
生命情報学 (7)

スケールフリーネットワーク

阿久津 達也

京都大学 化学研究所
バイオインフォマティクスセンター

内容

- 背景
- グラフとネットワーク
- スモールワールド
- スケールフリーネットワーク
- スケールフリーネットワークの構成法
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

背景

■ システム生物学

- 生命をシステムとして理解
 - 相互作用、ネットワーク推定
 - シミュレーション
 - 安定性解析、制御

■ ネットワーク生物学

- 生命をネットワークとして理解
 - スモールワールド(1998)
 - スケールフリーネットワーク(1999)
 - ネットワークモチーフ(2002)
- ネットワークの構造上の特徴の解析
- ネットワークの動的挙動の解析

内容

- 背景
- グラフとネットワーク
- スモールワールド
- スケールフリーネットワーク
- スケールフリーネットワークの構成法
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

グラフとネットワーク

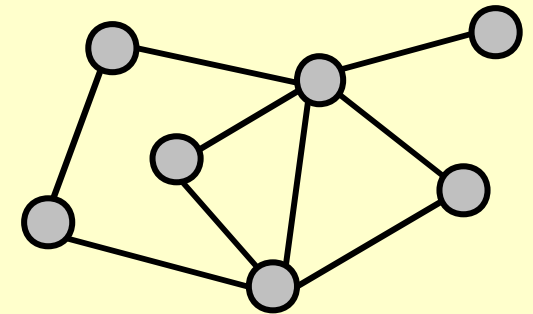
■ グラフ

- 情報科学や離散数学における基礎概念
- グラフは頂点集合と辺集合から構成される
 - 頂点 \Leftrightarrow 物 (例: 化合物)
 - 辺 \Leftrightarrow 2個の物の間の関係 (例: 化学反応)
- **無向グラフ**: 辺に方向無し
- **有向グラフ**: 辺に方向有り

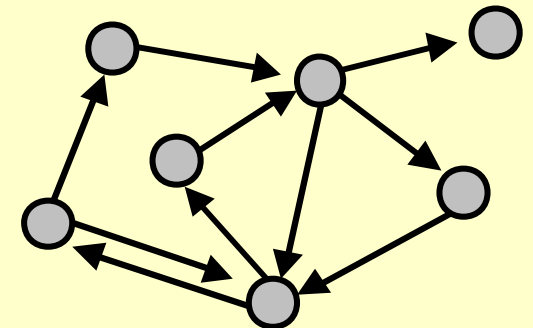
■ ネットワーク

- グラフの辺などに意味や量などのついたもの
- 本講義ではグラフとネットワークを区別しない

無向グラフ



有向グラフ

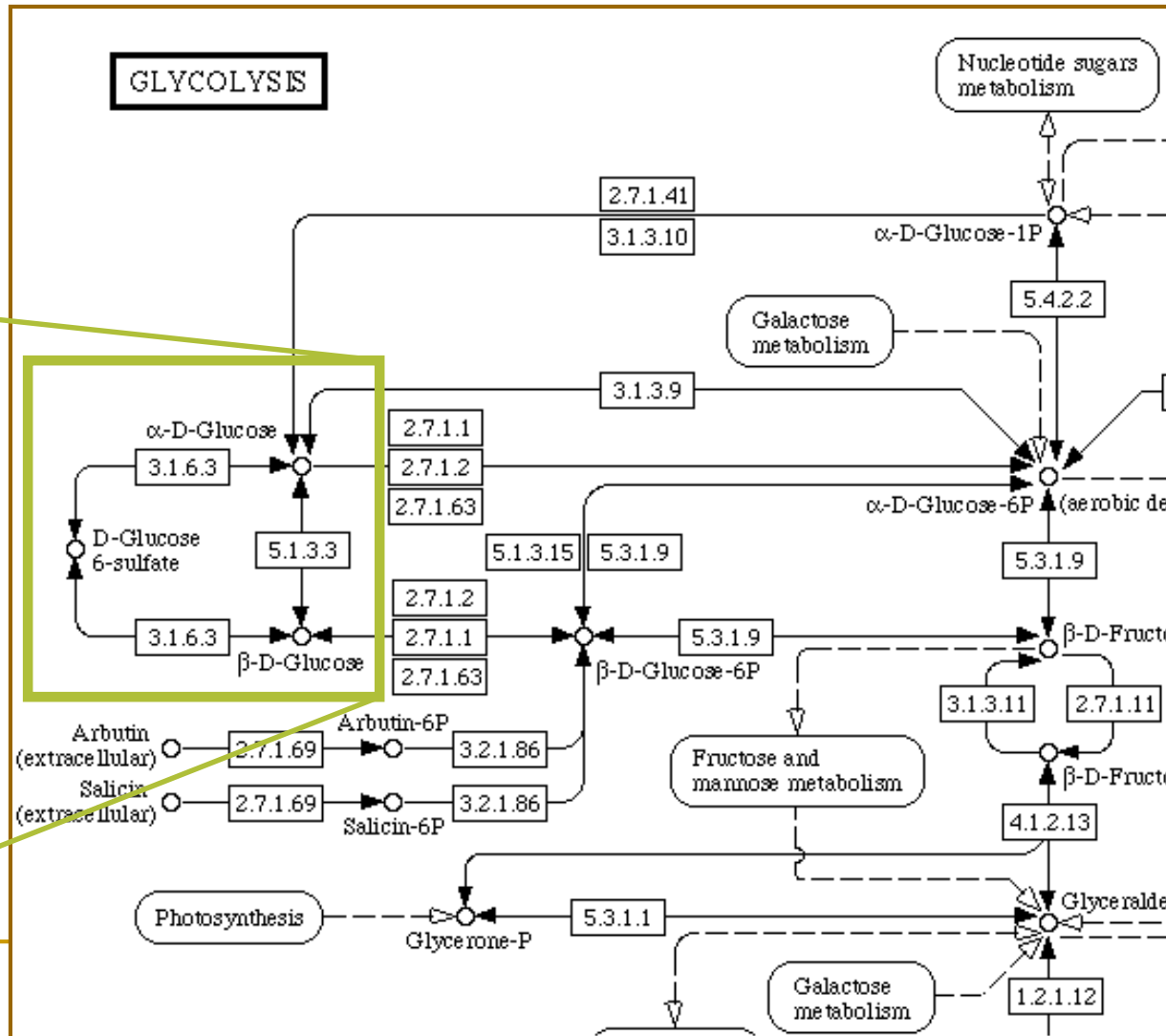
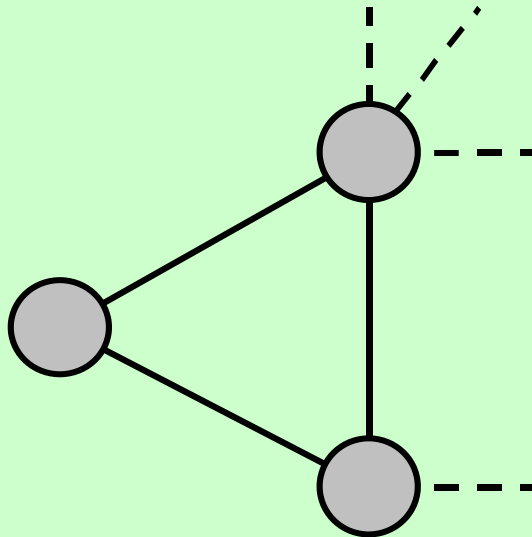


