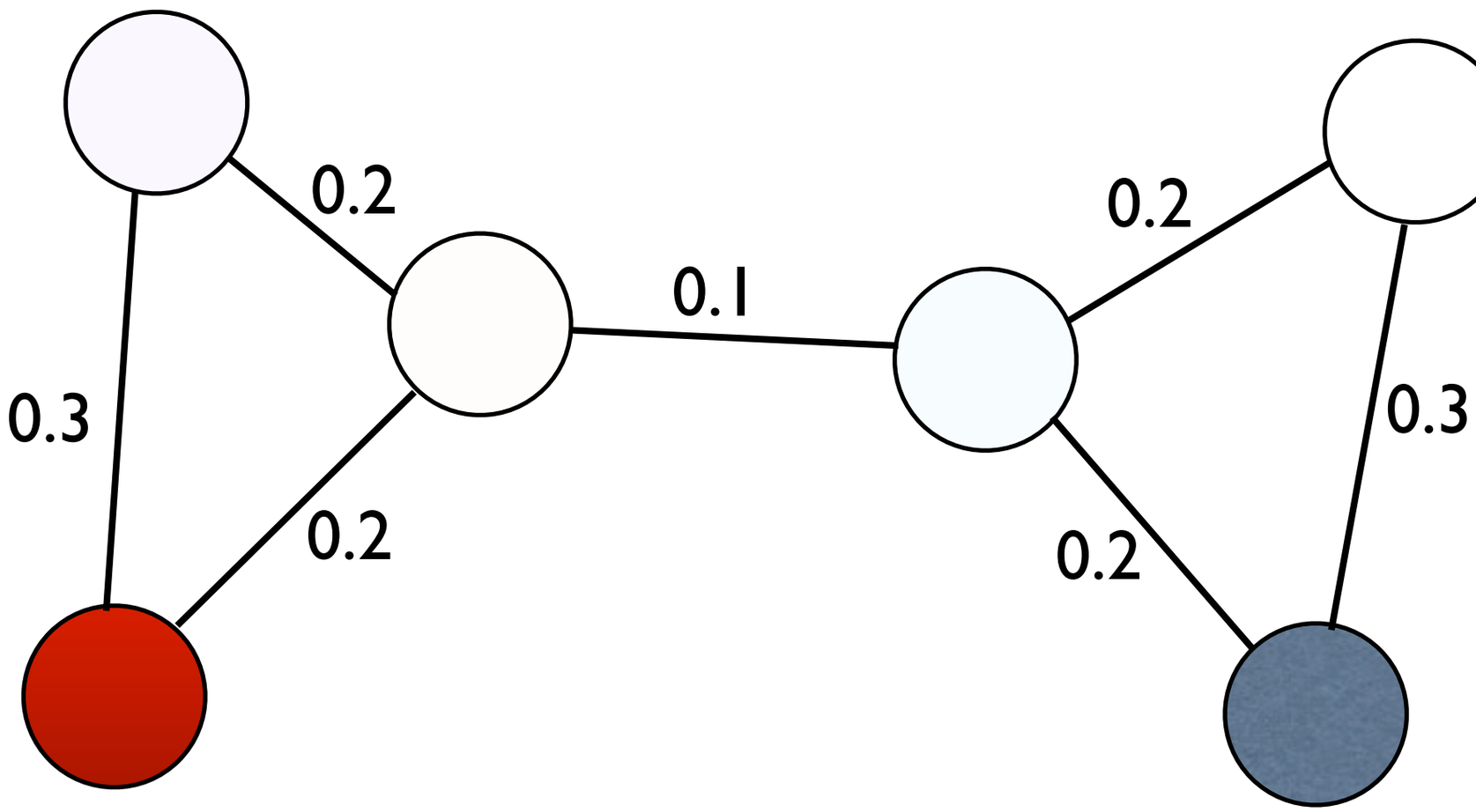
Graph-based Semi-Supervised Learning as Optimization

