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—

浜 口 幸 弘

1. はじめに

与えられた次数列を満たすランダムグラフの規模を決定する基準量に関して、それを扱う論文は多くはないが、Molloy and Reed (2000) はグラフ論のみならず他の自然科学分野からも非常に多く引用されている。この条件下でのランダムグラフは $G(n, M)$ モデルの特別な型なので、辺を選択する際の確率が条件付き確率になり、これが理論化を複雑にしている。よって、この問題にうまく対応することが鍵となる。本稿では、Molloy and Reed (2000) の論文を中心に議論を進め、その問題点を指摘しつつ、独自の分析評価を行うことにする。また、2つの結果を比較しやすくするために、用いる概念および記号は Molloy and Reed (2000) に準ずるものとする。

2. Molloy and Reed の定理の概括

ここでは、本稿で考察の対象とする Molloy and Reed (2000) による定理を概括する。次数列が与えられた頂点数 n のランダムグラフ G に対して、次数 i の頂点の個数を $d_i(n)$ と記し、その次数列を $\mathbf{D} = (d_0(n), d_1(n), \dots)$ とする ($i \leq n-1$)。一方、頂点 v の次数は $\deg(v)$ と記す。また、 $n = \sum_i d_i(n)$ は十分に大きな数とする。そして、次数列 $\mathbf{D}$ が満たす条件を次のように与えている：

次数列 $\mathbf{D}$ に対して、 $\lim_{n \rightarrow \infty} \frac{d_i(n)}{n} = \lambda_i$ となるような定数 $\lambda_i \geq 0$ が存在し、また、 $\frac{\sum_{i \geq 1} i d_i(n)}{n} = K + o(1)$ $= K + o(1)$ となる定数 K が存在する。また、次数列 $\mathbf{D}$ は “well-behaved” (Molloy and Reed (2000) 参照) である。

次に、 G に大規模な連結成分 (component) が含まれることを評価する基準量として、 $Q(\mathbf{D}) = \sum_{i \geq 1} i(i-2)\lambda_i$ を設定している。この基準量が議論の中心をなすことになる。また、最初の試行で次数 i の

頂点が選択される確率を $p_i(n) = \frac{id_i(n)}{\sum_{j \geq 1} jd_j(n)} = \frac{i\lambda_i}{K} + o(1)$ としている。なお、主張がほとんどすべてのグラフ (almost every graph) において成り立つ場合は、(a.e.) を付与する。この定義のもと、Molloy and Reed (2000) による主要な定理は以下のようになる。

定理 1 Molloy and Reed (2000)

次数列 $\mathbf{D}$ は、任意の n 、ある $\varepsilon > 0$ に対して、 $i > n^{\frac{1}{4}-\varepsilon}$ のとき $d_i(n) = 0$ となるような、前述の $\mathbf{D}$ の条件を満たす次数列とする。また、 G は頂点数 n で、次数 i の頂点を $d_i(n)$ 個もつランダムグラフとする。このとき、

(a) $Q(\mathbf{D}) > 0$ ならば、ある定数 $\zeta_1 > 0$ および $\zeta_2 > 0$ に対して、ほとんどすべての G は少なくとも $\zeta_1 n$ 個の頂点および $\zeta_2 n$ 個の閉路をもつ連結成分を含む。また、 $Q(\mathbf{D})$ が有限ならば、ほとんどすべての G は、ある定数 $\gamma > 0$ に対して、 $\gamma \log n$ より大きい連結成分をただ 1 つもつ。

(b) $Q(\mathbf{D}) < 0$ かつ、ある $0 \leq \omega(n) \leq n^{\frac{1}{8}-\varepsilon}$ に対して、 $d_i(n) = 0$ ($i \geq \omega(n)$) ならば、ある定数 $R > 0$ に対して、ほとんどすべての G は $R\omega(n)^2 \log n$ 個以上の頂点をもつ連結成分をもたない、また、ほとんどすべての G は $2R\omega(n)^2 \log n$ 個以上の閉路をもたない。また、ほとんどすべての G において、どの連結成分も高々 1 個の閉路しかもたない。

さて、本稿では大きな連結成分が生成される (a) の場合を対象にするが、論文中の複数の補題には誤植、あるいは誤りと思われる点がいくつかあるので、適宜考察を与えることにする。

3. コンフィギュレーションモデルの概要

次数列が与えられた頂点数 n のランダムグラフ G の構成は、一般にコンフィギュレーションモデルによって実現される。 n 頂点および次数列 $\mathbf{D}$ のランダムなコンフィギュレーション F は次のように定義される。

各頂点 v に対して、その次数分 $\deg(v)$ 個のコピーを作り、全体のコピーの集合を L とする。そして、 L の 2 つのコピーをランダムに選んでマッチングを構成する。すべてのマッチングがなされたとき、コンフィギュレーション F が完成する。

