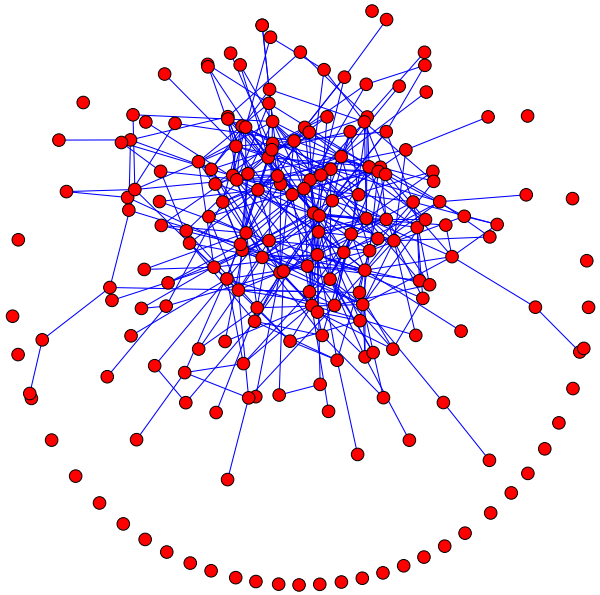
Dec. 18–20, 2006
in DEX-SMI 2006（大手町サンケイプラザ）



複雑ネットワーク生成と統計力学

大久保 潤

東北大学大学院情報科学研究科・日本学術振興会特別研究員 DC

http://www.smapip.is.tohoku.ac.jp/~jun/

in collaboration with M. Yasuda and K. Tanaka

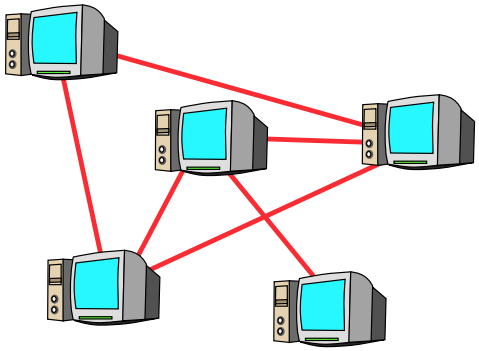
DEX-SMI

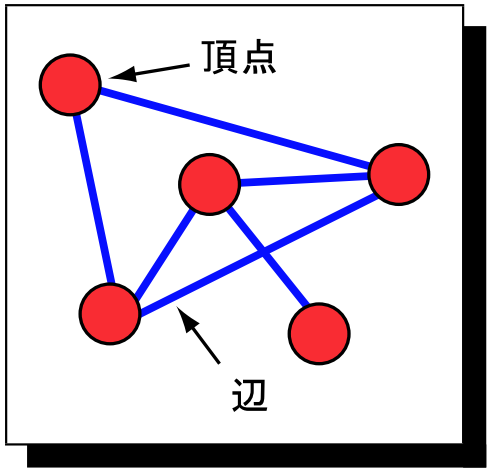
講演の内容

- 複雑ネットワーク研究とは何か
  - どのような対象を扱うのか
  - 研究のための基礎概念（次数分布，次数相関係数）
- 複雑ネットワークの生成問題
  - 代表的なネットワーク生成モデルの紹介
  - 新しいクラスのネットワーク生成モデルの提案とその統計力学的解析
- ネットワークとデータマイニング
  - コミュニティ検出問題
  - 統計物理学的な概念との関連性
- まとめ

