

スパースかつ低ランク制約を用いた 時変ネットワーク構造推定

平山 淳一郎 国際電気通信基礎技術研究所(ATR)

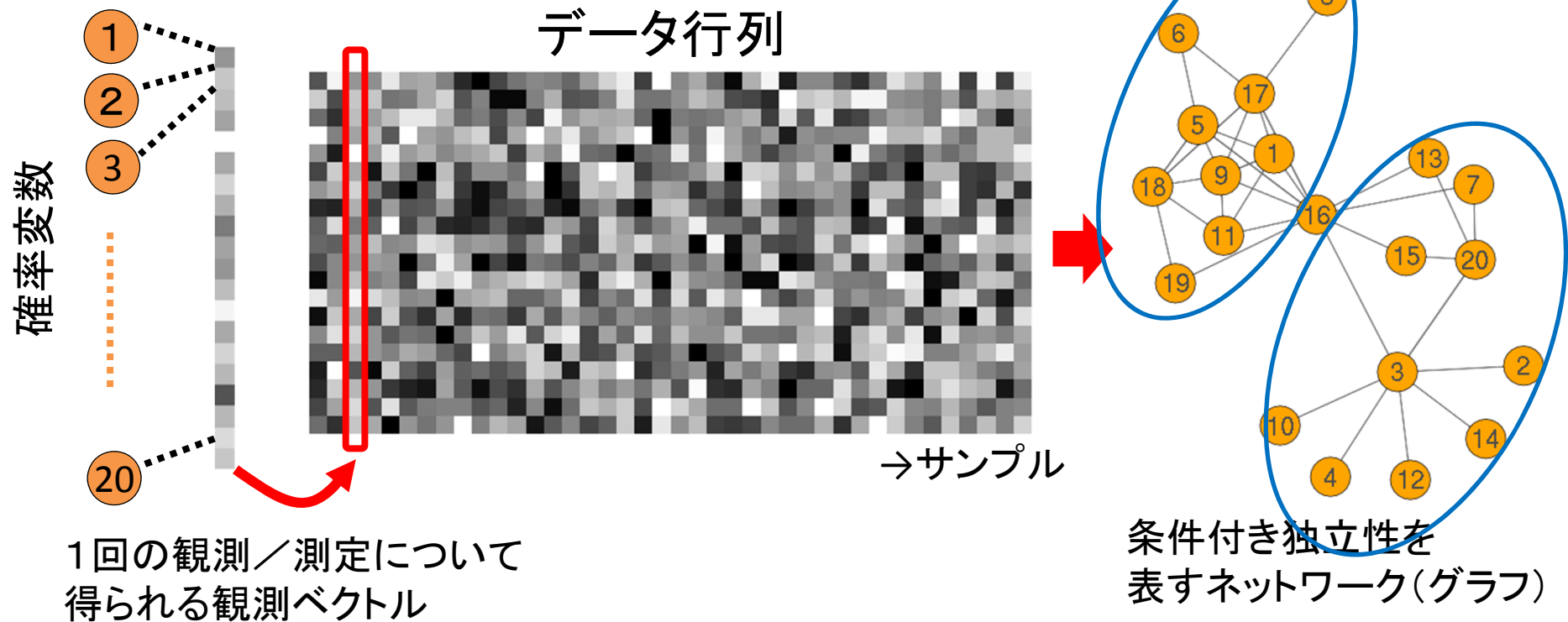
Aapo Hyvärinen University of Helsinki

石井 信 京都大学／ATR

2012年9月24日

第3回Latent Dynamics ワークショップ

ネットワーク構造推定

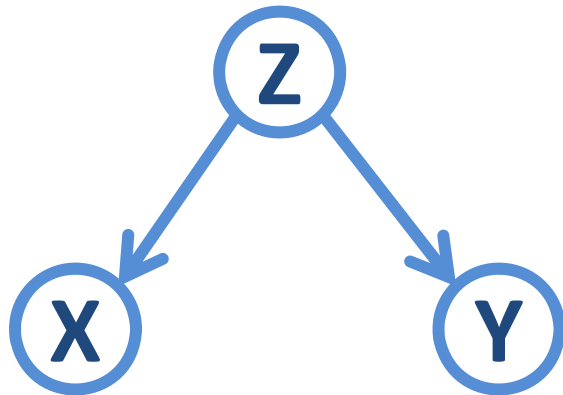


- ▶ 与えられたデータ行列から、変数(ノード)間の**潜在的な依存関係**を表すネットワーク(グラフ)を決定する
- ▶ 単なる相関解析では見えない直接的な関係がわかる

確率モデルに基づく方法

今回はこちら

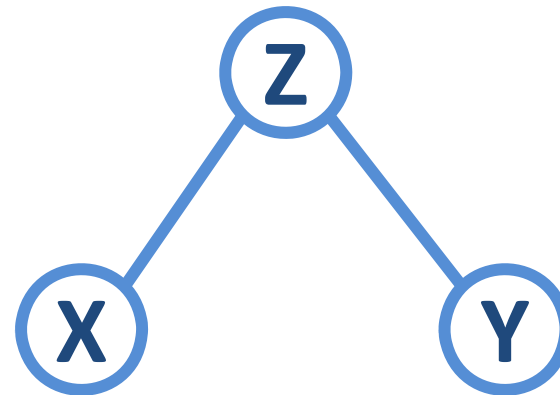
▶ ベイジアンネットワーク



有向非巡回グラフ

$$P(x,y,z) = P(x|z)P(y|z)P(z)$$

▶ マルコフネットワーク (マルコフ確率場)

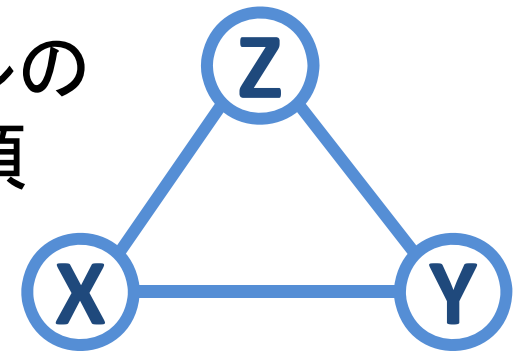


無向グラフ

$$P(x,y,z) \propto f(x,z)f(y,z)$$

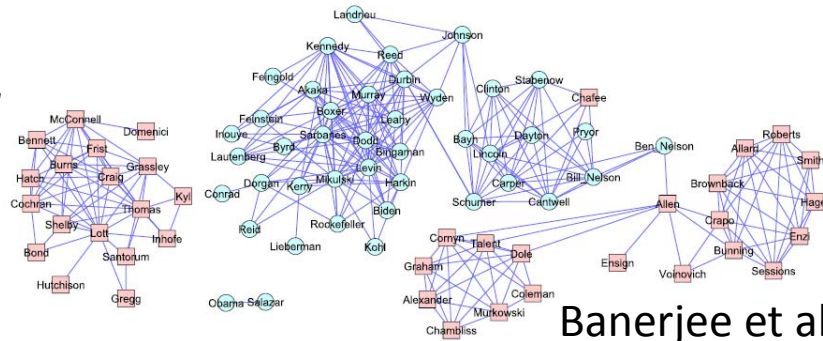
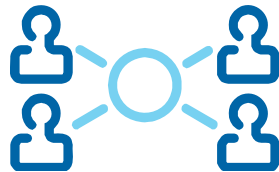
スパース正則化

