

複雑ネットワーク

増田直紀

東京大学大学院 情報理工学系研究科
数理情報学専攻(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。

ネットワークの数理： つながりを解く

(1) 世の中のネットワークが持つ普遍的な構造

増田直紀
工学部 計数工学科

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

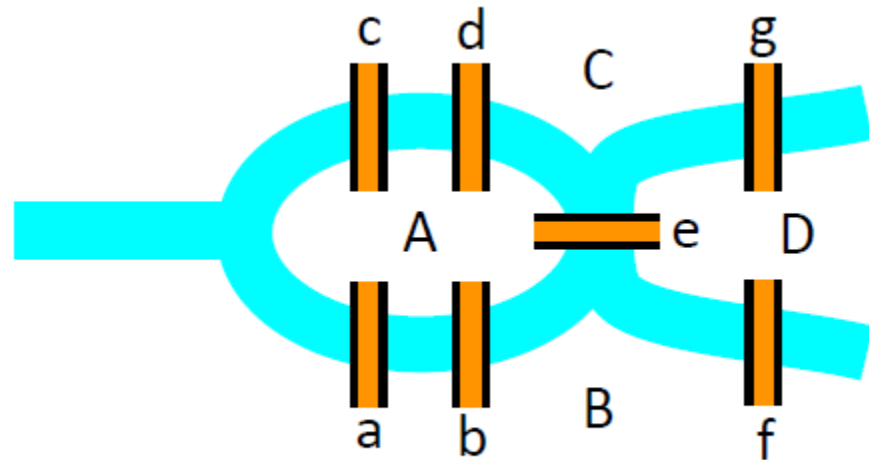
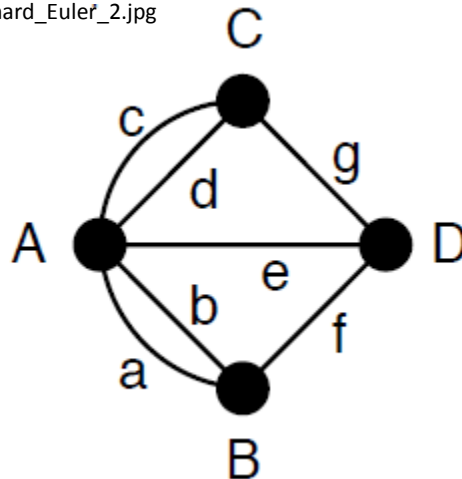
グラフ理論の誕生



(Leonhard Euler: 1707—1783)

Wikipediaより転載

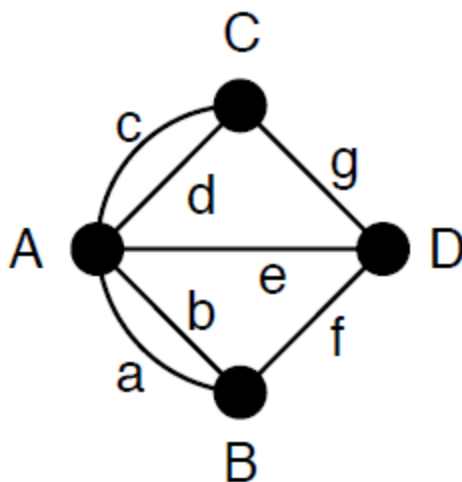
http://ja.wikipedia.org/wiki/ファイル:Leonhard_Euler_2.jpg



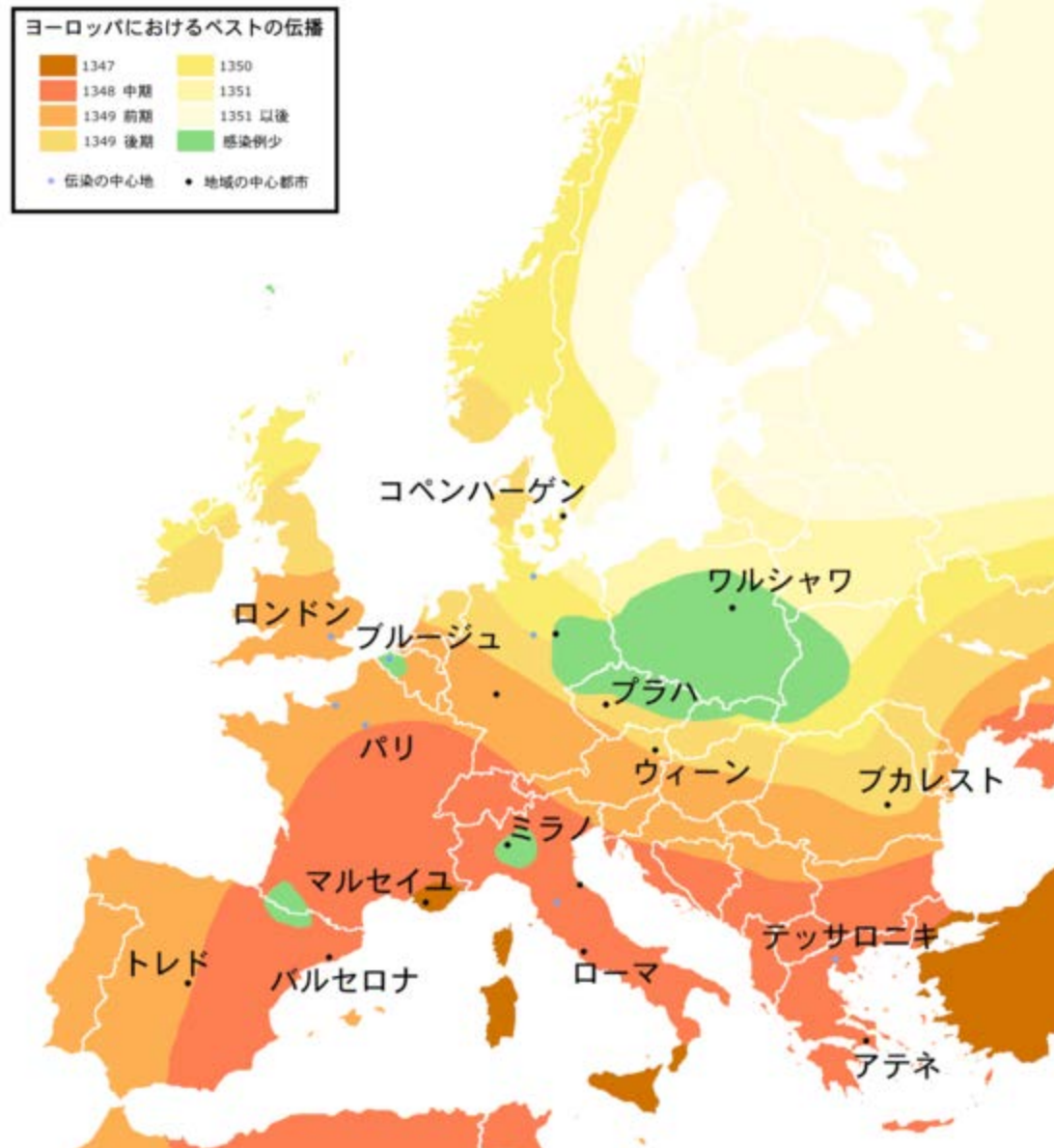
Königsberg (1736)