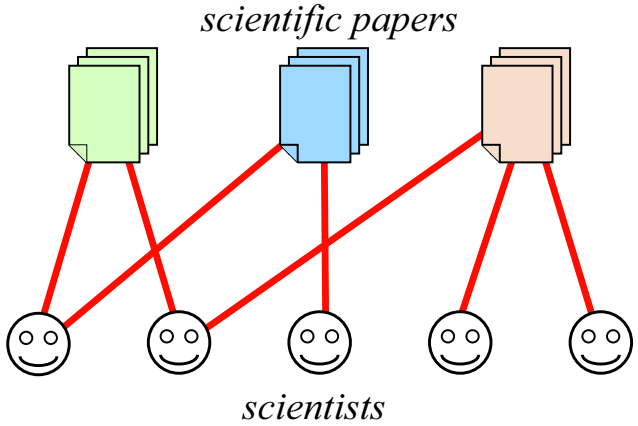
複雑ネットワーク研究とは何か

インターネット

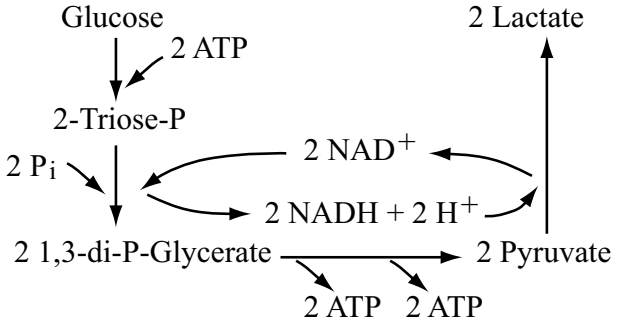




人間関係（研究論文の共著関係）



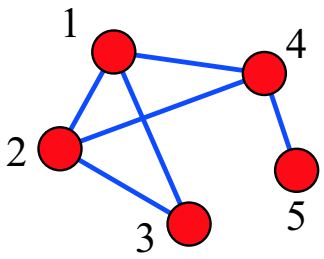
化学反応系（代謝系）



複雑なシステム ⇒ 頂点と辺から構成される「ネットワーク」

DEX-SMI

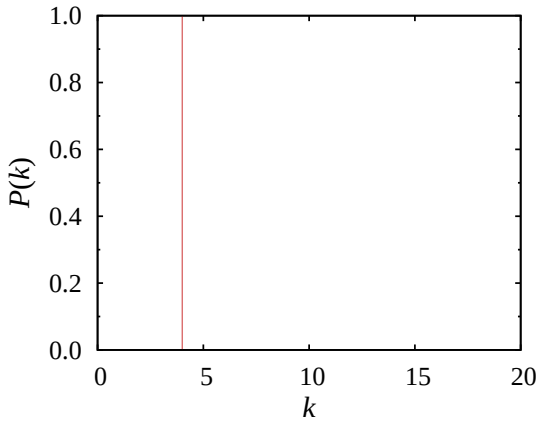
次数分布 ⇒ **スケールフリー性**



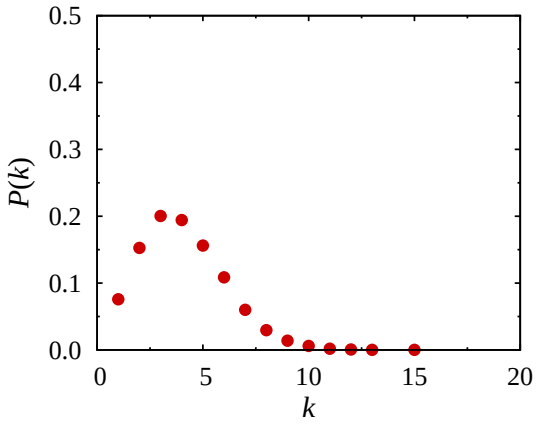
次数: 頂点に結ばれている辺の本数　(e.g., $k_1 = 3,\ k_3 = 2,\ k_5 = 1$)

次数分布: ある頂点に k 本の辺が結ばれている確率　$P(k) = \frac{1}{N}\sum_{i=1}^{N} \delta_{k_i,k}$

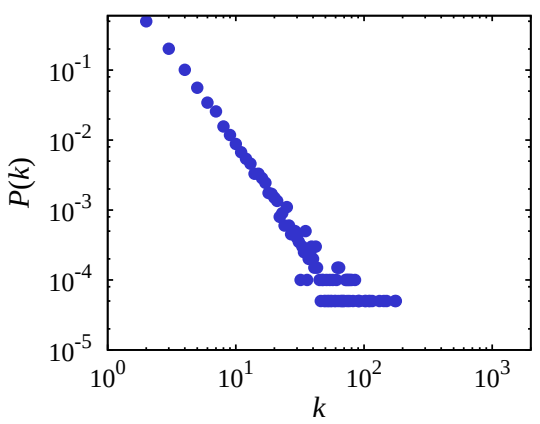
2 次元正方格子



ランダムネットワーク



スケールフリーネットワーク



- インターネット，WWW，人間関係，捕食・被捕食者関係，化学反応系 etc. ⇒ スケールフリー性
- スケールフリー性を特徴づける指数 γ : $P(k) \propto k^{-\gamma}$

DEX-SMI

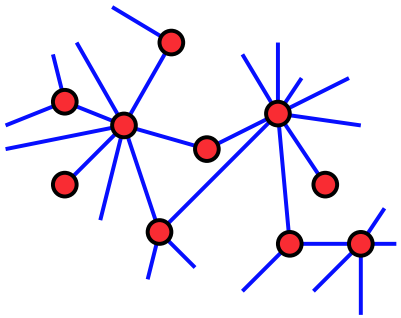
次数相関係数

N: 頂点の総数
M: 辺の総数
k_i: 頂点 i の次数
a_{ij}: 隣接行列

次数相関係数 r：頂点間の次数の相関

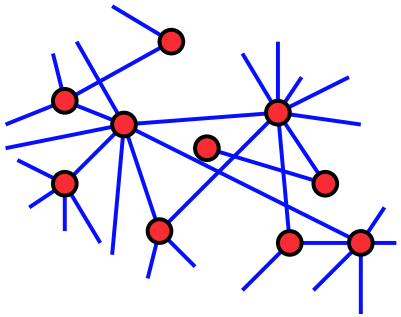
$$r = \frac{4M\sum_{i,j}^{N} k_i k_j a_{ij} - [\sum_{i,j}^{N}(k_i + k_j)a_{ij}]^2}{2M\sum_{i,j}^{N}(k_i^2 + k_j^2)a_{ij} - [\sum_{i,j}^{N}(k_i + k_j)a_{ij}]^2}$$

- 大きい次数の頂点と小さい次数の頂点が結ばれる
  ⇒ 次数相関は負



生物・工学関係のネットワーク

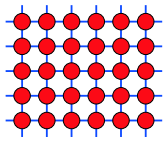
- 同程度の次数の頂点同士が結ばれやすい
  ⇒ 次数相関は正



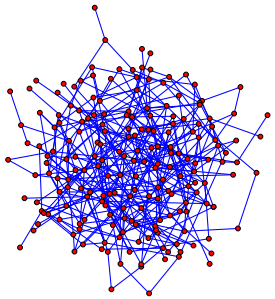
社会・人間関係のネットワーク

DEX-SMI

「複雑ネットワーク研究」という視点



正方格子



ランダムネットワーク

これまで:

統計物理学 ⇒ 空間構造での議論（連続 or 格子）

現在:

情報と物理の融合 ⇒ 関係性の科学

“情報系” を対象とする時には，backbone としての構造を「より現実的なもの」にする必要性

⇓

複雑ネットワーク研究

複雑ネットワーク研究の主要なテーマ

- 「測る」: 複雑ネットワークを測る指標・データマイニング
- 「作る」: 複雑ネットワークの生成問題
- 「使う」: 複雑ネットワーク上でのダイナミクス

DEX-SMI

複雑ネットワークの生成問題

現実に存在するネットワーク構造はどのように生成されるか？

BA **モデル** (growing, dynamical)	**「優先的選択」「成長」**
成長しないネットワーク生成モデル (nongrowing, dynamical)	**「優先的選択」「ランダム性」**

特にスケールフリー性にのみ注目

- **成長するネットワーク生成モデル（**BA **モデル）** $\Rightarrow$ **成長という要因が重要**
- nongrowing, dynamical **なモデルにおいて，どのようにすればスケールフリー性が発現するか？**
  （$\Rightarrow$ **平衡統計力学）**

