

# 二部グラフからの射影により得られた ネットワークの次数分布について

ホセ ナチエル

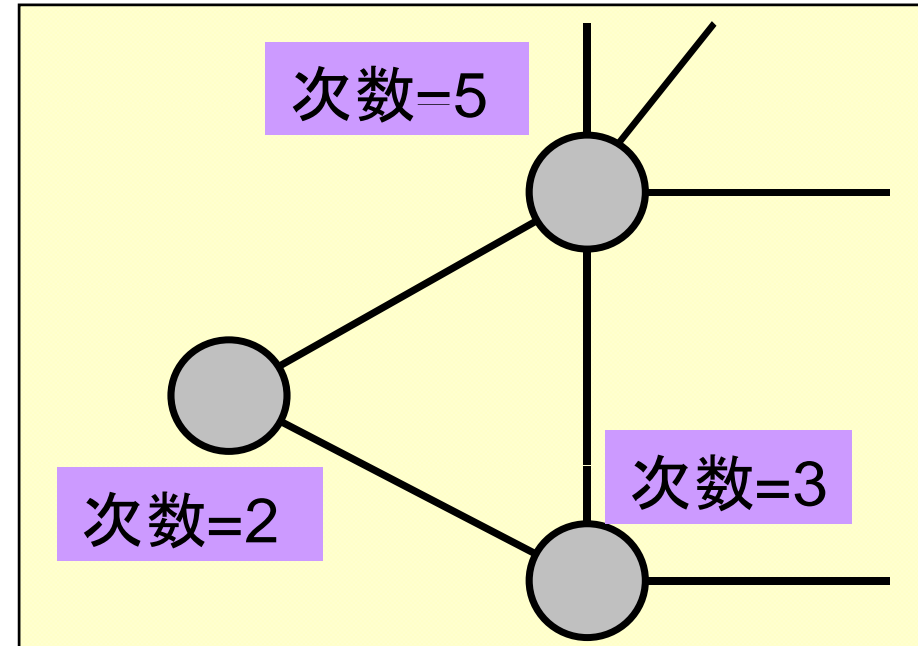
公立はこだて未来大学 複雑系知能学科

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

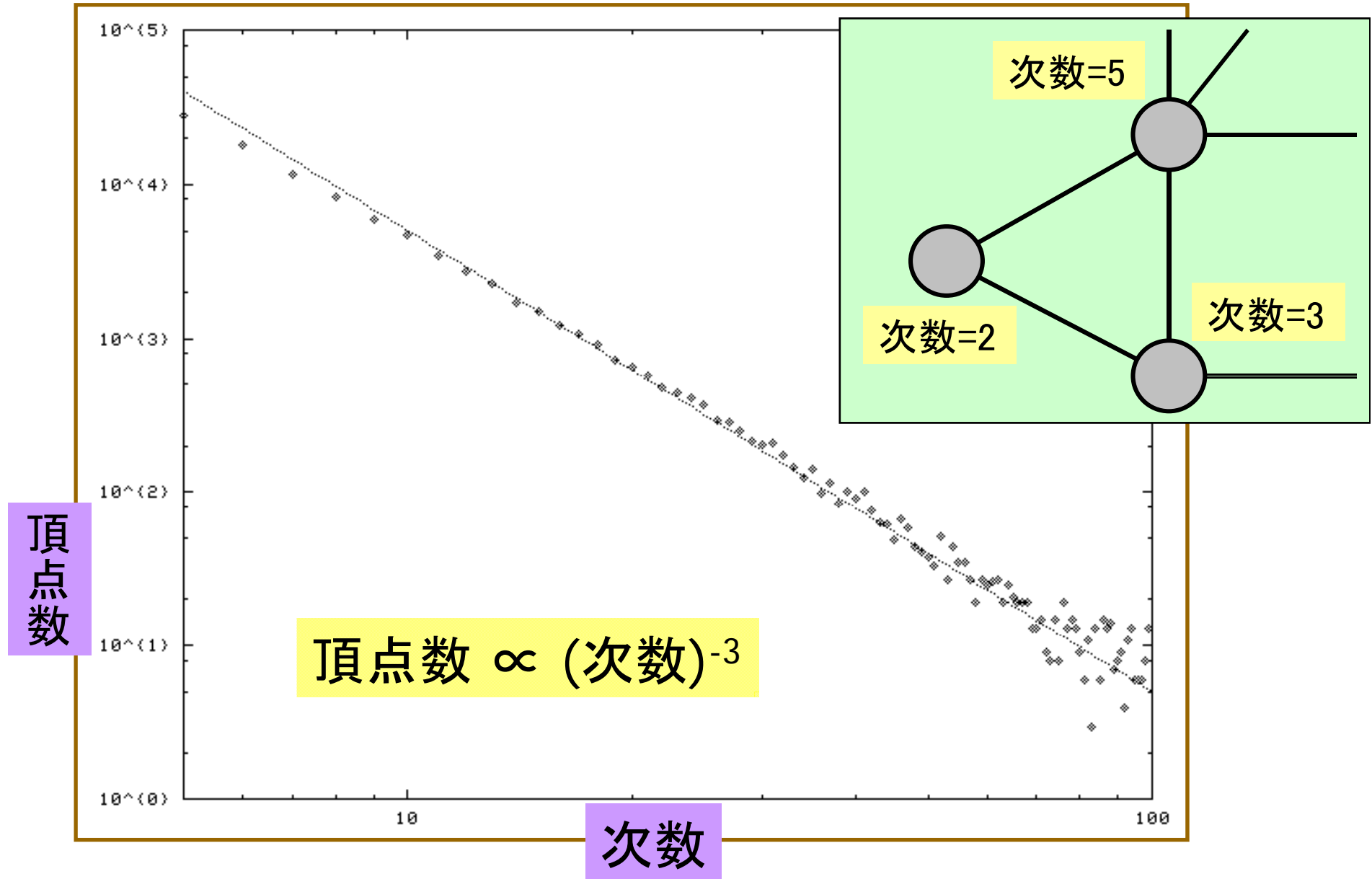
# スケールフリーネットワーク(1)