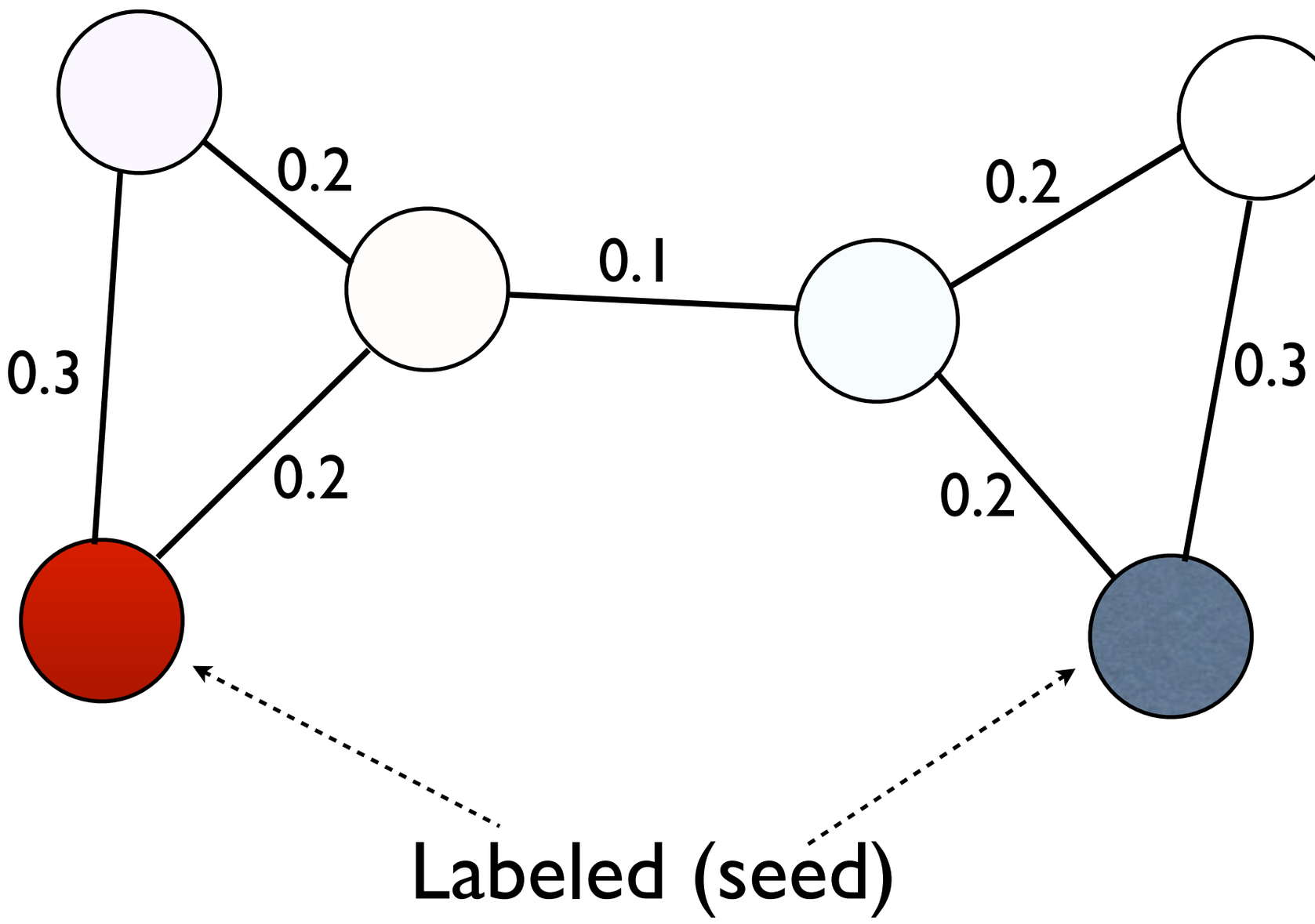
Partha Pratim Talukdar
CMU

Machine Learning with Large Datasets (10-605)
April 3, 2012

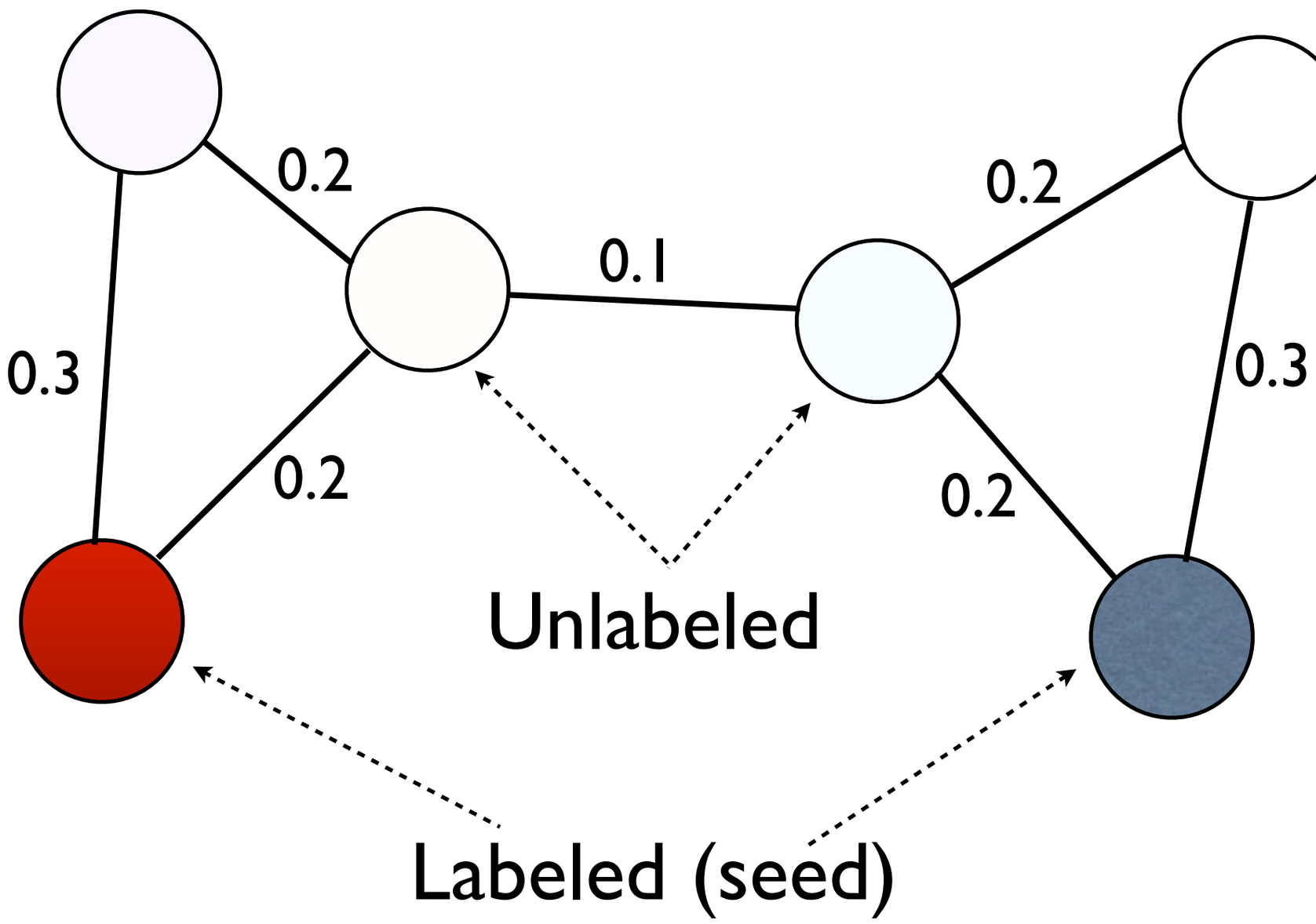
Graph-based Semi-Supervised Learning



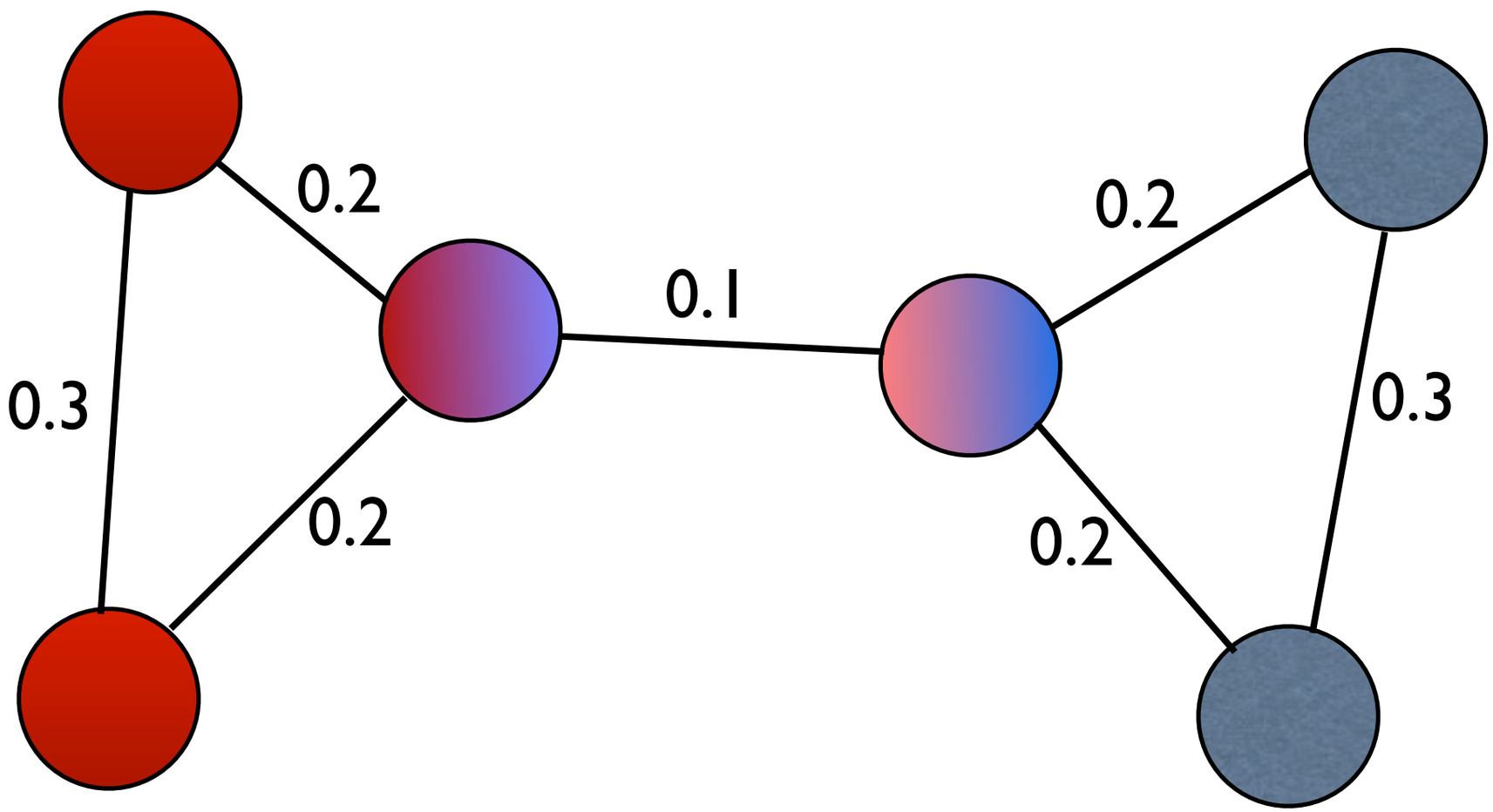
Graph-based Semi-Supervised Learning



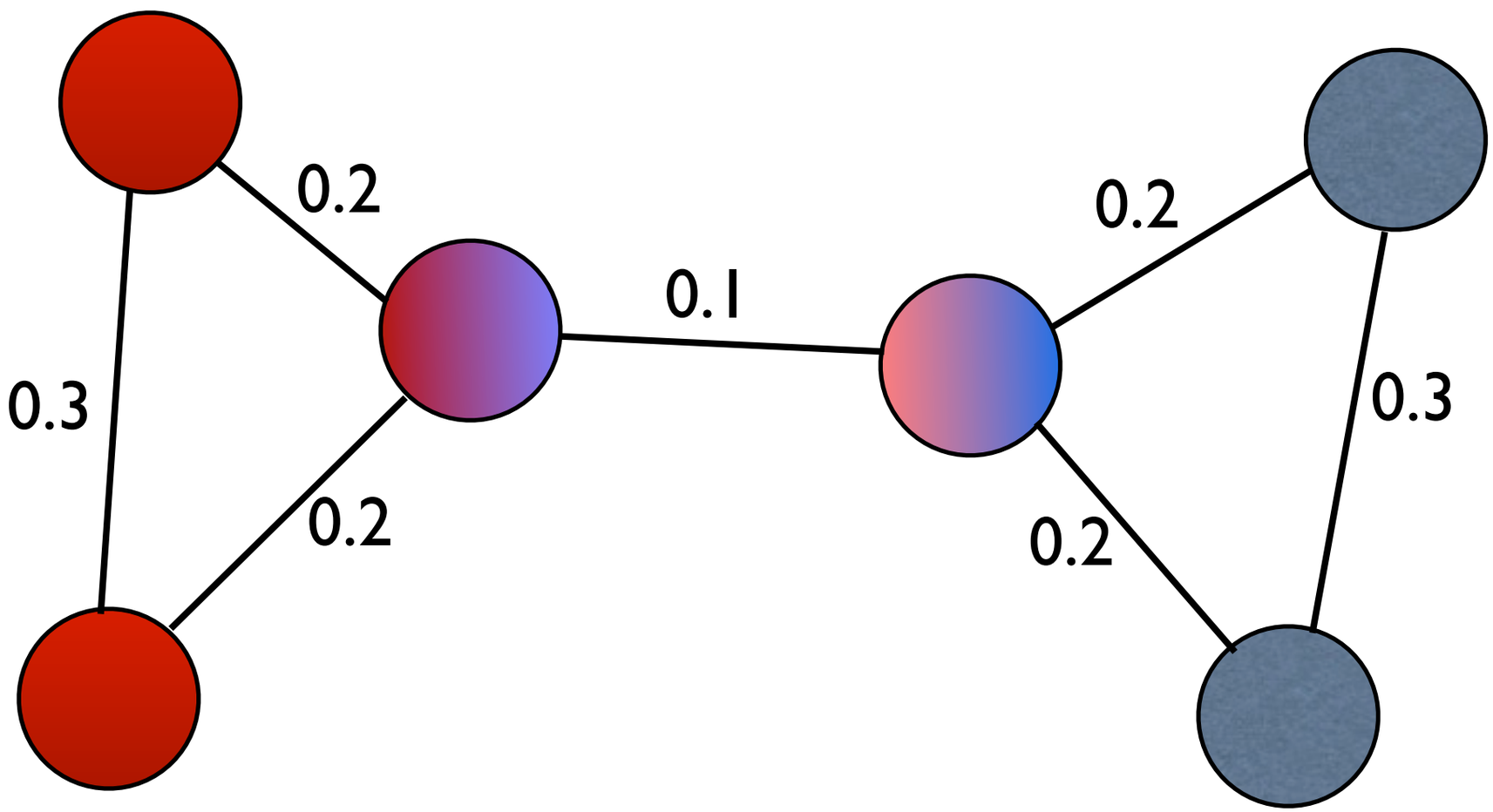
Graph-based Semi-Supervised Learning



Graph-based Semi-Supervised Learning

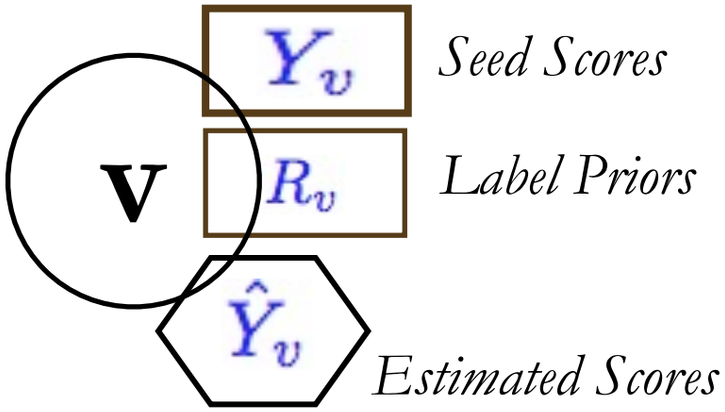


Graph-based Semi-Supervised Learning



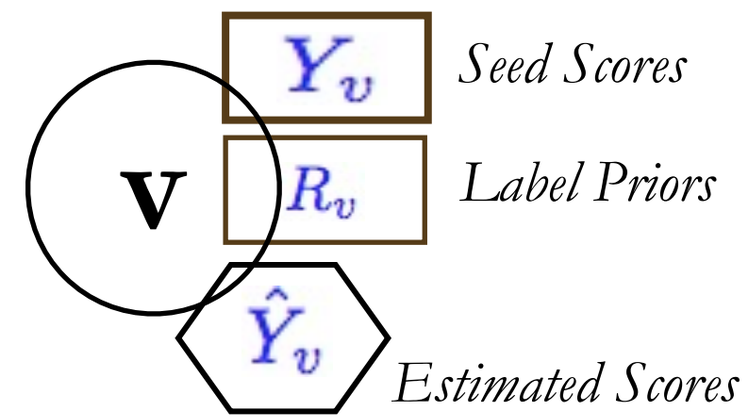
Various methods:
LP (Zhu et al., 2003); QC (Bengio et al., 2007); Adsorption (Baluja et al., 2008)

Notations



Notations

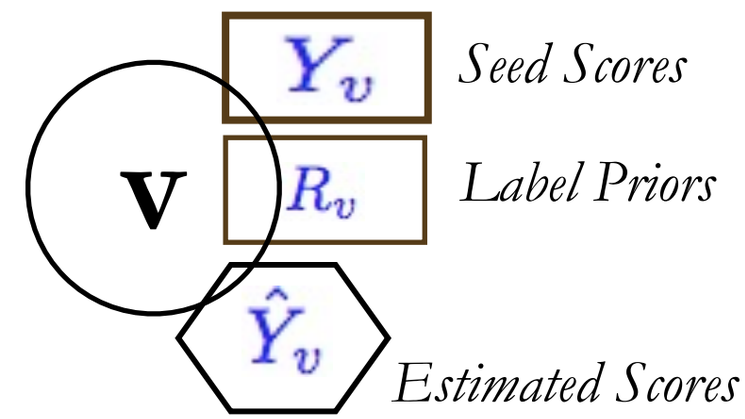
$\hat{Y}_{v,l}$: score of estimated label l on node v



Notations

$\hat{Y}_{v,l}$: score of estimated label l on node v

$Y_{v,l}$: score of seed label l on node v

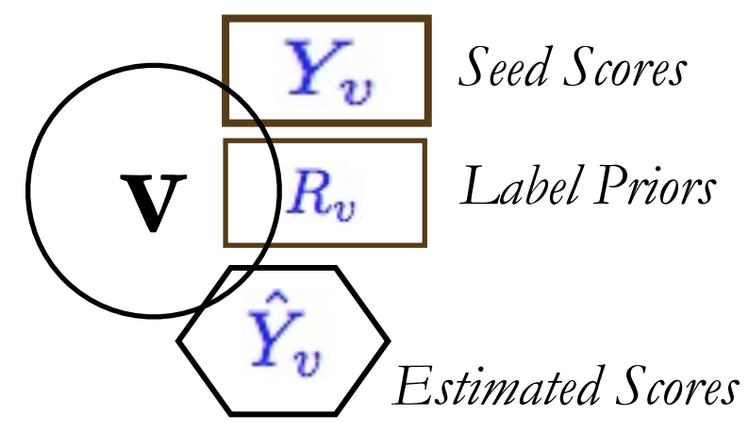


Notations

$\hat{Y}_{v,l}$: score of estimated label l on node v

$Y_{v,l}$: score of seed label l on node v

$R_{v,l}$: regularization target for label l on node v



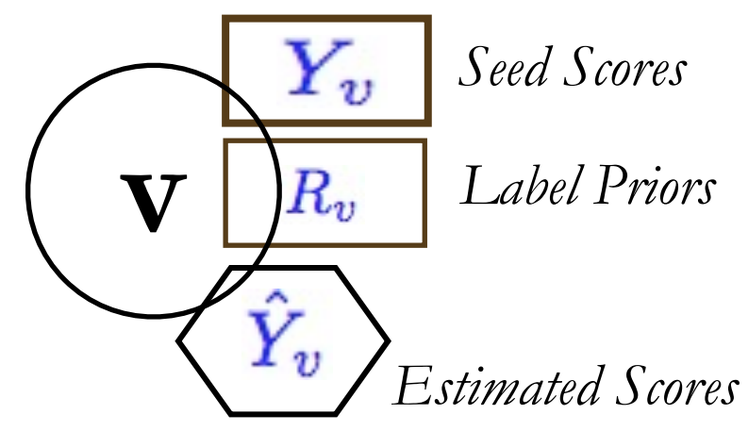
Notations

$\hat{Y}_{v,l}$: score of estimated label l on node v

$Y_{v,l}$: score of seed label l on node v

$R_{v,l}$: regularization target for label l on node v

S : seed node indicator (diagonal matrix)



Notations

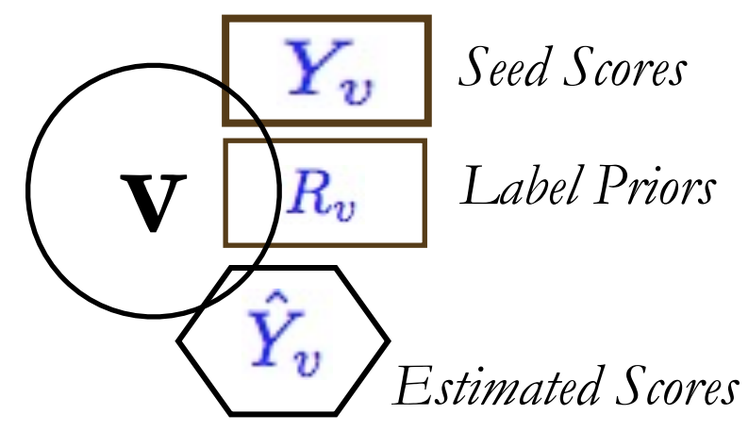
$\hat{Y}_{v,l}$: score of estimated label l on node v

