

「インターネットはスケールフリー」論再考

一井 信吾†

† 東京大学大学院数理科学研究科 〒153-8914 東京都目黒区駒場 3-8-1

E-mail: fuchii@ms.u-tokyo.ac.jp

あらまし 複雑ネットワーク研究の興隆の中で、インターネットのトポロジ（ルータレベル及びASレベル）は、その次数分布がベキ則を示すことから、所謂スケールフリーネットワークの代表格の一つとして扱われてきた。しかし、その元となるデータの測定方法やスケールフリーモデルによる議論等は、調べていくとかなり怪しい部分がある。簡単なシミュレーションの結果をふまえ、現状について批判的な観察と、今後の研究の進むべき方向についての提案を示す。

キーワード インターネットトポロジ、次数分布、ベキ則、複雑ネットワーク

“Internet is scale-free” Theory Revisited

Shingo ICHII†

† Graduate School of Mathematical Sciences, The University of Tokyo 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8914, Japan

E-mail: fuchii@ms.u-tokyo.ac.jp

Abstract In the boom of the complex network research, Internet topologies (both of the router level and of the AS level) has been accepted as one of the typical examples of so-called scale free networks and are heavily discussed by many researchers. However, the measurement methodology to obtain the data as the basis of those studies and the logical structure of the discussions using scale free models are not quite free from theoretical and practical doubts. In this report, we show critical observations based on simple simulations and then a proposal for future research.

Key words Internet topology, degree distribution, power-law, complex networks

1. はじめに

近年の複雑ネットワーク研究[1]の中で、インターネットトポロジ^(注1)はその代表格のひとつとして扱われて、さまざまな研究や考察の対象となってきた。中でも、次数を始めとするいくつかの指標がベキ則に従うことを示した Faloutsos-Faloutsos-Faloutsos [2] (以下 FFF とする) の先駆的な研究、及びこれを一つの有力な事例とした Barabási-Albert [3] の発表以来、「インターネットはスケールフリー^(注2)である」という認識が広まり[4]、一般向け解説書や論文のイントロダクション等で話の枕として使われるに至っている。

Barabási-Albert はスケールフリーなネットワークのモデル

(注1)：本稿でインターネットトポロジという場合は、インターネットにおけるAS間相互接続ないしルータ間相互接続のトポロジを指すものとする。すなわち、WWWにおけるページ間の参照関係や電子メール送受信関係、P2Pネットワーク上のノード間接続関係等のトポロジを対象としない。

(注2)：本稿では“スケールフリー”という言葉は次数のベキ則に対する印象的な標語として扱う。

として選択的結合 (preferential attachment) の戦略に基づいて発展するネットワークのモデル (以下 PA とする) を与え、次数分布のベキ則を説明するとともに、そのモデルによって生成されたネットワーク上でのダイナミックスの研究を進めた。そこでこれによって示された「ランダム攻撃に対して高い耐性を示す」[5] [6] 「感染伝播に対する閾値が存在しない」[7] といった結果がインターネットにおいて持つ含意が議論されている。

インターネットトポロジを適切にモデル化しようとする研究は短くない歴史を持つ[8]が、複雑ネットワーク研究に基づくモデルは新しい観点を導入したのもとして高く評価できる。また、インターネットトポロジをそれだけのものとして単独にとらえるのではなく、自然・社会・人工物などに見られるネットワーク構造との関係において広い視野の元に研究することは、社会学・物理学その他の多分野交流という観点からも有益であろう[9]。しかしその反面、インターネットの技術的・社会的特殊性を捨象した議論に対しては、その有効性に関して批判も存在する[9] [10]。

本稿は、FFFによって示されたとされるベキ則に関する素

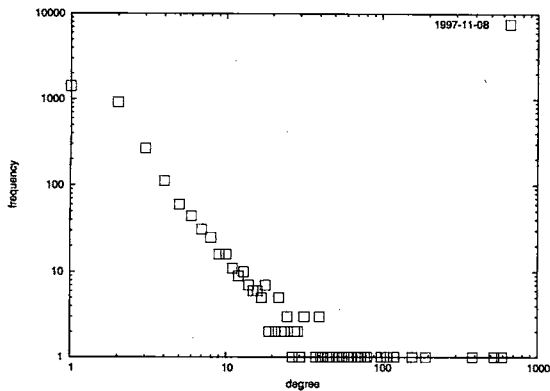


図1 1997年11月8日の度数分布

Fig.1 Degree distribution on Nov. 8, 1997.