グラフと生物情報ネットワーク

代謝ネットワーク (KEGG)

グラフ

- 点と線で構造を表す



グラフと実際のネットワークの対応



■ 代謝ネットワーク

- 頂点 \Leftrightarrow 化合物、 辺 \Leftrightarrow 代謝反応

■ タンパク質相互作用ネットワーク

- 頂点 \Leftrightarrow タンパク質、 辺 \Leftrightarrow 相互作用

■ 遺伝子ネットワーク

- 頂点 \Leftrightarrow 遺伝子、 辺 \Leftrightarrow 遺伝子間制御関係

■ WWW

- 頂点 \Leftrightarrow WEBページ、 辺 \Leftrightarrow リンク

■ 共著関係

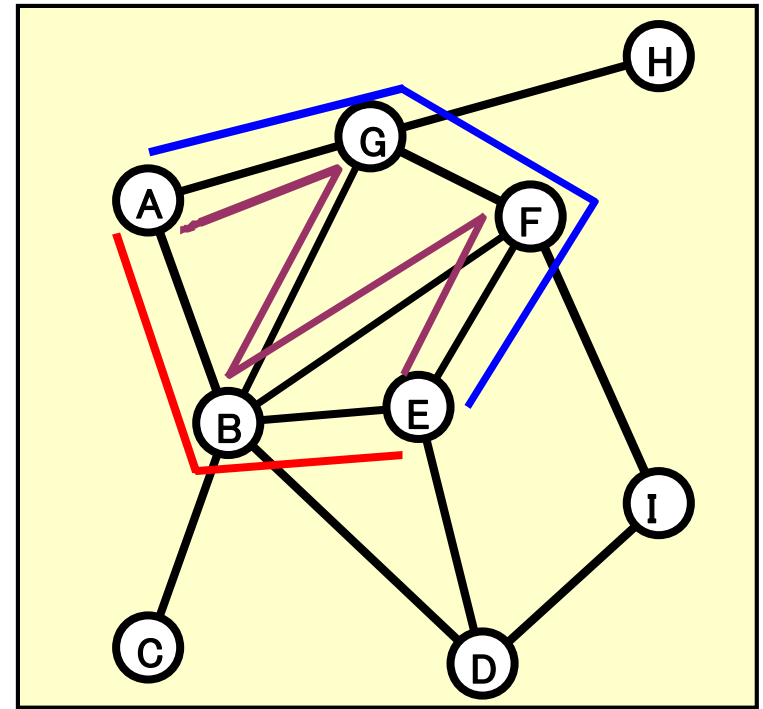
- 頂点 \Leftrightarrow 研究者、 辺 \Leftrightarrow 共著論文の有無

内容

- 背景
- グラフとネットワーク
- **スモールワールド**
- スケールフリーネットワーク
- スケールフリーネットワークの構成法
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

スモールワールド: 頂点間の距離

- 頂点間のパス
 - 二つの頂点をつなぐ辺の列
- パスの長さ
 - パス中の辺の個数
- 頂点間の距離
 - 長さが最短のパスの長さ



- AとEの間のパスの例
 - **パス1**: (A,G), (G,B), (B,F), (F,E) ⇒長さ=4
 - **パス2**: (A,G), (G,F), (F,E) ⇒長さ=3
 - **パス3**: (A,B), (B,E) ⇒長さ=2
- AとEの距離 = 2 (AとIの距離 = 3、CとHの距離 = 3)

スモールワールド：クラスター係数

■ クラスター係数

$$C_i = \frac{2m_i}{k_i(k_i - 1)}$$

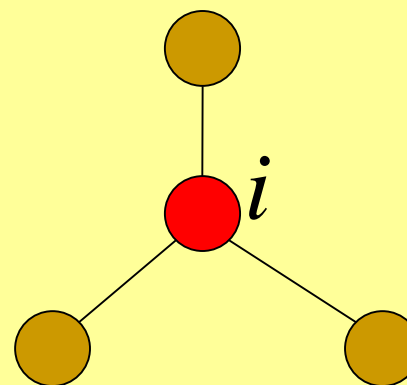
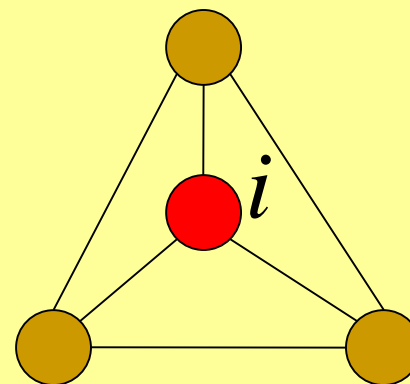
□ m_i : 頂点 i に隣接する頂点間の辺の個数