よって、次数の総和は偶数となる必要があるので、以降、この条件は満たされているものとする。また、1 つのマッチングの対がグラフにおける 1 本の辺に対応するので、得られる各 F は多重グラフを表している。すなわち、1 頂点におけるループや 2 つの頂点を結ぶ 2 本以上の辺を認めている。しかし、一定条件下において、ランダムな多重グラフの性質 P を対象にすることで、ランダムな単純グラフの性質 P を議論できるということは、よく知られた事実である。例えば、Molloy and Reed (2000) の場合、補題 1 と 2 がその説明をなしている。なお、本稿では後述のように $Q(\mathbf{D})$ は有限とするので、補題 2 に相当する。

そこで, Molloy and Reed (2000) による F の構成アルゴリズムを説明する。まず, 次のように各用語を定義する。頂点 v に対して, そのすべてのコピーがマッチングされているならば, v は “completely exposed (完全開示)” 状態にあるといい, コピーのすべてではなく一部がマッチングされているならば, v は “partially exposed (部分開示)” 状態にあるという。よって, それ以外の頂点は “unexposed (未開示)” 状態にある。また, 部分開示の頂点のコピーにおいて, まだマッチングされていないコピーを “open” 状態にあるという。

F の構成アルゴリズム :

1. 各頂点 v に対して, その次数分 $\deg(v)$ 個のコピーを作り, 全体のコピーの集合を L とする。
2. L の要素が尽きるまで以下の手続きを繰り返す。
 - (a) L の任意の要素 1 つを選択し, 次に, その対となるもう 1 つの要素を L からランダムに選択する。これら 2 つの要素 (マッチング) を L から除外する。
 - (b) 部分開示の頂点がある限り以下の手続きを繰り返す。尽きれば (a) に戻る。

部分開示にある頂点のオープンコピーを選択し (ランダムでなくてもよい), その対となる要素を L からランダムに選択する。これら 2 つの要素 (マッチング) を L から除外する。

このアルゴリズムによれば, 1 つの連結成分が完成してから, 新たな次の連結成分の構成に取り掛かることになる (手続き 2 (a) に戻る)。また, 各頂点の 2 つのコピー間で起こり得るマッチングはすべて等確率である。なお, Molloy and Reed (2000) では, 2 (a) における任意の 1 つの要素の選択は, 通常, ランダムな確率的選択としているので, 本稿でもそれに従うものとする。

次にいくつかの確率変数を定義する。 F の構成アルゴリズムにおいて, i 番目の対が決定された直後のオープンな頂点コピーの個数を X_i とする。よって, 手続き 2 (b) において 2 番目に選択されるコピーの頂点 (次数 d とする) が未開示ならば, X_i は $d-2$ 増加し, 部分開示ならば, X_i は 2 減少する。したがって, アルゴリズムを通して $X_i \geq 0$ である。また, 後者の頂点コピーの対を “backedge” と呼び, バックエッジの総数を Y_i とする。さらに, アルゴリズムの最初の i 個のステップにおいて, 少なくとも 1 つのコピーが選択された連結成分の個数を C_i とする。このとき, 完全開示または部分開示にある各頂点 v に対して, $W_i = \sum_v (\deg(v) - 2)$ を定義する。よって, $W_i = X_i + 2Y_i - 2C_i$ となる。実際, 第 i ステップから第 $i+1$ ステップに進むとき, 第 i ステップにおいて連結成分が完成してない場合, 第 $i+1$ ステップで 2 番目に選択されるコピーの頂点 v が部分開示ならば, W_i は変化せず (X_i と Y_i の変化分が打ち消しあう), 未開示ならば, W_i は $\deg(v) - 2$ だけ増加する。一方, 第 i ステップにおいて連結成分が完成する場合, C_i は 1 増加し, 第 $i+1$ ステップで選択されるコピーの頂点が 1 個 (2 つの選択コピーの頂点は同一) または 2 個 (2 つの選択コピーの頂点は異なる) の場合に従ってそれぞれ考えれば, 帰納的に前述の式が成り立つことがわかる。もう 1 つ変数を定義する。第 j 番目の頂点まで開示 (完全または部分のどちらか) 状態にあるとき, これら各頂点 v の次数に対して, $Z_j = \sum_v (\deg(v) - 2)$ を定義する。このとき, 次数 i の未開示頂点が $r_i(j)$ 個あるならば, $Z_{j+1} = Z_j + (i-2)$ となるようにコピーを選択する確率は $ir_i(j) / \sum_i ir_i(j)$ である。

