A max-margin framework for supervised graph completion

Stuart Andrews and Tony Jebara
Columbia University
andrews@cs.columbia.edu

Graph Completion

- given
  - features for each potential edge

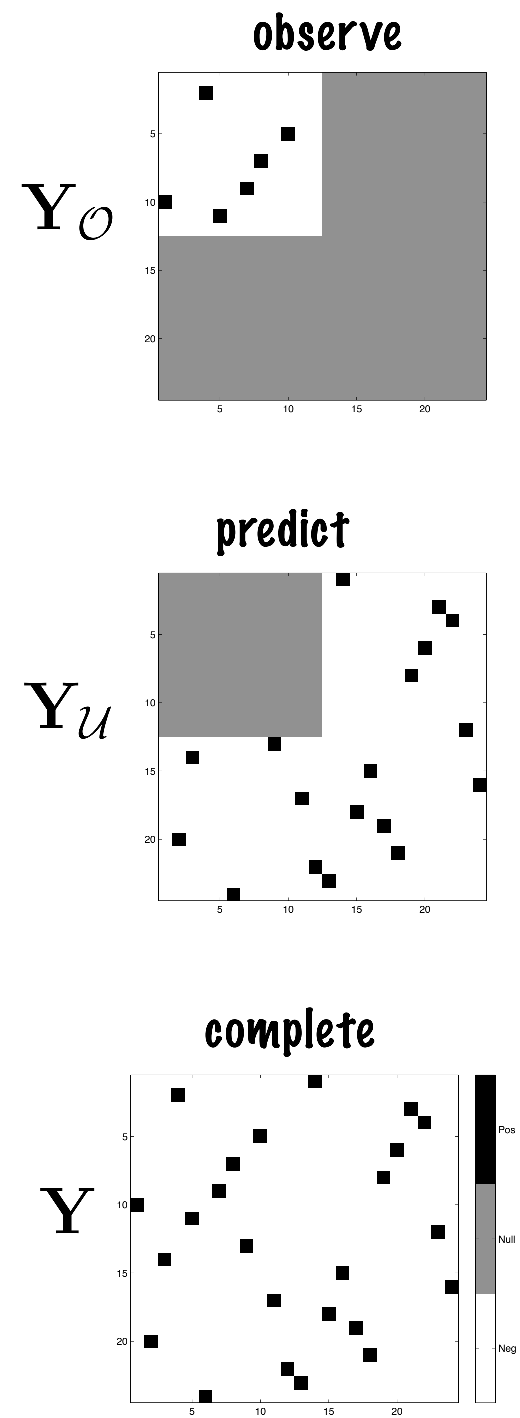
    $$\mathbf{X} = \{\mathbf{x}_{j,k}\}$$

  - partially observed connectivity

    $$\mathbf{Y}_\mathcal{O} \quad \text{from} \quad \mathbf{Y} = \{y_{j,k}\} \qquad y_{j,k} \in \{0, 1\}$$

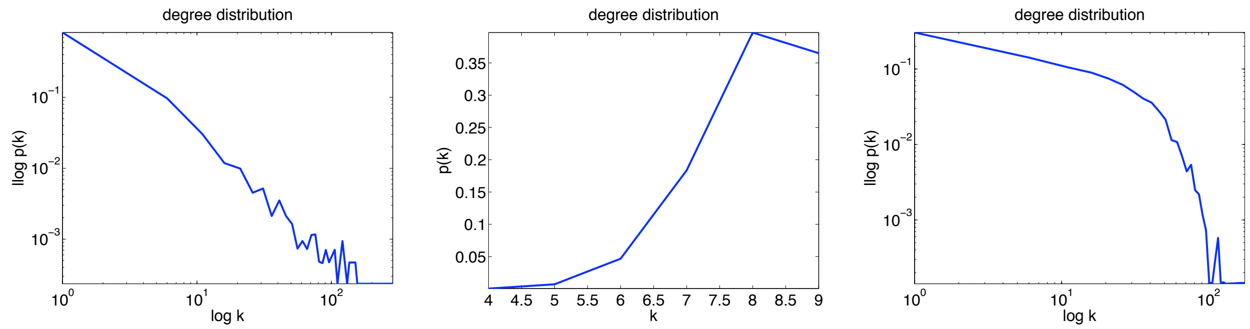
- task
  - complete connectivity

    $$\mathbf{Y} = \mathbf{Y}_\mathcal{O} \cup \mathbf{Y}_\mathcal{U}$$



Graph Completion

- approach 1: topology
  - topology of complete connectivity is similar to observed connectivity
  - e.g. Liben-Nowell and Kleinberg, 2003