$Y_{v,l}$: score of seed label l on node v

$R_{v,l}$: regularization target for label l on node v

S : seed node indicator (diagonal matrix)

W_{uv} : weight of edge (u, v) in the graph



LP-ZGL (Zhu et al., ICML 2003)

LP-ZGL (Zhu et al., ICML 2003)

$$\arg\min_{\hat{Y}} \sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2$$

such that $Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1$

LP-ZGL (Zhu et al., ICML 2003)

Smooth

$$\arg\min_{\hat{Y}} \boxed{\sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2}$$

such that $Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1$

LP-ZGL (Zhu et al., ICML 2003)

Smooth

$$\arg\min_{\hat{Y}} \boxed{\sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2}$$

such that $\boxed{Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1}$

Match Seeds (hard)

LP-ZGL (Zhu et al., ICML 2003)

Smooth

$$\arg\min_{\hat{Y}} \boxed{\sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2} = \sum_{l=1}^{m} \hat{Y}_l^T L \hat{Y}_l$$

Graph Laplacian
L = D - W (PSD)

such that $\boxed{Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1}$

Match Seeds (hard)

LP-ZGL (Zhu et al., ICML 2003)

Smooth

$$\arg\min_{\hat{Y}} \boxed{\sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2} = \sum_{l=1}^{m} \hat{Y}_l^T L \hat{Y}_l$$

Graph Laplacian
L = D - W (PSD)

such that $\boxed{Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1}$

Match Seeds (hard)

- Smoothness
  - two nodes connected by an edge with high weight should be assigned similar labels

LP-ZGL (Zhu et al., ICML 2003)

Smooth

$$\arg\min_{\hat{Y}} \boxed{\sum_{l=1}^{m} W_{uv}(\hat{Y}_{ul} - \hat{Y}_{vl})^2} = \sum_{l=1}^{m} \hat{Y}_l^T L \hat{Y}_l$$

Graph Laplacian
L = D - W (PSD)

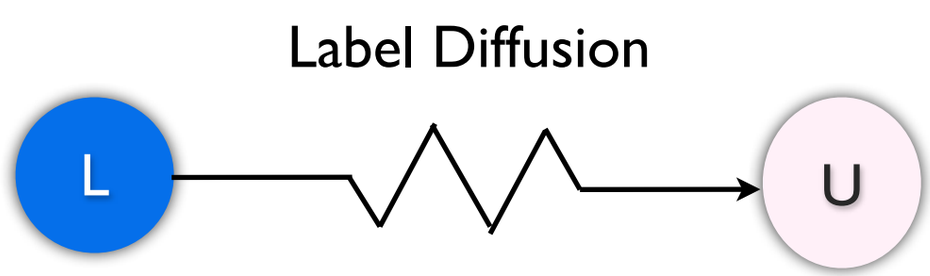
such that $\boxed{Y_{ul} = \hat{Y}_{ul}, \ \forall S_{uu} = 1}$

Match Seeds (hard)

- Smoothness
  - two nodes connected by an edge with high weight should be assigned similar labels
- Solution satisfies harmonic property

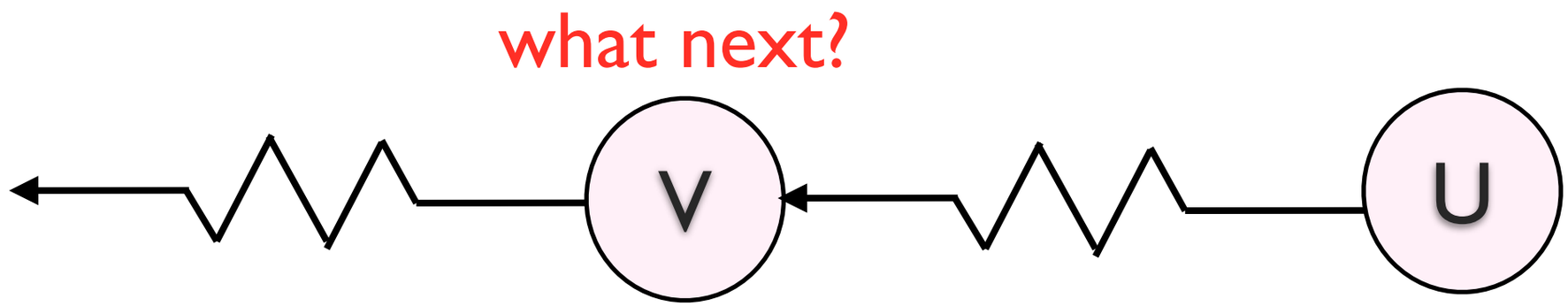
Two Related Views

Two Related Views

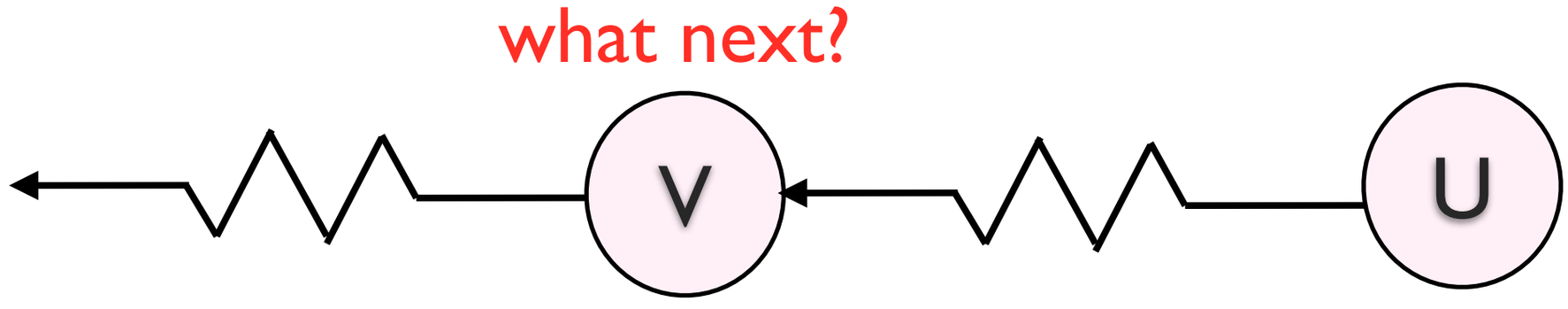


Random Walk View

Random Walk View



Random Walk View



- Continue walk with prob. p_v^{cont}
- Assign V's seed label to U with prob. p_v^{inj}
- Abandon random walk with prob. p_v^{abnd}
  - assign U a dummy label

Discounting Nodes

Discounting Nodes

- Certain nodes can be unreliable (e.g., high degree nodes)
  - do not allow propagation/walk through them

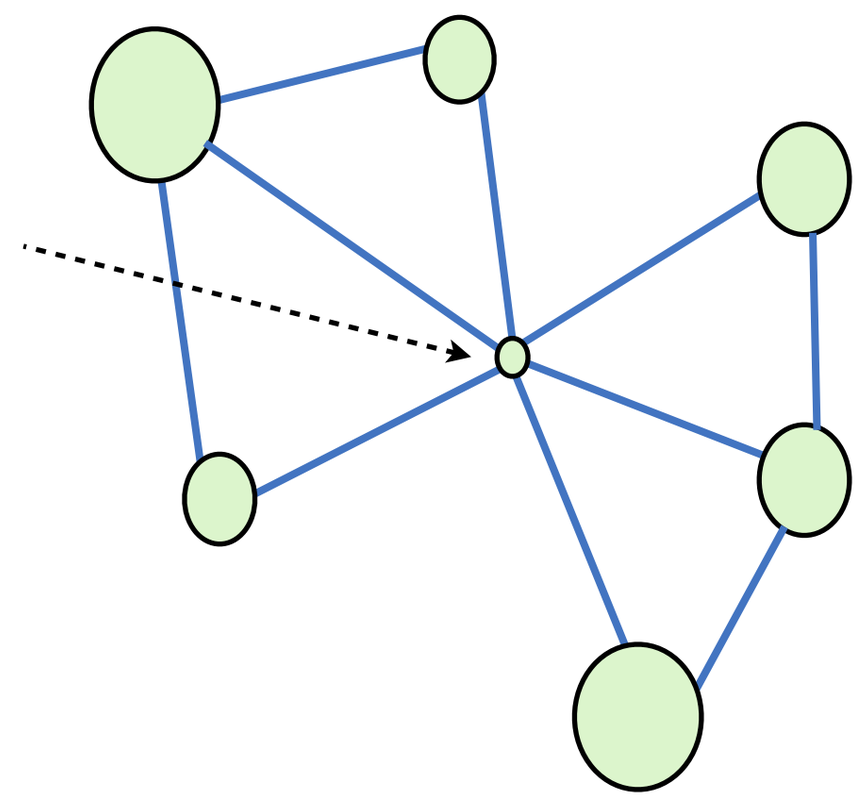
Discounting Nodes

- Certain nodes can be unreliable (e.g., high degree nodes)
  - do not allow propagation/walk through them
- Solution: increase abandon probability on such nodes:

Discounting Nodes

- Certain nodes can be unreliable (e.g., high degree nodes)
  - do not allow propagation/walk through them
- Solution: increase abandon probability on such nodes:

$$\mathbf{p_v^{abnd}} \propto \mathrm{degree(v)}$$



Redefining Matrices

Redefining Matrices

$$W'_{uv} = p_u^{cont} \times W_{uv}$$

New Edge Weight

Redefining Matrices

$$W'_{uv} = p_u^{cont} \times W_{uv}$$

New Edge Weight

$$S_{uu} = \sqrt{p_u^{inj}}$$

Redefining Matrices

$$W'_{uv} = p_u^{cont} \times W_{uv}$$

New Edge Weight

$$S_{uu} = \sqrt{p_u^{inj}}$$

$$R_{u\top} = p_u^{abnd}$$, and 0 for non-dummy labels

Dummy Label

Modified Adsorption (MAD)

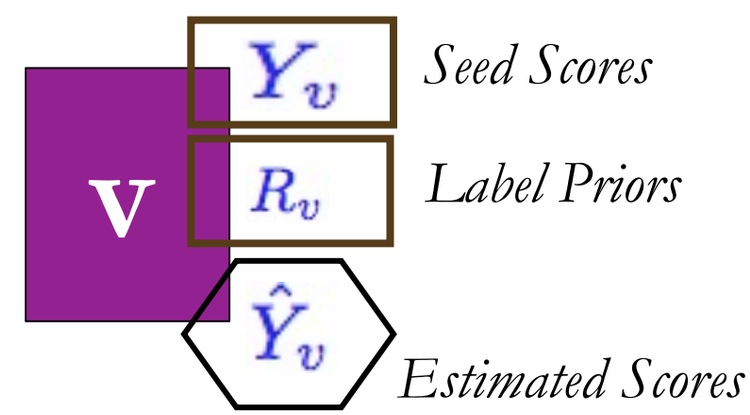
[Talukdar and Crammer, ECML 2009]

Modified Adsorption (MAD)

[Talukdar and Crammer, ECML 2009]

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2 + \mu_1 \sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2 + \mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2 \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{W}^{\dagger\top} + \boldsymbol{W}'$ is the symmetrized weight matrix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v



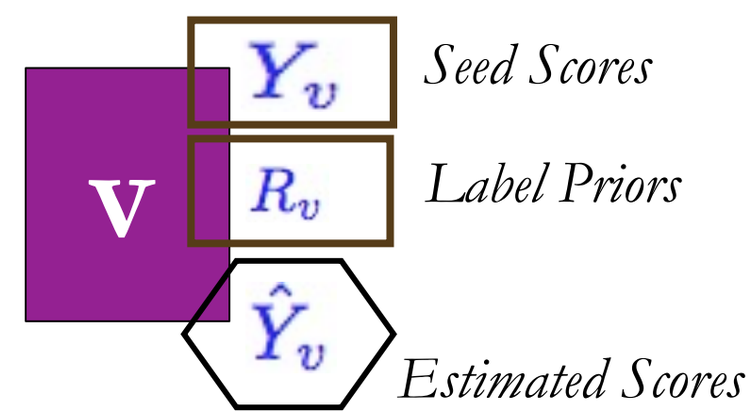
Modified Adsorption (MAD)

[Talukdar and Crammer, ECML 2009]

Match Seeds (soft)

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\boxed{\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2} + \mu_1 \sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2 + \mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2 \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{W}^{\dagger\top} + \boldsymbol{W}'$ is the symmetrized weight matrix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v



