cdots \vee A_m) = 1 - o(n^{-0.1})$$

このことから、青色グラフにおいて与えられた m 個の頂点は、確率 $1 - o(n^{-0.1})$ で少なくとも $m \log n/2$ 個かつ高々 $3 \log n/2$ 個の隣接点をもつことを得る。これは証明中の結果とやや異なるが、以下の議論から問題ないと考えられる。

頂点数 $n-r$ のグラフにおいて与えられた m 個の頂点は、少なくとも $m \log n/2$ 個かつ高々 $3 \log n/2$ 個の隣接点をもつという事象を R とし、頂点数 $n-r$ のグラフにおいて、その直径は少なくとも $D+1$

となる事象を Q とする。定理の証明における議論の展開を大雑把に式で表すと、以下となる。なお、近似的に G'_p の代わりに G'_p で考えても同じ結果を得るので、後者で考える。

$$P(G'_p \in Q) = P(G'_p \in \neg R)P(G'_p \in Q | G'_p \in \neg R) + P(G'_p \in R)P(G'_p \in Q | G'_p \in R)$$

定理では上記 2 番目の確率項について議論を進めている。このとき前述の不等式 $P(G_M \text{ has } Q) \leq 3P(G_p \text{ has } Q)$ を用いると、 $P(G'_p \in Q) \rightarrow 0$ を示せば、 $P(G'_M \in Q) \rightarrow 0$ となり、 $G_{\tau_0} \in \mathbf{g}(n, \tau_0)$ の直径は高々 D であることを導ける。ここでもし、定理 2.2 (iii) を適用するならば、 $P(G_M \text{ has } Q) \leq 3M^{1/2}P(G_p \text{ has } Q)$ なので、先ほどの議論から得られる $P(G'_p \in \neg R) = o(n^{-0.1})$ のために、 $P(G'_p \in Q) \rightarrow 0$ は必ずしも保証されない。よって、グラフの直径は一定の下限と上限をもつという、単調性の性質が証明の鍵になっていることがわかる。

さて、この定理によれば、グラフの直径がスモールワールドの場合より小さくなっており、ランダムグラフはスモールワールド現象を表していないといえる。

3. スケールフリーランダムグラフのモデル化

Bollobás and Riordan (2004) はスモールワールドをランダムグラフでモデル化して、その直径の下限と上限を理論的に導出している。そのモデル化についての概略は、以下のとおりである。グラフ G に新しい頂点を 1 個ずつ加えるとき、その頂点から m 本の辺をグラフ G に接続する。このとき、多重辺とループも認めるものとする。

さて、 $m=1$ の場合について、厳密に定式化すると以下ようになる。グラフに追加していく頂点列を v_1, v_2, \dots と固定し（以下、頂点の表記は $v_i=i$ とも記す）、グラフ G における頂点 v の次数を $d_G(v)$ と記す。ランダムグラフのプロセス $(G'_t)_{t \geq 0}$ を次のように帰納的に定義すると、 $\{v_i : 1 \leq i \leq t\}$ 上のグラフ G'_t が構成される。すなわち、1 個の頂点と 1 つのループを持つグラフ G'_1 から出発する。そして、 G'_1 まで構成されたとき、次に追加する頂点 v_t と $v_i \in V(G'_1)$ を 1 本の辺で結ぶ。このとき頂点 i (確率変数) は次のようにランダムに選択される。

$$P(i=s) = \begin{cases} d_{G'_1}(s)/(2t-1) & 1 \leq s \leq t-1 \\ 1/(2t-1) & s=t \end{cases}$$

これは、Barabási and Albert (1999) のモデルを定式化したものとなっている。

$m>1$ の場合については、一度に v_t から m 本の辺が G'_1 に接続される。よって、プロセス (G'_m) については、プロセス (G'_m) において、最初の頂点から m 個の頂点ごとに統合して 1 個の頂点にすれば (G'_m) が得られることになる。このとき、グラフ G'_m の含まれる確率空間を \mathbf{g}_m^n と記す。

さて、 G'_1 と同じ分布を実現する別の方法がペアリングの点から示されている (Bollobás and Riordan (2000))。すなわち、集合 $\{1, 2, \dots, 2n\}$ を n 個のペア (弦となる) に分割する方法 (n -paring) であり、全部で $(2n)!/(n!2^n)$ 通りある。それぞれを「線形化された弦のダイアグラム (linearized chord diagram or LCD)」と呼び、 L で表す。このとき、左から数字を昇順に並べて、1 つのペアをなす 2 個の数字 (要素) が弦で結ばれるようにする。また、各弦において小さい数を左端点 (left endpoint)、大

