Seeing stars when there aren't many stars

Graph-based semi-supervised learning for sentiment categorization

Andrew B. Goldberg, Xiaojin “Jerry” Zhu
Computer Sciences Department
University of Wisconsin-Madison

Goodbye Academia, Hello Hollywood!

Not quite!

Sentiment Categorization

“...captivating... special effects were amazing...” ☆☆☆☆ ?

“...excellent acting... quite good and believable...” ☆☆☆☆ ?

“...weak, lame attempts at humor...bland as can be...” ☆☆☆☆ ?

Sentiment Categorization

"...captivating... special effects were amazing..."

★★★★

"...excellent acting... quite good and believable..."

★★★☆

"...weak, lame attempts at humor...bland as can be..."

★☆☆☆

Graph-Based Learning

- This is rating inference [Pang+Lee05]
- We use graph-based semi-supervised learning
- Main assumption encoded in graph:
  Similar documents should have similar ratings

[Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*.]

Graph-Based Learning

Labeled

"...captivating... special effects were amazing..."

★★★★

"...excellent acting... quite good and believable..."

★★★☆

"...weak, lame attempts at humor... bland as can be..."

★☆☆☆

Unlabeled

"...captivating piece of work from an excellent team..."

"...excellent acting makes up for bland, lame scenery..."

"...preview quite good, but acting was weak, way off..."

"...weak, bad... pathetically lame... acting way off..."

Graph-Based Learning

Labeled

"...captivating... special effects were amazing..."

★★★★

"...excellent acting... quite good and believable..."

★★★☆

"...weak, lame attempts at humor... bland as can be..."

★☆☆☆

Unlabeled — **Truth**

"...captivating piece of work from an excellent team..." — ★★★★

"...excellent acting makes up for bland, lame scenery..." — ★★★☆

"...preview quite good, but acting was weak, way off..." — ★★☆☆

"...weak, bad... pathetically lame... acting way off..." — ★☆☆☆

Graph-Based Learning

Labeled	Unlabeled	Truth	Supervised
"...captivating... special effects were amazing..." ★★★★	"...captivating piece of work from an excellent team..."	★★★★	★★★★ ✓
"...excellent acting... quite good and believable..." ★★★☆	"...excellent acting makes up for bland, lame scenery..."	★★★☆	
"...weak, lame attempts at humor... bland as can be..." ★☆☆☆	"...preview quite good, but acting was weak, way off..."	★★☆☆	
	"...weak, bad... pathetically lame... acting way off..."	★☆☆☆	

Graph-Based Learning

Labeled

"...captivating... special effects were amazing..."
★★★★

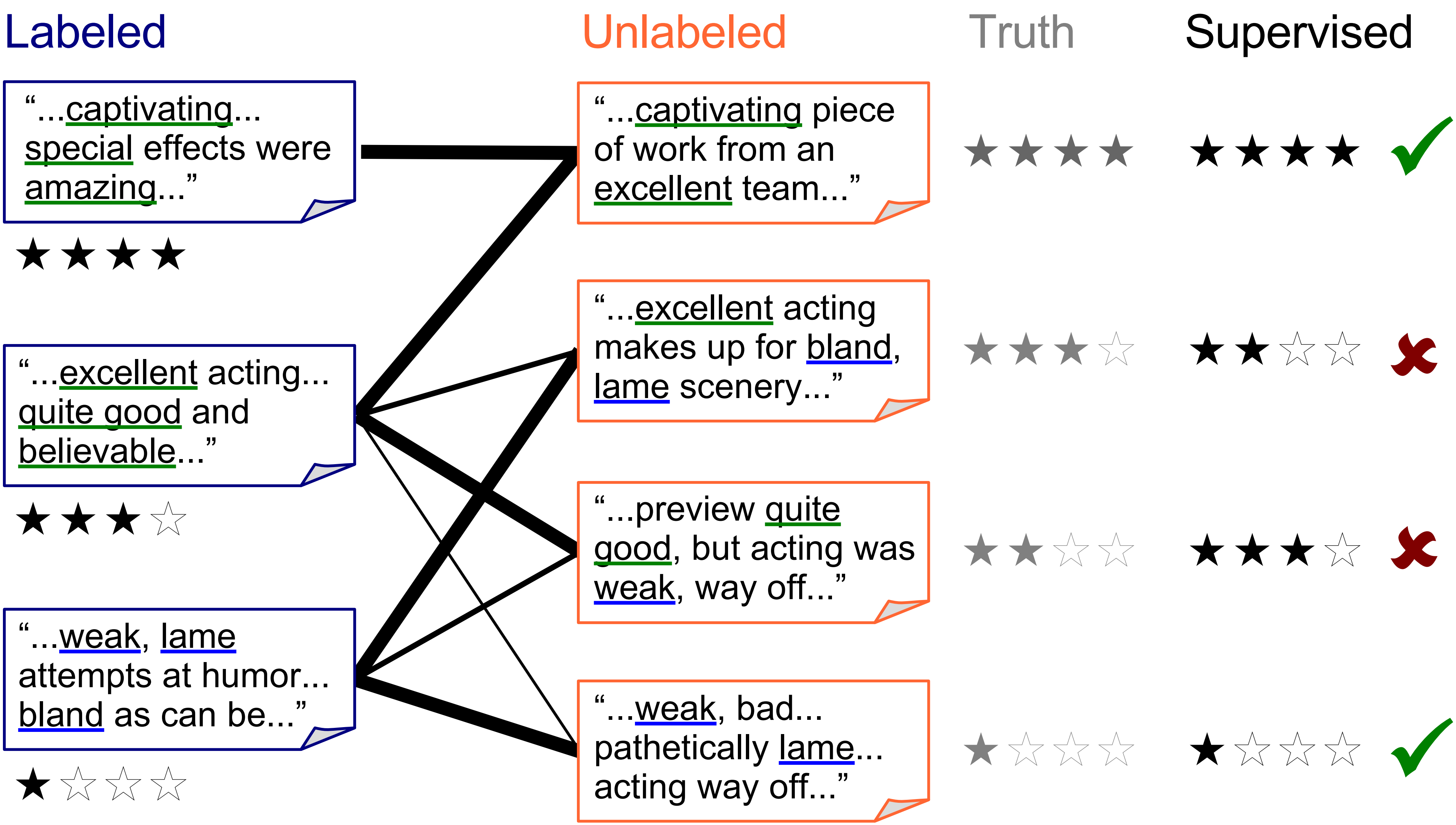
"...excellent acting... quite good and believable..."
★★★☆

"...weak, lame attempts at humor... bland as can be..."
★☆☆☆

Unlabeled

Unlabeled	Truth	Supervised
"...captivating piece of work from an excellent team..."	★★★★	★★★★
"...excellent acting makes up for bland, lame scenery..."	★★★☆	★★☆☆
"...preview quite good, but acting was weak, way off..."	★★☆☆	
"...weak, bad... pathetically lame... acting way off..."	★☆☆☆	