点と線

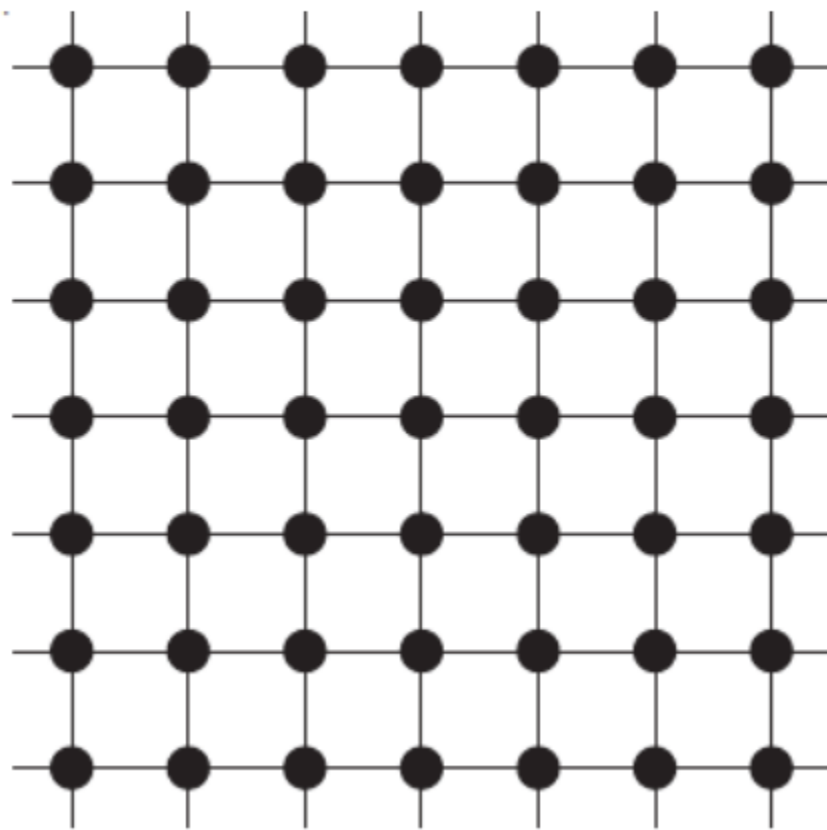
- 点 (●)
- 頂点, ノードなどと呼ばれる
- 線 (—)
- 枝, リンク, エッジ, 紐帯, 弧などと呼ばれる
- 二人関係で全てを表す
- 疎であることが通常の暗黙の仮定



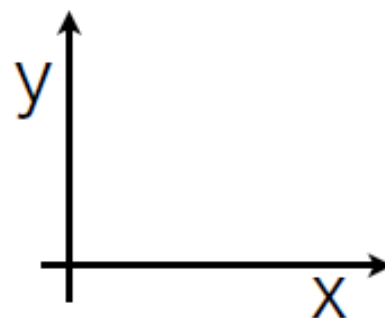
黒死病(ペスト)



正方格子



でも，偏微分方程式で記述できる

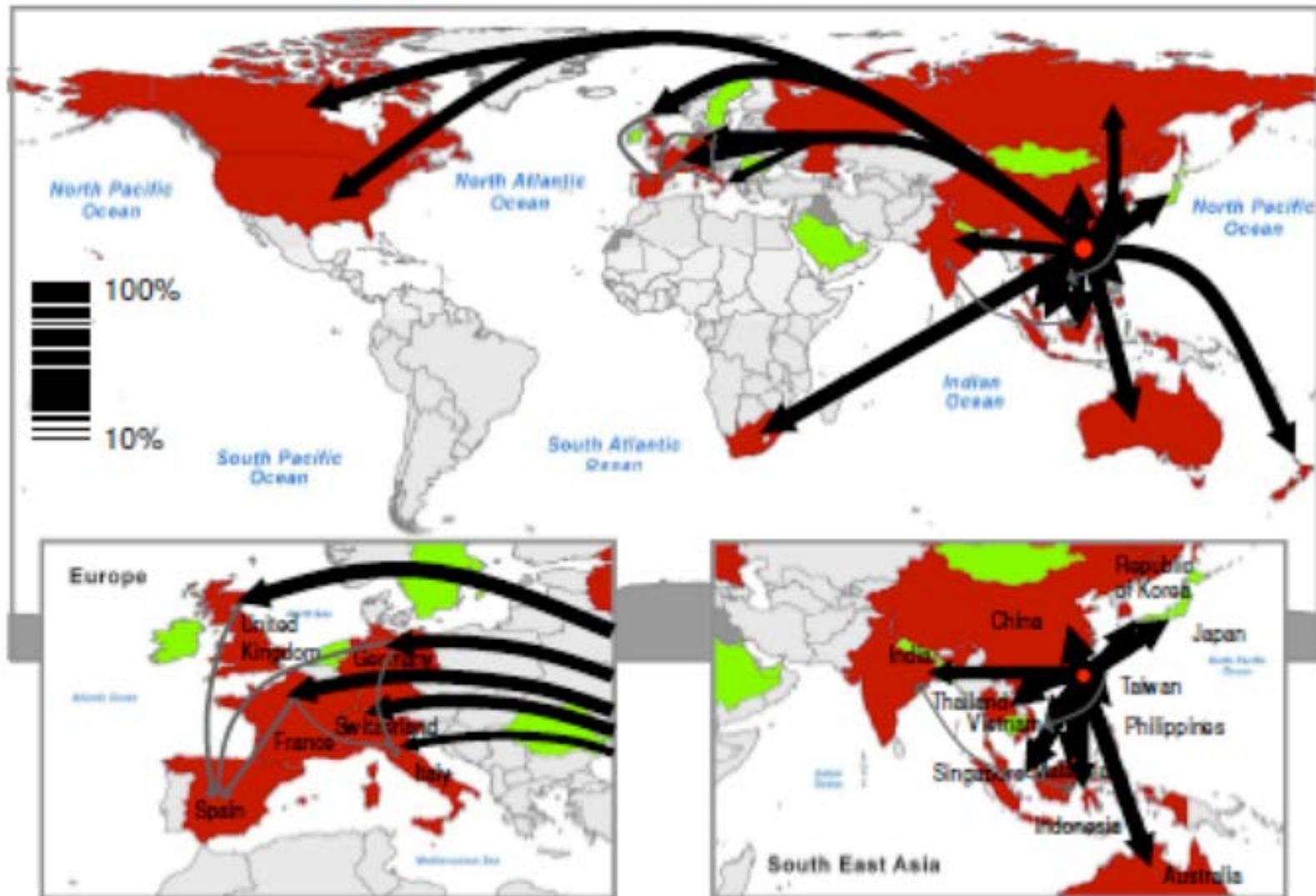


$$\begin{aligned}\frac{\partial S(x, y, t)}{\partial t} &= D \left(\frac{\partial^2 S(x, y, t)}{\partial x^2} + \frac{\partial^2 S(x, y, t)}{\partial y^2} \right) - \lambda S(x, y, t) I(x, y, t), \\ \frac{\partial I(x, y, t)}{\partial t} &= D \left(\frac{\partial^2 I(x, y, t)}{\partial x^2} + \frac{\partial^2 I(x, y, t)}{\partial y^2} \right) + \lambda S(x, y, t) I(x, y, t) - \mu I(x, y, t).\end{aligned}$$

人の動き

(Nobel, 1974)

ところが



†

Colizza et al. (2007) Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study, *BMC Medicine* 5:34, Fig.7