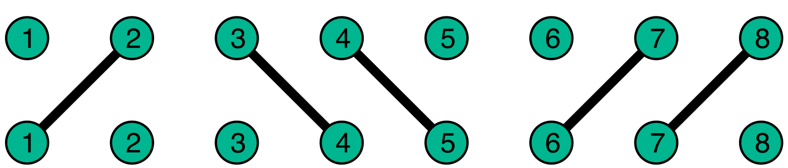


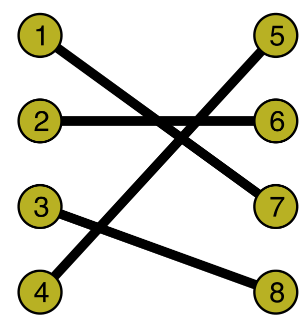
- approach 2: attribute-value learning
  - “on” / “off” property of edges is related to their features
  - e.g. Yamanishi, Vert, and Kanehisa, 2003

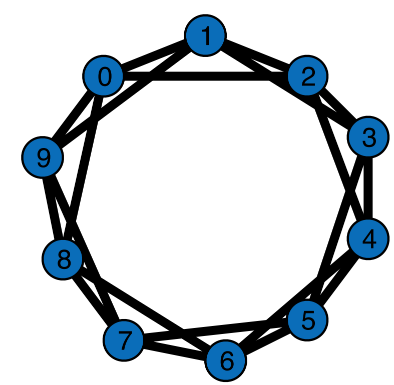
- idea: a marriage of structured-output models and degree-constrained subgraphs

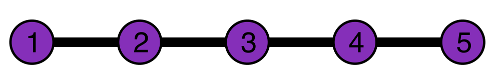
Structured Output Models

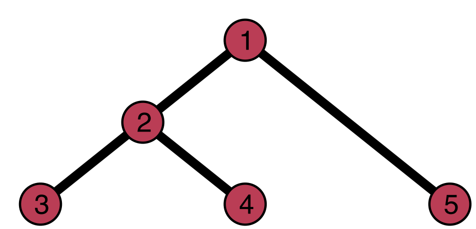
- multivariate interdependent outputs
  - chains, trees, alignments, matchings, DCS
  - parts $\mathbf{X} = \{\mathbf{x}_i\}$
  - labels $\mathbf{Y} = \{y_i\} \in \mathbb{Y}$ $\qquad \mathbb{Y} \subsetneq \{0,1\}^N$











Structured Predictions

- part scores

$$s_i = \mathbf{w}^T \mathbf{x}_i$$

- structure scores

$$\sum_{i \in \text{parts}} \mathbf{w}^T \mathbf{x}_i z_i \qquad \mathbf{Z} = \{z_i\} \qquad \mathbf{Z} \in \mathbb{Y}$$

- predictions

$$f(\mathbf{X}) = \underset{\mathbf{Z} \in \mathbb{Y}}{\operatorname{argmax}} \sum_{i \in \text{parts}} \mathbf{w}^T \mathbf{x}_i z_i$$

  - dynamic programming, combinatorial optimization, ...