Graph-Based Learning

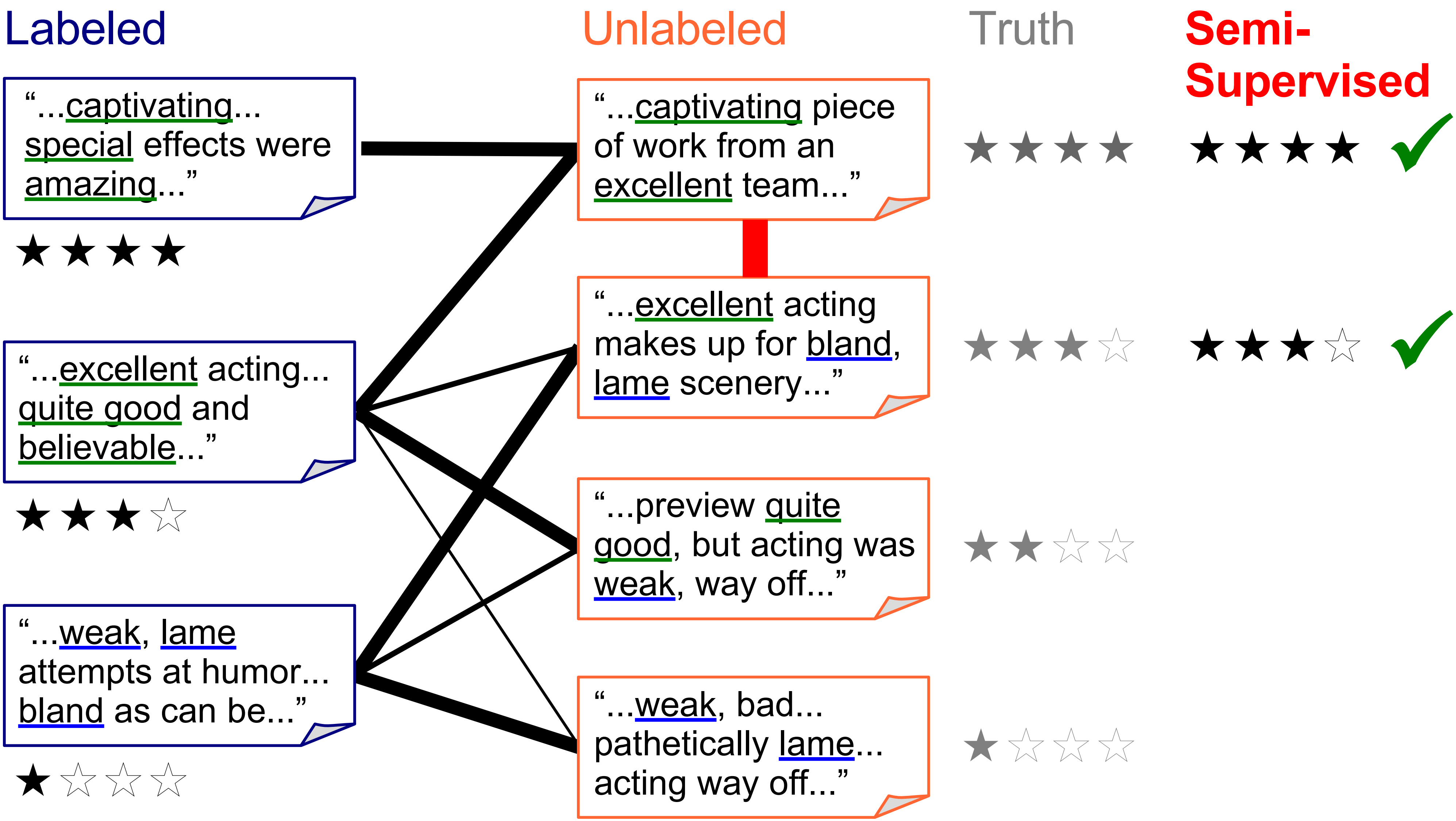


Graph-Based

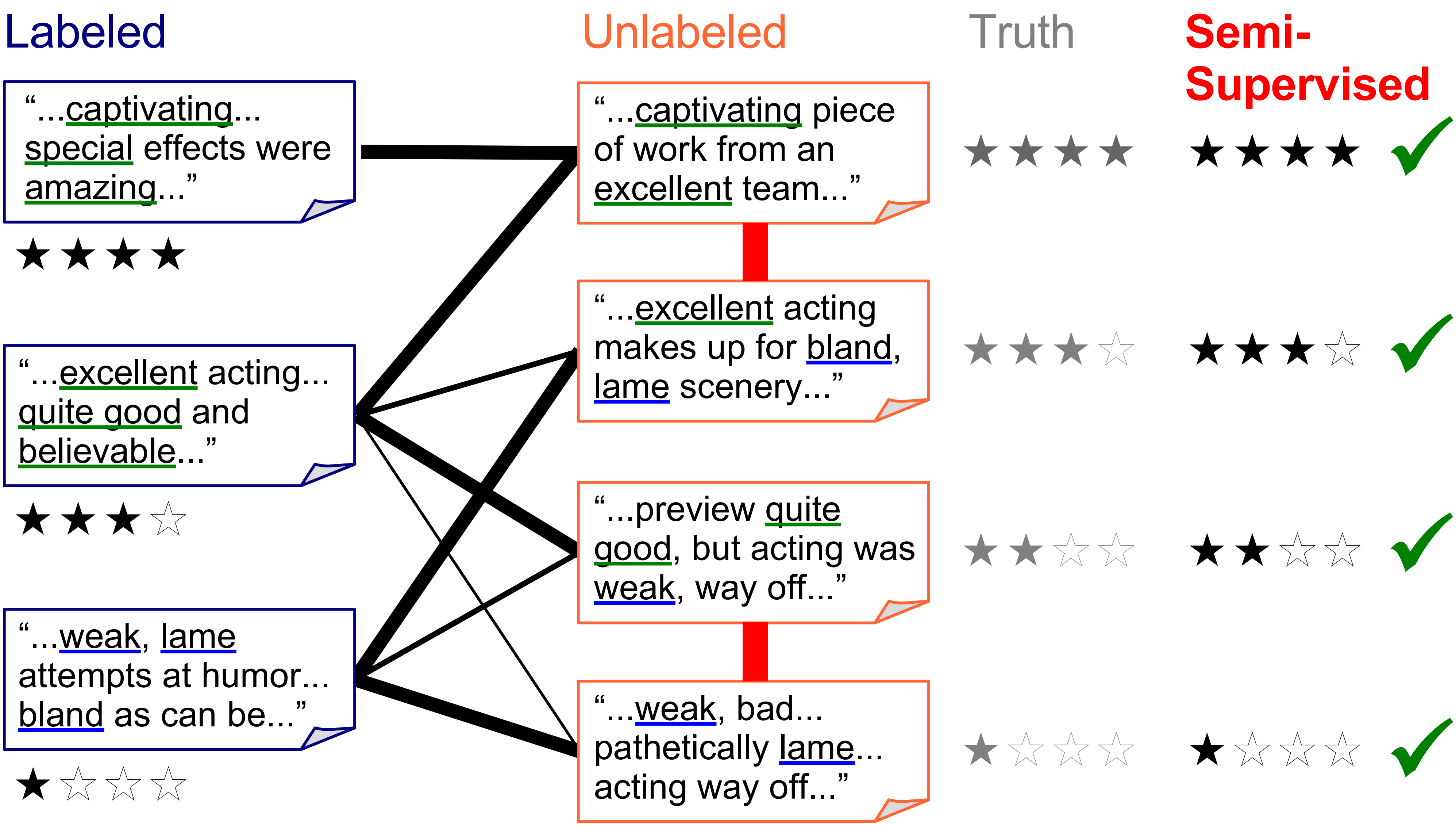
50% Accuracy ☹

Labeled	Unlabeled	Truth	Supervised	
"...captivating... special effects were amazing..." ★★★★	"...captivating piece of work from an excellent team..."	★★★★	★★★★	✓
"...excellent acting... quite good and believable..." ★★★☆	"...excellent acting makes up for bland, lame scenery..."	★★★☆	★★☆☆	✗
"...weak, lame attempts at humor... bland as can be..." ★☆☆☆	"...preview quite good, but acting was weak, way off..."	★★☆☆	★★★☆	✗
	"...weak, bad... pathetically lame... acting way off..."	★☆☆☆	★☆☆☆	✓

