Data Analysis with Graphs

Elizabeth Purdom

Department of Biostatistics, UC, Berkeley

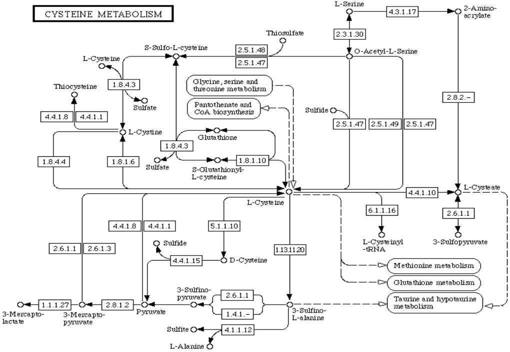
June 28, 2008

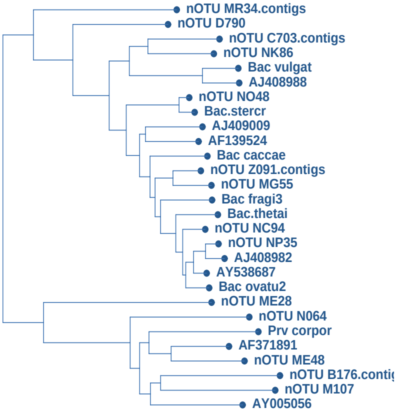
Outline

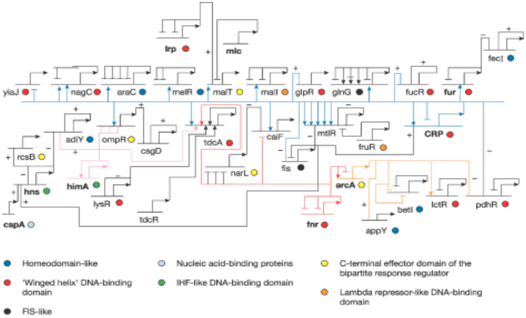
Outline

Graphical Data in Biology

- Phylogenetic trees
- Transcription networks
- Metabolic networks
- Signalling pathways