以上の準備のもとに, Molloy and Reed (2000) の定理を考察するが, 本稿で注目する点は,

$Q(\mathbf{D}) > 0$ の場合において形成される、頂点数が大きな連結成分の生成可能性を評価することである。今回は、前述の定理を導くうえで鍵となる変数 Z_j と X_i のオーダーを主として確率的に評価する。

4. 新たな命題の提案

本稿での新たな視点に基づく命題を提案する前に、Molloy and Reed (2000) による定理の導出方法を評価検討しておく。

F の構成アルゴリズムの各ステップにおいてコピーを選択する確率は、それ以前のコピー選択過程に依存するので、条件付き確率として扱う必要がある。しかし、前述の各確率変数に関する条件付き確率を扱うことはかなり困難である。この問題に対して、Molloy and Reed (2000) は最も扱いやすい変数である Z_j に関して、以下のような独立試行の手法を与えている。ただし、 Z_j の定義により選択したコピーをもつ頂点は除外するものとする。

最初に次数 i の頂点のコピーを選択する確率を $p_i^1 = \frac{id_i}{K} = \frac{id_i(n)}{\sum_{j \geq 1} jd_j(n)}$ とする。また、 j 回目 (j 番目の新しい頂点でのコピー選択) における、次数 i の頂点のコピー選択確率を p_i^j (条件付き確率) と記す。そして、独立試行の確率 ψ_i に基づくコピー選択モデルを次のように与える。

次数 i の頂点のコピーを選択する確率：

ある J に対して、任意の $j \leq J$ をとると、 $\psi_i < p_i^j$ ($2 \leq i \leq i'$) かつ $\psi_{i'} > p_{i'}^j$ となるような J (できる限り大きい) および ψ_i を設定する。

こうして Molloy and Reed (2000) では、 p_i^j の代わりに ψ_i を用いて議論を行い、独立試行モデルを各補題の証明の中心に据えている。ここで、条件付き確率に基づく場合または独立試行の確率に基づく場合において、 k 回の試行後、選択されたコピーの頂点次数の総和を表す確率変数を前者の場合 S_k 、後者の場合 S_k^* とすれば、 $S_k = Z_k + 2k$ である ($S_k^* = Z_k^* + 2k$ も同様)。以降、記述を簡潔にするために、 Z_k の代わりに次数の総和 S_k で考える (Z_k^* も同様)。

さて、Molloy and Reed (2000) は以下の不等式が明らかに成り立つものとして、補題 8 の証明に用いている (この考え方は以降の補題でも用いている)。

$$\text{任意の正の整数 } r \text{ に対して, } Pr[S_k \geq r] \geq Pr[S_k^* \geq r]$$

しかし、これに関して 2 つの問題を指摘できる。まず、グラフが 3 種類以上の次数から構成される場合、 S_k^* の確率分布を考えることは理論的に困難と思われる (Chernoff bound による評価は可能であるが)。また、たとえグラフが 2 種類の次数から構成されていても、 S_k の分布を考えることは一般に困難であり、先の不等式は容易に証明できないと思われる。

そこで本稿では前提を単純化して、グラフが次数 1 と次数 l (定数) の頂点から構成される場合を考え、上述の不等式によるのではなく、直接、両者の確率を比較することにより証明を進めるものとする。特別な場合であるが、よく知られている二項分布に関する諸定理を適用でき、一般の場合に関しても、結果をある程度推測できるという利点がある。また、これらの次数の組み合わせをとりあげる理由は以下

のようになる。次数 1 の頂点はコピーを 1 つ持つだけなので、そのコピーが 1 回選択されたら頂点は完全開示になり、次にその頂点を選択されることはない。よって、次数 1 の頂点の占める割合が連結成分の形成に大きな影響を及ぼすことになる。実際、 $Q(\mathbf{D})$ において、負の効果をもつのは次数 1 の頂点だけである。

以下では、グラフが次数 1 と次数 t の頂点から構成される場合について修正命題を与えるが、その前に若干の考察を行う。

まず、 $Q(\mathbf{D}) = -\lambda_1 + t(t-2)\lambda_t > 0$ であるから、 $-\lambda_1 + t(t-2)\lambda_t = \alpha$ とすれば ($\alpha > 0$)、定義により、 $t(t-2)d_t(n) = d_1(n) + \alpha n$ である。 $d_1(n) + d_t(n) = n$ を考えれば、以下を得る (簡単のため整数をとるものとする)。

$$d_1(n) = \frac{t(t-2) - \alpha}{(t-1)^2} n, \quad d_t(n) = \frac{\alpha + 1}{(t-1)^2} n$$

