

Application of network theory to ball games:  
short review of network theory and likelihood ratio test for power-law distribution

橫山慶子\*\*

Keiko YOKOYAMA\*\*

という構造によって機能が異なることを示すことになり、これがグラフ理論の先駆けとなる。

このようにケーニヒスベルク問題は、頂点とその次数

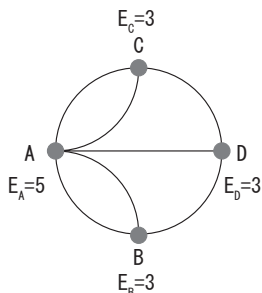


図 1 Königsberg の 7 つの橋問題とオイラーの解法

## 2.1 6次の隔たり：小さな世界

Figure 1 consists of two parts: (a) a line graph and (b) a network diagram.

(a) Line graph titled "Migrant 1967". The x-axis is labeled "最終的なあて先までに介した人数 (人)" (Number of people introduced to the final addressee, in people) and ranges from 0 to 12. The y-axis is labeled "最後までたどり着いた手紙の数" (Number of letters that reached the end) and ranges from 0 to 12. The graph shows a fluctuating trend with a peak of 11 at x=6.

最終的なあて先までに介した人数 (人)	最後までたどり着いた手紙の数
1	0
2	2
3	4
4	9
5	8
6	11
7	0
8	1
9	2
10	2
11	0

(b) Network diagram showing the transmission path from the final person to the first person. The central node is "最終的なあて先 持人 Sharon 叔". It is connected to "Jones" (10), "Jacobs" (16), and "Brown" (5). These three nodes are connected to a set of three intermediate nodes (labeled 2, 2, 2). These intermediate nodes are connected to a set of three nodes (labeled 3, 3, 3). These three nodes are connected to a final set of nodes (labeled 17). The diagram illustrates the flow of information from the final person to the first person.

図2 Milgram (1967) の実験の中でネブラスカ州オマハの差出人から、マサチューセッツ州シャロンに住む株仲間という最終あて先までに要した人数 (a) と、最後の一人となった人の分布 (b)。最後の一人となったうち16通もの手紙を最後に送った人物がいた。

そんな名前呼び合える友人に転送するというものであった。結果は予想よりはるかに少ない知人を通して最終的なあて先まで届いた。ネブラスカの実験では160通のうち44通が最後まで届く。そのとき最も多い経由が2人から10人までで6人が最頻値であった。そして届いた44通のうち16通を最後に送った鍵になる人がいたというものである。

これが「小さな世界」と呼ばれるもので、後に6次の隔たり (six degrees of separation) と呼ばれた (Gladwell, 1999)。つまり、アメリカ国内という広大な物理的空間の中で見ず知らずの2人の距離が、わずか6人でつながるということである。これはネットワークとしてはいわゆるランダムなつながりがあるからであると考えられた。

## 2.2 弱い絆

Granovetter (1973) は、人々がどのようにして現職を得たかを調査した結果、緩やかなつながり、いつも行動を共にするような強い絆で結ばれたグループの人以外の知人、すなわち年に一度会う程度の知人が大きな役割を果たしていることを見出し、それを「弱い絆の強さ (the strength of weak ties)」と呼んだ。つまり、人には親しい友人がいて、その友人同士はほとんどお互いに友人であり、社会構造として緊密な関係にある。その緊密な集団内では同じ情報が共有されており、職を得るといった情報に関してはそれほど緊密ではない知人からの情報が有用となるというものである (図3)。

これは、緊密な関係にある友人同士がクラスターを形成しているが、さらに大きなネットワークになるとこうしたいくつかのクラスターが弱い絆でつながれることによって、クラスター間の距離が短縮することを意味す

