

ネットワークの数理： つながりを解く

(2) 不平等性の数理

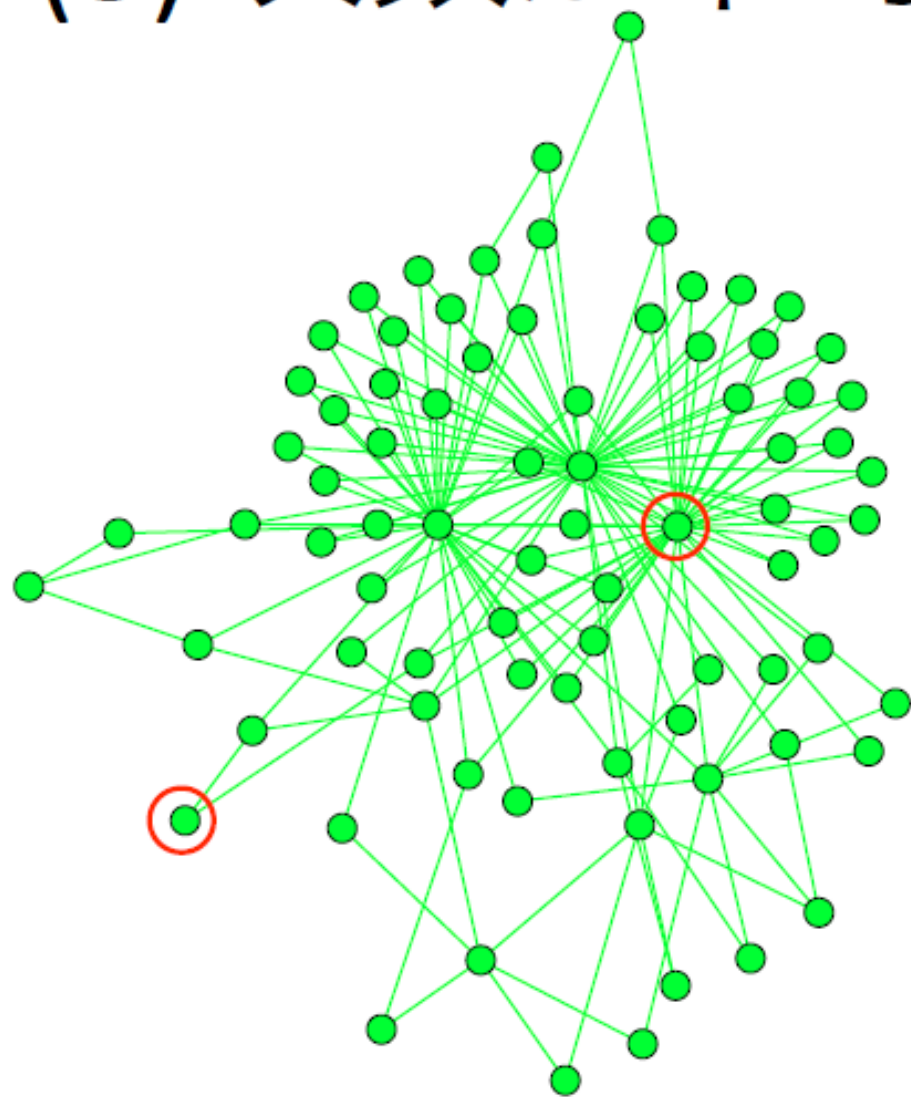
増田直紀
工学部 計数工学科

<http://www.stat.t.u-tokyo.ac.jp/~masuda>

※: このマークが付してある著作物は、第三者が有する著作物ですので、同著作物の再使用、同著作物の二次的著作物の創作等については、著作権者より直接使用許諾を得る必要があります。

しかし

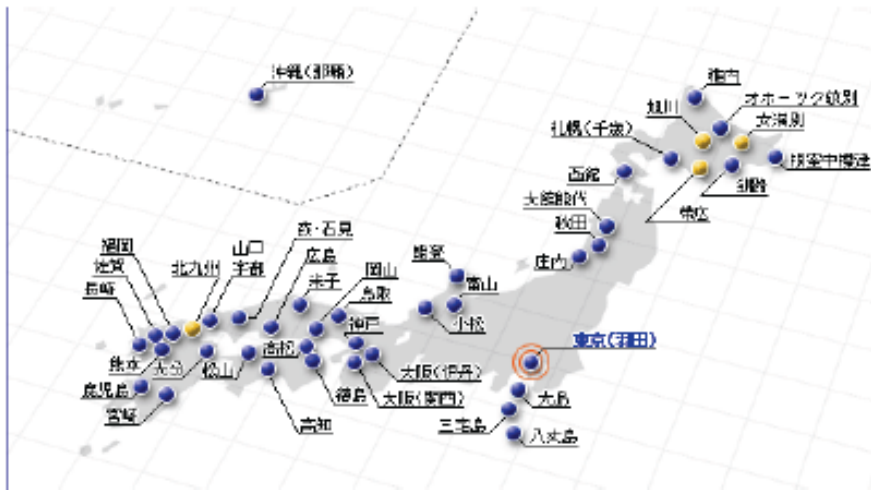
(3) 次数が不均一



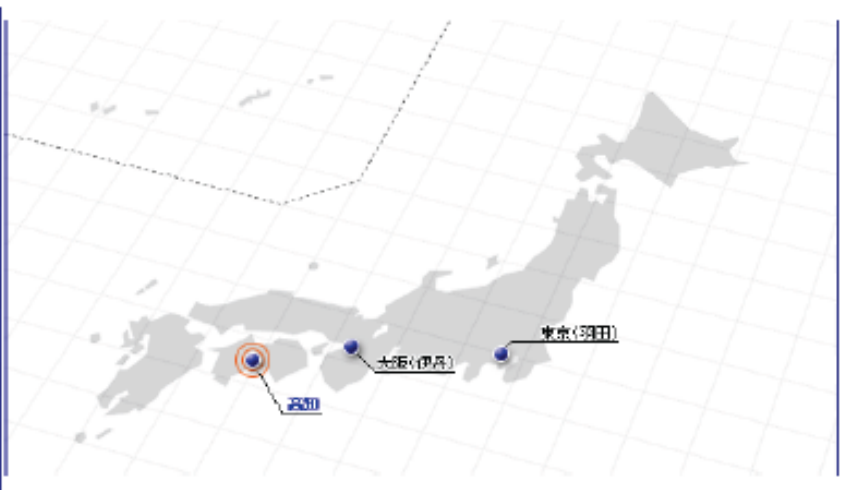
航空網のハブ

羽田空港はハブ空港

たいていの空港は小さい



✚ <http://www.ana.co.jp/dom/airinfo/networkmap/haneda.html>



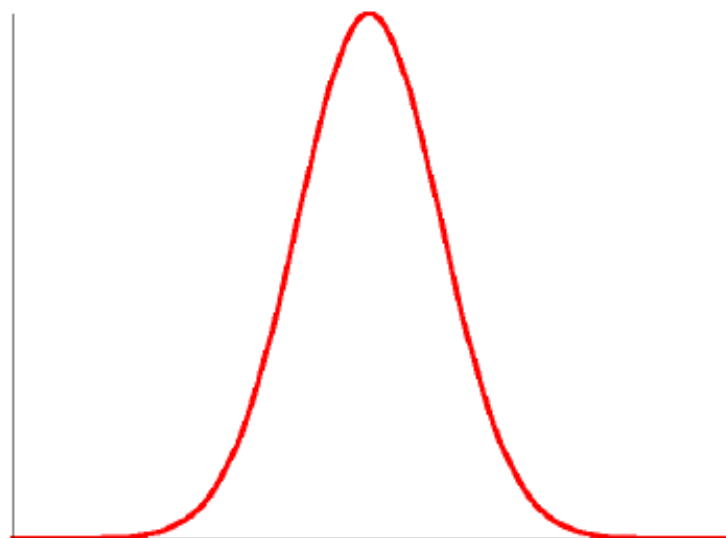
✚ <http://www.ana.co.jp/dom/airinfo/networkmap/kochi.html>