次に、証明では適宜近似をするので、Molloy and Reed (2000) よりも若干条件を強くして、 $Q(\mathbf{D}) = -\lambda_1 + t(t-2)\lambda_t > 2t$ すなわち、 $\alpha > 2t$ としなければならない。このため、 t の範囲を狭めて、 $t \geq 5$ とする。加えて、次数 t の頂点が相対的に多い場合は、大きな連結成分が生成されやすいことはほぼ明白なので、本稿ではそうでない場合に興味がある。よって、 $d_1(n) \geq d_t(n)$ すなわち、 $t(t-2) - 1 \geq 2\alpha$

の場合を対象とする。したがって、 α の範囲は、 $2t < \alpha \leq \frac{t^2 - 2t - 1}{2}$ である。

また、次数 1 の頂点を確率 ψ_1 、次数 t の頂点を確率 ψ_t で選択する k 回の独立試行に対して、 $S_k^* = T_1 + \dots + T_k$ とし、 T_i を i 回目を選択する頂点次数の確率変数とする。さらに、命題 1 では、次の近似評価を用いることにする。

$$0 < x < \frac{\varepsilon}{2} \quad (0 < \varepsilon < 1) \quad \text{に対して,}$$

$$-x > \log(1-x) > -(1+\varepsilon)x \quad \text{および} \quad x > \log(1+x) > (1-\varepsilon)x$$

が成り立つ。

命題 1

ランダムグラフが次数 1 と次数 $t (\geq 5)$ の頂点から構成され、 $t(t-2)d_t(n) = d_1(n) + \alpha n$ が成り立つとする ($d_1(n) \geq d_t(n)$)。このとき、任意の実数 $0 < \delta < \frac{1}{8t^3}$ に対して、 $\varepsilon_1 \delta n < Z_{\lceil \delta n \rceil} < \varepsilon_2 \delta n$ となるようなある正の定数 ε_1 と ε_2 が存在する (*a.e.*)。

証明

ここでは簡単のために δn は正の整数として考える (厳密な証明は、 $\lceil \delta n \rceil - 1 < \delta n \leq \lceil \delta n \rceil$ として評価すればよい)。確率変数 S_k の最小の実現値 k から始めて r を 1 つずつ大きくして、確率 $Pr[S_k = r]$ と $Pr[S_k^* = r]$ に対して、 $Pr[S_k = r] \leq Pr[S_k^* = r]$ となるような r の大きさを評価したい。これらの確率を考えるに当たっては、Molloy and Reed (2000) のアルゴリズムに従い、未開示の頂点のコピーだけをランダムに選択する状況 (すなわち、新しい頂点の選択) を考えればよい。

そこで、初期状態におけるコピーの総数を $m = d_1(n) + td_t(n)$ とする (ここでは以降、 $d_1 = d_1(n)$ および $d_t = d_t(n)$ と記す)。そして、 $0 < \delta < \frac{1}{8t^3}$ に対して、コピーの選択回数を $k = \delta n$ とする。また、選択される次数 1 の頂点コピー数を i_0 、次数 t の頂点コピー数を j_0 とし、 $k = i_0 + j_0$ とする ($n = d_1 + d_t$)。このとき、次数の合計は $r = i_0 + tj_0$ であり、 $r \geq k$ となる。そして選択される頂点の次数列を $(a_1, a_2, \dots, a_k)$ と記す。次数 r が与えられたとき、コピーを選択する確率は $Pr[S_k = r]$ であり、合計が r となる任意の次数列 $(a_1, a_2, \dots, a_k)$ に対するコピーの選択確率 L_2/L_1 (L_1 はとりうるすべての個数) を考える。 L_2 は頂点の選択順序によらず一定なので、 L_1 を考えると、

$L_1 = m(m - a_1)(m - (a_1 + a_2)) \dots (m - (a_1 + \dots + a_{k-1}))$ となる。このとき、 $a_1 = 1, \dots, a_{i_0} = 1, a_{i_0+1} = t, \dots, a_{k-1} = t$ の場合、任意の $i (1 \leq i \leq k-1)$ に対して、 $a_1 + \dots + a_i$ は最小となるので、 L_1 は最大となる。同様に、 $a_1 = t, \dots, a_{j_0} = t, a_{j_0+1} = 1, \dots, a_{k-1} = 1$ の場合、 L_1 は最小となる。よって、条件付き確率の最大値を P_{max} とすれば、次のようになる。

$$P_{max} = \frac{td_t(td_t - t) \dots (td_t - (j_0 - 1)t) d_1(d_1 - 1) \dots (d_1 - (i_0 - 1))}{m(m-t) \dots (m - (j_0 - 1)t) (m - j_0 t) (m - j_0 t - 1) \dots (m - j_0 t - (i_0 - 1))}$$

以下では、選択回数 k を固定して考える。また、 $r = k$ から始めて r を 1 つずつ大きくし、 i_0 と j_0 がともに整数となるような組を対象にするものとする。また前述の不等式を利用する際は、 $\varepsilon = \frac{1}{t^2}$ とする。

