

# 世界の大きさ 世界は意外と狭い

## The size of the world —It is a small world—

荒関 仁志

Araseki Hitoshi

### 概要

Can you believe that everyone is at most six steps away from any other person on the Earth? This phenomenon, which is called "small world phenomenon", was found by social psychologist S. Milgram in 1967. However, we could not recognize that it was important network structure until to prove by D.J.Watz and S.H.Strogatz in 1998. Recently, it has been understood that the small world phenomenon is a general network structure in various fields. In this paper, we discuss the small world network by simple mathematics.

キーワード： small world, scale free, random graph

## 1 はじめに

皆さんは、人と人とのつながりがりの距離がどのくらいか知っているでしょうか。ここでは、先ず「ケヴィン・ベーコンゲーム」というゲームを使って、人と人との距離について説明します。

ケヴィン・ベーコンは、ちょっと地味で個性的な俳優で、映画『激流』などに出演しています。なぜ地味なケビン・ベーコンに注目するかと言うと、ケビン・ベーコンは多くの映画に出演している名脇役なので、多くの俳優と共演していることが予想できるからです。ここでいう「ケヴィン・ベーコンゲーム」とは、全米の俳優の全てと彼とのある距離を調べた研究に基づいているものです。

ある距離とは以下のように定義されているものです。ここでこの距離のことを「ケビンベーコン指数： $k$ 」と呼びます。

$k = 0$ : ケヴィン・ベーコン本人です。

$k = 1$ : ケヴィン・ベーコン本人と競演した俳優。例えば、ショーン・ペン (Sean Penn) は「ミスティック・リバー (Mystic River)」で競演しました。

$k = 2$ : ケヴィン・ベーコンと競演した俳優と競演した俳優です。例えば、ウディ・アレン (Woody Allen) はショーン・ペンと「ギター弾きの恋 (Sweet and Lowdown)」で競演しました。

$k = 3$ : ...

では、あなたの知っている俳優のケヴィン・ベーコン指数はいくつでしょうか。例えば、渡辺謙はどうでしょうか。答えは  $k = 2$  です。これは 2003 年公開の映画「The Last Samurai」の Tom Cruise との共演が大きな役割を果たしているようです。ケビンベーコン指数は、以下のホームページで計算することができます [1]。

<http://oracleofbacon.org/>

ここでは、アメリカ映画に出演したほぼすべての俳優が登録されているようです。ちなみに、このホームページの「Oracle」とは、世界最大のデータベースの会社です。

例えば、渡辺謙を入力すると、以下のような「Ken Watanabe (I) has a Bacon number of 2.」と表示されます。

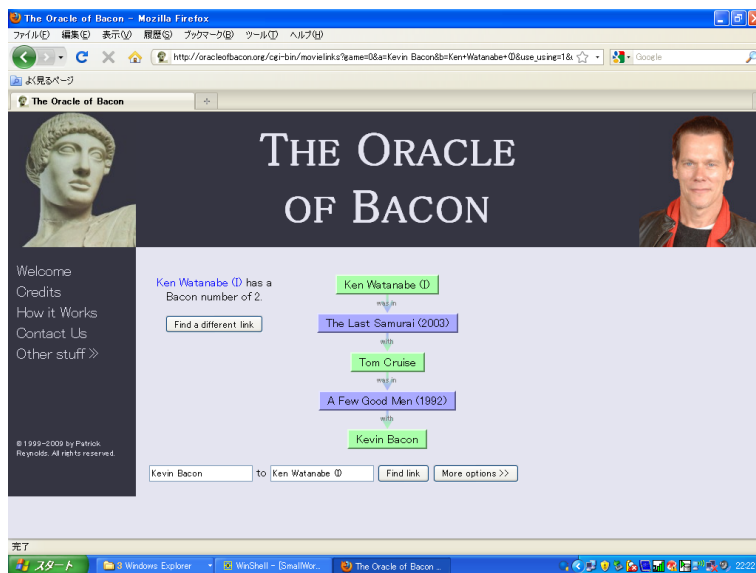


図 1: 渡辺謙のケビンベーコン指数

先ほど、全米の俳優の全てとケヴィン・ベーコンとのある距離（指数）を調べた研究に基づいていると述べましたが、その結果はほぼ全ての俳優が、ケヴィン・ベーコン指数  $6 (k \leq 6)$  以内であることが報告されています。この結果は驚きです。全米にどれだけの俳優がいるかは知りませんが、数万人以上だと思われるので、それらの俳優達がたった  $6$  (ケヴィン・ベーコン指数  $6$ ) しか離れていないということなのです。