社会

空手クラブ

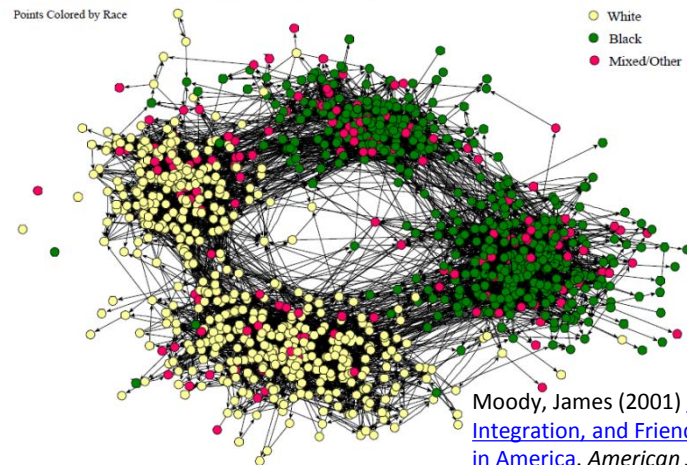


data from Zachary (1977)

アメリカの高校

The Social Structure of "Countryside" School District

Points Colored by Race



Moody, James (2001) [Race, School Integration, and Friendship Segregation in America](#). *American Journal of Sociology* 107(3) 679:716
Image courtesy of James Moody

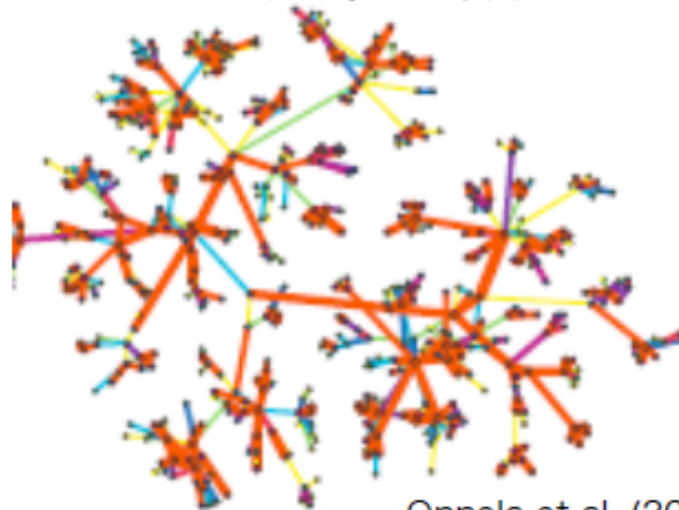
研究者の共著関係



Girvan & Newman (2002)

✚ M. Girvan and M. E. J. Newman (2002) Community structure in social and biological networks, *PNAS* 99 (12) : 7821–7826, p.7825 Fig.6.
Copyright (2012) National Academy of Sciences, U.S.A.

携帯電話

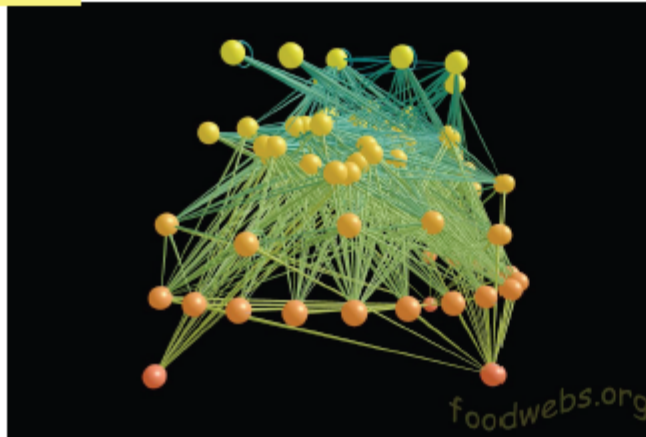


Onnela et al. (2007)

✚ J. P. Onnela et al. (2007) Structure and tie strengths in mobile communication networks, *PNAS* 104(18) : 7332-7336, p.7333 Fig.2(B).
Copyright (2012) National Academy of Sciences, U.S.A.

生物

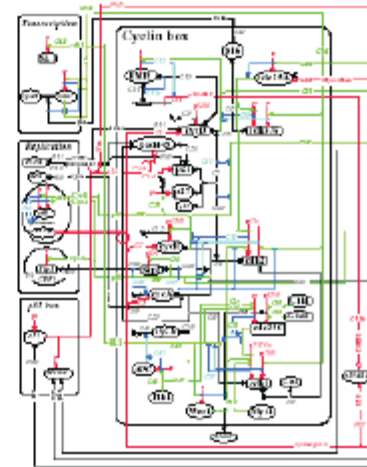
食物網



≠

Image produced with FoodWeb3D, written by R.J. Williams and provided by the Pacific Ecoinformatics and Computational Ecology Lab (www.foodwebs.org, Yoon et al. 2004)

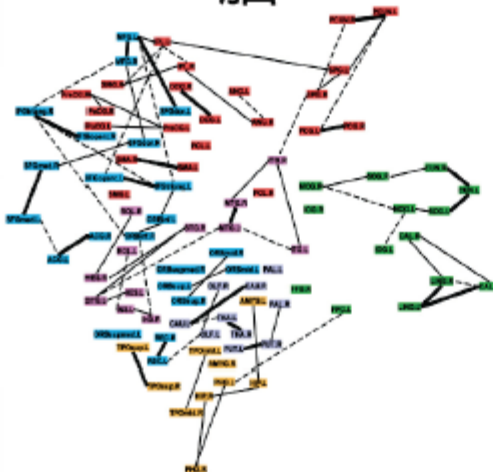
分子の相互作用



≠

Reprinted by permission from Macmillan Publishers Ltd: Steven H. Strogatz, *Nature* 410, 268-276, copyright (2001)

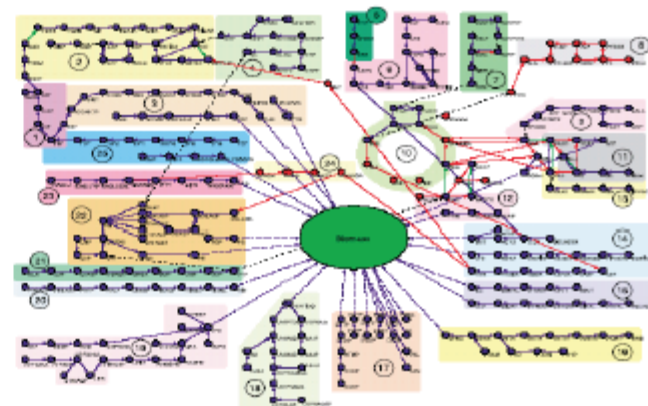
脳



≠

Salvador et al.(2005) Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain, *Cerebral Cortex* 15(9) :1332-1342, p.1340 Fig.4.

代謝



≠

Reprinted by permission from Macmillan Publishers Ltd: E. Almaas et al., *Nature* 427, 839-843, copyright (2004)

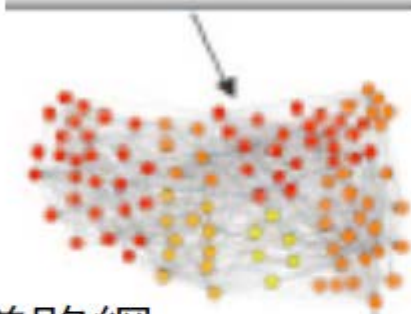
工学, 交通

航空網



✚

Vittoria Colizza et al.(2006)
The role of the airline
transportation network in
the prediction and
predictability of global
epidemics, *PNAS* 103(7) :
2015-2020, p.2017 Fig.2.
Copyright (2006) National
Academy of Sciences, U.S.A.



