

# 複雑ネットワーク科学 を活用した実データ解析

大西立顕      ohnishi@rikkyo.ac.jp  
立教大学 大学院人工知能科学研究科

2021/7/30 実践データ駆動科学オンラインセミナー

# 社会・経済やその他諸分野のビッグデータ

情報通信技術 (ICT) の驚異的な性能向上 [5年毎に10倍]

旧来の学問体系，社会・経済の仕組み，政策・制度は，高精度・高頻度・大量データが活用できない時代の枠組み (概念・思想重視でデータ軽視) に留まっている



時系列データ 金融市場，人の行動，センサ，脳波

テキストデータ TV，ニュース，新聞，ウェブ

ネットワーク 企業間取引，振込，貿易，銘柄間相関

時空間データ 電話帳，国勢調査，不動産，商業統計，  
ID付き Twitter，ID付き POS，人流 (位置)

データ科学，機械学習，社会・経済物理学，ネットワーク科学，超並列計算などを用いた実証科学的解析

ビッグデータ + データ駆動型科学 → 新たな知見・知識

# Outline

- 1 複雑ネットワーク科学とは
- 2 ネットワークの頑強性の応用例
- 3 中心性の応用例
- 4 まとめ

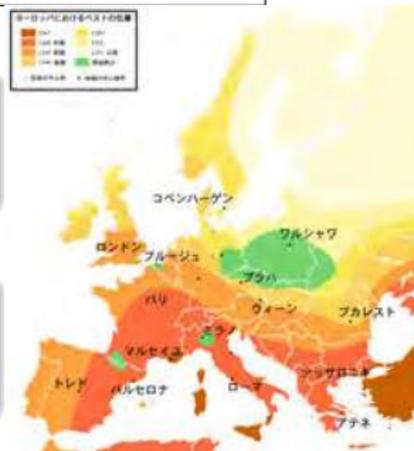
# Outline

- 1 複雑ネットワーク科学とは
- 2 ネットワークの頑強性の応用例
- 3 中心性の応用例
- 4 まとめ

# 新型インフルエンザの世界流行 (パンデミック)

ペストの伝播 (14 世紀) →  
単純な格子 (近接相互作用)

新型インフルエンザ (今世紀) ↓  
感染が拡がりやすい



[http://upload.wikimedia.org/wikipedia/ja/f/f5/Pestilence\\_spreading\\_Japan.png](http://upload.wikimedia.org/wikipedia/ja/f/f5/Pestilence_spreading_Japan.png) より

## 警戒水準5に引き上げ

### 新型インフルでWHO

**世界的な大流行を警告**  
日本政府 万全の対策を指示

新しいWHOのフェーズ分類  
WHOは、新型インフルエンザの流行を5段階に分類し、警戒水準を5に引き上げた。これは、世界的な大流行の恐れがあることを示している。

1. 動物から人に感染する可能性があるが、現時点では人に感染する能力はない。
2. 動物から人に感染し、人に感染する能力があるが、現時点では人に感染する能力はない。
3. 人に感染し、人に感染する能力があるが、現時点では人に感染する能力はない。
4. 人に感染し、人に感染する能力があるが、現時点では人に感染する能力はない。
5. 人に感染し、人に感染する能力があるが、現時点では人に感染する能力はない。

日本経済新聞

特報

2009年5月1日

03-3561-4000

www.nikkei.com



<http://news.tbs.co.jp/ram/news4125905.11.ram> より

# 「つながる」って便利？

## 常磐・宇都宮・高崎線と東海道線がつながる

(日本経済新聞 エコノ探偵団 2012/3/4)

出典：<https://style.nikkei.com/article/DGXDZ039258650S2A300C1W14000>



相互乗り入れ先で故障や事故があると影響が出る  
東日本大震災やタイの洪水ではサプライチェーン  
(供給網)が途切れ、大きな被害をもたらした  
1997年通貨危機，2009～2011年金融危機

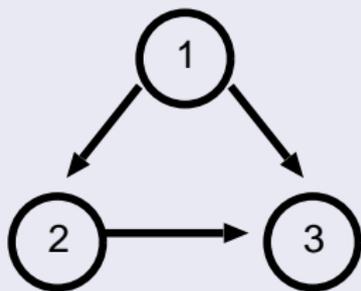
ショックも連鎖しやすくなり便利な半面危険も増える

複雑な相互作用：脳神経細胞，遺伝子と代謝物質など

# ネットワーク(頂点と枝)の特徴量

ネットワークには共通の原理・法則がある(普遍性)

複雑なネットワーク構造  
理解困難な大きな情報

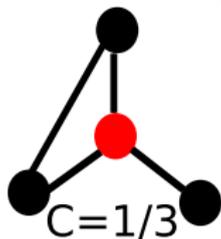
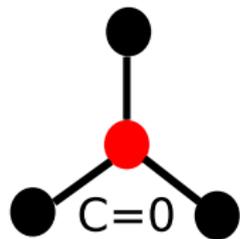


$$M = \begin{array}{c} \text{始} \\ \text{点} \\ (1) \\ (2) \\ (3) \end{array} \begin{array}{c} \text{終点} \\ (1) \quad (2) \quad (3) \\ \left( \begin{array}{ccc} \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right) \end{array}$$

ネットワーク特徴量  
理解可能な小さな情報

- クラスター係数  $C$
- 平均距離  $L$
- 次数分布
- 頑強性
- 中心性
- モチーフ
- 次数相関
- コミュニティ構造
- $\vdots$