どっちが現実？

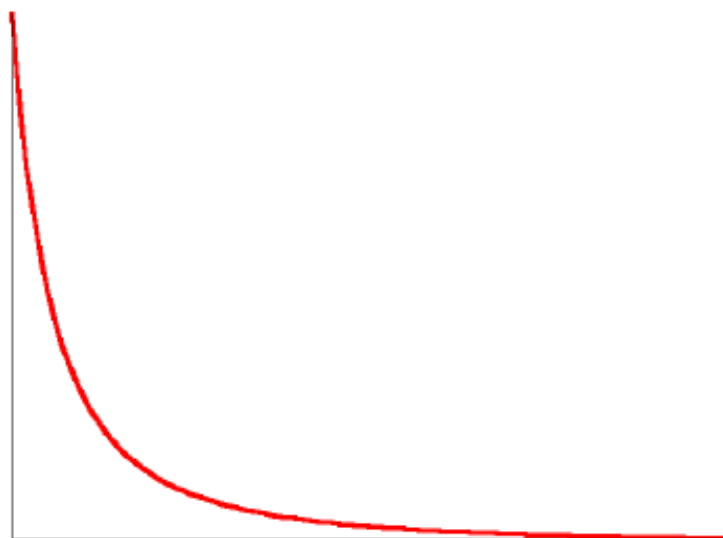
頻度

正規分布

べき分布（べき則）



大きさ



答：両方



しかし、あまり注目されて
いなかった！

べき則とは？

正規分布

$$p(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k - \langle k \rangle)^2}{2\sigma^2}\right)$$

$\langle k \rangle$: 平均

σ : 標準偏差

べき分布

$$p(k) = ck^{-\gamma}$$

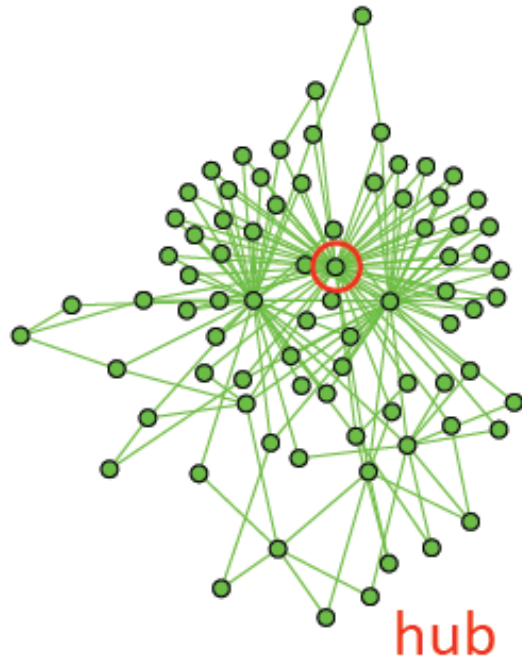
γ : べき指数

c : 定数 (重要でない)

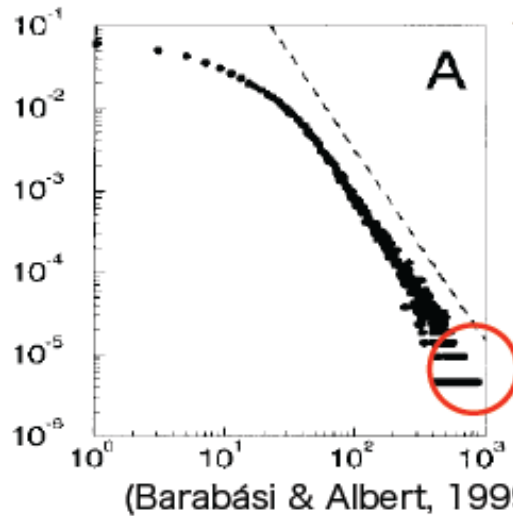


$$\log p(k) = \log c - \gamma \log k \quad \rightarrow \quad \text{両対数グラフで直線に乗る}$$

世の中の多くのネットワークは、 “スケールフリー・ネットワーク”



役者の共演関係



† Albert-László Barabási and Réka Albert (1999) Emergence of Scaling in Random Networks, *Science* 286(5439):509-512, p.510 Fig.1(A)

Internet

著作権の都合により、
ここに挿入されていた画像を削除しました。

M.Faloutsos et al.,(1999)
On power-law relationships of the Internet
topology, *ACM SIGCOMM Computer
Communication Review*
29(4): 251-262, p.255 Fig.4(a)

(Faloutsos et al., 1999)

$$p(k) = ck^{-\gamma}$$



$$\log p(k) = \log c - \gamma \log k$$

頻度

頻度

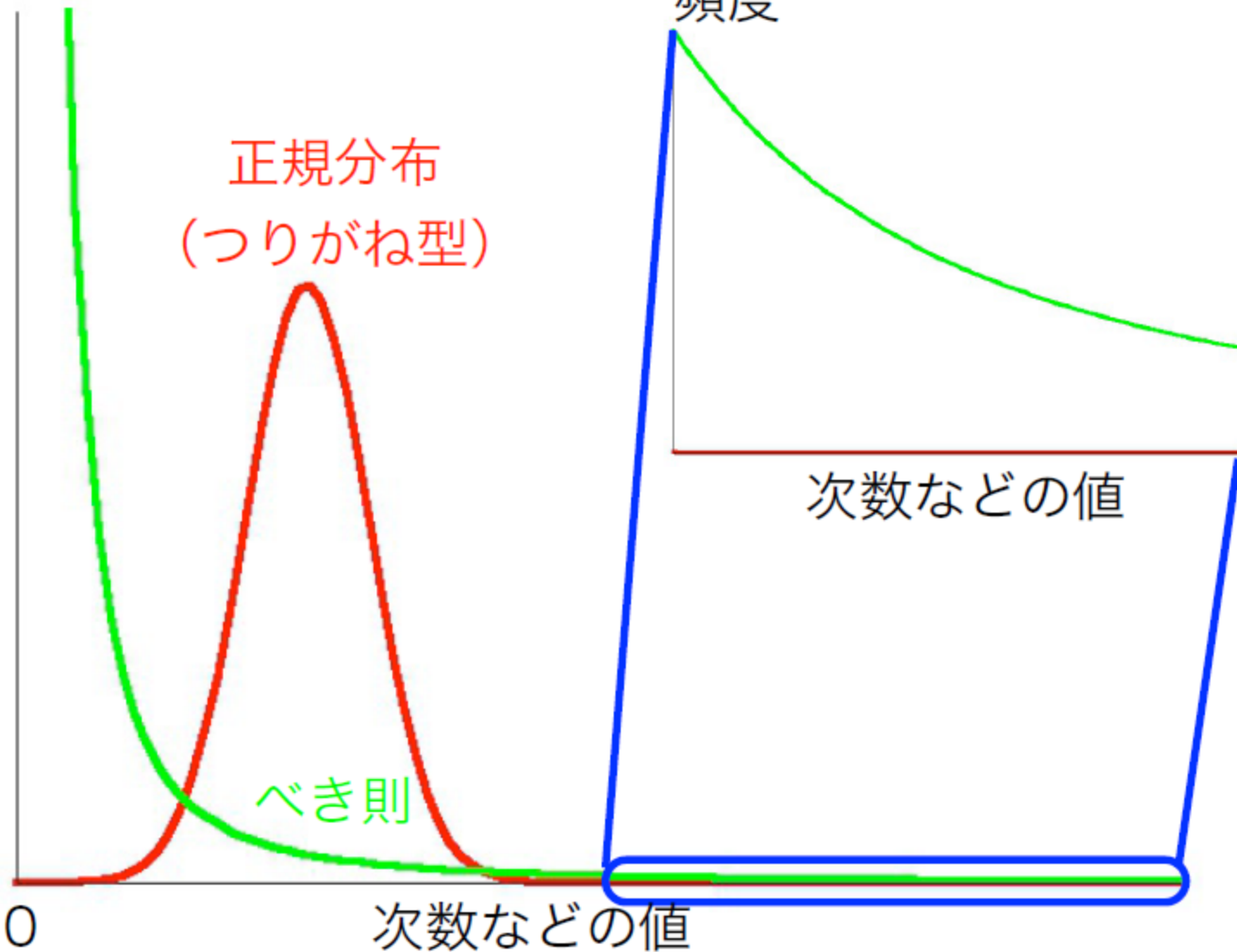
正規分布
(つりがね型)