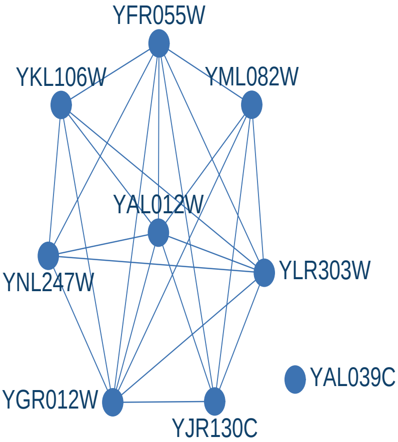






Data connected to Graphs

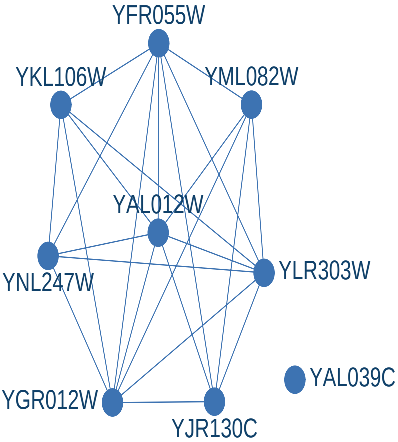
Known Graph on p genes



Data connected to Graphs

Data on p genes

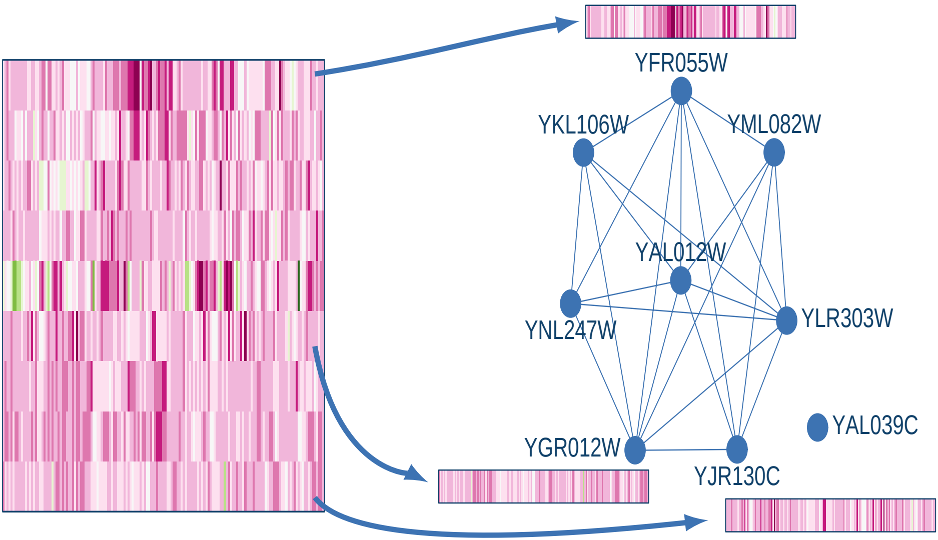
Known Graph on p genes



Data connected to Graphs

Data on p genes

Known Graph on p genes

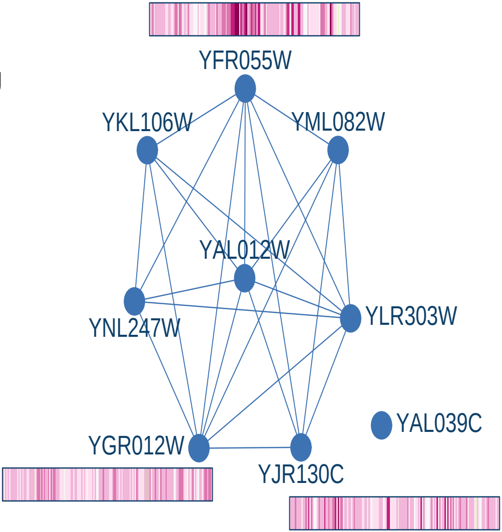


Data connected to Graphs

Data on p genes

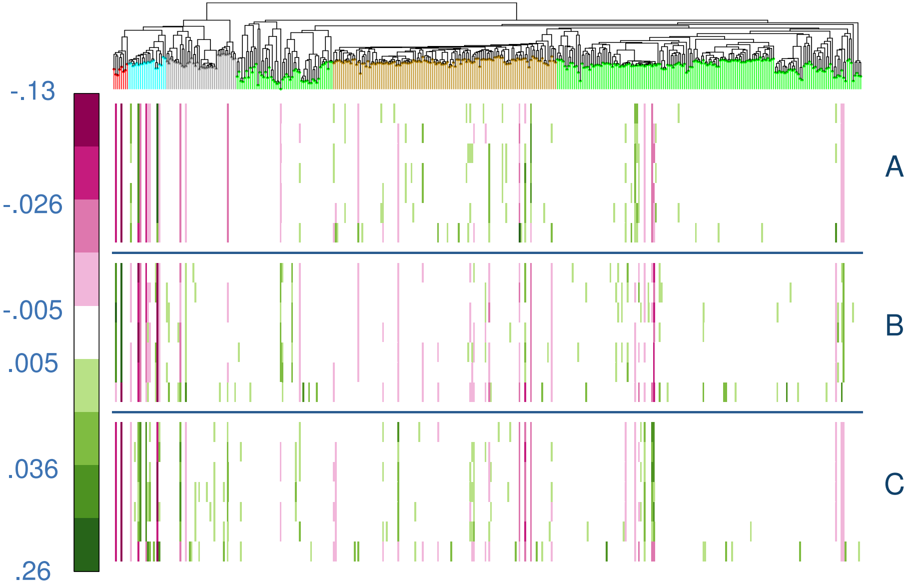
- Improved interpretability of inference/hypothesis testing
- Relevant Features for Prediction of Outcome
- Classify objects onto graph

Known Graph on p genes



Outline

Intestinal Bacterial Data



Eckburg et al. (2005)

Outline

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

$$\Downarrow$$

Transform based on $\mathcal{G}$, e.g. various summaries

- Averages over k subgraphs
- Difference between $\mathbf{x}_i$ and its neighbors
- Contrasts between subgraphs

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

Transform based on $\mathcal{G}$, e.g. various summaries

- Averages over k subgraphs
- Difference between $\mathbf{x}_i$ and its neighbors
- Contrasts between subgraphs

$$\Downarrow$$

Linear combinations based on graph: $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

Transform based on $\mathcal{G}$, e.g. various summaries

- Averages over k subgraphs
- Difference between $\mathbf{x}_i$ and its neighbors
- Contrasts between subgraphs