# クラスター係数 $C$ : 2人の友人同士は友人か？



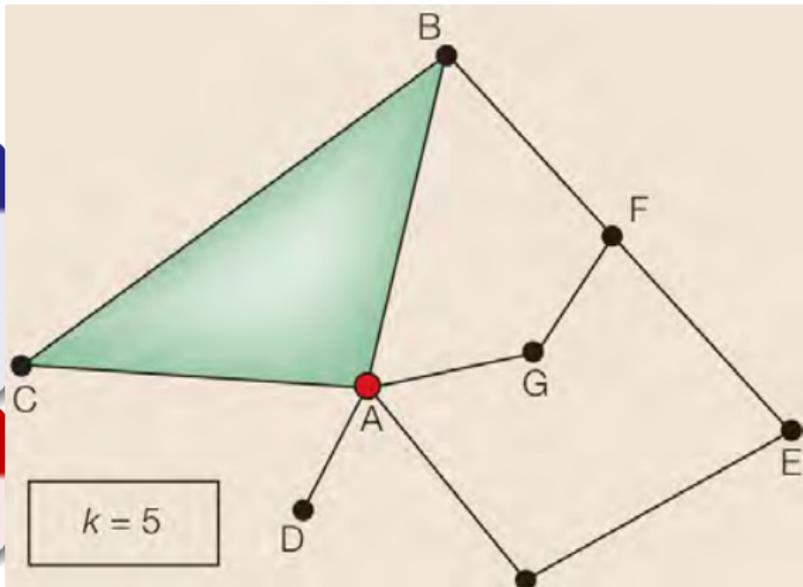
$$C = \frac{\text{の数の数}}{\text{の数の数} + \text{の数の数}}$$

## 頂点の $C$

$$C_A = \frac{1}{\frac{5(5-1)}{2}}, C_F = 0$$

## ネットワークの $C$

$$C = \langle C_i \rangle$$

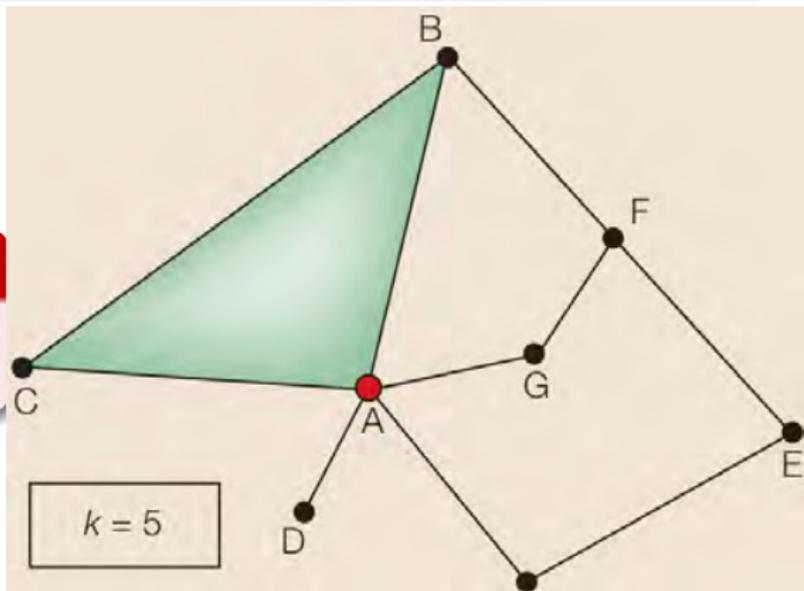


# 平均距離 $L$ : 2 頂点間の最小の枝数

A から E へ行くのに必要な最小の枝数 :  $L_{A,E} = 2$

ネットワークの  $L$

$$L = \langle L_{i,j} \rangle$$



A.L.Barabasi & L.N.Oltvai Nature Rev. Genet.(2004)

# 複雑ネットワーク登場以前 過度な単純化

増田直紀他, 応用数理 (2006)

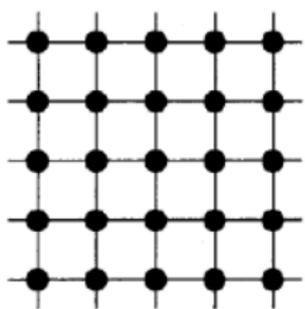


図1 格子

C 大 (三角格子) L 大

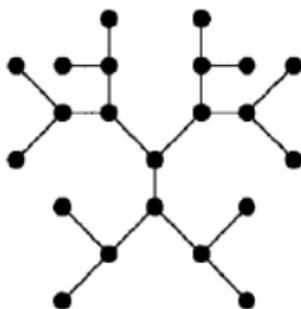


図2 木

C 小 L 大

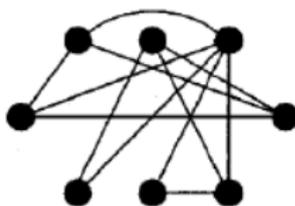


図3 ランダム・

C 小 グラフ L 小

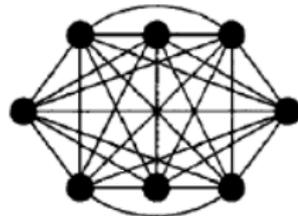


図4 完全グラフ

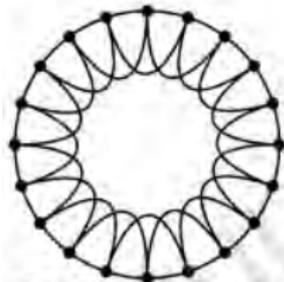
C 大 L 小

複雑ネットワークは Small-world (C 大で L 小)  
で Scale-free (頂点を持つ枝の本数がベキ分布)