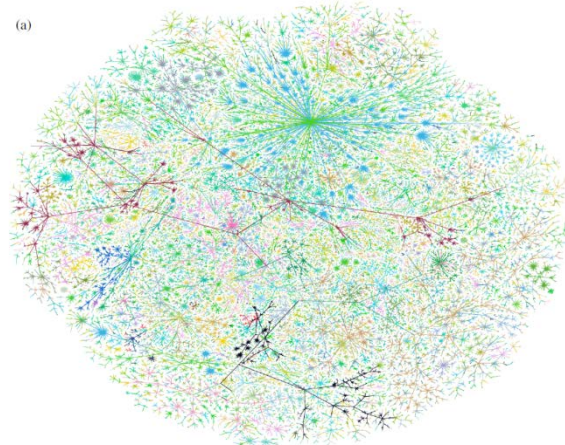
道路網



✚

Paolo Crucitti et al. (2006) Centrality measures in
spatial networks of urban streets,
Physical Review E 73(3) :036125, Fig.1(1)

インターネット



✚

Newman, M. E. J. (2003) The Structure and Function of Complex
Networks. *SIAM Review* 45 (2) : 167-256, p.170 Fig.1.2(a)
Image by Hal Burch and Bill Cheswick, courtesy of Lumeta Corporation

電気回路 (テレビ)

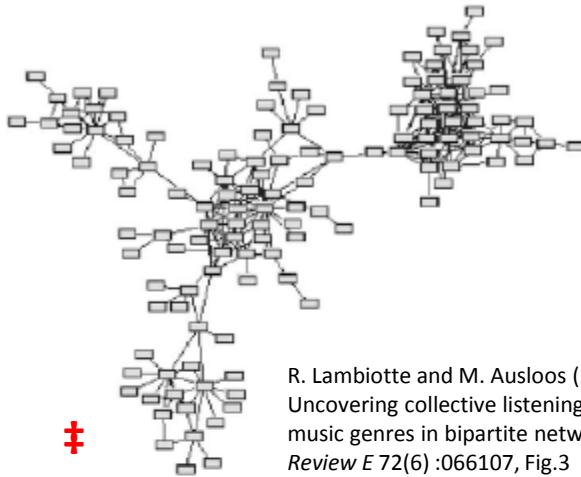


✚

Ramon Ferrer i Cancho et al. (2001) Topology of
technology graphs: Small world patterns in electronic
circuits, *Physical Review E* 64(4) :046119, Fig.1

その他

曲の好み



R. Lambiotte and M. Ausloos (2005)
Uncovering collective listening habits and
music genres in bipartite networks, *Physical
Review E* 72(6) :066107, Fig.3

英単語

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。

Ramon Ferrer i Cancho and Richard V. Solé (2001)
The small world of human language,
Proceedings of the Royal Society of London B, 268:2261-2265,
p.2262 Fig1(b)

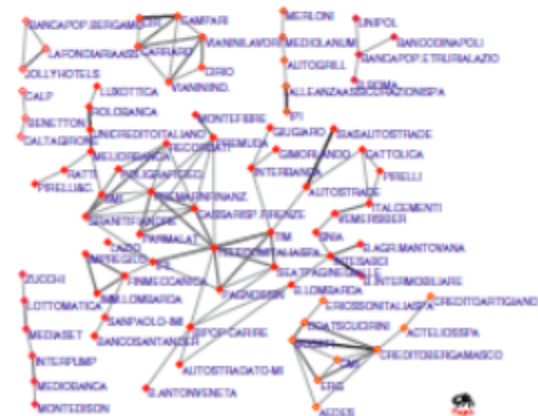
Ferrer i Cancho & Solé (2001)

Wikipedia



Wikipediaより転載
<http://ja.wikipedia.org/wiki/ファイル:WorldWideWebAroundWikipedia.png>

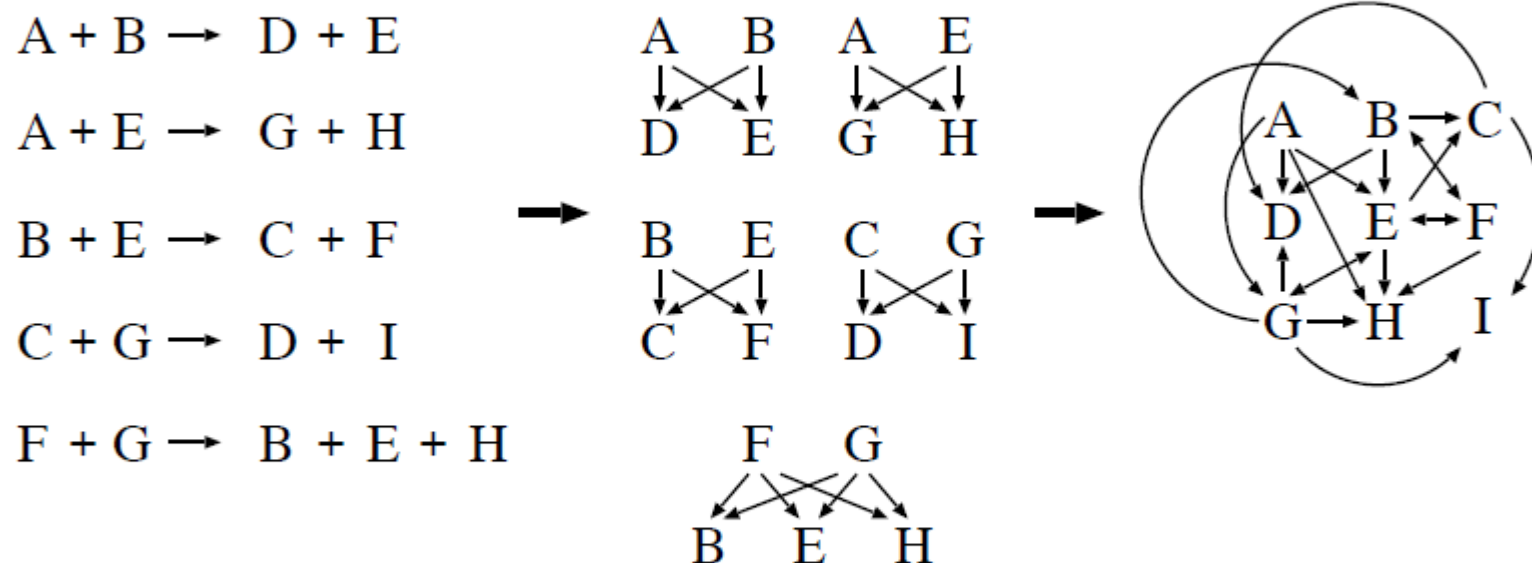
株式市場



Battison & Catanzaro (2004)

‡ Battison, S. and Catanzaro, M. (2004) Statistical properties of
corporate board and director networks, *European Physical
Journal B*, 38 (2): 345-352, p.345 Fig.1.
<http://publications.edpsciences.org/>

