

RA-001

スモールワールドネットワークにおけるストリング形式による q 次 の隔たりと一般化クラスタリング係数の考察

q Degrees of Separation and Generalized Clustering Coefficient In Small Word Networks by String Formulation

豊田 規人†
Norihiro Toyota

1. まえがき

半世紀前, Milgram は, 手紙の伝達に基づく社会実験によって“6次の隔たり” [1]という興味深い現象を発見した。この現象は多くの研究者の興味を引き, その後, 多くの研究者を刺激した[2]-[4]。彼のグループによる社会実験は, 全米の人々が, ほぼ6人の友人を介して繋がっているという状況証拠を与えた。この実験に対して, いくつかの批判があるが, その後, 追試の実験及び理論的考察がなされてきた[5]-[7]。その中でスモールワールドネットワーク [6],[7]やスケールフリーネットワーク [8]など, ネットワーク理論に大きなブレイクスルーがあり, この現象に対しても理解が進んできた[2],[3]。しかしその理解はまだ浅く, 特に, ネットワークのサイクル構造がどのようにその問題に影響を与えるのかは大きな課題である。

この問題は, まず, Newman によって3ノード及び4ノードを持つサイクル, すなわち三角形構造及び四角形構造の考察がなされてきた[9]。しかし, この方法では5ノード以上を持つサイクルへの一般化は困難である。一方, 青山等によって提案されたストリング形式により, 一般的議論ができる形式が整えられたが, 実際の考察はサイクル構造のない木近似でのみでなされている[10]。

最近, 我々はこのストリング形式を隣接行列を使い再定式化した。この方法では, 任意のノードからなるサイクル構造を考察することが可能であり[11],[12], 更に, 一般の(少なくとも6次の)隔たりまで考察可能になった。更に, この形式により三角形構造のみに着目した従来のクラスタリング係数を, 任意のノード数を持ったサイクル構造を反映させた一般化クラスタリングへ拡張ができるようになった。筆者は, この形式をスケールフリーネットワークに適用し, 6次の隔たりがスケール指数3と密接に関係していることを示した[13]-[15]。現実に存在する多くのスケールフリーネットワークのスケール指数がおおよそ3であることから, この結果は興味深いと考えられる。

この論文では, この手法をスモールワールドネットワークに適用し, スモールワールドネットワークにおける隔たり数, 及び, それとサイクル構造の関係について考察する。さらに, スケールフリーネットワークの場合と比較して, その様相がスケールフリーネットワークと決定的に違うことを示す。又, 隣接行列, 平均パス長, ミルグラム条件の数値計算を基にした考察により, 6次の隔たりに対して, ストリング形式が有効であることを示す。特に隔たり数と

ネットワークに内在するサイクル構造の関係の分析に対して, 隣接行列を使ったストリング形式が重要な方法論を与えることを示す。

2. 隣接行列に基づくストリング形式の再定式化

この章では, 青山等によって提案されたストリング形式 [10]の隣接行列を使った再定式化を与えた文献[11],[12]のレビューを与える。[11],[12]は, 主にこの再定式化のフォーマリズとその妥当性を論じたものであり, 6次の隔たりなどの応用は論じられていない。今, ネットワーク上で j ノードからなる連結した成分を, 青山に従い“ j -ストリング”と呼び, その本数を S_j^l とする。更に, j -ストリングのサブグラフにおいて, 多重連結やサイクル構造を持たないものを非縮退ストリングと呼び, その本数を S_j^u とする。但し j -ストリング全体のトポロジーがサークルと同相な場合も非縮退ストリングと定義する。しかし, 一般のグラフにおいて, S_j^l や S_j^u を理論的に求めることは困難であり, 目下, $j > 6$ の場合にそれらを求めることは不可能であると考えられる。

一方, このストリング形式を使うと, 通常の三角形構造を反映するクラスタリング係数 $C_{(3)}$ [6],[9]は, 以下の式のように表される。

$$C_{(3)} \equiv \frac{6 \times \text{number of triangles}}{\text{number of connected triplet}} = \frac{6\Delta_3}{S_3} \quad (1)$$