Degree-Constrained Subgraphs

- "DCS" have in-degree and out-degree constraints on each node

$$\left\{ y_{j,k} \in \{0,1\} \right.$$

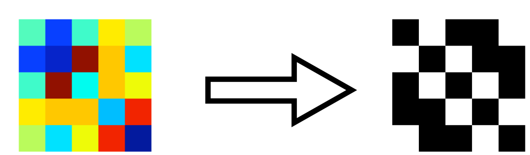
  - in-degree $\qquad \sum_j y_{j,k} \leq \boldsymbol{\delta}_k^{in}$
  - out-degree $\qquad \sum_k y_{j,k} \leq \boldsymbol{\delta}_j^{out}$

$$\left. \begin{array}{l} \sum_j y_{j,k} \leq \boldsymbol{\delta}_k^{in} \\ \sum_k y_{j,k} \leq \boldsymbol{\delta}_j^{out} \end{array} \right\} \qquad \mathbf{Y} \in \mathbb{Y}^{\text{DCS}}$$

  - degrees $\boldsymbol{\delta}_i^{in}$ $\boldsymbol{\delta}_i^{out}$ often available, or reliably estimated

- maximum-weight DCS
  - Given scores $s_{j,k}$ finding the maximum-weight DCS is easy $\mathcal{O}(n^3)$

$$\mathbf{Y} = \underset{\mathbf{Z} \in \mathbb{Y}^{\text{DCS}}}{\operatorname{argmax}} \sum_{j,k} s_{j,k} z_{j,k}$$



$s_{j,k}$ $\qquad$ $y_{j,k}$

Learning to Predict Graphs

- max-margin
  - margin $$\gamma(\mathbf{X},\mathbf{Y}) = \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} y_{j,k} - \max_{\substack{\mathbf{Z}\in\mathbb{Y}\\ \mathbf{Z}\neq\mathbf{Y}}} \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} z_{j,k}$$
  - omitting slack variables for simplicity

$$(\mathbf{X}^i,\mathbf{Y}^i) \sim P(\mathbf{X},\mathbf{Y})$$

$$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|^2$$

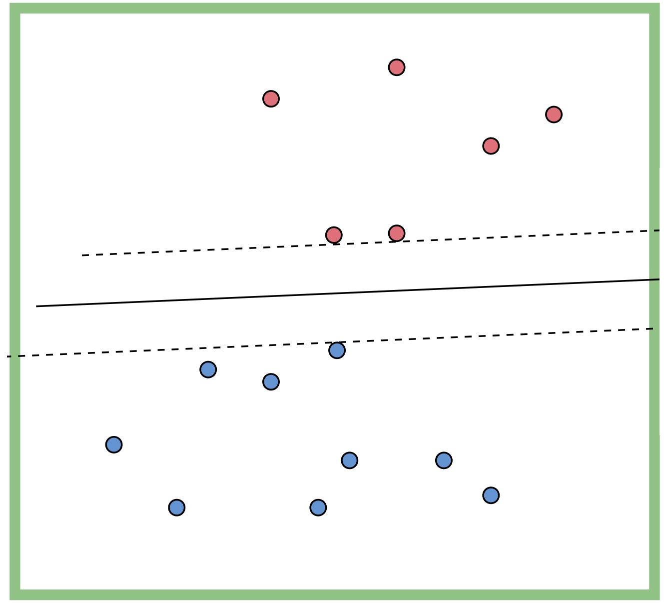
$$\text{s.t.} \sum_{j,k} \mathbf{w}^T \mathbf{x}^i_{j,k} y^i_{j,k} \geq \underbrace{\max_{\mathbf{Z}^i\in\mathbb{Y}} \sum_{j,k} \mathbf{w}^T \mathbf{x}^i_{j,k} z^i_{j,k} + \Delta^H\left(\mathbf{Y}^i,\mathbf{Z}^i\right)}\ ,\ \forall i$$

  - LHS: score of true structure
  - RHS: loss-augmented score of highest-scoring structure (all incorrect structures are ranked “much” lower than the true structure, if and only if, the highest-scoring incorrect structure is ranked “much” lower)
- algorithms
  - perceptron, exponentiated gradient, SMO, cutting-planes, extra & dual extragradient, search-based optimization