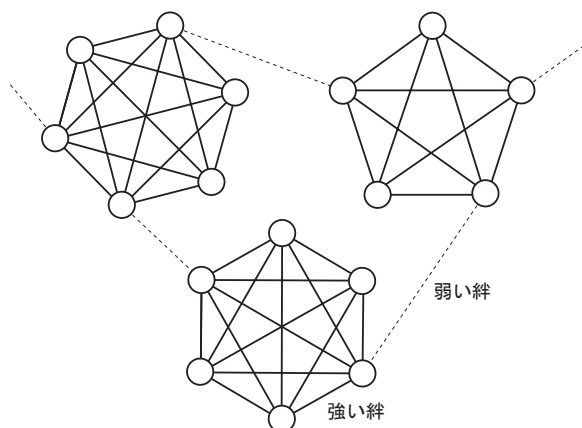


図3 友人同士の緊密な関係 (強い絆) とこれらのクラスター間の弱い絆。この弱い絆が社会構造には重要。

る。ネットワークの観点から言えば、すべての頂点がランダムにつながれているというよりは、いくつかのクラスターが形成され、そのクラスターのいずれかの頂点がある他のクラスターの頂点と結ばれることによって、ある頂点から別の頂点までの距離 (length of path) は短くなるというものである。

## 2.3 グラフ理論の革命：ランダム・ネットワーク

さまざまな秩序あるグラフの性質を探っていたグラフ理論 (graph theory) \*1が、大きな変換点を迎えたのはこのようなグラフあるいはネットワークがいかにして形成されるかという問題にエルデシュ (Paul Erdős (1913–1996)) とレーニイ (Alfréd Rényi (1921–1970)) が取り組んだことによる。彼らはそれまでのグラフ理論では、どの頂点も同じ次数を持つこと (正則グラフ) を仮定していたのに対し、頂点をランダムに結ぶことを繰り返すことによって、多くの頂点からなるつながりあった頂点の集合 (クラスター) ができ、さらにリンクを増やすことによってクラスター同士も結ばれ、ある臨界値に達すると巨大クラスターが出現し、それまでのネットワークとは根本的に異なる、いわゆる相転移やパーコレーション (percolation) と呼ばれる現象がみられることを見出した。

このランダムネットワーク理論の大前提は、徹底した平等主義である。つまり、どの頂点も同じ確率でリンクを獲得するために、十分大きなネットワークになるとそれぞれの頂点はほぼ近似的に同数のリンク、次数を持つことになる。つまり、大部分の頂点は平均的な頂点と同数のリンク、次数をもつということで、この次数分布がポアソン分布 (Poisson distribution) という、突出したピークを持つ分布になることがわかっている (Bollobás, 1981)。

それまでのグラフ理論が扱うネットワークは、完全な秩序を持っているネットワークであり、逆にエルデシュ・レーニイのランダム・ネットワーク理論は、実世界は複雑でその理解のために完全なランダムなネットワークを仮定し、グラフ理論あるいはネットワーク理論に新たな視点を与えたのである。

## 2.4 正則グラフとランダムグラフの狭間：スモールワールド・ネットワーク

6次の隔たりや弱い絆の強さが示すように、現実のネットワークはより複雑であり、これまでグラフ理論が扱ってきた完全に相互が結合した正則グラフ (regular graph) でもなければ、エルデシュ・レーニイが仮定し

\*1 有名な難問に「いかなる地図も、隣接する領域が異なる色になるように塗るには4色あれば十分である」ことを証明せよという問題があった。

た完全にランダムなつながりに基づくランダム・ネットワークでもない。つまり、6次の隔たりが示すように要素間の平均距離は予想以上に短く、と同時に弱い絆が示すようにネットワークにはクラスターが存在し、全てが緊密に結合されているのではなく、クラスター内の緊密な結合とクラスター間の弱い結合が存在する。

Watts and Strogatz (1998) は、この問題に対して、2つの頂点間の最短距離の平均として平均最短経路 (path length) と、ある頂点とつながっている頂点同士もつながっている確率であるクラスター係数 (cluster coefficient) の2つの指標で検討すべきであると考えた。