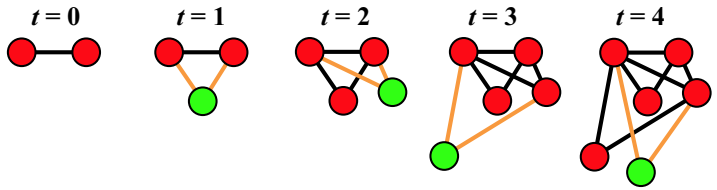
DEX-SMI

Barabási-Albert モデル (BA モデル)

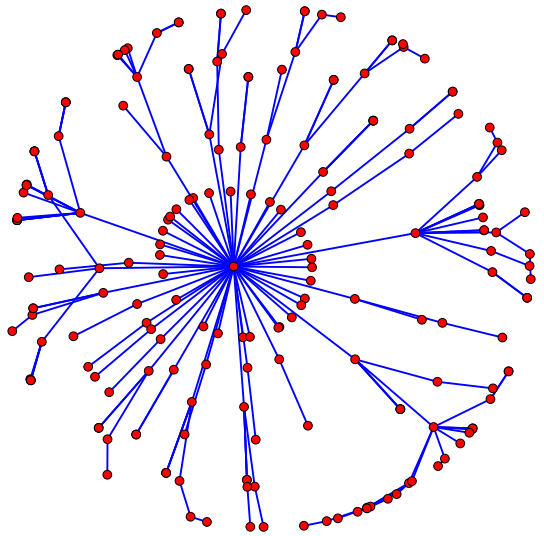
[Barabási and Albert, 1999]

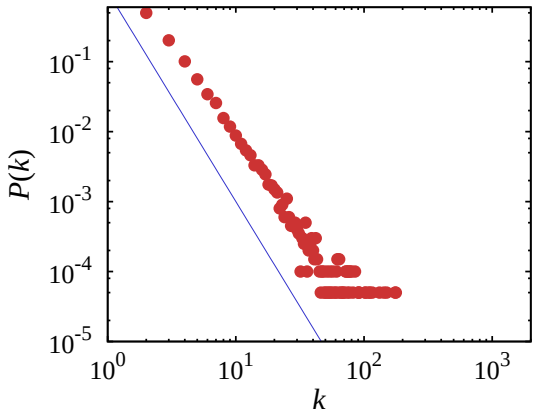
ポイント：「成長」「優先的選択」

1. m_0 個の頂点を互いに辺で結ぶ（初期ネットワーク・完全グラフ）．
2. m 個の頂点を $\Pi(i) \propto k_i$ の確率で選ぶ．ここで k_i は頂点 i の次数．
3. m 本の辺を持つ 1 つの頂点を追加し，m 個の頂点と結ぶ．
4. ステップ 2 と 3 を繰り返す（t 回繰り返した後，頂点数は $N = t + m_0$ になる）



$N = 200$
$\langle k \rangle = 2$





$N = 2000, \quad \langle k \rangle = 4$
averaged over 20 realizations

DEX-SMI

成長しないネットワーク生成モデル

[Ohkubo *et al.*, 2005]

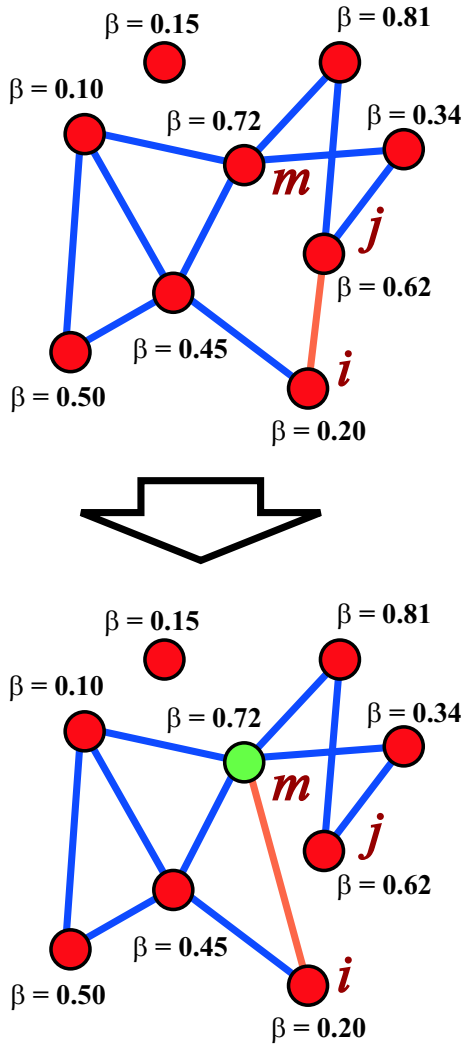
ポイント： **「優先的再結合」「ランダム性」**

1. M **本の辺を用いて** N **個の頂点をランダムに結ぶ．(平均次数は** $\langle k \rangle = 2M/N$**)**
2. **適応度分布** $\phi(\beta)$ **を用いて各々の頂点に適応度** $\{\beta_i\}$ **を割りあてる．**
3. **辺** l_{ij} **をランダムに選ぶ．**
4. **次の確率に従って頂点** m **を選択する．**

$$\Pi_m = \frac{(k_m+1)^{\beta_m}}{\sum_j (k_j+1)^{\beta_j}}$$

ここで k_m **は頂点** m **の次数．**

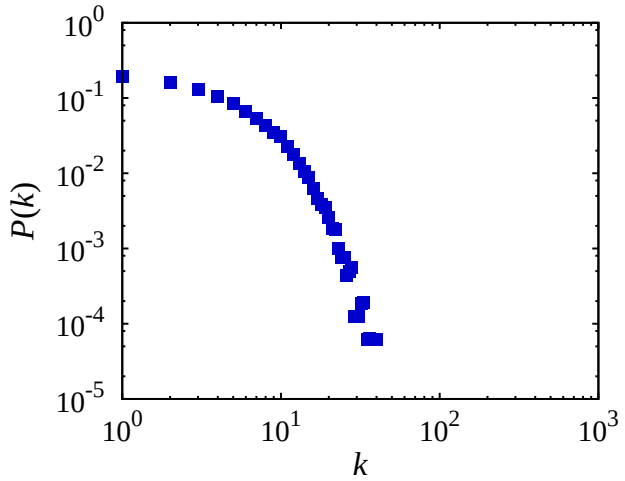
5. **辺** l_{ij} **を辺** l_{im} **へとつなぎ変える．**
6. **ステップ** 3,4,5 **を繰り返す（平衡状態に落ち着くまで）．**