- 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



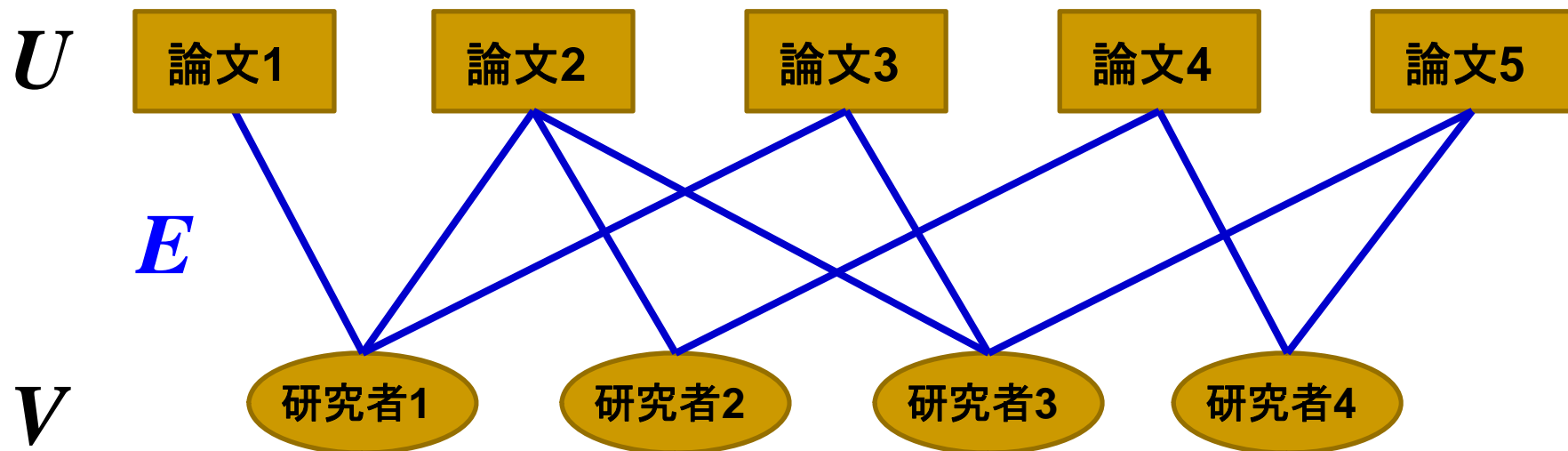
$$P(k) \propto k^{-\gamma}$$

# スケールフリーネットワーク (2)



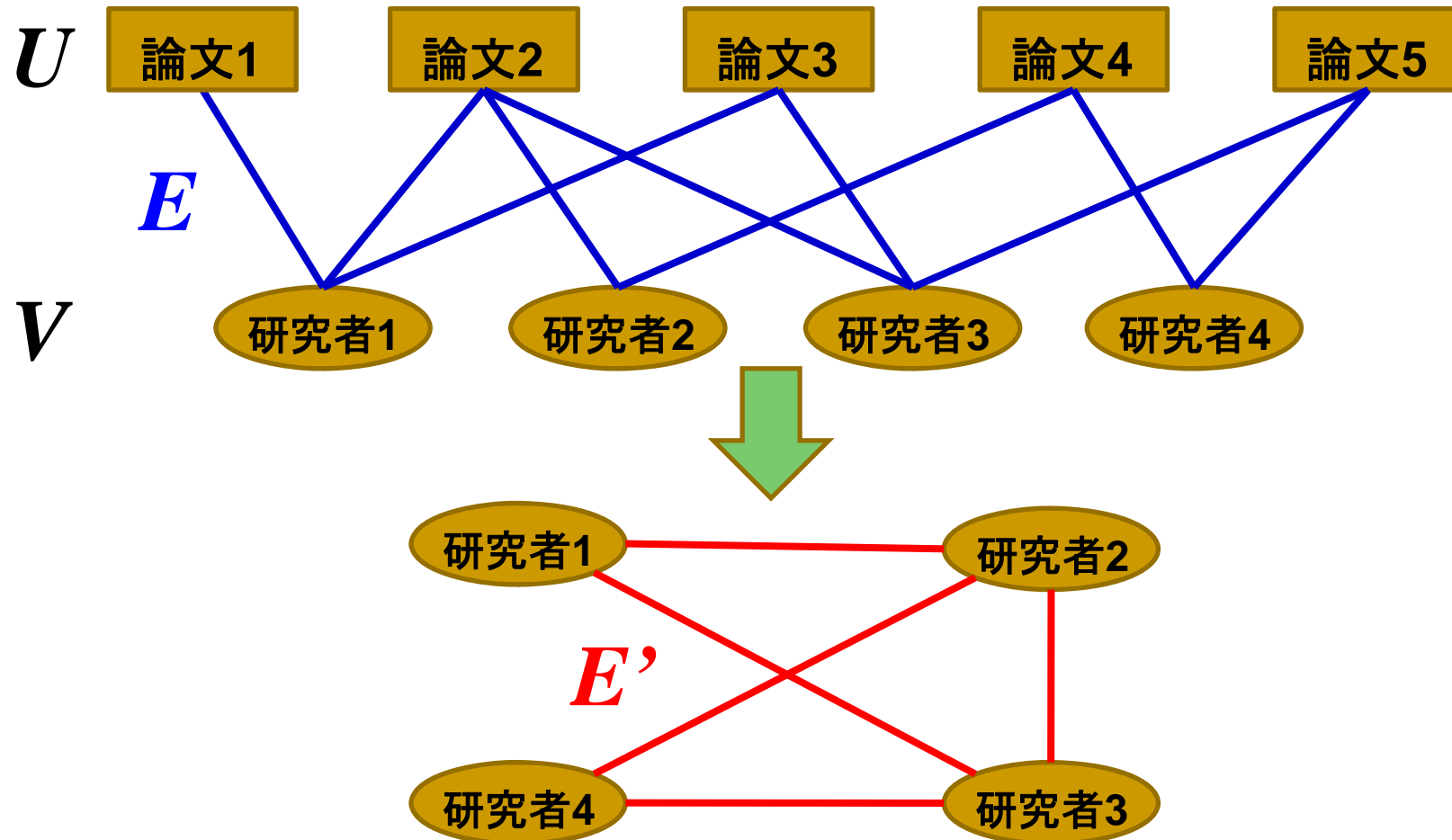
# 二部グラフ

- 多くのネットワークは二部グラフ表現 $G(U, V; E)$ を持つ
  - 代謝ネットワーク:  $U \Leftrightarrow$  化学反応、 $V \Leftrightarrow$  化合物
  - 共演関係:  $U \Leftrightarrow$  映画、 $V \Leftrightarrow$  俳優
  - 共著関係:  $U \Leftrightarrow$  論文、 $V \Leftrightarrow$  研究者