- ▶ グラフ構造の(組合わせ)最適化は難しい問題
- ▶ 最近では基本的なモデルについて「スパース正則化」と凸最適化による効率的な推定が利用可能に
 - e.g. Meinshausen & Bühlmann '06; Friedman et al. '08; Wainwright et al. '07; Banerjee et al. '08, Höfling et al.'09
- ▶ **L1ノルム正則化**: 「結合重み」ベクトルのL1ノルム(絶対値の和)をペナルティ項として加える→不要な重みがゼロに



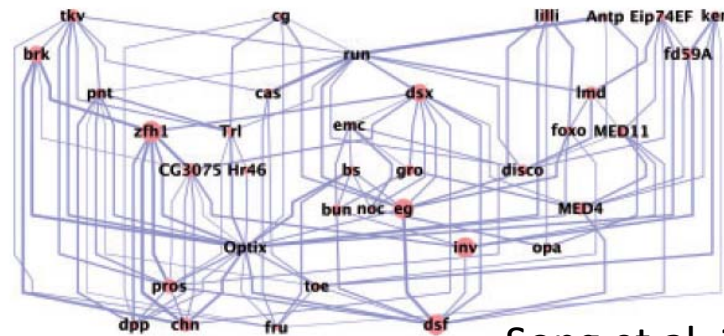
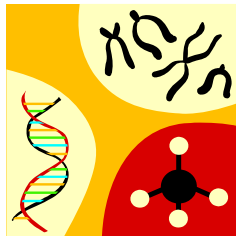
ネットワーク構造推定の応用例(1/2)

社会的つながりの可視化



Banerjee et al., 2008

遺伝子ネットワークの同定

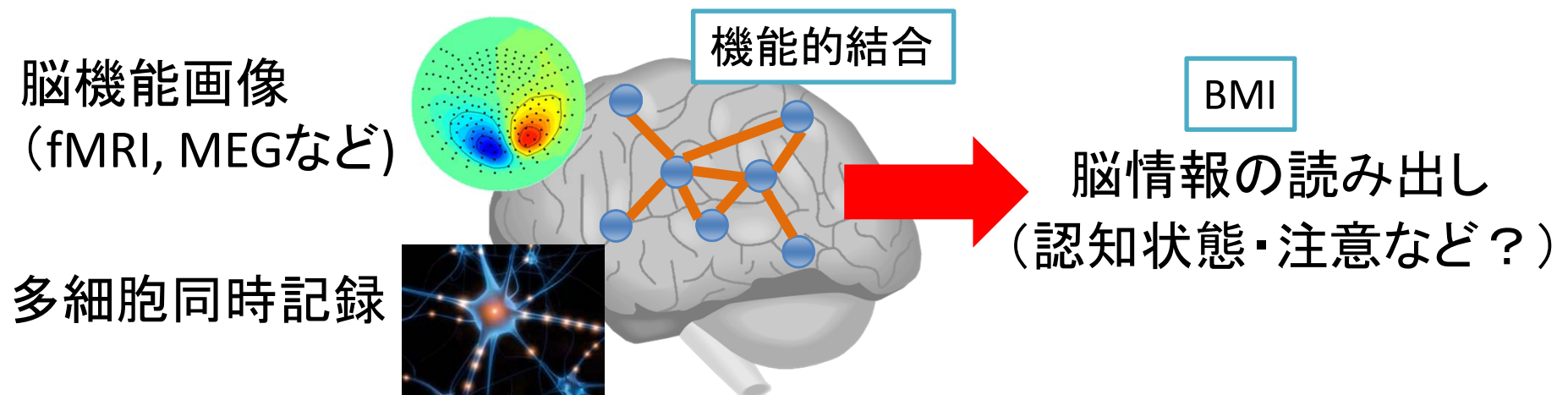


Song et al. 2009

- ▶ ネットワーク構造推定は、多数の要素からなる複雑なシステムを解明する有効なアプローチ

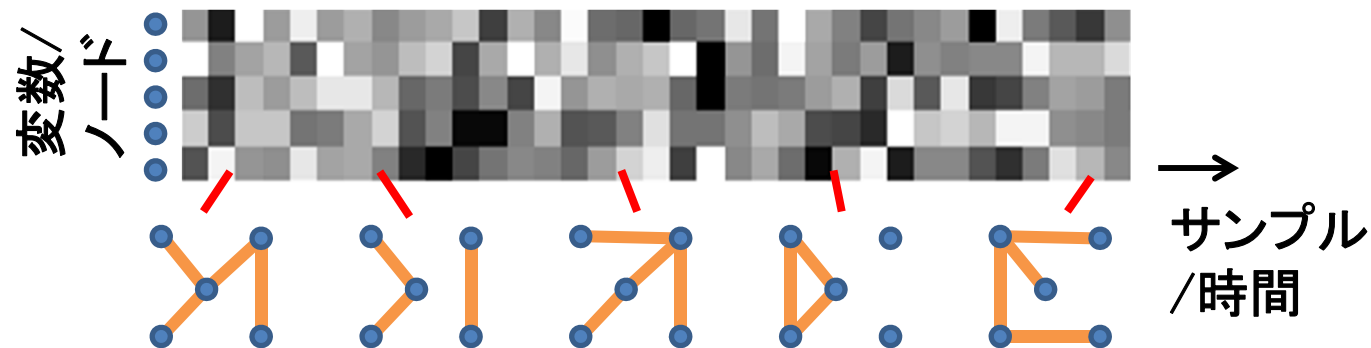
ネットワーク構造推定の応用例(2/2)

- ▶ とくに脳信号解析では、ネットワーク構造推定は領域レベルや神経細胞レベルでの機能的なつながり（機能的結合）を調べる強力な手段を提供する
- ▶ 工学的にも、推定された結合性に基づくブレイン・マシン・インターフェイス(BMI)等の応用が期待される



ネットワークの時間変化をとらえる

- ▶ 時間変化するネットワークの推定は最近のトピック



- 変化点・異常値検出 (e.g. Xuan et al.'07, Ide et al.'09, 早矢仕&山西'10)
- 社会／遺伝子ネットワークの変化 (e.g. Ahmed et al.'09)
- 同期神経発火の解析 (e.g. Shimazaki et al.'12)
- 機能的結合 (MEG) の変化と認知 (e.g. Iwayama et al.'12)

今回の発表内容

- ▶ 具体的なモデルを2値マルコフ確率場（イジングモデル）に限定して、時変ネットワーク推定に関する研究を紹介する
 - 連続データの場合（ガウシアングラフィカルモデル）に比べて研究が少ない
 - 遺伝子発現, 神経スパイク, 社会的つながり, etc.
- 1. 時間的な連続性を利用する方法（既存手法）
- 2. 加えて低ランク性を利用する方法（提案手法）