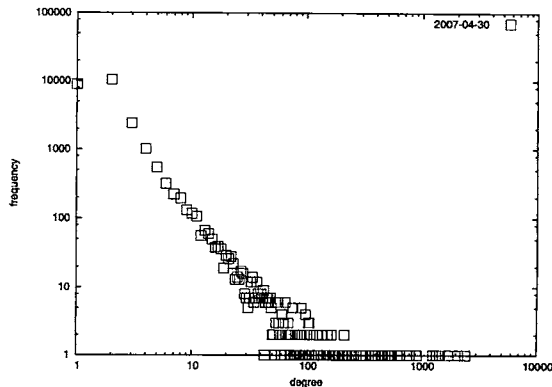


図2 2007年4月30日の度数分布

Fig.2 Degree distribution on Apr. 30, 2007.

朴な疑問から出発し、インターネットという現実の対象に即した観点からこれまでの複雑ネットワークの立場によるインターネットトポロジ研究が極めて不満足な状況にあることを議論する。さらにこれを踏まえて、今後進むべき方向について一つの提案を行う。

2. 発端

2.1 ベキ則

FFF では、National Laboratory for Applied Network Research [11] が収集公開したデータ^(注3)を元にした BGP 経路情報に基づく AS 間相互接続及び Pansiot and Grad [13] による traceroute に基づいたルータ間相互接続のトポロジが検討された。なお、Signaos-Faloutsos-Faloutsos-Faloutsos による 4 年後の論文 [14] は細部の修正を除き大筋に違はない。本節では主として AS 間相互接続に基づくトポロジについて考察する。

NLANR/Oregon Route-Views プロジェクトにおける BGP 情報の取得においては、フルルートを持つ一つのルータ (現在は route-views.routeviews.org.) を観測点とする。さらに、プロジェクトに協力する幾つかの他の AS と観測点ルータの間に BGP コネクションを確立して BGP による経路情報を受け取るにより、いわばインターネット上の複数の地点からの AS-Path を観測することができる体制をとっている。観測点ルータにおいて、定期的に BGP 経路表を取得する。その結果が生データとして公開されている。

AS 間相互接続のトポロジを調べるため、この生データから AS Path 情報を取り出す。AS Path 中、AS 番号が相前後して現われる場合、それらの AS 番号に対応する AS は隣接しているものと見なす。AS を頂点とし隣接する AS 間にリンクを張ることによって無向グラフを作成する。このグラフの各頂点の次数 (頂点につながったリンクの数) を求め、度数分布をプロットする。1997 年 11 月 8 日の観測について上記の操作を行った結果が図 1 である。(これは FFF の図 5(a) に対応するものであ

るが、ここでは新たに作成した。) ここで、X 軸 Y 軸ともに対数表示していること及び分布が右肩下がりの直線により近似されることより、

$$\text{次数 } d \text{ の頂点の出現頻度 } \propto d^{-\alpha}$$

($\alpha > 0$) と表されることが期待される。これが、度数分布のベキ則 (の一つの表現) である。

FFF では、1997 年 11 月、1998 年 4 月及び 12 月のデータについて度数分布を求め、ベキ則が常に成り立つこと、 α の値はわずかにしか変動していないことを示し、この性質がインターネットにおいて普遍的に成立するものであると主張された。

なお同様にして 2007 年 4 月 30 日 22:00 の観測について度数分布を求めると図 2 となる。これについても、FFF と同様に目視によればベキ則が成り立っていると認めることができるであろう。やはり、度数分布のベキ則は普遍的に成り立っていると考えてよいのであろうか。

2.2 素朴な疑問

AS 番号の付与基準は [15] [16] [17] に与えられている。これらの基準では、AS は少なくとも 2 本の外部リンクを持つ (“multi-homed”) のが原則である。ところが、図 1, 2 は、次数 1 の頂点が極めて多数観測されていることを示している。本来次数が 2 以上のものがこのように観測されているのだとすると、この結果は $1/2 \sim 1/3$ 近くのリンクが観測されていないことを意味する。果たしてこのデータは、本来のトポロジの度数分布を忠実に反映したものと推定することが可能であろうか?