□ k_i : 頂点 i の次数

■ 頂点のまわりのモジュール性の指標

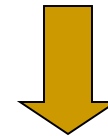
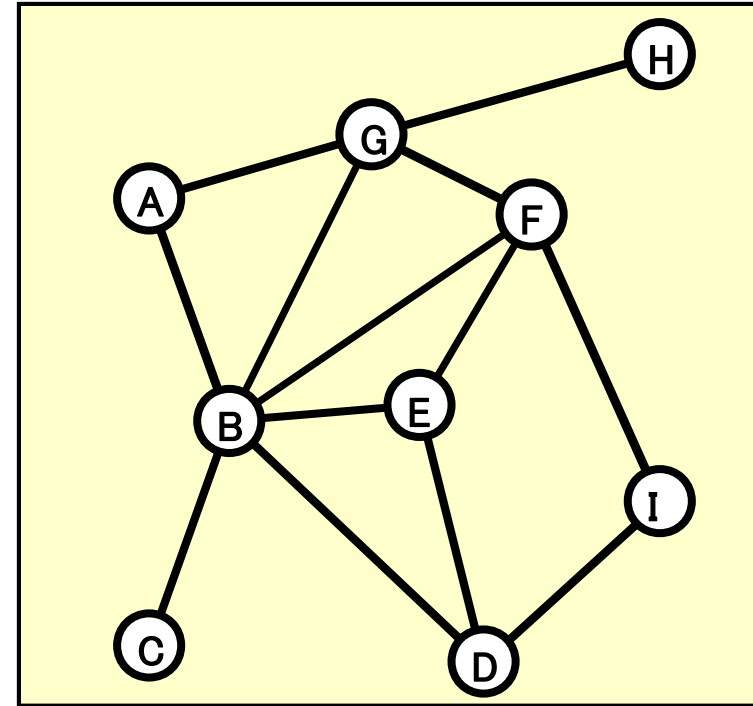
□ $C_i \doteq 1$ <—> クリークに近い

□ m_i の最大値は $\frac{k_i(k_i - 1)}{2}$



スモールワールド

- 任意の2頂点間の距離(最短経路)の平均値が小さく ($O(\log n)$ 以下)、かつ、クラスター係数の平均値が大きいグラフ
- 多くの現実のネットワークはスモールワールドとなる
- WWWの直径(各サイト間のリンク数の平均値)は? ⇒ **約19クリック** (Albert al., Nature, 1999)



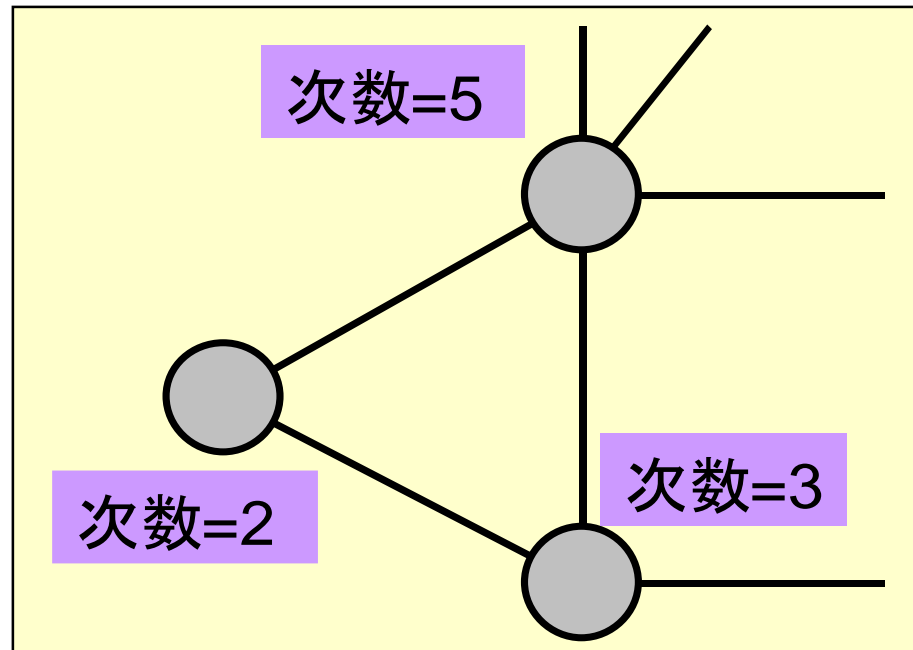
直径 ≤ 3

内容

- 背景
- グラフとネットワーク
- スモールワールド
- スケールフリーネットワーク
- スケールフリーネットワークの構成法
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

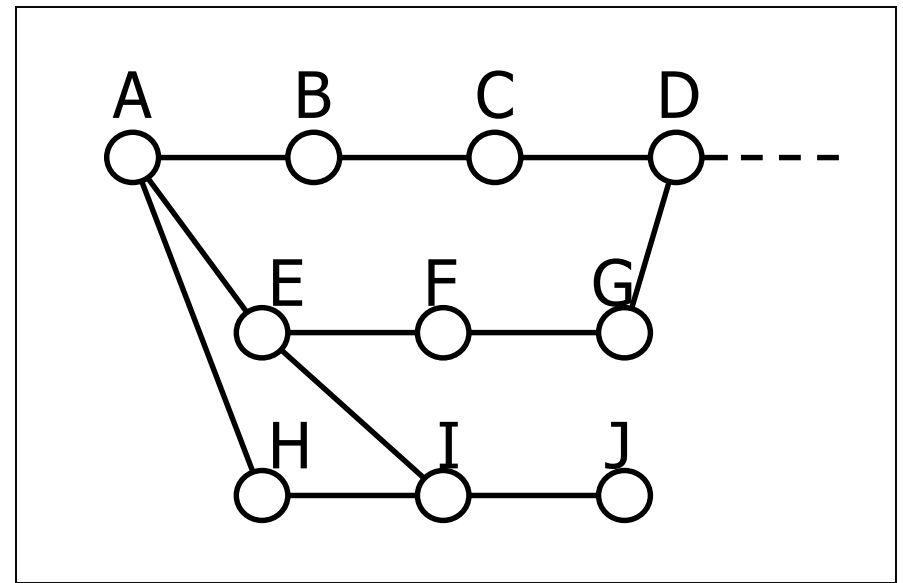
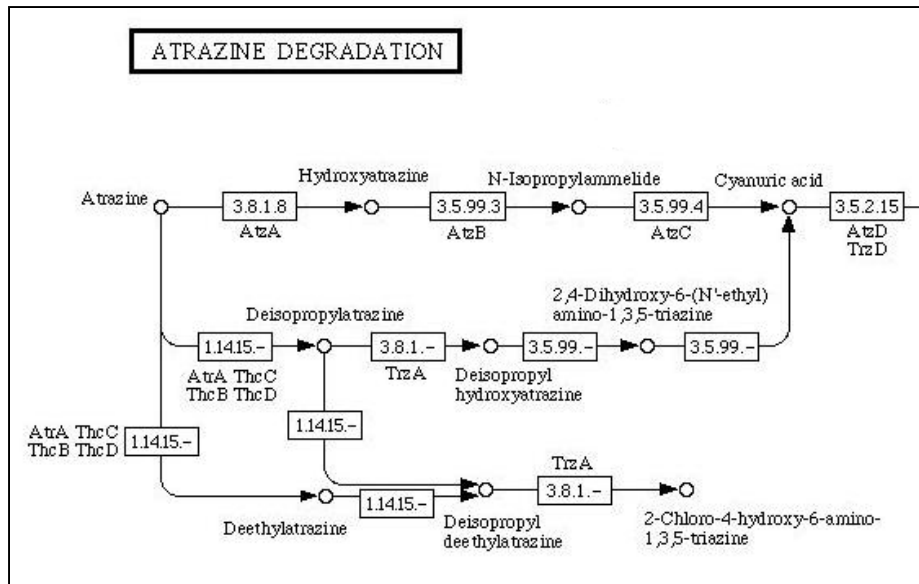
スケールフリーネットワーク(1)

