

# 革新的ハイパフォーマンス・コンピューティング・インフラ(HPCI)における研究者ネットワークの分析

峯尾 真一

一般財団法人 高度情報科学技術研究機構

mineo@rist.or.jp

## Analysis of Researcher Network in the High-Performance Computing Infrastructure in Japan (HPCI)

Shinichi Mineo

Research Organization for Information Science & Technology

### 概要

HPCI における研究者ネットワークをグラフとして可視化した結果、3000 人を超える巨大なクラスタが形成された。そこでこのグラフの分析を行い、重点的利用区分課題の介在と、計算機科学系研究者の異分野を含めた幅広い研究交流が、クラスタの成長を促進していることが分かった。また比較的少数の研究者が、中小規模の研究集団を結びつけ、ハブとして巨大クラスタを形成する大きな要因となっていることを明らかにした。

また HPCI の研究者ネットワークでスモール・ワールド現象が確認できることから、グラフの構造を分析し、その特徴が SNS のネットワークに近いことを示した。

## 1 はじめに

本分析のきっかけは、HPCI システム利用研究課題の選定過程で派生した問題であった。一般利用区分<sup>1</sup>に応募された研究課題のレビュー依頼時に、レビュアーと評価対象課題の参加者との利益相反関係を事前に知ることはできないか、というものである。そこで過去に同一の課題に参加した人をネットワークとして結び、グラフの最短経路問題として解けるのではないかと考えた。

当初は研究集団の中小クラスタが多数形成されると想定したが、実際にやってみると、意外にも 3,000 人を超える巨大なクラスタが出現した。すなわちほとんどの人がどこかで繋がっていることが分かった。そこで、このクラスタを作った要因は何か、またこれは、いわゆるス

モール・ワールド現象<sup>2</sup>なのか、という疑問に答えるべく、以下の分析を行った。

本論文では、クラスタが巨大化する要因として次の仮説を検証した。

- ・重点的利用区分課題<sup>1</sup>を媒介とした研究交流
- ・計算機科学系研究者を介した異分野交流
- ・多数の研究者を結び付けるハブ的な研究者の存在

また、形成されたクラスタをグラフとして分析し、スモール・ワールド現象との類似性を調べた。さらにソーシャル・ネットワーキング・サービス(SNS)と比較して、HPCI の研究者ネットワークの特徴を調べた。

### <sup>2</sup> スモール・ワールド現象

社会心理学者スタンレー・ミルグラムが 1967 年に行ったスモール・ワールド実験で検証された現象。この実験ではアメリカ合衆国国民から 2 人ずつの組を無作為に抽出し、その 2 人がつながっている場合には、平均すると 5 人の知り合いを介していることを示した [1][2]。この実験は、それ以前から文学作品等で知られていた「6 次の隔たり (Six Degrees of Separation)」という仮説の裏付けになったと言われている。