ここで Δ_p はネットワークの中にある一般的な p 本の枝からなる p 角形の個数である。従って, Δ_3 はネットワーク上の3角形の個数を表す。今, ネットワークに内在する一般的な多角形構造を考察したいので, それらの個数を反映した指標が必要である。我々は, (1)式から直接拡張して, 一般化 p 次クラスタリング係数 $C_{(p)}$ として以下の量を定義した [11],[12]。

$$C_{(p)} = \frac{2p \times \text{number of polygons}}{\text{number of connected } p\text{-plet}} = \frac{2p\Delta_p}{S_p} \quad (2)$$

次に, これを隣接行列 $A=(a)_{ij}$ を使って再定式化する。そのことによって, 一般化クラスタリング係数をシステマチックに求めることができる。一般的に隣接行列の冪乗 A^n は, n ステップで繋がっているノード間の情報を表すが, その行列要素は, その連結度合いの多重度を表す。従って A^n のままでは非縮退ストリングの本数を考察するには不向きである。そこでこういった縮退を解いた連結性を表す行列として A^n の代替として R^n を導入する。行列 R^n の i_0, i_1 成分は, n ステップで二つのノード i_0 と i_1 が連結しているかどうかについて, 多重度の情報を消し去った後の情報を与える行列である。具体的に $R^n(n > 1)$ は, 以下で与えられる

†北海道情報大学経営情報学部 Faculty of Business Administration and Information Science, Hokkaido Information University. (社)情報処理学会, IPSJ

[10],[11];

$$[R^n]_{i_0 i_n} = \sum_{i_1, \dots, i_{n-1}} a_{i_0 i_1} a_{i_1 i_2} \dots a_{i_{n-1} i_n} \frac{\prod_{0 \leq k, j, k-j > 1} (1 - \delta_{i_k i_j})}{(1 - \delta_{i_0 i_n})} \quad (3)$$

分子のクロネッカーの δ を含む積は非縮退ストリングのみ抽出する役割を持つが、分母はサークルに同相なストリングは残すという役割を持っている。(3)式を、このままの形で計算すると、 Σ 記号から来る多重のループ計算が生じ、莫大な計算時間がかかる。又、(3)式を展開すると $2^{n(n-1)/2}$ という膨大な項数になる。6次の隔たりに必要な $n=6$ では32768項にのぼり、このままではネットワークサイズが少しでも大きい場合、電子計算機でも現実的時間内での計算は不可能である。しかし、(3)式を展開し注意深く整理すると、ある程度コンパクトな表式が得られることが判明した。それにも関わらず、その表式は2, 3ページほどの紙面を必要とするので、具体的表式は、ここでは割愛する。なお、 $R^2 \sim R^6$ の具体的表式は[11],[12]に挙げられている。

この R^p の表式が、 S_q のシステムチェック計算を可能にさせ、更に、クラスタリング係数の一般化に対して以下のようにシステムチェックな表式を与える;

$$S_p = \sum_{i,j} (R^{p-1})_{ij} \quad (4)$$

$$C_{(p)} = \frac{\text{Tr } R^p}{\sum_{i,j} R^{p-1}} \quad (5)$$

ちなみに $C_{(3)}$ は

$$[R^2]_{ij} = [A^2]_{ij} - [A^2]_{ii} \delta_{ij}, \quad \|A\| \equiv \sum_{i,j} A_{ij} \quad (6)$$

を使って、以下のようになる。

$$C_{(3)} = \frac{\text{Tr } R^3}{\sum_{i,j} (A^2)_{ij} - (A^2)_{ii} \delta_{ij}} = \frac{\text{Tr } A^3}{\|A\| - \text{Tr } A^2} \quad (7)$$

3. q 次の隔たりへの応用

3.1 Milgram条件

次にこの形式を使って q 次の隔たりを考察する。 q 次の隔たりについて青山らは、ネットワークサイズが N の時、以下で表されるMilgram条件を提案している [10];

$$M_q \equiv \frac{S_q}{N} \approx O(N) \quad (8)$$

ただし、(8)式において、サークルに同相な閉ストリングの寄与は $N \rightarrow \infty$ の極限で無視できることを考慮している。この条件が q 次の隔たりが成立するための条件を表しているのである。すなわち1ノード当たりの q -ストリングの平均的本数が、ほぼネットワークサイズに等しいという条件であるが、このことはネットワーク全体が、基本的に q 次で繋がっているということの自然な条件を与えていると考えられる。 S_q が大きいほど、より精確には、 N が固定されている場合、 M_q が大きいほど q 次の隔たりが満たされる度合いが大きくなる。(8)式に現れる S_q は(49式から R^p 行列を使って、評価できるのである。