二値マルコフ確率場 (MRF)

- 観測ベクトル (二値) $t=1,2,\dots,T$

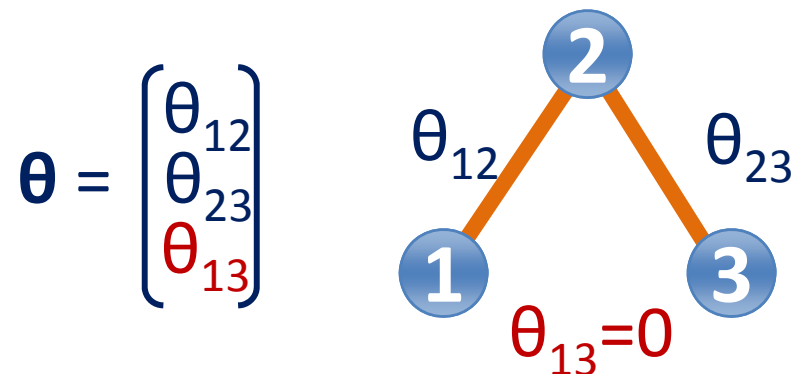
$$\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_D(t)) \in \{-1, 1\}^D$$

- 確率分布

$$p(\mathbf{x}(t) \mid \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp \left(\sum_{i < j} \theta_{ij} x_i(t) x_j(t) \right)$$

確率分布の正規化のための定数

- 無向グラフ



離散MRFの(構造)推定の難しさ

- ▶ 正規化定数が多重和を含むため、尤度関数やその勾配の計算に大きなコストがかかる

$$Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \{-1,1\}^D} \exp \left(\sum_{i < j} \theta_{ij} x_i x_j \right)$$

$$\log p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \sum_{i < j} \theta_{ij} x_i x_j - \log Z(\boldsymbol{\theta})$$

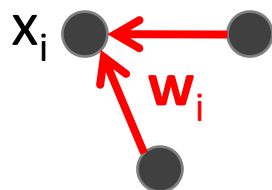
$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{x} \mid \boldsymbol{\theta}) = (x_i x_j - \mathbb{E}[x_i x_j \mid \boldsymbol{\theta}])$$

- ▶ 近似計算法(MCMC, 変分下限, etc.)は多いが, 大規模問題では尤度の利用はあまり効率的ではない

スパース回帰に基づく方法

(Wainwright et al., '06; Ravikumar et al., '10)

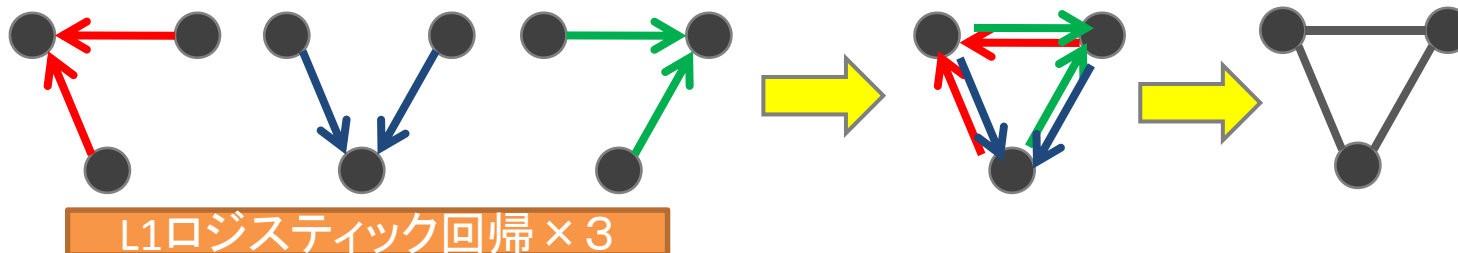
- ▶ 一方で, 条件付き尤度 $p(x_i | \mathbf{x}_{\setminus i})$ は容易に計算可能
- ▶ これはロジスティック回帰と等価 ($\mathbf{x}_{\setminus i}$: x_i 以外すべての変数)



$$-\log p(x_i | \mathbf{x}_{\setminus i}) = \xi(x_i, \mathbf{w}_i^\top \mathbf{x}_{\setminus i})$$

$$\text{ロジスティック損失 } \xi(x, y) = \log(1 + e^{-2xy})$$

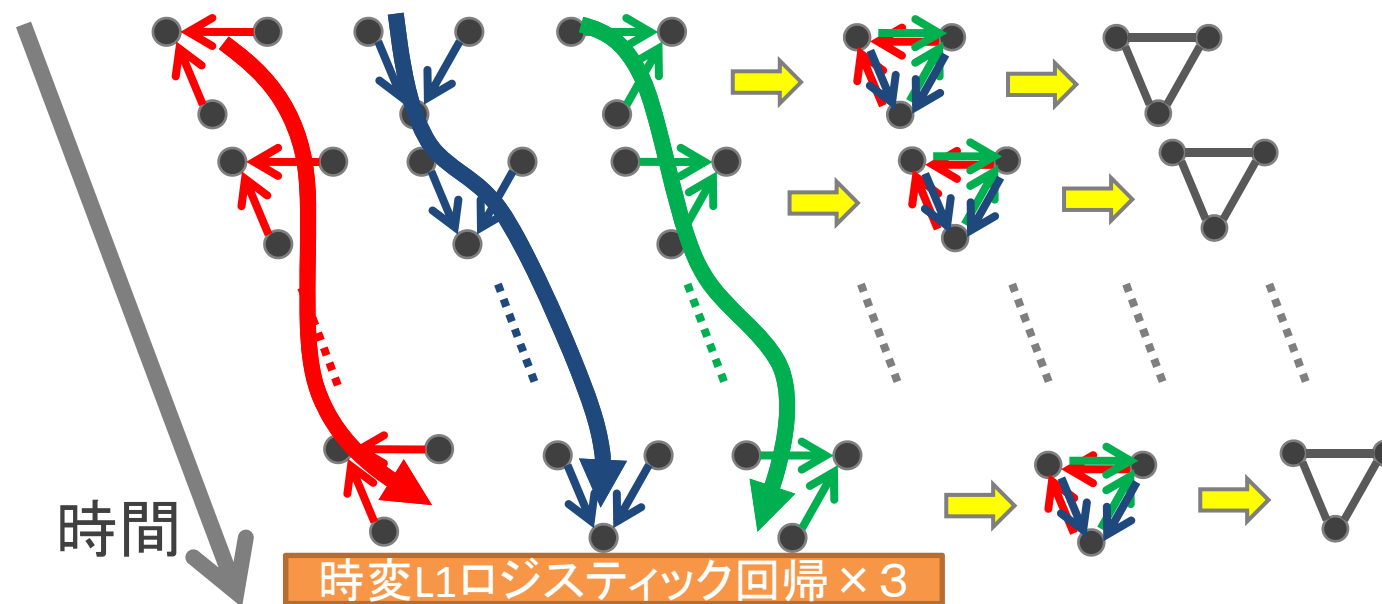
- ▶ スパースロジスティック回帰によるグラフ構造推定
 1. 各変数に関してL1正則化ロジスティック回帰
 2. 2重の推定値を後処理(絶対値のmin またはmax)