2.3 シミュレーション

BGP により AS 間のリンクを観測する際の効果を評価するため、シミュレーションを行った。基本的な手順は以下の通りである。

- (1) トポロジを生成する。
- (2) 観測点を設定する。
- (3) n 個のプロープを設定し、観測点とリンクを張る。
- (4) BGP プロトコルを動作させ、BGP ルーティングテーブルを生成する。

(注3) : 現在は Oregon Route-Views プロジェクト [12] に引き継がれている。

(5) 図1, 2を作成したのと同じ手順により, トポロジを再現し, 度数分布を求める。

2.3.1 トポロジ

トポロジ生成には, インターネットのシミュレーションに用いられることが多いBRITTE[18]を利用した。BRITTEは, ASレベルのネットワーク及びルータレベルのネットワークを想定し, ランダムグラフやPAなど幾つかのアルゴリズムに基づくトポロジを生成することができる。

ここではまず, 図1, 2に見られる通りきわめて大きな次数の頂点が存在することに基づき, 度数分布が近似的にベキ則を示し, 大きな次数の頂点を生成することができるPAを基本アルゴリズムとして選択した。なお, BRITTEのデフォルト設定では生成されるネットワークにおいて各頂点の次数は全て2以上となる。インターネットのトポロジに関して妥当な設定である。

さらに, これに加えて観測されていないリンクが存在することを想定して, PAにランダムにリンクを加えたトポロジについても扱うこととした。

2.3.2 観測点とプローブ

観測点となるASを任意の一つ選び, さらにそのASと n 個の任意に選んだASとの間にリンクを張った。ここではこれをプローブと呼ぶ。Oregon Route-Viewsプロジェクトにおいて協力関係にあるAS及びそれとのBGPコネクションに対応する。 n を0から増やしていく時, 観測される度数分布にどのような変化があるかを見る必要がある。

2.3.3 BGPシミュレーション

生成したトポロジ場でBGPプロトコルを動かすために, ns-2をベースとするBGP++[19][20]を用いた。BGP++はzebraのBGP実装をns-2に移植したもので, 実運用に近いシミュレーションを行うことができるという特徴がある。ただし, ネットワークのサイズには制約があり(ns-2の実装に依存), 現実のインターネットのように大きなネットワークのシミュレーションを行うことはできない。今回行ったところでは, 頂点数200程度が限界であった。

なお, 今回のシミュレーションではネットワークにポリシを入れていない。従って, 各ASでの経路選択においてパス長最小の経路が選択される。この点は実際のインターネットとは大きく異なる。

2.3.4 シミュレーション結果

以下のシミュレーションでは頂点数は全て200である。

まず, PAから出発した場合について, 元のトポロジとBGPによって観測されたトポロジについての度数分布を図3に示す。BGPによって観測されたトポロジでは, 分布が全体に左方向に移動している, すなわち観測されなかったリンクが存在することがわかる。特に, 元のトポロジには存在しなかった次数1の頂点が多数観測されていることが注目される。また, プローブ数 n を増やすと, 元のトポロジに近づくことがわかる。

次に, PAにリンクをランダムに加えたネットワークについて, 同様のシミュレーション結果を図4に示す。元のトポロジは次数4付近に山ができており, ベキ則とは言えないが, プローブ数が小さい時観測されたトポロジの度数分布はベキ則と

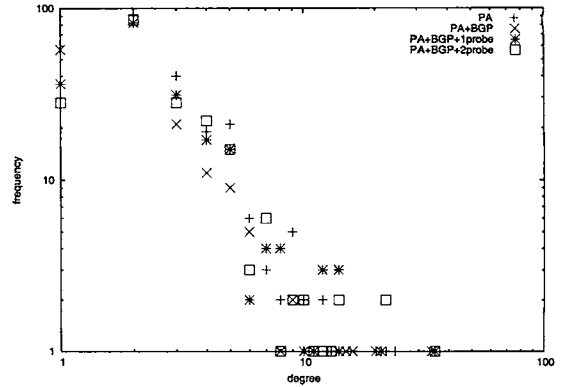


図3 PAに基づくBGPシミュレーション結果

Fig. 3 BGP simulation results based on the preferential attachment model.