- 頂点の**次数**
 - その頂点につながっている辺の個数
- $P(k)$
 - 次数分布
 - 次数 k の頂点の頻度
- **スケールフリーネットワーク**
 - $P(k)$ がべき乗則に従う



$$P(k) \propto k^{-\gamma}$$

代謝マップ, グラフ, 次数



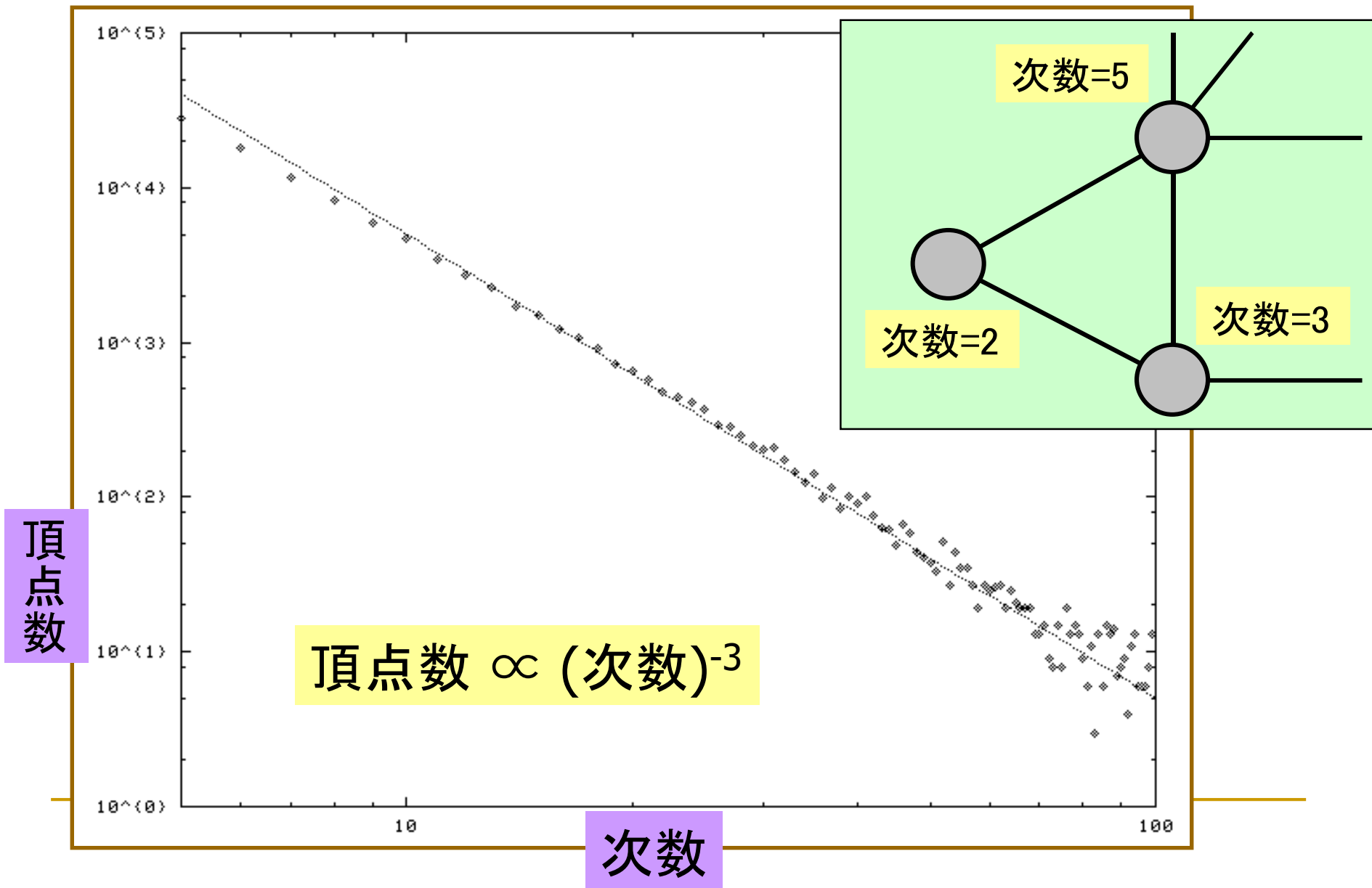
■ 次数

- 次数1の頂点: J
- 次数2の頂点: B, C, D, F, G, H
- 次数3の頂点: A, E, I

■ 次数分布: $P(k)$

- $P(1)=0.1, P(2)=0.6, P(3)=0.3, P(4)=P(5)=P(6)=\dots=0$

スケールフリーネットワーク (2)