<sup>1</sup>一般利用区分と重点的利用区分

HPCI システム利用研究課題の利用区分には、公募される一般利用区分と、別途国が定めたプロセスで選ばれる重点的利用区分がある。

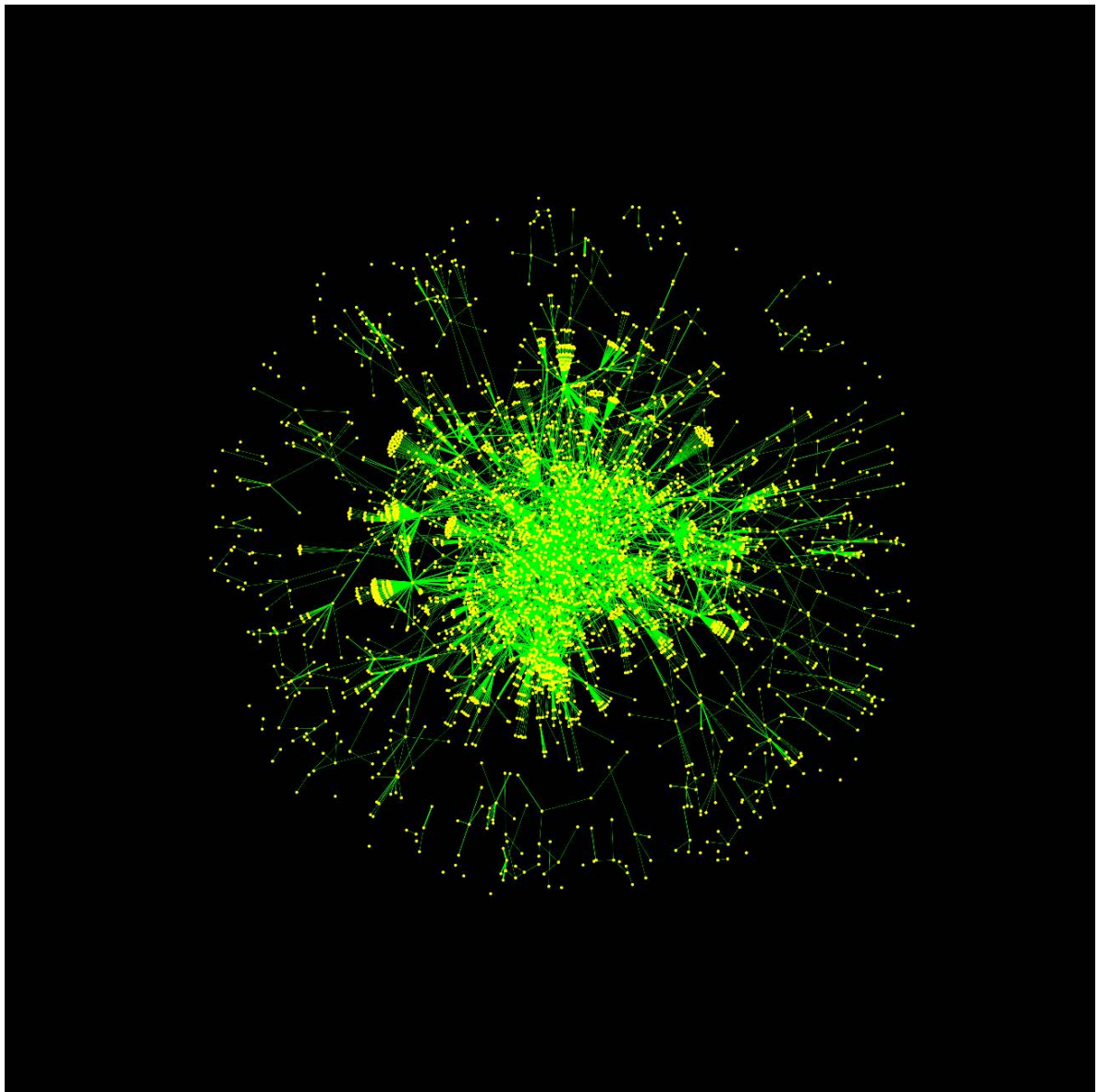


図 1 全課題参加者のグラフ G

者と他の参加者を結ぶリンクを表す。

## 2 グラフの作成

2012年9月に「京」を中核とする HPCI システムの共用が開始され、以来毎年、定期および随時の研究課題募集が行われている。そこで、一般利用区分<sup>1</sup>に応募された課題および重点的利用区分に選定された課題の参加者をノードとし、課題代表者とその他の参加者の間をリンクとしてスター型に結ぶ。課題代表者以外の参加者ノード同士の結合はしない。使用したデータは、2018年B期までの全課題で、課題数2,256、のべ参加者16,485人となった。

作成されたグラフ G を図 1 に示す。黄色の点が課題参加者ノードを表し、緑の線が課題代表

表 1 グラフ G の特徴量

項目	G
ノード数(全課題参加者数)	3996
クラスタ数	177
最大クラスタ内ノード数	3443
ノード平均リンク数	3.04
ノード平均クラスタ係数 <sup>3</sup>	0.16

グラフ G 上でできたクラスタ（リンクで繋がっているノードの集合）は 177、その内最も大きなクラスタには 3443 人（ノード）が属している。この様な大きなクラスタができたということは、研究者が分野を超えて直接・間接に繋がっていることを意味する。

このグラフの基本的な特徴量を表 1 に、クラスタの人数分布を多い順に図 2 に示す。全課題参加者 3996 人の 86% に当たる 3443 人が、一つの巨大なクラスタを形成していることが分かる。2 位以下は 20 人未満である。