時変ネットワークへの拡張

(Song et al. '09, Ahmed et al. '09, Kolar et al.'10)

- ▶ 時変係数ロジスティック回帰による時変構造推定

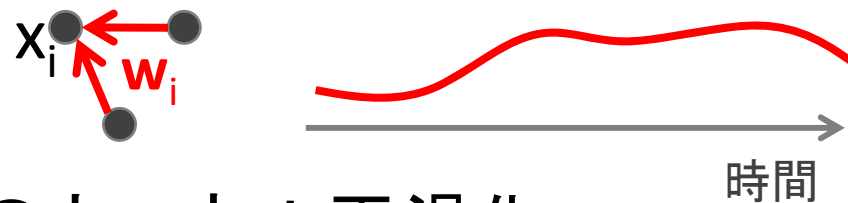


- ▶ 重みの変化に仮定をおき, 凸最適化にもちこむ
 - 仮定の入れ方が異なる2つのアプローチ

カーネル平滑化

(Song et al. '09, Kolar et al.'10)

- ▶ 仮定: 重みは時間的に滑らかに変化する



- ▶ 損失関数のカーネル平滑化

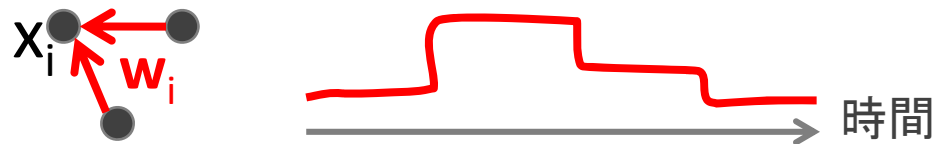
$$\underset{w_i(t)}{\text{minimize}} \quad \sum_{t'=1}^T K_w(t' - t) \xi(x_i(t'), w_i(t)^\top x_{\setminus i}(t')) + \lambda \|w_i(t)\|_1$$



正則化による方法

(Ahmed et al. '09, Kolar et al.'10)

- ▶ 仮定: 重みはだいたい一定 & ときどき大きく変化



- ▶ 時変係数の“Total Variation” (fused Lasso) 正則化

$$\underset{w_i(1), w_i(2), \dots, w_i(T)}{\text{minimize}} \quad \frac{1}{T} \sum_{t=1}^T \xi(x_i(t), w_i(t)^\top x_{\setminus i}(t)) + \lambda \sum_{t=1}^T \|w_i(t)\|_1$$

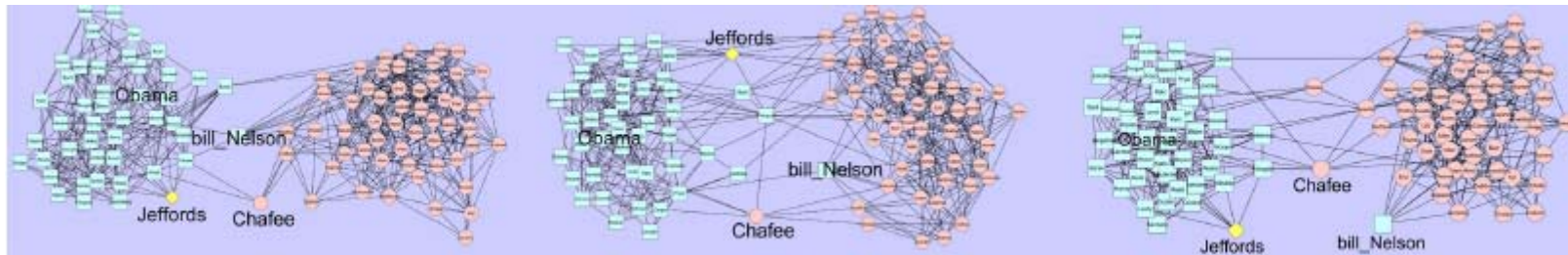
各時刻tについてL1ロジスティック回帰

$$+ \gamma \sum_{t=2}^T \|w_i(t) - w_i(t-1)\|_1$$

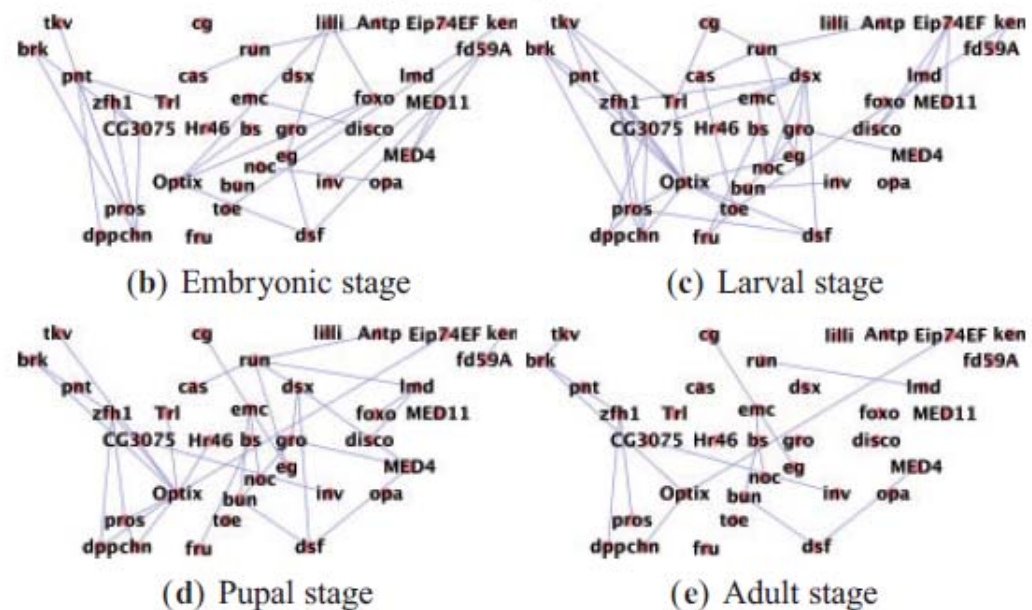
TV正則化(重みの変化量をスパースに)

結果の例

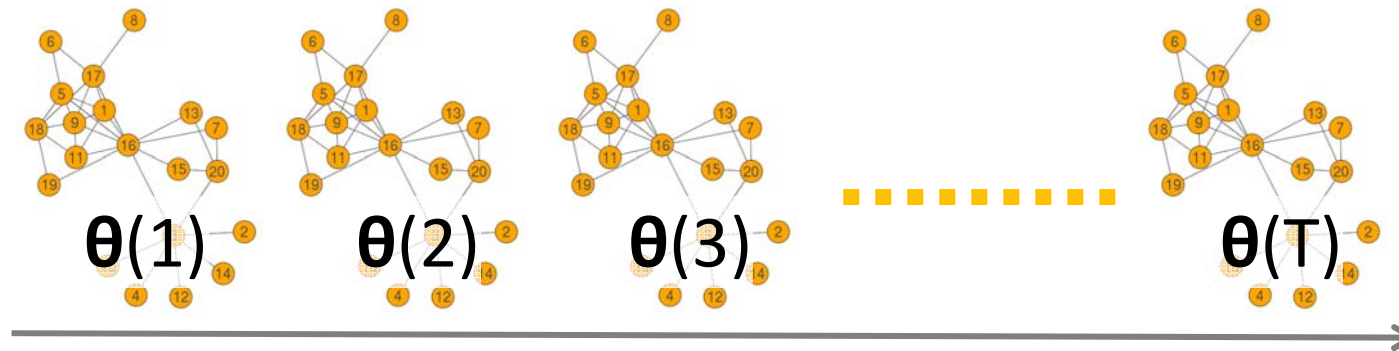
- 米上院議員の時変ネットワーク構造 (Ahmed et al.'09)