べき則

次数などの値

0

次数などの値



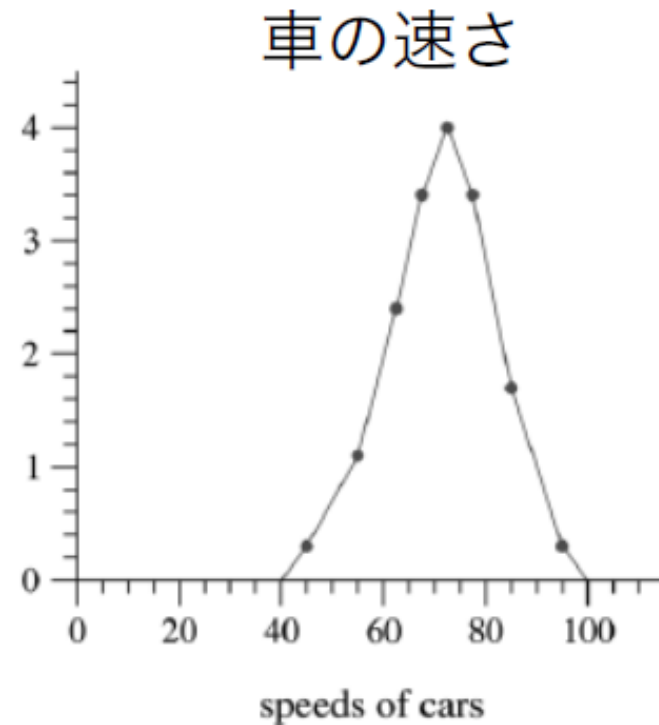
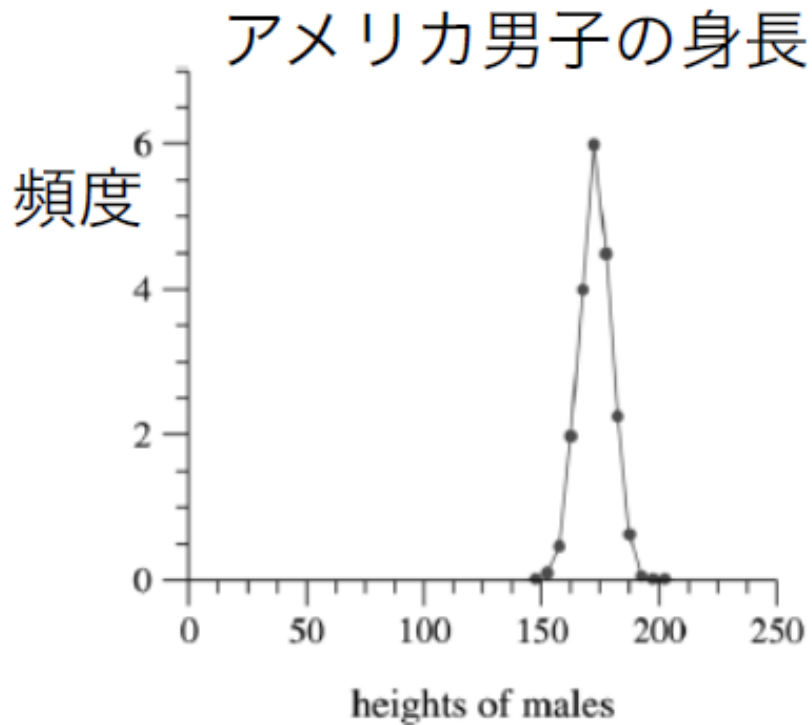
べき則の減り方

x	x^{-2}	正規分布（平均 1, 標準偏差 5）
1	1	1
2	0.25	0.9802
3	0.11	0.9231
5	0.04	0.7261
10	0.01	0.1979
100	0.0001	7.4×10^{-86}
1000	10^{-6}	3.0×10^{-8669}

← $x=1$ での値が 1 になる
ように「規格化」する

- 減り方が遅い

正規分布に近い量

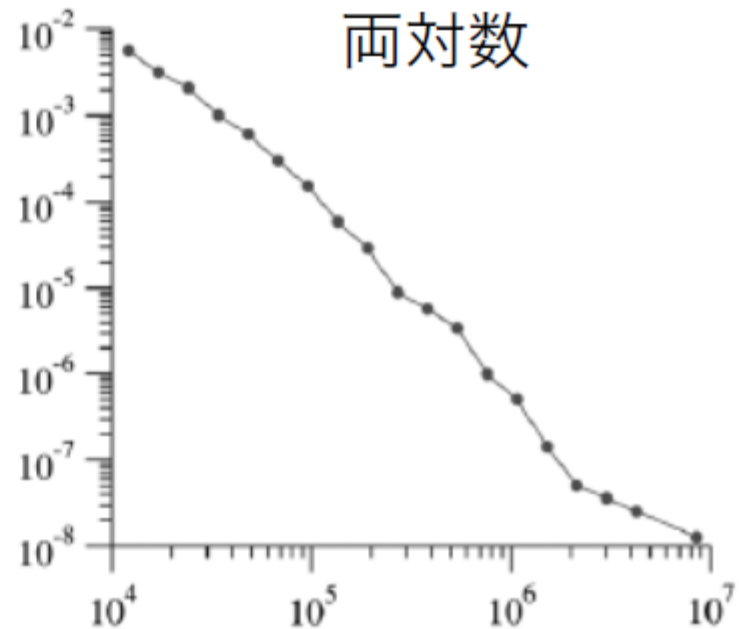
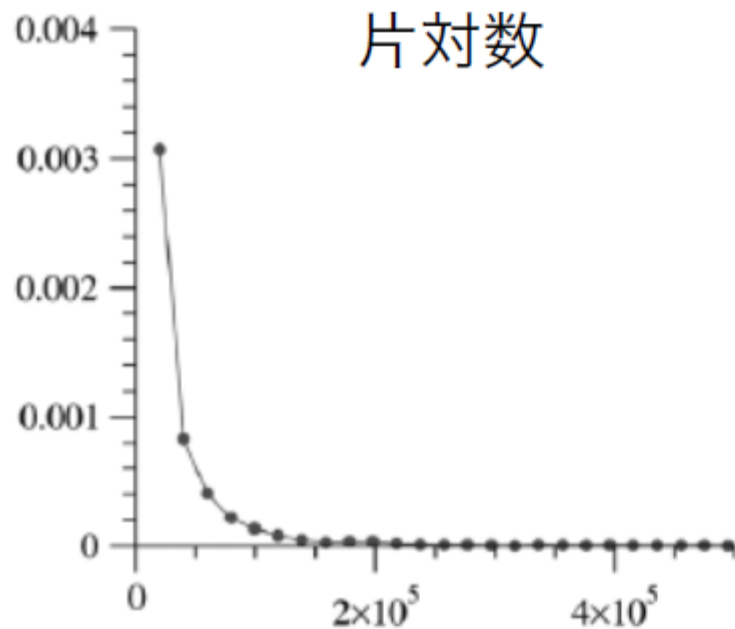


(Newman, 2005)



アメリカの都市サイズ (人口 ≥ 10000 人)

頻度



$$p(k) = ck^{-\gamma}$$



$$\log p(k) = \log c - \gamma \log k$$

(Newman, 2005)

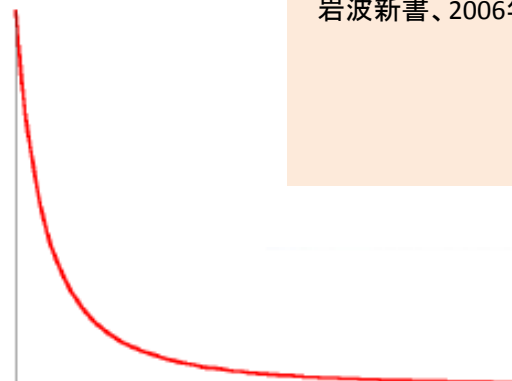


べき則の普遍性

- 収入 (Pareto の法則, 1896) ← 格差社会
- 都市の人口 (Zipf の法則, 1949; Auerbach, 1913)
- 研究者の書く論文数 (Lotka の法則, 1926)
- 本の売れ行き
- 土地の価格
- 会社の規模
- 単語の頻度
- ウェブページが訪問される回数
- 地震の規模
- 月のクレーターの大きさ
- 戦争の被害規模