反応ネットワークとは

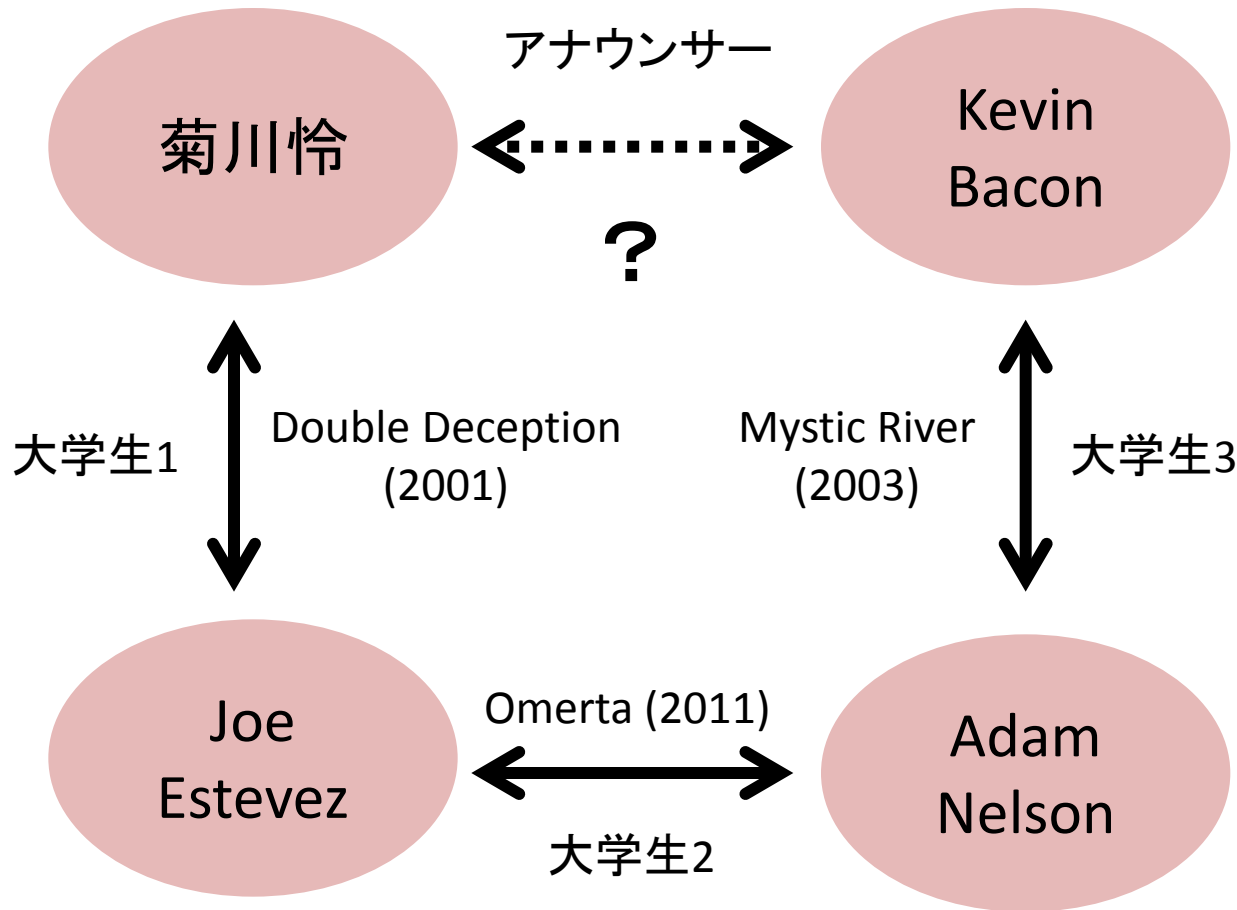


- A, B, C, ...: 細胞内分子（代謝物質, 酵素など）.
- 水や ATP は当たり前すぎるので除くことも多い.

- 代謝ネットワーク (metabolic network)
 - 頂点 = 代謝物質
 - 枝 ← 代謝物質の間の生化学反応
 - 遺伝子の発現は気にしない
- 遺伝子発現調節ネットワーク (genetic regulatory network)
 - 遺伝子の発現（と、対応するタンパク質の合成）は、他のタンパク質に影響される
 - 遺伝子 $\approx$ タンパク質
 - タンパク質が頂点（遺伝子を含めることもある）
 - 方向付きの枝. 枝の「太さ」は負のこともある (negative feedback)
- タンパク質相互作用ネットワーク (PIN)
 - タンパク質同士の物理的な相互作用 (bind)
 - 方向なしの枝

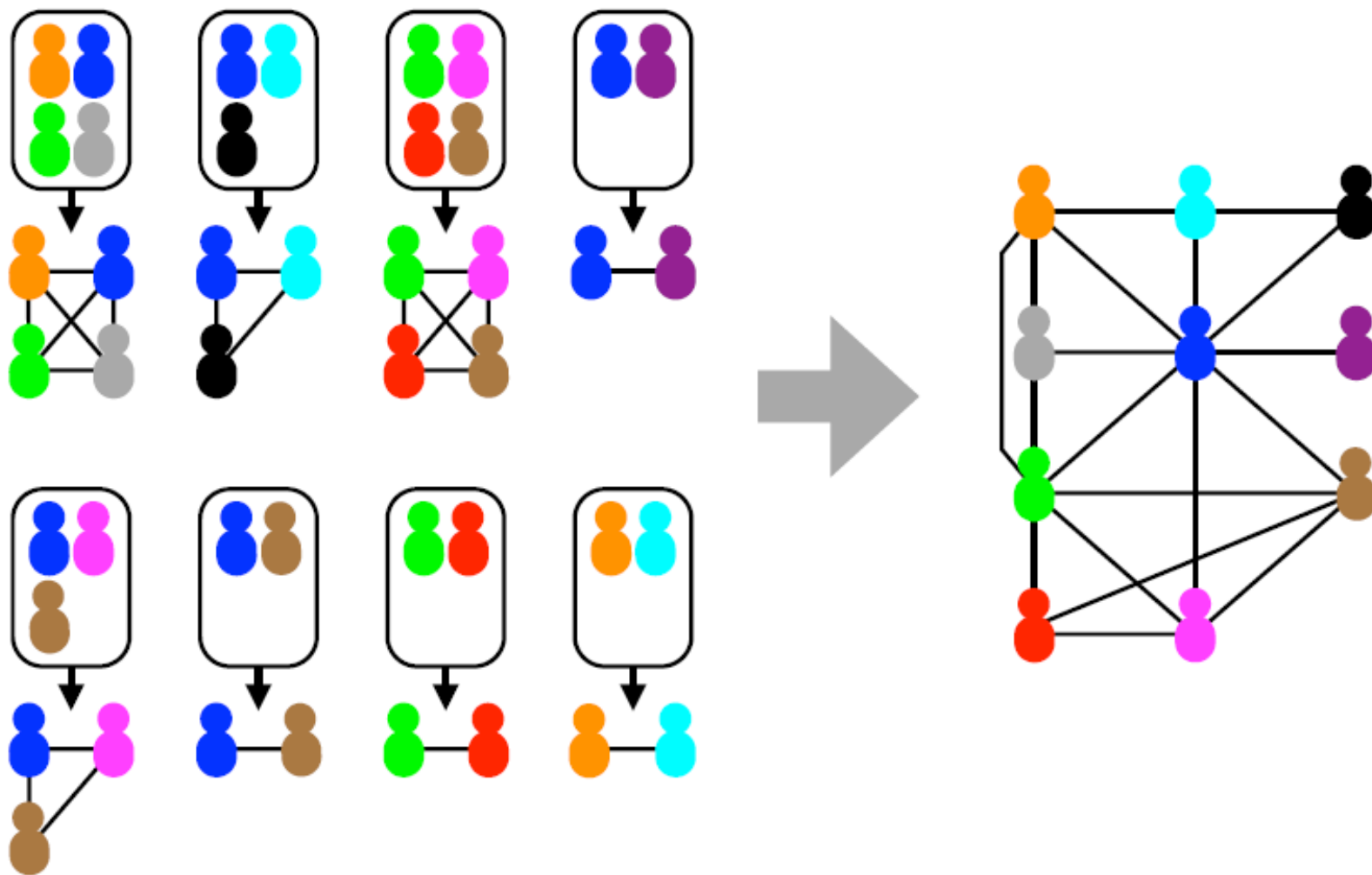
(1) 6次の隔たり

Kevin Bacon game (1994)



