

第2回「ネットワーク生態系と空間デザイン」シンポジウム

ネットワーク構造の可視化

平成16年3月5日

NTT コミュニケーション科学基礎研究所

研究のねらい

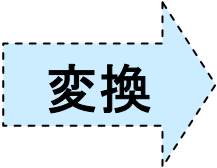
- 複雑なデータはネットワークで表現できる
 - ウェブページのハイパーリンクネットワーク
 - 生物学の遺伝子制御ネットワーク
 - 社会学のソーシャルネットワーク
 - 単語間の意味ネットワーク
- ネットワーク(関係データ)を可視化する意義
 - ネットワーク全体構造の直感的な把握支援
 - 直接「ブラウジング」による新知識の発見支援
 - 異常現象の早期発見システムに向けたコア技術

隣接行列と距離行列

提案法：隣接行列を直接用いた可視化法

ネットワークを表す
隣接行列
 $A = (a_{i,j})$
隣接 $\Rightarrow 1$
非隣接 $\Rightarrow 0$

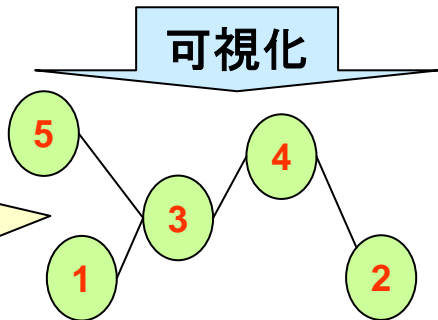
		ノード番号				
		1	2	3	4	5
ノード番号	1	1	0	1	0	0
	2	0	1	0	1	0
	3	1	0	1	1	1
	4	0	1	1	1	0
	5	0	0	1	0	1



		ノード番号				
		1	2	3	4	5
ノード番号	1	0	3	1	2	2
	2	3	0	2	1	3
	3	1	2	0	1	1
	4	2	1	1	0	2
	5	1	3	1	2	0

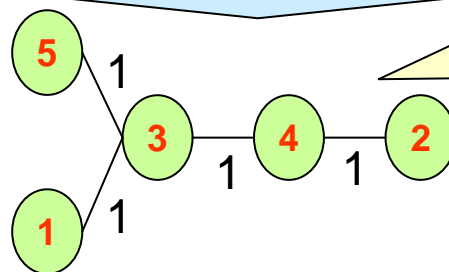
距離行列
ノード間の
グラフ距離
(最短経路長)を
成分とする行列

より柔軟な配置
が可能だが、
アルゴリズムの
実装が課題



隣接ノードを非隣接ノードより
近くに配置

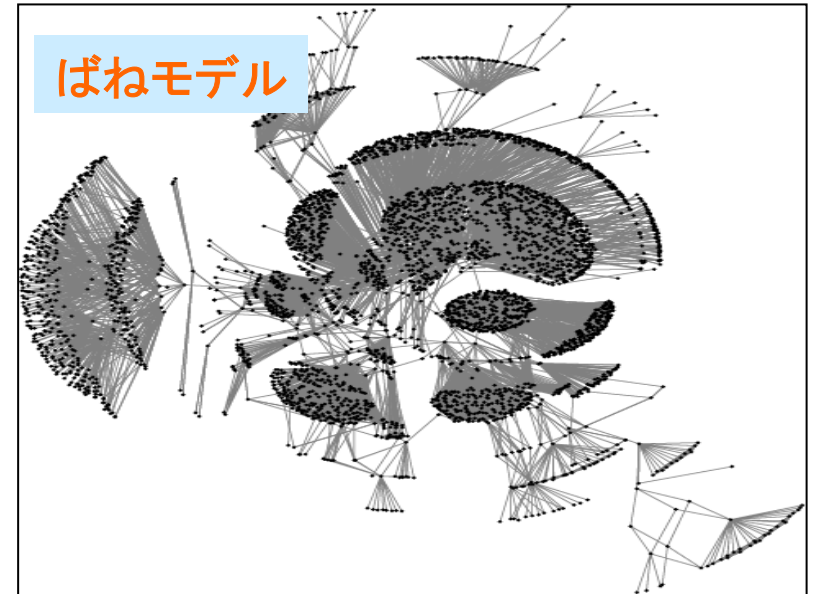
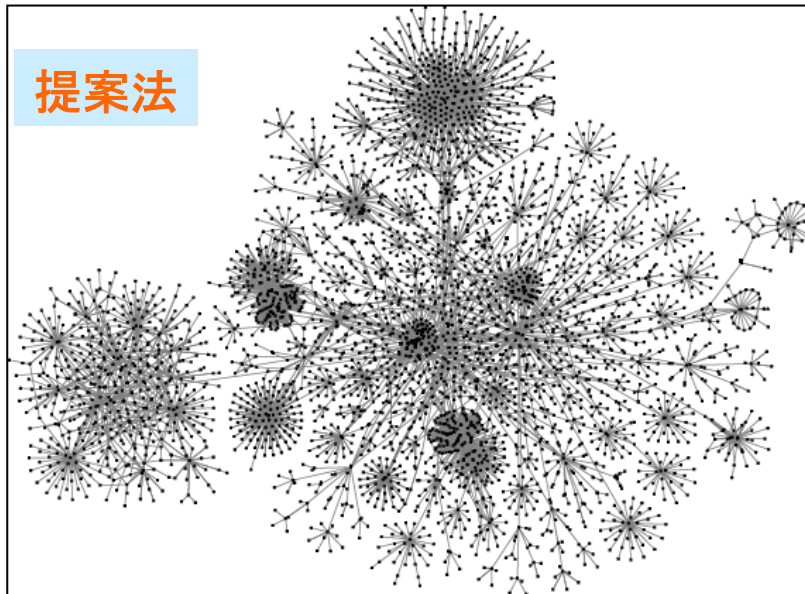
可視化