Regular

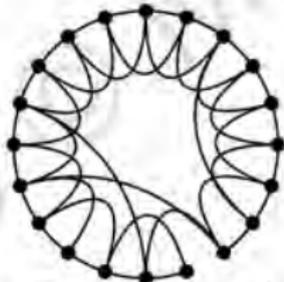
Small-world

Random

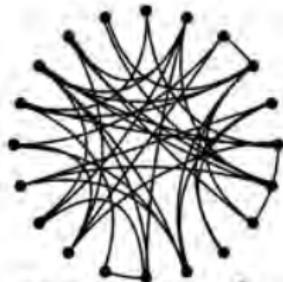


C 大 L 大

$p = 0$



C 大 L 小



C 小 L 小

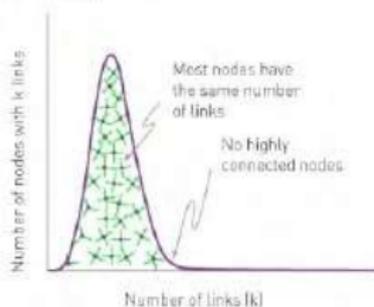
$p = 1$

Increasing randomness

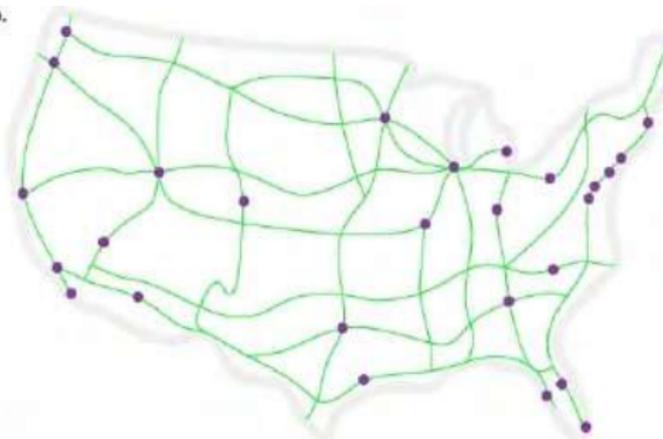
Watts, D.J., and Strogatz, S.H., Nature (1998) / 29

# ランダム , スケールフリーネットワーク

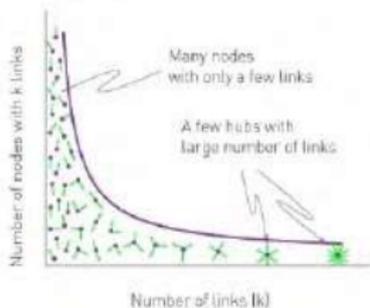
a. POISSON



b.



c. POWER LAW

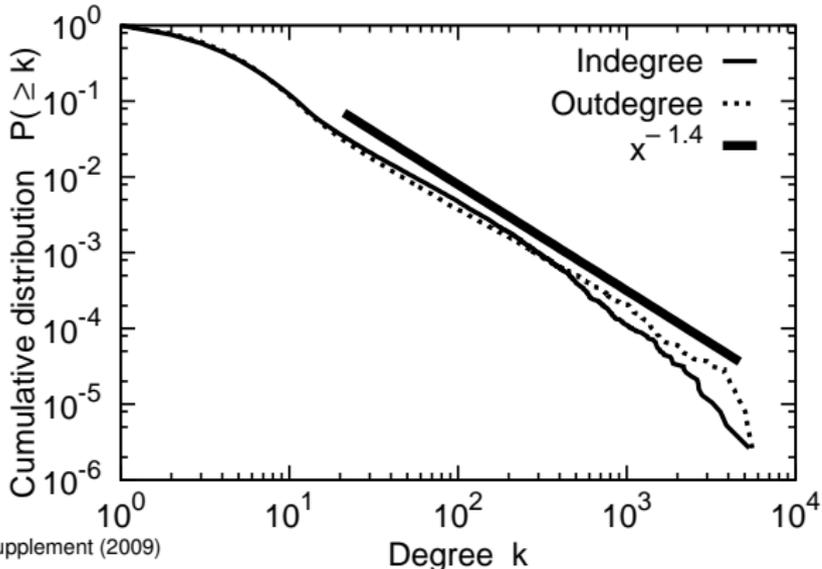
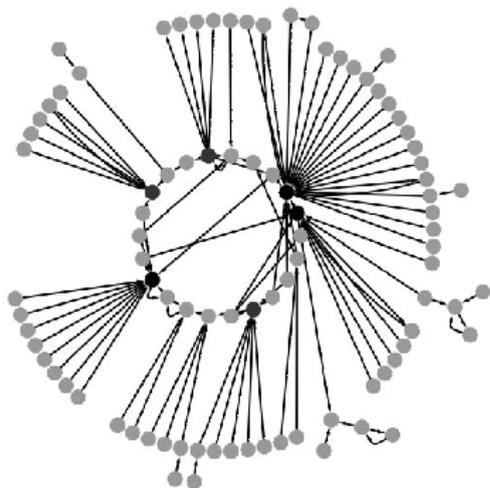


d.



# 企業間取引のネットワーク

どの企業から仕入れてどの企業に販売していますか？  
⇒ 約 100 万社について約 400 万リンクのネットワーク



T. Ohnishi et al., Progress of Theoretical Physics Supplement (2009)

取引先数はほとんどは数社だが，数千社の企業も存在  
スケールフリーネットワーク (リンク数がベキ分布)