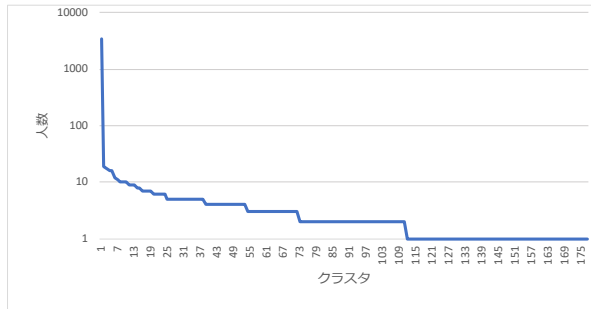


図 2 グラフ G クラスタの人数分布

### 3 巨大クラスタ形成の要因分析

HPCI において、このような巨大なクラスタが形成される要因として以下の通り 3 つの仮説を立て、各要因の影響度を分析した。

#### 3.1 重点的利用区分課題を媒介とした研究交流

重点的利用区分課題は、HPCI を活用した成果の創出と、研究推進・研究支援や人材育成等を進めていくための体制整備という観点で選定された課題である。必然的に各研究分野で中核となる人材や、計算機科学者等分野横断的に活躍できる人材を含んでおり、この課題を介して研究者ネットワークが広がった可能性がある。

そこで、重点的利用区分の課題を除去したグラフ G1 を作成し、基本的な特徴量の変化を調べた。(表 2) また、研究者の個人的な貢献度を調べるため、重点的利用区分課題への参加経験者 1265 人に対し、元のグラフ G における基本的な特徴量を算出して比較した(表 3)。

#### 3 クラスタ係数

次数(ノードの持つリンク数)  $k$  のノードのクラスタ係数とは、そのノードに隣接しているノード同士を結ぶリンク数(すなわちそのノードを頂点とした三角形ができる数)を、そのリンク数のとりうる最大値である  $k(k-1)/2$  で割った値。ネットワークの凝集性を表現しており、グラフのノード平均クラスタ係数が大きい程、そのネットワークは密であると言える。

#### 表 2 グラフ G1 の特徴量

項目	G	G1	増減
ノード数	3996	3350	-16.2%
クラスタ数	177	209	18.1%
最大クラスタ内ノード数	3443	2679	-22.2%
ノード平均リンク数	3.04	2.66	-12.5%
ノード平均クラスタ係数	0.16	0.13	-16.9%

#### 表 3 重点的利用区分課題参加者ノードの特徴量

項目	G	対象ノード	増減
ノード数	3996	1265	-68.3%
ノード平均リンク数	3.04	4.69	54.4%
ノード平均クラスタ係数	0.16	0.24	54.0%

重点的利用区分課題が無かったとしたら、研究者ネットワークのクラスタは 177 から 209 に 32 増え、最大クラスタが 3,443 人から 2,679 に 764 人縮小する。ノード数の減少率よりもクラスタ数の増加率や最大クラスタの縮小率の方が大きいことは、重点的利用区分課題にはネットワークの成長促進効果があることを意味する。その効果により、新たに 764 人の研究者を含む 32 の研究グループが結びついたことになる。

重点的利用区分課題の参加者を見ると、平均リンク数、平均クラスタ係数ともに全体平均より 50% 以上大きく、研究者間の緊密な連携と交流により複雑系ネットワークを成長させる要因となりうることが分かる。

#### 3.2 計算機科学系研究者を介した異分野交流

課題参加者各々の専門分野についての情報が無いため、残念ながら計算機科学系研究者を特定することができず、その影響を正確に測定することはできない。ここでは、HPCI システム利用研究課題のレビューアとして登録された研究者の内、計算機科学を専門分野としており、かつ過去にいずれかの研究課題に参加している 50 人について、ノードの基本的な特徴量を計算し、全体平均と比較した(表 4)。

#### 表 4 計算機科学の課題参加者ノードの特徴量

項目	G	対象ノード	増減
ノード数	3996	50	-98.7%
ノード平均リンク数	3.04	9.56	214.7%
ノード平均クラスタ係数	0.16	0.14	-12.5%