まず、 S_k^* の確率が $Pr[S_k = r] \leq Pr[S_k^* = r]$ となるような r の範囲 $k \leq r \leq r_1$ を求める。このために、

$$\psi_t = \frac{td_t - a}{m - a}, \quad \psi_1 = \frac{d_1}{m - a} \text{ とし、 } a = 2 \text{ とする。このとき、 } r \text{ が } k \text{ に近いならば、 } Pr[S_k = r] \leq Pr[S_k^* = r]$$

となることが予想されるので、 $i_0 \geq j_0$ とする。そして、 $Pr[S_k^* = r] = \binom{k}{i_0} \psi_1^{i_0} \psi_t^{j_0}$ および $\binom{k}{i_0} P_{max} > Pr[S_k = r]$ となるから、 $\psi_1^{i_0} \psi_t^{j_0} \geq P_{max}$ となるための i_0 と j_0 の満たすべき関係を求める。 δ

は $0 < \delta < \frac{1}{8t^3}$ なので、 $\frac{kt}{m} = \delta t \frac{n}{m} < \frac{1}{8t^2}$ である。また、 $d_t > \frac{n}{t}$ となるから、 $\frac{j_0}{d_t} < \frac{kt}{n} = \delta t < \frac{1}{8t^2}$ で

あり、 $d_1 \geq \frac{n}{2}$ より、 $\frac{kt}{d_1} \leq 2\delta t < \frac{1}{4t^2}$ である。さらに、 $\log \frac{1 - \frac{jt}{m}}{1 - \frac{j}{d_t}} > 0$ である。以上のことを考慮すれば、

$$\begin{aligned} \log \psi_1^{i_0} \psi_t^{j_0} &= i_0 \log d_1 + j_0 \log td_t - k \log m + j_0 \log \left(1 - \frac{a}{td_t}\right) - k \log \left(1 - \frac{a}{m}\right) \\ \log P_{max} &= \sum_{j=0}^{j_0-1} \log \frac{td_t - jt}{m - jt} + \sum_{i=0}^{i_0-1} \log \frac{d_1 - i}{m - j_0 t - i} = i_0 \log d_1 + j_0 \log td_t + \sum_{i=1}^{i_0-1} \log \left(1 - \frac{i}{d_1}\right) \\ &+ \sum_{j=1}^{j_0-1} \log \left(1 - \frac{j}{d_t}\right) - \left(k \log m + \sum_{j=1}^{j_0-1} \log \left(1 - \frac{jt}{m}\right) + \sum_{i=1}^{i_0-1} \log \left(1 - \frac{j_0 t + i}{m}\right) + \log \left(1 - \frac{j_0 t}{m}\right)\right) \end{aligned}$$

であり、以下のようになる。

$$\begin{aligned}
 & \log \psi_1^{i_0} \psi_t^{j_0} - \log P_{\max} \\
 &= -i_0 \log \left(1 - \frac{a}{m}\right) - \sum_{i=1}^{i_0-1} \log \left(1 - \frac{i}{d_1}\right) + \sum_{i=1}^{i_0-1} \log \left(1 - \frac{j_0 t + i}{m}\right) + \log \left(1 - \frac{j_0 t}{m}\right) \\
 & \quad + j_0 \left(\log \left(1 - \frac{a}{td_t}\right) - \log \left(1 - \frac{a}{m}\right) \right) - \sum_{j=1}^{j_0-1} \log \left(1 - \frac{j}{d_t}\right) + \sum_{j=1}^{j_0-1} \log \left(1 - \frac{tj}{m}\right) \\
 &> i_0 \frac{a}{m} + \sum_{i=1}^{i_0-1} \frac{i}{d_1} - (1+\varepsilon) \sum_{i=1}^{i_0-1} \frac{j_0 t + i}{m} + j_0 \left(\log \left(1 - \frac{a}{td_t}\right) - \log \left(1 - \frac{a}{m}\right) \right) + \log \left(1 - \frac{j_0 t}{m}\right) \\
 &> i_0 \frac{a}{m} + \sum_{i=1}^{i_0-1} \frac{i}{d_1} - (1+\varepsilon) \sum_{i=1}^{i_0-1} \frac{j_0 t + i}{m} + i_0 \left(\log \left(1 - \frac{a}{td_t}\right) - \log \left(1 - \frac{a}{m}\right) \right) - \frac{j_0 t(1+\varepsilon)}{m} \\
 &> i_0 \left(\frac{2a}{m} - \frac{(1+\varepsilon)a}{td_t} - \frac{td_t - \varepsilon d_1}{2md_1} + \frac{i_0(td_t - \varepsilon d_1)}{2md_1} - \frac{j_0 t(1+\varepsilon)}{m} \right)
 \end{aligned}$$

