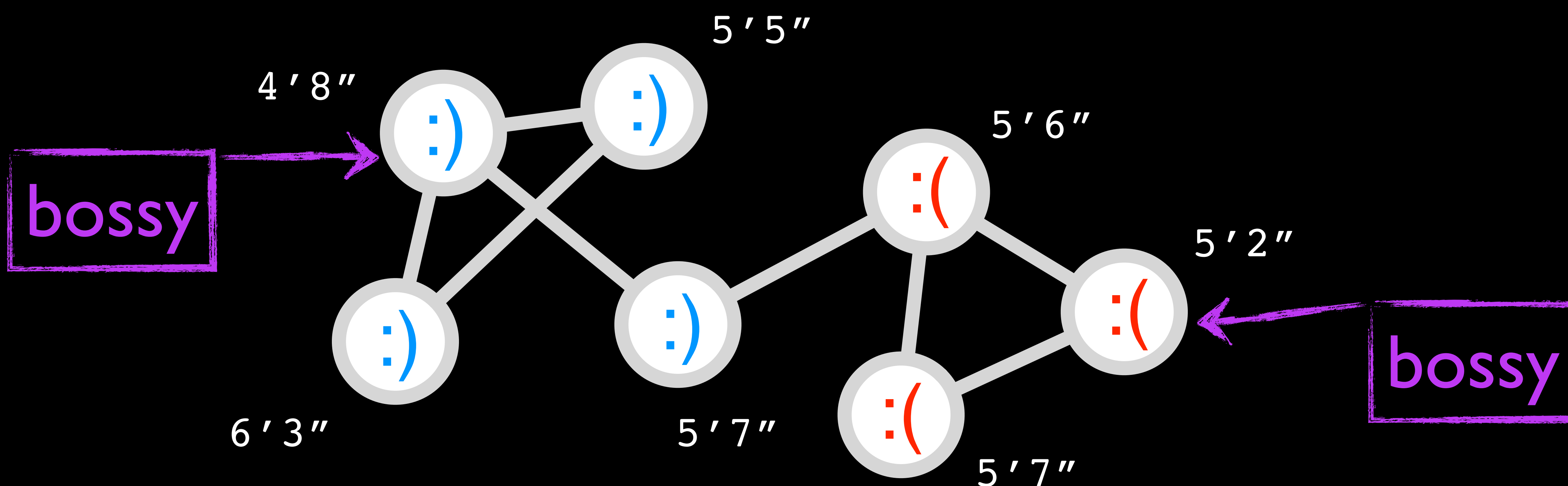
Learning a Degree-Augmented Distance Metric from a Network

Bert Huang, U of Maryland
Blake Shaw, Foursquare
Tony Jebara, Columbia U

Beyond Mahalanobis: Supervised Large-Scale Learning of Similarity
NIPS Workshop, Sierra Nevada, Dec. 16, 2011