佐藤俊樹
『不平等社会日本:
さよなら総中流』
中公新書、2000年

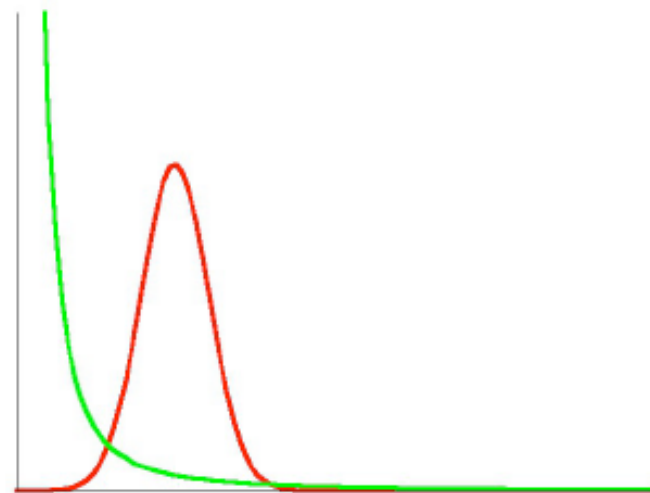
橘木俊
『格差社会:何が問
題なのか』
岩波新書、2006年



べき則の驚き (1)

平均から何倍も離れた人がある

- ・ 収入
 - ・ 平均年収 500 万円
 - ・ 1 億円プレーヤー
(いなくはない)
- ・ ひとり勝ち
- ・ 正規分布 (など) では説明できない.



$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \langle x \rangle)^2}{2\sigma^2}\right)$$

べき則の驚き (2)

平均がない

$$p(x) = (\gamma - 1)x_{\min}^{\gamma-1}x^{-\gamma} \quad (x_{\min} \leq x < \infty)$$

規格化定数 (確率なので和を1にする)

$$\langle x \rangle = \int_{x_{\min}}^{\infty} xp(x)dx = \begin{cases} \infty & (\gamma \leq 2) \\ \frac{\gamma-1}{\gamma-2}x_{\min} & (\gamma > 2) \end{cases}$$

期待値 (平均値)

べき則の驚き (3)

大半の人が平均以下

- 中央値 < 平均値
- 平均以下でも慌てない $\int_{x_{1/2}}^{\infty} p(x) dx = \frac{1}{2} \int_{x_{\min}}^{\infty} p(x) dx$

- 平均は（あるとしても）あまり意味を持たない
- $p(x) = (\gamma - 1)x_{\min}^{\gamma-1}x^{-\gamma}$
を代入すると

$$x_{1/2} = 2^{\frac{1}{\gamma-1}} x_{\min}$$

- 先入観を変えよう！

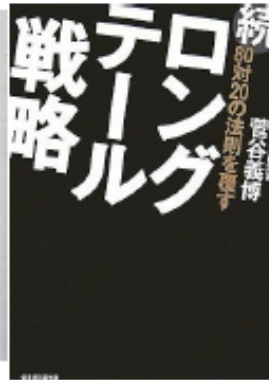
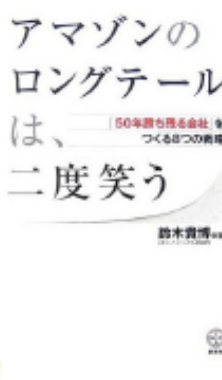
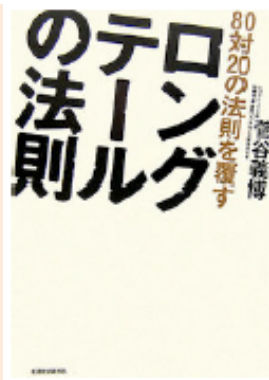
$$\gamma = 3 \implies x_{1/2} = 1.41 x_{\min}, \langle x \rangle = 2 x_{\min}$$

$$\gamma = 2.5 \implies x_{1/2} = 1.59 x_{\min}, \langle x \rangle = 3 x_{\min}$$

ビジネス応用

- 20%の商品から売上の80%, 80%の商品から売上の20%
- 90:10や70:30でもよい。
- **80対20の法則** → 20%の商品に注目
 - 在庫を整理すべき
 - 20%の仕事時間から80%の成果 → たいていの仕事はやめるべき
- **ロングテールの法則** → 80%の商品に注目（在庫コストがほぼ0）
 - 書籍 → 注文がきてから印刷
 - 音楽、映画（例：アップル）

リチャード・コッチ
『人生を変える80対
20の法則』
仁平 和夫訳、阪急
コミュニケーションズ、
1998年



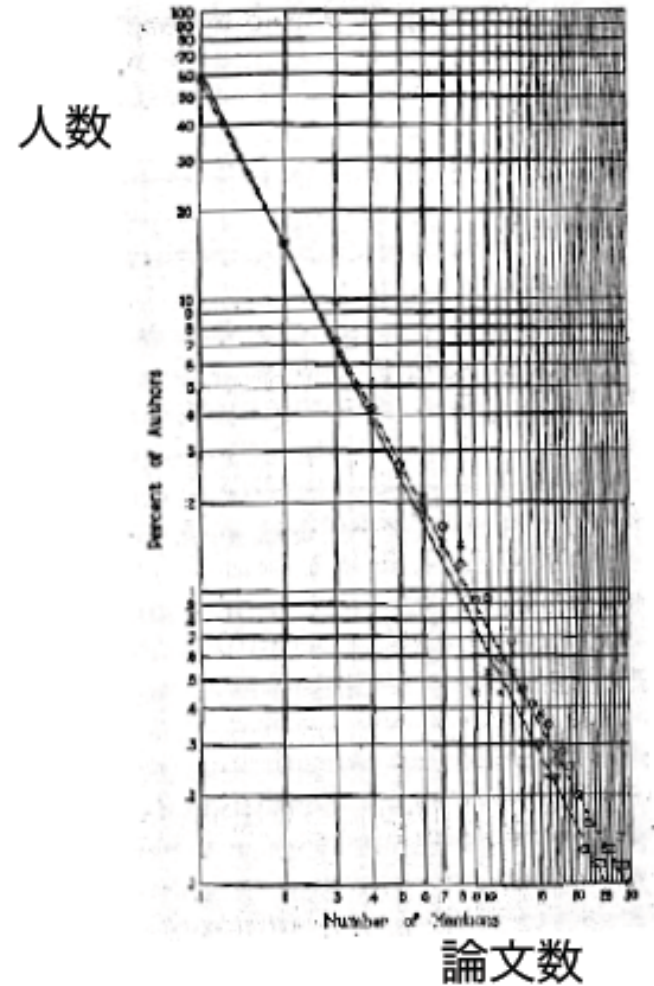
日本人は正規分布がお好き？

- 偏差値 → 正規分布
 - $= 50 + 10 \times (\text{得点} - \text{平均点}) / \text{標準偏差}$
 - 平均 = 50
 - 標準偏差 = 10 → 40～60 に 68%, 30～70 に 95% の人が入る.
- 直観的にべき則よりわかりやすい
- 平等的 → 日本人の精神に合う？



Lotka の法則 (1926)

- 仕事の生産性のべき則
- 一部の天才, 秀才の存在
 - 芸術, スポーツ, 科学
 - 仕事能力も?
- タブーなので触れにくい
 - 出る杭は打つ
 - 機会平等の勘違い