Inductive vs. Transductive

- inductive learning
  - maximize margin $\gamma(\mathbf{x}, y) = y(\mathbf{w}^T\mathbf{x} + b)$ over labeled examples

$$(\mathbf{x}^i, y^i) \sim P(\mathbf{x}, y) \Rightarrow f(\mathbf{x})$$

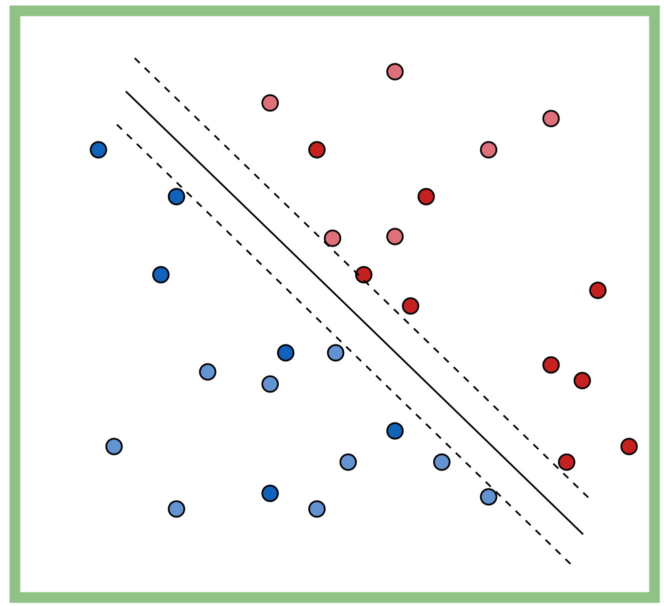


- transductive inference
  - define self-reinforcing margin for unlabeled examples

$$\gamma(\mathbf{x}, *) = \max_{y=\pm 1} y(\mathbf{w}^T\mathbf{x} + b)$$

  - maximize margin over ALL data

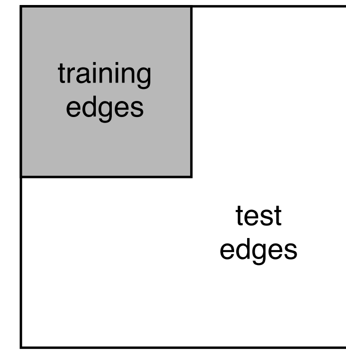
$$\begin{matrix}(\mathbf{x}^i, y^i) \\ (\mathbf{x}^j, *)\end{matrix} \Rightarrow \begin{matrix}(\mathbf{x}^i, y^i) \\ (\mathbf{x}^j, \hat{y}^j)\end{matrix}$$



Learning to Complete Graphs

- max-margin for partially-observed graphs
  - similar to transductive inference

$$(\mathbf{X}, \mathbf{Y}_{\mathcal{O}}, *) \mapsto (\mathbf{X}, \mathbf{Y}_{\mathcal{O}}, \hat{\mathbf{Y}}_{\mathcal{U}}) \qquad \mathbf{Y} = \mathbf{Y}_{\mathcal{O}} \cup \mathbf{Y}_{\mathcal{U}}$$



  - define self-reinforcing margin

$$\gamma(\mathbf{X}, \mathbf{Y}_{\mathcal{O}}) = \max_{\substack{\mathbf{V} \in \mathbb{Y}^{\mathrm{DCS}} \\ \mathbf{V}_{\mathcal{O}} = \mathbf{Y}_{\mathcal{O}}}} \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} v_{j,k} - \max_{\substack{\mathbf{Z} \in \mathbb{Y}^{\mathrm{DCS}} \\ \mathbf{Z}_{\mathcal{O}} \neq \mathbf{Y}_{\mathcal{O}}}} \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} z_{j,k}$$

  - maximize margin over ALL examples

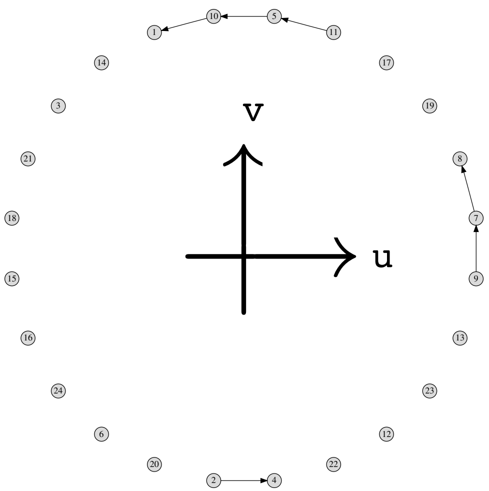
$$\min_{\mathbf{w},\xi} \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{s.t. } 0 \geq \min_{\substack{\mathbf{V} \in \mathbb{Y}^{\mathrm{DCS}} \\ \mathbf{V}_{\mathcal{O}} = \mathbf{Y}_{\mathcal{O}}}} \max_{\mathbf{Z} \in \mathbb{Y}^{\mathrm{DCS}}} \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} z_{j,k} + \Delta^H (\mathbf{V}, \mathbf{Z}) - \sum_{j,k} \mathbf{w}^T \mathbf{x}_{j,k} v_{j,k}$$