## 2 スモールワールド

そこで、世界の様々なものの関係を調べてみると、このケヴィン・ベーコンゲームのようになっているということが分かってきました。意外と指数  $6$  以内で世界の多くのものが関係しているということです。意外と世界は狭いですね。これを「スモールワールド現象 (small world phenomenon, small world effect)」と呼び、20 世紀後半から数理的に盛んに研究されるよう

になって来ました。

この研究の発端は、社会心理学者スタンレー・ミルグラム (Stanley Milgram) が 1967 年に行ったスモールワールド実験 (small world experiment) に基づいています。この実験ではアメリカ合衆国の国民から無作為に二人を選び、一人にもう一方の名前と職業だけを教えて、知り合いのネットワークを使って手紙を届けるという実験を行いました。この実験の結果、平均すると 6 人の知り合いを介して、その二人が繋がっていることが実験的に検証されました。その後この仮説をもとに 6 次の隔たり (Six Degrees of Separation) というフレーズが生まれました。

この「スモールワールド」現象が発見されてから、数学的にランダムグラフ (ここではエルデシュ (P. Erdős) とレニイー (A. Rényi) によるランダムグラフ [6]) を使って、この「スモールワールド」を説明しようと試みられましたがうまく説明ができませんでした。どうも「ランダムグラフ」と「スモールワールド」とは構造が異なっているようです。この「ランダムグラフ」とは、簡単に言うと、たくさんのボタンの中からある確率でランダムにボタンを 2 つ選び出し、それらをつなぐことに対応します。このランダムグラフは、数理的には取り扱いが比較的簡単なグラフモデルなので、数理的解析を行う場合にはよく利用されます。また、構造も比較的一般的な (特殊ではない) ものなので、その点においても解析に有効なグラフです。

この「スモールワールド」を数学的に解明したのが、1998 年に発表されたワッツとストロガッツ (Duncan J. Watts and Stevn H. Strogatz) の論文 [3] です。

### 3 世界を小さくしている犯人

ここでは先ず「ランダムグラフの特徴」について説明します [2]。詳しい計算方法は付録 A を参考にしてください。ここでは、グラフの特徴量の説明をします。先ず、次数 ( $k$ ) とは 1 つのボタンにつながる系の数の分布です。また、平均距離 ( $L$ ) とはあるボタンから他のボタンまでの距離の平均で、クラスター係数とは、1 つのボタンにつながっている他の 2 つのボタンが、お互いにつながっている数のことです。(簡単に言うと、自分の友達同士が友達である確率です。詳細は付録 A を参照してください)。平均クラスター係数とはそのクラスター係数の平均を表しています。

「ランダムグラフ」では、グラフを表す特徴量は以下のようにになっています。

- 次数の確率分布は「ポアソン分布」:

$$P(k) \sim \frac{\bar{k}^k e^{-\bar{k}}}{k!}$$

- 平均距離:  $L \sim \log n$
- 平均クラスター係数:  $C \sim \frac{1}{n}$

ここで大集団 (膨大なボタン数) を考えると  $n \gg 1$  となるので、

- 平均距離:  $L \rightarrow$  大

- 平均クラスター係数:  $C \rightarrow 0$

となります。

この「ランダムグラフ」の特徴に対して、「スモールワールド」の研究で分かったことは

- 次数の確率分布は「スケールフリー (Free-scale)」:  $P(k) \sim k^{-\nu}, 2 < \nu < 3$
- 平均距離: 比較的小さい  $L$
- 平均クラスター係数: 大きい  $C$