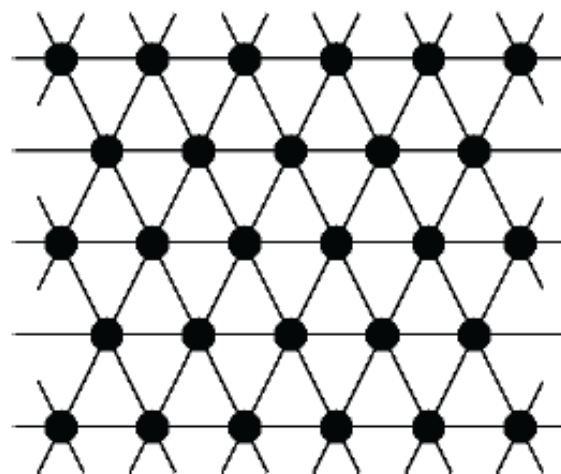
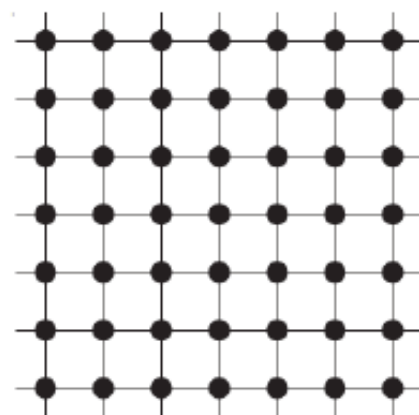
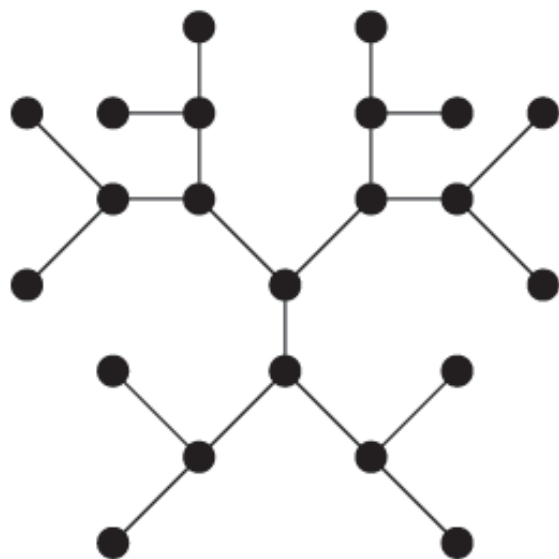


(Lotka, 1926)



Lotka, Alfred J.(1926) The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, 16 (12): p317-323, p.322 Fig.2.

格子や木では頂点次数は均一



ランダム・グラフでは

次数 二項分布 ポアソン分布

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}$$

where $\langle k \rangle = (N-1)p$ 平均次数

- べき則ほど裾野が長くない
- 正規分布の仲間

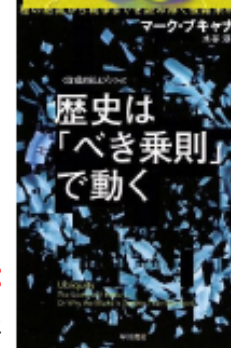
べき則が出る仕組み

- 相転移と自己組織化
 - 1990年代くらいの複雑系ブーム
 - でも、いつでもそうなのか？
- ある種の変数変換
- ランダム・ウォークの「再帰」時間
- 乗算過程
 - 金融データなど
- そもそも能力に差がある
- ...
- 教訓：1つの普遍則があるわけではない。
- 「統一」主義にだまされないように。



講談社、2009年

今野紀雄
『複雑系 (図解雑学)
第2版』
ナツメ社、2006年



早川書房、2009年



新潮社、2000年

(1) 変数変換

$p(y)$: 0 をまたいで分布する変数 y の密度

$$x = \frac{1}{y}$$

$\Downarrow$

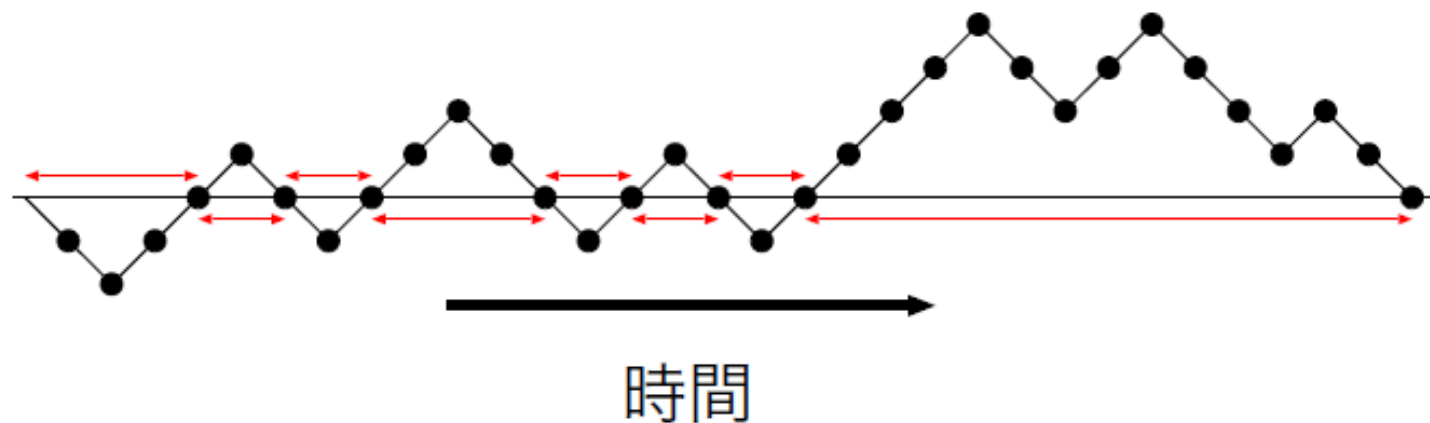
$$p(x) = p(y) \frac{dy}{dx} = -\frac{p(y)}{x^2}$$

x の大きい所では

$$p(x) \approx -\frac{p(0)}{x^2}$$

参考 : Newman (2005)

(2) ランダム・ウォーク



t : ウォーカーが初めて原点に戻ってくる時間 (偶数)

$$p(t) = \frac{\binom{t}{t/2}}{(t-1)2^t} = \frac{t!}{(t/2)!(t/2)!(t-1)2^t}$$

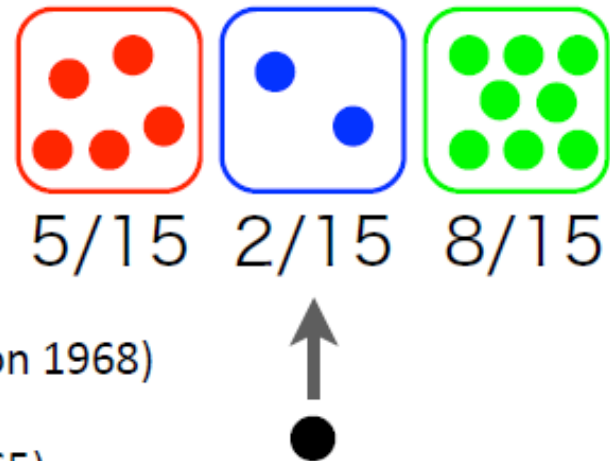
Sterling の公式 $t! \approx t^{t+\frac{1}{2}} e^{-t}$ を用いると

$$p(t) \approx \frac{1}{t^{\frac{1}{2}} \left(t - \frac{1}{2}\right)} \propto t^{-\frac{3}{2}}$$

例：睡眠

(3) The rich get richer

- 富める者はますます富む. 様々な別名
 - Yule process (1920s)
 - Gibrat 則
 - バンドワゴン効果
 - マタイ効果 (Simon, 1955; Merton 1968)
 - Cumulative advantage (Price, 1965)
 - 優先的選択 (Barabási & Albert, 1999)
- 都市の人口, 論文の被引用数, キーボードの配列, Windows vs Mac ...