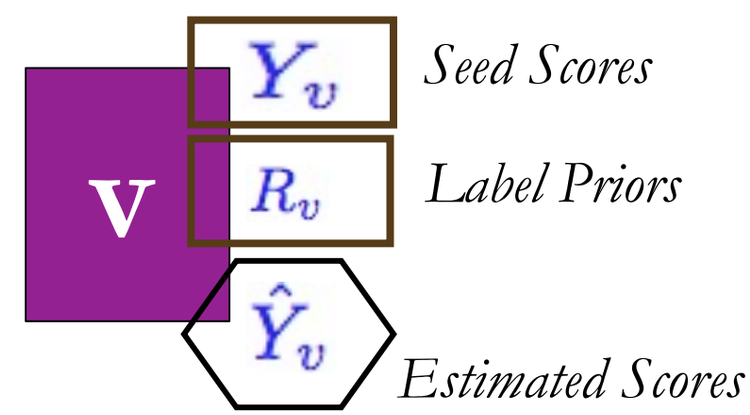
Modified Adsorption (MAD)

[Talukdar and Crammer, ECML 2009]

Match Seeds (soft) Smooth

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\boxed{\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2} + \mu_1 \boxed{\sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2} + \mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2 \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{W}^{\dagger\top} + \boldsymbol{W}'$ is the symmetrized weight matrix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v



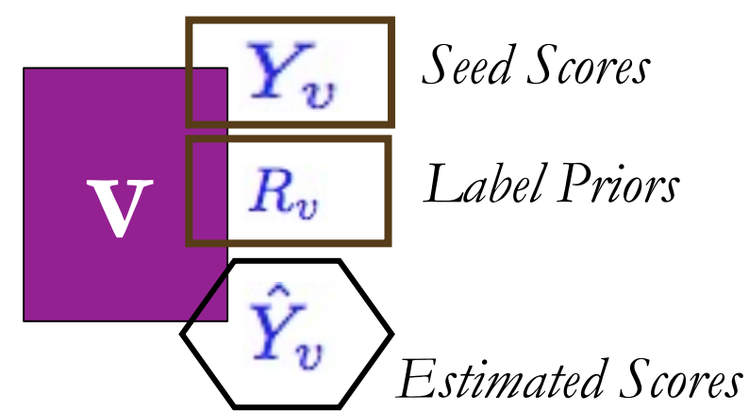
Modified Adsorption (MAD)

[Talukdar and Crammer, ECML 2009]

Match Seeds (soft) Smooth Match Priors (Regularizer)

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\boxed{\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2} + \mu_1 \boxed{\sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2} + \boxed{\mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2} \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{W}^{\dagger\top} + \boldsymbol{W}'$ is the symmetrized weight matrix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v



Modified Adsorption (MAD)

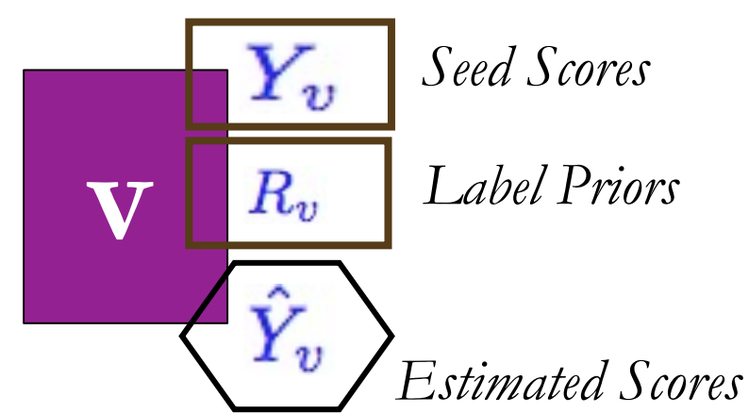
[Talukdar and Crammer, ECML 2009]

Match Seeds (soft) | Smooth | Match Priors (Regularizer)

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\boxed{\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2} + \mu_1 \boxed{\sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2} + \boxed{\mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2} \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{V}$ … trix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v

used to indicate *none-of-the-above* label



Modified Adsorption (MAD)

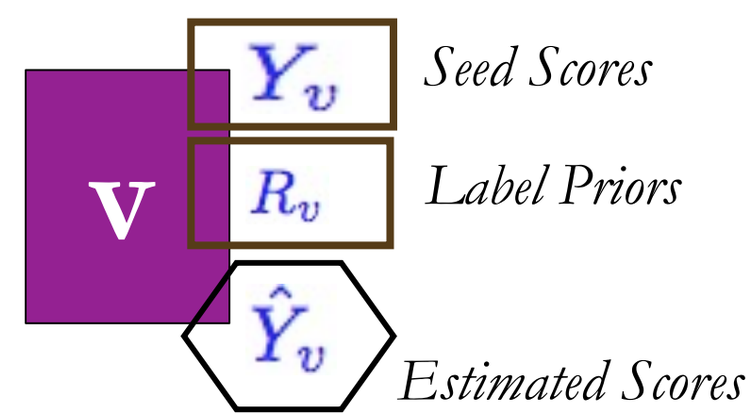
[Talukdar and Crammer, ECML 2009]

Match Seeds (soft) — Smooth — Match Priors (Regularizer)

$$\arg\min_{\hat{\boldsymbol{Y}}} \sum_{l=1}^{m+1} \left[\|\boldsymbol{S}\hat{\boldsymbol{Y}}_l - \boldsymbol{S}\boldsymbol{Y}_l\|^2 + \mu_1 \sum_{u,v} \boldsymbol{M}_{uv}(\hat{\boldsymbol{Y}}_{ul} - \hat{\boldsymbol{Y}}_{vl})^2 + \mu_2 \|\hat{\boldsymbol{Y}}_l - \boldsymbol{R}_l\|^2 \right]$$

- m labels, $+1$ dummy label
- $\boldsymbol{M} = \boldsymbol{V}$ trix
- $\hat{\boldsymbol{Y}}_{vl}$: weight of label l on node v
- $\boldsymbol{Y}_{vl}$: seed weight for label l on node v
- $\boldsymbol{S}$: diagonal matrix, nonzero for seed nodes
- $\boldsymbol{R}_{vl}$: regularization target for label l on node v

used to indicate *none-of-the-above* label



MAD has extra regularization compared to LP-ZGL (Zhu et al, ICML 03); similar to QC (Bengio et al, 2006)

MAD (contd.)

Inputs $\boldsymbol{Y}, \boldsymbol{R} : |V| \times (|L|+1)$, $\boldsymbol{W} : |V| \times |V|$, $\boldsymbol{S} : |V| \times |V|$ diagonal

$\hat{\boldsymbol{Y}} \leftarrow \boldsymbol{Y}$

$\boldsymbol{M} = \boldsymbol{W}' + \boldsymbol{W}'^{\dagger}$

$Z_v \leftarrow \boldsymbol{S}_{vv} + \mu_1 \sum_{u \neq v} \boldsymbol{M}_{vu} + \mu_2 \quad \forall v \in V$

repeat

  for all $v \in V$ **do**

    $\hat{\boldsymbol{Y}}_v \leftarrow \frac{1}{Z_v} \left((\boldsymbol{S}\boldsymbol{Y})_v + \mu_1 \boldsymbol{M}_{v\cdot} \hat{\boldsymbol{Y}} + \mu_2 \boldsymbol{R}_v \right)$

  end for

until convergence

MAD (contd.)

Inputs $\boldsymbol{Y}, \boldsymbol{R} : |V| \times (|L|+1)$, $\boldsymbol{W} : |V| \times |V|$, $\boldsymbol{S} : |V| \times |V|$ diagonal

$\hat{\boldsymbol{Y}} \leftarrow \boldsymbol{Y}$

$\boldsymbol{M} = \boldsymbol{W}' + \boldsymbol{W}'^{\dagger}$

$Z_v \leftarrow \boldsymbol{S}_{vv} + \mu_1 \sum_{u \neq v} \boldsymbol{M}_{vu} + \mu_2 \quad \forall v \in V$

repeat

  for all $v \in V$ **do**

    $\hat{\boldsymbol{Y}}_v \leftarrow \frac{1}{Z_v} \left((\boldsymbol{S}\boldsymbol{Y})_v + \mu_1 \boldsymbol{M}_{v\cdot} \hat{\boldsymbol{Y}} + \mu_2 \boldsymbol{R}_v \right)$

  end for

until convergence

- Extends Adsorption with well-defined optimization
- Importance of a node can be discounted
- Easily Parallelizable: Scalable

MapReduce Implementation of MAD

MapReduce Implementation of MAD

- Map
  - Each node send its current label assignments to its neighbors

MapReduce Implementation of MAD

- Map
  - Each node send its current label assignments to its neighbors
- Reduce
  - Each node updates its own label assignment using messages received from neighbors, and its own information (e.g., seed labels, reg. penalties etc.)

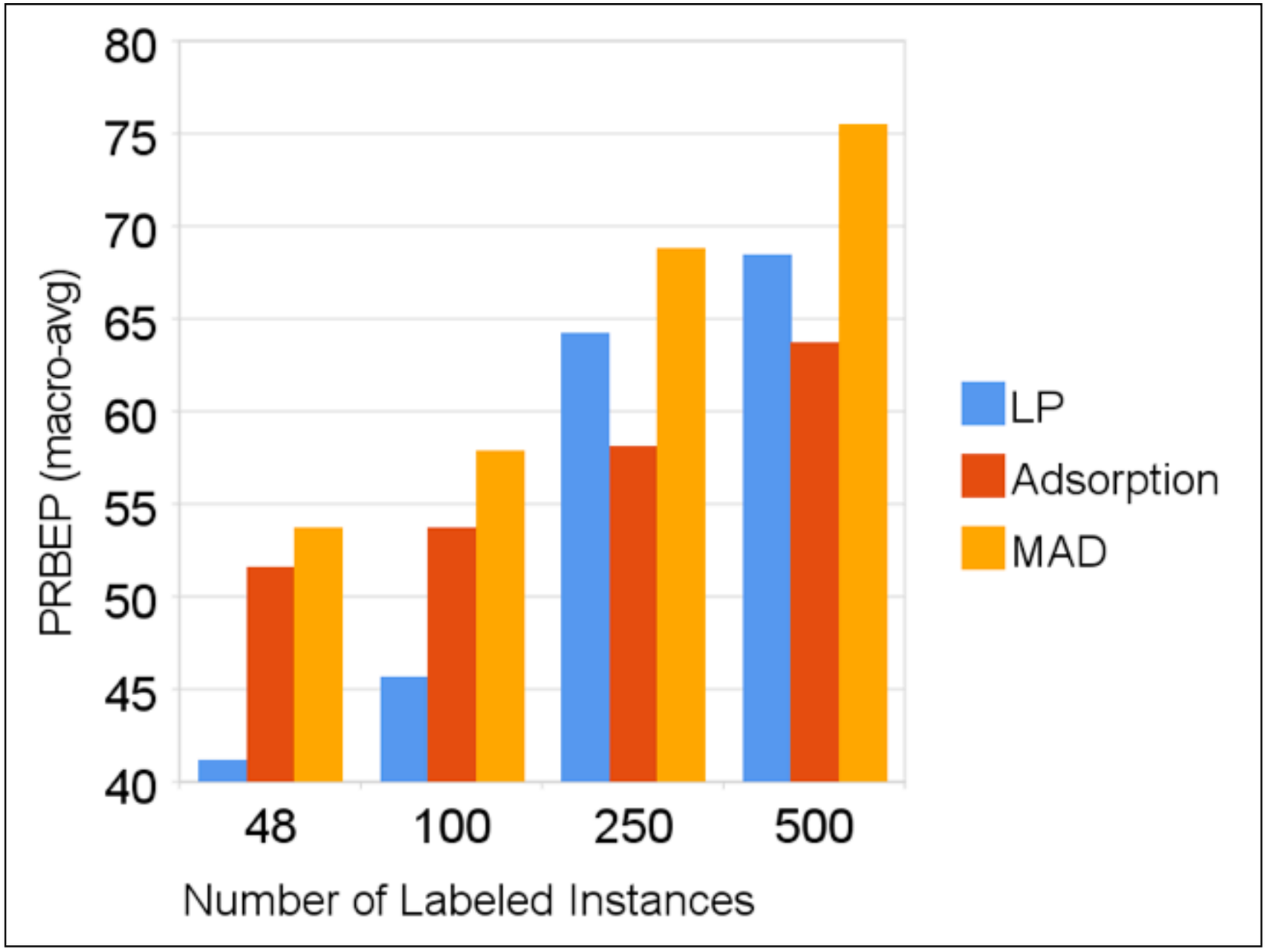
MapReduce Implementation of MAD

- Map
  - Each node send its current label assignments to its neighbors
- Reduce
  - Each node updates its own label assignment using messages received from neighbors, and its own information (e.g., seed labels, reg. penalties etc.)
- Repeat until convergence