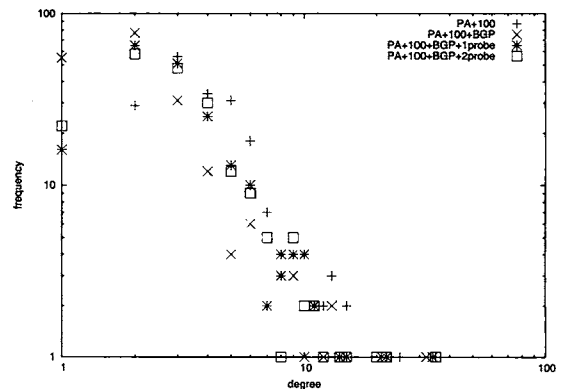


図4 ランダムリンクを加えた場合のBGPシミュレーション結果

Fig. 4 BGP simulation results with links added randomly.

いえる形になっている。この場合も, プローブ数を増やすと元のトポロジに近づくことがわかる(ただし, 完全に回復してはいない)。

3. 観測

前節のシミュレーション結果を踏まえて, BGPによる観測の特質について考察する。

BGPプロトコルの元で, 各ASは隣接ASから受け取った経路情報を元に各デスティネーションIPプレフィックスに対する最適経路を選択し, その最適経路を隣接ASに送出する。従ってある一つのASから見た場合経路情報のAS-Pathにはループはなく, 木構造をなす。AS間相互接続は冗長性のため複数経路を持たせることが普通であるが, このようにして観測されるトポロジが木構造であるということは, 観測されないリンクが必然的に存在することを意味する^(注4)。

(注4): 本稿ではBGPによるAS間相互接続のトポロジを主に扱っているが, tracerouteによるリンクの観測については大きなバイアスがあり[21], ランダ

このことは当然に予測されていたことであって、Oregon Route-Views プロジェクトでは複数のプローブを置き、いわば複数の地点から見た木構造を重ね合わせることによってこの問題に対処しようとしている。シミュレーション結果からも分かる通り、この方策は有効であって、複数のプローブを置くことは確かに未観測のリンクを減らす効果がある。しかしどれだけの規模のネットワークにおいて幾つのプローブを置けば十分かは全く未検討である。

プローブに関して言えば、数だけでなくどこに設置するかも重要な要素であるはずである。シミュレーションではランダムに選択したが、Oregon Route-Views プロジェクトではその性質上米国の研究ネットワークや IX、主要 ISP 中心の配置となっている。もとより、設置場所の最適な選択はまさしく観測対象となるネットワークのトポロジに依存するはずのものであり、予め適切な設置場所を指定することができるような性質のものではない。

以上を総合すると、BGP 経路情報に基づくトポロジの推定は、精度が不明でキャリブレーションも行われていない装置による観測結果に基づくものであり、十分信用できるものとは言えない。

なお、それにしても図 1, 2 に見られるようなきれいなグラフがある以上、ベキ則を考えることに意味があるのではないかという立場もあり得るが、ベキ則かどうかの認定をグラフに基づいて行うことは非常に危険であり、かつ、ベキ則は正規分布ほどではないにせよそれほど“特別な”分布ではないことから、このことに重要性を与えることは適当ではない。これらのことについては、付録で略述した。

4. モデル

観測された度数分布その他の指標の原因を探るため、さまざまなモデルが提案されている。既に前節で利用した PA もそのようなモデルの（再初期の、かつ代表的な）一つである。また、これらのモデルによって生成されたトポロジを用いて、ネットワーク上のダイナミクスの研究が行われている。本節では、このような研究の方法が果たして妥当かどうかについて検討する。