Pólya の壺

- 問題：5個の白玉と3個の赤球の入った壺から1個ずつ球を取り出し、以下の条件のとき、4回目に白球が出る確率を求めなさい。

【条件】取り出した球が

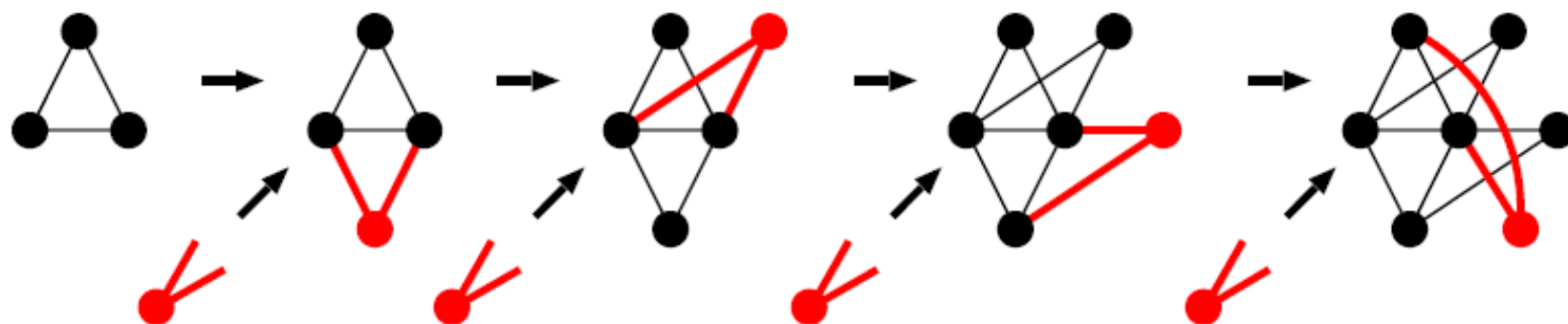
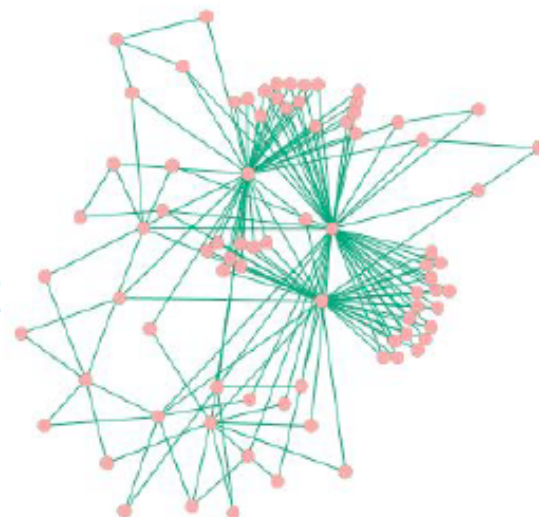
- ①白球⇒その白球と新たに白球を1個追加して壺に戻す
- ②赤球⇒その赤球と新たに赤球を1個追加して壺に戻す

(たけしのコマ大・数
学科の第152回目)

- 直観でわかること：貧富の差がつくこと
- 数学でわかること：具体的な法則

Barabási & Albert のスケールフリー・ ネットワーク(1999)

- スケールフリー = べき則
- ネットワークの成長 (インターネット, WWW, 航空網)
- 優先的選択 $\Pi(k_i) = \frac{k_i}{\sum_{j=1}^N k_j}$ k_i : 頂点 i の次数
 - 富めるものは富む
 - ハブはますますハブになる



BA モデルの性質

- スケールフリー： $p(k) \propto k^{-3}$
 - （この枠組みでは）優先的選択は必要
- L : 小さい
- c : 小さい, "現実" に合わない
- たくさんの変形版がある

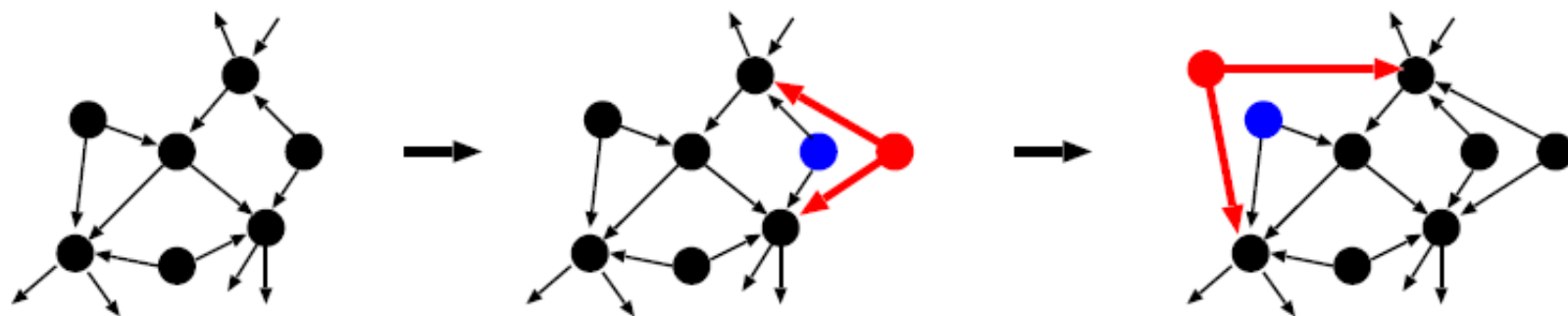
BA モデルの落とし穴

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^N k_j} \quad k_i: \text{頂点 } i \text{ の次数}$$

- N (頂点数) は, 通常大きい
- どうやって既存の頂点の次数を調べる?

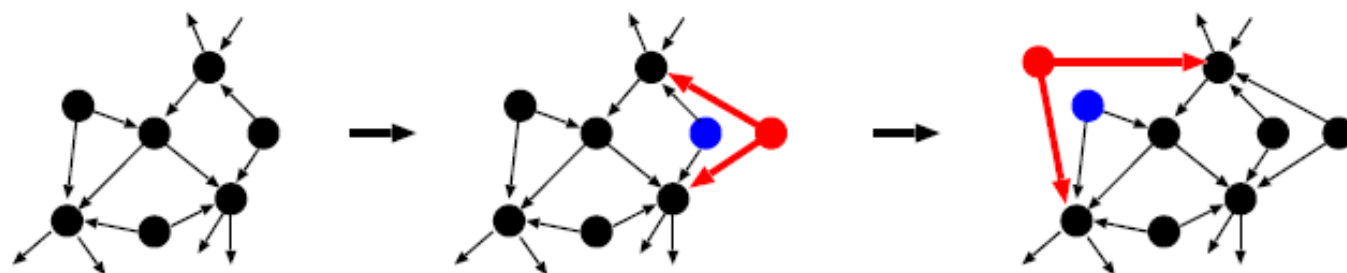
頂点コピーモデル

- 便宜上、枝に方向があるとする（無くてもよいようにできるが）。
- 新しい頂点は、ある「親」頂点を選び
 - 確率 $1-\alpha$ で、「親」の行き先を引き継ぐ
 - 確率 α で、無作為に行き先を決める



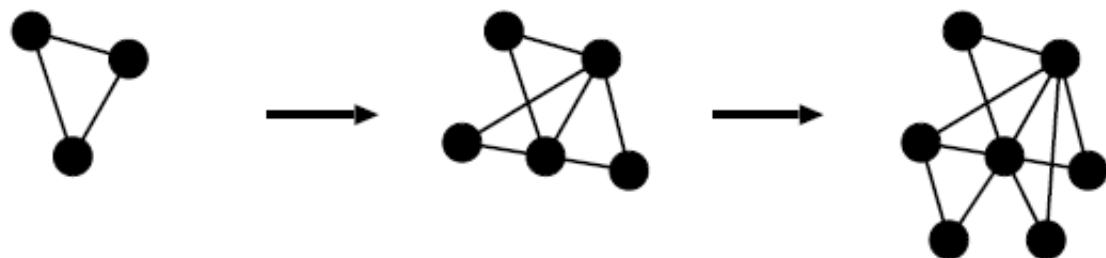
べき則になる！

- 隣にはハブがしやすい
 - ネットワークを歩けばハブに当たる
 - 自分の友人は自分より有名人
- タンパク質の相互作用ネットワーク (PIN)
 - 複製直後は、(タンパク質を作る) 新しい遺伝子と元の遺伝子は隣接点が全て同一
 - その後、突然変異を経て進化、多様化
- WWW
 - 新しいウェブサイトは、あるサイト (のリンク) をひな形から作られやすい
 - その後、徐々にサイトを書き換えて、リンク先も変えていく