Linear combinations based on graph: $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$

$$\Downarrow$$

Decompose $\mathbf{x}$: $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \sum_i \tilde{x}_{(i)}\mathbf{v}_i$

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

Transform based on $\mathcal{G}$, e.g. various summaries

- Averages over k subgraphs
- Difference between $\mathbf{x}_i$ and its neighbors
- Contrasts between subgraphs

Linear combinations based on graph: $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$

Decompose $\mathbf{x}$: $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \sum_i \tilde{x}_{(i)}\mathbf{v}_i$

$$\Downarrow$$

Reweight directions: $f(\mathbf{x}) = \sum_i \boldsymbol{\lambda}_i^{1/2}\tilde{x}_{(i)}\mathbf{v}_i$

Transform Data Based on Graph

Observe $\mathbf{x} \in \mathbb{R}^p$ Graph $\mathcal{G}$ linking p variables

Transform based on $\mathcal{G}$, e.g. various summaries

- Averages over k subgraphs
- Difference between $\mathbf{x}_i$ and its neighbors
- Contrasts between subgraphs

Linear combinations based on graph: $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$

Decompose $\mathbf{x}$: $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \sum_i \tilde{x}_{(i)}\mathbf{v}_i$

Reweight directions: $f(\mathbf{x}) = \sum_i \boldsymbol{\lambda}_i^{1/2}\tilde{x}_{(i)}\mathbf{v}_i$

$$\Downarrow$$

New metric space with $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$: $\langle \mathbf{x}_k, \mathbf{x}_\ell \rangle_{\mathbf{S}} = \mathbf{x}_k^T\mathbf{S}\mathbf{x}_\ell$

Representing Graphs

- Adjacency Matrix: $\mathbf{A}$

$$\exp(\alpha \mathbf{A}) = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k / k!$$

$$(\mathbf{I} - \alpha \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k$$

Number of Paths

- Laplacian: $\mathbf{L} = \mathbf{D} - \mathbf{A}$

$$\exp(\alpha \mathbf{L}) = \sum_{k=0}^{\infty} \alpha^k \mathbf{L}^k / k!$$

$\leftarrow$ *Heat Kernel*

$$(\mathbf{I} + \alpha \mathbf{L})^{-1} = \sum_{k=0}^{\infty} \alpha^k (-\mathbf{L})^k / k!$$

$$\mathbf{L}^+$$

$\leftarrow$ *Commute Distance*

- Normalized Laplacian $\mathcal{L} = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2}$

Smola and Kondor (2003); Fouss et al. (2003)

Representing Graphs

- Adjacency Matrix: $\mathbf{A}$

$$\exp(\alpha\mathbf{A}) = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k / k!$$

Number of Paths

$$(\mathbf{I} - \alpha\mathbf{A})^{-1} = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k$$

- Laplacian: $\mathbf{L} = \mathbf{D} - \mathbf{A}$

$$\exp(\alpha\mathbf{L}) = \sum_{k=0}^{\infty} \alpha^k \mathbf{L}^k / k!$$

$\leftarrow$ *Heat Kernel*

$$(\mathbf{I} + \alpha\mathbf{L})^{-1} = \sum_{k=0}^{\infty} \alpha^k (-\mathbf{L})^k / k!$$

$$\mathbf{L}^{+}$$

$\leftarrow$ *Commute Distance*

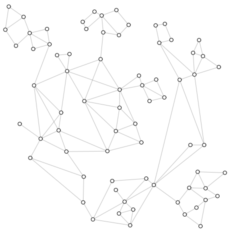
- Normalized Laplacian $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$

Analysis of Graphs

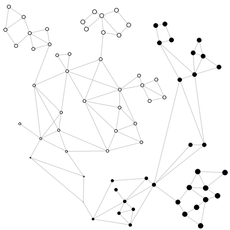
Data $\rightarrow$ Graph: Cut Algorithms, Laplacian Eigenmaps

Smola and Kondor (2003); Fouss et al. (2003)