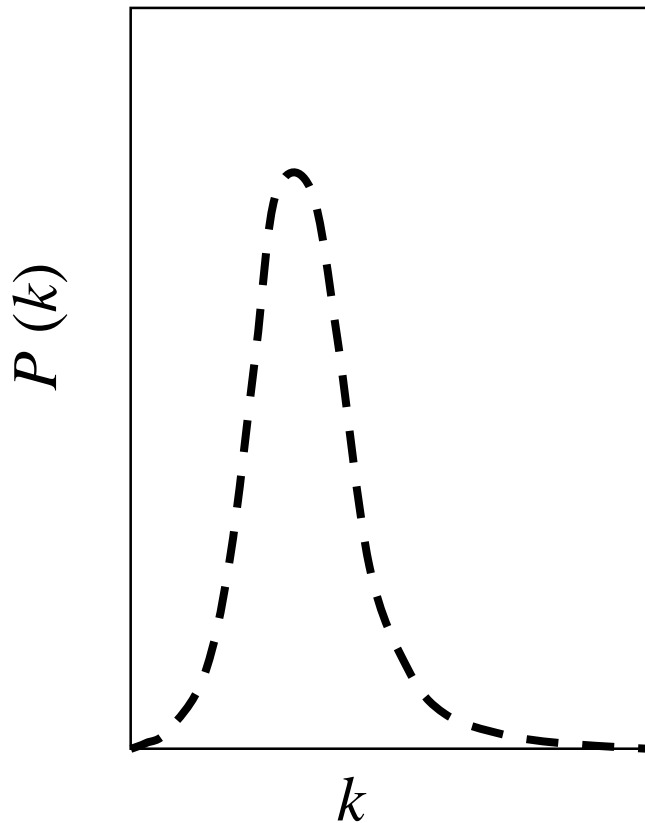


スケールフリーネットワーク (3)

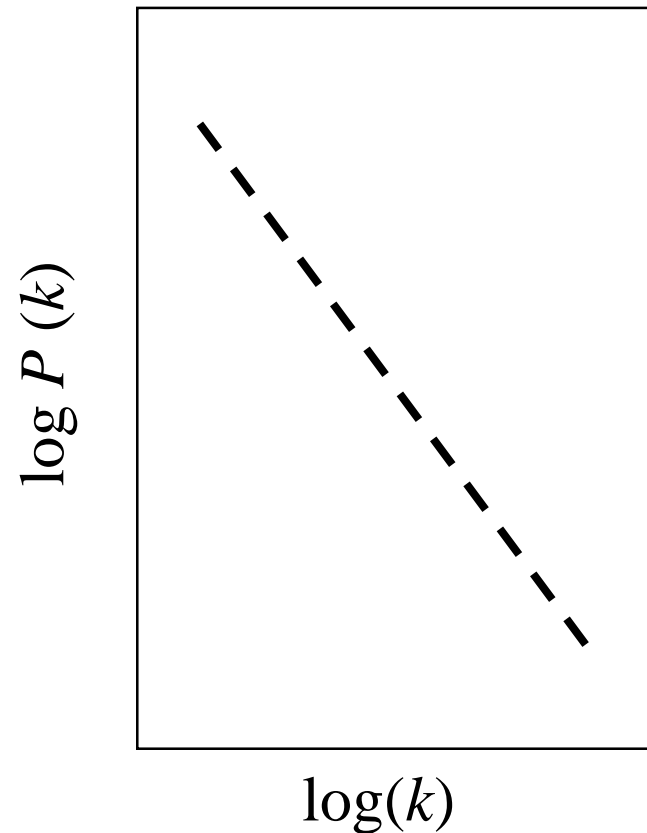
- 次数 k の頂点の個数が $k^{-\gamma}$ に比例 (べき乗則)
 - ランダムな場合 (ポアソン分布: $e^{-\lambda}\lambda^k/k!$) と大差
- Barabasi らが1999年頃に発見。以降、数多くの研究
- 特徴: 有力な頂点 (ハブ) に多くの頂点が連結
- 実際のネットワークにおける $k^{-\gamma}$
 - タンパク質相互作用: $\gamma \doteq 2.2$ (生物種により異なる)
 - 代謝ネットワーク: $\gamma \doteq 2.24$ (生物種により異なる)
 - 映画俳優の共演関係: $\gamma \doteq 2.3$
 - WWW: $\gamma \doteq 2.1$
 - 送電網: $\gamma \doteq 4$

ポアソン分布とべき乗分布

ポアソン分布
(ランダムグラフ)



べき乗分布
(スケールフリーグラフ)



タンパク質ネットワークの解析

- タンパク質相互作用のネットワークもべき乗則に従う(酵母の場合)
 - 頂点:タンパク質
 - 辺:相互作用の有無
- 次数5以下の頂点(全体の93%)
 - 21%程度が必須(生存に必要)
- 次数16以上の頂点(全体の0.7%)
 - 62%程度が必須
 - 次数の高い頂点は**ハブ**と呼ばれ、重要な役割を果たすものが多い

内容

- 背景
- グラフとネットワーク
- スモールワールド
- スケールフリーネットワーク
- **スケールフリーネットワークの構成法**
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

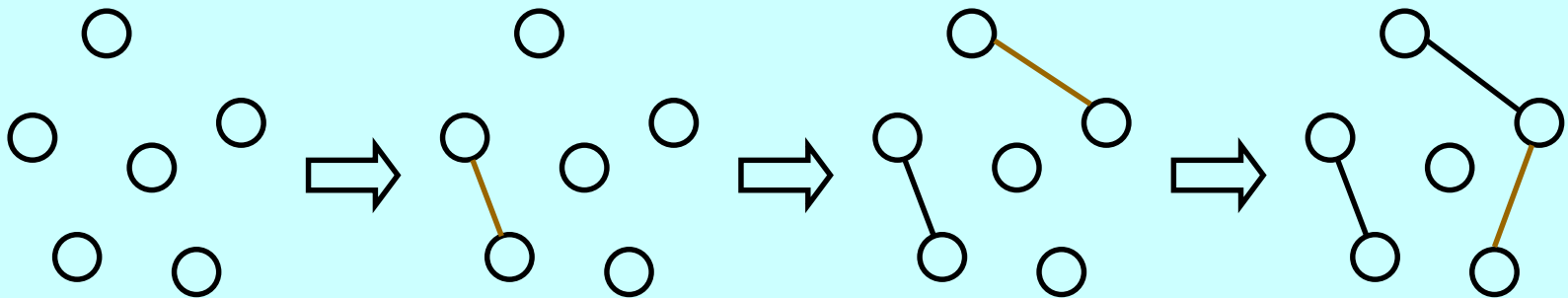
スケールフリーネットワーク構成法: 優先的選択法

- 優先的選択法(優先的選択型成長モデル) [Barabasi & Albert 1999]
 - 別名: Rich-get-richer モデル
- 構成法(ほぼ、 k^{-3} のべき乗則従うネットワークを生成)
 - m_0 個の頂点から成るグラフを構成する
 - 以下のステップを必要なだけ繰り返す
 - 現在のグラフに新たな頂点 v を追加する
 - v から既存の頂点に、 $\text{deg}(v_i)/(\sum_j \text{deg}(v_j))$ に従う確率で、ランダムに辺を張る(全部で m 本の辺を張る)
- 参考: ランダムグラフの構成法
 - N 個の頂点を配置
 - 以下の操作を辺の個数が指定の数になるまで繰り返す
 - 任意の2頂点をランダムに選んでは辺を追加