となっていて「ランダムグラフ」では説明できないことが分かりました。

ちなみに、「スケールフリー」のような次数分布では、図2のポアソン分布のような、分布の偏りを特徴付ける平均的な尺度（スケール：平均や分散など）といったものが存在しません。分布がこのような性質を持つことを「スケールフリー」と呼びます。また、このような確率分布のときの分散は無限大となります。

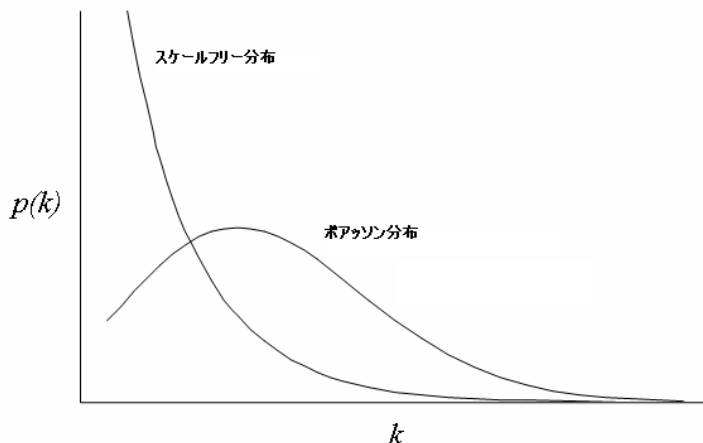


図 2: ポアソン分布とスケールフリー分布

この「平均距離が小さく」て「平均クラスター係数が大きい」とはどのような構造なのでしょう。

- 平均距離が小さいとは、 $n \gg 1$  でも、全ての点（ボタンや個人）が比較的小さい距離でつながっていることを表します。
- 平均クラスター係数が大きいとは、全集団が比較的大きなクラスター（集団）に分離されていることを表しています。

では、どのような構造が世界を小さくしているのでしょうか。ワッツとストロガッツは図3のような計算機実験を行いました。この計算機実験では、ランダムネス ( $p$ ) を変化させること

で、「Regular(規則的)」から「Small-world (スモールワールド)」を経て「Random (ランダムグラフ)」へ移行する様子が描かれています。これはすべての点(ボタンに対応)を規則的(全ての最近接点と2つ隣の点)に結んだ「規則的なグラフ ( $p = 0$ )」から、次に、結んだ糸の一部をはずして適当(ランダム)な点を結んでいきます。それを繰り返すことで、最終的には全ての点がランダムに結合された「ランダムグラフ ( $p = 1$ )」が作られる様子を示しています。

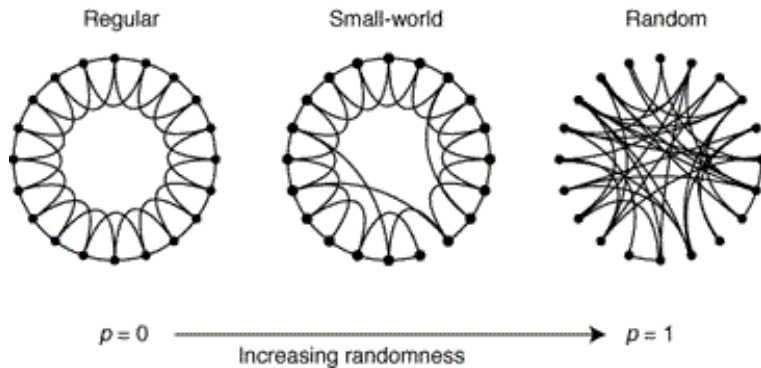


図 3: ランダムネス ( $p$ ) の変化によるスモールワールド from Nature[3]