計算機科学を専門とする研究者の顕著な特徴は、ノード平均リンク数が非常に大きい、す

なわち多くの課題に参加する傾向が強いことである。このことから、計算機科学を媒介とした異分野を含めた研究交流が実現されている可能性が高いと考えられる。

それを確かめるために、この 50 人について参加した研究課題の分野<sup>4</sup>の種類数を調査した(図3)。一人で8種類の研究分野に参加した実績を初め、異分野の研究課題に積極的に参加している様子が伺える。

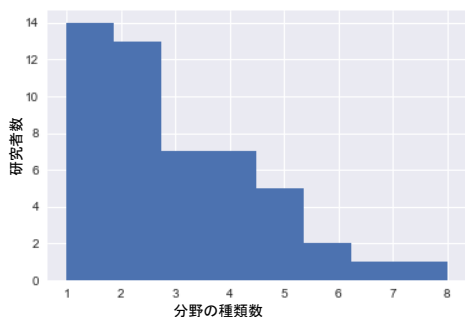


図3 参加した研究課題の分野の種類数

またグラフ G 上の、この 50 人のノードについて、リンク数の多い順に外して行き、クラスタの増加数を調べた(図4)。赤丸で示した少数の人の影響が顕著に見え、131 に分解されることが分かる。

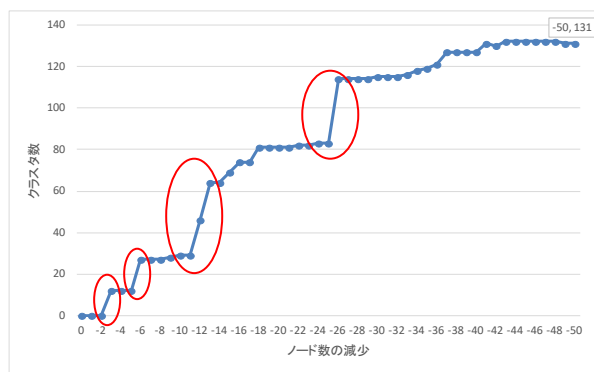


図4 ノード減少によるクラスタ数の増加

### 3.3 多数の研究者を結び付けるハブ的な研究者の存在

巨大クラスタに含まれる各ノードのリンク数を数えてヒストグラムにしたものが図5であ

#### 4 研究分野

- ①数理科学、②物理・素粒子・宇宙、③物質・材料・化学、④工学・ものづくり、⑤バイオ・ライフ、⑥環境・防災・減災、⑦情報・計算機科学、⑧エネルギー、⑨社会システム科学、⑩その他、の10種類

る。上位2人のノードのリンク数は、110と87であり、ハブ的な効果を発揮する研究者が実在することが分かる。

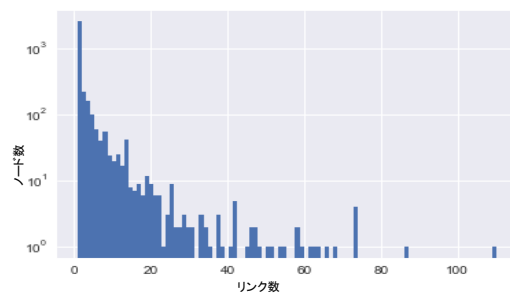


図5. 巨大クラスタにおける各ノードのリンク数

そのハブとしての効果を調べるために、上位10人の研究者をリンク数の多い順に外して、このクラスタがどの様に分割されていくかを調べた(図6)。

最初の一人は74のクラスタ(研究集団)を結びつけていることが分かる。また10人のノードを外しただけで、研究集団が322に分裂することから、ハブ的な研究者の存在が非常に大きな影響を持つことが分かる。なお、この10人は、3.2の計算機科学系の研究者とは重複していない。また10人の内6人は重点的利用区分課題の経験者、4人は一般利用区分のみの課題経験者であり、利用区分を問わずハブ的な研究者が存在することが分かる。