ここで、 $\lim_{n \rightarrow \infty} \left(\frac{2a}{m} - \frac{(1+\varepsilon)a}{td_t} - \frac{td_t - \varepsilon d_1}{2md_1} \right)$ は 0 となるから ($i_0 \geq k/2$), $\psi_1^{i_0} \psi_t^{j_0} \geq P_{\max}$ となるには、

以下の式が成り立てばよい。

$$\lim_{n \rightarrow \infty} \left(\frac{i_0(td_t - \varepsilon d_1)}{2md_1} - \frac{j_0 t(1+\varepsilon)}{m} \right) = \sigma \quad (\sigma \text{ はある正の定数})$$

よって、 $k = i_0 + j_0$ および $r = i_0 + tj_0$ により、 $r_1 = k + (t-1)j_0 = k + \frac{(t-1)(td_t - \varepsilon d_1)}{td_t + (2(1+\varepsilon)t - \varepsilon)d_1} k$ ($td_t - \varepsilon d_1 > 0$)

とすれば、

$$\begin{aligned}
 & (t-1)(td_t - \varepsilon d_1) - (td_t + (2(1+\varepsilon)t - \varepsilon)d_1) \\
 &= t(t-2)d_t - (t(2+3\varepsilon) - 2\varepsilon)d_1 \\
 &= an + d_1 - (t(2+3\varepsilon) - 2\varepsilon)d_1 > an - 2td_1 > 0
 \end{aligned}$$

となるから、 $r_1 > 2k$ である。このとき、十分小さな σ をとると、ある i_0 と j_0 の対は $\psi_1^{i_0} \psi_t^{j_0} \geq P_{\max}$ を満たし、 $r = i_0 + tj_0 < r_1$ に対して、 $r > 2k$ である。

ここで、 S_k^* の平均を μ とし分散を σ^2 とすると、任意の定数 $0 < \lambda < 1$ に対して、Chebyshev の不等式を適用し、 $Pr[|S_k^* - \mu| \geq \lambda\mu] \leq \frac{\sigma^2}{\lambda^2 \mu^2}$ となる。 μ と σ^2 はともに n のオーダーなので、十分大きな n に対して、左辺の確率は十分小である。そして、 $\mu > 2k$ となるから、ある正の定数 ε_1 に対して、ほとんどすべての G は、 $(\varepsilon_1 + 2)k < S_k$ となる。

次に、 S_k^* の確率が $Pr[S_k = r] \leq Pr[S_k^* = r]$ となるような r の範囲 $r \geq r_2$ を求める。このために、

$$\psi_t = \frac{td_t}{m-a}, \quad \psi_1 = \frac{d_1 - a}{m-a} \text{ とし、 } a = 2tk \text{ とする。このとき、 } r \text{ が } k \text{ より十分に大きいならば、 } Pr[S_k = r] \leq$$

$Pr[S_k^* = r]$ となると予想されるので、 $j_0 \geq (t+2)i_0$ とする。そこで、 $\psi_1^{i_0} \psi_t^{j_0} \geq P_{\max}$ となるための i_0 と j_0 の満たすべき関係を求める。前述と同様に考えて、

$$\begin{aligned}
& \log \psi_1^{i_0} \psi_t^{j_0} - \log P_{\max} \\
&= -k \log \left(1 - \frac{a}{m}\right) + i_0 \log \left(1 - \frac{a}{d_1}\right) + \sum_{j=1}^{j_0-1} \log \frac{1 - \frac{jt}{m}}{1 - \frac{j}{d_1}} + \sum_{i=1}^{i_0-1} \log \frac{1 - \frac{j_0 t + i}{m}}{1 - \frac{i}{d_1}} + \log \left(1 - \frac{j_0 t}{m}\right) \\
&> a \left(\frac{k}{m} - \frac{(1+\varepsilon)i_0}{d_1}\right) + \sum_{i=1}^{i_0-1} \log \frac{1 - \frac{j_0 t + i}{m}}{1 - \frac{i}{d_1}} + \log \left(1 - \frac{j_0 t}{m}\right) \\
&> 2ti_0 \left(\frac{k}{m} - \frac{(1+\varepsilon)i_0}{d_1}\right) + i_0 \left(\frac{i_0(td_t - \varepsilon d_1)}{2md_1} - \frac{td_t - \varepsilon d_1}{2md_1} - \frac{j_0 t(1+\varepsilon)}{m}\right) \\
&> \frac{i_0}{2md_1} (4ktd_1 - 4(1+\varepsilon)ti_0m + i_0(td_t - \varepsilon d_1) - 2j_0t(1+\varepsilon)d_1 - (td_t - \varepsilon d_1))
\end{aligned}$$