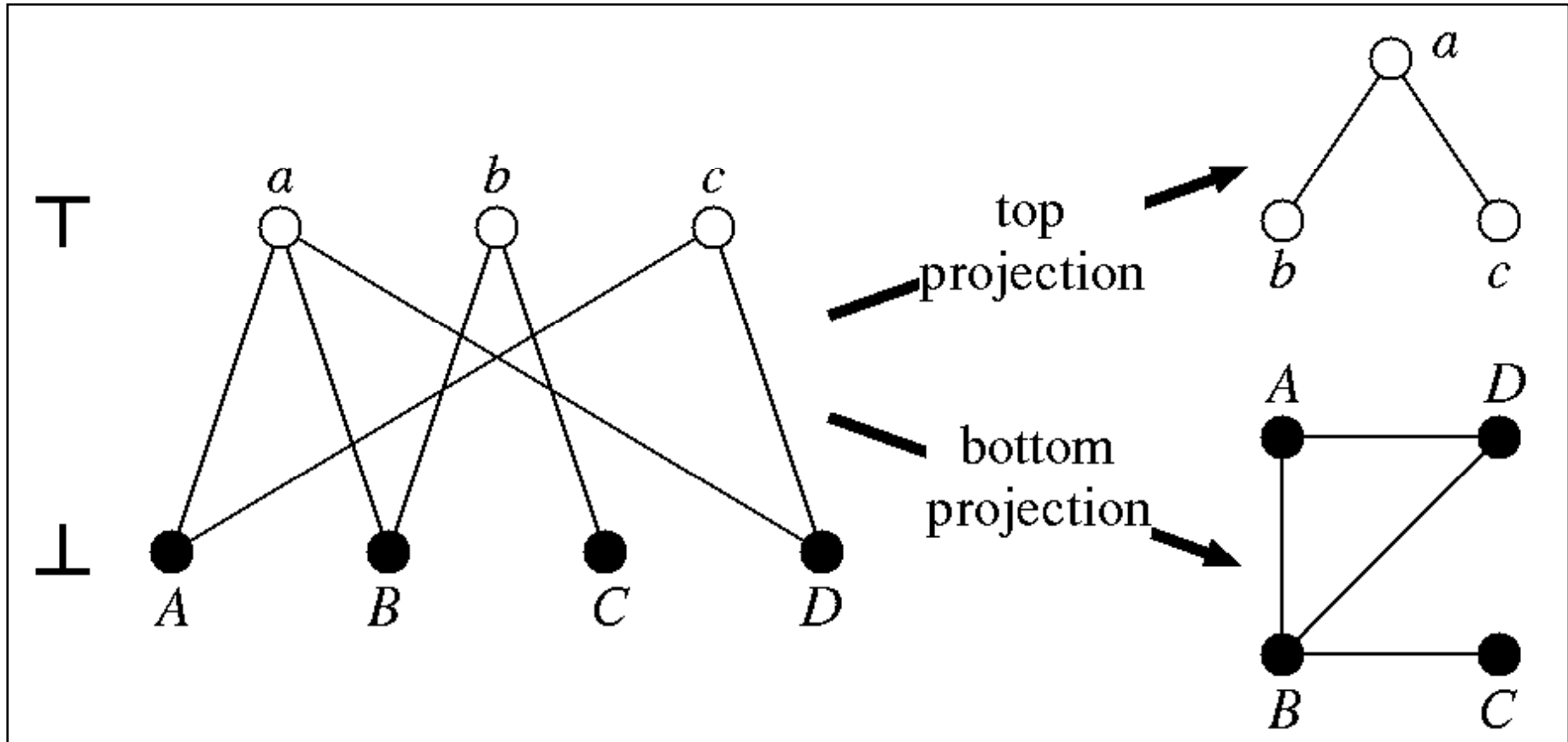
# 二部グラフからの射影

- $G(U, V; E)$  から以下により、 $G'(V, E')$  を構成
  - $\{v_1, v_2\} \in E'$  iff  $(\exists u \in U)(\{u, v_1\} \in E \text{ and } \{u, v_2\} \in E \text{ and } v_1 \neq v_2)$



# Top Projection と Bottom Projection

- Top Projection: 上側の頂点のみを残す
- Bottom Projection: 下側の頂点のみを残す



# 射影と次数分布

以降は、 $G(\top, \perp; E)$  を用いて次のとおりとする

$G'(\top, E')$ : Top Projection

$G'(\perp, E')$ : Bottom Projection

$P_{\top}(k)$ : 上側頂点 ( $\top$ ) の次数分布

$P_{\perp}(k)$ : 下側頂点 ( $\perp$ ) の次数分布

次の3種類の場合に対する射影 (bottom projection) を考える

(ES)  $P_{\top}(k) \propto \exp(-\lambda k)$  かつ  $P_{\perp}(k) \propto k^{-\gamma_2}$

(SE)  $P_{\top}(k) \propto k^{-\gamma_1}$  かつ  $P_{\perp}(k) \propto \exp(-\lambda k)$

(SS)  $P_{\top}(k) \propto k^{-\gamma_1}$  かつ  $P_{\perp}(k) \propto k^{-\gamma_2}$

# 理論的結果

$$(ES) \quad P_{\top}(k) \propto \exp(-\lambda k) \text{ かつ } P_{\perp}(k) \propto k^{-\gamma_2}$$

$$(SE) \quad P_{\top}(k) \propto k^{-\gamma_1} \text{ かつ } P_{\perp}(k) \propto \exp(-\lambda k)$$

$$(SS) \quad P_{\top}(k) \propto k^{-\gamma_1} \text{ かつ } P_{\perp}(k) \propto k^{-\gamma_2}$$