さい数を右端点 (right endpoint) と呼ぶ。

次に, L からグラフ $\phi(L)$ を構成する方法について略述する。まず小さい数から順に並んだ L に対して, 一番左の端点から出発して最初の右端点に達したら, それまでのすべての端点と右端点を含めて 1 つの頂点に統合する。次に, 残りの端点を左から進んで最初の右端点に達したら, 先ほどと同様にして 2 番目の頂点を作る。辺については, 各弦を右端点に対応する頂点と左端点に対応する頂点とを結ぶ辺で置き換える。

ここで, L が n 個の弦をもつ $(2n)!/(n!2^n)$ 個の LCDs から一様ランダムに選択されるならば, $\phi(L)$ の分布がスケールフリーグラフと同じ分布になることを n に関する帰納法で証明する。 L に関する主張の前提が成り立つとして, L から, L の一番右端にある右端点とそれに対応する左端点を結ぶ弦をとり, 結果としてできる LCD を L' とする。 $n-1$ のとき主張が成り立つと仮定するので, $\phi(L')$ の分布はスケールフリーグラフと同じ分布になる。さて除外した弦を改めて L' に加えるとする。右端点を L' の一番右端にし, 左端点を L' の入り得る位置 (各端点の両隣) に等確率で入れると, その確率は $1/(2n-1)$ である。このとき, L' が決定されている条件の下で, L は一様ランダムに構成される。上述のグラフ $\phi(L)$ の構成方法に従えば, この追加する弦に対応する辺は, 頂点次数に比例した確率 (すなわち, 前述の式で示す頂点 i が選択される確率) で $\phi(L')$ の 1 つの頂点に接続されることになり, 主張は成り立つ。

4. グラフの直径の下限に関する考察

まず, G_m^n の直径を評価する Bollobás and Riordan (2004) の定理を以下に示す。

[Bollobás and Riordan (2004) の定理]

$m \geq 2$ を整数とし, ε を正の実数として固定する。このとき, ほとんどすべての G_m^n は連結であり, その直径 $\text{diam}(G_m^n)$ は以下の不等式を満たす。

$$(1 - \varepsilon) \frac{\log n}{\log \log n} \leq \text{diam}(G_m^n) \leq (1 + \varepsilon) \frac{\log n}{\log \log n}$$

この結果は Bollobás (2001) の定理と類似しているが, $m=1$ の場合には, $\text{diam}(G_m^n)$ はほぼ $\log n$ になることも証明されている。

さて, $\text{diam}(G_m^n)$ の下限を導く場合には, 前述の LCD を用いる方法は必要としない。ランダムグラフ理論においてよく用いられる期待値による方法を考えればよい。まず Bollobás and Riordan (2004) は $\text{diam}(G_m^n)$ の下限を得るために次の補題を導いている。

[Bollobás and Riordan (2004) の最初の補題]

$1 \leq i \leq j$ に対して, 以下の式が成り立つ。

$$(1) \quad P(g_j = i) = O((ij)^{-1/2})$$

また, $1 \leq i < j < k$ に対して, 以下の式が成り立つ。

$$(2) \quad P(g_j = i, g_k = i) = O(i^{-1}(jk)^{-1/2})$$

この補題は前述の定理の不等式の下限を証明する際に用いられている。ここで, 上記補題について検

討する。この補題の証明では、次の式が成り立つとして、それを証明に用いている。

$$P(g_j=i, g_k=i | G_1^{k-1}) = I_{\{g_j=i\}} \frac{d_{k-1,i}}{2k-1}$$

ここで、 I_A は事象 A の指示関数 (indicator function) であり、 $d_{t,i} = d_{G_1^t}(i)$ はグラフ G_1^t における頂点 i の次数を表す。本文中の(3)式から、右辺は以下となる。

$$I_{\{g_j=i\}} \frac{d_{k-1,i}}{2k-1} = I_{\{g_j=i\}} P(g_k=i | G_1^{k-1})$$

さて条件付確率の定義により、事象 A, B, C に対して、 $P(A \wedge B | C) = P(A | C) P(B | C \wedge A)$ となるので、上の式から以下を得る。

$$I_{\{g_j=i\}} = P(g_j=i | g_k=i \wedge G_1^{k-1})$$

このとき、左辺は 0 または 1 の値となるが、右辺は G_1^{k-1} かつ $g_k=i$ を満たすプロセスの条件下で $g_j=i$ を満たすプロセスの出現確率であり、これは 0 と 1 の間をとる。両辺は一致せず、この点において疑問が提示できるが、実は、これに続く補題ではこの補題の(1) だけで十分なので、結果的に問題ないと言える。