- 発達過程によって異なる遺伝子ネットワーク (Song et al.'09)

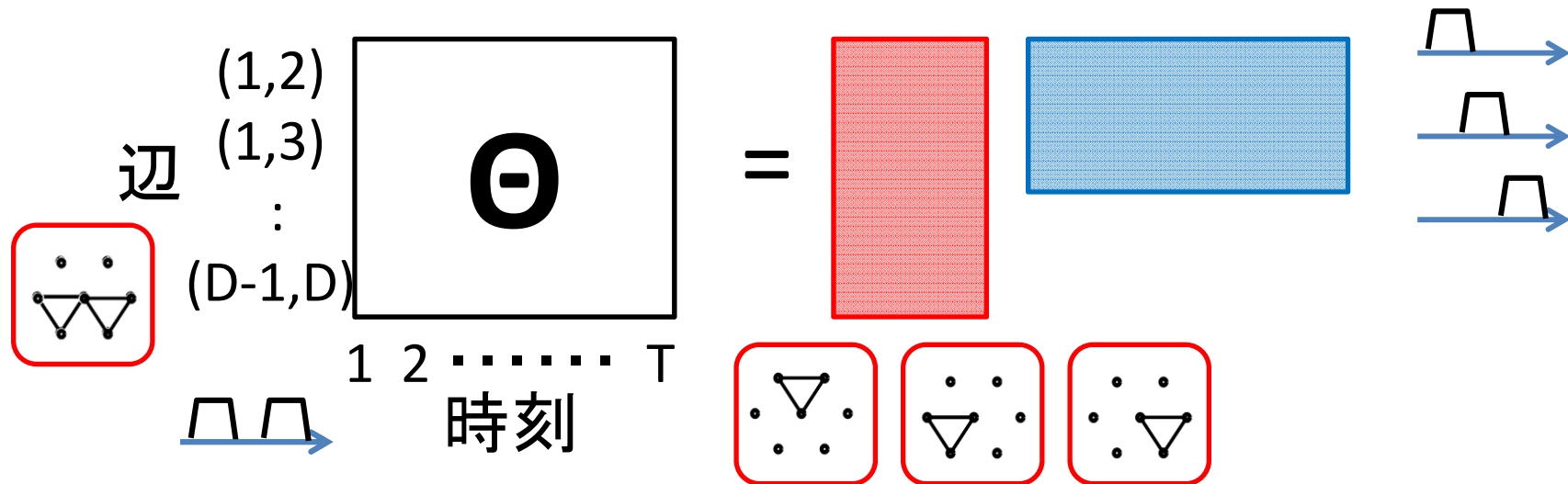


低ランク性の導入 (Hirayama et al.'10)



- ▶ 時変ネットワークでは、推定すべきパラメータ(結合重み)の数が多くなりがち
- ▶ スパース性や変化の滑らかさなど、適切な事前知識(仮定)を用いることで推定の自由度を下げうる
- ▶ ここでは、新たにパラメータ行列 $\Theta = (\theta(1), \theta(2), \dots, \theta(T))$ の低ランク性を仮定する

低ランク性が意味すること



- ▶ すべての時刻のグラフ(重み)が少数の異なるパターンの和で表される
- ▶ すべての辺の重み時系列が少数の異なるパターンの和で表される

凸最適化問題として定式化

カーネル平滑化に基づく損失関数
(時間的な滑らかさ)

各列のL1ノルムの和
(重みのスパース化)

$$\underset{\Theta}{\text{minimize}} \quad f(\Theta) + \lambda \|\Theta\|_1 + \eta \|\Theta\|_*$$

Θ のトレースノルム
(特異値の和 = 特異値ベクトルのL1ノルム)

- ▶ **トレースノルム正則化**: 行列のトレースノルムをペナルティ項として加える→不要な特異値がゼロに(低ランク化)

損失関数について

- ▶ ここでは以下のカーネル平滑化**擬似尤度**を用いる

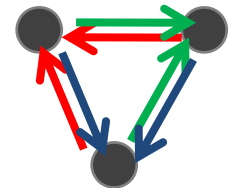
$$f(\Theta) = \frac{1}{T} \sum_{t=1}^T \sum_{t'=1}^T K_w(t - t') l(\mathbf{x}(t'), \boldsymbol{\theta}(t))$$

- Besag(1975)の擬似尤度 = 条件つき尤度の積

$$l(\mathbf{x}, \boldsymbol{\theta}) = - \sum_{i=1}^D \log p(x_i \mid \mathbf{x}_{\setminus i}, \boldsymbol{\theta})$$

cf. Höfling et al.'09
L1正則化擬似尤度による
(時変でない)構造推定

- ▶ ロジスティック回帰法と比較して,
1) パラメータ数が約半数に, 2) 後処理が不要



最適化の方法

$$\underset{\Theta}{\text{minimize}} \quad f(\Theta) + \lambda \|\Theta\|_1 + \eta \|\Theta\|_*$$

- ▶ 滑らかでない関数を含む凸最適化問題
- ▶ トレースノルムは定義に特異値分解を含む
- ▶ 正則化項が1つの場合は効率的な解法が有名
 - Iterative Shrinkage 法
 - Dual Augmented Lagrangian (DAL) (Tomiooka et al.'09,'10)

$$\underset{\Theta}{\text{minimize}} \quad f(\Theta) + \lambda \|\Theta\|_1$$

$$\underset{\Theta}{\text{minimize}} \quad f(\Theta) + \eta \|\Theta\|_*$$

交互方向乗数法 (ADMM)

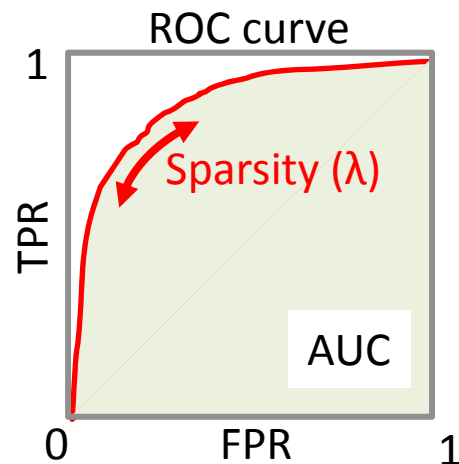
- ▶ Alternating Direction Method of Multiplier (ADMM) 法は2つ以上の正則化項に対処する1つのアプローチ
- ▶ DAL等と同じく拡張ラグランジュ法の一種で, 対象とする問題を比較的簡単な部分問題に分解して解く
- ▶ 大規模問題や並列化に向くとして機械学習分野で最近注目されている (cf. Boyd et al. 2010)
- ▶ 今回の問題では, 部分問題が比較的簡単に解けるのがポイント (Iterative Shrinkage やDALに共通)
 - 特異値分解としきい値処理
- ▶ アルゴリズムの詳細は省略