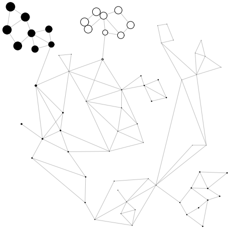
Sample of Eigenvectors for $\mathbf{L}$



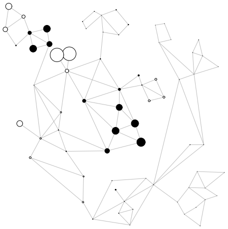
$\lambda_0 = 0$



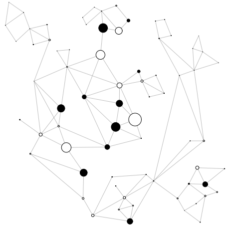
λ_1



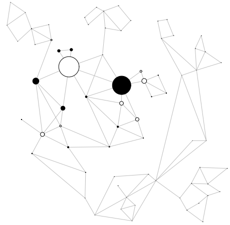
λ_2



λ_{12}

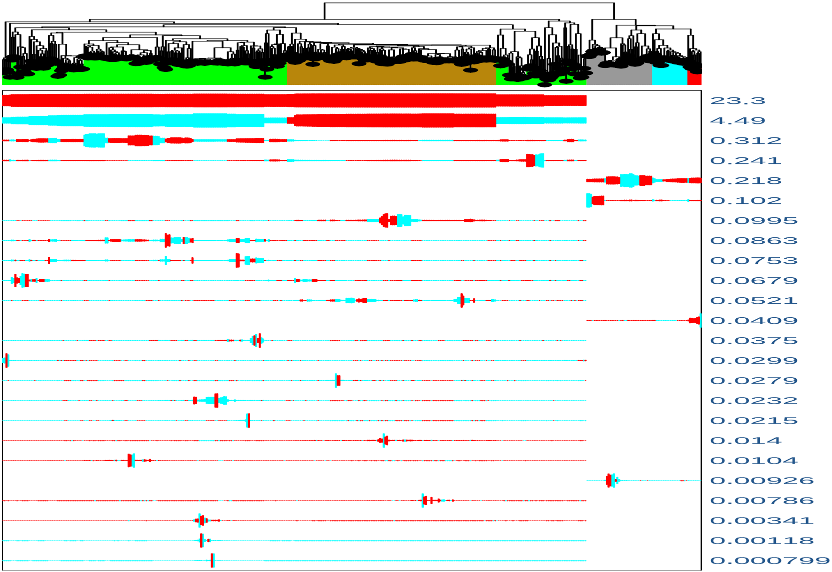


λ_{52}



λ_{71}

Sample of Eigenvectors of $\Sigma_{\mathcal{T}}$



Outline

Generalized PCA

Metric space with $\mathbf{S}_{p\times p}$, $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{X}_{n\times p}$

Find $\mathbf{b} \in \mathbb{R}^p$:

$$\max_{\|\mathbf{b}\|^2_{\mathbf{S}}=1} \text{var}(\langle \mathbf{b}, \mathbf{x} \rangle_{\mathbf{S}}) \quad \longrightarrow \quad \begin{gathered} \mathbf{USb}_i = \lambda_i \mathbf{b}_i \\ \mathbf{U} = \widehat{\text{cov}}(\mathbf{x}) = \mathbf{X}^T \mathbf{Q} \mathbf{X} \end{gathered}$$

Escoufier (1973)

Generalized PCA

Metric space with $\mathbf{S}_{p\times p}$, $\mathbf{x}\in\mathbb{R}^p$, $\mathbf{X}_{n\times p}$

Find $\mathbf{b}\in\mathbb{R}^p$:

$$\max_{\|\mathbf{b}\|^2_{\mathbf{S}}=1} \text{var}(\langle \mathbf{b},\mathbf{x}\rangle_{\mathbf{S}}) \quad \longrightarrow \quad \begin{aligned} \mathbf{USb}_i &= \lambda_i \mathbf{b}_i \\ \mathbf{U} = \widehat{\text{cov}}(\mathbf{x}) &= \mathbf{X}^T\mathbf{Q}\mathbf{X} \end{aligned}$$

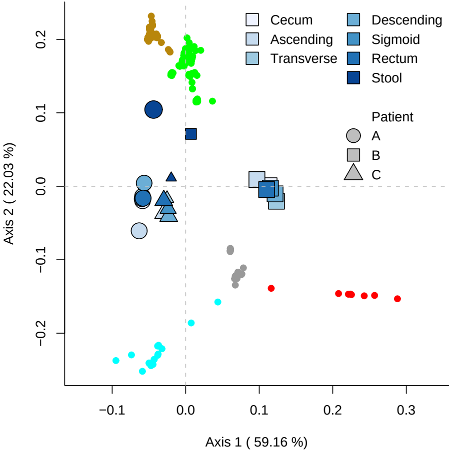
e.g.