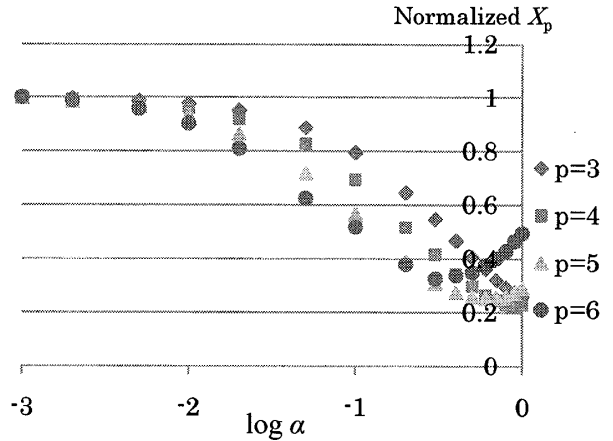


Fig.1 Normalized X_p

3.2 スモールワールドネットワークへの適用

この条件をスケールフリーネットワークに適用した場合は[13]-[15]で議論されている。この論文では、ミルグラム条件をスモールワールドネットワークに適用し、その結果を分析しスケールフリーネットワークの場合との違いについて考察する。又、これらを通して、サイクル構造と隔たり数 q の分析において、ここで与えられたストリング形式の有用性が示される。

スモールワールドネットワークの作成は Newman-Watts に従う [16]。基本となるネットワークは、1次元のサークルに同相な次数4の規則ネットワークとする。又、このモデルのパラメータである、枝の付加率を α とし、 $N=200$ とする。 R^6 の計算量が大きいため、現実的な計算時間を考慮すると N を大きくすることができないが、このサイズでも $C_{(3)}$ 、及び、平均パス長 L はスモールワールドネットワークの特徴を表すことは確認されている(実際後述の図1や図4がそれらを表している)。

6次の隔たりを調べる時、サイクル構造を担う指標として $C_{(3)} \sim C_{(6)}$ の情報が必要である。そこで以下のような量を定義する;

$$X_p = \sum_{i=3}^p C_{(i)} \quad (9)$$

$C_{(1)}, C_{(2)}$ はサイクル構造と無縁なので和は、 $i=3$ からとられている。従って、 X_p は p 個以下のノードからなるすべてのサイクル構造の情報を担っている。

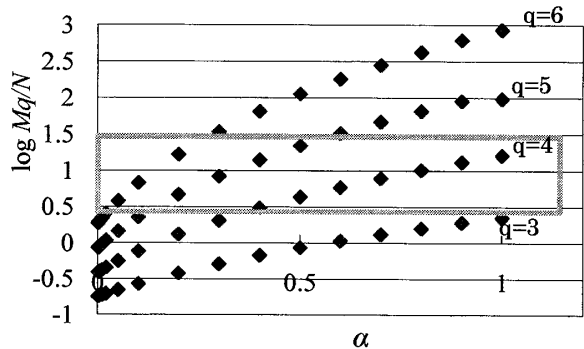


Fig.2 Milgram Condition with $q=3-6$

$p=3\sim 6$ における X_p を, α を変えて横軸の片 \log 表示したのが図1である. この論文では, 一貫して \log の底は 10 を用いている. 図1の縦軸 X_p の値は p ごとに X_p の最大値が 1 になるように規格化された値を表示している. 一番上方のダイヤモンドマークで表示されたデータ系列点は $p=3$ で, それより下方の点列は, それぞれ $p=4, 5, 6$ の時のものである. 全体的に X_p は通常の C_3 と類似した振る舞いを示していることが分かる. そういった意味でこのネットワークは, “一般化スモールワールドネットワーク” といつてよいであろう. $\alpha=0$ の時はランダムグラフであるが, そのあたりで, $X_{5,6}$ は若干増加している. これは, スモールワールドネットワークの構成が, 比較的規模の小さい円環を基にしたため, 有限サイズ効果が表れたためである. 又, 図1において, p が大きくなるにつれて, α による変化が小さくなるということは, 最初に用意した格子ネットワーク ($\alpha=0$) において, 3 より多くのノードを持つサイクル構造が無かったことを考えると自然に理解できる.