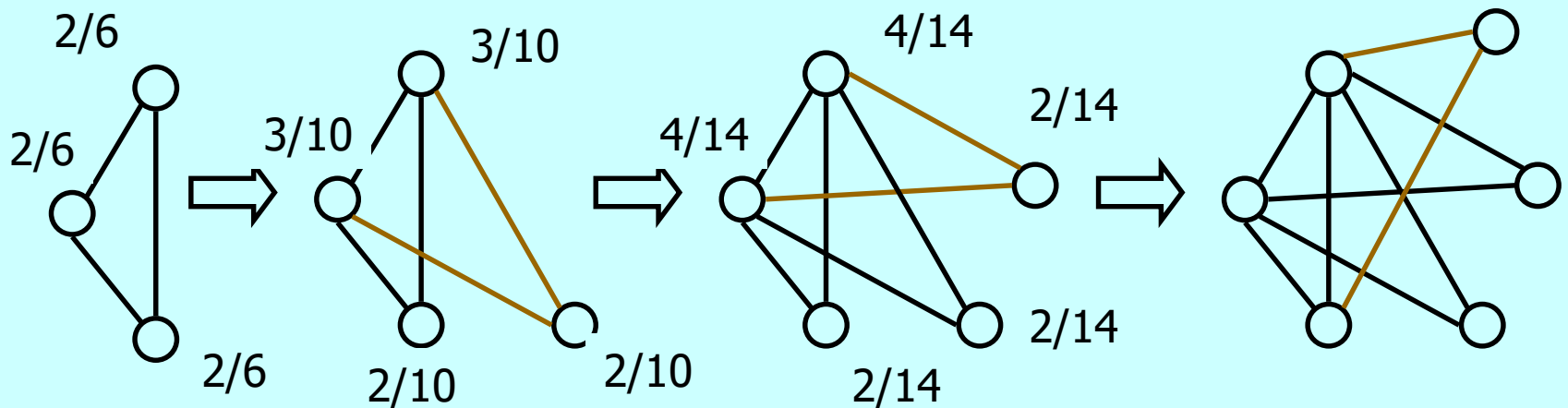
(もしくは、一様な確率 p で任意の2頂点間に辺を引く)

ランダムネットワーク vs. スケールフリーネットワーク

ランダムネットワーク



スケールフリーネットワーク



優先的選択法の平均場近似による解析

- $k_i(t)$: (時刻 t_i) に追加された頂点 i の時刻 t における次数

- 時刻 t までに追加された辺の個数 $\doteq mt$

- 時刻 t において頂点 i の次数が増加する確率は $\frac{\partial k_i(t)}{\partial t} = \frac{mk_i(t)}{2mt}$