前節で既に AS レベルのトポロジの度数分布がベキ則であるとは言えないことを示したが、仮にベキ則であるとしても付録で述べたようにモデルに対する制約としては十分とは言えない。

また、PA では古い（初期条件として与えられた、または成長の初期に加えられた）頂点が度数大となり、度数が減少することはない。かつ度数大の頂点はネットワークの“中心”に位置する。

ところが、明らかに全く異なった形状のトポロジが同一の度数分布を示すことがある [23]。実際のインターネットとの対応で見ると、ルータレベルのネットワークでは最大度数を持つものは ISP において顧客へのリンクを収容するエッジルータであり、“中心”が高い度数を持つ PA とは対応しない。

これに対し、AS レベルのネットワークでは度数大の頂点は大規模な ISP でありネットワークの“中心”に位置すると考えることもできそうである。Oregon Route-Views プロジェクト開始以来最大度数の AS は AS701 (UUNET) である。残念ながら 1997 年以前にさかのぼっての AS レベルのトポロジに関する系統的な情報は存在しないが、ISP である AS701 が最も古い AS の中に含まれる可能性は低い。また、度数 2 位以下の AS の順位は少なからず変動している（急速に順位を下げた AS もある）ので、これについても PA の性質は適合しない。

モデルの適合性がこのような状況である以上、モデルに基づいたダイナミクスの議論に信用が置けないことは明らかである。

「ランダム攻撃に対して耐性がある」[5] [6] とされる理由は、攻撃対象としてランダムに頂点を選んだ場合殆どの頂点は度数が小さくかつネットワークにおいて周局的であるからネットワーク全体の結合性に大きな影響を与えないという事情に基づいている。裏返すと度数大の頂点を攻撃すれば脆弱だという結論（“アキレスの踵”論）[5] にもつながるが、ルータレベルのネットワークに関してこの結論が適用できないことは上記の観察から明らかである。AS レベルのネットワークに関して、AS 全体が攻撃を受けて停止してしまうということはどういうことであるか不明であり、あまり意味を持たない議論である。

なお「感染率に対する閾値が存在しない」[7] ことについては、ルータや AS がウイルスに感染したりウイルスを伝播させたりするのではないから、本稿の議論には適合しない。

5. 考察

現状のトポロジ観測及びモデルが満足できない状況であることを確認した上で、今後どのように研究を進めるべきかについて考察する。

5.1 「真の」トポロジ?

一つの考え方は、真のトポロジに迫るべくより精密な観測方法を開発し、モデル選択に利用できるより詳細なパラメータを発見するよう努めることであろう。

例えば Oregon Route-Views プロジェクトの場合であれば、プローブを増やし、より多数の経路情報を得ることができれば、いずれは真のトポロジに十分近いトポロジを得ることができるかもしれない。しかしこの場合、本当に真のトポロジに近づいているかどうかを判定する方法が存在しないため、確証を与えることは難しい。^(注5)

あるいは、文献 [24] にあるように whois データ等他の種類のデータを加えることによって総合的にトポロジをとらえることも考えられる。しかしこの場合は相矛盾する情報があつた場合どのような処理するのか、時間的な精度や信頼性に相違がある多種の情報をどのように扱うのがよいかなどの問題があり、決定的な結論に至ることは難しいように思われる。

しかしながら改めて考えてみると、我々がネットワークトポロジを研究する目的は様々なダイナミクスがネットワーク上でどのように振る舞うかを知るためであり、はたしてどのような

ムグラフに対してもベキ則の度数分布を観測してしまうこと [22] が報告されている。BGP による観測についても事情は本質的には同様であろうと考えられる。

(注5)：実際にはプローブ数は近年やや減少傾向にあるようである。

目的にも用いることができるような一般的な「真の」トポロジなるものが存在するかどうか、疑問になる。

例えばプライベートピアリングを行うための AS 間のリンクは BGP による観測では見いだされないことが多いだろうが、特に大規模 ISP などを含むトラフィック分配を考える場合には必要な情報である。また、AS がスタブなのかトランジットサービスを行っているかの区別は単純なトポロジ情報からは失われているが、耐故障性を考える上では必要である（あるリンクがバックアップリンクとして使えるか否か）。当然、ポリシーの他にもダイナミクスを動かす上でリンクの容量や遅延などの情報が必要な場合もあるであろう。