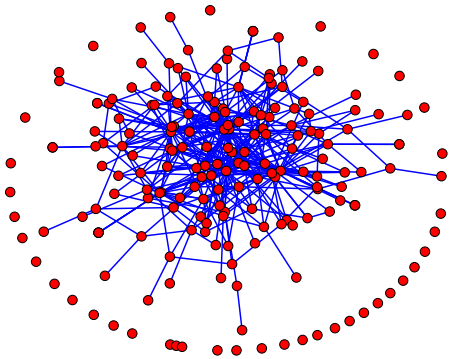
DEX-SMI

生成されたネットワークの次数分布

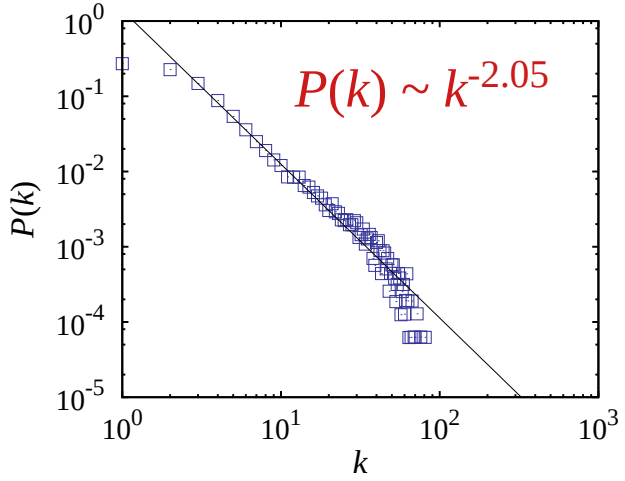
ランダム性がない場合: $\phi(\beta) = \delta(\beta - 1)$



– **指数関数的な減衰**



ランダム性がある場合: $\phi(\beta) = 1, \quad (0 \leq \beta \leq 1)$



- **べき則** (fat-tailed behavior)

べき則を発現させるために，ランダム性が重要である可能性

DEX-SMI

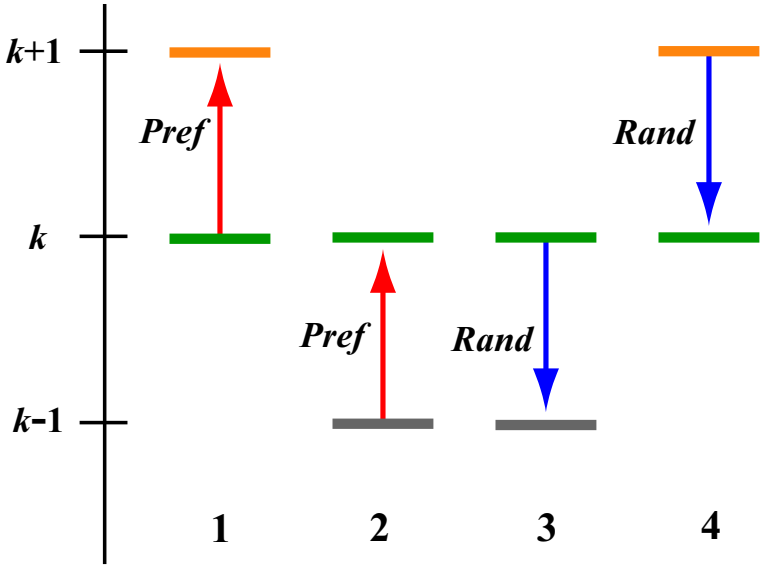
マスター方程式による解析

$f_k(\beta, t)$: **適応度** β **（**$[\beta, \beta + d\beta]$**）を持ち，次数** k **を持つ頂点の個数を表す確率**

（時間に対して連続極限）

$$\frac{\partial f_k(\beta,t)}{\partial t} = -\frac{(k+1)^\beta}{Z(t)} f_k(\beta,t) + \frac{k^\beta}{Z(t)} f_{k-1}(\beta,t) - \frac{k}{N\langle k\rangle} f_k(\beta,t) + \frac{(k+1)}{N\langle k\rangle} f_{k+1}(\beta,t)$$

$$Z(t) = \int d\beta \sum_k (k+1)^\beta f_k(\beta,t)$$



DEX-SMI

問題点

- スケールフリー性を発現させるために何が必要か？
  - BA モデル ⇒ 「優先的選択」「成長」 ⇒ 成長しないネットワークに対しては？
  - 成長しないモデル ⇒ 「成長」がなくても「ランダム性」があればべき則は生成される

- 成長しないモデルの解析について
  - ランダム性が存在することにより，次数分布を解析するのは難しい
    （マスター方程式において，遷移確率が時間に依存する）