そこで、これらのグラフを基にランダムネス ( $p$ ) を指標として、「平均距離:  $L(p)$ 」と「平均クラスター係数:  $C(p)$ 」を計算したのが図 4 です。このランダムネス  $p$  は、次数  $k$  と同じものだと考えてください。ただし、図 4 では、表示を確率で表すために規格化していますが、本質的ではないので無視してください(図 4 では  $L(0)$  や  $C(0)$  で割ることで規格化を行っています)。ここで分かることは、ランダムネス ( $p$ ) が中間(「スモールワールド」の付近)で、

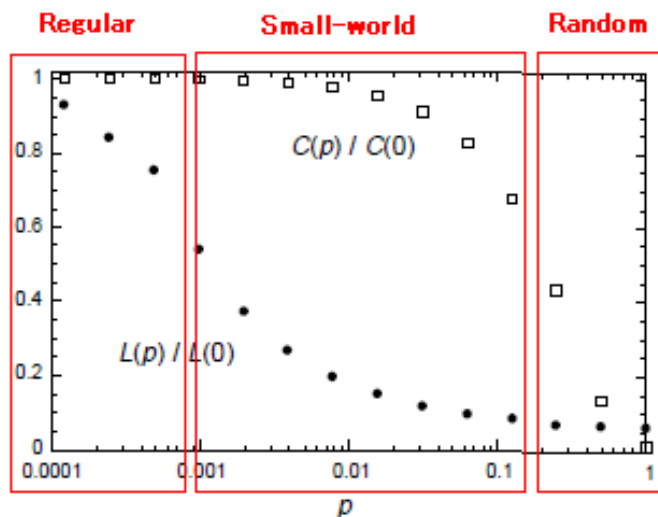


図 4: 規格化された平均距離と平均クラスター係数 from Nature[3]

小さな  $L(p)$  と大きな  $C(p)$  を持つことが分かります。これは、世界が比較的サイズが大きなグループに分割されていて、グループとグループの間の距離は比較的短い「スモールワールド」が実現していることを示しています。このことから「スモールワールド」は「ランダムグラフ」に比べて「疎なグラフ構造」といわれています。

その後の研究で、このような「スモールワールド」的世界（小さな  $L$  と大きな  $C$ ）は、我々の社会には比較的沢山あることが分かってきました。

ワッツとストロガッツの研究から、このような「スモールワールド」は、比較的大きい集団（国家や会社、家庭などの結合が強い集団）とそれら集団間を結ぶ関係（HUB と呼ばれています）からできていることが分かりました。

## 4 スケールフリーな分布

ワッツとストロガッツの研究では、クラスター係数 ( $C$ ) と距離 ( $L$ ) を説明することはできませんが、残念ながら次数の確率分布の「スケールフリー性」を説明することはできませんでした。この「スケールフリー性」を説明したのが、1999年に発表されたアルバート・ラズロ・バラバシとレカ・アルバート (A.-L. Barabási and R. Albert) の研究です [7, 8]。この研究では、ネットワークを形成する際に、次の操作を加えました。

ネットワークが時間とともに成長することを考えます。ワッツとストロガッツのネットワークのようにつながるノード（ボタンの数）の数が固定されているのではなく、時間と共に増えるということを考えます。また、新しいノードは、優先的に次数の多いノードにつながるようにします（優先的選択）。このようなネットワークを考えると、「スケールフリー性」が現れることを発見しました。

このことは、ひとたび次数が大きくなったノードは、時間と共にさらに次数が増える事になります。このことは我々の社会がスモールワールド的だとすると、富めるものが増々富むという結果になると解釈することができます。この富んでいるノードが「HUB」に対応します。ちなみに、会社では2割の社員が働いて、残りの8割の社員が遊んでいるという「2割8割の法則（パレートの法則）」も、このバラバシのモデルで説明ができます。これはまさしく我々の社会構造に対応するものです。

様々な分野での「スモールワールド」の研究から、我々社会は、このHUB構造によって世界中の人たちと比較的短い距離（ペーコン指数）でつながっていることが分かってきました。もっと興味深いことには、最近の研究では、このHUBを形成するのは人間関係においては「親友」ではなく「友達」（比較的疎な関係）だということです。違う言い方をすると、世界を広くするには「密な関係（親友）」ではなく、「比較的疎な関係（友達）」が大きな役割を果たしていることが分かってきました。