また、物理的な攻撃・障害に対する耐性を考えるのであれば、AS レベルにせよルータレベルにせよ IP レイヤ以上の論理的なトポロジを考えただけでは到底議論できない。（この議論は本稿の本来の範囲を逸脱するが、ルーティングシステムとしてのインターネットの設計や性能評価に関心を持つ立場からは自然な懸念である。）

以上のように、汎用的・一般的な「真の」インターネットトポロジを求めるといふ方向は、必ず壁にぶつからざるを得ない。

5.1.1 応用指向性

上記より明らかになったことは、関心を持つダイナミクスに応じて、ネットワークに関する情報のうち何が必要かということはある、ということである。ネットワークについて議論する場合は、例えば大災害に耐性を持つネットワーク構築を行いたい、全国にテレビ番組を配信するために必要なバックボーン設計を行いたい、など具体的な目的があるはずである。このような目的を達成することを考えれば、

- (1) 目的を明確にする
- (2) 目的に影響を与えるネットワーク指標を特定する
- (3) その指標を必要な精度で観測・測定する

という手順を踏むべきである。特に、上記項目の (2) がこれまで欠けていたのではないと思われる。以前から行われている社会学的ネットワーク理論や近年の複雑ネットワーク研究の過程を通じてネットワークに関する指標・パラメータがさまざまに提案されているが、それらの指標・パラメータがダイナミクス等にどのように影響するかを定量的に求めた上で観測・測定を行うという姿勢は決して一般的なものであったとは言えない。今後応用上有意義な成果を得るためには、このような観点からの目的志向の計画的な研究が必要であろう。

6. 関連研究

Chen *et al.* [25] は、AS レベル、ルータレベルのトポロジに加え、WHOIS データを利用してベキ則次数分布の原因に迫っている。この論文では WHOIS データを利用した場合ベキ則とは言えない形になることを指摘しながら、それが必ずしも信用できるものではないという理由から重視していない。

D'Souza *et al.* [26] [27] はむしろ WHOIS データによる次数分布を積極的に認め、ベキ則とは異なった分布を説明するモデルを提案している。

いずれも、そのようにして得られた次数分布がダイナミクス

に対していかなる意味を持つのかについて踏み込んでいないため、有効性が限られたものになっている。

Chang *et al.* [28] は AS を tier-1, 2 等に分類することで BGP による次数分布推定の事後的な評価を行おうとする。しかし分類後も観測結果には大きなばらつきがあり、観測値を「真の」値に近づけるための補正には用いることができない。

Doyle *et al.* [29] はルータレベルのトポロジに関して PA のような確率的に生成されるネットワークとインターネットは根本的に異なった構造を持つことをある指標を用いて示している。また、スケールフリーネットワークという術語を定義し直して、インターネットはスケールフリーでないとして述べているが、ベキ則が成り立つことは認めているため、本稿とは異なる立場である。

7. 結 論

インターネットトポロジに注目が集まるきっかけとなった FFF 及び複雑ネットワーク研究者達の功績は大きい、ベキ則あるいはスケールフリーネットワークというキーワードに依存したトポロジに関する認識は大きく改める必要がある。

真のネットワークトポロジを得さえすれば全てが解決されるという幻想を捨て、個別具体性を踏まえつつインターネットの現実に応じた応用に適合した研究を進めるべき時が来たと考えべきである。