目的：平衡状態における次数分布を求める

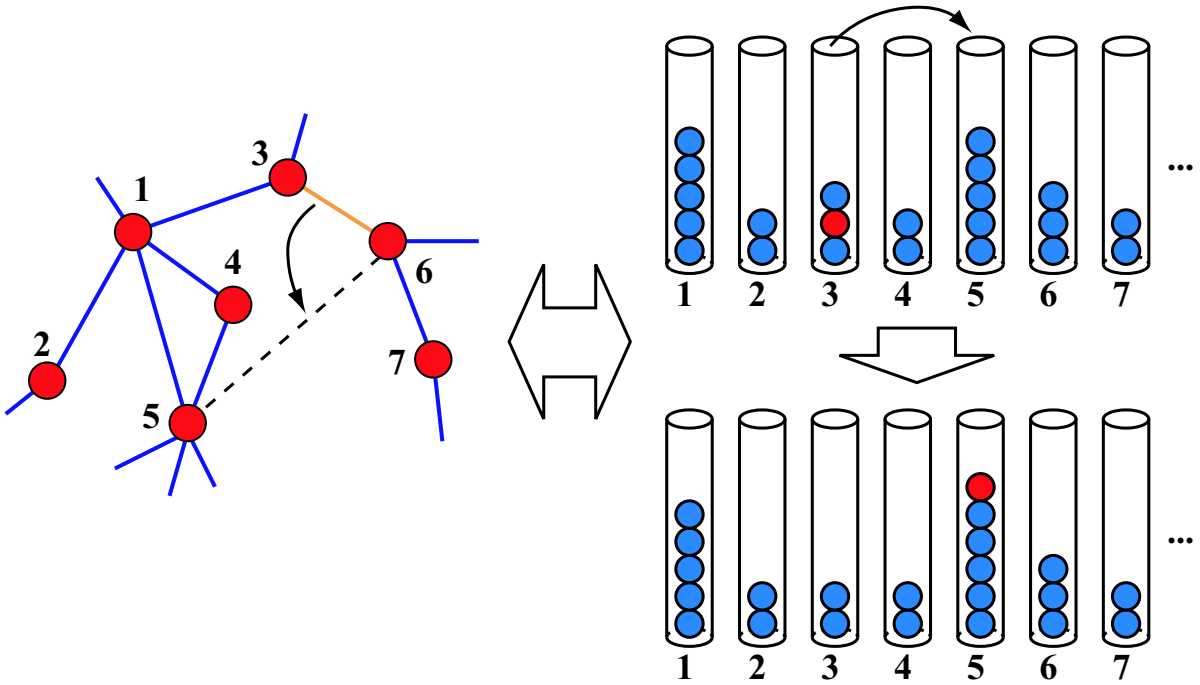
⇓

「壺モデル」と呼ばれる確率モデルの導入と分配関数による解析

ネットワークとランダム系の統計力学との接点

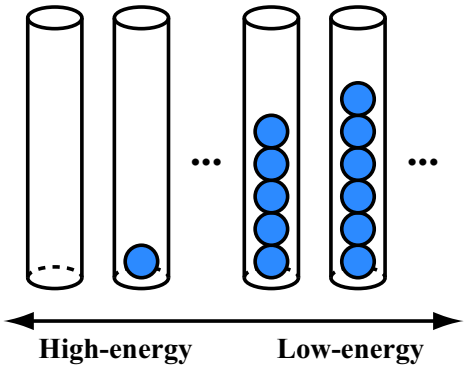
ネットワークと壺モデルの重要な関係

ネットワークにおける次数分布　⇔　壺モデルにおけるボールの占有分布

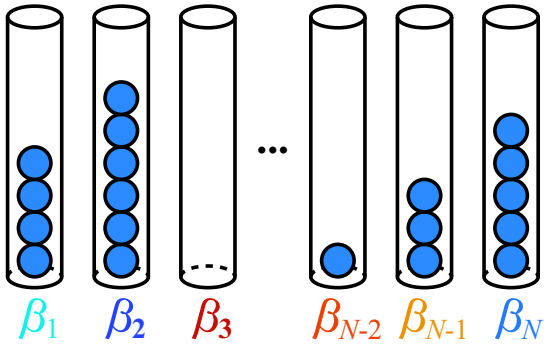


DEX-SMI

壺モデルにおけるエネルギーの定義



- **壺の個数** N**，ボールの個数** M**，密度** $\rho = M/N$
- **各々の壺のエネルギー**:
  $$E(n_i) = -\ln(n_i!)$$
- **ボルツマン因子**:
  $$p_{n_i} = e^{-\beta_i E(n_i)} = (n_i!)^{\beta_i}$$
  $\{\beta_i\}$: **局所逆温度 （**$\phi(\beta)$ **から選ばれる）**
- **熱浴法によるダイナミクス**:
  $$W_{n_l \to n_l+1} \propto (n_l + 1)^{\beta_l}$$



- **成長しないネットワーク生成モデルと同じ遷移確率**
- **このエネルギーの定義を持つ体系の平衡状態におけるボールの占有分布 ⇔ 次数分布**
- **適応度 ⇔ 逆温度**

DEX-SMI

分配関数による壺モデルの解析のポイント 1

分配関数

$$Z_1 = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \frac{p_{n_1}}{n_1!} \cdots \frac{p_{n_N}}{n_N!} \delta\left(\sum_{i=1}^{N} n_i, M\right) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} (n_1!)^{\beta_1 - 1} \ldots (n_N!)^{\beta_N - 1} \frac{1}{2\pi i} \oint dz\, z^{\sum_{i=1}^{N} n_i - M - 1}$$

ある配位 (configuration) において壺 1 に k 個のボールが入っている確率

$$f_k^{(\beta_1)} = \frac{1}{Z_1} \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \delta(n_1, k)(n_1!)^{\beta_1 - 1} \ldots (n_N!)^{\beta_N - 1} \delta\left(\sum_{i=1}^{N} n_i, M\right) = (k!)^{\beta_1 - 1} \frac{Z_2}{Z_1}$$

配位平均において壺 1 に k 個のボールが入っている確率

$$P(k, \beta_1) = \left\langle f_k^{(\beta_1)} \right\rangle_{\{\beta_2, \ldots, \beta_N\}} = (k!)^{\beta_1 - 1} \left\langle \frac{Z_2}{Z_1} \right\rangle_{\{\beta_2, \ldots, \beta_N\}}$$

次数分布

$$P(k) = \int d\beta\, \phi(\beta) P(k, \beta) = \int d\beta\, \phi(\beta) (k!)^{\beta - 1} \left\langle \frac{Z_2}{Z_1} \right\rangle_{\{\beta_2, \ldots, \beta_N\}}$$