これまでの研究で分かってきたのは、「ランダムグラフ (exponential network)」と「スモールワールド (scale-free network)」との違いがはっきりとしてきたことです。これらのネット

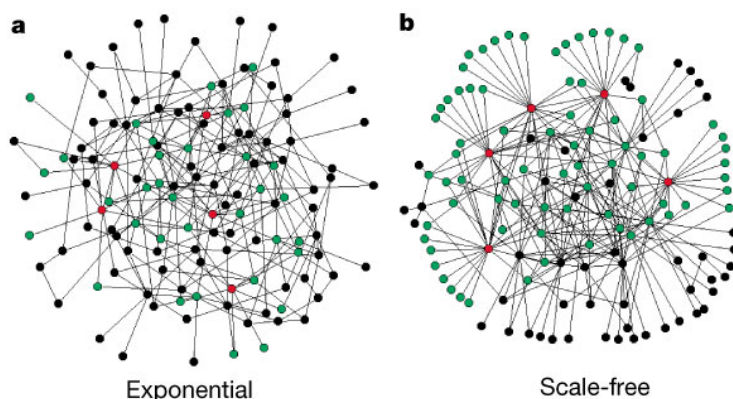


図 5: Visual illustration of the difference between an exponential and a scale-free network. from Nature[4]

ワークを視覚的に分かりやすいように図にしてみると、その構造の違いがはっきりとわかります。図5を見ると、「スケールフリー（スモールワールド）」は、我々の回りにみられる社会構造のように見えませんか。あなたの属している社会は、どちらに属しているでしょうか。興味があったら調べてみてください。

## 5 まとめ

世の中の多くの関係には、スモールワールド（スケールフリー）的關係が成り立っていることが分かってきました。特に、人間関係ではスモールワールドで重要な要素である「HUB」の存在が、情報やウィルスなどの流行や蔓延を左右します。最近の研究では、社会構造やインターネットの構造、学术论文の引用関係、感染症の拡散など、社会の様々なネットワーク構造が、スモールワールド（スケールフリー）的な構造を成していることが分かってきました。このようなスケールフリーなネットワークでは、情報伝播の「要」はHUBに集中しています。ですから感染症の拡大を阻止するためには、HUBとなる人間や環境を集中的に管理することが、感染拡大を防止するのに重要となります。

さらに生物進化の研究から、動物の進化（DNAの進化）や脳細胞の記憶能力にはスモールワールド構造が関係していることが示唆されています [9, 10, 11]。例えば、進化の初期過程では、小さなDNAの断片しかない世界でも、それらが長い時間にわたって、つながったり離れたりすることで、巨大なDNAのクラスター、すなわちスモールワールド的構造が生まれ、それが進化速度を加速するという考えです。同様に、脳細胞も神経細胞がスモールワールド的構造をしていることが分かっており、比較的短い距離で多くの神経細胞とつながっていることが、多くの情報処理を瞬時にいたり、大量の情報を保存することができるのだと考えられています。

この巨大なDNAのクラスターが生成する過程は、物理学で研究が進んでいる「パーコレー

ション現象 [12]」と関係があり、物理学的取り扱いが生物進化のメカニズムに応用ができそうだと期待されています。

この「スモールワールド」の研究は、ネットワークの研究では今一番ホットな研究分野で、社会学や心理学にとどまらず、生物学、物理学、脳科学など様々な分野で研究が始まっています。