- $P(k)$ : bottom projection後の次数分布とすると

$$(ES) \text{ の場合: } P(k) \propto k^{-\gamma_2}$$

$$(SE) \text{ の場合: } P(k) \propto k^{-\gamma_1+1}$$

$$(SS) \text{ の場合: } P(k) \propto k^{\max(-\gamma_1+1, -\gamma_2)}$$

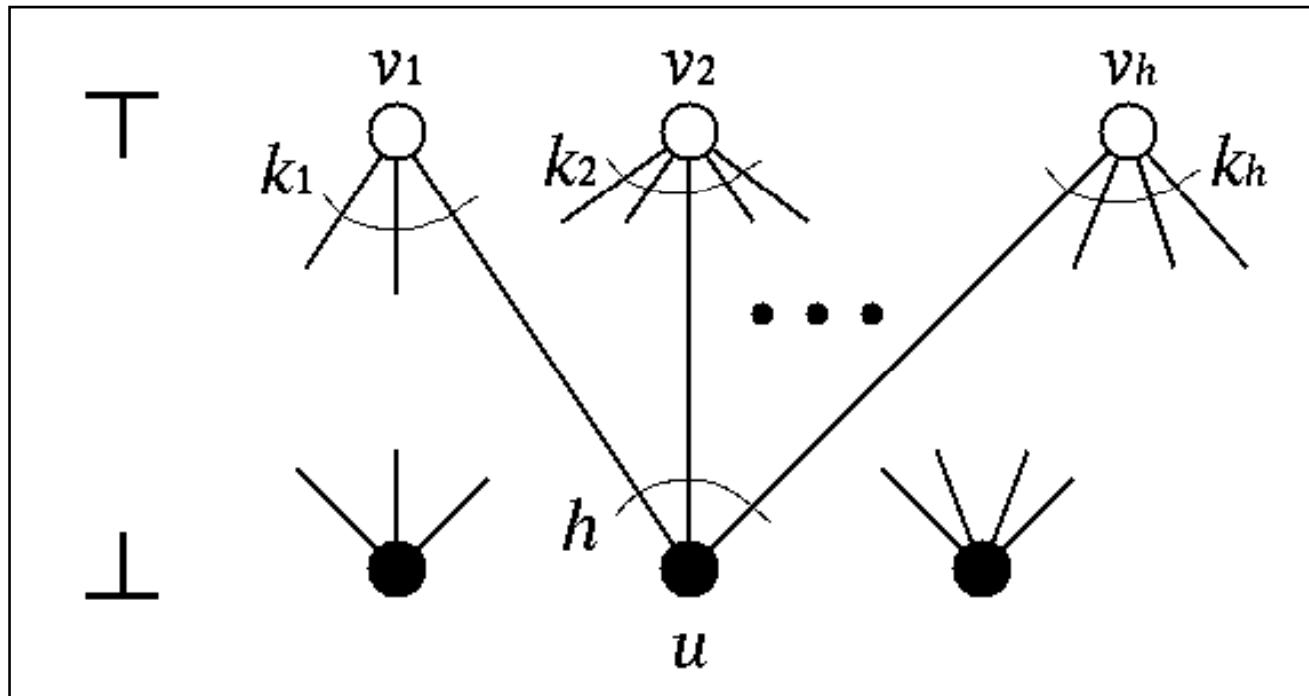
- (ES)の場合は既知 ([Guillaume & Latapy 2006] [Birmele 2009])
- top projectionの結果もこれから導かれる
- (ES)においては $\gamma_1 > 2$ 、(SE)(SS)においては $\gamma_1 > 4$   
(ただし、実際にはより小さい $\gamma_1$ でも適用可と考えられる)



# 理論解析

- 下側からランダムに選んだ頂点  $u$  が射影後に次数  $k$  を持つ確率は

$$\sum_{h=1}^{\infty} P_{\perp}(h) \sum_{k_1+k_2+\dots+k_h=k+h} (k_1 \cdot k_2 \cdot \dots \cdot k_h)^{1-\gamma_1}$$



# (SE)の理論解析(1)

- 以下を近似する式を計算

$$g_{\lambda, \gamma}^{SE}(k) = \sum_{h=1}^{\infty} \exp(-\lambda h) \sum_{k_1+k_2+\dots+k_h=k+h} \left(\frac{1}{\zeta(\gamma)}\right)^h (k_1 \cdot k_2 \cdots k_h)^{-\gamma}$$

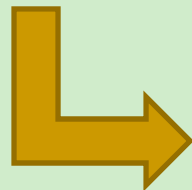
- $f(h, k)$  を次のように定義し、

ただし、 $\gamma = \gamma_1 - 1$

$$f(h, k) = \sum_{k_1+k_2+\dots+k_h=k} \left(\frac{1}{\zeta(\gamma)}\right)^h (k_1 \cdot k_2 \cdots k_h)^{-\gamma}$$

- Sum of power law について既知の性質を利用

$$f(h, k) \propto \begin{cases} hk^{-\gamma} & \text{if } k \text{ is large} \\ \frac{1}{\sqrt{2\pi h\sigma}} e^{-\frac{(k-h\mu)^2}{2h\sigma^2}} & \text{otherwise} \end{cases}$$



$$\hat{f}(h, k) = \begin{cases} hk^{-\gamma} & \text{if } k > h^{1+\delta} \\ 1 & \text{otherwise} \end{cases}$$

## (SE)の理論解析(2)

- 以下により  $P(k) \propto k^{-\gamma+1}$  を得る(下限の導出は省略)