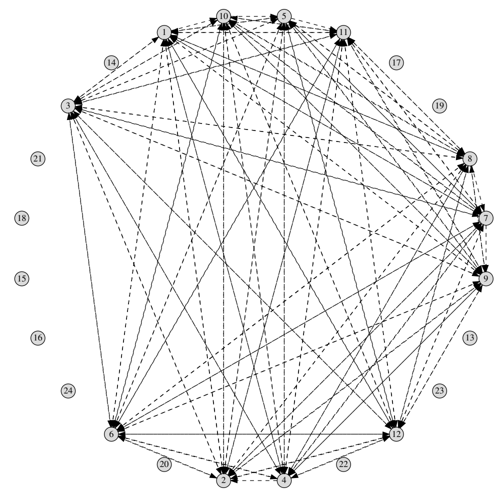
- algorithms
  - perceptron, cutting-planes, dual extragradient (DXG)

Circle

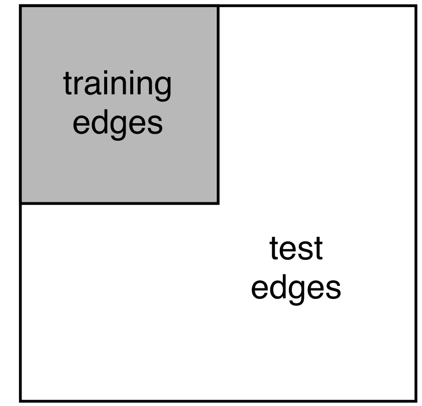
[nodes: 24, features: nodes $\mathbf{x}_j = (u_j, v_j)$, edges $\mathbf{x}_{j,k} = \mathbf{x}_j \mathbf{x}_k^T$]