現実のネットワークを、同じ頂点数と平均度数に基づくランダムグラフとの対比で表したが表1である。これから、平均最短経路はランダムグラフと仮定した場合とそれほど大きな差がないが、クラスター指数においてはランダムグラフの場合の方が明らかにクラスターが小さいことがわかる。このことから、現実のネットワークは正則グラフでもなくランダムグラフでもなく、その中間に位置する、すなわち規則性とランダム性を併せ持ったネットワークを仮定するのが妥当であるとし、それをスモールワールドネットワーク (small-world network) と呼んだ。

これは、 $n$  個の頂点からなる度数  $k$  の  $k$ -正則グラフから出発し、確率  $p$  で、ランダムに辺をつなぎ変えていく。このときに二重のリンクや自分自身への再帰的リンクは禁止するものとする (図4)。すると、図5に示すように、つなぎ変えの確率が高くなると平均最短経路は

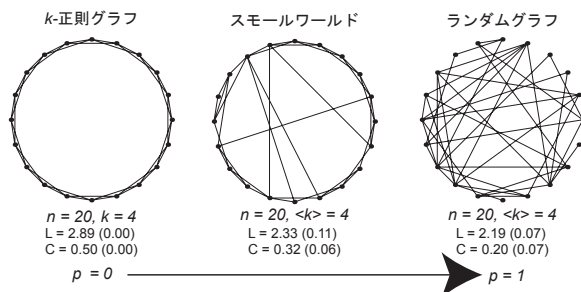


図4  $n = 20, k = 4$  の  $k$ -正則グラフから出発して、確率  $p$  で辺をつなぎ変えていき、 $p = 1$  ではランダムグラフとなる。この例では中央が  $p = 0.2$  でつなぎ変えてある。全て平均度数  $k$  は4である。

表1 現実のネットワークの特徴 (Watts & Strogatz, 1998より)

ネットワーク	N	$\langle k \rangle$	$L_{actual}$	$L_{random}$	$C_{actual}$	$C_{random}$
映画俳優	225,226	61	3.65	2.99	0.79	0.00027
電力網	4,941	2.67	18.7	12.4	0.080	0.005
シー・エレガンス (線虫) の細胞	282	14	2.65	2.25	0.28	0.05

\*2 このサイトでは俳優の名前を入れるとペーコン数が出てくる。松島菜々子は "Poppugurupu koroshiya (2000)" で Scott Show と共演し、Scott Show は "Player, The (1992)" で Andie MacDowell と共演し、Andie MacDowell は "Beauty Shop (2005)" で Kevin Bacon と共演しているためペーコン数が3である。

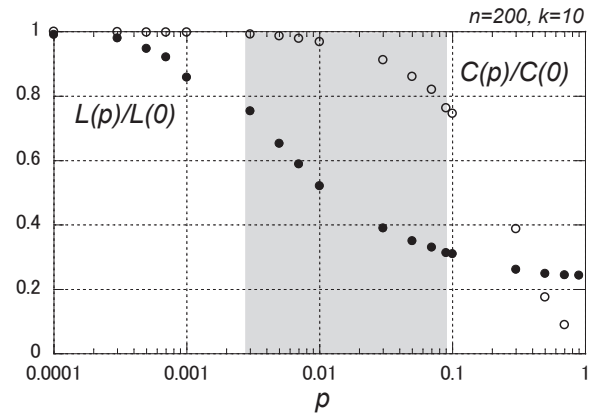


図5  $n = 200; k = 10$  の  $k$ -正則グラフから出発して、確率  $p$  で辺をつなぎ変えていった場合の平均最短経路  $L(p)$  と、クラスター係数  $C(p)$  を、 $p = 0$  を基準に基準化して示してある。中央のグレーの部分が平均最短距離が短くなるが、クラスター係数は大きいスモールワールドの部分である。

急激に減少するが、クラスター指数の減少は見られず、結果的に平均最短経路が小さく、かつクラスター係数が大きい、クラスター化された状態が生じ、現実のネットワークと類似した特徴が見られるというものである。