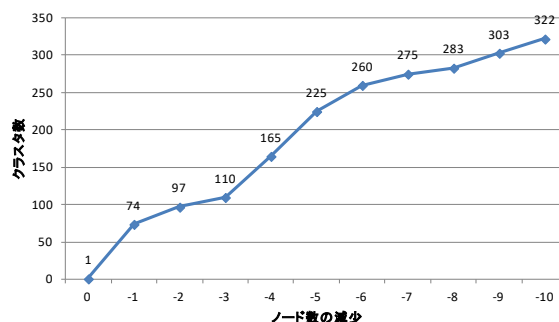


図6. リンク数の多い順にノードを削除した結果

## 4 スモール・ワールド現象との類似性

ミルグラムのスモール・ワールド実験で検証されたのは6次の隔たりであった。HPCIの巨大クラスタのみを取り出し、グラフG2として、その特徴量を測定した結果を表5に示す。平均最短経路長は5.66となり、このグラフG2においても「6次の隔たり」が確認できた。

表 5 最大クラスタのグラフ G2 の特徴量

項目	G2
ノード数	3443
ノード平均エッジ数	3.30
ノード平均クラスタ係数	0.18
平均最短経路長	5.66

1998年に Watts らが、small-world network を定式化して作成する方法を発表した[3]。その数学的モデル中で指摘された small-world network の特徴は、平均経路長が短くクラスタ係数が大きいということである。一方グラフ G2 の平均クラスタ係数は0.18と大きくはない。これはグラフ作成の時に、課題参加者同士のリンクを張っておらず、全てスター型のネットワークとなっていることを考慮すると不自然ではない。Watts らの論文にある small-world network の实例を見ると、クラスタ係数が 0.080~0.79 となっており、HPCI が small-world network であると言うことに問題はないと考える。

またこの論文は、ノードが多数あっても、少数のショートカットさえあれば、スモール・ワールド現象が起こりえると指摘している。HPCI の場合は、3.で検証した計算機科学系研究者を介した異分野交流や、多数の研究者を結び付けるハブ的な研究者の存在がショートカットを形成していることが容易に想像できる。

## 5 HPCI の研究者ネットワークと SNS の比較

現実世界のネットワークである SNS の構造的な特徴は[4]、

**特徴 1:** 次数分布がべき乗則<sup>5</sup>に従う (有向グラフである SNS では両方向)

**特徴 2:** 次数が低いノードの集まりである多数のクラスタが、少数の次数の高いノードを介して繋がる (その結果、クラスタ係数はノードの次数に反比例する)

<sup>5</sup> べき乗則(power law)

次数:k、次数がkの頂点の割合:p(k)、べき指数:γ、とした時、 $P(k) \propto k^{-\gamma}$ と表せること。

**特徴 3:** 大きく高密度で繋がったコアを持つ (ハブ的なノードが存在し、それを介して接続経路が短縮される)

とされる。ここでは、この参考文献[4]で実測された SNS や Web の特徴量と比較して、HPCI の研究者ネットワークの特徴を調べる。

既に分析された結果(3.2 および 3.3)から、HPCI の研究者ネットワークは、特徴 3 を持ち、スモール・ワールド現象を引き起こす要因となっていることが分かる。HPCI で測定された平均最短経路長は 5.66、一方 SNS の例では 4.25~5.10 となっており、ほぼ SNS と等しい。

次に、特徴 1 について調べるために、HPCI の研究者ネットワークのグラフ G の各ノードの次数の分布図を描いた (図 7)。両軸ともにログスケールであり、直線状に分布することから次数分布がべき乗則に従っていることが分かる。

このべき乗則のべき指数γを参考文献[5]に従って推定すると 2.71 となった。SNS の例 (1.50~1.99) よりも大きく、web の例 (2.09~2.67) に近い。すなわち SNS よりも次数の偏りが大きい (次数の小さなノードはより多く、次数の大きなノードはより少なく存在する) ことが分かる。HPCI の利用研究課題は、明確な意図を持って意識的に作られたグループであるという意味で、SNS のグループと相似性があると考えられるが、今回のグラフ作成では、課題参加者同士をリンクで結ばなかったことが、次数の小さなノードが多くなった原因と考えられる。