- 菊川怜: 3
- 松田聖子: 2
- 竹中直人: 2
- 安室奈美恵: 4
- 安達祐実: 2
- 中村獅童: 3
- 北野武: 2
- 藤原紀香: 3

Erdős number



平均距離

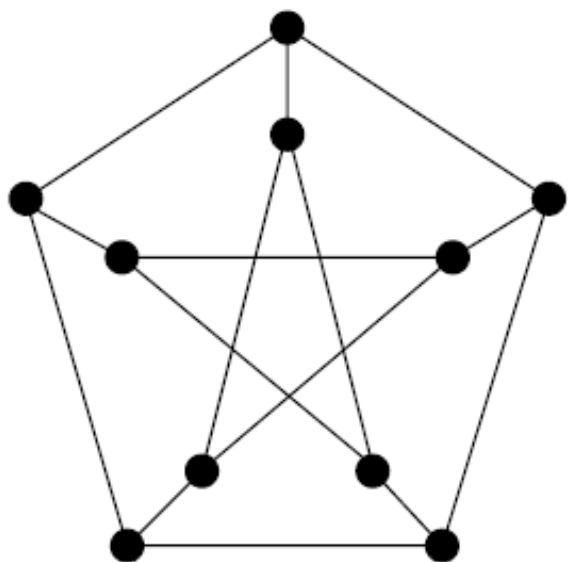
- L = “2頂点間距離の、全ての頂点ペアに渡る平均”

例 1: ピーターセン・グラフ

$$10 \times 9 / 2 = 45 \text{ 頂点ペア}$$

$$L = 1 \cdot 15/45 + 2 \cdot 30/45 = 5/3$$

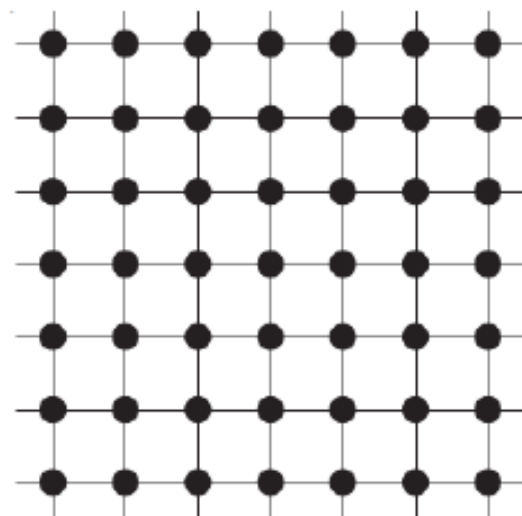
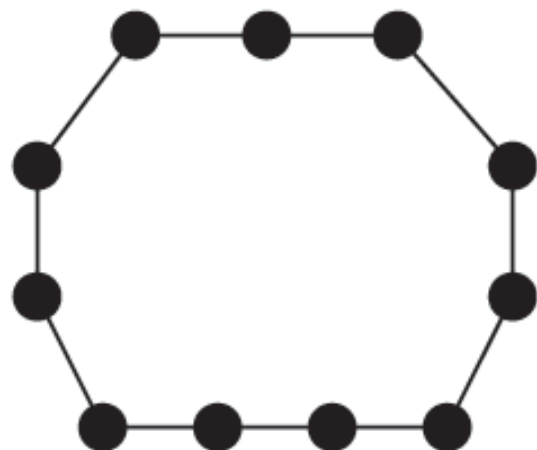
大抵の実データ: $L \propto \log N$



small-world 実験

- Milgram (1960s)
 - Boston のある人が「ゴール」
 - アメリカ中部や Boston の「誰か」から「スタート」
 - 手紙を投げる. 全体のうち 18 通がゴールまで到達
 - $L = 6$
- Dodds, Muhamad & Watts (2003).
 - 世界中の色々な人が「ゴール」
 - E メールを投げる. $384/98847 \approx 0.3\%$ がゴールまで到達
 - $L = 5$ (同国内), $L = 7$ (異国間)
 - 捨てられる Eメールの分もちゃんと考慮して計算

例: サイクル

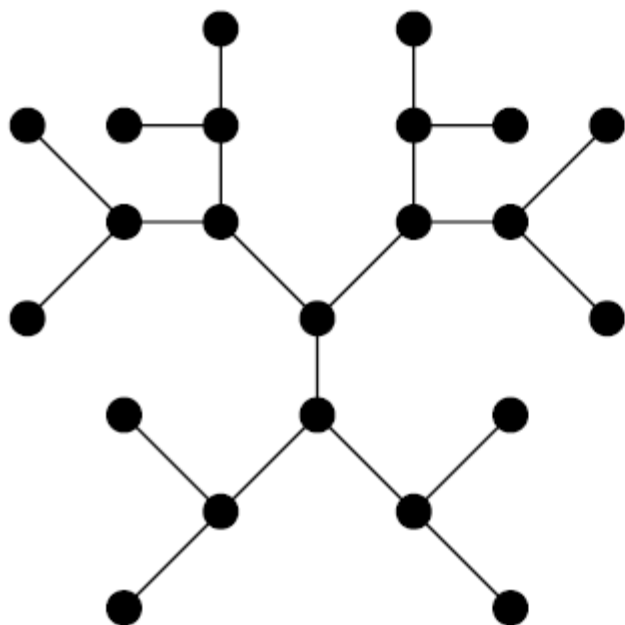


$$L = 1 \cdot 2/10 + 2 \cdot 2/10 + 3 \cdot 2/10 + 4 \cdot 1/10 + 5 \cdot 2/10 = 3$$

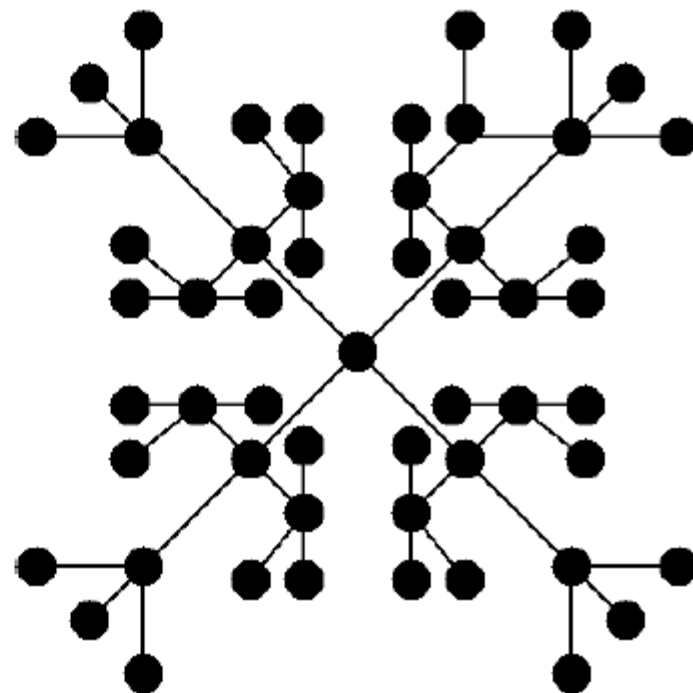
実際には, L は約 $N/4$. 大きすぎる!