## 2.5 貴族主義的ネットワーク：スケールフリー・ネットワーク

Watts と Strogatz のモデル (Watts & Strogatz, 1998) は現実のネットワークをよく反映しているように思えたのだが、 $k$ -正則グラフから辺をつなぎ変えていくために、つなぎ変えの確率が高くなるとランダムネットワークに近づき、度数分布も基本的にはランダム・ネットワークモデルと同様である。つまり、大部分の頂点は平均的な頂点と同数のリンクを持つ分布になる。このことは、ミルグラムの実験で最後の一人になった人が16通もの手紙を最後のあて先に送っていたような、いわゆるハブ (hub) と呼ばれるような、大きな度数を持った頂点が数少なくあり、度数が小さい頂点は数多くあるという現実のネットワークとは少し異なる。

こうしたハブの存在はさまざまなネットワークで確認されており、人的ネットワークではランダムネット

ワークを提案した数学者エルデシュとの共著論文からのつながりを表すエルデシュ数や、ハリウッドの映画界におけるケヴィン・ベーコンとの共演数をからそのつながりを表すベーコンゲーム (<http://oracleofbacon.org/>)\*2などが有名である。

Albert, Jeong, and Barabási (1999) は、ノートルダム大学 (nd.edu) のウェブを調べ、325,729のドキュメントと1,469,680のリンクを全て調べたところ、次数分布がベキ則 (power law) に従うことを見出した。つまり、出て行くリンクのベキ指数は  $\gamma_{out} = 2.45$ 、入ってくるリンクでは  $\gamma_{in} = 2.1$  とどちらもベキ則に従っていた。一方、平均最短経路は  $\langle d_{nd.edu} \rangle = 11.2$  で、 $\langle d \rangle = 0.35 + 2.06 \log(N) = 11.6$  という予測値に極めて近く、これは平均最短経路は頂点の対数に比例するということである。したがって、世界中の頂点 (ドキュメント) を  $8 \times 10^8$  とすると、世界中の www の最短平均経路は18.59となり、約19回でどこへでもたどり着けることになるが、これは頂点の数が急激に増えても、その平均最短距離の増加は少なく、小さな世界が維持されることを意味する。また、このことは現実のネットワークがベキ則に従うために、ランダムグラフの次数分布に見られるような平均的な次数を持つ頂点という特徴が見られないことから、スケールフリー\*3ネットワークと呼ばれる。

さらに、Albert, Jeong, and Barabási (2000) は、スモールワールドネットワークとスケールフリーネットワークのランダムに生じるエラーと意図的な攻撃に対する頑健性に関して検討し、スケールフリーネットワークの方がランダムに生じるエラーに関しては頑健で、意図的な攻撃に対する脆弱性は、スモールワールドネットワーク

のランダムなエラーや意図的な攻撃に対する脆弱性と同じであることを見出している。この結果は実際のインターネットや WWW の脆弱性とも一致しており、現実のネットワークの次数分布がベキ則に従うスケールフリーネットワークであることの論拠とされている。これが、平等主義的なランダムネットワークと対比され、貴族主義的ネットワークと呼ばれるゆえんである。

### 2.5.1 スケールフリーネットワークの特徴

現実のネットワークは変化する。WWW は日々その頂点やリンクが増加している。そして、大きなハブ (例えば Google などの検索エンジン) の次数も増加している。こうした現実のネットワークの変化に対して Barabási and Albert (1999) は、成長 (growth) と優先的選択 (preferential attachment) という2つの要素を組み込み、スケールフリーネットワークのモデルを構築した。

成長とは、ある時間単位で新たな頂点に加わり、その頂点は既存の頂点のいずれかと結合することである。そして、その結合の際には、ランダムに結合するのではなく、既存の頂点の次数に依存する (式1) という優先的選択を行うということである。

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i} \quad (1)$$

## 3. ボールゲームのパス行動へのネットワーク理論の適用

ネットワーク理論をボールゲームのパス行動に適用した例 (Yamamoto & Yokoyama, 2011) をもとに、実際のパス行動をネットワークとして分析し、ベキ則が成り立つことを確認する手順を以下に示す。