上の式の展開では、 $j_0 \geq (t+2)i_0$ によって、 $\frac{k}{m} - \frac{(1+\varepsilon)i_0}{d_1} = \frac{j_0 d_1 - (1+\varepsilon)i_0 td_t - \varepsilon i_0 d_1}{md_1} > 0$ であることに注意する。また、 $td_t - \varepsilon d_1 > 0$ によって、 $i_0(td_t - \varepsilon d_1) \geq td_t - \varepsilon d_1$ である ($i_0 = 0$ の場合は明らかに主張が成り立つ)。よって、 $4ktd_1 - 4(1+\varepsilon)ti_0m - 2j_0t(1+\varepsilon)d_1 \geq 0$ となれば、 $\psi_1^{i_0} \psi_t^{j_0} \geq P_{\max}$ である。この式が成り立つような i_0 と j_0 の関係は、

$$2j_0t(1+\varepsilon)d_1 \geq 4(1+\varepsilon)ti_0m - 4ti_0d_1 \text{ すなわち、 } (1-\varepsilon)d_1j_0 \geq 2((1+\varepsilon)td_t + \varepsilon d_1)i_0 \text{ となる。}$$

この不等式に対して、先の条件 $j_0 \geq (t+2)i_0$ を考慮すると、求める関係式は、 $j_0 \geq (t+2)i_0$ となる。

このとき、 $k = i_0 + j_0$ および $r = i_0 + tj_0$ により、 $r_2 = k + (t-1)j_0 = k + \frac{(t-1)(t+2)}{t+3}k$ とすれば、 $r \geq r_2$

において $\psi_1^{i_0} \psi_t^{j_0} \geq P_{\max}$ となる。また、 $r_2 > 2k$ である。前と同様に考えれば、ある正の定数 ε_2 に対して、ほとんどすべての G は、 $S_k < (\varepsilon_2 + 2)k$ となる。以上まとめると、 $Z_k = S_k - 2k$ により、 $\varepsilon_1 k < Z_k < \varepsilon_2 k$ となる。□

次の命題に取り掛かる前に、もう 1 つの変数 I_j を設定する。これは、 j 番目の頂点が部分開示 (完全開示も含む) な状態になるまでに開示されるマッチングの数、すなわちステップ数を表す。よって、 $W_{I_j} = Z_j$ である。

さて、Molloy and Reed (2000) の補題 9 は前述のように独立試行の確率に基づくコピー選択モデルを用いているので、本稿ではこれを修正した証明方法に基づく命題を提案する。なお前の補題と同じ記号を用いる。

命題 2

ある定数 $\varepsilon (0 < \varepsilon < 1)$ に対して、 $\delta' = \frac{K}{((\log n)^\varepsilon + 1)(3 + \varepsilon_2)}$ とする。このとき、任意の実数

$0 < \delta < \delta'$ に対して、 $X_i \geq \frac{\varepsilon_1 \delta n}{2}$ となるようなある $1 \leq i \leq I_{\lceil \delta n \rceil}$ が存在する (a.e.)。

証明

以前と同様に, δn は正の整数とする。 n は固定して考える。 1 より小さい任意の定数 $\sigma > 0$ に対して, $\gamma = \frac{\varepsilon_1 \delta}{2} + \sigma$ とする。 また, $I_{\delta n}$ 個のコピーの組がマッチングされるまでに生成されるバックエッジの個数を W とする。 そこで, 次のことを示す。

任意の $0 < \delta < \delta'$ に対して, $W < \frac{\gamma}{2} n$ (a.e.) となるような δ' をとれる。

部分開示な状態になる第 j 番目の頂点に対して, $j \leq \delta n$ として, 以下では, $1 \leq i \leq I_{\delta n}$ の任意のステップ i を対象とする。 $X_i \leq \frac{\varepsilon_1 \delta n}{2}$ の場合を考えればよい。

前述の定義から, $Z_j = W_{I_j} = X_{I_j} + 2Y_{I_j} - 2C_{I_j}$ である。 よって,

$$I_j = j + Y_{I_j} \leq j + \frac{Z_j - X_{I_j} + 2C_{I_j}}{2} \leq j + \frac{Z_j}{2} + C_{I_j} \leq \left(\frac{3\delta}{2} + \frac{\varepsilon_2 \delta}{2} \right) n \quad \text{よって,}$$

$I_{\delta n} \leq \left(\frac{3\delta}{2} + \frac{\varepsilon_2 \delta}{2} \right) n$ となる。 ステップ i においてオープンな頂点コピーが選択される (すなわち,