d 次元なら, $L \propto N^{1/d}$

木



次数=3



次数=4

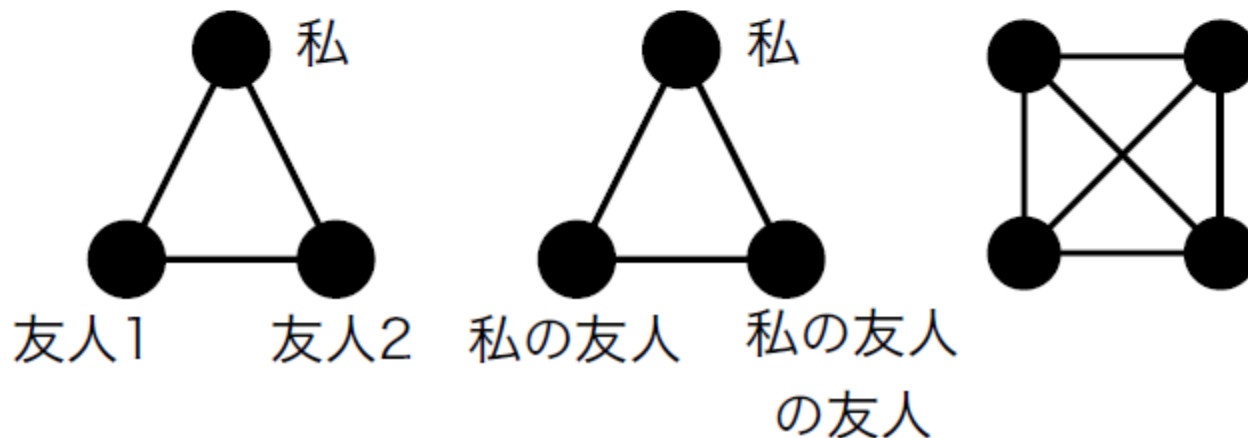
$\sqrt{L}$ は小さい ($\propto \log N$)

感染症はどう広がるか

	昔	今
平均距離 L	大きい	小さい
ペスト	進行波	ジャンプしながら急速に広がる
新型インフルエンザ, SARS	進行波	ジャンプしながら急速に広がる

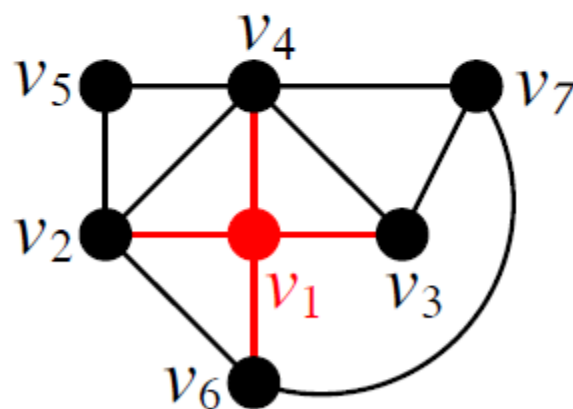
(2) クラスター性

- 私たちは内輪で群れやすい
- 私の友人2人同士が友人
- 友人の友人はやはり私の友人
- → 全て同じこと. 三角形で代表する.



クラスター係数

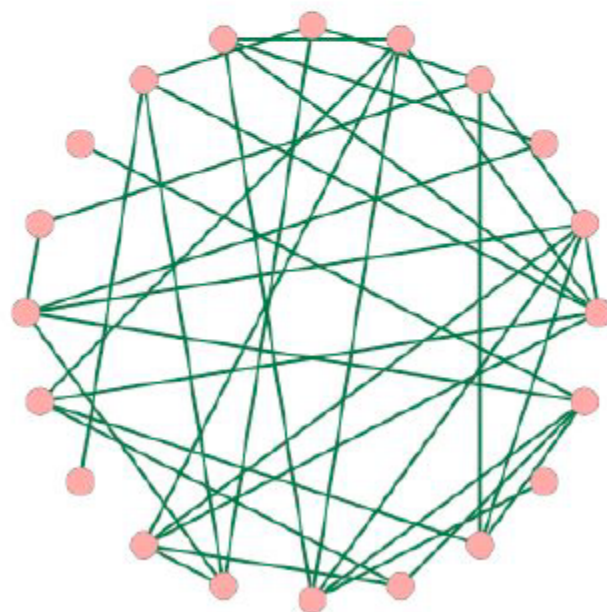
- $k_i = i$ の次数
- $C(v_i) = (\text{v}_i \text{ を含む三角形数}) / [k_i(k_i-1)/2]$
 - $C(v_1) = 3/[4*3/2] = 1/2$
 - $C(v_3) = 2/[3*2/2] = 2/3$
- $C = (C(v_1) + \dots + C(v_N)) / N$
 - $C = 8/15$
 - 木の $C = 0$



ランダム・グラフ

(Erdős & Rényi, 1959)

- 頂点は N 個
- 各頂点对の間に、確率 p で枝がある
- 同じ N と p でも毎回異なるネットワークが出てくる
- 何も仮定がない場合の標準モデルの 1 つ



ランダム・グラフの性質

- L は小さい

- c は小さい

$$C = p = \frac{\langle k \rangle}{N-1} \rightarrow 0$$

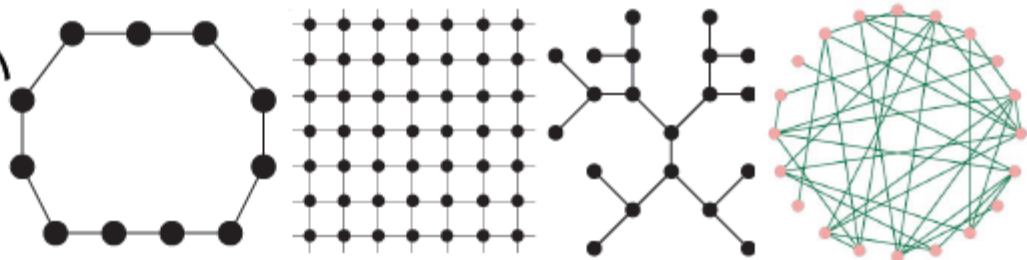
- 数学で証明の道具に使われたりする.
- ある性質が $N \rightarrow \infty$ で正の確率で起こる.

ここまでのまとめ

- 6 次の隔たり
- L が小さい
- 三角形が豊富
(クラスター性)

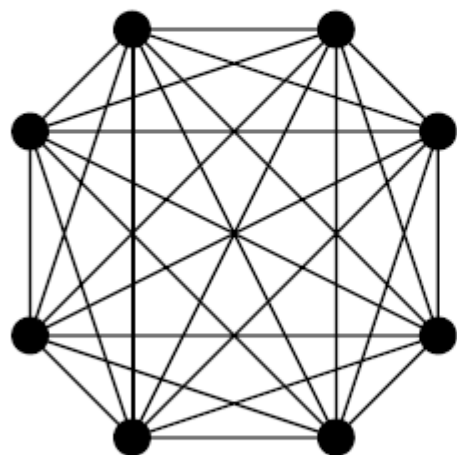
ネットワー ク	$p(k)$	L	C
輪	$k=2$ のみ	大	小 (大きく はできる)
正方格子 (碁盤目)	$k=4$ のみ	大	小 (大きく はできる)
木	一定	小	0
ランダム・ グラフ	あまり散ら ばってない	小	小

- C が大きい



小さい L + 大きい C

完全グラフ



$L = 1$ (考えられる最小値)

$C = 1$ (考えられる最大値)