## 参考文献

- [1] Tom Siegfried, もっとも美しい数学 ゲーム理論, 文藝春秋, 2008 年
- [2] Rick Durrett, Random Graph Dynamics, Cambridge Univ. Press, 2007
- [3] Duncan J. Watts and Stevn H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393, 440-442 (4 June 1998)
- [4] Reka Albert, Hawoong Jeong and Albert-Laszlo Barabasi, Error and attack tolerance of complex networks, Nature 406, 378-382 (27 July 2000)
- [5] C. Song, S. Havlin and H. A. Makse, Self-similarity of complex networks, Nature 433, pp.392-395 (30 November 2004)
- [6] P.Erdős and A. Rényi, On the evolution of random graph, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5, 17, 1960
- [7] A.-L. Barabási, and R. Albert, Emergence of scaling in random networks, Science 286, pp.509-512 (1999)
- [8] A.-L. Barabási, 新ネットワーク思考—世界の仕組みを解く—, NHK 出版, 2002 年
- [9] Stuart A. Kauffman, The Origins of Order: Self-Organization and Selection in Evolution, Oxford Univ Press, 1993 年 6 月
- [10] Stuart A. Kauffman, 自己組織化と進化の理論, 日本経済新聞社, 1999 年 9 月
- [11] Stuart A. Kauffman, カウフマン、生命と宇宙を語る, 日本経済新聞社, 2002 年 9 月
- [12] 増田 直紀, 今野 紀雄, 複雑ネットワークの科学, 産業図書, 2005 年



## 付録 A ランダムグラフの数学的表現

ここではエルデシュとレニーによるランダムグラフ [6] を数学的に説明します。今、点の総数を  $n$  とすると、点と点を結ぶ線（結線）の可能な組み合わせは、 $n$  個の点から 2 個取り出す組み合わせですから、以下のように計算できます。

$${}_n C_2 = \frac{n!}{2!(n-2)!} = \frac{1}{2}n(n-1) \tag{1}$$

この結線に関して、各々つながる確率を  $p(0 \leq p \leq 1)$ 、つながない確率を  $(1-p)$  とします。

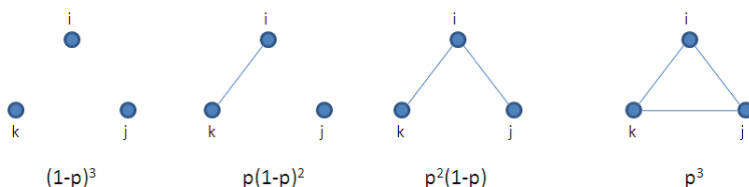


図 6: ランダムグラフ ( $n = 3$ )

ある点の次数（ある点の結線数）が  $k$  である確率  $P(k)$  は、他の  $n-1$  個の点から  $k$  個を選ぶ組み合わせ  $({}_{n-1} C_k)$  と、結ぶ確率  $p^k$ 、残りの  $(n-1-k)$  個の点と結ばない確率  $(1-p)^{n-1-k}$  から、以下のように計算することができます。

$$P(k) = {}_{n-1} C_k p^k (1-p)^{n-1-k} \tag{2}$$

これを使って、すべての点で平均化すると、全体の平均次数  $\bar{k}$  は、以下のように簡単な形に表現することができます。

$$\bar{k} = \sum_{k=0}^{n-1} k P(k) = (n-1)p \tag{3}$$

式 (3) は、式 (2) を使って、以下のように導き出します。

$$\begin{aligned} \bar{k} &= \sum_{k=1}^{n-1} k P(k) = \sum_{k=0}^{n-1} k {}_{n-1} C_k p^k (1-p)^{n-1-k} \\ &= \sum_{k=1}^{n-1} \frac{(n-1)!}{(k-1)!(n-1-k)!} p^k (1-p)^{n-1-k} \\ &= (n-1)p \sum_{m=0}^{n-2} \frac{(n-2)!}{m!(n-2-m)!} p^m (1-p)^{n-2-m} \\ &= (n-1)p(p+1-p)^{n-2} = (n-1)p \end{aligned}$$