Graph-Based Learning

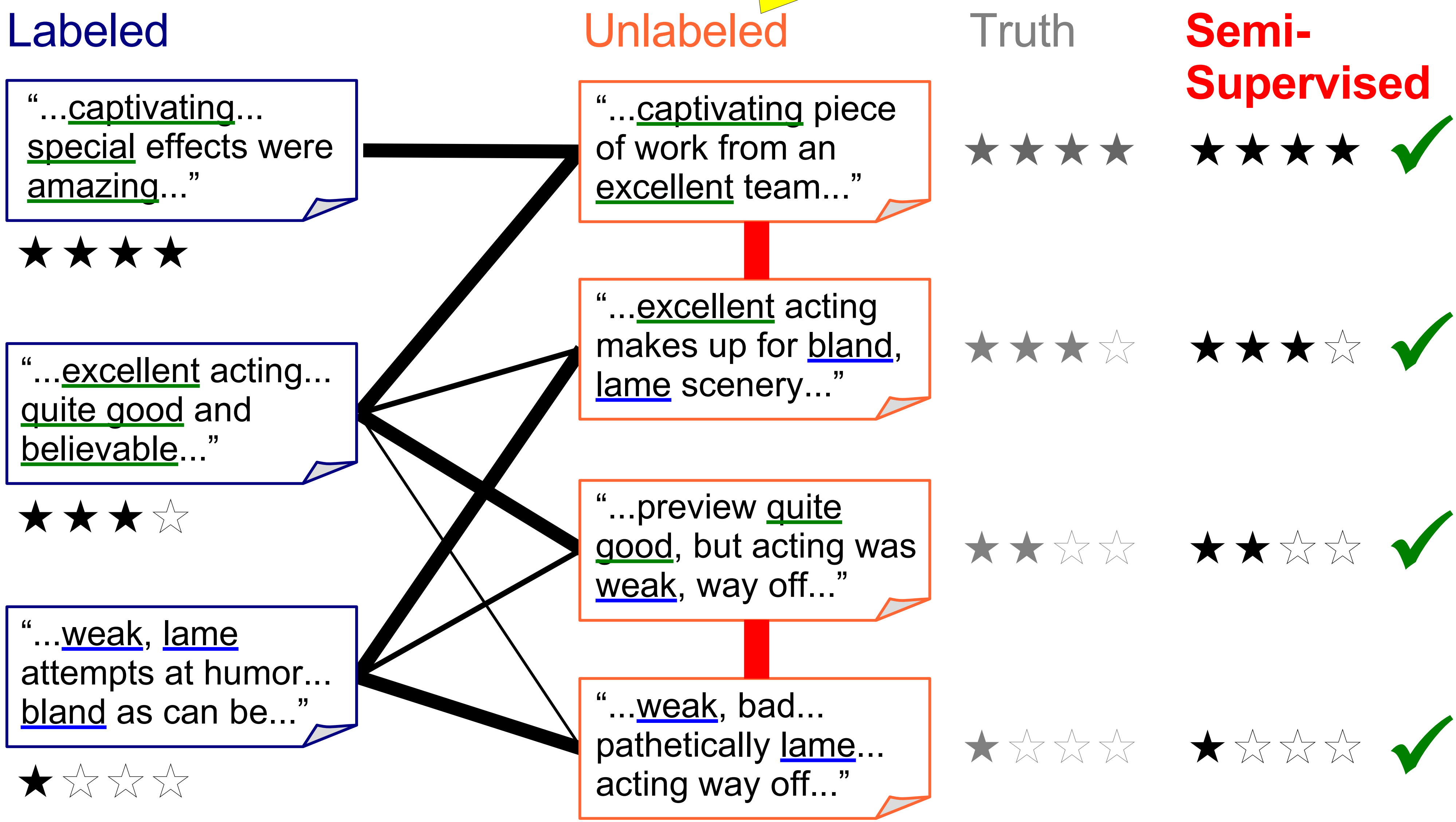


Graph-Based Learning



Graph-Bas

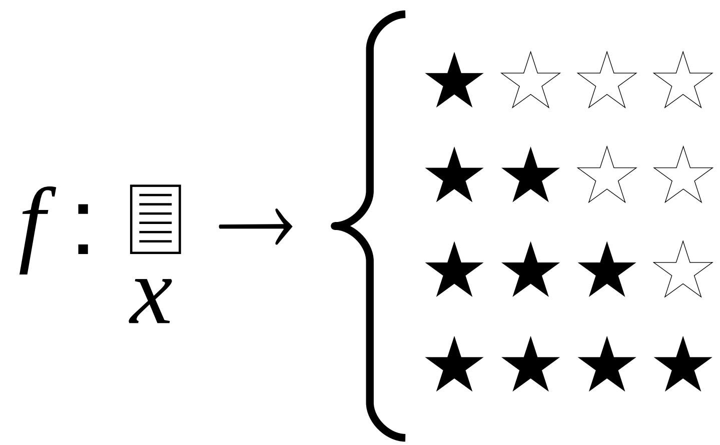
100% Accuracy ☺



How it really works

Goal

- Assign a discrete numeric rating *f(x)* to each document *x*

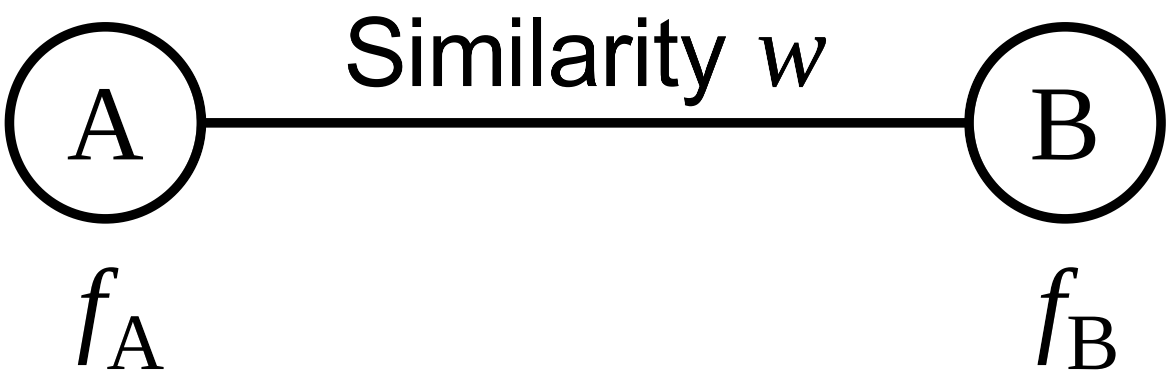


Our Approach

- Get initial predictions using SVM
- Improve predictions using graph-based SSL
  - Nodes = reviews
  - Edges = assumed relations between reviews
  - Find the optimal f over the graph