計算機実験

- ▶ 構造推定の性能評価
 - 人工データ2種類
- ▶ 予測性能の評価
 - 米上院・議決投票データ

構造推定の性能評価

- ▶ 人工データ(真のグラフ構造が既知)を用いて, 二値判別問題(辺あり／なし)として評価する
 - 例題数 = 辺の数 × 時系列の長さ
- ▶ 評価値: スパース性の強さを変化させた際のROCカーブの下側面積(AUC値)

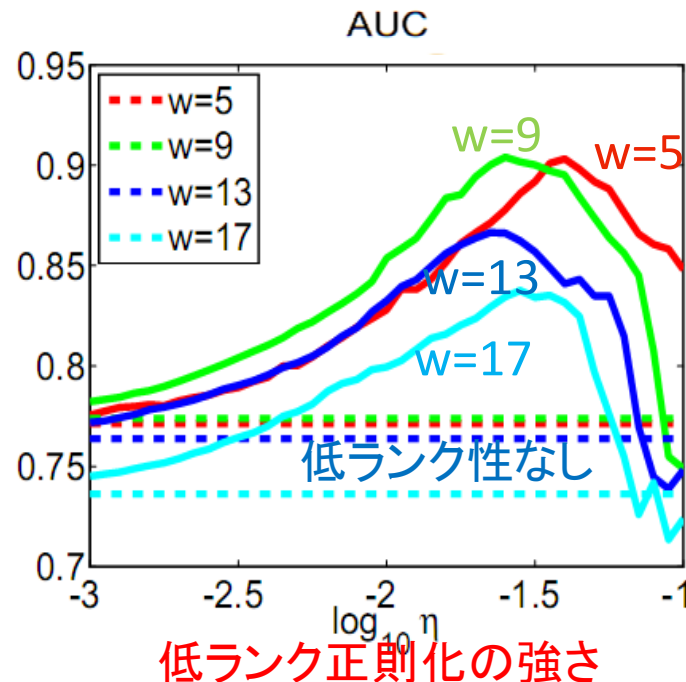
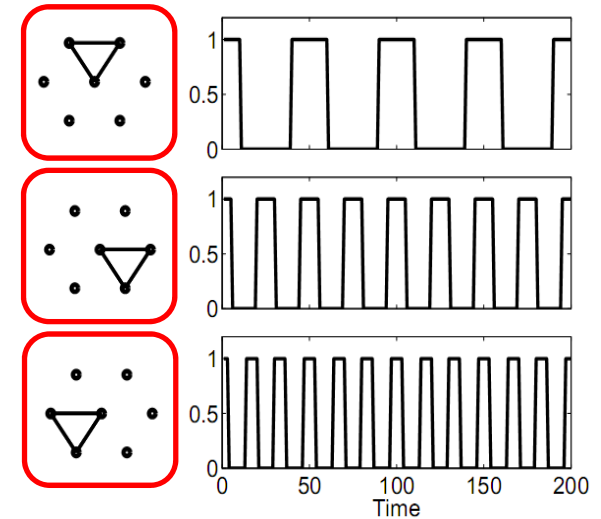


- ▶ False Positive Rate = $FP / (TN + FP)$
- ▶ True Positive Rate = $TP / (TP + FN)$

推定 \ 真実	有り	無し
有り	TP	FP
無し	FN	TN

実験1

- ▶ 3つの部分グラフからなる時変MRFで人工的にデータ生成
 - 7変数・200サンプル
 - 重みは定数



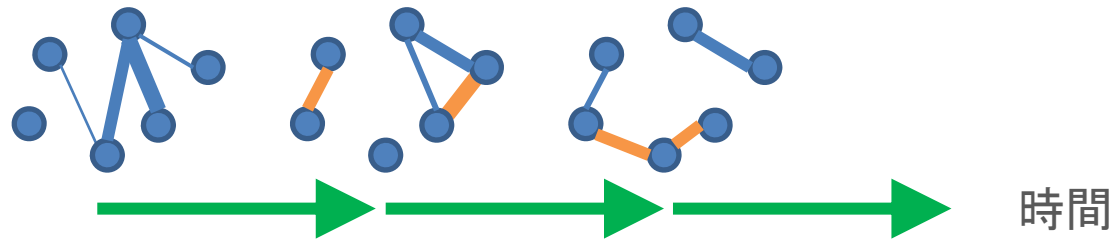
- 適切な低ランク性の導入によりAUCが改善しうる

低ランク正則化の強さ

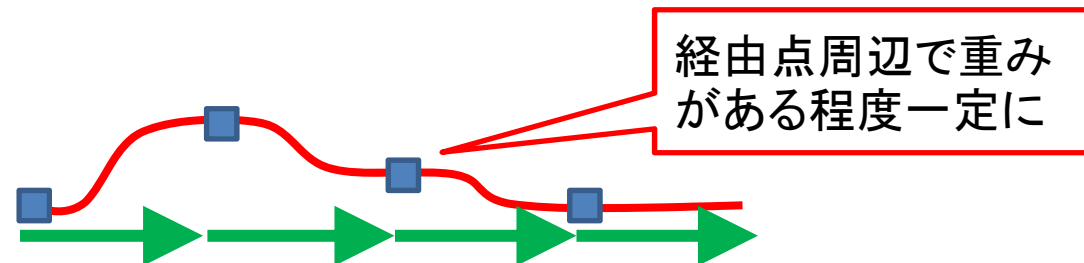
実験2

▶ ランダムなグラフ重み系列の作成(20変数360時刻)

1. 30辺をランダムに選び、重みを乱数で初期化する
2. 固定ステップ数(ブロック)毎に半数の辺を消去し、別の辺をランダムにつなぐ+すべての辺の重みをランダムに初期化