DEX-SMI

分配関数による壺モデルの解析のポイント 2

- 分母分子について同時に配位平均を取るのは困難
- 計算を進めると，分配関数の対数の配位平均 $\langle \ln Z_1 \rangle_{\{\beta_2,\ldots,\beta_N\}}$ の計算をすればよいことがわかる
- 分配関数の対数の配位平均も一般的には計算が困難 $\Rightarrow$ **レプリカ法**

レプリカ法における恒等式

$$\langle \ln Z_i \rangle_{\{\beta_2,\ldots,\beta_N\}} = \lim_{m\to 0} \left(\frac{\langle Z_i^m \rangle_{\{\beta_2,\ldots,\beta_N\}} - 1}{m} \right) \qquad (i = 1, 2)$$

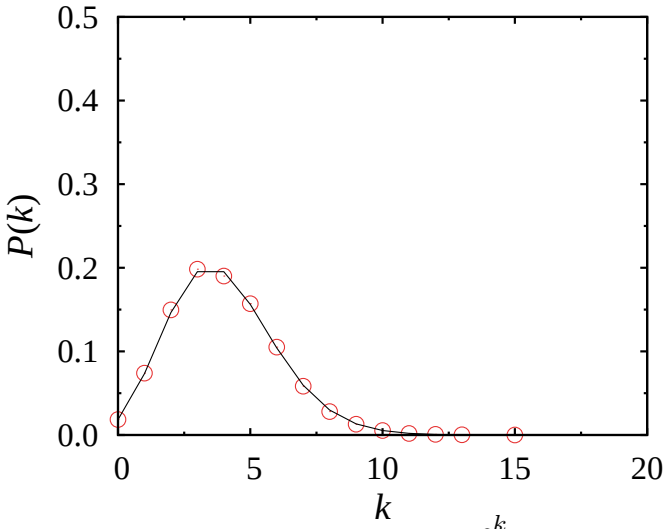
鞍点方程式 $\rho = \dfrac{z_{\mathsf{s}}}{G(z_{\mathsf{s}})} \dfrac{d}{dz_{\mathsf{s}}} G(z_{\mathsf{s}})$ $\left(G(z) = \displaystyle\int d\beta\ \phi(\beta) \sum_{n=0}^{\infty} (n!)^{\beta-1} z^n \right)$

$$P(k) = \int \phi(\beta) \frac{(k!)^{\beta-1} z_{\mathsf{s}}^k}{\sum_{n=0}^{\infty} (n!)^{\beta-1} z_{\mathsf{s}}^n} d\beta$$

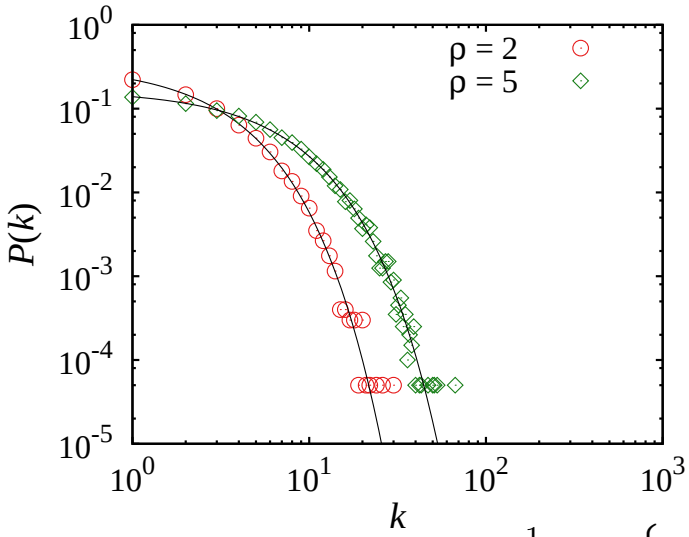
DEX-SMI

解析結果と数値実験

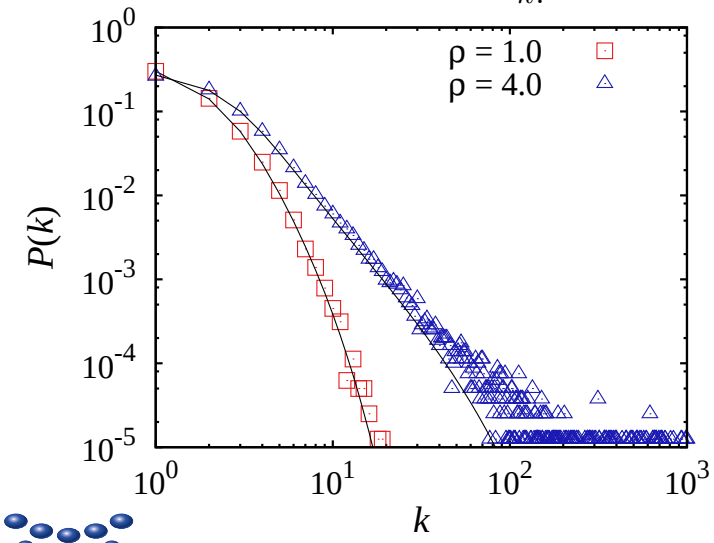
$N = 1000$ $(N = 4000)$
averaged over 20 realizations



$\phi(\beta) = \delta(\beta), \qquad P(k) = e^{-\rho}\dfrac{\rho^k}{k!}$



$\phi(\beta) = \delta(\beta - 1), \qquad P(k) = \dfrac{1}{1+\rho}\exp\left\{-k\ln\left(1+\dfrac{1}{\rho}\right)\right\}$



$\phi(\beta) = 1, \quad (0 \le \beta \le 1),$

$$P(k) = \int_0^1 \frac{(k!)^{\beta-1} z_{\mathrm{s}}^k}{\sum_{n=0}^{\infty}(n!)^{\beta-1} z_{\mathrm{s}}^n} d\beta$$

$$\Rightarrow P(k) \sim k^{-2}(\ln k)^{-2} \qquad (\rho \gg 1)$$

$z_{\mathrm{s}} = 0.660 \ (\rho = 1)$

$z_{\mathrm{s}} = 0.967 \ (\rho = 4)$

DEX-SMI

生成問題のまとめ

様々なクラスの生成モデル

- growing (dynamical)
- nongrowing (static)
- nongrowing & dynamical $\Rightarrow$ **平衡統計力学・ランダム系の統計力学**