Experiments

Text Classification

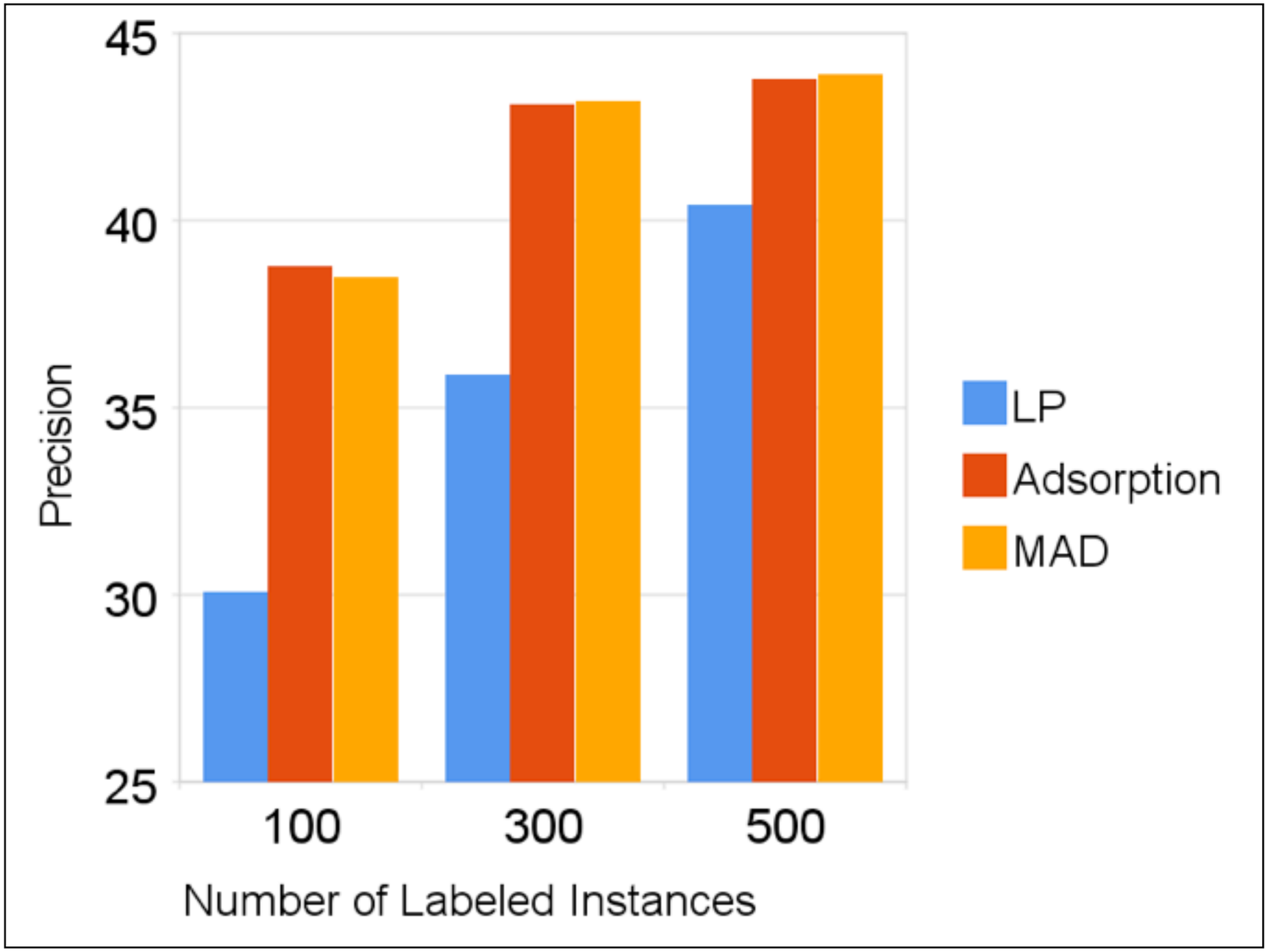
Text Classification



PRBEP (macro-averaged) on WebKB Dataset, 3148 test instances

Sentiment Classification

Sentiment Classification



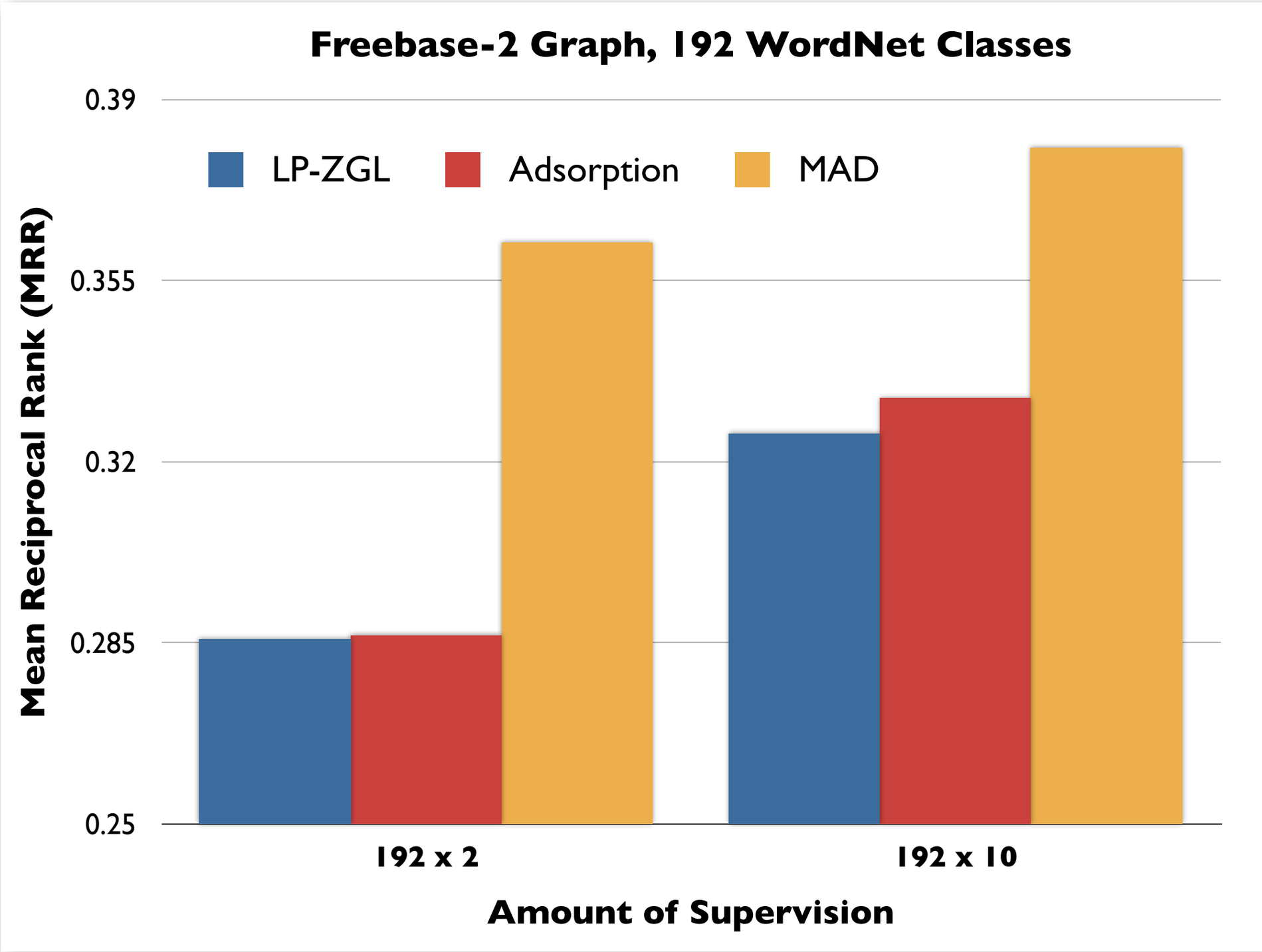
Precision on 3568 Sentiment test instances

Class-Instance Acquisition

Class-Instance Acquisition

Graph with 303k nodes, 2.3m edges.

Class-Instance Acquisition



Graph with 303k nodes, 2.3m edges.

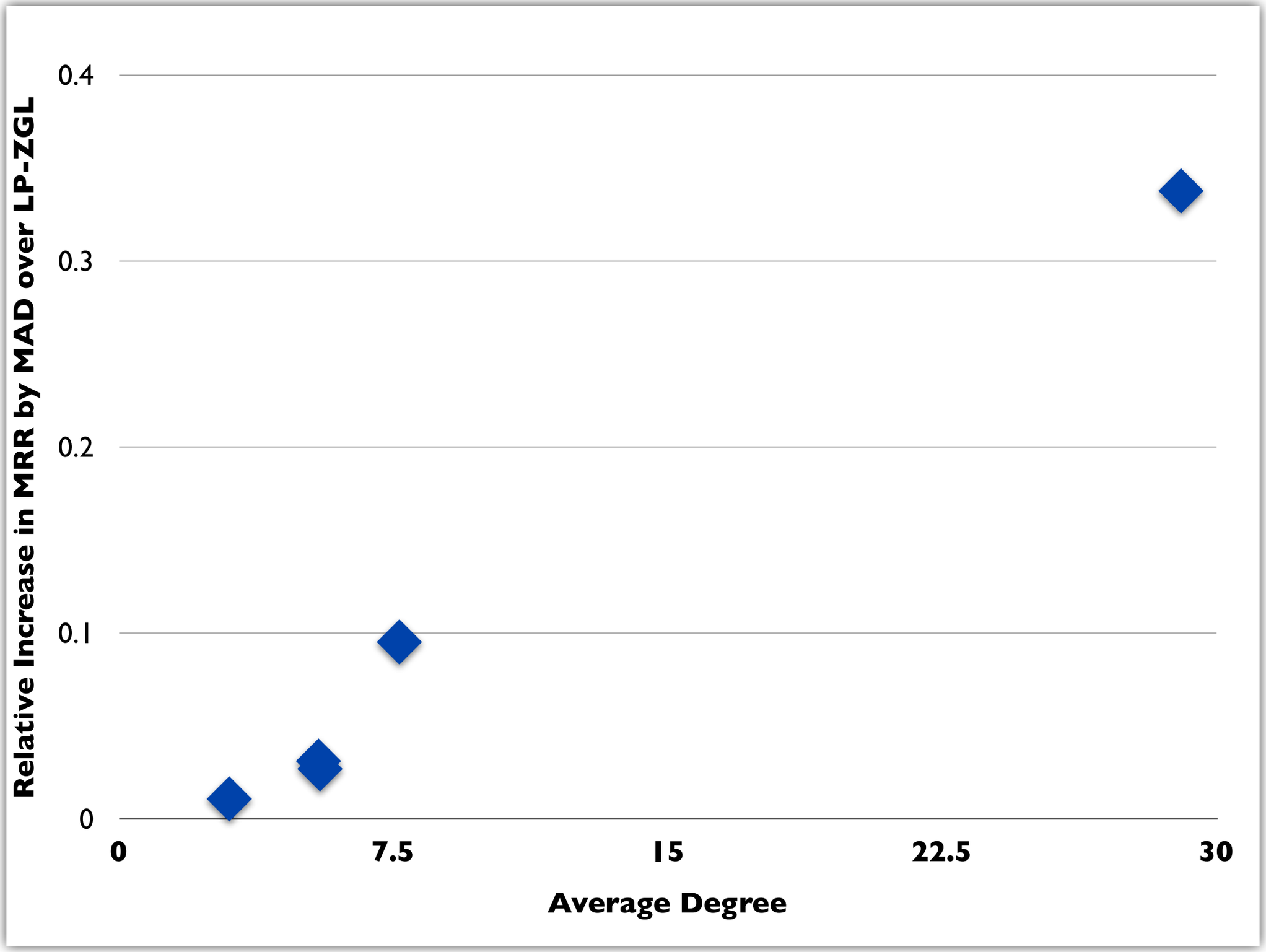
New (Class, Instance) Pairs Found

Class	**A few non-seed Instances found by Adsorption**
Scientific Journals	Journal of Physics, Nature, Structural and Molecular Biology, Sciences Sociales et sante, Kidney and Blood Pressure Research, American Journal of Physiology-Cell Physiology, …
NFL Players	Tony Gonzales, Thabiti Davis, Taylor Stubblefield, Ron Dixon, Rodney Hannan, …
Book Publishers	Small Night Shade Books, House of Ansari Press, Highwater Books, Distributed Art Publishers, Cooper Canyon Press, …