社会・経済・生命の実ネットワークでよく観測される

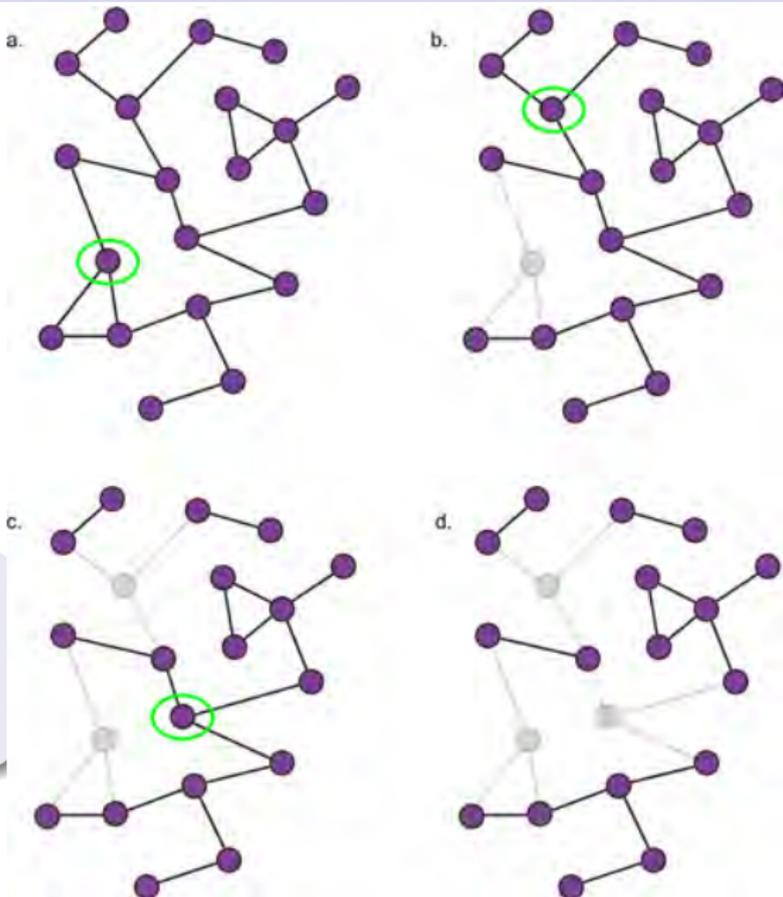
# Outline

- 1 複雑ネットワーク科学とは
- 2 ネットワークの頑強性の応用例**
- 3 中心性の応用例
- 4 まとめ

# 頂点を除去していったときの影響

- 1つの頂点を除去しても限定的な影響しか与えない。
- いくつかの頂点を除去すると、数個の連結成分に分解される。

何個の頂点 ( $100p\%$  の頂点) を除去するとバラバラになるか  
最大連結成分の大きさの  $p$  依存性に注目



# ランダム除去と選択的除去

クラスター性や次数相関が無視できるなら

臨界確率  $p_C$  (バラバラになる  $p$ ) は次数分布で決まる

## ランダムに頂点を除去

ランダムネットワーク :  $p_C = 1 - \frac{1}{\langle k \rangle} \rightarrow 1 (k \rightarrow \text{大})$

スケールフリーネットワーク :  $p_C \sim 1$

ほぼ全ての頂点を除去するまで連結成分が保持される

- WWW は故障に強い
- 感染症は患者全員の治療が必要

## 次数の大きい順に頂点を除去

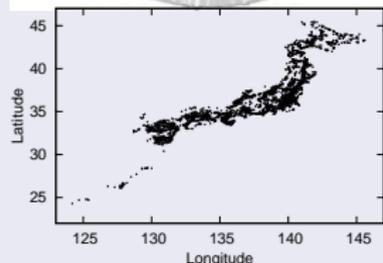
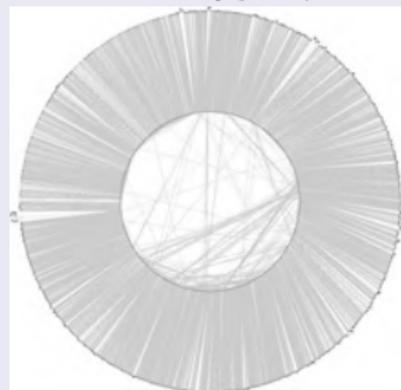
スケールフリーネットワーク :  $p_C < 0.03$  程度

- WWW はハブ攻撃に弱い
- 感染症はハブとなる患者の治療で制圧できる

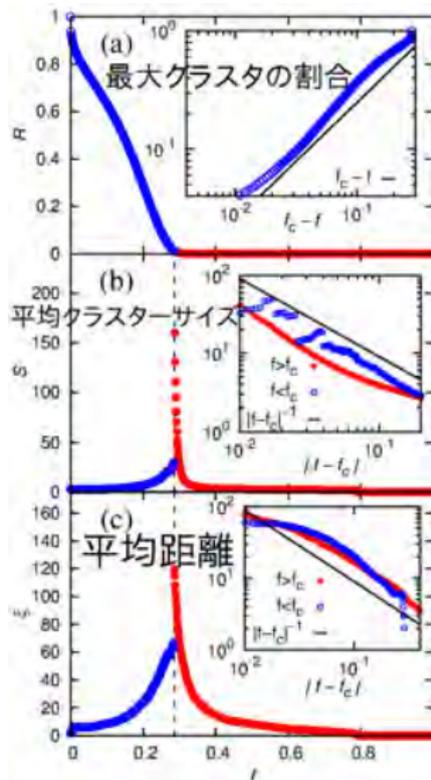
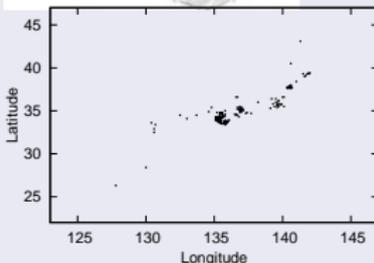
# サプライチェーンのリスクを評価