謝 辞

本研究は平成 18 年度科学研究費補助金（課題番号 18650011）の援助を受けて行われた。

文 献

- [1] M. Newman, A.-L. Barabási and D. J. Watts Eds.: "The Structure and Dynamics of Networks", Princeton University Press (2006).
- [2] M. Faloutsos, P. Faloutsos and C. Faloutsos: "On power-law relationships of the Internet topology", *Computer Communication Review*, **29**, 4, pp. 251-262 (1999).
- [3] A.-L. Barabási and R. Albert: "Emergence of scaling in random networks", *Science*, **286**, pp. 509-512 (1999).
- [4] R. Pastor-Satorras and A. Vespignani: "Evolution and Structure of the Internet: A Statistical Physics Approach", Cambridge University Press (2004).
- [5] R. Albert, H. Jeong and A.-L. Barabási: "Error and attack tolerance of complex networks", *Nature*, **406**, pp. 378-381 (2000).
- [6] R. Cohen, K. Erez, D. ben Avraham and S. Havlin: "Resilience of the Internet to random breakdowns", *Physical Review Letters*, **85**, 21, pp. 4626-4628 (2000).
- [7] R. Pastor-Satorras and A. Vespignani: "Epidemic spreading in scale-free networks", **86**, 14, pp. 3200-3203 (2001).
- [8] V. Paxson and S. Floyd: "Difficulties in simulating the internet", *IEEE/ACM Transactions on Networking*, **9**, 4, pp. 392-403 (2001).
- [9] D. Krioukov, F. Chung, kc claffy, M. Fomenkov, A. Vespignani and W. Willinger: "The workshop of Internet topology (WIT) report", *Computer Communication Review*, **37**, 1, pp. 69-73 (2007).
- [10] W. Willinger, R. Govindan, S. Jamin, V. Paxson and S. Shenker: "Scaling phenomena in the Internet: Critically examining criticality", *Proceedings of the National Academy of Sciences*, **99**, pp. 2573-2580 (2002).

- [11] N. L. for Applied Network Research Routing Data: <http://moat.nlanr.net/Routing/rawdata/>.
- [12] U. of Oregon Route Views Project: <http://www.routeviews.org/>.
- [13] J.-J. Pansiot and D. Grad: "On routes and multicast trees in the Internet", *Computer Communication Review*, **28**, 1, pp. 41-50 (1998).
- [14] G. Siganos, M. Faloutsos, P. Faloutsos and C. Faloutsos: "Power laws and the AS-level Internet topology", *IEEE/ACM Transactions on Networking*, **11**, 4, pp. 514-524 (2003).
- [15] ARIN: "ARIN number resource policy manual", <http://www.arin.net/policy/nrmp.html>.
- [16] RIPE NCC: "Autonomous System (AS) number assignment policies and procedures", ripe-389 (2006).
- [17] APNIC: "Policies for Autonomous System number management in the Asia Pacific region", APINC-094 (2002).
- [18] A. Medina, A. Lakhina, I. Matta and J. Byers: "BRITE: Universal topology generation from a user's perspective", Technical report, Boston University (2001).
- [19] X. A. Dimitropoulos and G. F. Riley: "Creating realistic BGP models", *Proceedings of MASCOTS'03*, pp. 64-70 (2003).
- [20] X. A. Dimitropoulos and G. F. Riley: "Large-scale simulation models of BGP", *Proceedings of MASCOTS'04*, pp. 287-294 (2004).
- [21] D. Achlioptas, A. Clauset, D. Kempe and C. Moore: "On the bias of traceroute sampling", *Proceedings of STOC*, ACM, pp. 694-703 (2005).
- [22] A. Clauset and C. Moore: "Traceroute sampling makes random graphs appear to have power law degree distributions", *arXiv:cond-mat/0312674* (2004).
- [23] L. Li, D. Alderson, W. Willinger and J. Doyle: "A first-principles approach to understanding the Internet's router-level topology", *Computer Communication Review*, **34**, 4, pp. 3-14 (2004). *Proceedings of ACM SIGCOMM'04 Conference*.
- [24] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, kc claffy and A. Vahdat: "The Internet AS-level topology: Three data sources and one definitive metric", *Computer Communication Review*, **36**, 1, pp. 17-26 (2001).
- [25] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker and W. Willinger: "The origin of power laws in Internet topology revisited", *Proceedings of IEEE INFOCOM 2002*, pp. 608-617 (2002).
- [26] R. M. D'Souza, C. Borgs, J. T. Chayes, N. Berger and R. D. Kleinberg: "Emergence of tempered preferential attachment from optimization", *Proceedings of the National Academy of Sciences*, **104**, 15, pp. 6112-6117 (2007).
- [27] R. M. D'Souza, C. Borgs, J. T. Chayes, N. Berger and R. D. Kleinberg: "Fitting the WHOIS Internet data", *arXiv:cs.NI/0701198* (2007).
- [28] H. Chang, R. Govindan, S. Jamin, S. J. Shenker and W. Willinger: "Towards capturing representative AS-level Internet topologies", *Computer Networks*, **44**, pp. 737-755 (2004).
- [29] J. C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka and W. Willinger: "The 'robust yet fragile' nature of the Internet", *Proceedings of the National Academy of Sciences*, **102**, 41, pp. 14497-14502 (2005).
- [30] M. Mitzenmacher: "A brief history of generative models for power law and lognormal distributions", *Internet Mathematics*, **1**, 2, pp. 226-251 (2003-2004).
- [31] M. E. J. Newman: "Power laws, Pareto distributions and Zipf's law", *arXiv:cond-mat/0412004* (2006).
- [32] V. Pareto: "Cours d'économie politique professé à l'université de Lausanne", Vol. 2, F. Pichou, Paris (1897).
- [33] R. Perline: "Strong, weak and false inverse power laws", *Statistical Science*, **20**, 1, pp. 68-88 (2005).
- [34] B. B. Mandelbrot: "Fractals and Scaling in Finance: Discontinuity, Concentration, Risk", Springer (1997).
- [35] B. V. Gnedenko and A. N. Kolmogorov: "Limit Distributions for Sums of Independent Random Variables, Revised Edition", Addison-Wesley (1968).
- [36] B. Mandelbrot: "New methods in statistical economics", *The Journal of Political Economy*, **71**, 5, pp. 421-440 (1963).
- [37] N. L. Johnson, S. Kotz and N. Balakrishnan: "Continuous univariate distributions", Vol. 1, Wiley, 2nd edition (1994).