ノード数が
少ない場合
は有効

ノード間のグラフ距離(最短経路長)
≡ ユークリッド空間における距離

提案法の基本アイデアとメリット

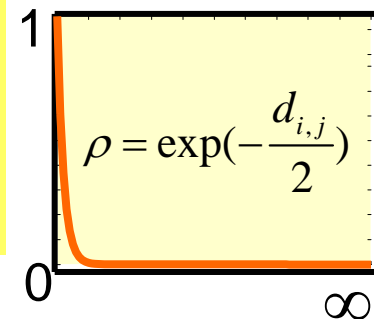


$A = (a_{i,j})$:隣接行列, $\rho(d_{i,j}) \in [0,1], \rho(0) = 1, \rho(\infty) = 0$:距離に基づく類似度

$\rho(d_{i,j})$ と $a_{i,j}$ とのクロスエントロピー $E_{i,j}$ の和を
ニューラルネットのオンライン学習で最小化

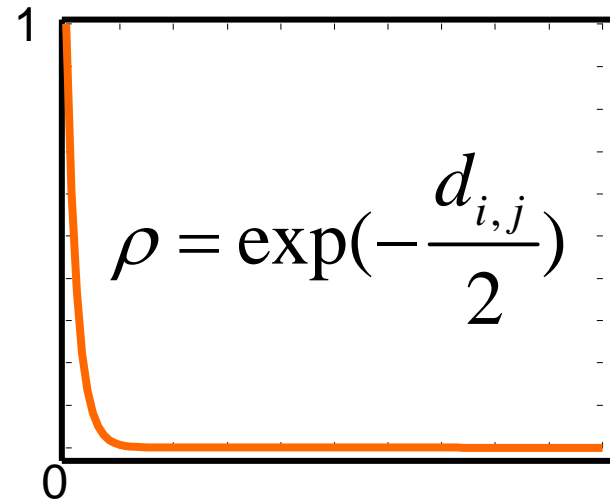
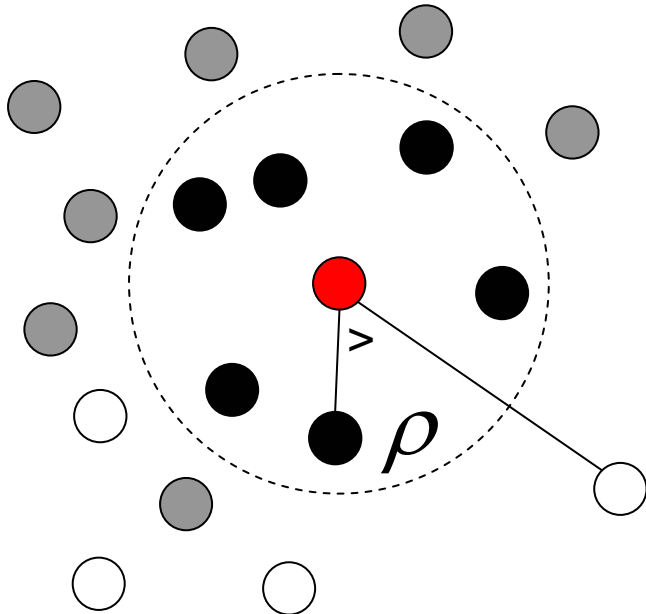
$$E_{i,j} = -a_{i,j} \ln(\rho(d_{i,j})) - (1 - a_{i,j}) \ln(1 - \rho(d_{i,j}))$$