- Correspondence Analysis $\Rightarrow$ $\mathbf{S}$, $\mathbf{Q}$ diagonals
  SALSA Algorithm
- PCA on $f(\mathbf{x}_i)$ $\Rightarrow$ $\mathbf{Q}=\mathbf{I}$
- DPCoA $\Rightarrow$ $\mathbf{S}=\mathbf{\Sigma}_{\mathcal{T}}$ covariance for trees, $\mathbf{Q}$ diagonal

Kernel PCA

Generalized PCA of Intestinal Data

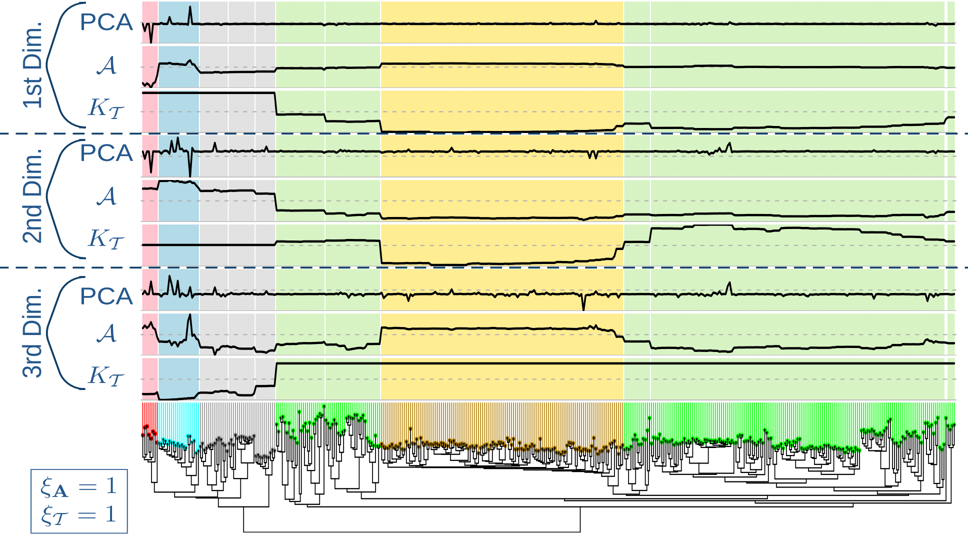
$\mathbf{x} \in \mathbb{R}^p$, $\mathbf{S} = \boldsymbol{\Sigma}_{\mathcal{T}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ PCA on $f(\mathbf{x}) = \sum_i \lambda^{1/2}\tilde{x}_{(i)}\mathbf{v}_i$



PCA as Transformations of $\mathbf{x}$

Regular PCA: $\quad \mathbf{x} \quad \Rightarrow \quad \hat{\mathbf{x}}_{PCA} = \mathbf{A}_{PCA}^T \mathbf{x}$

Generalized PCA: $\quad f(\mathbf{x}) \quad \Rightarrow \quad \hat{\mathbf{x}}_F = \mathbf{A}_{\mathcal{G}}^T f(\mathbf{x}) = \mathcal{A}^T \mathbf{x}$



$\xi_{\mathbf{A}} = 1$
$\xi_{\mathcal{T}} = 1$

Outline

Supervised Classification/Prediction

Outcome: y
Predictor variable: $f(\mathbf{x}) \in \mathbb{R}^p$

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} \ell(\boldsymbol{\beta}^T f(\mathbf{x}_i), y_{(i)}) + C\|\boldsymbol{\beta}\|^2$$

$$\Leftrightarrow {}^{*}$$

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \sum_{i=1}^{n} \ell(\langle \boldsymbol{\alpha}, \mathbf{x}_i \rangle_{\mathbf{S}}, y_{(i)}) + C\|\boldsymbol{\alpha}\|_{\mathbf{S}}^2$$