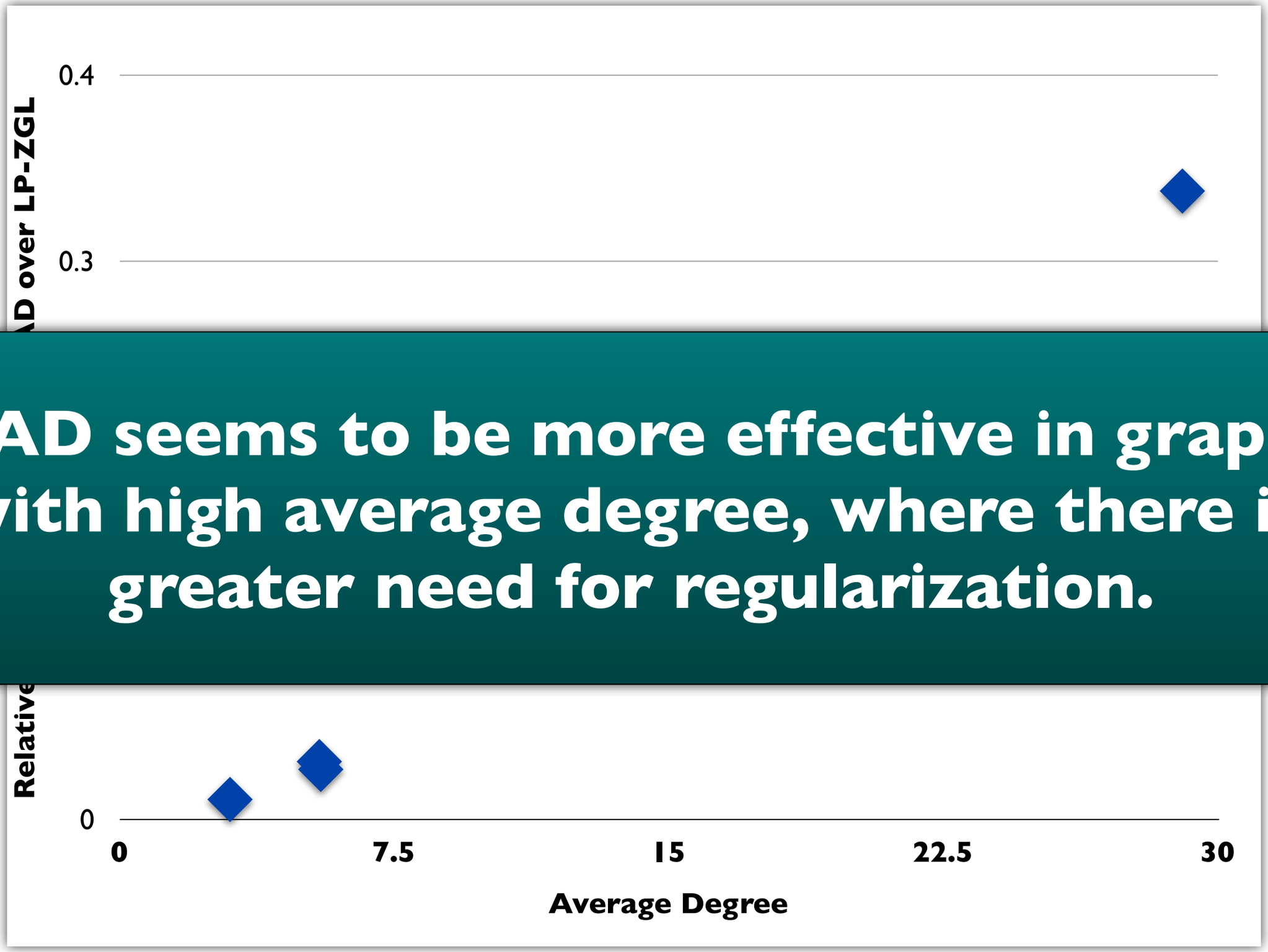
Total classes: **9081**

When is MAD most effective?

When is MAD most effective?



When is MAD most effective?



MAD seems to be more effective in graphs with high average degree, where there is greater need for regularization.

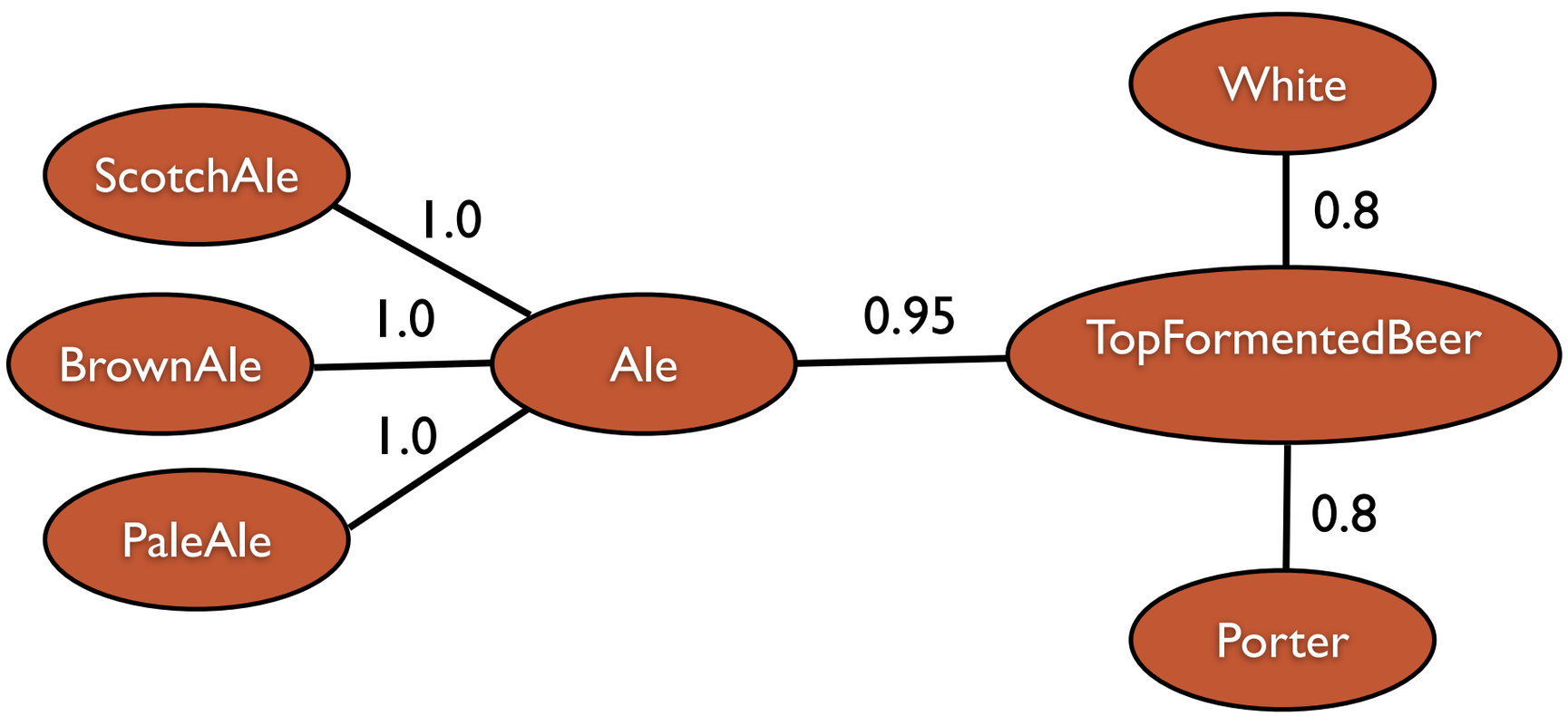
Benefits of Having an Optimization Framework

Extension to Dependent Labels

- Labels are not always mutually exclusive.

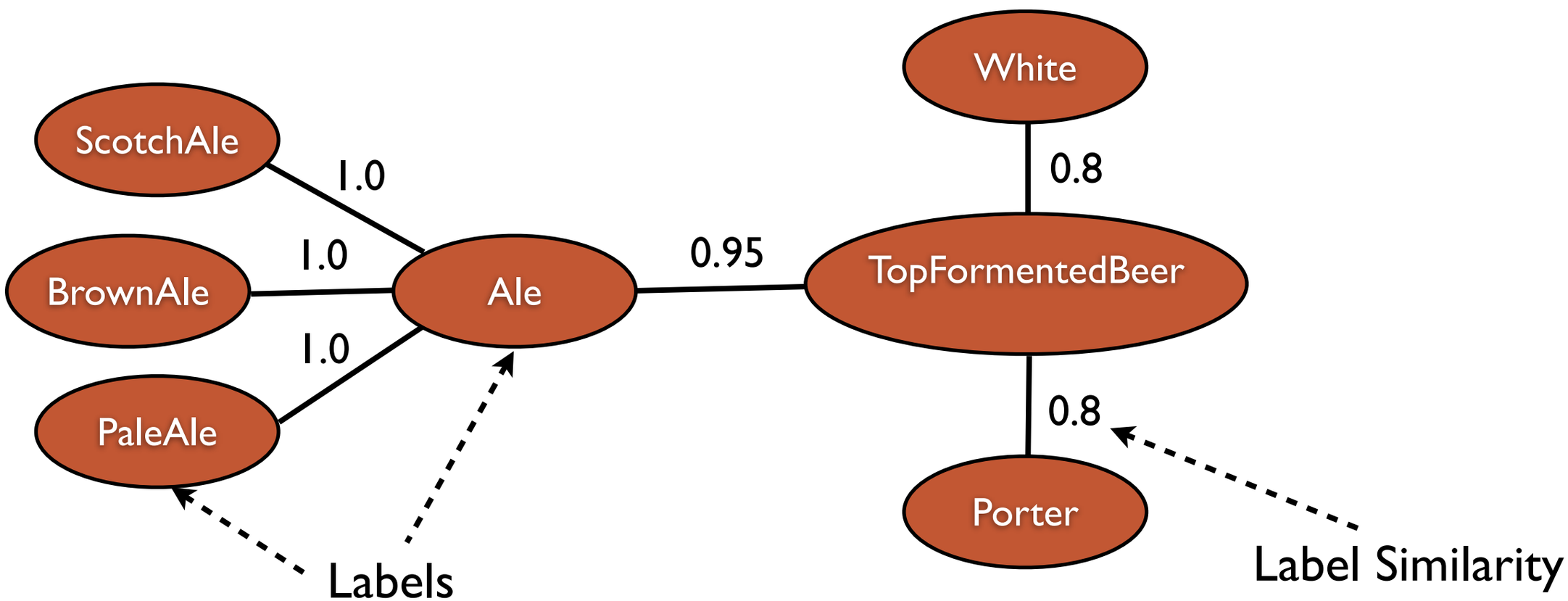
Extension to Dependent Labels

- Labels are not always mutually exclusive.



Extension to Dependent Labels

- Labels are not always mutually exclusive.

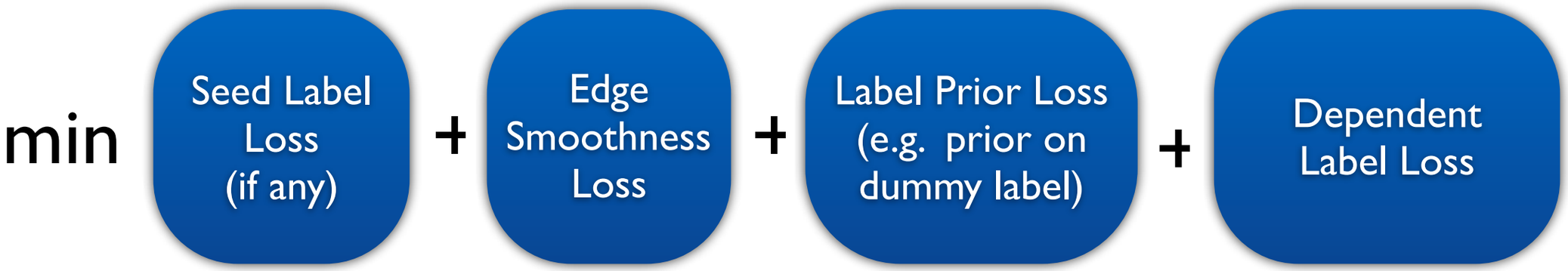


Label Graph

MAD with Dependent Labels (MADDL)

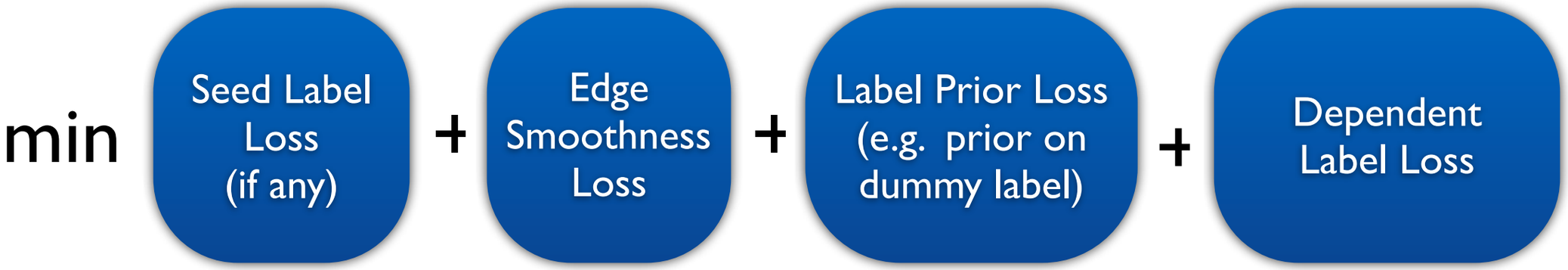
min Seed Label Loss (if any) + Edge Smoothness Loss + Label Prior Loss (e.g. prior on dummy label)

MAD with Dependent Labels (MADDL)



MAD with Dependent Labels (MADDL)

MADDL Objective



MAD with Dependent Labels (MADDL)