Motivation: Similarity in Networks

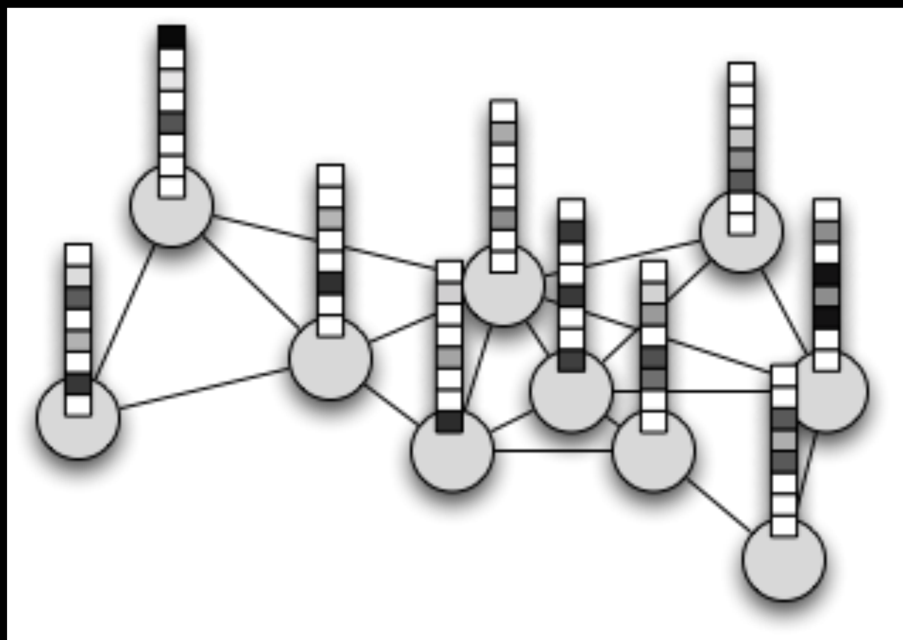


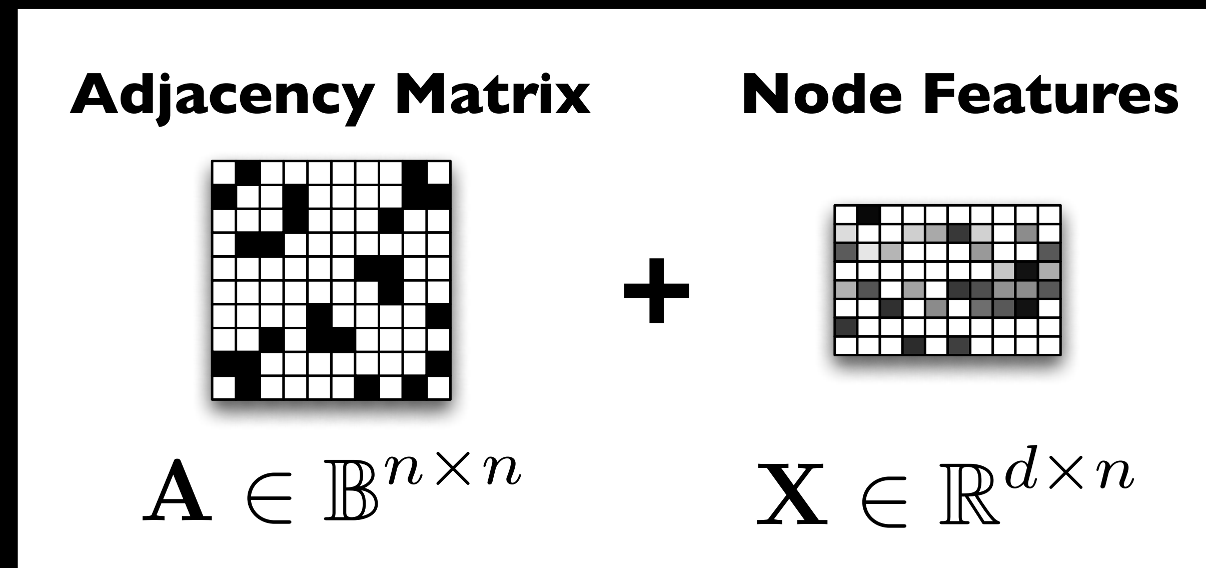
- Homophily occurs in natural networks: neighbors are similar
- Learning must account for structural nature of networks

Outline

- Problem formulation
- **S**tructure **P**reserving **M**etric **L**earning
- **D**egree **D**istributional **M**etric **L**earning
- Analysis
- Experiments

Problem Formulation





- Given: node feature matrix X and adjacency matrix A
- Learn the inherent distance metric related to the homophily of the network