- この微分方程式を条件 $k_i(t_i)=m$ のもとで解くと $k_i(t) = m \left(\frac{t}{t_i} \right)^{0.5}$

- 時刻 t_n にネットワークが完成したとすると、
次数 k の頂点の生成時刻は、 $k_i(t_n)=k$ を解いて、

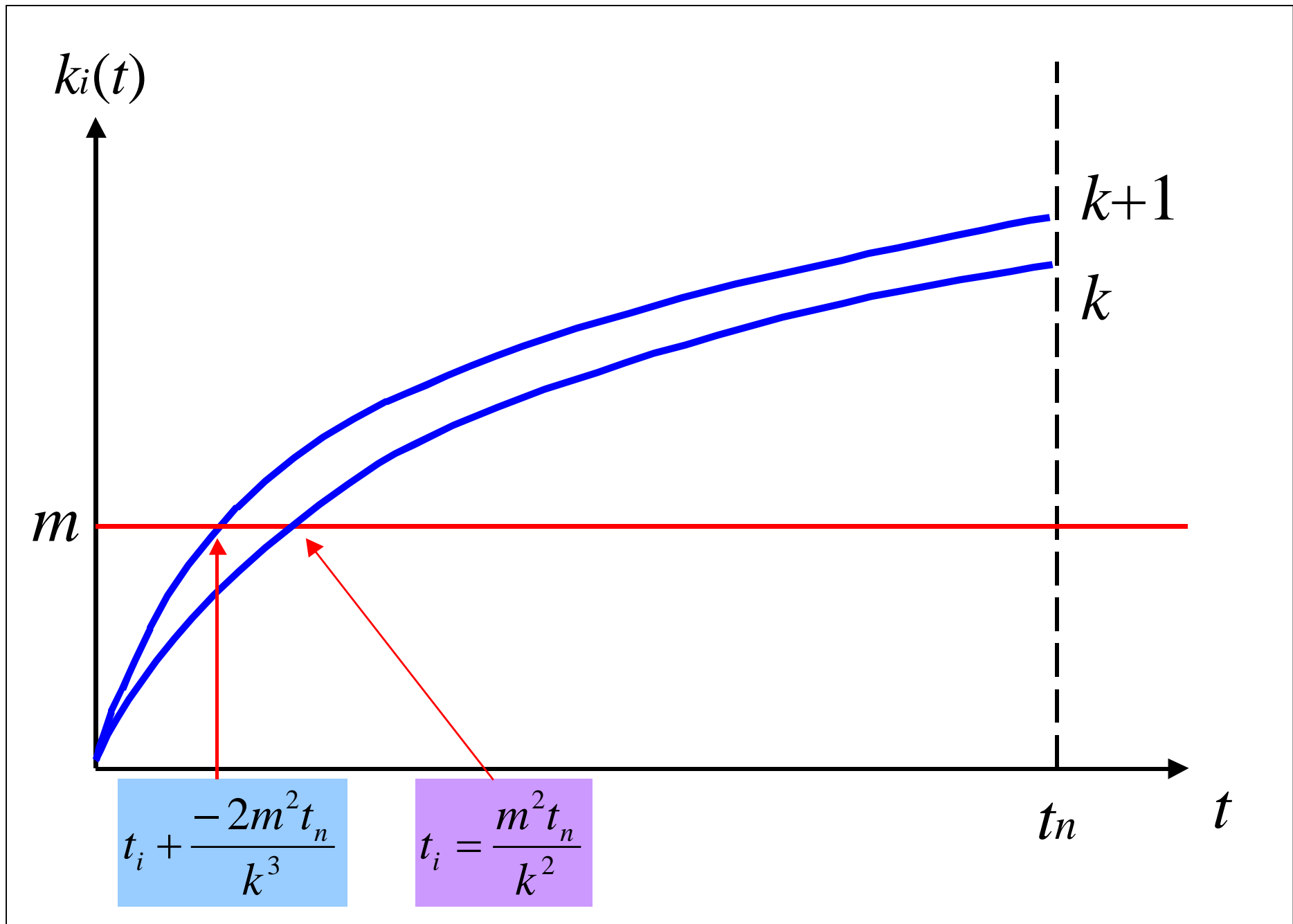
$$t_i = \frac{m^2 t_n}{k^2}$$

- ここで、 k が1だけ増えると、 t_i がどれくらい減るかは、
上の式を k で微分することにより、

$$\frac{-2m^2 t_n}{k^3}$$

- よって、時刻が $2t_n m^2 k^{-3}$ だけ異なると k が1変わる

- よって、次数 k の頂点は $2t_n m^2 k^{-3}$ のオーダーの個数存在



内容

- 背景
- グラフとネットワーク
- スモールワールド
- スケールフリーネットワーク
- スケールフリーネットワークの構成法
 - Preferential Attachment (Rich-get-Richer)
- ネットワークモチーフ

ネットワークモチーフ

■ モチーフ

- 配列解析において現れる機能と関連した配列パターン
- 例: L-x(6)-L-x(6)-L-x(6)-L (ロイシンジッパーモチーフ)

■ ネットワークモチーフ

- (ランダムなネットワークと比べて) 実際のネットワークにおいて頻出する(統計的に有意に)ネットワークのパターン
- ネットワークのパターン: 部分グラフ
- ランダムネットワークの作成: 辺の交換の繰り返し