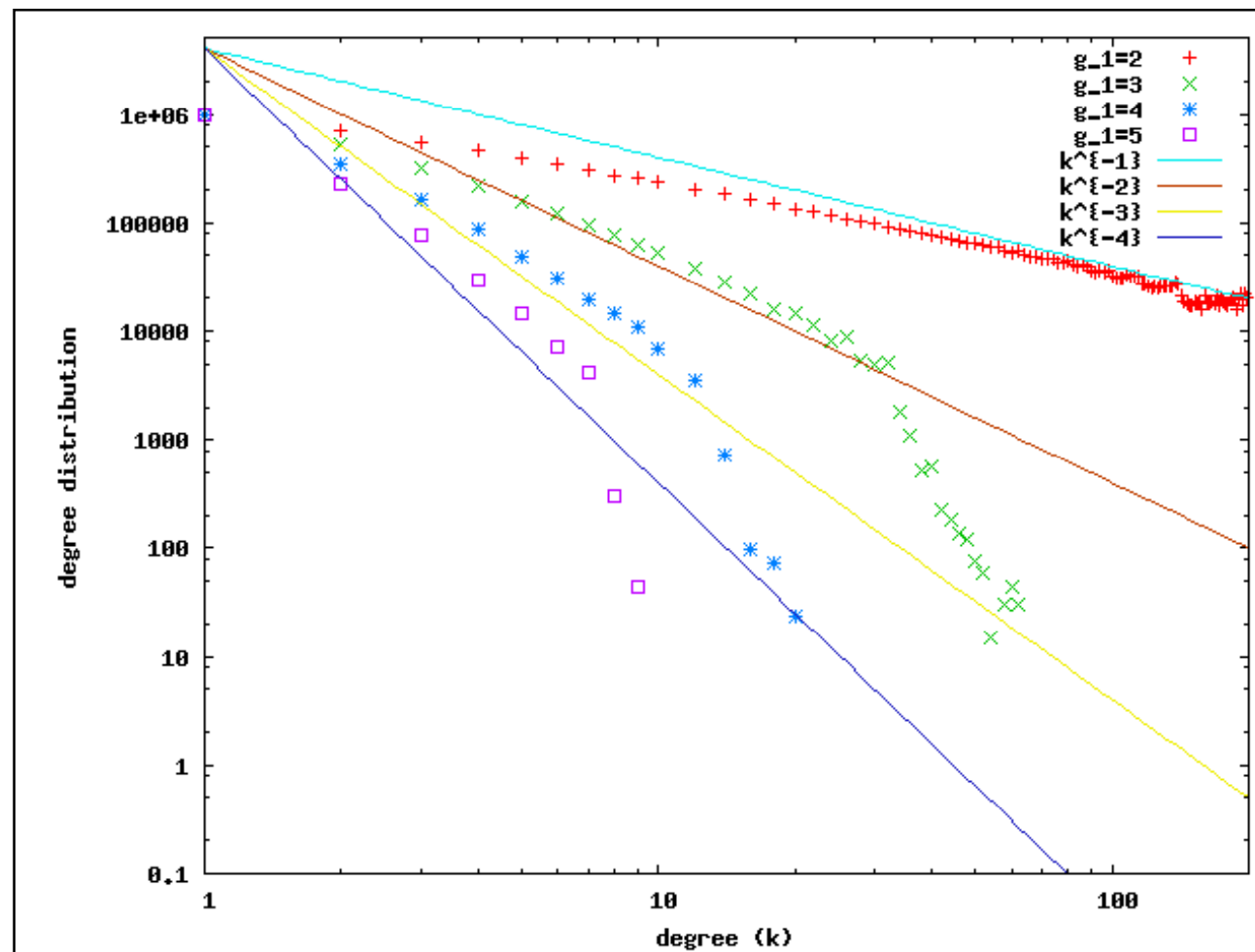
$$\begin{aligned} \sum_{h=1}^{\infty} e^{-\lambda h} \hat{f}(h, h+k) &\leq \sum_{h=1}^{\lceil k^{1/(1+\delta)} \rceil} e^{-\lambda h} h(h+k)^{-\gamma} + \sum_{h=\lceil k^{1/(1+\delta)} \rceil}^{\infty} e^{-\lambda h} \\ &\leq \sum_{h=1}^{\infty} e^{-\lambda h} h \cdot k^{-\gamma} + \sum_{h=\lceil k^{1/(1+\delta)} \rceil}^{\infty} e^{-\lambda h} \\ &\approx \left( \sum_{h=1}^{\infty} e^{-\lambda h} h \right) k^{-\gamma} + \int_{k^{1/(1+\delta)}}^{\infty} e^{-\lambda h} dh \\ &\approx c \cdot k^{-\gamma} + \frac{1}{\lambda} \cdot e^{-\lambda k^{1/(1+\delta)}} \leq 2c \cdot k^{-\gamma} \end{aligned}$$

(SS)の理論解析:

省略 (上記技法に加えて累積分布を使う)

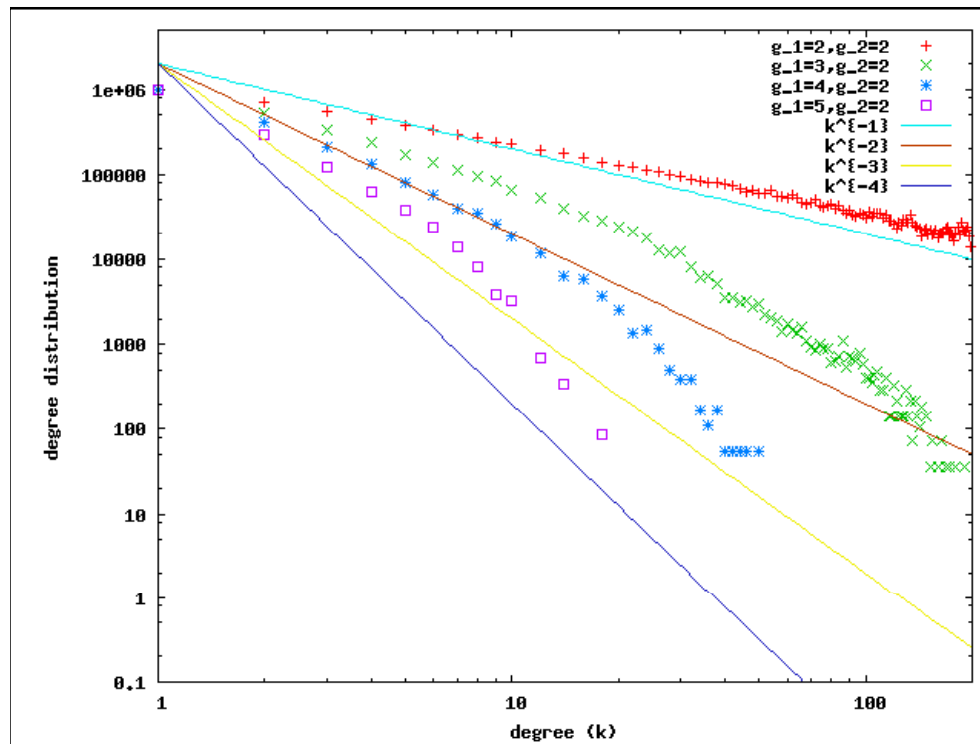
# 人工ネットワークにおける結果(1)