次数の大きい順に企業を除去していく  
核（最大連結成分）がどの程度壊れるか

20% 除去

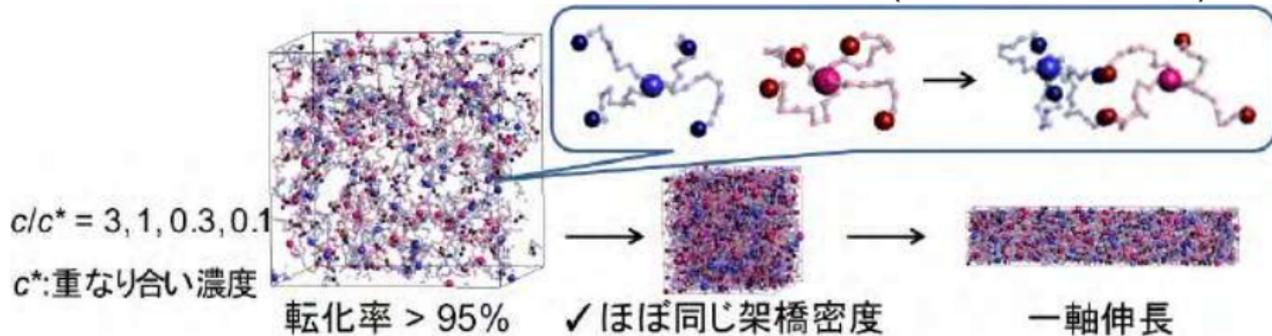


30% 除去 臨界点

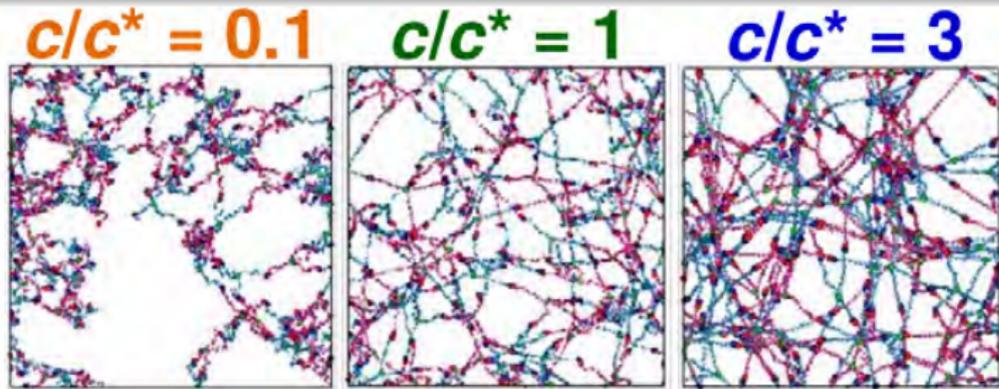


# MDシミュレーション (Kremer-Grest モデル)

2種類の4分岐の星形高分子から調製 (Tetra-PEGゲル)

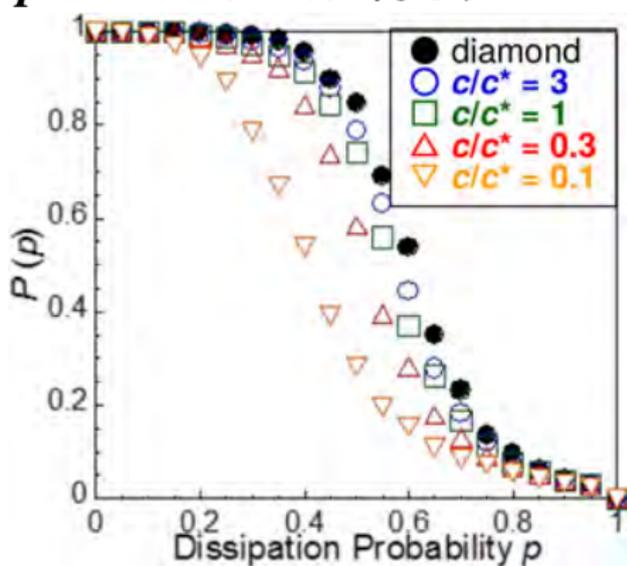
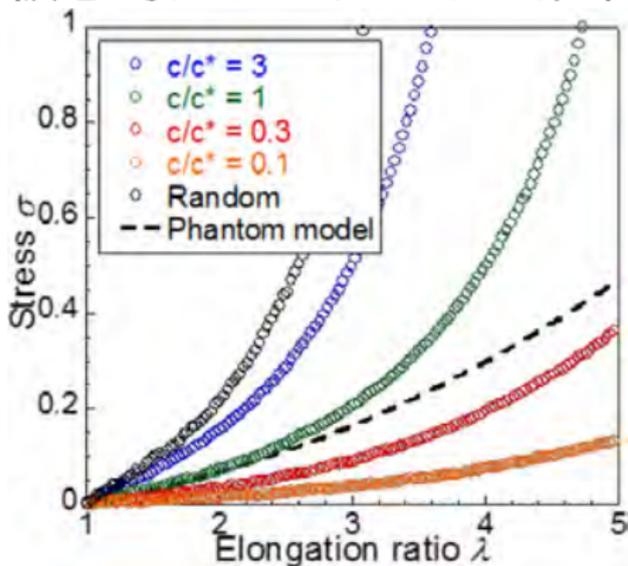


架橋濃度  $c/c^*$  は異なるが、分岐数と架橋密度は同じ



# ネットワークポリマーの脆弱性・頑強性

仮想的にリンクをある確率  $p$  でランダムに除去



次数分布は同一だが高濃度になるほど  $p_C$  は大きくなる

$p_C$  が大きい (最大連結成分を維持) ほど高い応力を示す

ネットワークの頑強性 (構造)  $\iff$  伸長時の力学物性

# Outline

- 1 複雑ネットワーク科学とは
- 2 ネットワークの頑強性の応用例
- 3 中心性の応用例**
- 4 まとめ

# 中心性

次数中心性：次数 (枝の本数)

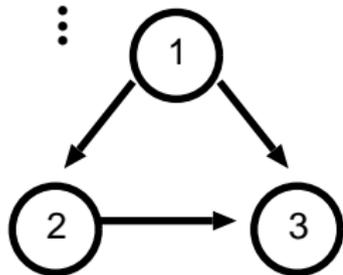
近接中心性： $1/\langle d(i, j) \rangle_{j \neq i}$  ← 他の頂点までの距離の平均

媒介中心性： $\left\langle \frac{s \text{ から } t \text{ への最短路で } i \text{ を通る数}}{s \text{ から } t \text{ への最短路の数}} \right\rangle_{s, t \neq i}$