枝の使いすぎ

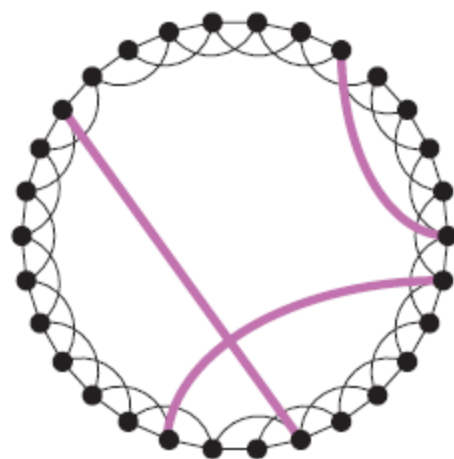
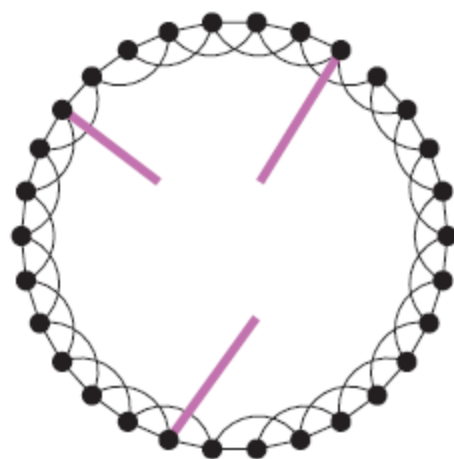
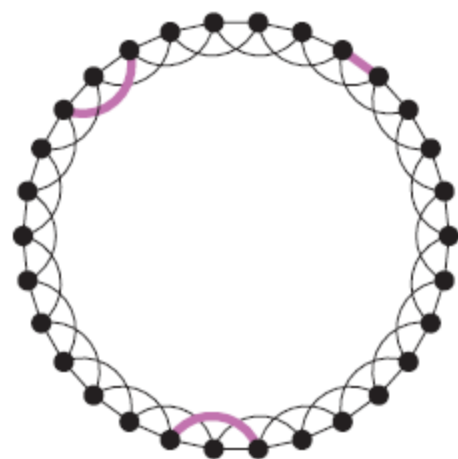


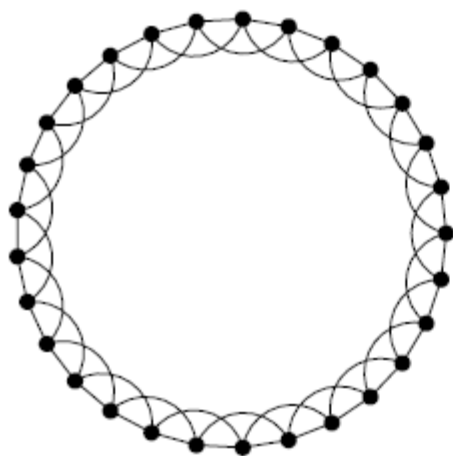
暗黙の前提：

「次数」は大きくない $k \ll N$

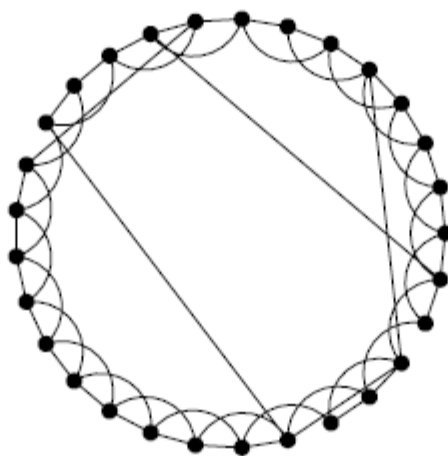
Watts-Strogatz モデル (1998)

- 「輪」を用意
- 割合 p の枝をつなぎ変える
- 近道

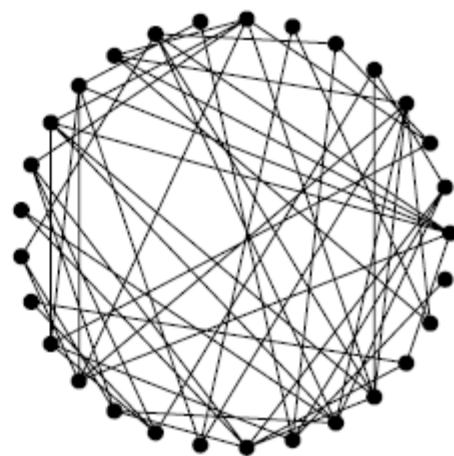




$p = 0$



$p = 0.1$



$p = 1$

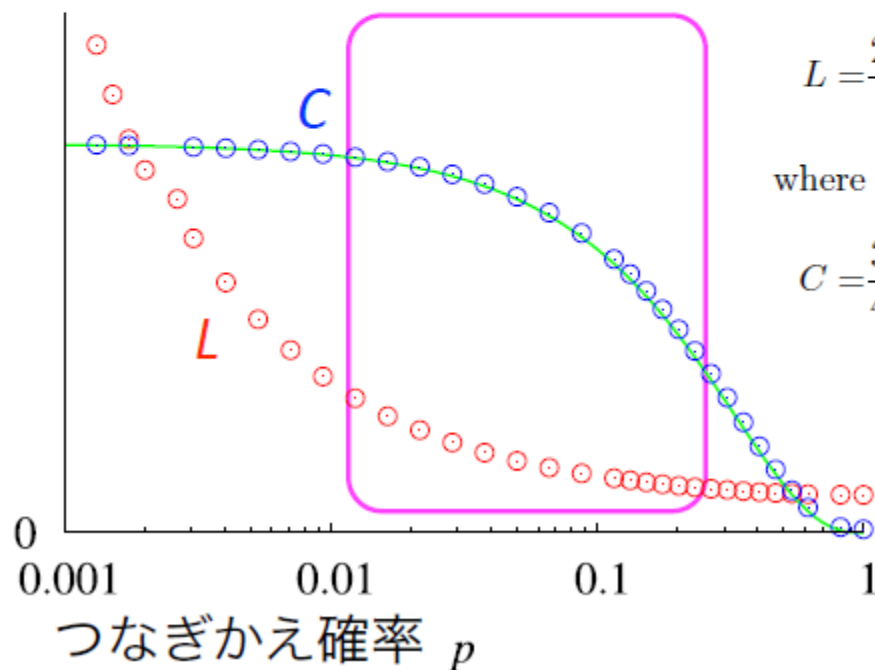
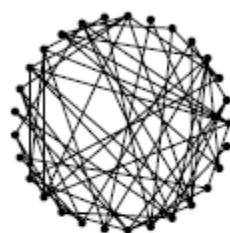


小

つなぎかえ確率 p

大

WS モデルの性質



$$L = \frac{2N}{k} f\left(\frac{Nkp}{2}\right)$$

$$\text{where } f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1}\left(\sqrt{\frac{x}{x+2}}\right)$$

$$C = \frac{3k-6}{4k-4}(1-p)^3$$

- ✓ 枝が多すぎもしない
- ✓ いいとこどり可

弱い紐帯の強さ (Granovetter, 1973)

- 紐帯=枝、リンク
- 有利な転職情報が、実は、いつも近くにいる友人よりも、弱い紐帯（近道）を介した遠い知人から入りやすい。
- 別の価値観の人から新鮮な情報
- 内輪で固まるとチャンスを逃す
- 仲がとてよくなくてもいい
- 就職情報の話に限らない
- ある程度「異質」を心がけたい