- SEネットワークをランダムに生成して射影を計算
- 理論式  $P(k) \propto k^{-\gamma+1}$  と、比較的良く一致

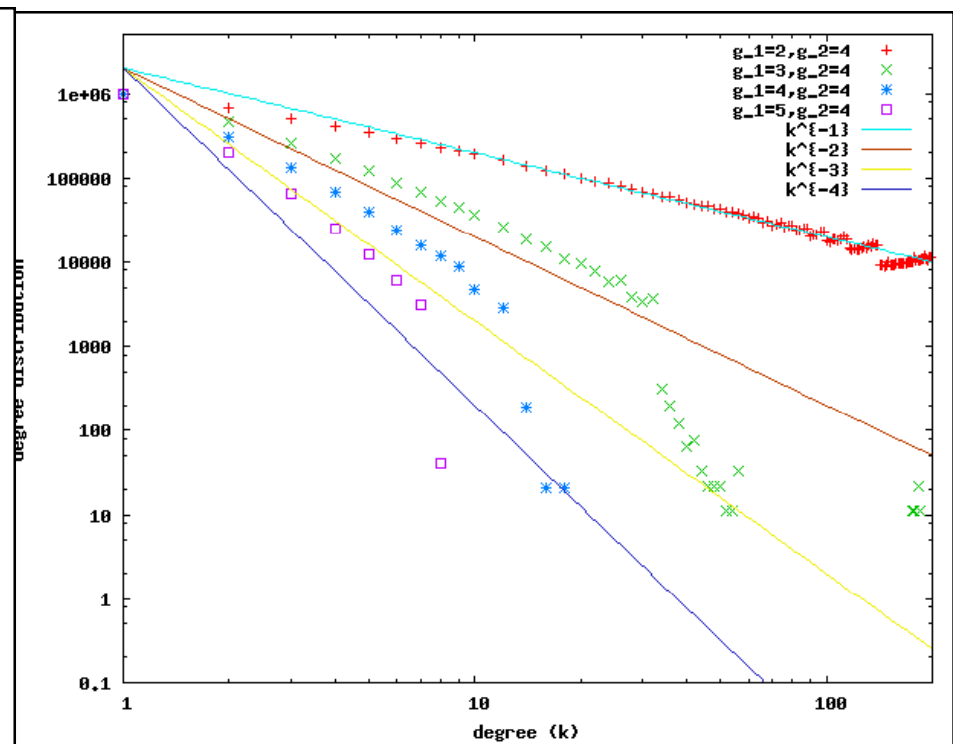


# 人工ネットワークにおける結果(2)

- SSネットワークをランダムに生成して射影を計算
- 理論式  $P(k) \propto k^{\max(-\gamma_1+1, -\gamma_2)}$  と、妥当な一致

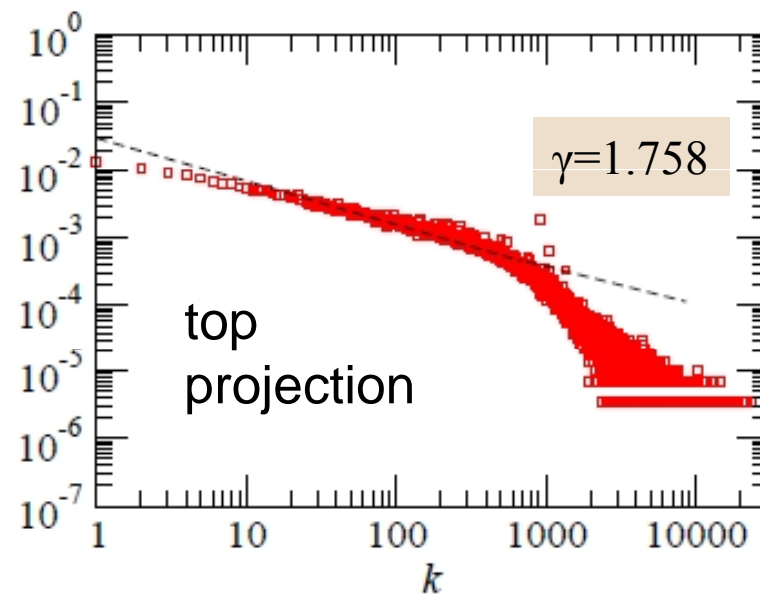
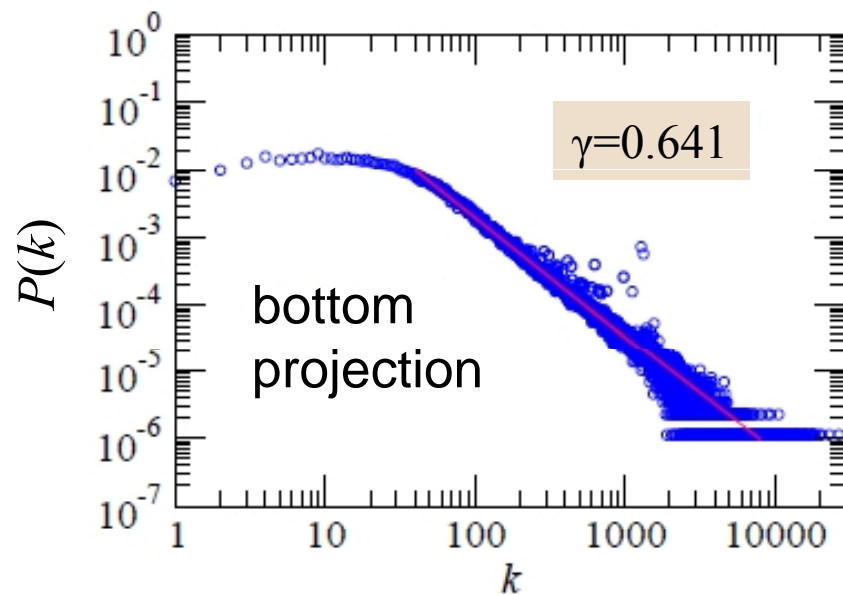
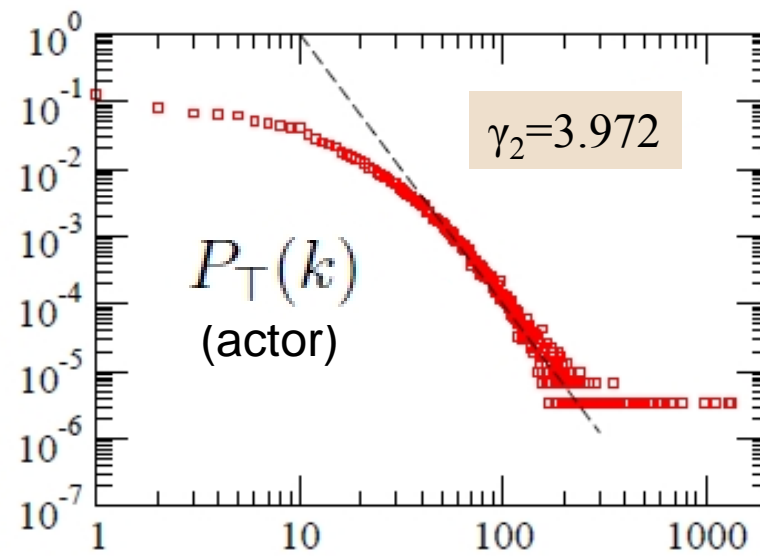
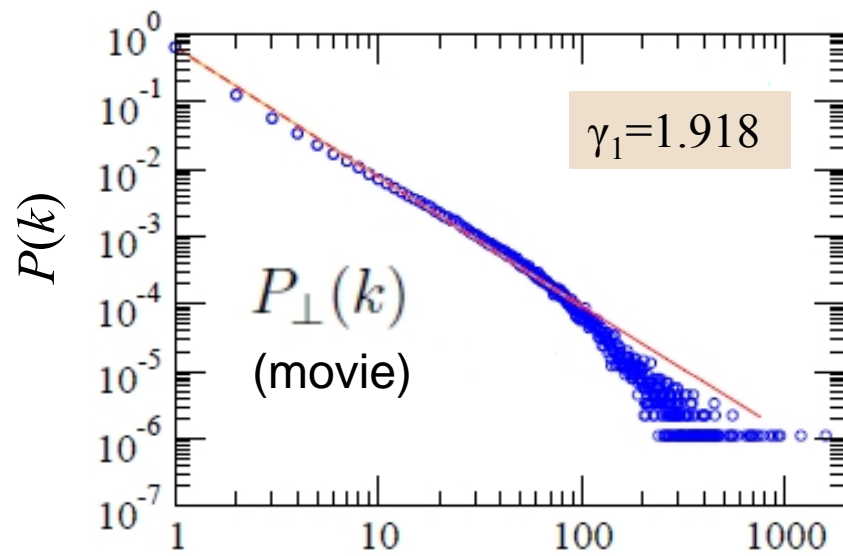