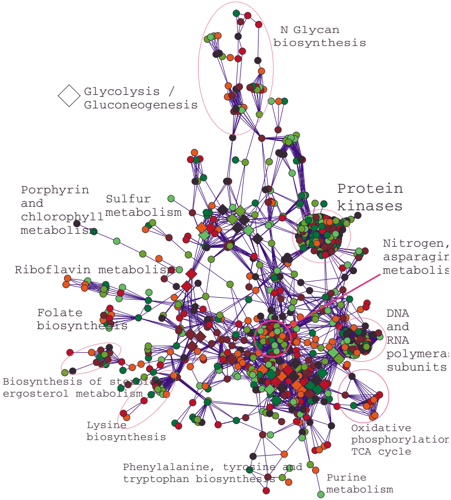
${}^{*}\boldsymbol{\beta} = \mathbf{S}^{1/2}\boldsymbol{\alpha}$ Implicitly assume: $\boldsymbol{\beta} \in \text{span}\{\mathbf{v}_i\}_{i:\boldsymbol{\lambda}_i > 0}$

Rapaport et al. (2007)

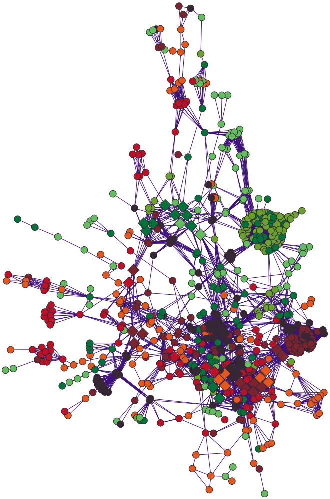
SVM Decision Rule

Predicting Irradiated Yeast Cultures from Microarray Data

Without Graph



With Graph



Rapaport et al. (2007, Figure 3)

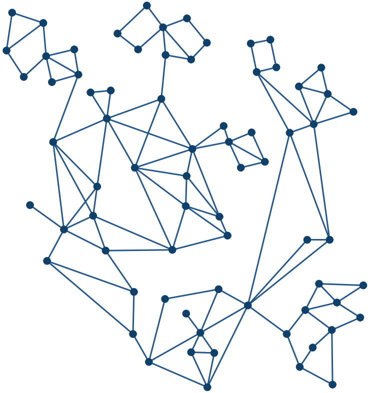
Outline

Predictions related to $\mathcal{G}$:

- New node, predict edges based on $\mathbf{x}$

Data $\mathbf{X}$, graph $\mathcal{G}_n$ on n observations

$$\mathbf{X} = \begin{pmatrix} x_{1(1)} & \cdots & x_{1(p)} \\ x_{2(1)} & \cdots & x_{2(p)} \\ & \vdots & \\ x_{n(1)} & \cdots & x_{n(p)} \end{pmatrix}$$



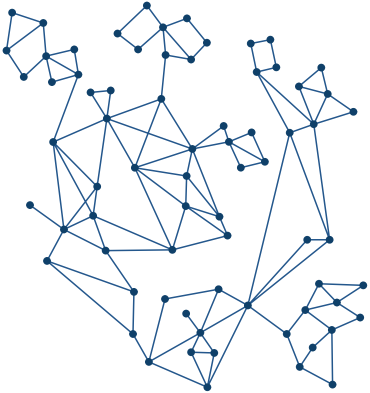
Predictions related to $\mathcal{G}$:

- New node, predict edges based on $\mathbf{x}$

Data $\mathbf{X}$, graph $\mathcal{G}_n$ on n observations

$$\mathbf{X} = \begin{pmatrix} x_{1(1)} & \cdots & x_{1(p)} \\ x_{2(1)} & \cdots & x_{2(p)} \\ & \vdots & \\ x_{n(1)} & \cdots & x_{n(p)} \end{pmatrix}$$

$$\mathbf{X}_* = \begin{pmatrix} x^*_{1(1)} & \cdots & x^*_{1(p)} \\ x^*_{2(1)} & \cdots & x^*_{2(p)} \\ & \vdots & \\ x^*_{m(1)} & \cdots & x^*_{m(p)} \end{pmatrix}$$



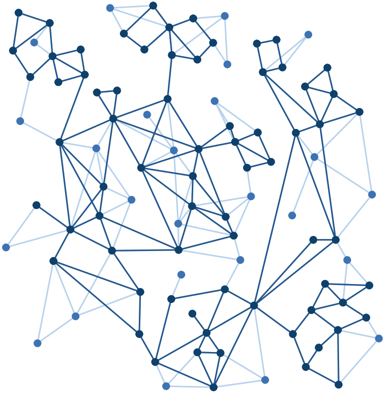
Predictions related to $\mathcal{G}$:

- New node, predict edges based on $\mathbf{x}$

Data $\mathbf{X}$, graph $\mathcal{G}_n$ on n observations $\quad \Rightarrow \quad \mathcal{G}_{n+m}$

$$\mathbf{X} = \begin{pmatrix} x_{1(1)} & \cdots & x_{1(p)} \\ x_{2(1)} & \cdots & x_{2(p)} \\ & \vdots & \\ x_{n(1)} & \cdots & x_{n(p)} \end{pmatrix}$$

$$\mathbf{X}_* = \begin{pmatrix} x^*_{1(1)} & \cdots & x^*_{1(p)} \\ x^*_{2(1)} & \cdots & x^*_{2(p)} \\ & \vdots & \\ x^*_{m(1)} & \cdots & x^*_{m(p)} \end{pmatrix}$$

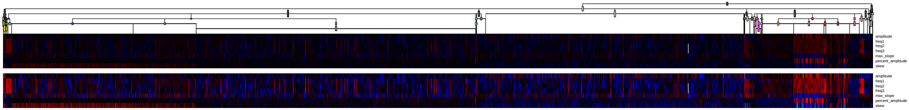


Predictions related to $\mathcal{G}$:

- New node, predict edges based on $\mathbf{x}$
- New observation, assign to node of $\mathcal{G}$ based on $\mathbf{x}$

Data $\mathbf{X}$, graph $\mathcal{G}_d$ on n observations

$$\mathbf{X} = \begin{pmatrix} x_{1(1)} & \cdots & x_{1(p)} \\ x_{2(1)} & \cdots & x_{2(p)} \\ & \vdots & \\ x_{n(1)} & \cdots & x_{n(p)} \end{pmatrix}$$

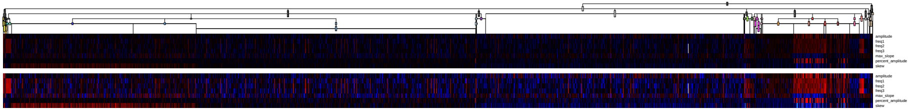


Predictions related to $\mathcal{G}$:

- New node, predict edges based on $\mathbf{x}$
- New observation, assign to node of $\mathcal{G}$ based on $\mathbf{x}$

Data $\mathbf{X}$, graph $\mathcal{G}_d$ on n observations $\quad \Rightarrow \quad$ assign to d classes

$$\mathbf{X} = \begin{pmatrix} x_{1(1)} & \cdots & x_{1(p)} \\ x_{2(1)} & \cdots & x_{2(p)} \\ & \vdots & \\ x_{n(1)} & \cdots & x_{n(p)} \end{pmatrix}$$



Unifying the Problems

Predict New Edges:
$\mathbb{1}\{i \sim j\} = f(\mathbf{x}_i, \mathbf{x}_j)$

Classify New Observations:
$\mathbb{1}\{i \in v\} = f(\mathbf{x}_i)$

Unifying the Problems

Predict New Edges:
$\mathbb{1}\{i \sim j\} = f(\mathbf{x}_i, \mathbf{x}_j)$

Classify New Observations:
$\mathbb{1}\{i \in v\} = f(\mathbf{x}_i)$

$$\Downarrow$$

Instead Predict Similarities:

$$\begin{aligned} S_{\mathcal{G}}(i,j) &= f(\mathbf{x}_i, \mathbf{x}_j) + \epsilon_{ij} \\ &= g(S_x(i,j)) + \epsilon_{ij} \end{aligned}$$

Simple fitting: $\min \sum_{i,j} (S_{\mathcal{G}}(i,j) - c_{ij} S_x(i,j))^2$

Unifying the Problems

Predict New Edges:
$\mathbb{1}\{i \sim j\} = f(\mathbf{x}_i, \mathbf{x}_j)$

Classify New Observations:
$\mathbb{1}\{i \in v\} = f(\mathbf{x}_i)$

$$\Downarrow$$

Instead Predict Similarities:
$$\begin{aligned} S_{\mathcal{G}}(i,j) &= f(\mathbf{x}_i, \mathbf{x}_j) + \epsilon_{ij} \\ &= g(S_x(i,j)) + \epsilon_{ij} \end{aligned}$$