### 3.1 分析方法と対象

2006年に日本で開催された、サッカーの日本代表が

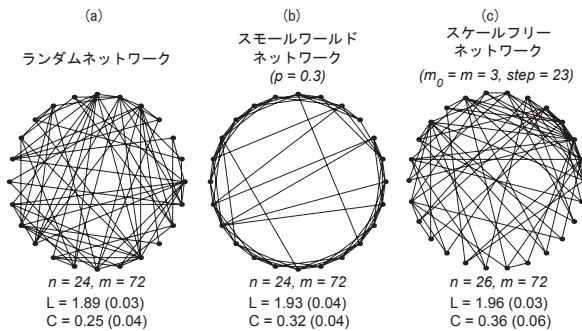


図6 頂点と辺の数がそれぞれ  $n = 24, m = 72$  の (a) ランダムグラフ、(b) スモールワールドと、 $n = 26, m = 72$  の (c) スケールフリーネットワーク。平均最短経路 (L) とクラスター係数 (C) はそれぞれ100回の平均とカッコ内に標準偏差を示してある。ランダムネットワークよりも他の2つはクラスター係数が大きく、またスモールワールドよりもスケールフリーネットワークの方がハブが生じていることが分かる。

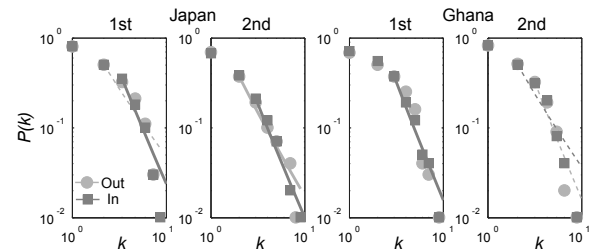


図7 左が前半、右が後半で、横軸にボールに触れた回数、縦軸にその頻度を示してある。

\*3 スケール不変性、あるいは膨張収縮対称性とも呼ばれ、小さな構造も大きな構造も共通した構造を有し、そのことが空間的にはフラクタル (自己相似性) の性質である (松下, 2001; 西山, 2001)



行った、日本対ガーナのゲームを対象にして、パスの出し手と受け手の選手を記録し、5分ごとに各選手がボールに触れた頻度を求めた。なおこの際に、パスが途中でインターセプトされたり、シュートを打つなどパスとして出し手と受け手が特定できないものは除外した。そして、前後半別に、パスを出した回数とパスを受けた回数について、その回数とその頻度（確率）を求め両対数プロットを行った。

### 3.2 分析結果

次数分布を各チームごとにパスを出した回数と受けた回数として前後半別に見たものが、図7である。

これは5分ごとにパスを出した数、受けた数を横軸にとり、その人数の割合を縦軸にとり、両対数グラフで表したものである。この傾きが直線になれば、ベキ則が成り立つことになる。

### 3.3 ベキ則への当てはまりの良さと対立仮説との比較

ここでは Clauset, Shalizi, and Newman (2009) の提唱する方法により、ベキ則への当てはまりの良さ、および対立仮説との比較をし、実際に得られた次数分布においてベキ則が成り立つか否かを検討する以下の三段階の手続きを踏んで検証することを示す。

なおこの手順を踏むためのプログラム群も提供されており\*4、ここではその中で MATLAB と R のプログラムを利用した。

■手順1 パラメータ  $x_{min}$  を変更しながら、ベキ則の  $\alpha$  を最尤推定法 (maximum likelihood estimator: MLE) を用

いて以下の式で推定する。

連続データの場合：

$$\hat{\alpha} = 1 + n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1} \quad (2)$$

離散データの場合：

$$\hat{\alpha} \simeq 1 + n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{min} - \frac{1}{2}} \right]^{-1} \quad (3)$$

なおこの際に  $x_{min}$  は、次式で示す Kolmogorov-Smirnov 検定を用いて、 $D$  が最も小さくなるように  $x_{min}$  を推定する。

$$D = \max_{x \geq x_{min}} |S(x) - P(x)| \quad (4)$$

ここで  $S(x)$  は観測値の累積分布関数 (cumulative distribution function) で、 $P(x)$  はベキ則モデルに当てはめた際の累積分布関数である。