$$\gamma_2 = 2$$



$$\gamma_2 = 4$$

# 共演関係ネットワークに対する結果

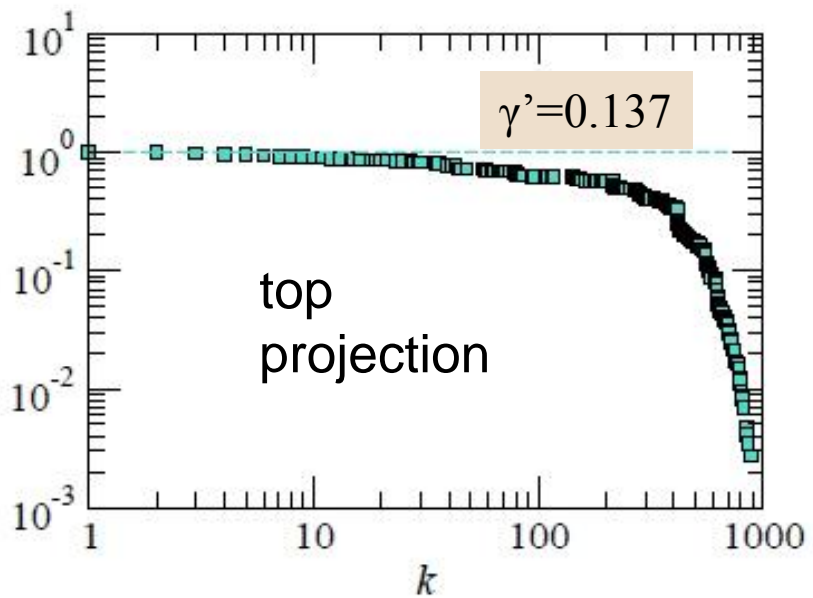
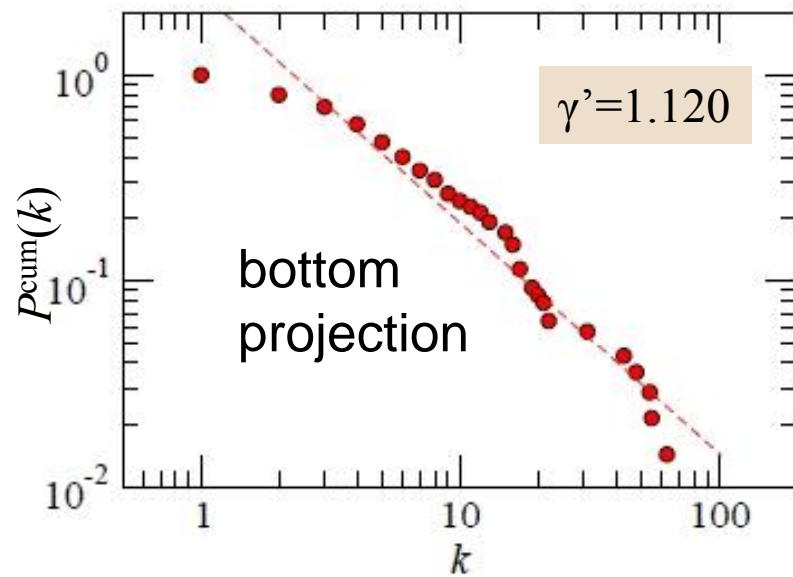
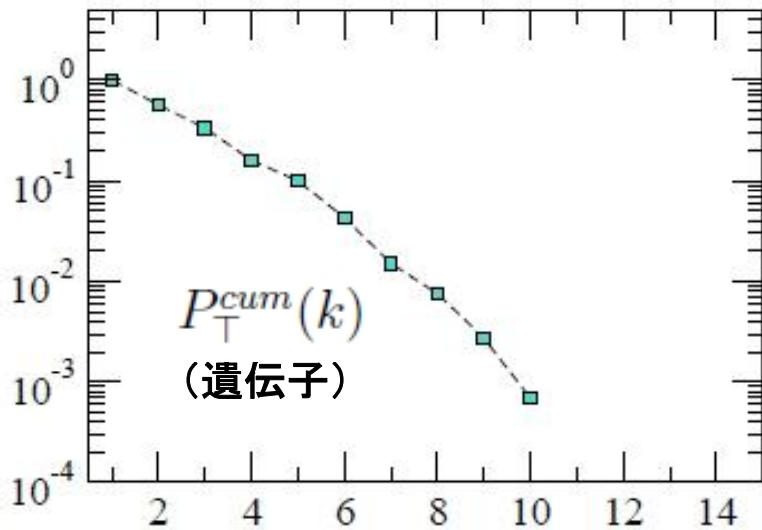
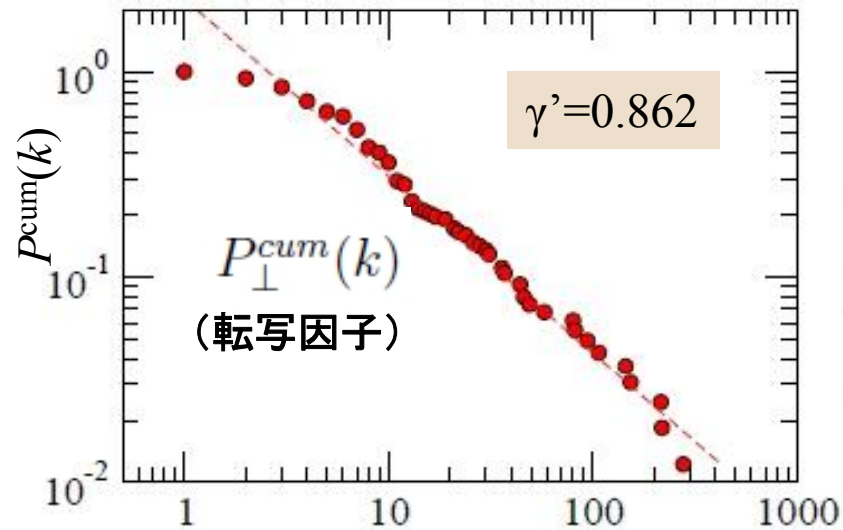