5. グラフの直径の上限に関する考察

Bollobás and Riordan (2004) によれば、グラフの直径の上限については前述の LCD では導けないということで、彼らは新たな LCD を定義している。以下では、この LCD の構成方法について概略する。いま $N=mn$ とし、 N 個の弦からなるベアリングを構成することを考える。このとき、 x 軸上で右端点の位置を示す確率変数を r_1, \dots, r_N とする。これらは区間 $(0, 1)$ 上で密度関数 $2x$ に従う独立変数である。そして、各 r_1, \dots, r_N に区間 $(0, 1)$ に属す実数をランダムに割り当てたら、これらを昇順に並べ替えて、小さいほうから R_1, \dots, R_N とする。次に、区間 $[0, R_i], 1 \leq i \leq N$ 上で一様分布する独立変数を L_i とし、弦 $\{L_i, R_i\}$ を構成して LCD を得る。この LCD を用いて G_1^{mn} を構成する方法は、前述の L からグラフ $\phi(L)$ を構成する方法と同じで、 $R_{k-1} < L_i < R_k (R_0=0)$ を満たす i に対して、各頂点 i から頂点 k までを統合して 1 つの頂点にする。 G_m^n を構成する方法についても前述の方法と同様である。すなわち、 m 個の頂点ごとに 1 つの頂点に統合するので、 $W_i=R_{mi}$ とし、 R_i の代わりに W_i を対象にする。このとき、 $W_{k-1}=R_{m(k-1)} < L_{m(i-1)+j} < W_k=R_{mk} (1 \leq i \leq n, 1 \leq j \leq m)$ に対して、 $L_{m(i-1)+j}$ は、 x 軸上を右に進んで最初の右端点である $R_{m(k-1)+s}$ に 1 つの頂点になるように統合され、さらに R_{mk} に統合されて、また $R_{m(i-1)+j}$ は R_{mi} に統合される結果、 G_m^n において頂点 i と頂点 k は辺で結ばれる。

この定義の下で、先ほどのようにグラフ $\phi(L)$ の分布がスケールフリーグラフと同じ分布になるかどうか検討してみる。先ほどと同じ記号を用い、簡単のために $m=1$ の場合について前述の帰納法による証明方法に従い考える。 L から弦 $\{L_n, R_n\}$ を除くとする (R_n は最も右端にある端点)。 L_n がどの区間(隣り合う右端点の間)に含まれていたか考えると、一様分布するという新たな LCD の定義によれば、その確率は区間幅 $|R_k - R_{k-1}|$ に比例する。いま、 $k+2 < k'$ として、同じ区間幅である $[R_{k-1}, R_k]$ と

$[R_{k'-1}, R_{k'}]$ を考えると、それぞれの区間に L_n は同じ確率で含まれていたことになる。ここで、右端点 R_i と R_j ($i > j$) に対して、 $R_{k'} < R_i$ および $R_k < R_j < R_{k'-1}$ とする。このとき、 R_i に対応する左端点 L_i は両区間に同じ確率で含まれるが、 R_j に対応する L_j は $[R_{k'-1}, R_{k'}]$ に含まれることはない。すなわち、左側にある区間の方が弦の左端点を含む可能性が高いと思われる。これは R_k の方が $R_{k'}$ よりも、 $\phi(L)$ において頂点の次数が高くなることを意味するので、スケールフリーグラフの構成方法に反する。したがって、スケールフリーグラフについて論じていないかもしれないという疑問を提示できる。

6. 今後の課題

引き続き、Bollobás and Riordan (2004) の証明方法を検討するとともに、上限について他の証明方法を模索する必要もあろう。しかし、任意の 2 頂点の距離がある値以下になることを示すには、その組み合わせのオーダー n^2 を掛けることになるので、補題 3 の (1) の確率では、そのオーダーが大きく、下限と同様の証明方法は使えない。また、この確率は精度の高い近似なので、簡単には改善することはできない。よって、この確率を用いる場合には、別の点から証明しなければならない。

参考文献

- A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science* (1999)
 R. Albert, H. Jeong and A.-L. Barabási, "Diameter of the world-wide web", *Nature* (1999)
 A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web", *Physica A* (2000)
 B. Bollobás, "Random Graphs (2nd ed.)", *Cambridge University Press* (2001)
 B. Bollobás and O. Riordan, "Linearized chord diagrams and an upper bound for Vassiliev' invariants", *J. Knot Theory Ramifications* 9 (2000)
 B. Bollobás and O. Riordan, "The diameter of a scale-free random graph", *Combinatorica* (2004)
 A. Frieze and M. Karoński, "Introduction to Random Graphs", *Cambridge University Press* (2015)