ここで点の数を増やす ( $n \gg 1$ ) ことで、式 (2) は以下のように計算することができます。計算の結果、この (二項) 分布はポアソン分布で近似できることが分かります。

$$P(k) = \frac{(n-1)(n-2)\cdots(n-k)}{k!(n-1)^k} \bar{k}^k \left(1 - \frac{\bar{k}}{n-1}\right)^{n-1-k}$$

$$\begin{aligned}
 &= \frac{\bar{k}^k}{k!} \cdot 1 \cdot \left(1 - \frac{1}{n-1}\right) \left(1 - \frac{2}{n-1}\right) \cdots \left(1 - \frac{k-1}{n-1}\right) \left(1 - \frac{\bar{k}}{n-1}\right)^{n-1-k} \\
 &\approx \frac{\bar{k}^k}{k!} \left\{ \lim_{z \rightarrow \infty} \left(1 + \frac{1}{z}\right)^z \right\}^{\bar{k}} \approx \frac{\bar{k}^k e^{-\bar{k}}}{k!}
 \end{aligned}$$

### 付録 A.1 平均距離の計算

この  $\bar{k}$  と  $P(k)$  が計算できると、平均距離  $L$  を計算することができます。この計算は、点の数が  $n$  個の完全  $k$  分木の”木の高さ”  $\log_k n$  を計算することです。この平均距離  $L$  とは、グラフ上の全ての点の間の距離の平均を表します。ここで点の総数  $n$  は、平均次数  $\bar{k}$  を使うと概算で以下のように計算できます。

$$1 + \sum_{z=1}^L \bar{k}^z = \sum_{z=0}^L \bar{k}^z \approx n \tag{4}$$

式(4)の最初の1は出発点を、その後は出発点に接続されている点の数を表し、 $L$ は出発点から最終点(最遠点)までの接続回数(距離)です。この計算は等比数列の和の計算ですから、以下のように計算できます。

$$f(\bar{k}) = \sum_{z=0}^L \bar{k}^z = \frac{(1 - \bar{k}^{L+1})}{(1 - \bar{k})} \tag{5}$$

この式(5)の値が、点の総数  $f(\bar{k}) \approx n$  にほぼ等しいと置くことで、以下のように  $L$  について解くことができます。ここでは、両辺に  $\log$  を演算することで  $L$  について解きます。

$$L \approx \frac{\log n}{\log \bar{k}} + \frac{(\bar{k} - 1)}{\bar{k}} - 1$$

ここで、 $n \gg 1$  とすると、ランダムグラフの平均距離  $L$  は、 $\log n$  で表現することができます。

$$L \sim \frac{\log n}{\log \bar{k}} \sim \log n \tag{6}$$

### 付録 A.2 平均クラスター係数の計算

次に、平均クラスター係数  $C = \frac{1}{n} \sum_{i=1}^n C_i$  を計算してみます。

まず、 $i$  番目の点のクラスター係数  $C_i$  は、 $i$  番目の点(起点)から  $k_i$  個の接線数の中で、お互いがさらに接続されている数を  $n_i$  を使って  $C_i = n_i / k_i C_2$  と定義できます。この係数を簡単に言うと最近接点同士が接続されている確率を表します。多くの書籍では、このクラスター係数のことを自分の友人同士が友人である確率と説明しています。つまり「友達の友達は友達か」ということです。

ここで  $n_i$  は、 $i$  番目の点に接続された点 ( $k_i$  個)の中から2つを選び、それらがお互いに接続されてる確率  $p$  をかけることで計算することができます。

$$n_i = k_i C_2 p$$

このクラスター係数を全ての点で平均したものが平均クラスター係数  $C$  となります。

ただし、同じ点を接続する（重複）ことも許していますが、概算計算なので問題はないと思います。したがって、平均クラスター係数  $C$  は、式 (7) のように計算することができます。

$$C = \frac{1}{n} \sum_{i=1}^n \frac{n_i}{k_i C_2} = \frac{1}{n} \sum_{i=1}^n \frac{k_i C_2 p}{k_i C_2} = p \quad (7)$$

式 (7) と式 (3) から、 $n \gg 1$  の条件を使うと、式 (8) のように計算できます。

$$C = p = \frac{\bar{k}}{(n-1)} \sim \frac{\bar{k}}{n} \sim \frac{1}{n} \quad (8)$$