固有ベクトル中心性

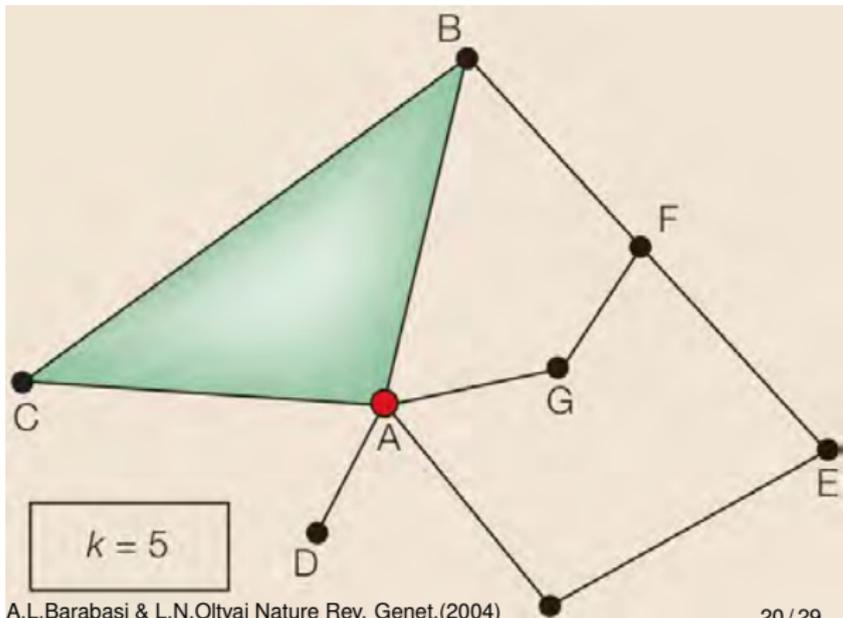
ページランク

⋮



		終点		
		(1)	(2)	(3)
始点	(1)	0	1	1
	(2)	0	0	1
	(3)	0	0	0

$M =$

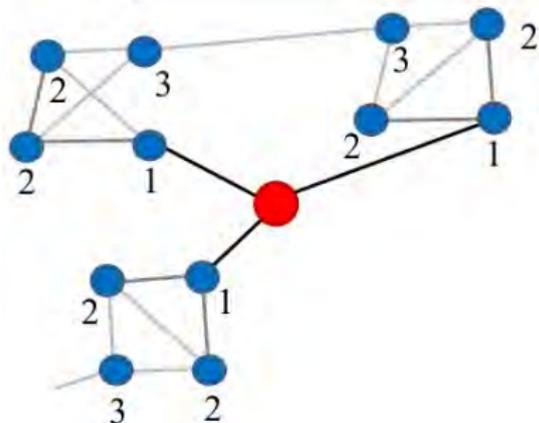


# 近接中心性

「近接中心性」  
の定義

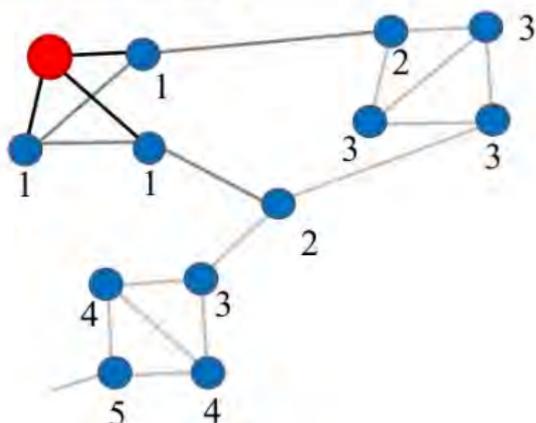
$$C_c(i) = \frac{1}{n} \sum_{j=1}^n d_{ij}$$

$d_{ij}$ : Shortest link number  
between node  $i$  and  $j$



$$C_c(i) = \frac{1}{1 \times 3 + 2 \times 6 + 3 \times 3} = 0.0417$$

中心性 高

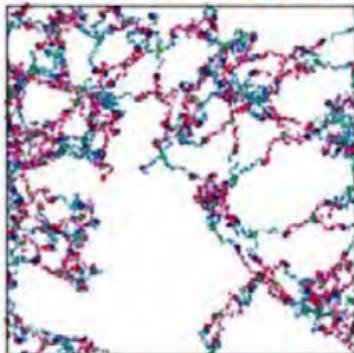


$$C_c(i) = \frac{1}{1 \times 3 + 2 \times 2 + 3 \times 4 + 4 \times 2 + 5 \times 1} = 0.0286$$