研究のポイント

- **従来の生成モデルとは異なるクラスのモデルにおいて，スケールフリー性が発現するか模索**
- **平衡統計力学において解析が進められている他のモデルとの対応関係**

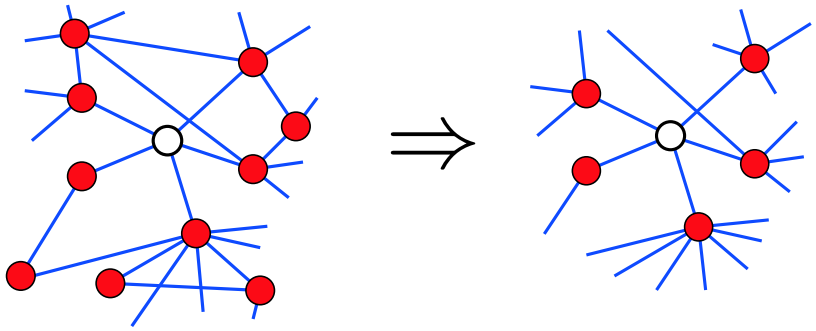
スケールフリー性を発現させるための条件

	概念 1	概念 2
BA モデル	優先的選択	成長
成長しない生成モデル	優先的選択	ランダム性

DEX-SMI

次数分布とダイナミクスとの関係

ここで特に注目するもの … 次数・次数分布



- ネットワーク全体
  - 隣接行列全体を扱う
  - 様々な相関が存在
  - 一般的に取り扱うのは困難
- 次数分布のみに着目
  - 隣接行列の情報を縮約（周辺化）
  - 数理的に取り扱いやすい
  - 次数を考慮するだけでも，ダイナミクスに大きな影響（⇐ 二体相互作用）

■ 数理的な生成モデル（定性的・定量的な知見）⇒ ネットワーク上でのダイナミクス

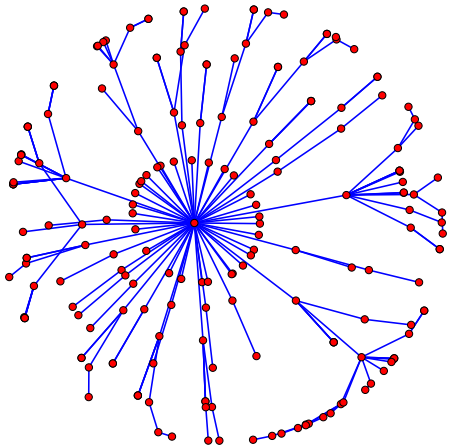
– 次数分布に着目するだけでどの程度数理システムに対する影響を考慮できるのか？

– 生成の数理モデル ⇒ ダイナミクス研究・応用へどのようにつなげるか？

DEX-SMI

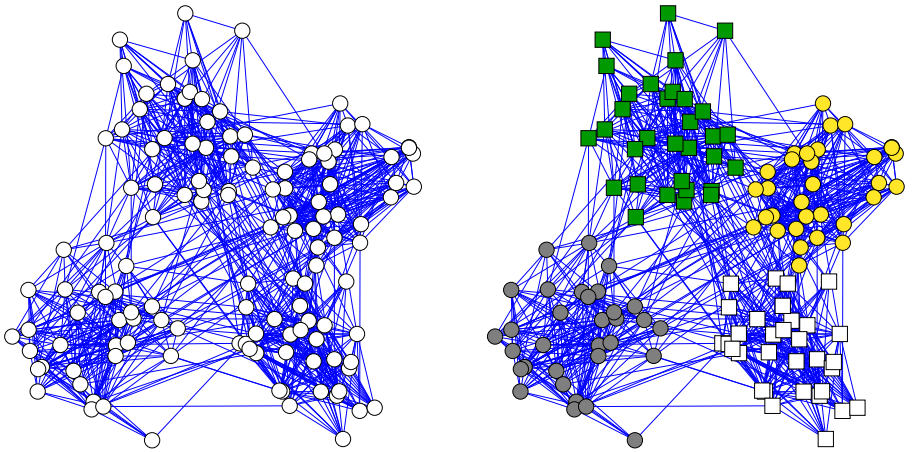
ネットワークとデータマイニング

- 『スケールフリー性』だけが複雑ネットワーク研究の論点ではない
- 非一様性
  ⇒ 頂点ごとに環境が「大きく」異なる



コミュニティ検出

- グラフ理論・アルゴリズム論
- 複雑ネットワーク研究の視点からの指標
- 「何をコミュニティとして考えるか」
  ⇒ 様々な手法が存在



DEX-SMI

Newman の手法

[Newman and Girvan, 2004]

統計的指標: modularity Q **の導入**

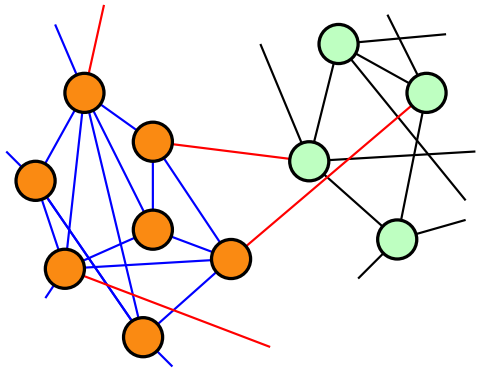
$$Q = \sum_{i=1}^{M} \left[\frac{e_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right]$$