Structure Preserving Metric Learning

- Connectivity algorithms: k-nn, b-matching, MST, ϵ-neighborhoods
- Distances are *structure-preserving* if the connectivity algo outputs the true connectivity [SJ09]
- Parameterize distances

$$D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j)$$

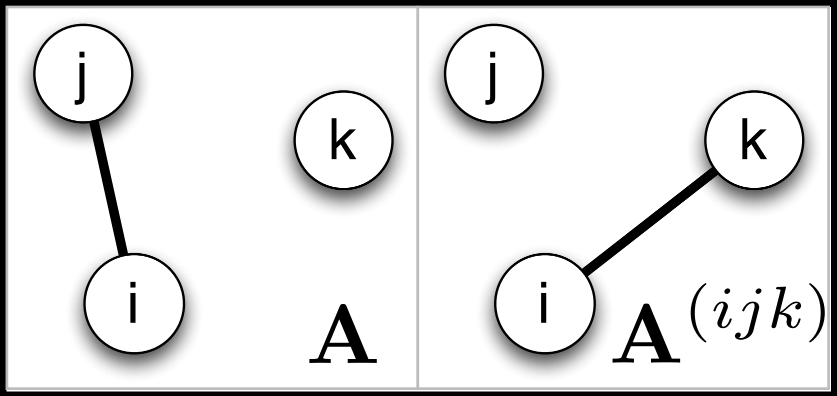
Structured Prediction Motivation

- Constraints: true $\mathbf{A}$ must score higher than any feasible adjacency matrix $\tilde{\mathbf{A}}$
- Frobenius regularization -> SVM
- Cutting plane doesn't scale: requires iterating SDP and separation oracle
- Relaxation:
    - (Optional) drop PSD constraint
    - only consider small changes to $\mathbf{A}$

Stochastic SPML (k-nn)

- Consider only changes along node-neighbor-impostor triplets

$$T = \{(i,j,k) | A_{ij} = 1, A_{ik} = 0\}$$



- Difference between scores is only along triplet edges

$$\frac{\lambda}{2}||\mathbf{M}||_{\mathrm{F}}^2 + \frac{1}{|T|} \sum_{(ijk)\in T} h(D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) - D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_k) + 1)$$

- Randomly sample triplets and follow stochastic subgradients

- (Periodically project to PSD)

Out-of-Sample Extension

- Connectivity algorithm is fixed: structural parameters must be known
  - e.g., k = degree of training nodes
- What degree should new nodes have?
  - Feature-dependent *degree preference functions*

Degree Distributional Metric Learning

- Simultaneously learn feature-dependent *degree preference functions* such that the connectivity algorithm maximizes

$$F(\mathbf{A}|\mathbf{X};\mathbf{M},\mathbf{S}) = -\sum_{ij} A_{ij} D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) + \sum_i g\left(c[i]|\mathbf{x}_i;\mathbf{S}\right)$$

- Linear deg. pref. score $g(k|\mathbf{x};\mathbf{S}) = \sum_{k'=1}^{k} \mathbf{x}^\top \mathbf{s}_{k'}$

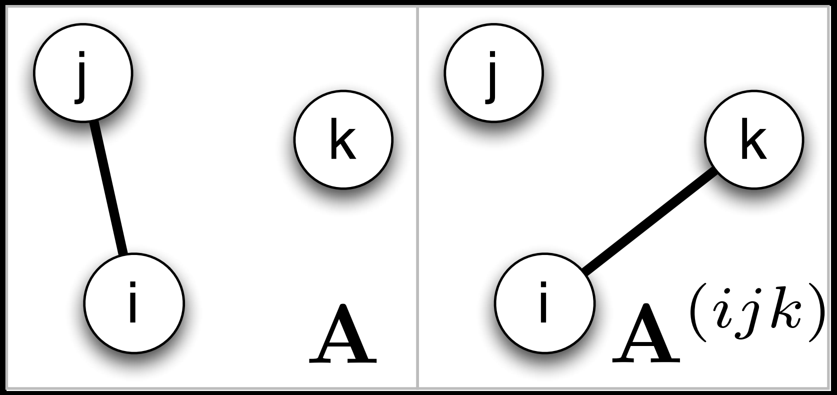
- Regularizing w/ $||\mathbf{S}||_{\mathrm{F}}^2$, DDML can also be a structural SVM

Stochastic DDML

- Triplet-based loss function:

$$\min_{\mathbf{M},\mathbf{S}} \frac{\lambda}{2}(||\mathbf{M}||^2 + ||\mathbf{S}||^2) +$$
$$\frac{1}{|T|} \sum_{ijk \in T} h(F(\mathbf{A}|\mathbf{X};\mathbf{M},\mathbf{S}) - F(\mathbf{A}^{(ijk)}|\mathbf{X};\mathbf{M},\mathbf{S}) + 1)$$