この X_p の振る舞いと 6 次の隔たりの関係を考察するために, M_q と α の関係を求める. 図2は各 α における M_q/N が縦軸に \log プロットされている. 4 種のデータ系列は, 下から $q=3,4,5,6$ の場合のデータ点列である. 図2中, 枠内は Milgram 条件がおおよそ満たされている領域である. これから, $q=6$ つまり 6 次の隔たりについては, 図1において, X_p の値がちょうど下がり始める時に Milgram 条件が満たされ始めることが分かる. つまり, 比較的わずかでも近道に対応する枝を加えることによって, 容易に 6 次の隔たりが実現しうることが分かる. しかし, 図2は, ランダムグラフ ($\alpha=1$) に近づけても, 4 次の隔たりにはなりうるが, 3 次で繋がることは厳しいことを示している.

このことを隣接行列を使った議論で補強する. 今, A^n の 0 以外の要素の個数の, 全要素数に対する比率を r_n とする. この値は n ステップでお互いに連結しているノード間の情報を与える. q 次の隔たりを考察する場合は, $n=q$ 以下のステップで連結しているノード間の情報が必要であるので, 以下の量を導入して考察する;

$$T_n = \sum_{i=1}^n r_n \quad (10)$$

様々な α の値で T_n を評価したものが図3に掲げられている. 図3における系列 i は $n=i+1$ ステップまでの連結を表す T_{i+1} を表している. T_n が 1 を超える場合があるのは, 二つのノードが異なったステップ数で連結している場合があるからである. 図3で, 0.5 (連結ステップ数の重複も含む) 以上

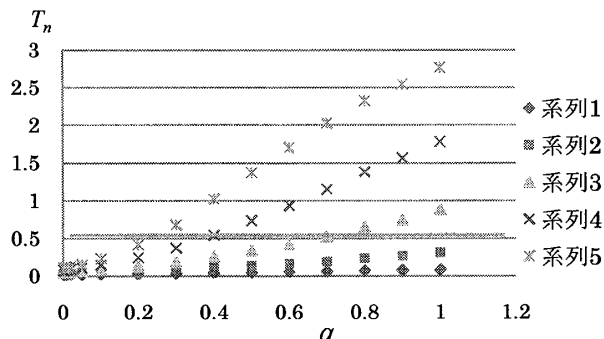


Fig.3 Numbers of nonzero elements in Adjacent Matrix

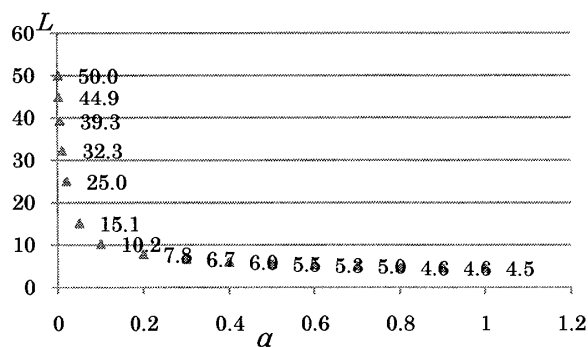


Fig.4 Average Distance

の確率でお互いのノードが n ステップ以内で連結しているところをボーダーとすると, 図2で Milgram 条件が満たされている領域のほぼ中央, つまり, $\log M_q/N$ が 1 以上の領域に対応していることが分かる. つまり互いの 2 対のノードが 50%以上の確率で繋がっている場合が, 丁度 Milgram 条件のクリティカルポイントに対応している.

次に平均パス長 L とパラメータ α の関係を考察する. 図4に, L が α とともにどう変化するか, L の数値データと共に描かれている. 通常のスモールワールドネットワークの特徴として, α を少し大きくしただけで L の値は急速に減少していくが, 図4からもその様子が見て取れる. 隔たり数との対応として $L \sim 10$ と考えると, 図2及び, 図3で言及された内容と無矛盾である. 従ってこれらの 3 種の考察は本質的に隔たり数 $q=n$ に関してほぼ同様な知見を与える. しかし, ネットワークのサイクル構造と隔たり数の議論においてクラスタリング係数の和 X_p は, 特に重要な役割をする. 従って我々の与えた定式化と(9)式を連動させることが, サイクル構造と q 次の隔たりを考察する上で重要な方法論を与えることになる.