ここまでの手順を、`plfit_yy.m`\*5を用いて、ベキ則が成り立たない少数の部分であるパラメータ  $x_{min}$  を Kolmogorov-Smirnov 検定を用いて推定し、スケーリング指数  $\alpha$  を求めた。その後、`plvar.m`によって、推定されたパラメータ  $x_{min}$  とベキ則分布が成り立っているところのデータ数 ( $n_{tail}$ ) の標準偏差を求めた。

■手順2 元データとベキ則の間の適合度 (goodness-of-fit) を Kolmogorov-Smirnov 検定において計算する。この場合に  $p$  値が0.1以上であるならば、ベキ則は採択される仮説であり、それ以外は棄却される。

表2 サッカーのパス行動におけるベキ則への適合度検定の結果。各列はパス総回数 ( $n$ )、平均パス数 ( $\langle x \rangle$ )、標準偏差 ( $\sigma$ )、and 最高パス数 ( $x_{max}$ )、ベキ則が成り立つ最小値 ( $x_{min}$ )、ベキ則が成り立つ範囲でのスケーリング指数 ( $\langle \alpha \rangle$ )、ベキ則が成り立つ範囲でのパス数 ( $n_{tail}$ )、適合度の確率 ( $p$ )。

			$n$	$\langle x \rangle$	$\sigma$	$x_{max}$	$x_{min}$	$\langle \alpha \rangle$	$n_{tail}$	$p$
Japan	1st	Out	197	17.9	5.05	24	$2 \pm 0.75$	2.50	$50 \pm 18.2$	0.00
		In	197	17.9	4.60	27	$3 \pm 0.58$	3.50	$35 \pm 12.9$	<b>0.45</b>
	2nd	Out	151	13.7	6.55	27	$2 \pm 0.60$	2.70	$37 \pm 13.3$	<b>0.21</b>
		In	151	13.7	7.62	30	$3 \pm 0.71$	3.32	$21 \pm 14.1$	<b>0.19</b>
Ghana	1st	Out	205	18.6	11.61	44	$4 \pm 0.87$	3.50	$25 \pm 11.3$	0.03
		In	205	18.6	10.88	40	$3 \pm 0.40$	3.41	$37 \pm 7.0$	<b>0.39</b>
	2nd	Out	197	17.9	8.68	36	$3 \pm 0.75$	3.50	$32 \pm 17.0$	0.04
		In	197	17.9	9.40	35	$2 \pm 0.93$	2.55	$50 \pm 19.8$	0.00

\*4 <http://tuvalu.santafe.edu/~aaronc/powerlaws/>,

[http://intersci.ss.uci.edu/wiki/index.php/Power-law\\_distributions](http://intersci.ss.uci.edu/wiki/index.php/Power-law_distributions)

\*5 配布されている `plfit.m` にベキ則分布が成り立っているところのデータ数 ( $n_{tail}$ ) を戻すように修正した。

\*6 ここでのいくつかのプログラムはC言語で書かれたルーチンを利用しており、さらにこのC言語ではGSL (GNU Scientific Library) が必要となる。

求めたパラメータ  $x_{min}$  とスケール指数  $\alpha$  を用いて、実際の元データとの間の適合度 (goodness-of-fit) を、`plpva.m` を用いて算出した。ここまでの結果をまとめたのが表2である。ここで表中では太字で示してある適合度の確率が  $p > 0.1$  においてベキ則が採択されるべき仮説であるといえる (Appendix 1 参照)。

■手順3 尤度比検定 (likelihood ratio test) によって他の仮説とベキ則を比較する。それぞれの仮説について計算された尤度比の  $p$  値が0.1以下で正の値であればベキ則、負の値であれば対立仮説がもっともらしいといえ、もし  $p$  値が0.1以上ならば、どちらがよりもっともらしいかは言えない (Vuong, 1989)。

ここでは、R 言語によって提供されているプログラムを用いて、指数分布 ( $e^{-\lambda x}$ ) と遮断を含むベキ則分布 ( $x^\alpha e^{-\lambda x}$ ) を対立仮説とした\*6 (Appendix 2 参照)。