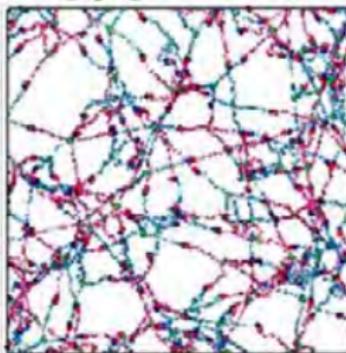
中心性 低

# 近接中心性

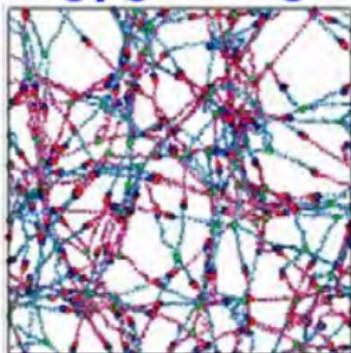
$c/c^* = 0.1$



$c/c^* = 1$



$c/c^* = 3$



近接  
中心性

0.149

<

0.247

<

0.260

架橋点にかかる応力 = 架橋点間距離，架橋点の揺らぎ

「近接中心性」と「初期の架橋点間距離」が重要

# 修正中心性：架橋点間の空間的距離を考慮

空間的な距離

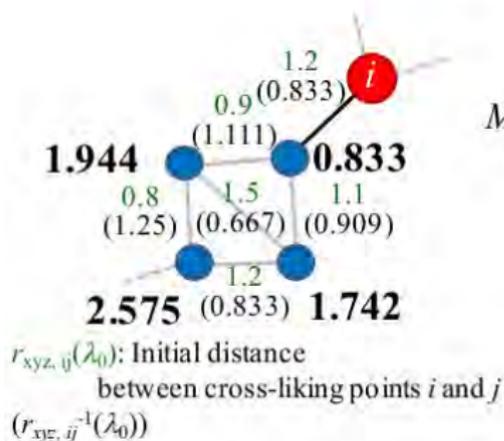
トポロジー

初期の架橋点間距離 $r(\lambda_0)$

近接中心性 $C_c(\lambda_0)$

統一的なパラメータ

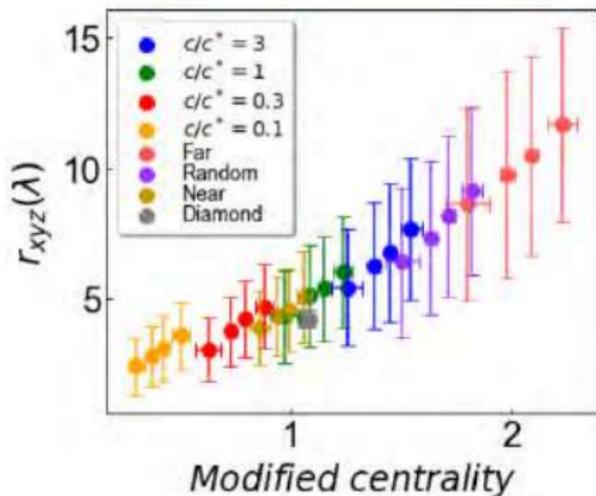
修正中心性



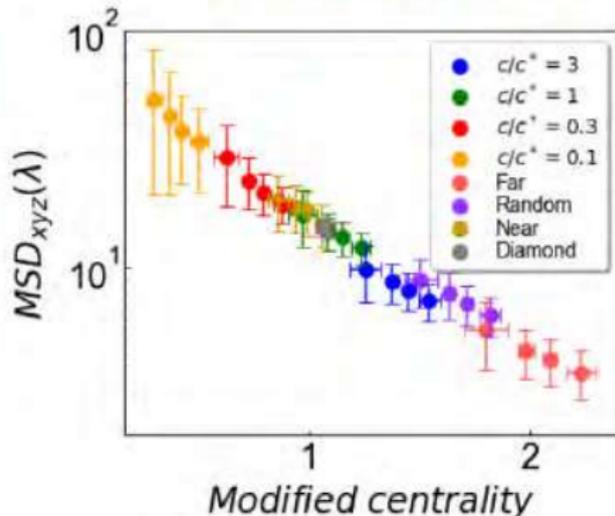
$$\begin{aligned} \text{Modified centrality}(i) &= \frac{1}{\sum_{j=1}^n \min_{\text{path}} \sum r_{xyz, ij}^{-1}(\lambda_0)} \\ &= \frac{1}{1.944 + 0.833 + 2.575 + 1.742} \\ &= 0.141 \end{aligned}$$