■ ネットワークモチーフの例

- フィードフォワード制御
- Single Input Module
- Dense Overlapping regulons

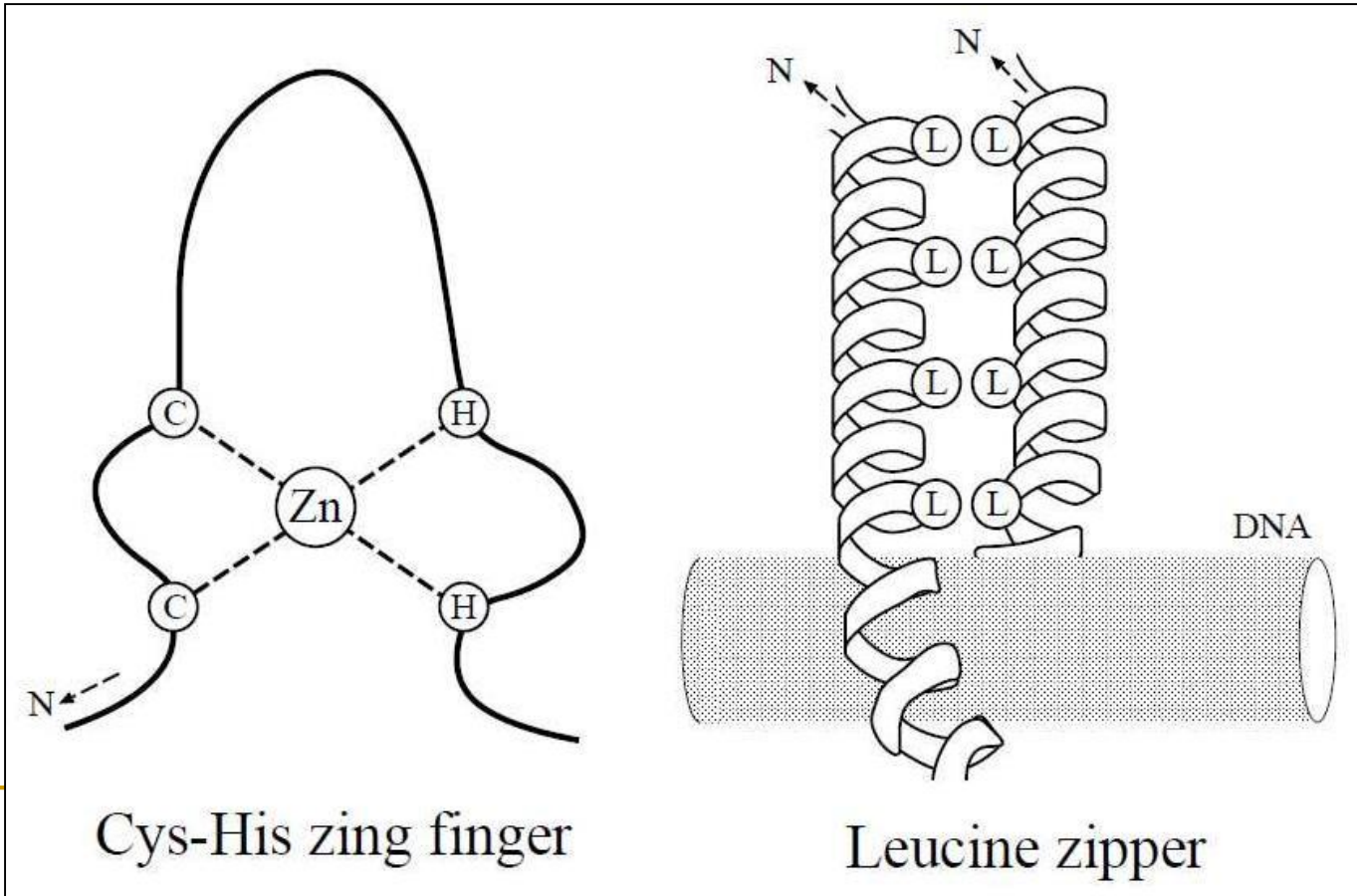
配列モチーフの例

- **ジンクフィンガーモチーフ**

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

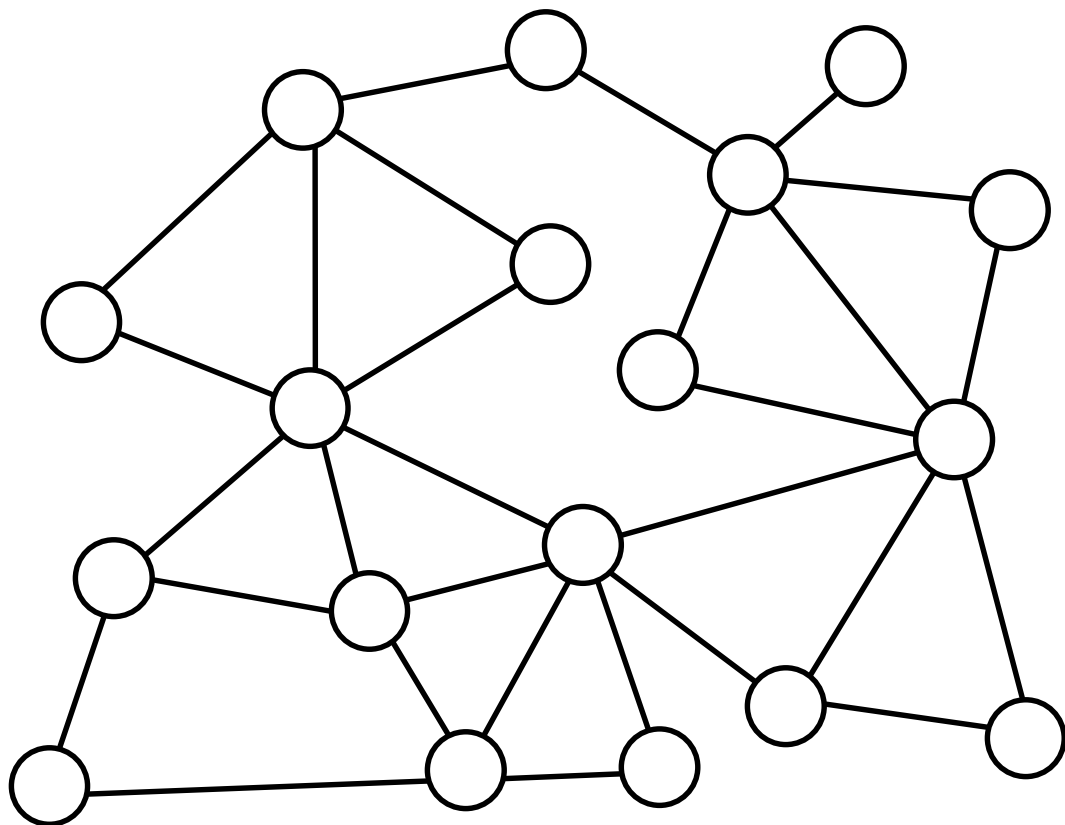
- **ロイシンジッパーモチーフ**

L-x(6)-L-x(6)-L-x(6)-L

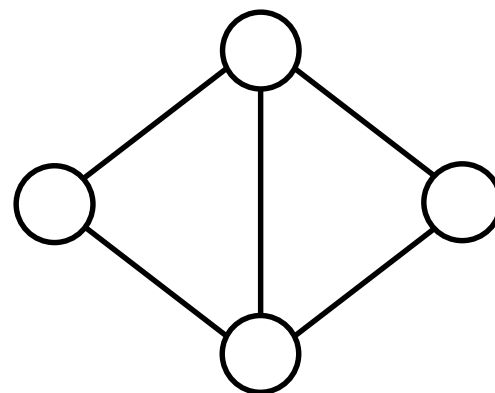


ネットワークモチーフの例 (1)

ネットワーク

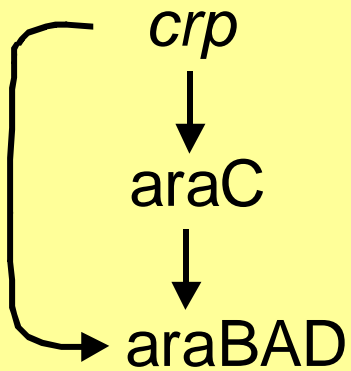
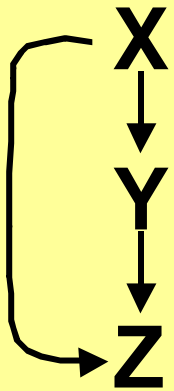


モチーフ

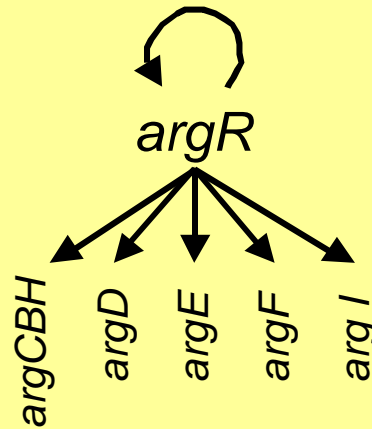
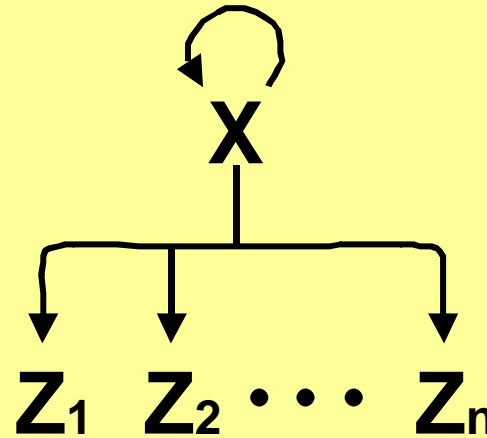


ネットワークモチーフの例 (2)

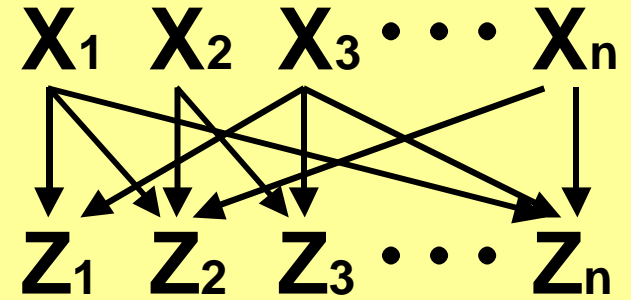
feedforward loop



single input module



dense overlapping regulons



まとめ

- グラフとネットワーク
 - 頂点と辺
- スモールワールド
 - 2頂点間の平均距離が短いグラフ
- スケールフリーネットワーク
 - 次数分布がべき乗則に従うネットワーク
- スケールフリーネットワークの構成法
 - 優先的選択型成長モデル
- ネットワークモチーフ