ここで、ベキ則分布への適合度の確率が  $p > 0.1$  で、かつ対立仮説の分布の尤度比の確率が  $p > 0.1$  であれば、ベキ則分布といえる。また、ベキ則分布への適合度の確率が  $p > 0.1$  で、かつ対立仮説の尤度比の確率が  $p < 0.1$  ならばベキ則分布とある程度は言えるという。ベキ則分布への適合度の確率が  $p < 0.1$  で、かつ対立仮説の尤度比の確率が  $p < 0.1$  で尤度比の符号が負であるならば、対立仮説のうちいずれか尤度比の大きな方の仮説を採用することになる。指数分布と遮断を含むベキ則分布との尤度比とその検定結果を表したのが表3である。ここでは8個の分布の内、一つ (Japan 2nd In) がベキ則分布といえ、3つ (Japan 1st In, 2nd Out, Ghana 1st In) がある程度はベキ則分布といえ、3つ (Japan 1st Out, Ghana 2nd Out, In) が遮断を含むベキ則分布といえた。残りの一つ (Ghana 1st Out) に関してはいずれの分布とも言えなかった。

#### 4. まとめ

本稿ではネットワーク理論の成立と発展を概説し、こ

のネットワーク理論をサッカーのパス行動へ適用し、ベキ則分布を検討するための方法を紹介した。ベキ則が成立することは、ハブの存在を確認することであり、自己組織化やフラクタルと関連が深い。しかしながら、われわれが扱うデータは、これまでネットワーク理論で扱われてきたデータよりも要素数 (頂点あるいは次数) がはるかに少ないため、完全なベキ則分布が成立することは難しいと考えられる。その中で、こうした尤度比検定等を用いてベキ則の存在を検討することは、集団ダイナミクスの解明の一助となるはずである。

#### 文 献

- Albert, R., Jeong, H., & Barabási, A. (1999). Diameter of the world-wide web. *Nature*, **401**, 130-131.
- Albert, R., Jeong, H., & Barabási, A. (2000). Error and attack tolerance of complex networks. *Nature*, **406**, 378-382.
- Barabási, A. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**, 509-512.
- Bollobás, B. (1981). Degree sequences of random graphs. *Discrete Mathematics*, **33**, 1-19.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.*, **51**, 661-703.
- Gladwell, M. (1999). Six degrees of Lois Weisberg. *New Yorker*, Jan. 11.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, **78**, 1360-1380.
- 松下貢 (2001). フラクタル. 「複雑系の辞典」編集委員会 (編), 『複雑系の辞典-適応複雑系のキーワード150-』, 323-332. 東京: 朝倉書店.
- Milgram, S. (1967). The small-world problem. *Psychology Today*, **1**, 61-67.
- 西山賢一 (2001). 自己組織臨界状態. 「複雑系の辞典」編集委員会 (編), 『複雑系の辞典-適応複雑系のキーワード150-』, 131-133. 東京: 朝倉書店.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307-333.
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, **393**, 440-442.
- Yamamoto, Y. & Yokoyama, K. (2011). Common and unique network dynamics in football games. *PLoS ONE*, **6**, e29638.

表3 ベキ則分布と対立仮説の分布との尤度比検定の結果。LR は2つの分布の尤度比を、その確率が  $p$ 。有意な  $p$  値に関しては太字で示す。

2006 Kirin cup			Power law	Exponential		Power law+cut-off		Support for
			$p$	LR	$p$	LR	$p$	power law
Japan	1st	Out	0.00	-4.159	<b>0.00</b>	-9.444	<b>0.00</b>	with cut-off moderate
		In	<b>0.45</b>	-2.318	<b>0.02</b>	-2.689	<b>0.02</b>	
	2nd	Out	<b>0.21</b>	-1.011	0.31	-1.833	<b>0.06</b>	moderate good
		In	<b>0.19</b>	-1.305	0.19	-1.351	0.10	
Ghana	1st	Out	0.03	-2.370	<b>0.02</b>	-1.778	<b>0.06</b>	none moderate
		In	<b>0.39</b>	-1.330	0.18	-1.726	<b>0.06</b>	
	2nd	Out	0.04	-2.720	<b>0.01</b>	-3.178	<b>0.01</b>	with cut-off with cut-off
		In	0.00	-3.552	<b>0.00</b>	-7.917	<b>0.00</b>	