あるバックエッジが生成される) 確率は, $\frac{X_i}{Kn-2i} + o(1)$ である。 そこで, 確率 p を

$\frac{X_i}{Kn-2i} + o(1) \leq \frac{X_i}{Kn-2I_{\delta n}} < \frac{\gamma}{K-2I_{\delta n}/n} = p$ と設定する。 そして, ステップ $I_{\delta n}$ までに生成されるバックエッジの期待値を $E(W)$ とする。 ここで, ステップ i においてバックエッジが生成されれば 1 , されないときは 0 をとる確率変数を B_i とする。 このとき, $W = B_1 + \dots + B_{I_{\delta n}}$ である。 そこで, Markov の不等式を適用して以下を得る。

$$\Pr \left(W \geq \frac{\gamma}{2} n \right) \leq \frac{2}{\gamma n} E(W) < \frac{2}{\gamma n} p I_{\delta n} = \frac{2I_{\delta n}}{Kn - 2I_{\delta n}}$$

$\lim_{n \rightarrow \infty} \Pr \left(W < \frac{\gamma}{2} n \right) = 1$ となる (すなわち, $W < \frac{\gamma}{2} n$ (a.e.)) ためには, 以下の不等式が成り立てばよい。

$$\text{ある定数 } \varepsilon (0 < \varepsilon < 1) \text{ に対して, } \frac{2I_{\delta n}}{Kn - 2I_{\delta n}} < \frac{1}{(\log n)^\varepsilon}$$

したがって, $2((\log n)^\varepsilon + 1)I_{\delta n} \leq ((\log n)^\varepsilon + 1)(3\delta + \varepsilon_2 \delta)n < Kn$ となればよい。 よって,

$$\delta < \frac{K}{((\log n)^\varepsilon + 1)(3 + \varepsilon_2)} \text{ となるから, } \delta' = \frac{K}{((\log n)^\varepsilon + 1)(3 + \varepsilon_2)} \text{ である。 } W = Y_{I_{\delta n}} \text{ を考えれば,}$$

$$X_{I_{\delta n}} \geq Z_{\delta n} - 2Y_{I_{\delta n}} > \varepsilon_1 \delta n - \gamma n = \left(\frac{\varepsilon_1 \delta}{2} - \sigma \right) n \text{ であり, } X_{I_{\delta n}} \geq \frac{\varepsilon_1 \delta n}{2} \text{ を得る。 } \square$$

得られる X_i のオーダーは $\Theta\left(\frac{n}{(\log n)^\varepsilon}\right)$ であるが、これはステップ数が異なることによるものである。

本稿のステップ数に対する X_i の比率は $\Theta(n)$ であり、これは Molloy and Reed (2000) の補題 9 と一致する。しかし、これは明らかに改善すべき問題である。

Molloy and Reed (2000) の補題 10 ついても同様に、独立試行の確率に基づくコピー選択モデルを用いており、さらにオープンコピーがバックエッジを形成する可能性を考慮せずに閉路の数を導いているので、本稿ではさしあたり連結成分の頂点数に関して、これを修正した証明方法に基づく命題を提案する。

命題 3

$\delta' = \frac{K}{((\log n)^\varepsilon + 1)(3 + \varepsilon_2)}$ とする。このときステップ $I_{\lceil \delta' n \rceil}$ において、ある ζ に対して、頂点数が

$\zeta \frac{n}{(\log n)^\varepsilon}$ であるような連結成分が存在する (a.e.)。

証明

オープンコピーをもつ頂点 (すなわち、部分開示) は 1 つの連結成分上にある。この頂点数を a とし、完全開示の状態にある頂点数を b とする。このとき、

$a + b = \delta' n$ であり、完全開示にある頂点において、1 つのバックエッジ当たり高々 2 個の頂点が対応するから、 $b < 2Y_{I_{\delta' n}} < \gamma n$ となる。よって、 $a \geq \left(1 - \frac{\varepsilon_1}{2}\right) \delta' n$ を得る。□

この結果も命題 2 と同様の評価になるが、 $Q(\mathbf{D})$ が有限の場合には、Molloy and Reed (2000) では、連結成分のオーダーは少なくとも $\log n$ となるので、本稿の方がより改善された結果を得ている。

5. 次回 II の展望と課題

今回 $Q(\mathbf{D})$ に関する条件を強くしたために評価できなかった、次数 $t=2, 3, 4$ の場合を検討しなければならない。また、生成される連結成分の大きさを再評価し、それに基づいて閉路の大きさを評価する必要がある。

参考文献

- Molloy, M. and Reed, B. "A Critical Point for Random Graphs with a Given Degree Sequence", *citeseerx.ist.psu.edu* (2000)
- Molloy, M and Reed, B, "The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence", *Combinatorics, Probability and Computing* 7, 295-306 (1998)
- Molloy, M and Reed, B, "A Critical Point for Random Graphs with a Given Degree Sequence", *Random Structures and Algorithms* 6, 161-180 (1995)