図1と図2から, クラスタリング係数が小さいほど 6 次の隔たりが起きやすい, あるいは隔たり数 q が小さいということが分かる. このことをもう少し定量的に調べる. 図5は α を変えていった時の X_n と M_n/N の両対数グラフである. それぞれの筋状のデータ点列は, $n=3\sim 6$ に対応している. データ点列はそれぞれほぼ直線状に並んでいることが分かる. このことから, 両者は, 次式で与えられる冪乗則に従っていることが分かる.

$$M_n \sim B(X_n)^{-a} \quad (11)$$

ただし, B と a は数値実験から n ごとに求められる定数である. この結果は, スケールフリーネットワークの時には指数則

$$M_n \sim \text{Exp}(cX_n) \quad (12)$$

に従っていた[13]-[15]ことと比べると対照的である. ここで定数 c 及び a は正の数である. 従って(11)と(12)の振る舞いは, 全く逆であることが分かる. つまり, スモールワールドネットワークでは, 総じて一般化クラスタリング係数が大きくなれば, Milgram 条件は満たされにくくなるが, スケールフリーネットワークでは急激に満たされやすくなる. スモールワールドネットワークでは, サイクル構造は情報

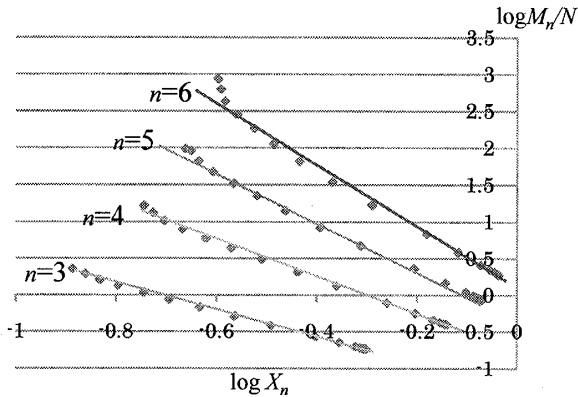


Fig.5 Milgram Condition and Generalized Clustering Coefficient

(Milgram の実験では手紙) は同じサイクルをぐるぐる回り、遠方への伝搬を阻害する方向に働く。逆にスケールフリーネットワークは、木構造に近いのであるが、サイクル構造が近道を提供し、情報の伝搬がスムーズになったと考えられる。このように、隔たり数 q はサイクル構造が多く含まれているほど、一律に大きくなる、あるいは小さくなるといった単純なものではない。 q 次の隔たりに対するサイクル構造の影響は、基本となるネットワークの構造に大きく依存することが分かった。

又、スモールワールドネットワークでは、 n を固定した時に、図 5 のように X_n と M_n/N の両対数グラフは一つの直線になっており、 n を変えると直線の傾きや切片が変わる。これは、ネットワークのトポロジー、すなわちネットワークに内在するサイクル構造を変化させた場合に、 q 次の隔たりの実現の仕方が、 $q=n$ ごとに違うということを意味している。一方、スケールフリーネットワークでは、次数 k の分布が以下で定義されるが、

$$P(k) \sim k^{-\gamma} \quad (13)$$

スケーリング指数 γ を変え、ネットワークトポロジーを変えると、 X_n と M_n の関係の振る舞いは n に殆どよらないことが分かっている[15]。つまり、 n を変えても、片 \log グラフにおいて一本の共通の直線上にデータ点が並んだグラフが見られる。このことは、スケールフリーネットワーク場合、 q 次の隔たりの実現仕方が、どの $q=n$ でもネットワークのトポロジー変化に対する応答が同じであることを意味している。

4. まとめ

この論文では、我々の開発した、ネットワークにおけるストリング形式を隣接行列による再定式化を使い、スモールワールドネットワークにおける 6 次の隔たり、更に一般的な隔たり数 q について考察した。この再定式化により、一般のサイクル構造の情報を担っている一般化クラスタリング係数をシステムチックに計算可能になる。又、我々の提案したこの定式化は、次数分布がポアソン分布になるランダムネットワーク[17]やスケールフリーネットワーク

などで、既知の性質を再現することが確認されていて、信頼できる形式である[11],[12]。

この形式による、スケールフリーネットワークへの解析によって、6 次の隔たりと次数分布のスケーリング指数の間の関係が明白になっている。特に現実世界によく見られるスケーリング指数が約 3 の場合に、丁度 6 次の隔たりと密接に関係しているという興味深い結果が得られている[13]-[15]。