$\Rightarrow \rho(d_{i,j})$ が $a_{i,j}$ にできるだけ一致する時最小となる



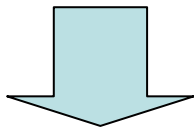
類似度関数

- N : ネットワークのノード数
- $A = (a_{i,j})$: 隣接行列 $a_{i,j} = 1$ or 0 , $a_{i,j} = a_{j,i}$ (無向を仮定)
- x_1, \dots, x_N : N 個のノードの座標, K : 埋め込み次元数
- $d_{i,j} = \|x_i - x_j\|^2$ (通常のユークリッド距離)



クロスエントロピーに基づく埋め込み法

隣接ノードを非隣接ノードより**近く**に配置

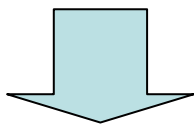


「隣接」 = 「 $a_{i,j} = 1$ 」

「近く」 = 「類似度 $\rho(d_{i,j}) \rightarrow 1$ 」

[類似度の近似] 各ノードにおいて連続類似度 $\rho(d_{i,j})$ が
離散類似度 $a_{i,j}$ の最も良い近似となるように配置

(ρ : i,j の類似度関数, x_1, \dots, x_N はパラメータで, これを調節)



「類似度 $\rho(d_{i,j})$ 」 \doteq 「 $a_{i,j}$ (離散類似度)」

$\rho(d_{i,j})$ と $a_{i,j}$ とのクロスエントロピーを最小化

$$E_{i,j} = -a_{i,j} \ln(\rho(d_{i,j})) - (1 - a_{i,j}) \ln(1 - \rho(d_{i,j}))$$

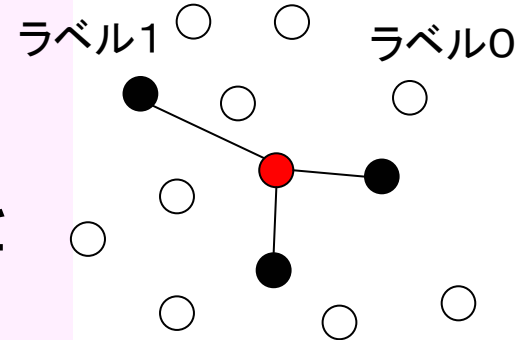
埋め込み法の目的関数

$$E_{i,j} = \frac{a_{i,j}}{2} \|x_i - x_j\|^2 - (1 - a_{i,j}) \ln(1 - \exp(-\frac{1}{2} \|x_i - x_j\|^2))$$

目的関数 $J = \sum_{i=1}^{N-1} \sum_{j=i+1}^N E_{i,j} + \frac{\mu}{2} \sum_{i=1}^n \|x_i\|^2$

正則化のためのWeight-Decay項

- i を固定したとき, $a_{i,j}$ はノード j の2値「ラベル」を定義
- N 個の2値分類問題を同時に解くことと解釈可能 (x_1, \dots, x_N がパラメータ)



学習アルゴリズム

(1) $t = 1$ とし, x_1, \dots, x_N をランダムに初期化

(2) 勾配ベクトル $J_{x_1}^{(1)}, \dots, J_{x_N}^{(1)}$ を計算

(3) $i = \arg \max_j \{|J_{x_j}^{(t)}|^2\}$ となる x_i を選択

(4) $|J_{x_i}^{(t)}|^2 < \varepsilon$ ならば x_1, \dots, x_N を出力し終了

(5) 変分ベクトル Δx_i を計算

$$\Delta x_i = \begin{cases} -H^{-1} J_{x_i}^{(t)} & \text{if } J^{(t+1)} < J^{(t)} \\ -\lambda J_{x_i}^{(t)} & \text{otherwise} \end{cases}$$

(6) $J_{x_1}^{(t)}, \dots, J_{x_N}^{(t)}$ を用いて $J_{x_1}^{(t+1)}, \dots, J_{x_N}^{(t+1)}$ を更新