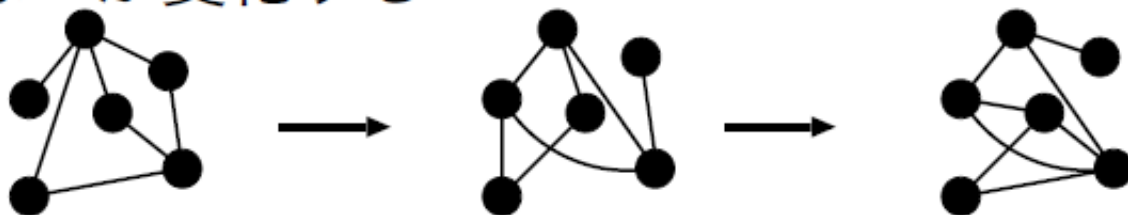


ネットワークは成長するとは限らない

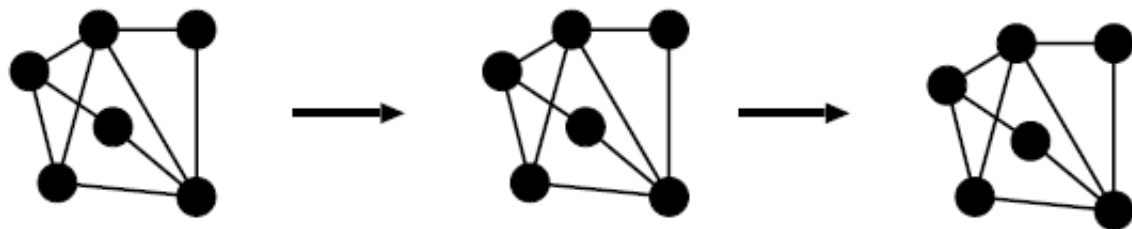
成長する



成長しないが変化する

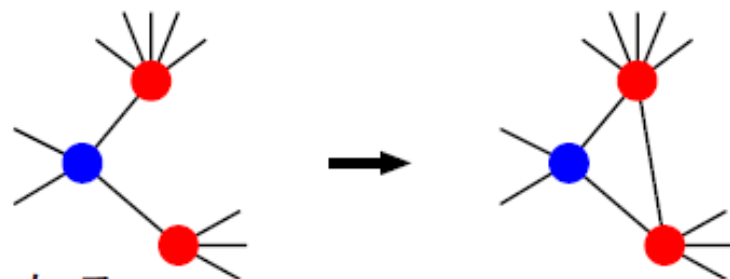


成長も変化もしない



知人紹介モデル

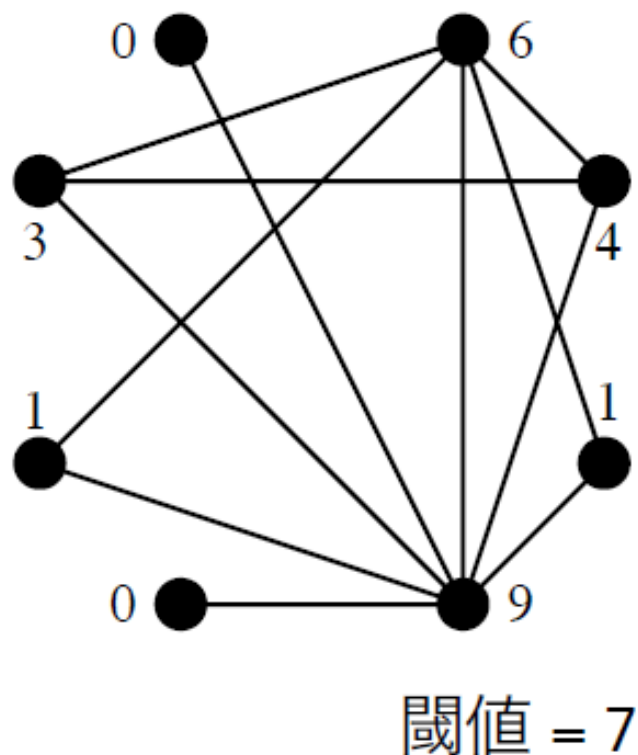
- 頂点数は固定
- 確率 $1-p$ で、ある頂点が自分の知人2人同士を紹介する
- 確率 p で、ある頂点を除去し、新しい頂点（と枝1本）を入れる
- p は小さいとする
- スケールフリー（風）になる
- 知人紹介を行うから三角形ができる。



(Davidsen, Ebel & Bornholdt, Physical Review Letters, 88, 128701, 2002)

閾値モデル

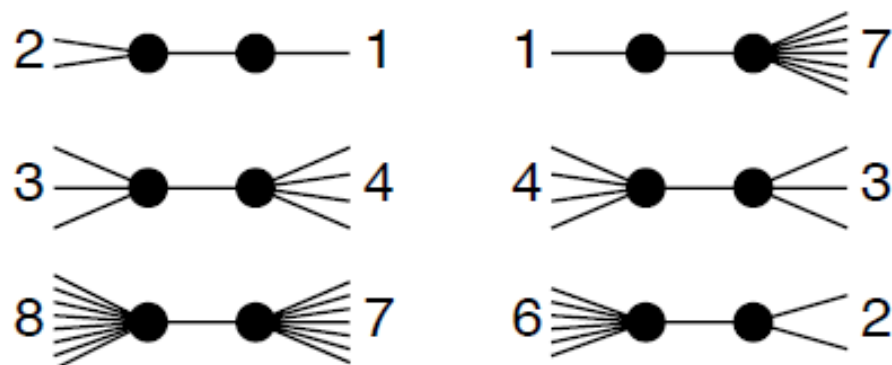
- i と j をつなぐ $\Leftrightarrow$
 $w_i + w_j \geq \text{“閾値”}$
- 成長しない
- スケールフリー
- w_i の振り方に
あまり寄らずに



ハブになるには？

- 先手を打つ
- 運
- そもそも能力が高い
- 放っておくと（次数の）格差は広がる

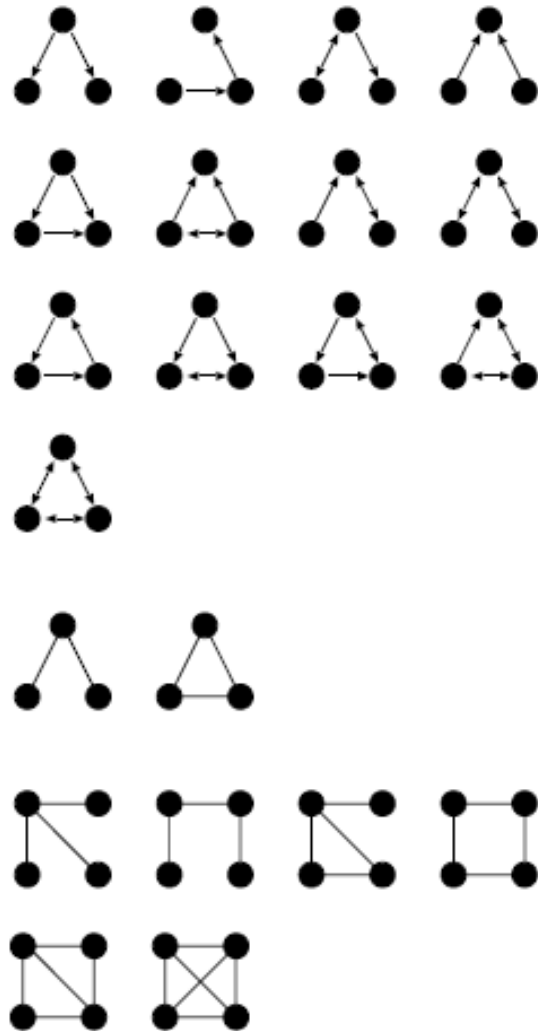
次数相関



- 正 (assortative)
 - 人間関係 ← homophily
- 負 (disassortative)
 - 生物系（食物網，神経，タンパク質）
 - 工学・情報系（Internet, WWW）
- しばしば，ネットワーク上の諸現象の行方を左右する