この論文では、この形式をスモールワールドネットワークに適用し、一般の隔たり数について分析した。更にスケールフリーネットワークの分析結果と比較考察した。その結果、一般化クラスタリング係数については、通常三角構造のみ考慮したクラスタリング係数とほぼ類似した振る舞いをする事が分かった。又、隔たり数に関する Milgram 条件の考察より、1 次元規則格子に対して、少数の近道を付加しただけで、6 次の隔たりに対する Milgram 条件は急速に満たして行くことが分かった。この点は平均パス長の振る舞いと同一である。又、この点は隣接行列の行列要素の分析からも支持される。Milgram 条件、隣接行列、平均パス長の考察は、その結果においてほぼ同じ知見を与えるが、サイクル構造と関係して議論できるのはストリング形式による Milgram 条件の分析である。このように、我々の提案した隣接行列に立脚したストリング形式は、隔たり数、より広くはネットワーク上の情報伝搬に対して、内在するサイクル構造を考察する上で大きな意義を持つことが分かった。このことがこの論文の主要な主張の一つである。

更に、今回の分析では、Milgram 条件を表す M_n と一般化クラスタリング係数の間に冪乗側が成り立つことが判明した。このことは、スケールフリーネットワークの時と対照的である。スモールワールドネットワークにおいては、サイクル構造が、隔たり数に対して抵抗力として働き、むしろそういった構造が無い方が隔たり数が小さくなる。このように、サイクル構造のあり様の、隔たり数に対する影響は、基本となるネットワークの性質・構成法に大きく依存することが判明した。

今後の課題は、大きな n に対する R^q のコンパクトかつシステムチックな表式を与え、6 次以上の隔たりに関する分析をおこなうことである。更に、より大規模なネットワークでの確認が必要になるであろう。いずれにしても、そのために膨大な計算が必要であり、高性能のコンピュータが必要になる。しかし、スケールフリーの場合、その本質的振る舞いは、ある程度ネットワークのスケールによらないことが筆者によって確認されている。このことは、“スケールフリー”という概念の定義から自然な帰結であると考えられる。

参考文献

- [1]S. Milgram, "The small world problem", *Psychology Today* 2, 60-67 (1967)
- [2]R. Albert and A.-L. Barabasi, "Statistical Mechanics of complex networks", *Rev. Mod. Phys.* 74, 47-97(2002)
- [3]M.E.J. Newman, A.-L. Barabasi and D. J. Watts, "The Structure and Dynamics of Networks", Princeton Univ. Press, 2006

- [4]J. Travers and S. Milgram, "An Experimental Study of the Small World Problem", *Sociometry* 32, 425 (1969)
- [5]C. Korte and S. Milgram, "Acquaintance edges between White and Negro populations: Application of small world method", *Journal of Personality and Social Psychology* 15(2), (1970)
- [6]D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, 393, 440-442(1998)
- [7]D. J. Watts, "Six degree-- The science of a connected age", W.W. Norton and Company, New York (2003)
- [8]A.-L.Barabasi and R.Albert, "Emergence of scaling in random networks", *Science*, 286, 509-512(1999)
- [9]M.E.J.Newman,"Ego-centered networks and the ripple effect or why all your friends are wired", *Social Networks* 25 (2003) p.83;arXiv. cond-mat/0111070
- [10]青山秀明, 家富洋, 池田裕一, 相馬亘, 藤原義久, 「経済物理学」, 3.4 章「つながり計算」, 共立出版, (2008)
- [11]豊田規人, " p 次クラスタリング係数のストリング形式 —6 次の隔たりに向けて—", 信学技法 IEICE Technical Report NLP2009-49(2009.Aug.)
- [12]N. Toyota, " p -th Clustering coefficients C_p and Adjacent Matrix for Networks: Formulation based on String", arXiv: 0912.2807
- [13]豊田規人, 坂本智晴, "隣接行列-ストリング形式による P 次の隔たり", 第 6 回ネットワーク生態学シンポ 2009.Dec.
- [14]豊田規人, 坂本智晴, ; "スケールフリーネットワークにおける一般化クラスタリング係数と 6 次の隔たり" 第 42 回計測自動制御学会北海道支部学術講演会 2010.Feb.
- [15]Norihito Toyota, "p-th Clustering coefficients and q-th degrees of separation based on String-Adjacent Formulation", Preprint arXiv:1002.3431
- [16]M.E.J. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model", *Phys.Lett.A*263, 341-346(1999)
- [17]P. Erdos and A. Renyi, "On random graphs I", *Publications Mathematicae Debrecen*6, 290-297, 1959