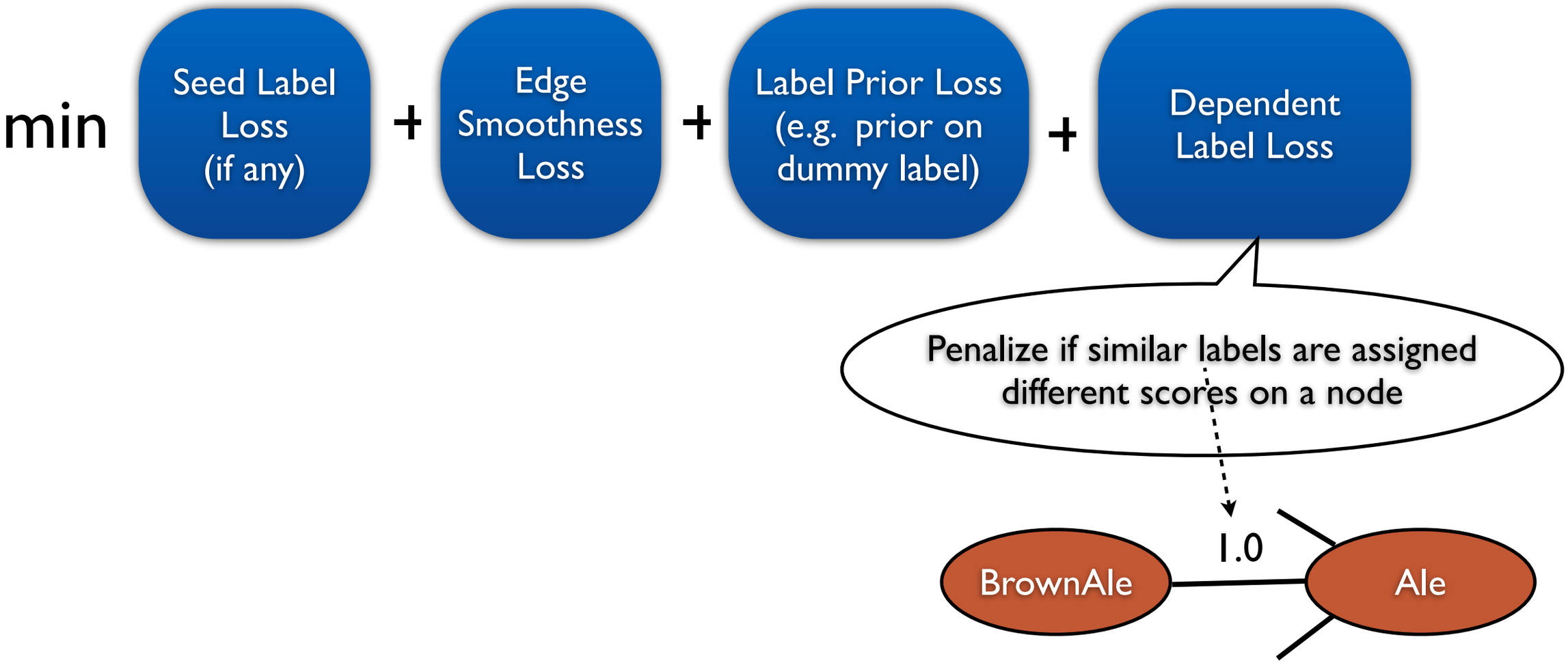
MADDL Objective

min [Seed Label Loss (if any)] + [Edge Smoothness Loss] + [Label Prior Loss (e.g. prior on dummy label)] + [Dependent Label Loss]

Penalize if similar labels are assigned different scores on a node

MAD with Dependent Labels (MADDL)

MADDL Objective

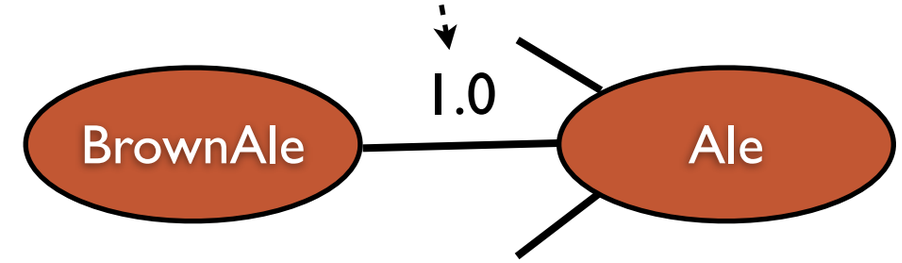


MAD with Dependent Labels (MADDL)

MADDL Objective

min [Seed Label Loss (if any)] + [Edge Smoothness Loss] + [Label Prior Loss (e.g. prior on dummy label)] + [Dependent Label Loss]

Penalize if similar labels are assigned different scores on a node



MADDL objective results in a scalable iterative update, with convergence guarantee.