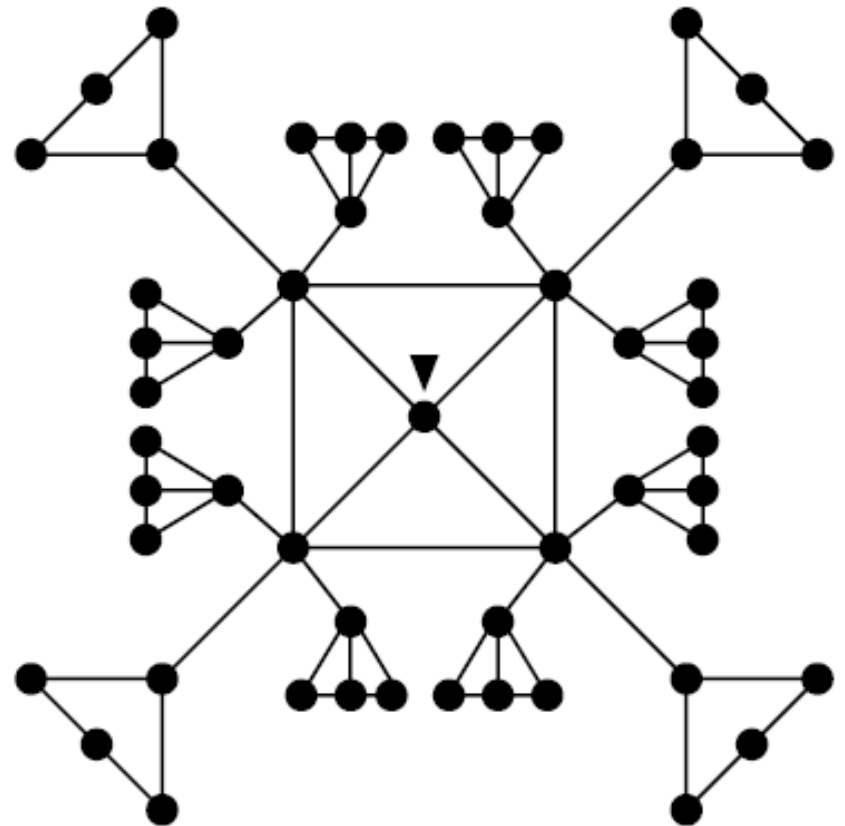
Network motifs

- 3 nodes $\rightarrow$ 13 motifs
- 4 nodes $\rightarrow$ 199 motifs
- 有向ネットに対して
使うことが多い

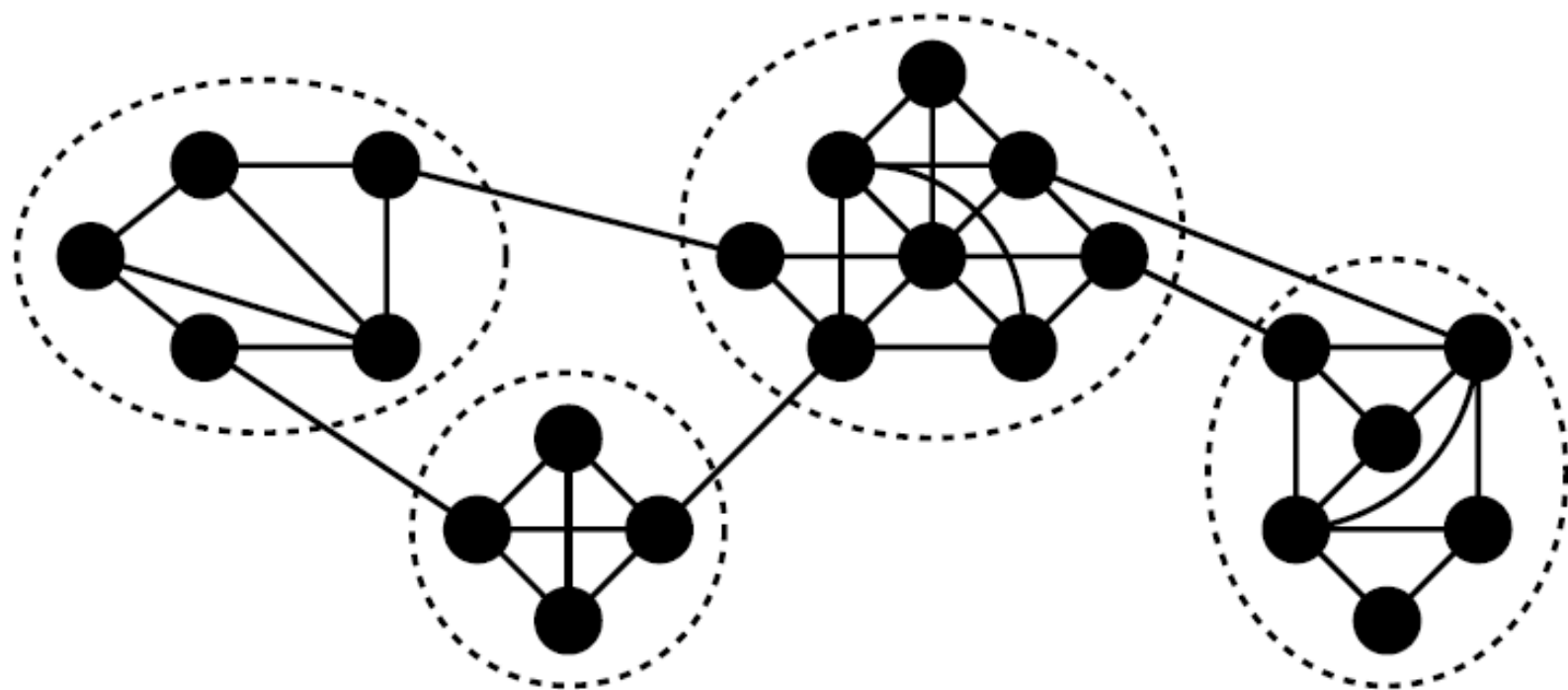


中心性

- 次数中心性
- 近接中心性
- 媒介中心性
- 固有ベクトル中心性
- ランダム・ウォーク（媒介）中心性
- サブグラフ中心性
- ページランク
- HITS
- ...
- 適切な定義は、応用による

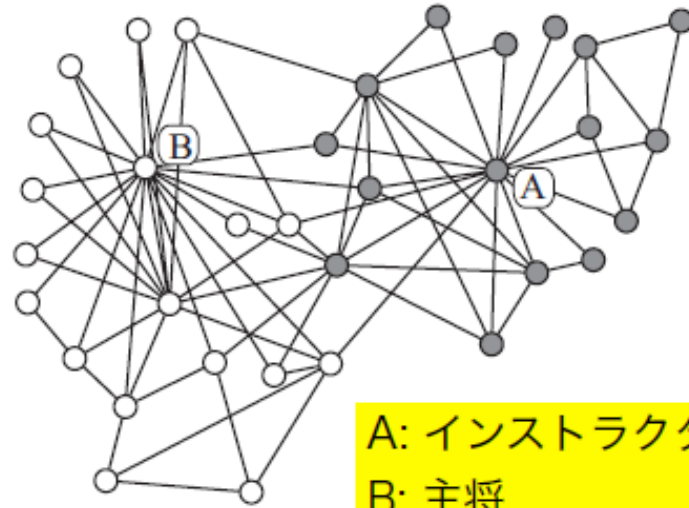


コミュニティ分割



Zachary の空手クラブ (1977)

- コミュニティ = グループ
 - 家族, 学級, 課, サークル
 - WWW のホスト, ドメイン
- 何のために分割?
 - 機能ごとのグループ分け
 - ともかく分類
 - 描画 (粗視化)
- 色々なアルゴリズム. 分割の答は 1 つでない
 - 社会学 (1950年代から)
 - 計算機科学
 - ネットワーク科学
- 計算量 (計算速度) もそれなりに大切



A: インストラクター
B: 主将

イルカ

著作権の都合により、
ここに挿入されていた画像を削除しました。

David Lusseau and M. E. J. Newman (2004)
Identifying the role that animals play in their social networks,
Proceedings of the Royal Society of London B, 271:S477-S481,
p.S478 Fig.1(a)

(Lusseau & Newman, 2004)

ネットワーク	頂点	枝	コミュニティ
空手クラブ	メンバー	友人関係	インストラクター 派, 主将派
ジャズ	ジャズ音楽家	バンドへの 所属	ニューヨーク 系, シカゴ系
共著関係	研究者	論文の共著	何々一派, 何々分野
単語ネット	英単語	同じ文脈で 使われる	科学系単語, 音楽系単語