Measuring Loss over the Graph

Similar reviews should get similar ratings

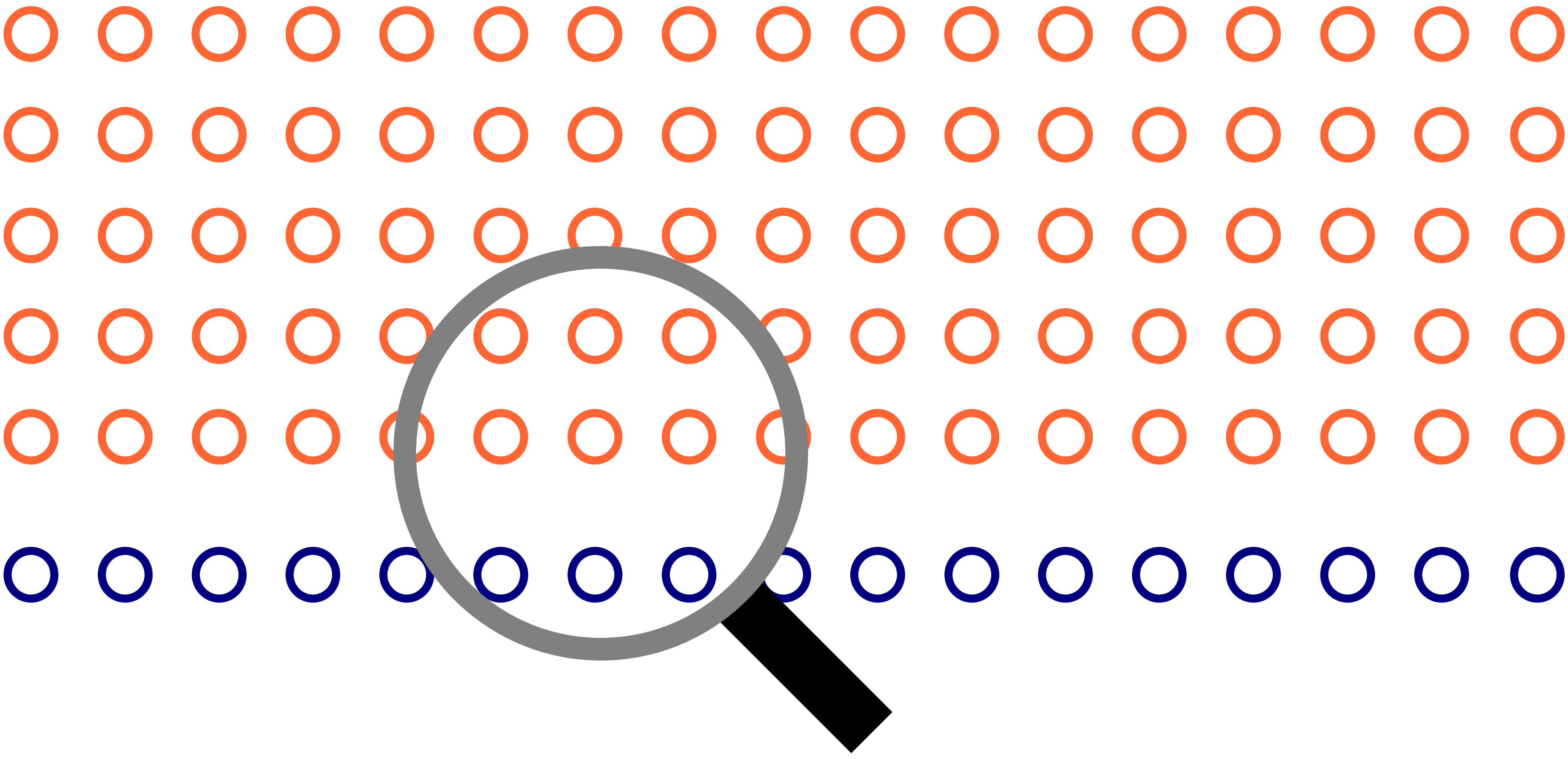


Loss over this edge $= w\,(f_A - f_B)^2$

Task: Minimize loss $L(f)$ over whole graph

Semi-Supervised Learning Graph

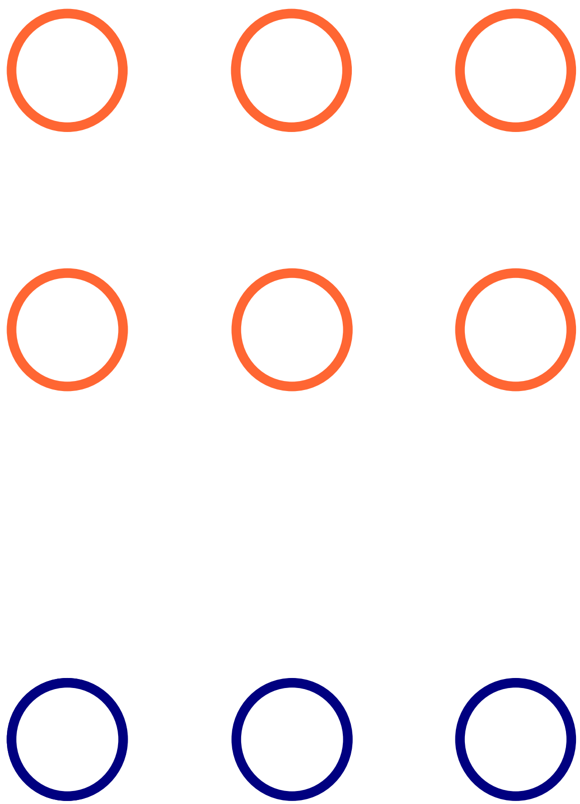
Unlabeled



Labeled

Semi-Supervised Learning Graph

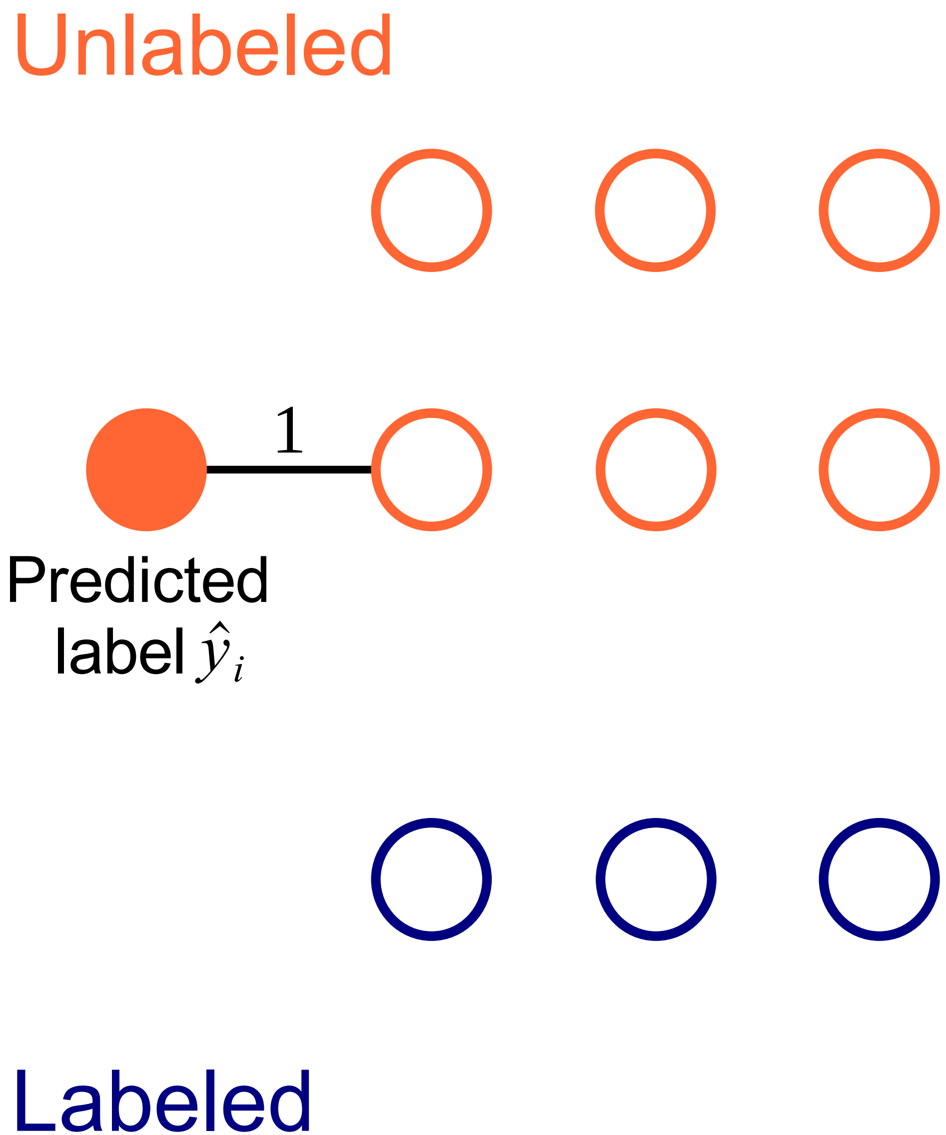
Unlabeled



Labeled

Loss $L(f) =$

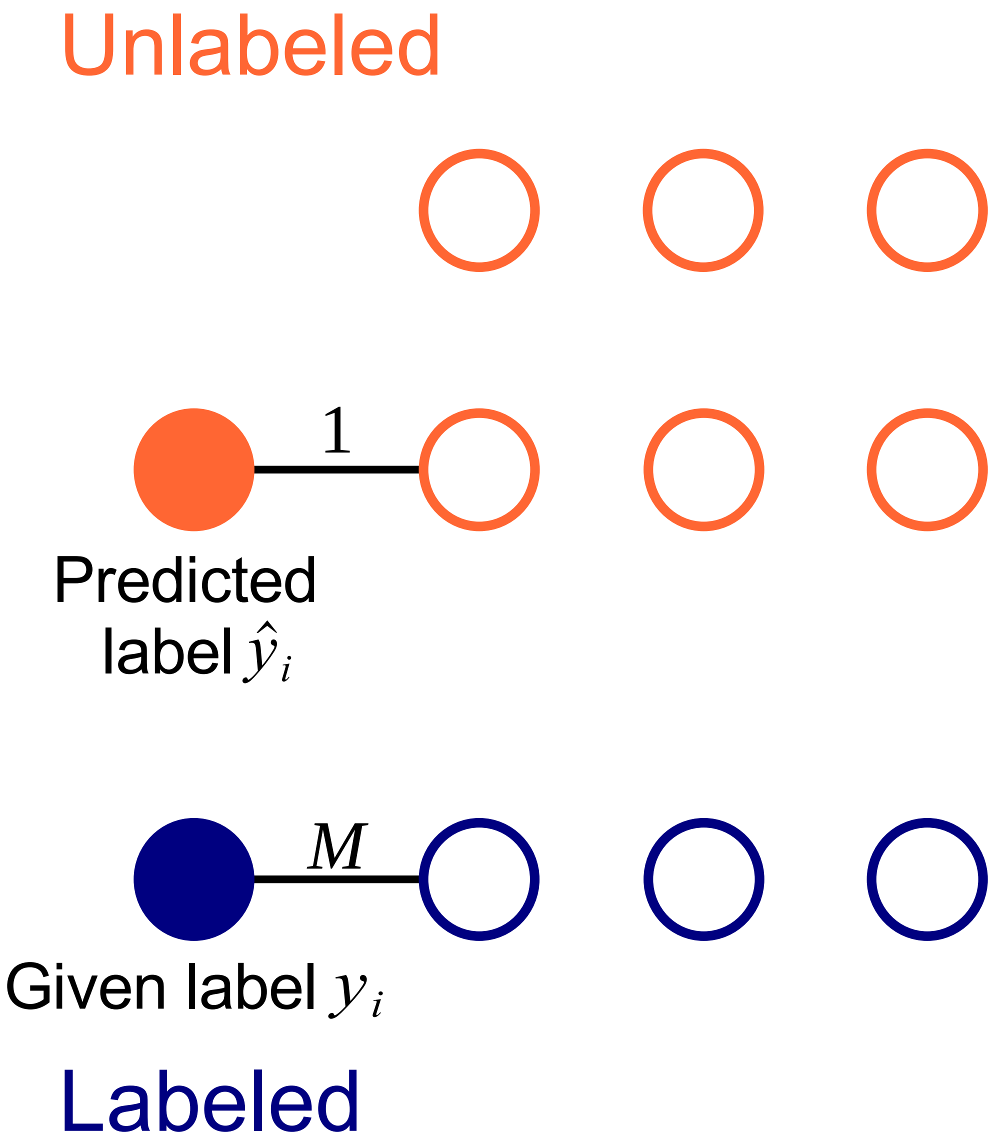
Semi-Supervised Learning Graph



Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2$$

Semi-Supervised Learning Graph

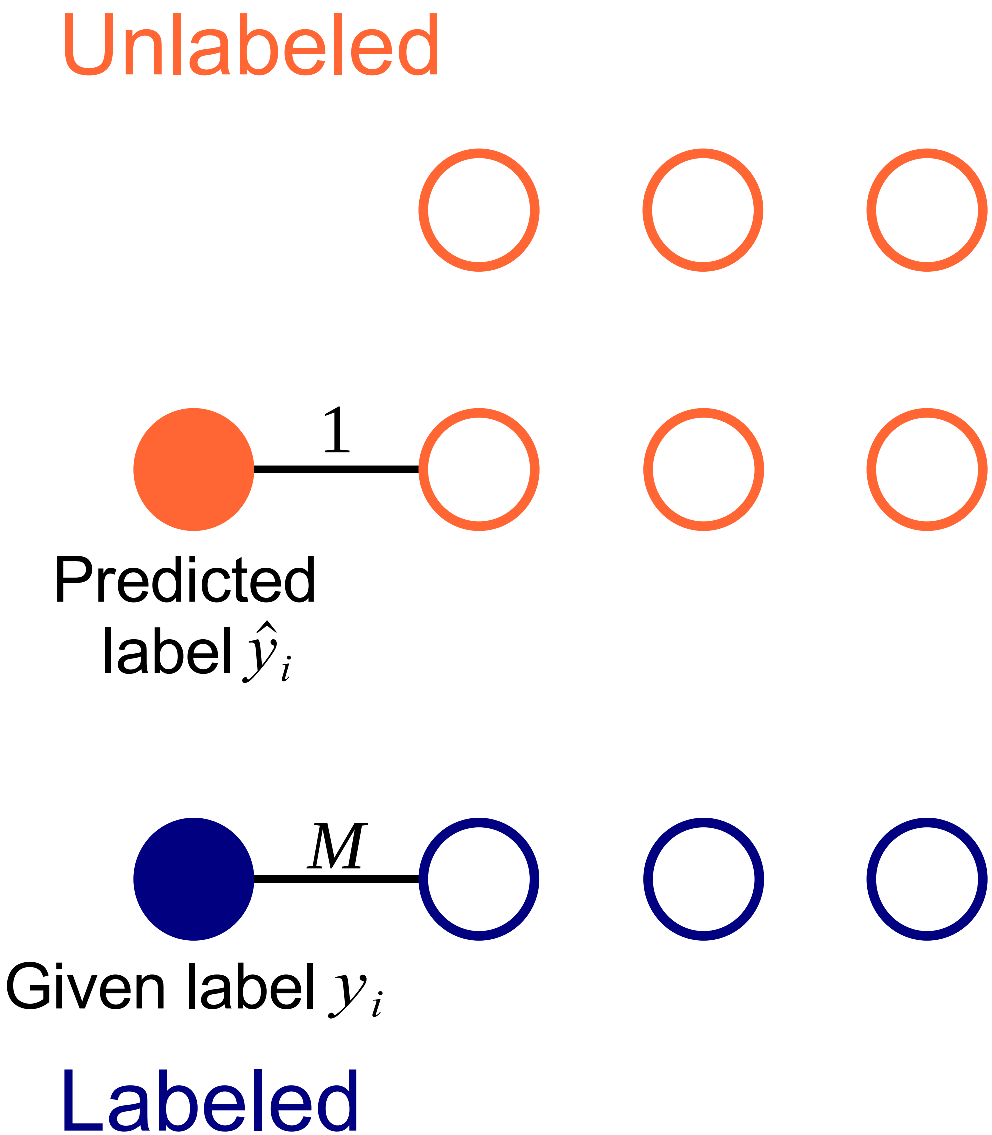


Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2 +$$

$$\sum_{i \in L} M (f_i - y_i)^2$$

Semi-Supervised Learning Graph



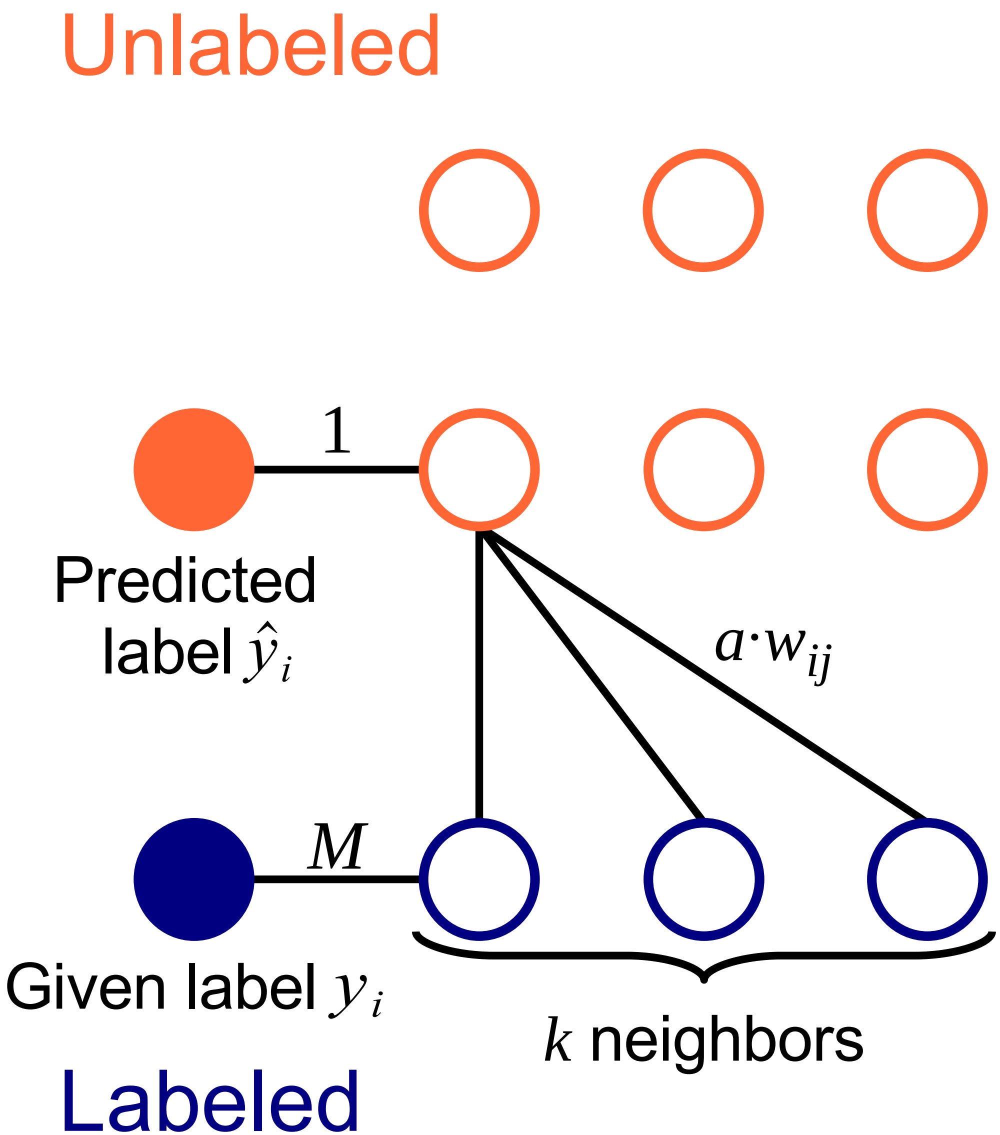
Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2 +$$

$$\sum_{i \in L} M (f_i - y_i)^2$$

Minimization is fairly trivial $f_i = \hat{y}_i$ or y_i

Semi-Supervised Learning Graph



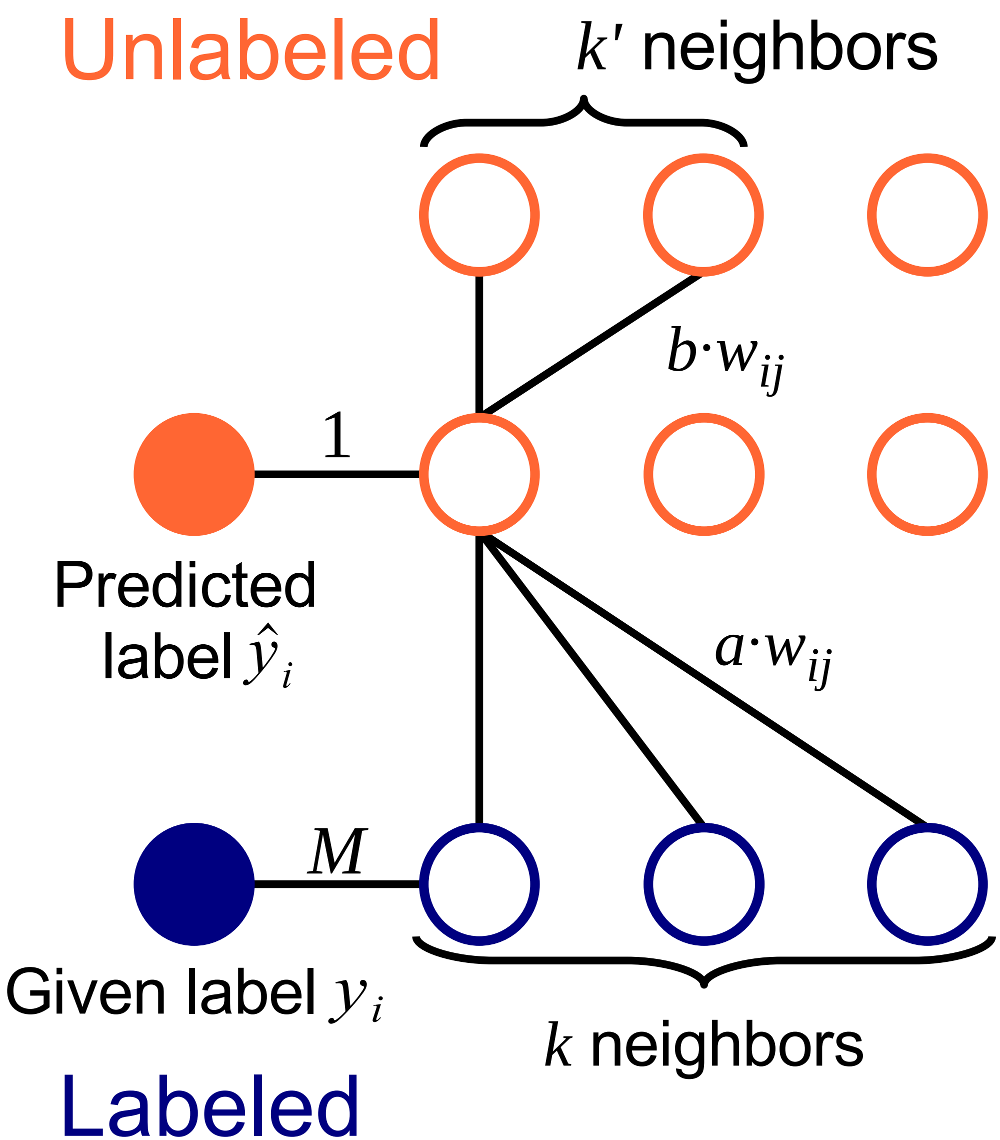
Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2 +$$

$$\sum_{i \in L} M (f_i - y_i)^2 +$$

$$\sum_{\substack{i \in U \\ j \in kNN_L(i)}} a w_{ij} (f_i - f_j)^2$$

Semi-Supervised Learning Graph



Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2 +$$

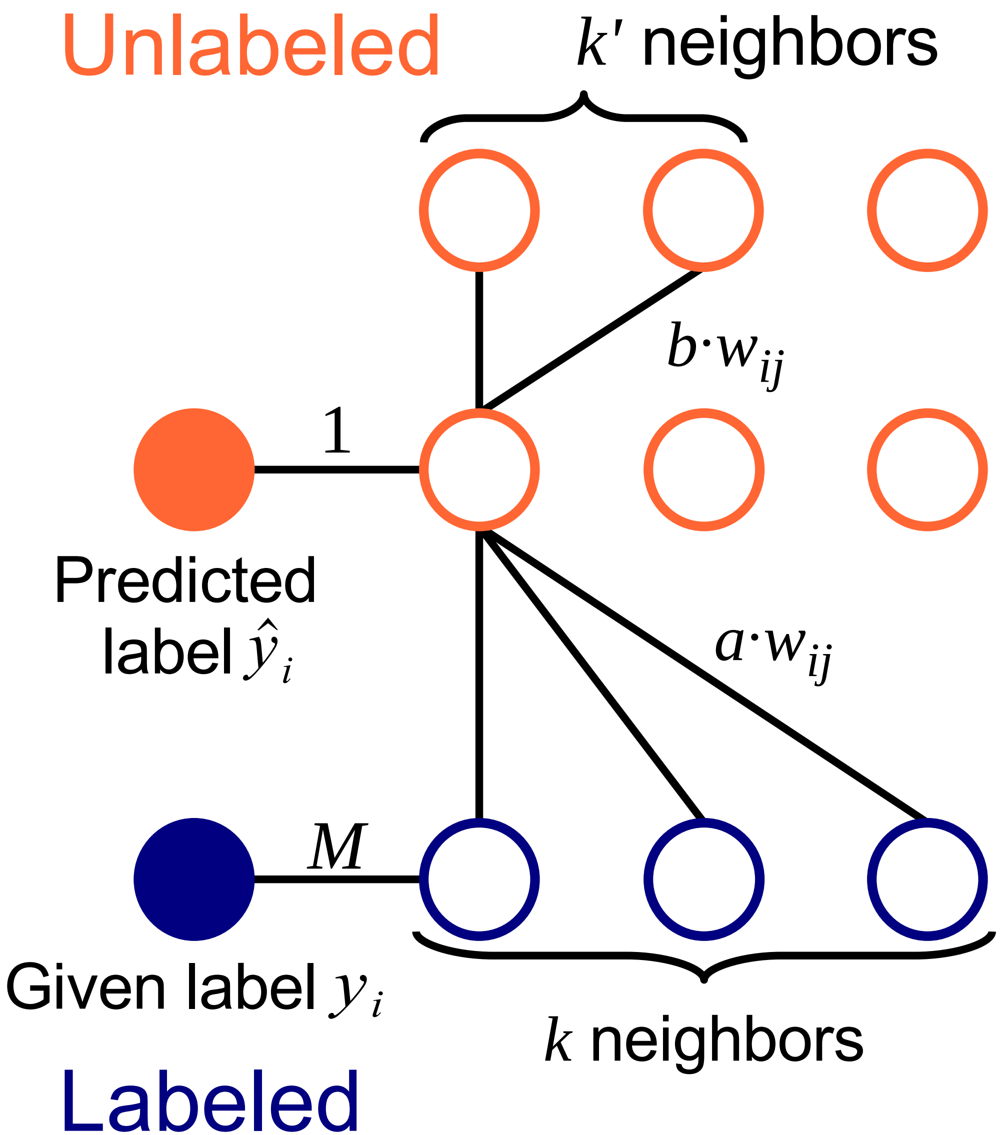
$$\sum_{i \in L} M (f_i - y_i)^2 +$$

$$\sum_{\substack{i \in U \\ j \in kNN_L(i)}} a w_{ij} (f_i - f_j)^2 +$$

$$\sum_{\substack{i \in U \\ j \in k'NN_U(i)}} b w_{ij} (f_i - f_j)^2$$

Semi-Supervised

Minimization is now *non-trivial*



Loss $L(f)=$

$$\sum_{i\in U}(f_i-\hat{y}_i)^2+$$

$$\sum_{i\in L} M(f_i-y_i)^2+$$

$$\sum_{\substack{i\in U\\ j\in kNN_L(i)}} a w_{ij}(f_i-f_j)^2+$$

$$\sum_{\substack{i\in U\\ j\in k'NN_U(i)}} b w_{ij}(f_i-f_j)^2$$

Finding a Closed-Form Solution

$$L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 +$$

$$\sum_{\substack{i \in U \\ j \in kNN_L(i)}} aw_{ij}(f_i - f_j)^2 + \sum_{\substack{i \in U \\ j \in k'NN_U(i)}} bw_{ij}(f_i - f_j)^2$$