(7) $x_i = x_i + \Delta x_i$ によって x_i を更新

(8) $t = t + 1$ とし, ステップ(3)へ

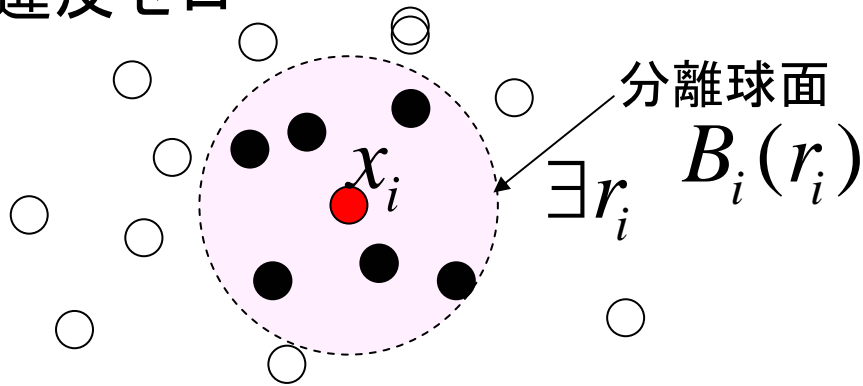
$$H = \frac{\partial^2 J^{(t)}}{\partial x_i \partial x_i^t}$$

$$J_{x_i}^{(t+1)} \approx J_{x_i}^{(t)} + \frac{\partial^2 J^{(t)}}{\partial x_i \partial x_i^t} \Delta x_i = 0$$

F値に基づく評価のアイデア

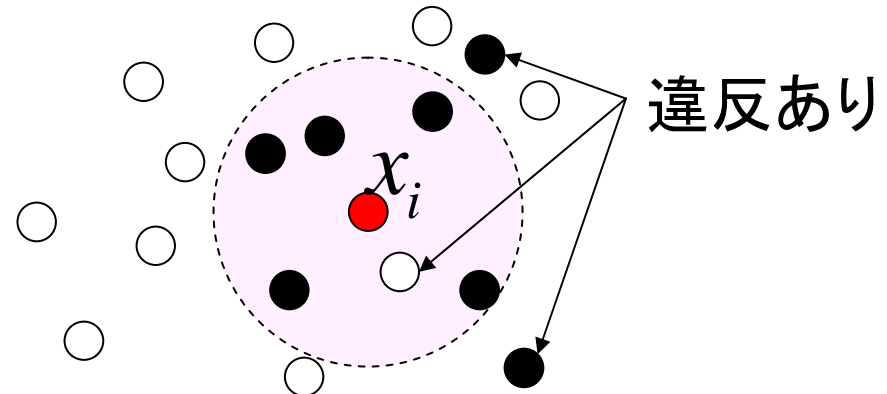
[理想的埋め込み]

違反ゼロ



各ノード i において違反ゼロ
の完全分離球面が存在

[現実の埋め込み]



違反ゼロにはできない

- 各ノード i にて F 値の意味で違反度最小となる $B_i(r_i)$ を構成
- 全ノードの F 値の平均で埋め込みの良さを評価

F値に基づく評価

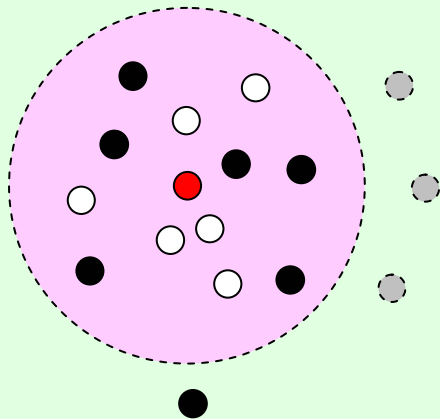
全ノードのF値の平均で埋め込みの良さを評価

F値: Precision と Recall の調和平均

$$F_i(r_i) = 1 / \left\{ \alpha \frac{1}{P_i(r_i)} + (1 - \alpha) \frac{1}{R_i(r_i)} \right\} \Rightarrow F = \sum_{i=1}^N \frac{F_i(\hat{r}_i)}{N} \quad \text{通常 } \alpha = 1/2$$