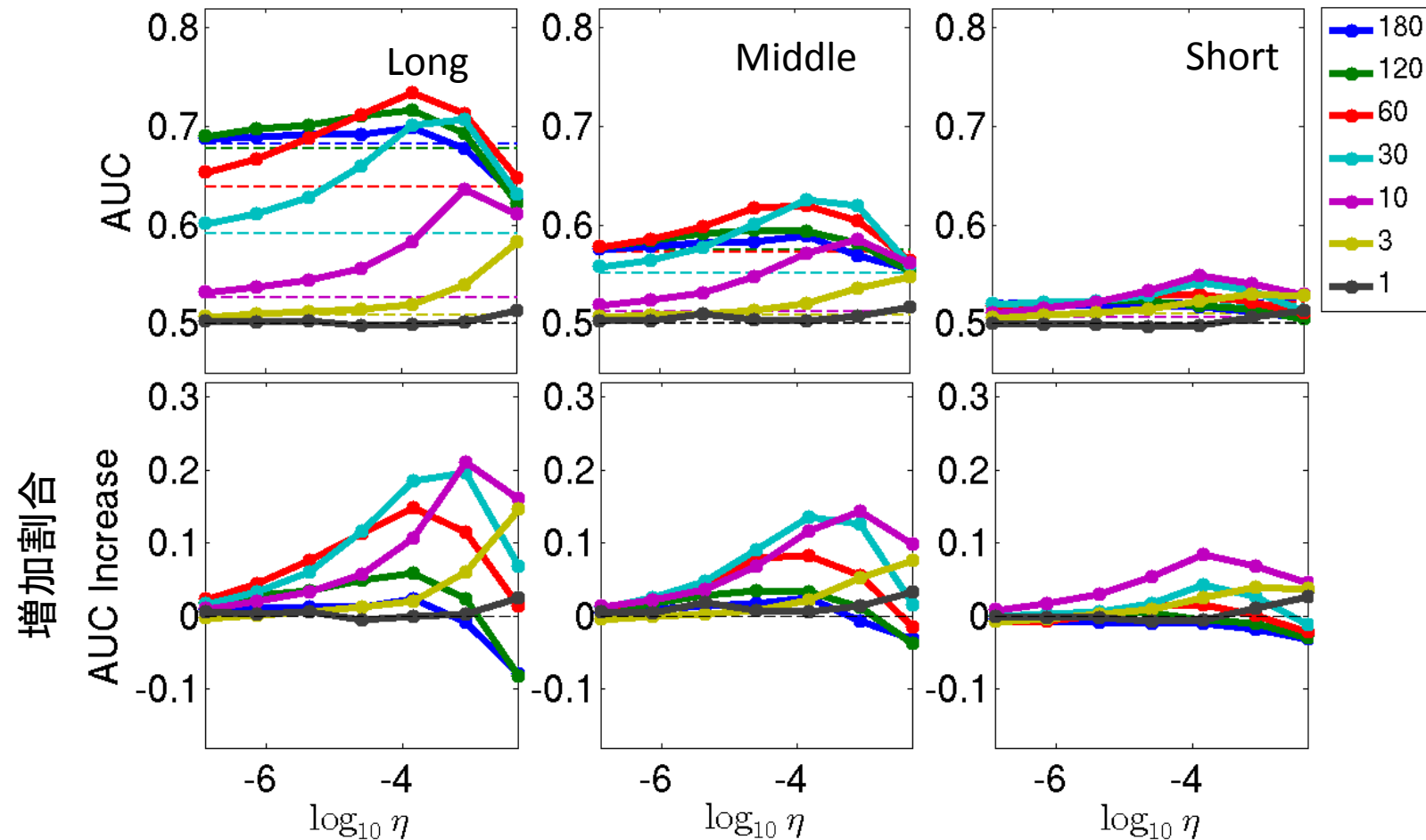


3. 各ブロック内でシグモイド曲線状に重みを変化させる



▶ ブロック長さ: 60 (Long), 20 (Medium), 5 (Short)