Finding a Closed-F

Yikes!

$$L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M (f_i - y_i)^2 +$$

$$\sum_{\substack{i \in U \\ j \in kNN_L(i)}} a w_{ij} (f_i - f_j)^2 + \sum_{\substack{i \in U \\ j \in k'NN_U(i)}} b w_{ij} (f_i - f_j)^2$$

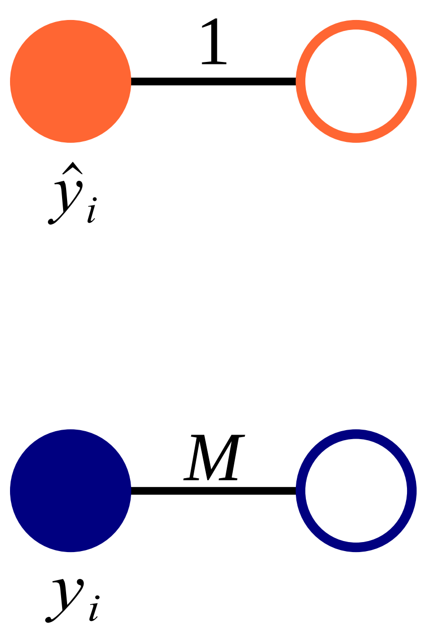
Can rewrite $L(f)$ in matrix notation as:

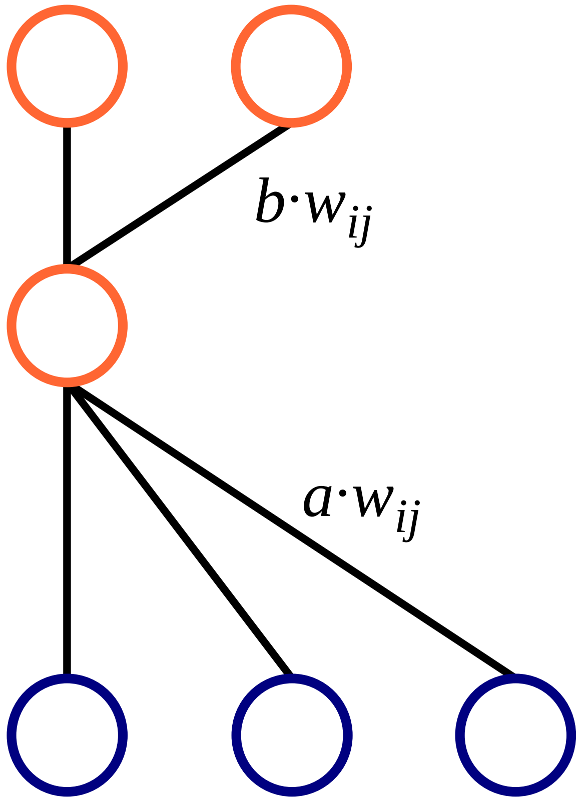
$$(\mathbf{f} - \mathbf{y})^{\mathsf{T}} C (\mathbf{f} - \mathbf{y}) + \eta \, \mathbf{f}^{\mathsf{T}} \Delta \, \mathbf{f}$$

Finding a Closed-Form Solution

Can rewrite $L(f)$ in matrix notation as:

$$(\mathbf{f}-\mathbf{y})^{\top} C (\mathbf{f}-\mathbf{y}) + \eta\, \mathbf{f}^{\top} \Delta\, \mathbf{f}$$





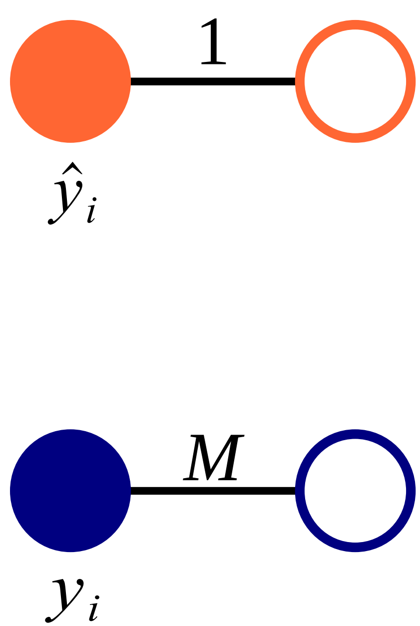
…n… Solution

…r… …tation as:

Vector of f values for all reviews

Vector of given labels y_i for labeled reviews and predicted labels $\hat{y}_i$ for unlabeled reviews

$$(\mathbf{f}-\mathbf{y})^{\mathsf{T}} C (\mathbf{f}-\mathbf{y}) + \eta\, \mathbf{f}^{\mathsf{T}} \Delta\, \mathbf{f}$$



$$C = \begin{array}{c} \overbrace{\qquad\qquad}^{\text{Labeled}} \; \overbrace{\qquad\qquad\qquad}^{\text{Unlabeled}} \\ \begin{bmatrix} M &&&&&&& \\ & \ddots &&&&&& \\ && M &&&&& \\ &&& 1 &&&& \\ &&&& 1 &&& \\ &&&&& \ddots && \\ &&&&&& 1 & \\ &&&&&&& 1 \end{bmatrix} \end{array}$$

All other matrix entries are 0

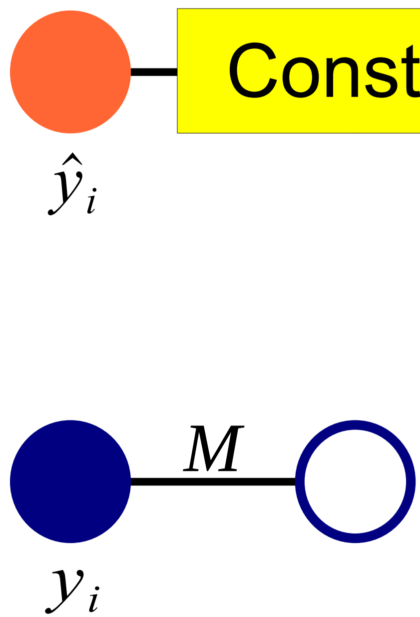
Finding a Closed-Form Solution

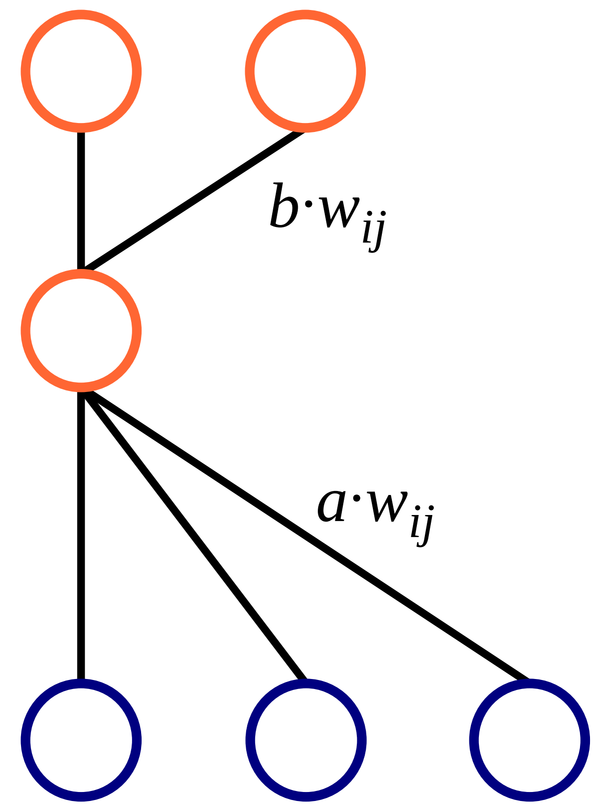
Graph *Laplacian* matrix

Can rewrite $L(f)$ in matrix notation

$$(\mathbf{f}-\mathbf{y})^{\mathsf{T}} C (\mathbf{f}-\mathbf{y}) + \eta\, \mathbf{f}^{\mathsf{T}} \Delta\, \mathbf{f}$$

Constant parameter





Graph *Laplacian* Matrix

Assume n labeled and unlabeled documents

W = n x n weight matrix

D = n x n diagonal degree matrix, where

$$D_{ii} = \sum_{j=1}^{n} W_{ij}$$

Graph *Laplacian* matrix is

$$\Delta = D - W$$

Finding a Closed-Form Solution

- $L(f)$ in matrix notation:

$$(\mathbf{f}-\mathbf{y})^{\top} C (\mathbf{f}-\mathbf{y}) + \eta\, \mathbf{f}^{\top} \Delta\, \mathbf{f}$$