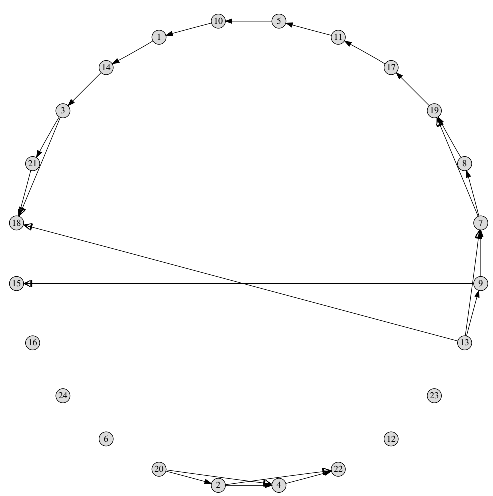


training
edges

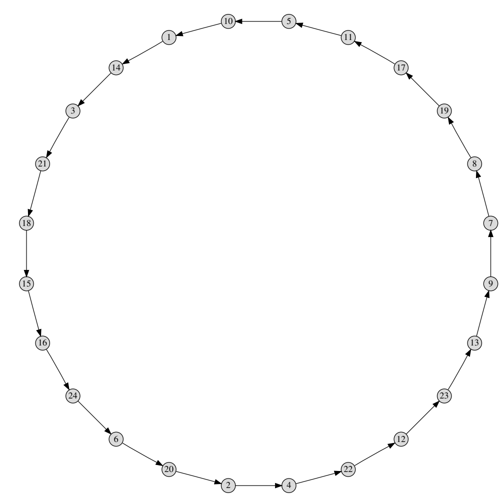


training
non-edges





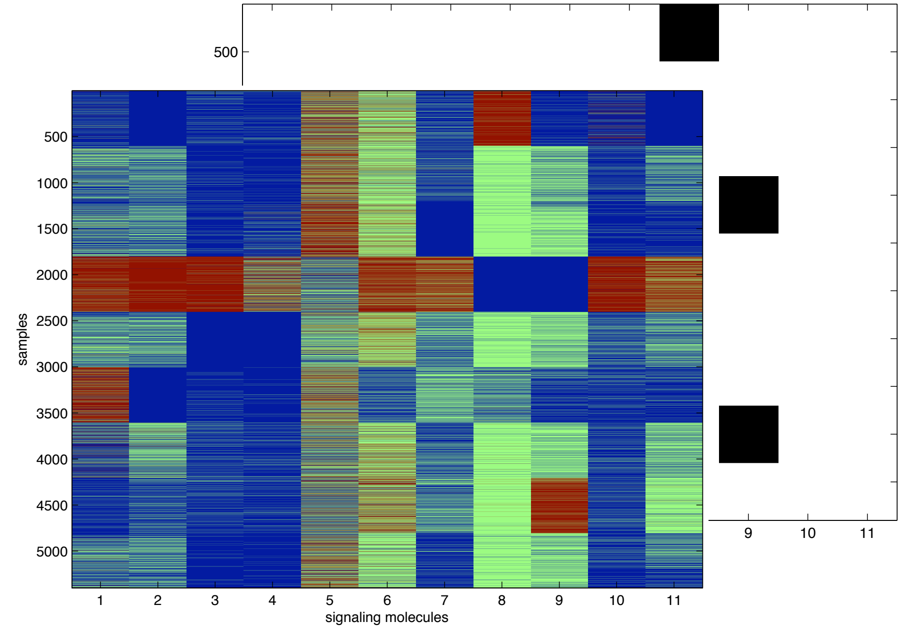
i.i.d. SVM



DXG
w/ transduction

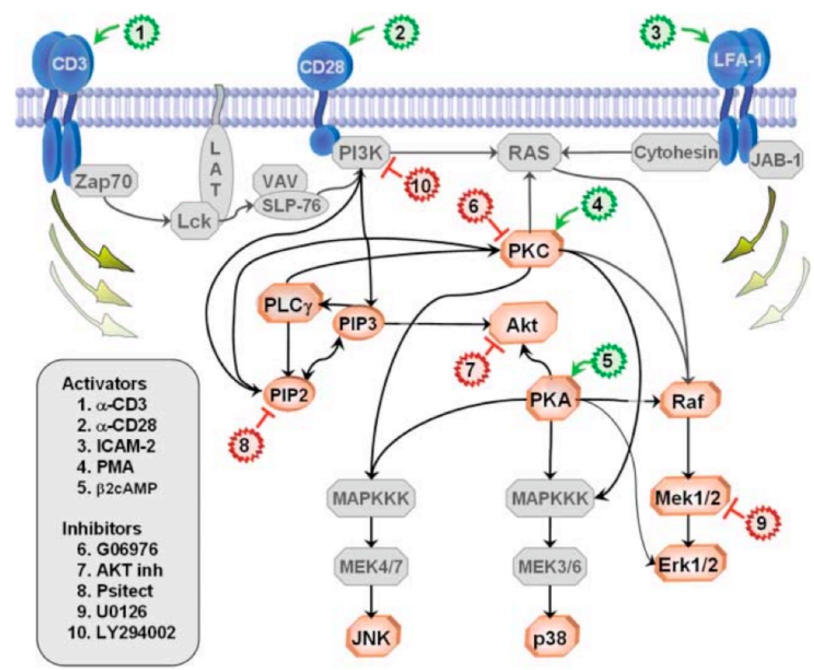
Signal Transduction

- flow cytometry expression levels $\mathbf{x}_j$ and intervention states $\bar{\mathbf{x}}_j$