Precision

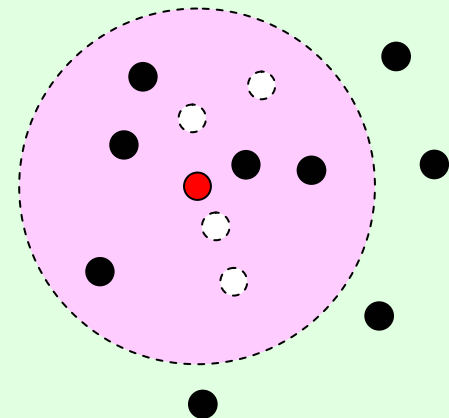
$$P_i(r_i) = \frac{6}{13}$$



半径が**小さい**ほうが有利

Recall

$$R_i(r_i) = \frac{5}{9}$$



半径が**大きい**ほうが有利

比較に用いた従来法

CMDS: Torgersonによる古典的MDS

$$J_{\text{CMDS}} = \text{trace}\left\{\left(-\frac{1}{2}\mathbf{Y}\mathbf{O}\mathbf{Y} - \mathbf{X}\mathbf{X}'\right)^2\right\}$$

$\mathbf{G}=(g_{i,j})$: グラフ距離行列

$\mathbf{O}=(o_{i,j})$: $o_{i,j} = g_{i,j} \mathbf{1} g_{i,j}$

\mathbf{Y} : N -次元 Young-Householder 変換行列

SC: スペクトラルクラスタリング

$$J_{\text{SC}} = \text{trace}\left\{\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{B}\mathbf{D}^{-\frac{1}{2}} - \mathbf{X}\mathbf{X}'\right)^2\right\}$$

$\mathbf{B}=(b_{i,j})$: $\exp(-g_{i,j}/2)$

$\mathbf{D}=(d_{i,j})$: 対角行列, $d_{i,i} = \mathbf{B}$ の第*i*行の全要素の和

$\hat{x}_{i,j} = \frac{x_{i,j}}{\sqrt{\sum_{j=1}^N x_{i,j}}}$ によって結果を球面上に表示

KK: Kamada & Kawaiによるばねモデル

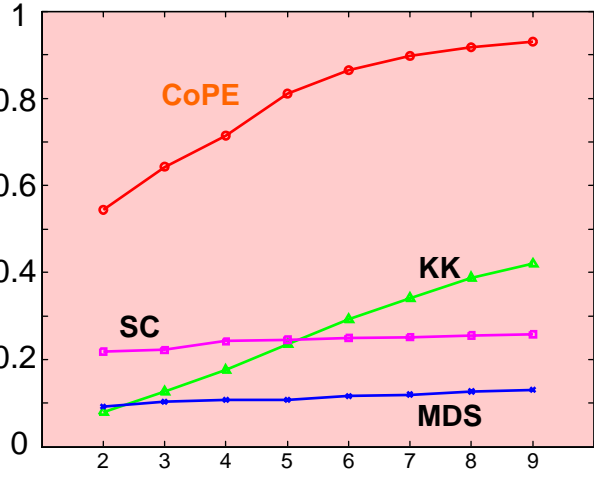
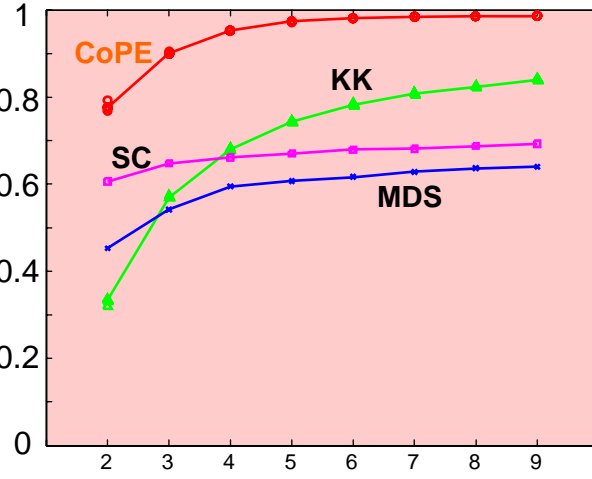
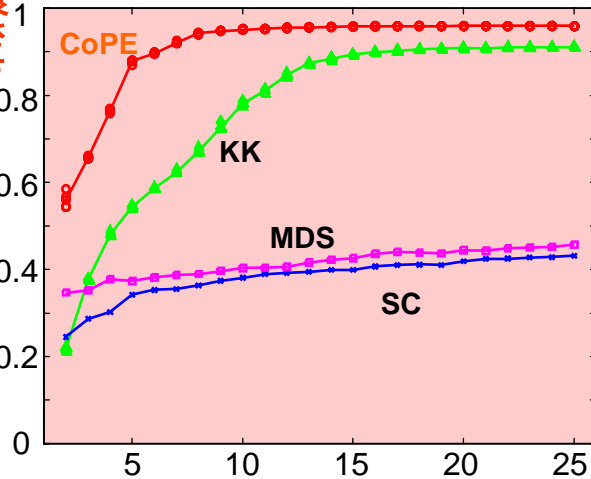
$$J_{\text{KK}} = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(g_{i,j} - \|x_i - x_j\|)^2}{g_{i,j}^2}$$