Simple fitting: $\min \sum_{i,j} (S_{\mathcal{G}}(i,j) - c_{ij} S_x(i,j))^2$

Regression Model: $\mathbf{y} = \mathbf{C}\mathbf{x} + \boldsymbol{\epsilon}$

Constraints on $\mathbf{C}$ (Reduced Rank Regression):

- Low rank
- Uncorrelated

Geometric Goal

Maximize similarity between clouds of data (RV-Coeff)

Transform $\mathbf{S}_x \quad \rightarrow \quad \widehat{\mathbf{S}}_{\mathcal{G}}$,

$$\max \left\langle \mathbf{S}_{\mathcal{G}}, \widehat{\mathbf{S}}_{\mathcal{G}} \right\rangle$$

- Gives interpretation of relation between data and graph
- Use for exploration

(Robert and Escoufier, 1976)

Regularization and Kernel Methods

- Clear $\mathbf{S}_{\mathcal{G}}, \mathbf{S}_x$ define kernels: Kernel CCA, PLS, RRR
  Yamanishi et al. (2004): yeast data
- Previous formulation: unregularized kernel RRR
- Implement trade-off between (empirically) uncorrelated features and the norm of the features

(Bach and Jordan, 2002; Rosipal and Trejo, 2001; Purdom, 2006)

Regularization and Kernel Methods

- Clear $\mathbf{S}_{\mathcal{G}}, \mathbf{S}_x$ define kernels: Kernel CCA, PLS, RRR
  Yamanishi et al. (2004): yeast data
- Previous formulation: unregularized kernel RRR
- Implement trade-off between (empirically) uncorrelated features and the norm of the features
- If $x = \mathbf{x} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{K}_X = \mathbf{X}^T\mathbf{X}$

Kernel CCA/RRR $(\mathbf{K}_{\mathcal{G}}, \mathbf{K}_X)$ $\Leftrightarrow$ Generalized PCA $(\mathbf{X}, \mathbf{Q}, \mathbf{S})$

(For appropriate regularization of Kernel CCA)

In Summary

- Computationally simple ways of using large graphs
- Automatically picks features of the data
- ’Smooths’ data rather than explicitly limit analysis to graph

In Summary

- Computationally simple ways of using large graphs
- Automatically picks features of the data
- 'Smooths' data rather than explicitly limit analysis to graph

- Adjacency matrix not deal with moderate details
  Directed Edges
  Different Nodes
- Eigenvectors $\neq$ subnetworks
  $\Rightarrow$ Kind of approximation, but not explicit

Acknowledgements

- Susan Holmes
- Relman Lab (Stanford)
- Bloom Lab (Berkeley)

References I

BACH, F. R. and JORDAN, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research* **3** 1–48.

ECKBURG, P. B., BIK, E. M., BERNSTEIN, C. N., PURDOM, E., DETHLEFSEN, L., SARGENT, M., GILL, S. R., NELSON, K. E. and RELMAN, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* **308** 1635–1638.

ESCOUFIER, Y. (1973). Le traitement des variables vectorielles. *Biometrics* **29** 751–760.

FOUSS, F., RENDERS, J.-M. and SAERENS, M. (2003). Some relationships between Kleinberg's hubs and authorities, correspondence analysis and the Salsa algorithm. In *Proceedings of the Third International Conference on Data Mining*. IEEE.

References II

PURDOM, E. (2006). *Multivariate Kernel Methods in the Analysis of Graphical Structures*. Ph.D. thesis, Stanford University.

RAPAPORT, F., ZINOVYEV, A., DUTREIX, M., BARILLOT, E. and VERT, J.-P. (2007). Classification of microarray data using gene networks. *BMC Bioinformatics* **8**.

ROBERT, P. and ESCOUFIER, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Applied Statistics* **25** 257–265.

ROSIPAL, R. and TREJO, L. J. (2001). Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research* **2** 97–123.

References III

SMOLA, A. and KONDOR, I. (2003). Kernels and regularization on graphs. In *Proceedings of the Annual Conference on Computational Learning Theory* (B. Schölkopf and M. Warmuth, eds.). Lecture Notes in Computer Science, Springer.

YAMANISHI, Y., VERT, J.-P. and KANEHISA, M. (2004). Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics* **20** i363–i370.