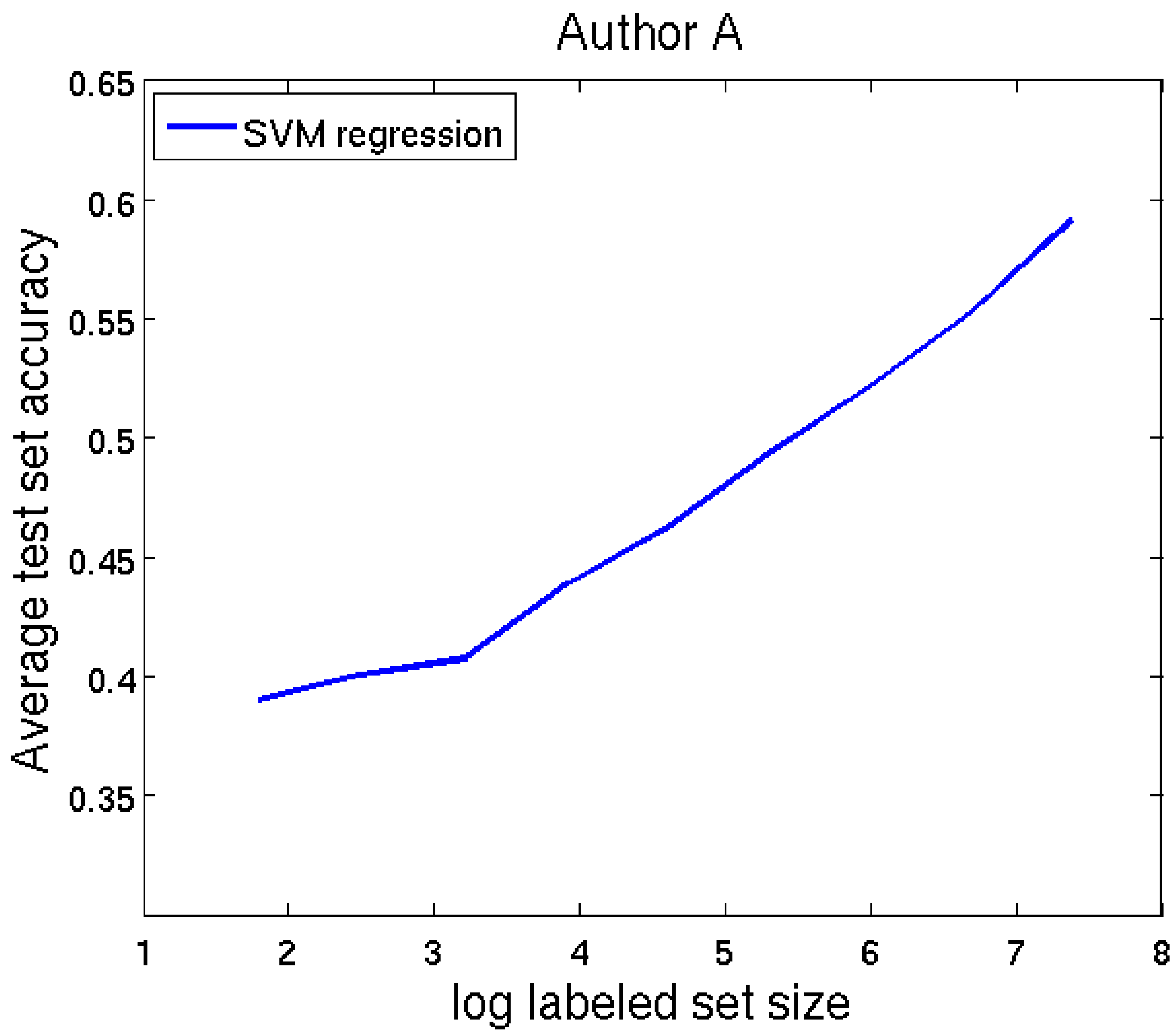
- Set gradient to zero and solve for $\mathbf{f}$:

$$\mathbf{f} = \left(C + \eta \Delta\right)^{-1} C\, \mathbf{y}$$

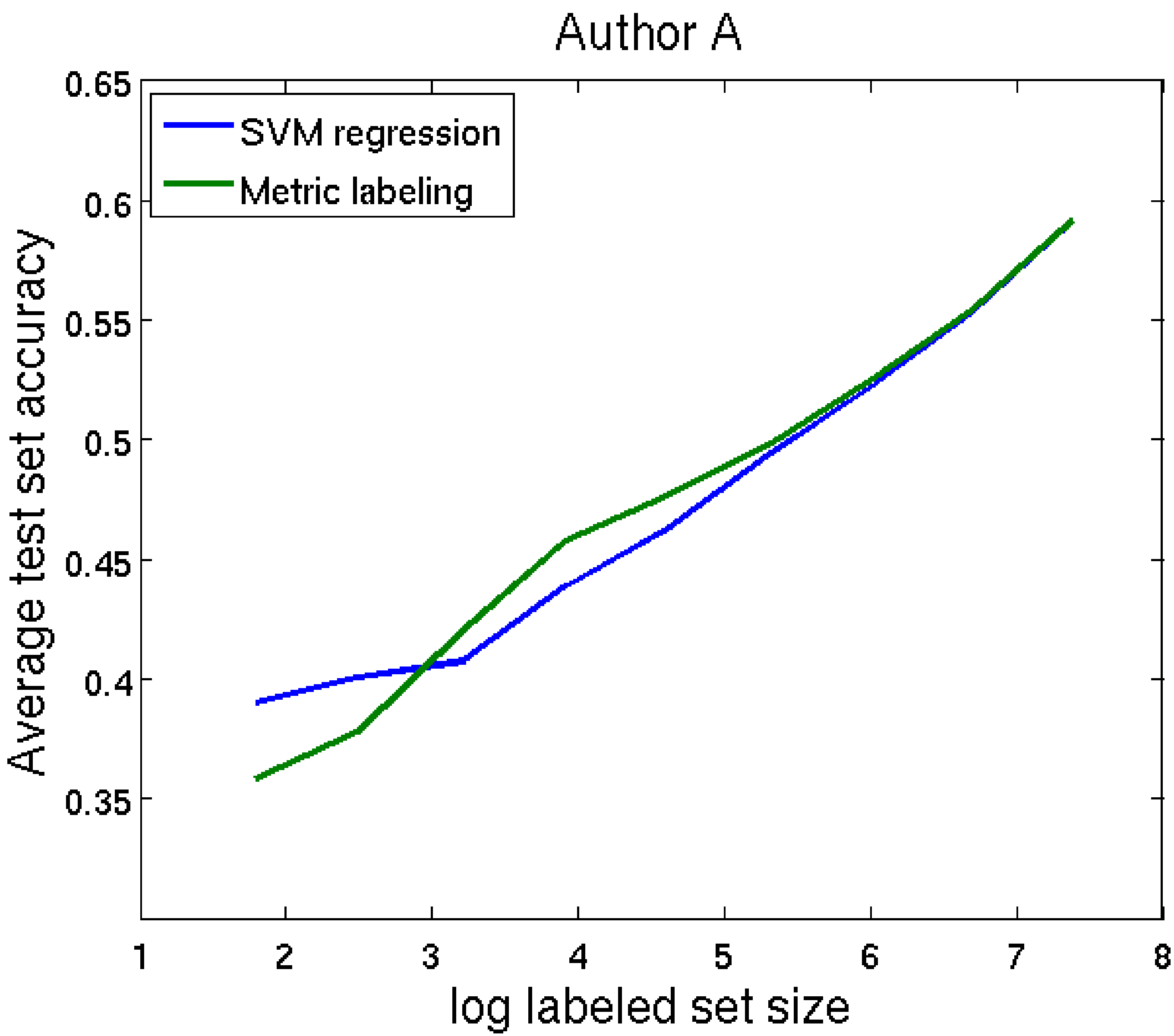
Experiments

- Task: Predict 1 to 4 star ratings for reviews
  - 4-author data used by [Pang+Lee05]
  - Predicted $\hat{y}_i$ values with SVM^{light} regression using {0,1} word vectors
  - Positive-sentence percentage (PSP) similarity [Pang+Lee05]
  - Tuned parameters with cross validation

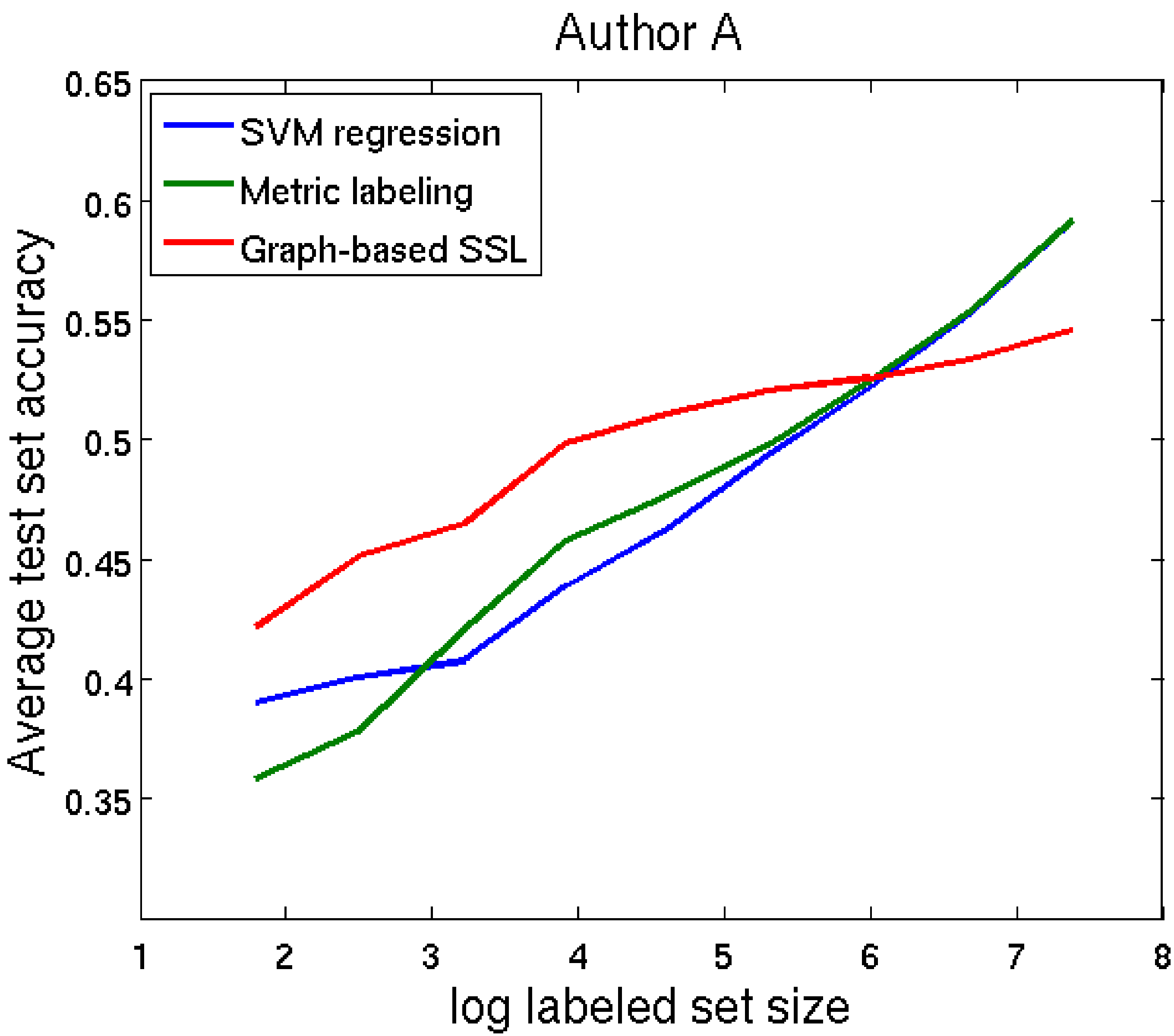
Results


Author A
SVM regression
Average test set accuracy
0.65
0.6
0.55
0.5
0.45
0.4
0.35
1
2
3
4
5
6
7
8
log labeled set size