- Score difference cancels except for four quantities:



$$F(\mathbf{A}|\mathbf{X};\mathbf{M},\mathbf{S}) - F(\mathbf{A}^{(ijk)}|\mathbf{X};\mathbf{M},\mathbf{S}) =$$
$$D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_k) - D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{x}_k^\top \mathbf{s}_{(c[k]+1)} - \mathbf{x}_j^\top \mathbf{s}_{(c[j]-1)}$$

- (Project toward concave degree prefs)

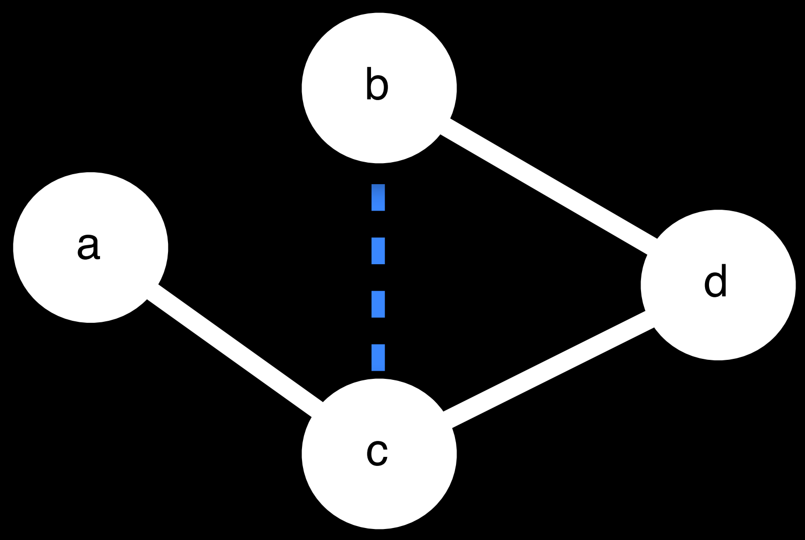
Learner Running Time

- Limit the maximum degree so the number of parameters for degree preference function is constant
- Subgradient computation: $O(d^2)$ for SPML and $O(d^2 + c_{\max})$ for DDML
- Learner reduces to PEGASOS algorithm (Shalev-Shwartz et al. ’07) on one-class SVM:

$O\left(\frac{1}{\epsilon\lambda}\right)-$ time for ϵ-convergence

Link Prediction from DDML

- For concave degree preferences, the connectivity algorithm reduces to an $O(N^3)$ combinatorial algorithm (Huang & Jebara, '09)
- Or rank edges by degree-augmented distance