## Appendix 1 MATLAB でのスケーリング指数推定と適合度検定例

```

clear all;
alpha_t = zeros(16,1); gof_t = zeros(16,1); p_t = zeros(16,1); xmin_t = zeros(16,1);

dat = dlmread('kcup.txt','\t',1,0); % データ読み込み
figure(1);

for i = 1:8
    x = dat(:,i);
    [alpha, xmin, nt, D] = plfit_yy(x); % 手順1
    [a_err, xm_err, nt_err] = plvar(x); % 手順2
    [p, gof] = plpva(x, xmin); % 手順3
    x = reshape(x, numel(x), 1);
    q = unique(x); n = length(x);
    c = hist(x,q)'./n;
    c = [[q; q(end)+1] 1-[0; cumsum(c)]]; c(find(c(:,2)<10^-10),:) = [];
    cf = ([xmin;q(end)]'.^-alpha)./(zeta(alpha) - sum([1:xmin-1].^-alpha));
    cf = [[xmin;q(end)+1] 1-[0; cumsum(cf)]];
    cf(:,2) = cf(:,2) .* c(find(c(:,1) == xmin),2);
    % 描画
    subplot(2,4,round(i/2)+4);
    if mod(i,2) == 0 ...
        loglog(c(:,1),c(:,2),'bs','MarkerSize',8,'MarkerEdgeColor','b','MarkerFaceColor','b');
        hold on;
        if p>0.1
            loglog(cf(:,1),cf(:,2),'b-','LineWidth',2);
        else
            loglog(cf(:,1),cf(:,2),'b--','LineWidth',1);
        end
    else
        loglog(c(:,1),c(:,2),'ro','MarkerSize',8,'MarkerEdgeColor','r','MarkerFaceColor','r');
        hold on;
        if p>0.1
            loglog(cf(:,1),cf(:,2),'r-','LineWidth',2);
        else
            loglog(cf(:,1),cf(:,2),'r--','LineWidth',1);
        end
    end
    alpha_t(i+8) = alpha; gof_t(i+8) = gof; p_t(i+8) = p; xmin_t(i+8) = xmin;
    D_t(i+8) = D; a_err_t(i+8) = a_err; xm_err_t(i+8) = xm_err;
    nt_t(i+8) = nt; nt_err_t(i+8) = nt_err;
end

```

## Appendix 2 R を用いた尤度比検定例

```

# データ読み込み
jout1<-c(1,1,2,3,5,2,4,2,1,1,1,1,0,1,1,1,3,2,1,3,3,0,1,5,5,3,3,3,5,4,4,4,2,1,0,0,1,
0,1,2,2,2,0,0,2,1,1,1,2,1,1,0,0,1,0,1,3,2,1,2,0,0,0,2,2,2,2,1,6,4,4,4,5,4,5,6,1,1,3,
1,1,0,0,1,0,2,1,0,5,1,5,6,3,4,4,0,3,2,0)
# ベキ分布最小値
jout1.min<-2
# ベキ則へのあてはめ
jout1.zeta<-zeta.fit(jout1,jout1.min)
# 指数分布へのあてはめ
jout1.discexp<-discexp.fit(jout1,jout1.min)
# 尤度比
jout1.zeta.discexp.llr<-zeta.exp.llr(jout1,jout1.zeta,jout1.discexp)
# 尤度比検定
vuong(jout1.zeta.discexp.llr)
# 遮断つきベキ則へのあてはめ
jout1.discpowerexp<-discpowerexp.fit(jout1,jout1.min)
# 尤度比検定
power.powerexp.lrt(jout1.zeta,jout1.discpowerexp)

```