- consensus network

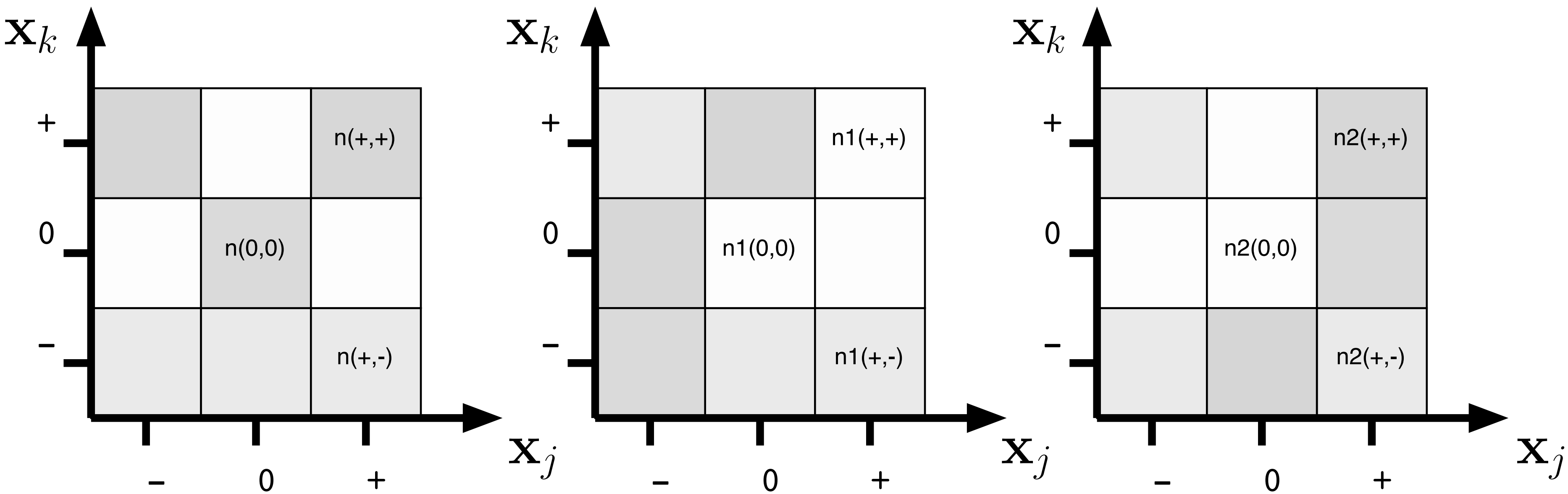
  $\mathbf{y}_{j,k}$



  - published by Karen Sachs et al. Science 308, 523 (2005) [formatted by Eaton and Murphy, AISTAT, 2007]

Signal Transduction

- edge features
  - based on expression level contingency tables

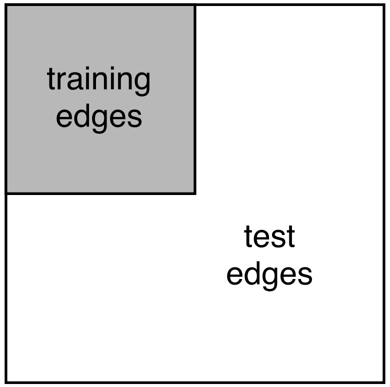


no interventions $\quad$ $\mathbf{x}_j$ intervened $\quad$ $\mathbf{x}_k$ intervened

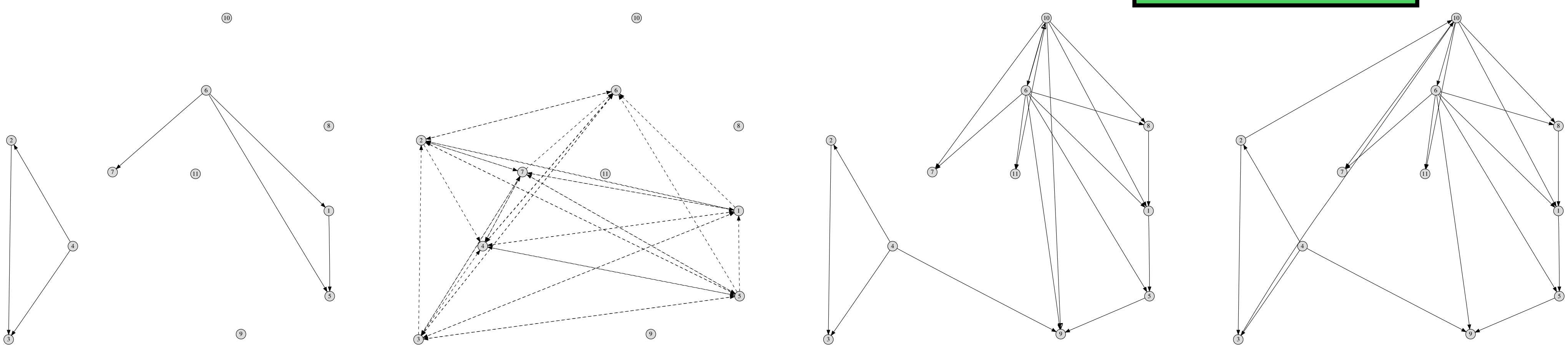
$$\mathbf{x}_{j,k} = (n_0(-,-), n_0(-,0), \ldots, n_2(+,0), n_2(+,+))$$