$$D_{\mathbf{M}}(b, c)$$

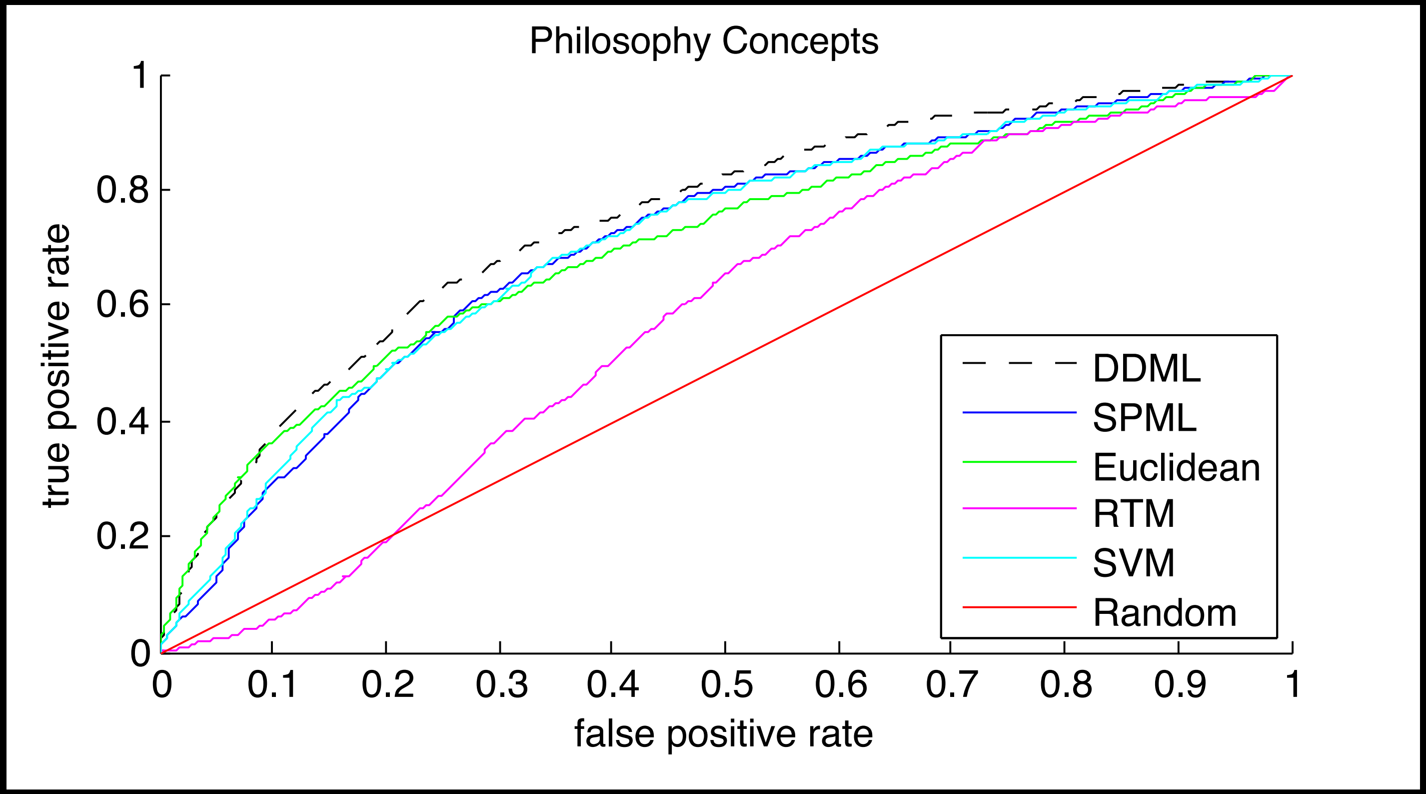
$$g(b, 2) - g(b, 1)$$

$$g(c, 3) - g(c, 2)$$

Experiments

- Wikipedia: word counts and hyperlinks
- Facebook: status, gender, major, dorm, year, and friendship links [Traub, et al., '11]
- Randomly hold out 20% of nodes and incident edges for testing

Experiment Results



AUC

		n	*m*	*d*	Euclid.	RTM	SVM	SPML	DDML
Wikipedia	Graph Theory	223	917	6695	0.624	0.591	0.610	**0.722**	0.691*
	Philosophy Concepts	303	921	6695	0.705	0.571	0.708	0.707	**0.746***
	Search Engines	269	332	6695	0.662	0.487	0.611	**0.742**	0.725*
	Philosophy Crawl	100k	4m	7702	0.547	–	–	**0.601**	0.562
Facebook	Harvard	1937	48k	193	0.764	0.562	0.839	**0.854**	0.848
	MIT	2128	95k	173	0.702	0.494	0.784	**0.801**	0.797
	Stanford	3014	147k	270	0.718	0.532	0.784	0.808	**0.810**
	Columbia	3050	118k	251	0.717	0.519	0.796	0.818	**0.821**

Summary + Open Problems + Thanks!

- SPML: metrics consistent with structural behavior of networks
- DDML: explicit degree preference functions that are feature-dependent
- Linear constraints and Frobenius regularization: **constant time convergence**

- Other applications of structure-preserving metric
- More natural regularizer?
- Stopping criterion
- Large-scale prediction