M: **コミュニティ（頂点のかたまり）の総数**

e_i: **コミュニティ**i **内に存在する辺の本数**

m: **ネットワーク全体に存在する辺の本数**

d_i: **コミュニティ**i **内のすべての頂点の次数**



Q を大きくするコミュニティ構造が最もらしい

コミュニティ内に存在する辺の本数が多く，コミュニティ間をつなぐ辺の本数は少ない

計算に使用しているものは…

- **コミュニティ内に存在する辺の情報**
- **コミュニティ間を結ぶ辺の情報**

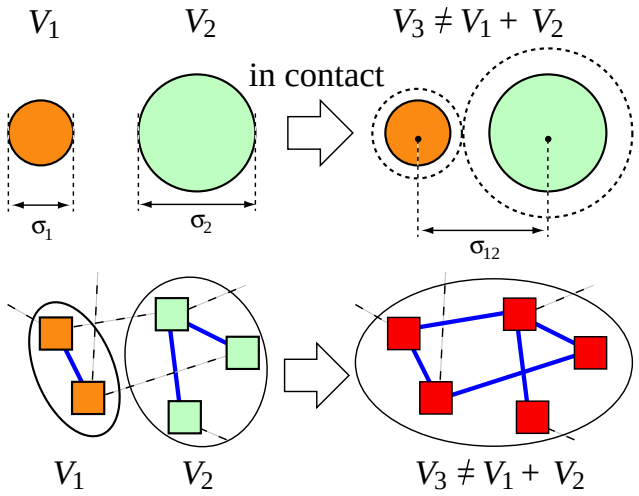
DEX-SMI

非加法的体積を用いた検出

[Ohkubo and Tanaka, 2006]

（物理的）指標: 非加法的体積 V の導入

$$V = n \times \frac{{}_nC_2}{e}$$



n: **頂点の個数**

e: **あるコミュニティの内部につながれた辺の本数**

例： $V_1 = 2,\ V_2 = 4.5,\ V_3 = 10$

V が小さいほど良いコミュニティである

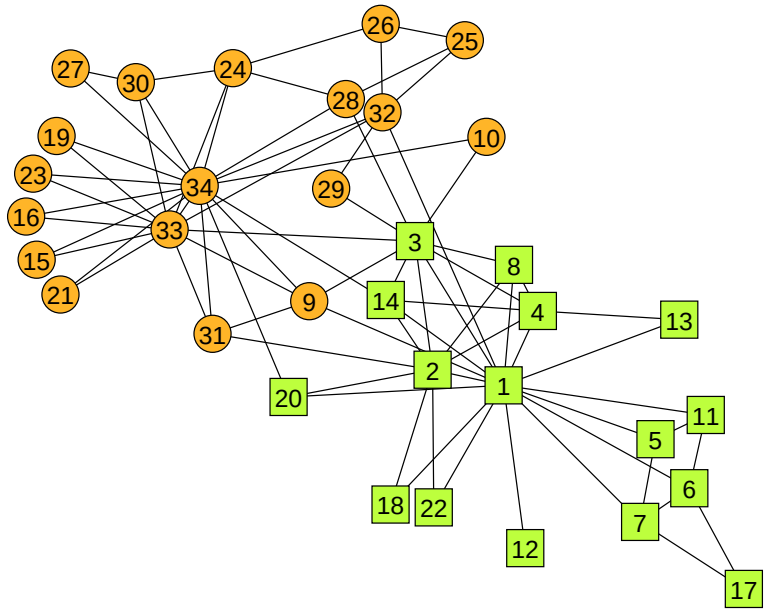
計算に使用しているものは...

- **コミュニティ内に存在する辺の情報**

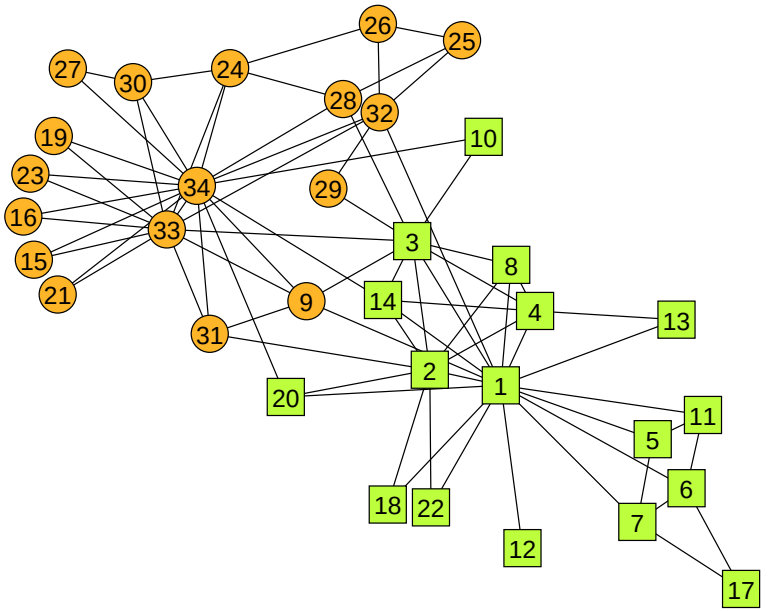
DEX-SMI

karate club ネットワーク

社会学者の分類：



数値実験の結果：



- **辺は** karate club **内における人間関係を示す**
- **ある時期に，**karate club **の管理者** 1 **と先生** 33 **の** 2 **グループに分裂**
- **分裂前の人間関係から，分裂後のグループを検出することが出来るか？**

頂点 10 **を除いて分裂したグループ構造を正しく検出**

DEX-SMI

まとめ

- 複雑ネットワーク研究の紹介
  - 「測る」: 複雑ネットワークを測る指標・データマイニング
  - 「作る」: 複雑ネットワークの生成問題
  - 「使う」: 複雑ネットワーク上でのダイナミクス
- 複雑ネットワークの生成問題
  - これまで提案されていたクラスとは異なる生成モデルの提案
  - 関連する確率モデル・平衡統計力学・ランダム系の統計力学

- 個々の対象へのアプローチのひとつとしての複雑ネットワーク
  - ニューラルネットワーク
  - ベイジアンネットワーク
  - バイオインフォマティクス

  etc.

DEX-SMI