Signal Transduction

- graph completion protocol
  - blockwise partition of adjacency matrix
  - training edges labeled “on”/“off”
  - transductive max-margin training on all edges
  - predict “on”/”off” for test edges



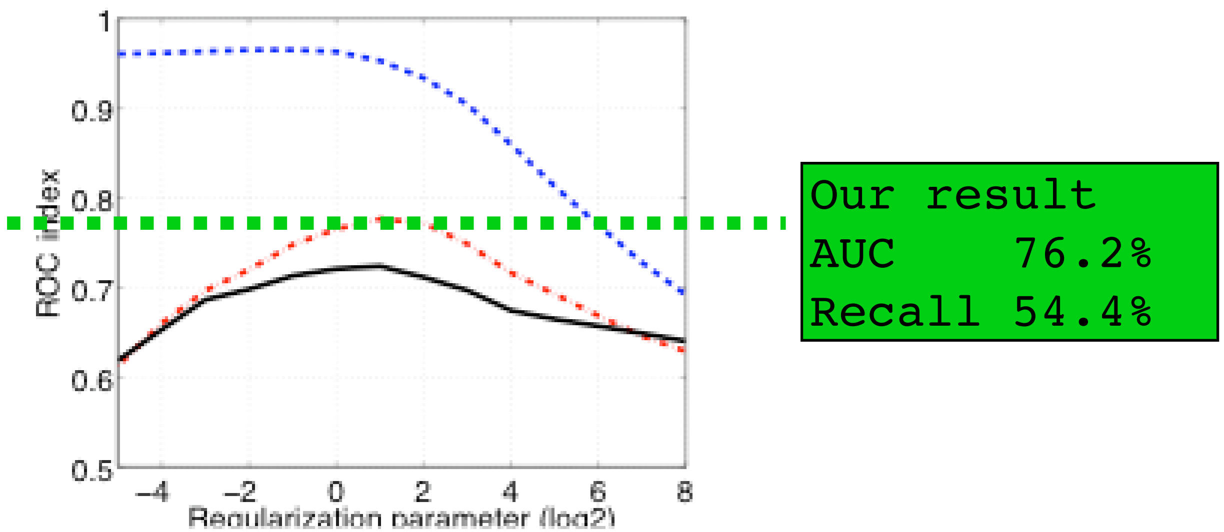
AUC 93%
Recall 80%



training edges | training non-edges | our completion | true completion

Protein-Protein Interaction

- interactions derived from several sources
  - Yamanishi, Vert and Kanehisa, 2004
  - features: expression, Y2H, localization, phylogenetic $x_{j,k}$
  - interactions: metabolic, physical and regulatory $y_{j,k}$
  - 769 nodes
  - state-of-the-art performance w/o optimizing parameters



(d) Integrated kernel

Vert and Yamanishi

Summary

- a marriage of structured-output models and degree-constrained subgraphs
  - topology and attribute-value learning
- evaluation on synthetic and real-world networks
  - task 1: complete circle network using x,y node positions
  - task 2: complete cell signaling network using flow cytometry expressions
  - task 3: complete protein-protein interactions using expression, Y2H, localization and phylogenetic features