# 応力は，修正中心性で統一的に説明できる

## 架橋点間距離



## 架橋点の揺らぎ



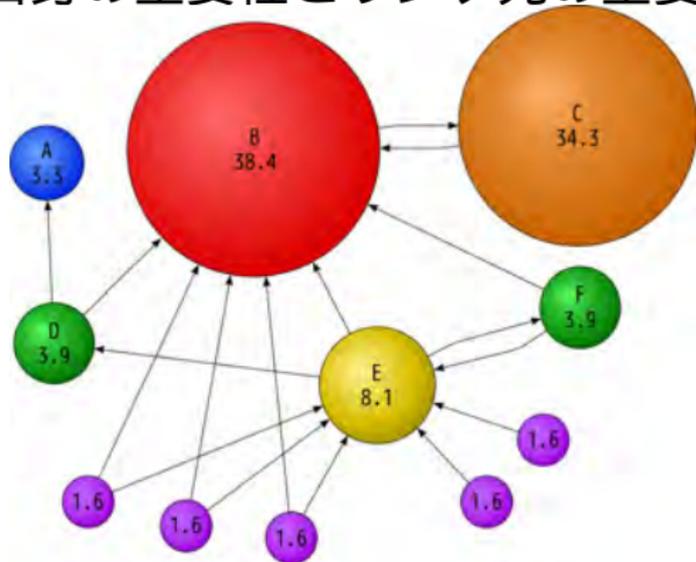
ネットワーク構造に依存しない

つながりに関して中心にいる架橋点ほど，かつ，初期の架橋点間距離が大きい架橋点ほど応力に寄与

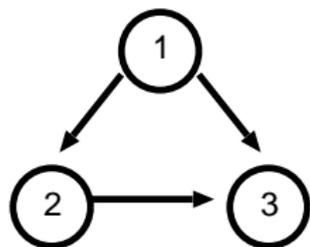
応力 = 架橋点間距離，架橋点の揺らぎ

# ページランク

自身の重要性とリンク元の重要性を加味してランク付け



出典: <http://en.wikipedia.org/wiki/PageRank>

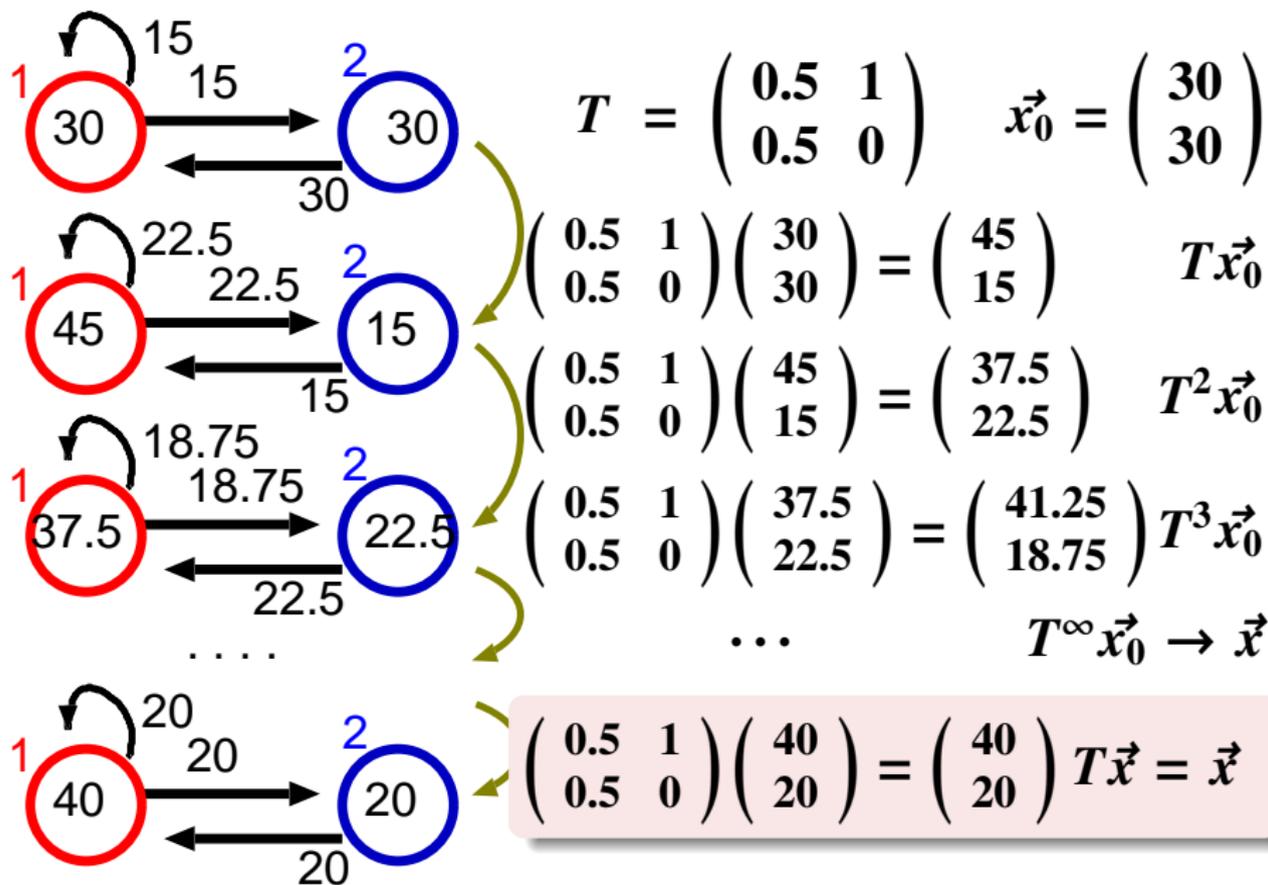


始点 (1) (2) (3)      終点 (1) (2) (3)

$$M = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$M^T$  の近似行列の固有ベクトル

# マルコフ過程の行く先：最大固有ベクトル



# 成長率 $r$ の条件つき期待値 $E(r | k_{in}, p)$

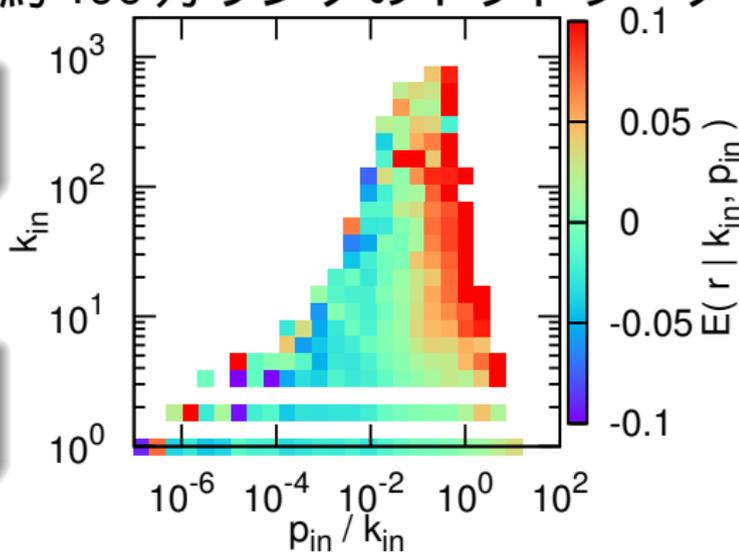
どの企業から仕入れてどの企業に販売していますか？  
⇒ 約 100 万社について約 400 万リンクのネットワーク

100 万社の企業をページランクでランキング

金流でみて、入次数が大きく

単位入次数あたりのページランク： $p/k_{in}$

が大の企業は成長する



$k_{in} \geq 20$  のノード： $r$  と  $p/k_{in}$  の Kendall's  $\tau = 0.22$

業種や都道府県に依存しない．物流ではみられない．

# Outline

- 1 複雑ネットワーク科学とは
- 2 ネットワークの頑強性の応用例
- 3 中心性の応用例
- 4 **まとめ**

## 複雑ネットワーク科学におけるネットワークの特徴量

- クラスタ係数が大きく，平均距離が小さい
- スケールフリー性 (次数がベキ分布)
- 頑強性 (パーコレーション)
- 中心性 (近接中心性，ページランク)
- ネットワークモチーフ
- 次数相関
- コミュニティ構造

「離散幾何学とその周辺 (仮)」 特集：離散数学に親しむ (仮)，数理科学，2021年12月号 (予定)

「ネットワーク学習から経済分析へ (仮)」 AI/データサイエンス：基礎から応用へシリーズ (仮)，サイエンス社，2022年中 (予定)