実データを用いた評価実験

CoPE: 提案法, MDS: 多次元尺度法, KK: ばねモデル, SC: スペクトラルクラスタリング

★ 縦軸は、隣接ノードが非隣接ノードより近くに配置されている割合、横軸は埋め込みの次元

配置の良し



次元

大腸菌の遺伝子制御ネットワーク

ノード数 328, リンク数 456

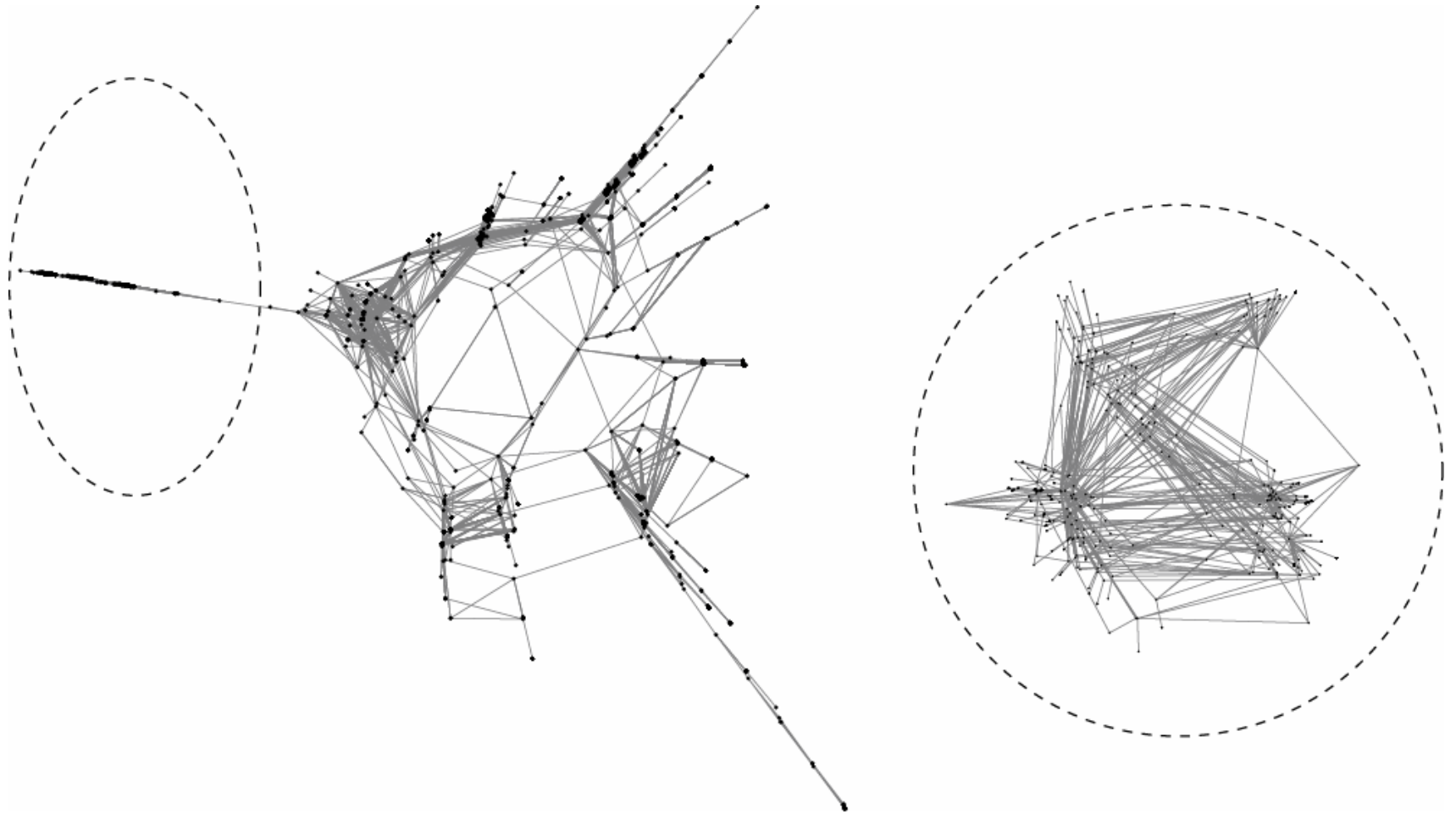