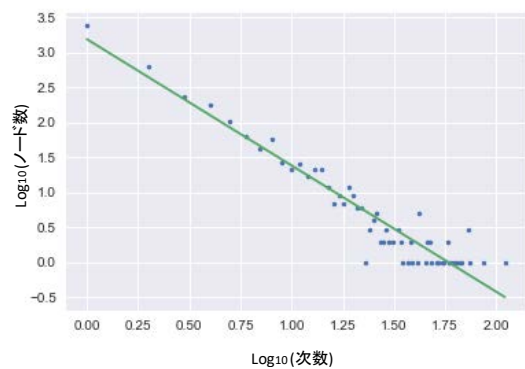


図 7. 次数とノード数の関係

次に特徴 2 について調べるために、まずはネットワークの平均クラスタ係数を比較すると、SNS が 0.136~0.330 に対して、HPCI は 0.18 であり、ほぼ SNS と等しい。

また、ノードの次数とクラスタ係数の関係を図示すると(図 8)、SNS とは違い、次数とクラスタ係数は逆比例せず、よりクラスタ係数の小さなノードが多いことが分かる。これは HPCI の場合は、最小単位のグループである研究課題がトポロジ的にスター型であるためであると考えられる。大局的には SNS と同様に、次数が低いノードの集まりである多数のクラスタが、少数の次数の高いノードを介して繋がる傾向が見える。

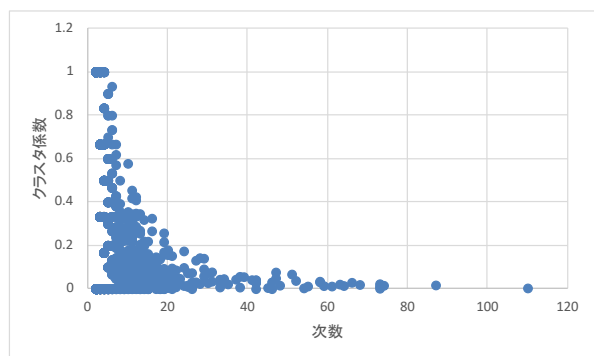


図 8. 次数とクラスタ係数の関係

## 6 まとめ

以上の分析から言えることは、HPCI はこの 6 年間で分野を超えて繋がる 3,000 人を超える規模の研究者ネットワークを築き上げたこと、その形成には重点的利用区分課題による複雑系ネットワークの成長促進効果と、計算機科学系研究者の異分野を含めた幅広い研究交流が貢献していると考えられることである。また比較的少数の研究者が、中小規模の研究集団を結びつけ、ハブとして巨大クラスタを形成する大きな要因となっていることが明らかになった。

生成されたネットワークではスモール・ワールド現象が確認できることから、そのグラフ構造を調べて、SNS に近いことを示した。この結果は、人間の社会的行動が本質的に変わらないことを示す一例であると考えられる。しかしその社会的行動を促し、今日のこの分野を超えた研究コミュニティを創り上げたことは、我が国 HPCI の大きな成果であったと言えるのではないだろうか。

今後の課題は、本分析の発端となったレビューアー選定への応用を試みること、および研究集

団のネットワーク構造と成果創出の関係等、さらに分析を進めることである。またこの様な分析結果を、将来の HPCI の研究コミュニティ作りに生かせないかとも考えている。

## 謝辞

本論文の作成にあたり的確な助言を頂いた、関昌弘理事長を初めとする一般財団法人高度情報科学技術研究機構の方々に感謝します。

## 参考文献

- [1] Stanley Milgram, “The Small-World Problem”, *Psychology Today*, vol.1, no.1, May 1967, pp61-67
- [2] Travers, Jeffrey, and Stanley Milgram. “An Experimental Study of the Small World Problem.” *Sociometry*, vol. 32, no. 4, 1969, pp. 425–443. JSTOR, JSTOR, [www.jstor.org/stable/2786545](http://www.jstor.org/stable/2786545).
- [3] Duncan J. Watts, et al., “Collective dynamics of ‘small-world’ networks”, *Nature* volume 393, pages 440–442 (04 June 1998)
- [4] Alan Mislove, et al, “Measurement and analysis of online social networks”, *IMC’07*, October 24-26, 2007, San Diego, California, USA.
- [5] Jeff Alstott, Ed Bullmore, Dietmar Plenz. (2014). powerlaw: a Python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(1): e85777 <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0085777>>