付録 ベキ則の同定について

ベキ則の歴史は古い [30] [31].

統計学上ベキ則が目されるようになったのは、1897年に発表された Pareto による収入の分布に関する研究 [32] からである。その後様々な現象においてベキ則が見いだされたが、殆どの場合、分布の全体でなく一部においてのみ成り立つものである [33]。またこのこととも関連して、両対数グラフ用紙にプロットした場合区別が付きにくい他の分布（特に対数正規分布）と比較して、どれが正しいかについてさまざまな論争の種になってきた [34].

実際、いろいろな形で歴史上何度も提案されてきた "the rich gets richer" principle (preferential attachment もその最新のモードである) は単純な形では対数正規分布を導くが、反射壁を設ける等の条件を与えることによりベキ則を導くこともできる [30].

またこれ以外にも実に様々な道筋によりベキ則を導くことができることも知られている [31]. 確率的なプロセスによってではなく、技術的な最適性を求めることによって結果的にベキ則が導かれることもある。

多く見いだされるベキ則 ($\alpha \leq 3$) は分散が発散するという性質を持ち、このことが非常に大きな値の出現 ("long tail") と関連づけられることもあるが、数学的には、このような 2 次のモーメントが存在しないような分布についての極限定理が知られている [35]. それによると、多数の独立な確率変数の和がベキ則をそのグループに含むような一群の分布に収束することが示せる。周知のように 2 次のモーメントが存在する場合は中心極限定理になり、正規分布が極めてありふれたものであることの原因となっているが、同様にベキ則も決して特別なものとは言えないことがこのことから分かる [36]. すなわち、ベキ則はモデルを弁別するための指標としては弱いということである。

なお、ベキ則を同定するにあたって α の値が議論の対象となることがあるが、FFF が行っている最小二乗法による推定法は不偏でなく適切でない [37]. PA を始めとする多くのモデル y では漸近挙動を得た後 α を与えているが、論文 [33] に指摘されているように分布の一部のみを用いて α の値を決めることは不定性が大きく適切でない。

以上のように、ベキ則は取り扱いが難しく、モデルの選択に用いることは不可能ではないにせよかなりの困難があることは確かである。