NIPS掲載論文の共著者ネットワーク

ノード数1061, リンク数 2080

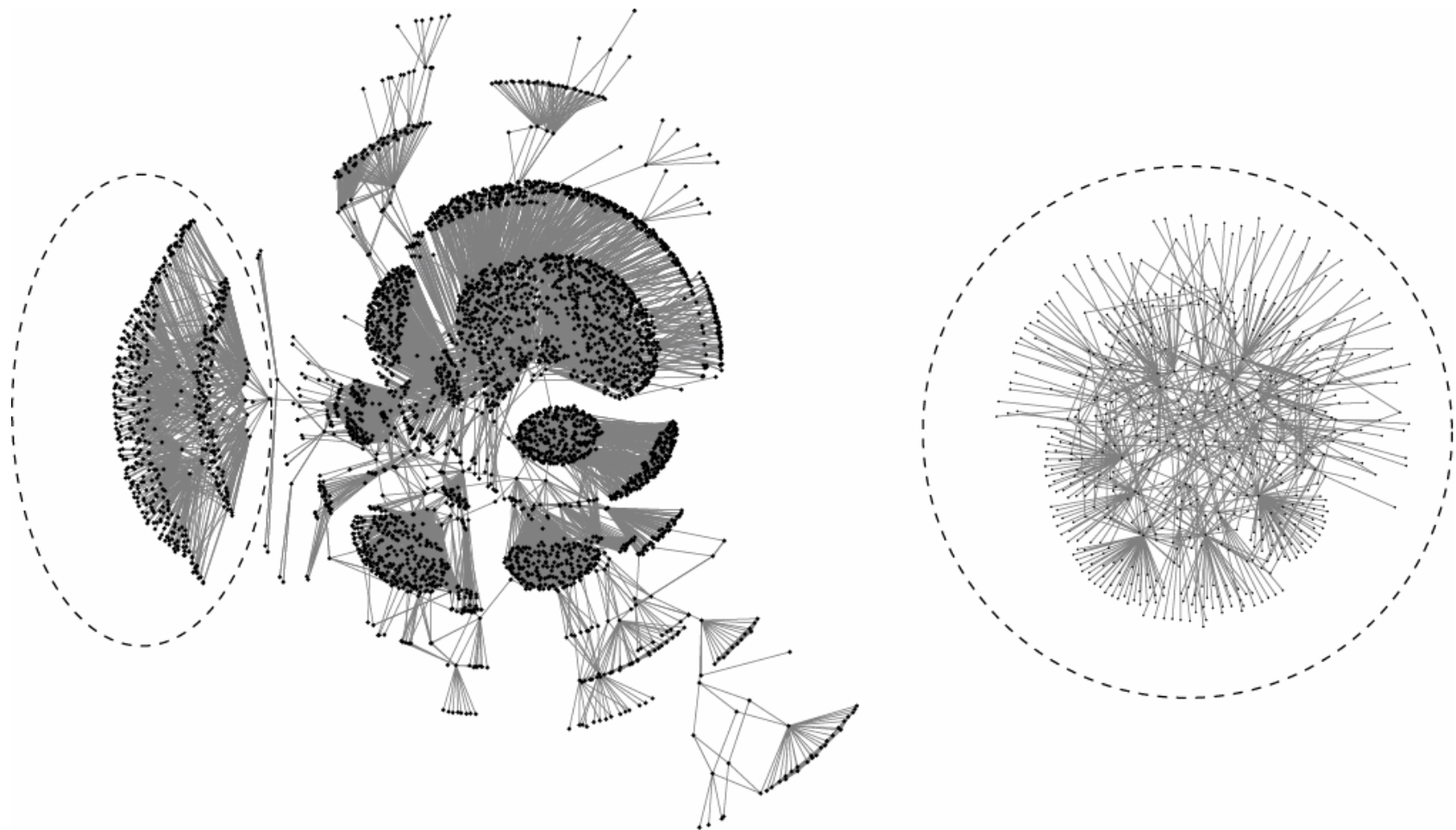
NTTドメインに属するWWWページ

ノード数 2870, リンク数 9337

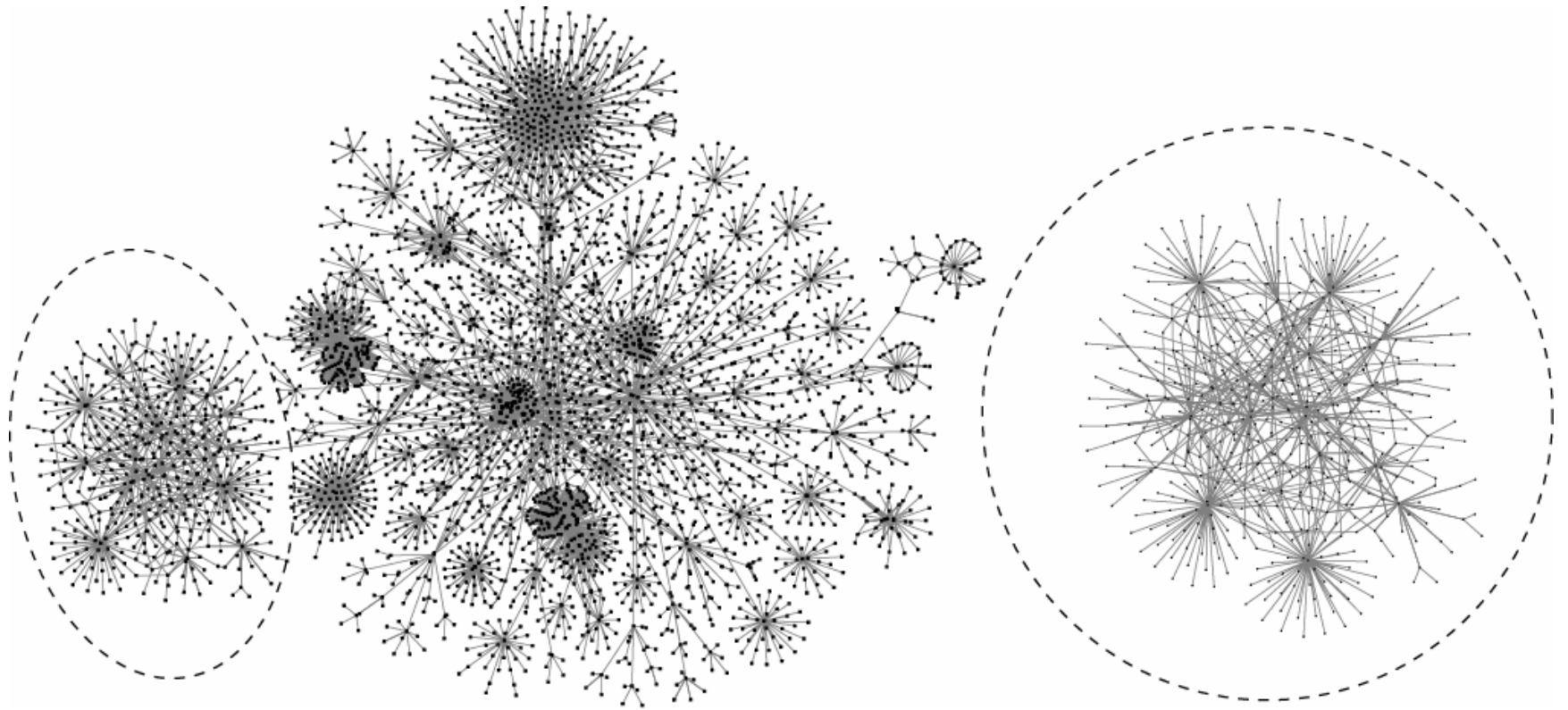
2次元への埋め込み結果(古典MDS)



2次元への埋め込み結果(バネモデル)



2次元の埋め込み結果(提案法)



3次元の埋め込み結果

[E.Coli](#): 大腸菌の遺伝子制御ネットワーク

[NIPS](#): NIPS掲載論文の共著者ネットワーク

[NTT](#): NTTドメインに属するWWWページ

[KECL](#): CS研ドメインに属するWWWページ