Results

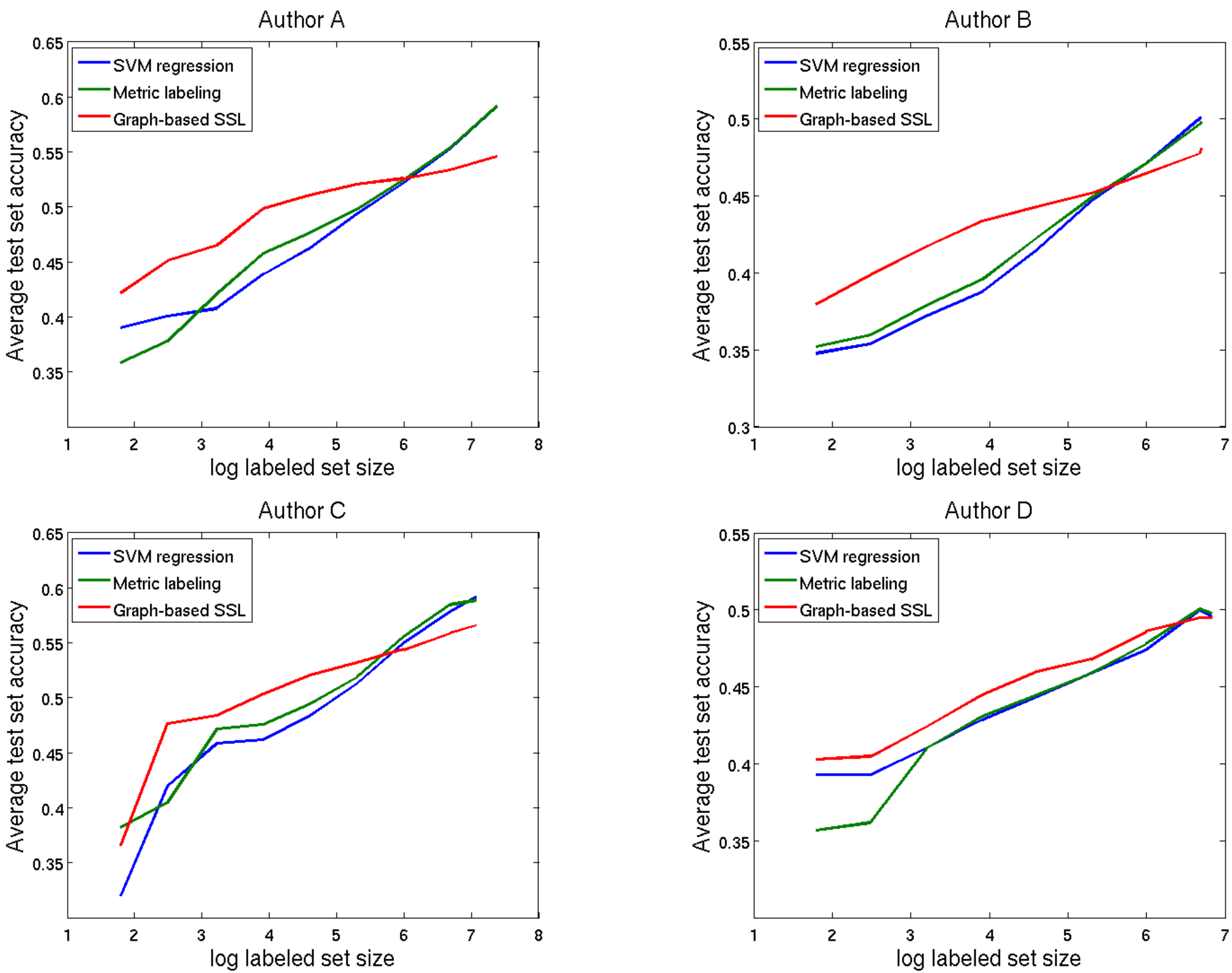

Author A
SVM regression
Metric labeling
Average test set accuracy
0.65
0.6
0.55
0.5
0.45
0.4
0.35
1
2
3
4
5
6
7
8
log labeled set size

Results

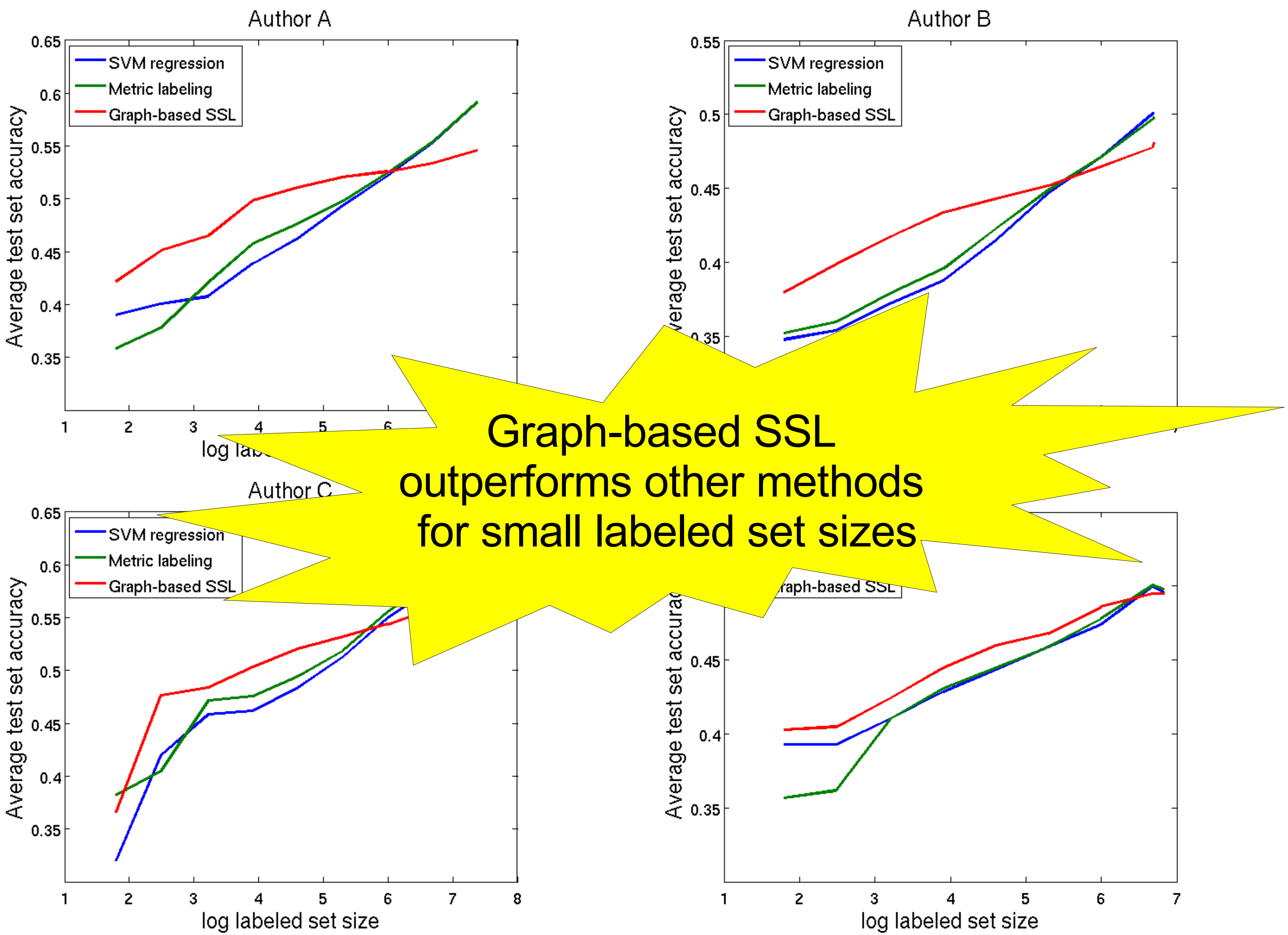


Author A
0.65
0.6
0.55
0.5
0.45
0.4
0.35
Average test set accuracy
SVM regression
Metric labeling
Graph-based SSL
1
2
3
4
5
6
7
8
log labeled set size

Results


Author A
Author B
Author C
Author D
SVM regression
Metric labeling
Graph-based SSL
Average test set accuracy
log labeled set size
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
1
2
3
4
5
6
7
8

Results



Author A
Author B
Author C
SVM regression
Metric labeling
Graph-based SSL
Average test set accuracy
log labeled set size
Graph-based SSL
outperforms other methods
for small labeled set sizes

Conclusions

- Adapted graph-based semi-supervised learning to sentiment analysis domain
- Designed a graph for rating inference
- Showed benefit of SSL using movie review data

Thank you! Any questions?