Smooth Sentiment Ranking

smooth predictions

Smooth Sentiment Ranking

smooth predictions

non-smooth predictions

Smooth Sentiment Ranking

Prefer

smooth predictions

over

non-smooth predictions

Smooth Sentiment Ranking

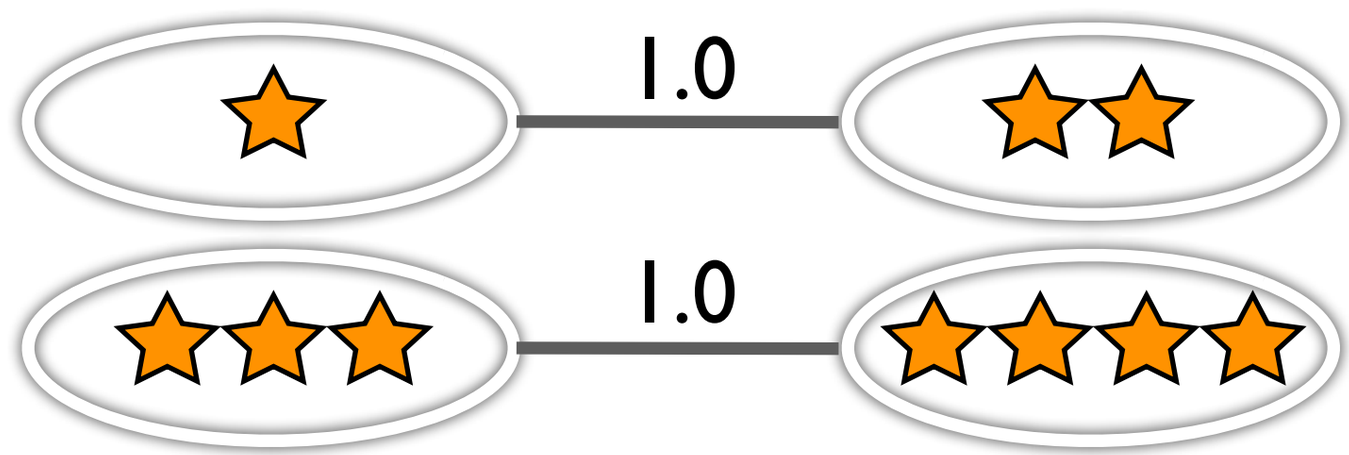
Prefer

smooth predictions

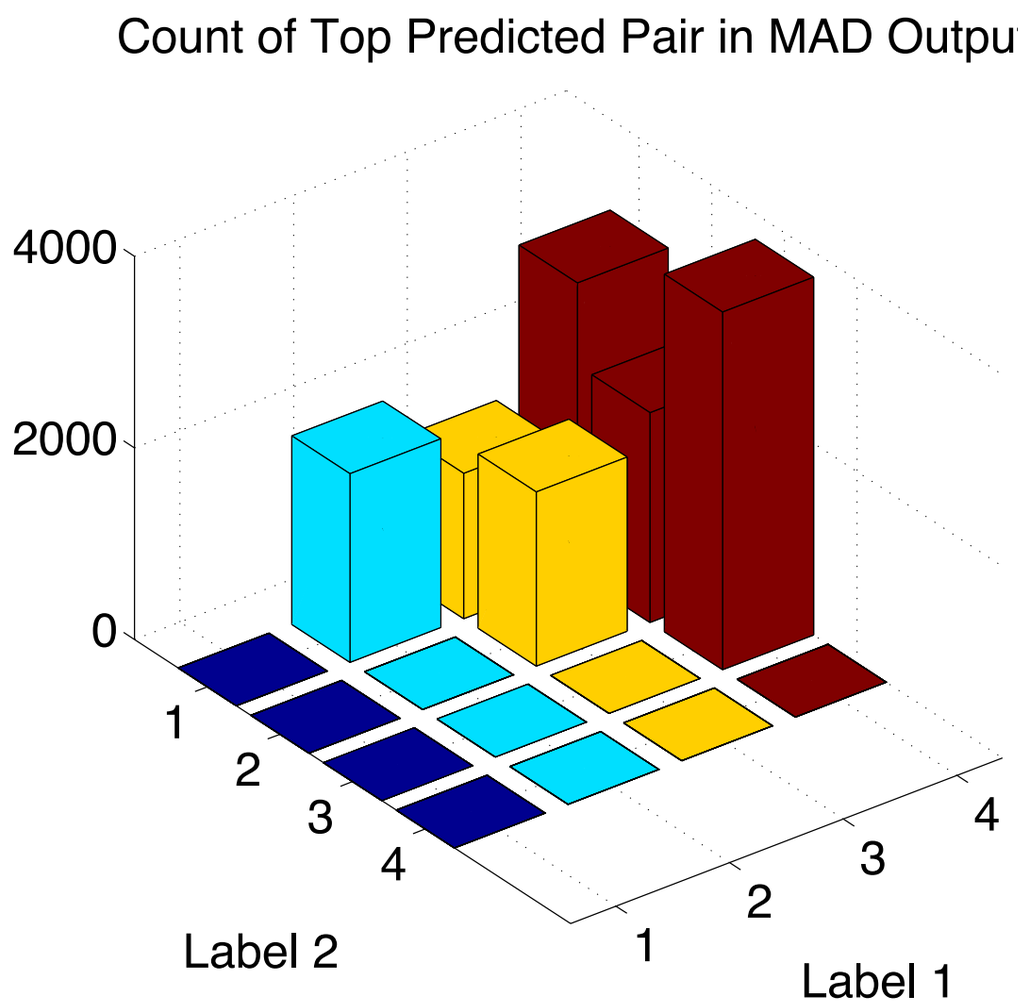
over

non-smooth predictions

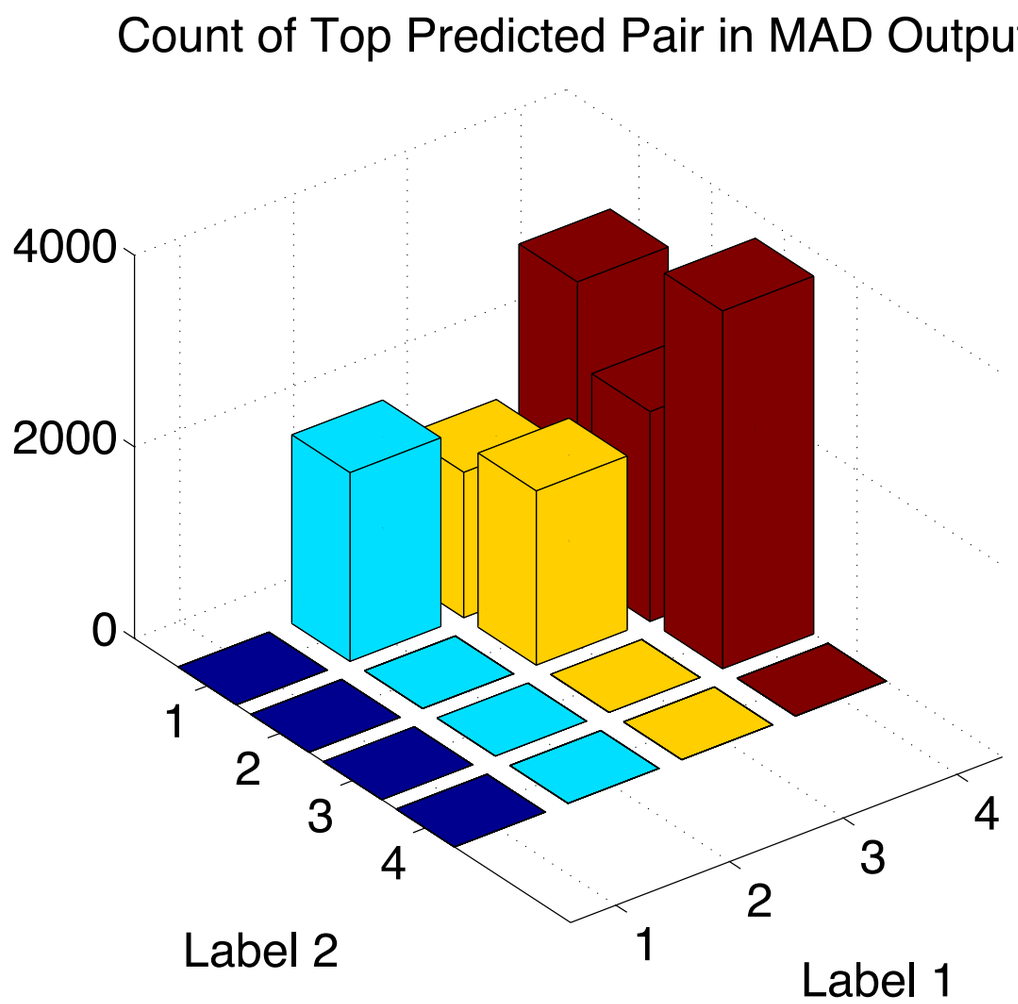
MADDL Label Constraints

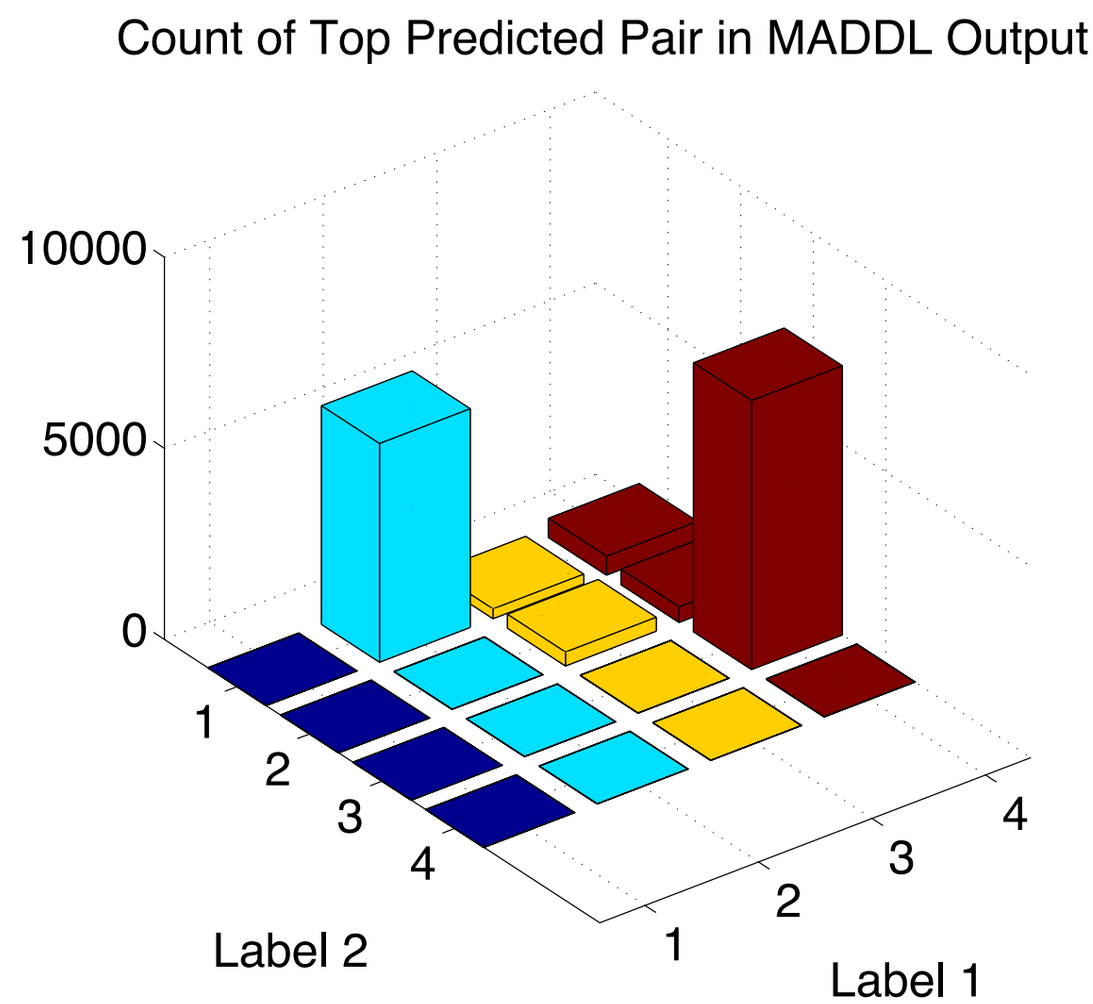


Smooth Sentiment Ranking

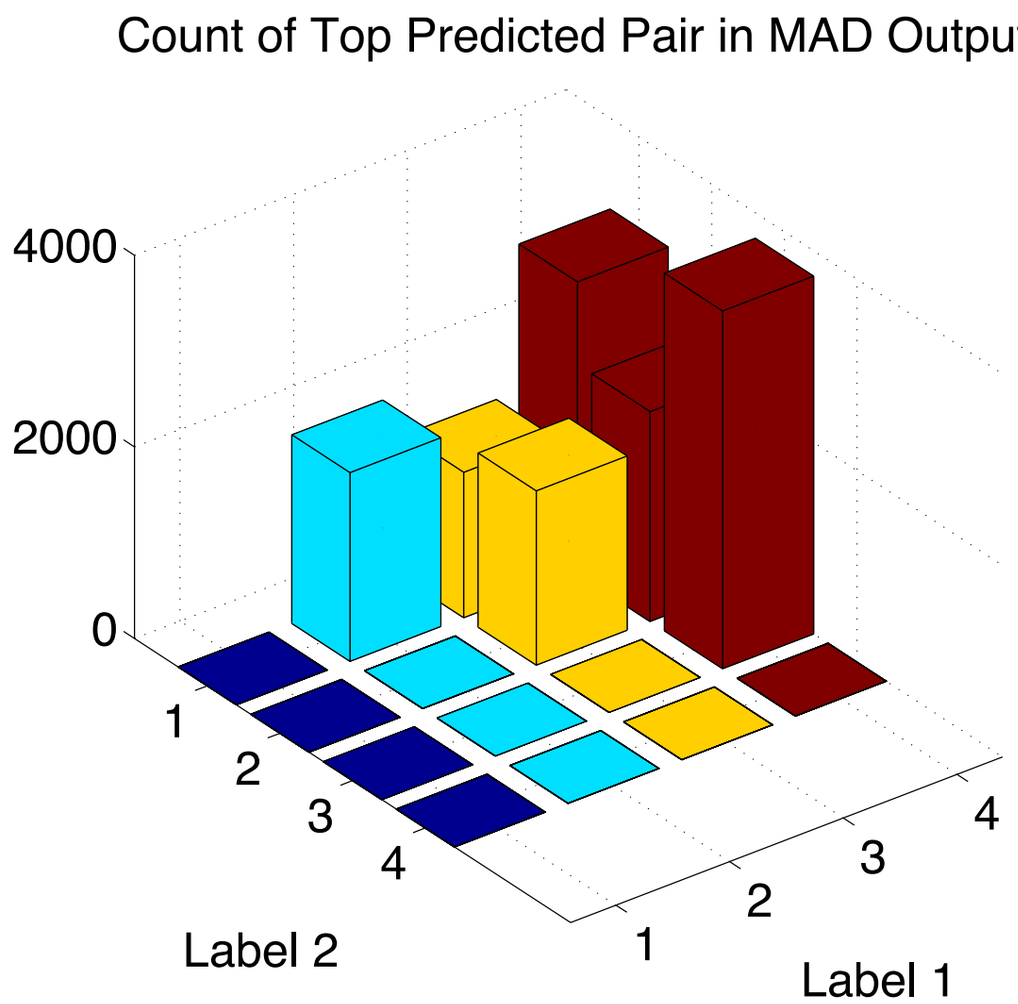


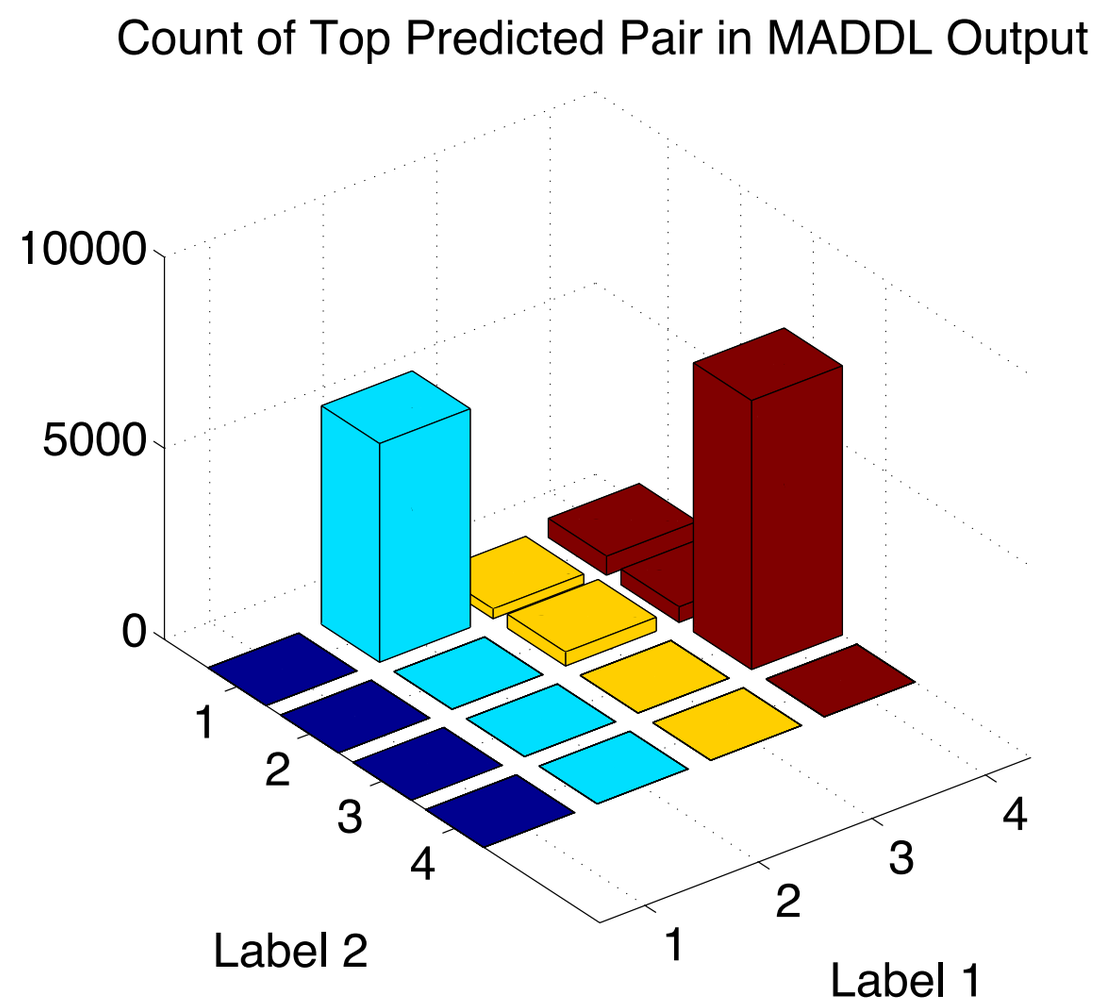
Smooth Sentiment Ranking





Smooth Sentiment Ranking





MADDL generates smoother ranking, while preserving accuracy of prediction.

Other Methods and Graph Construction

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]
  - KL-based loss

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]
  - KL-based loss
  - Results over 120m node graph

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]
  - KL-based loss
  - Results over 120m node graph
- Chapter 21 of the SSL book [Chappelle et al.]

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]
  - KL-based loss
  - Results over 120m node graph
- Chapter 21 of the SSL book [Chappelle et al.]
- Graph Construction

Other Methods and Graph Construction

- Measure Propagation [Subramanya and Bilmes, NIPS 2009]
  - KL-based loss
  - Results over 120m node graph
- Chapter 21 of the SSL book [Chappelle et al.]
- Graph Construction
  - Task-specific Graph Construction using Metric Learning [Dhillon, Talukdar, and Crammer, ACL 2010]

Graph-based SSL: Final Thoughts

Graph-based SSL: Final Thoughts

- Provide flexible representation
  - for both relational and IID data

Graph-based SSL: Final Thoughts

- Provide flexible representation
  - for both relational and IID data
- Scalable: MR/Hadoop-based implementation

Graph-based SSL: Final Thoughts

- Provide flexible representation
  - for both relational and IID data
- Scalable: MR/Hadoop-based implementation
- Can handle labeled as well as unlabeled data

Graph-based SSL: Final Thoughts

- Provide flexible representation
  - for both relational and IID data
- Scalable: MR/Hadoop-based implementation
- Can handle labeled as well as unlabeled data
- Can handle multiple-classes

Graph-based SSL: Final Thoughts

- Provide flexible representation
  - for both relational and IID data
- Scalable: MR/Hadoop-based implementation
- Can handle labeled as well as unlabeled data
- Can handle multiple-classes
- Effective in practice

Code in Junto Label Propagation Toolkit

(includes Hadoop-based implementation)

http://code.google.com/p/junto/

Code in Junto Label Propagation Toolkit
(includes Hadoop-based implementation)
http://code.google.com/p/junto/

Thanks!