理論式

$$P(k) \propto k^{\max(-\gamma_1+1, -\gamma_2)}$$

$$P(k) \propto k^{\max(-\gamma_2+1, -\gamma_1)}$$

# 転写ネットワーク(大腸菌)に対する結果



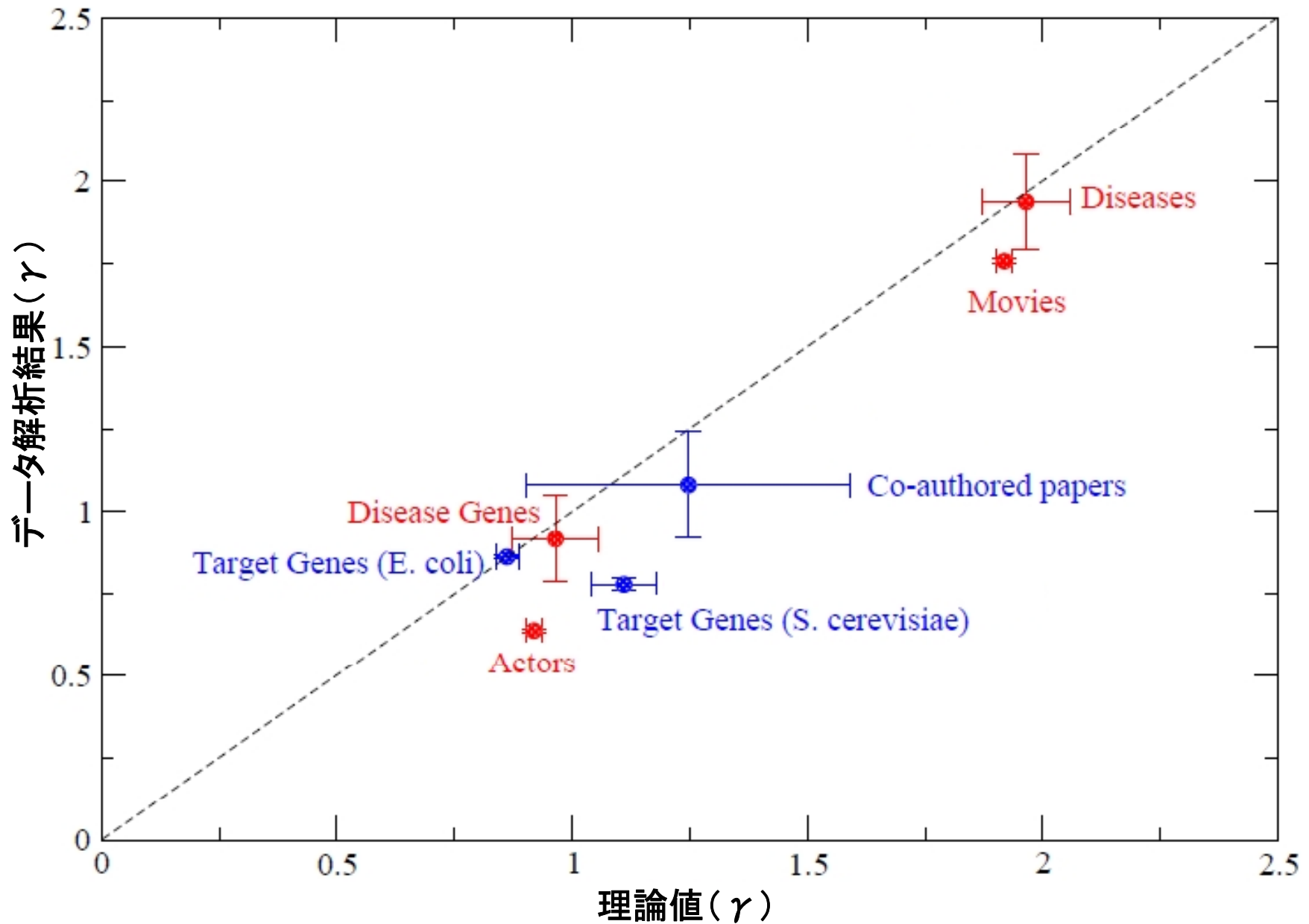
理論式

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

$$P(k) \propto k^{-\gamma+1}$$

ただし、 $\gamma' = \gamma - 1$

# 理論値とデータ解析結果の比較





# 結論

- 二部グラフからの射影 (bottom projection) により得られるグラフの次数分布について以下の関係を理論的に導出

$$\begin{aligned} \text{(SE)} \quad P_{\top}(k) \propto k^{-\gamma_1} \text{ かつ } P_{\perp}(k) \propto \exp(-\lambda k) \\ \implies P(k) \propto k^{-\gamma_1+1} \end{aligned}$$

$$\begin{aligned} \text{(SS)} \quad P_{\top}(k) \propto k^{-\gamma_1} \text{ かつ } P_{\perp}(k) \propto k^{-\gamma_2} \\ \implies P(k) \propto k^{\max(-\gamma_1+1, -\gamma_2)} \end{aligned}$$

- シミュレーション解析、および、データベース解析の結果と比較し、その妥当性を検証した

# 今後の課題

- より厳密な理論的解析
- 二部グラフの生成メカニズムの解明