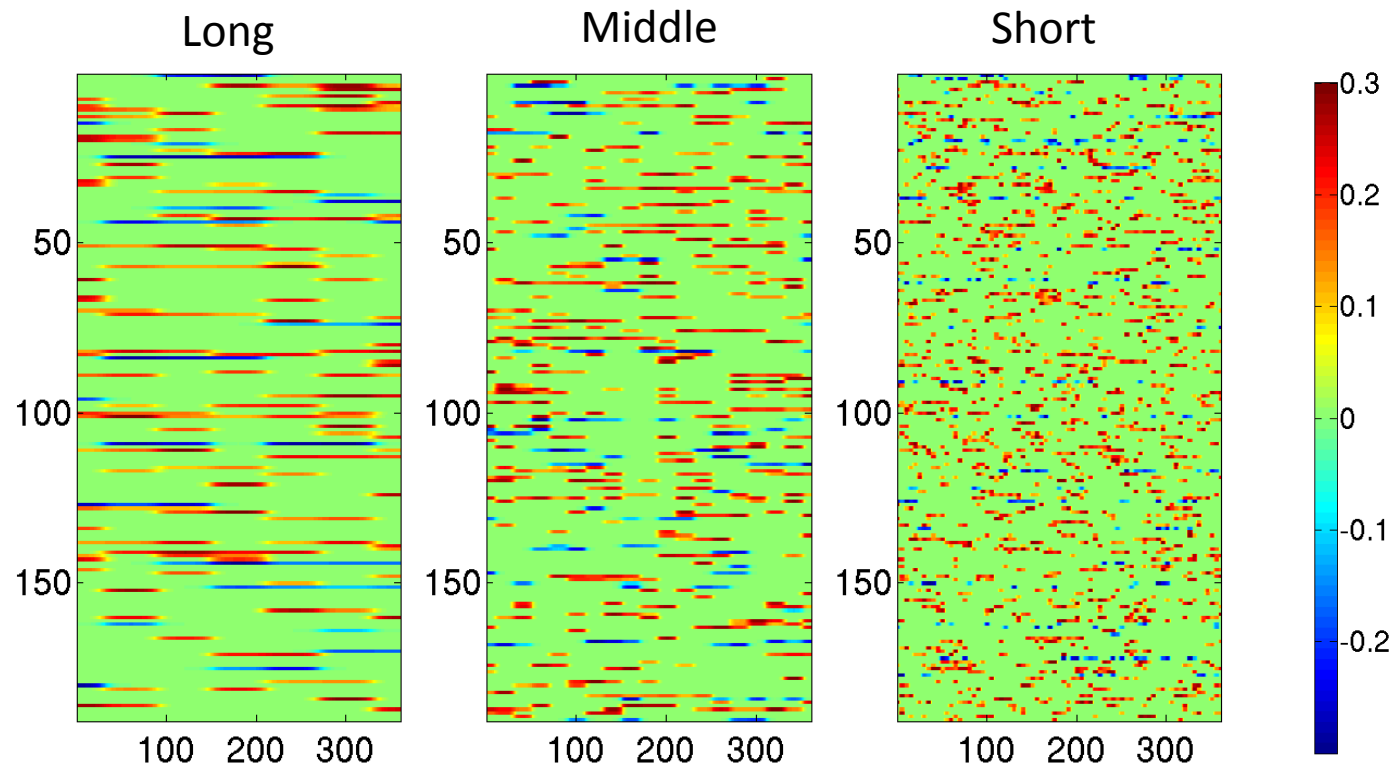
実験2: AUCの比較



- ▶ 定常性の強い条件ほど低ランク化の効果あり

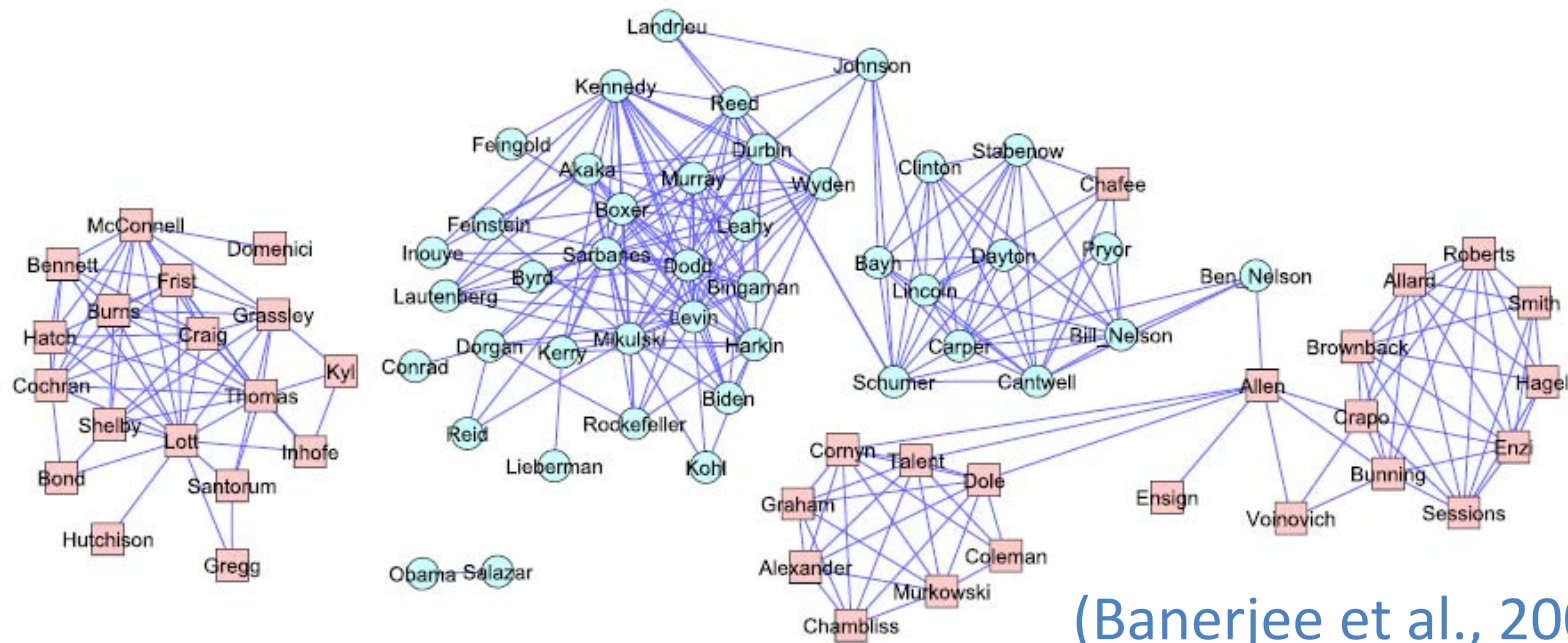
定常性と低ランク性

- ▶ 部分的な定常性が存在する場合，定常性の程度が強いほど，重み行列が低ランクになり易い



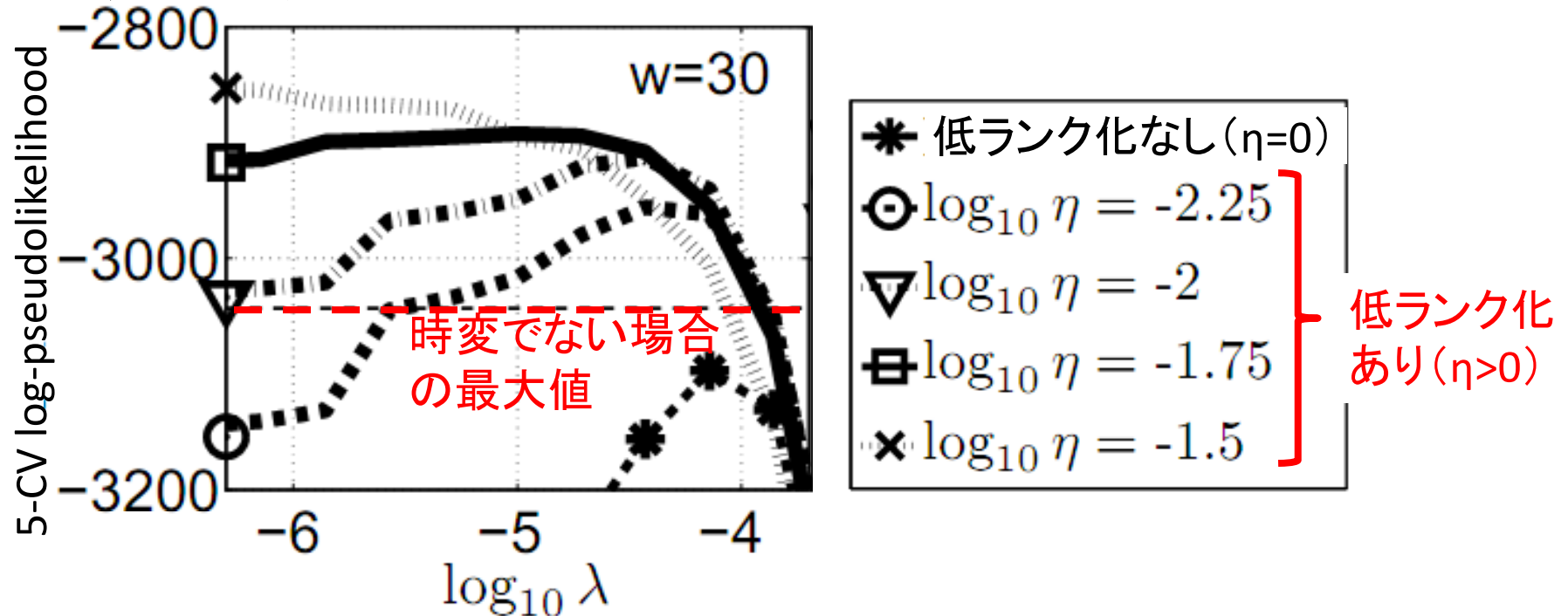
予測性能の評価

- ▶ 米上院議会，議決投票データ
 - 645個の法案についての賛成票／反対票の投票
 - 上院議員100名（100次元データ）



予測性能の評価：結果

- ▶ 擬似尤度をクロスバリデーションで評価



- ▶ 低ランク性の導入により、カーネル幅が短くとも予測性能の低下を避けうる

まとめ

- ▶ 2値マルコフ確率場に焦点をあて、時変ネットワーク構造推定に関する研究を紹介した
- ▶ 既存の方法は重みの時間的連続性を主に仮定する
- ▶ 新たに提案した手法はさらに重み行列の低ランク性を導入し、凸最適化(ADMM法)による推定を行う
- ▶ 人工データと実データを用いて、低ランク正則化の有効性を示した

今後の課題

- ▶ 正則化パラメータの決定法, 効率的な探索法
- ▶ イジングモデル以外への適用
- ▶ 脳計測データ・BMIへの具体的な応用