Graph-Based Active Learning: A New Look at Expected Error Minimization

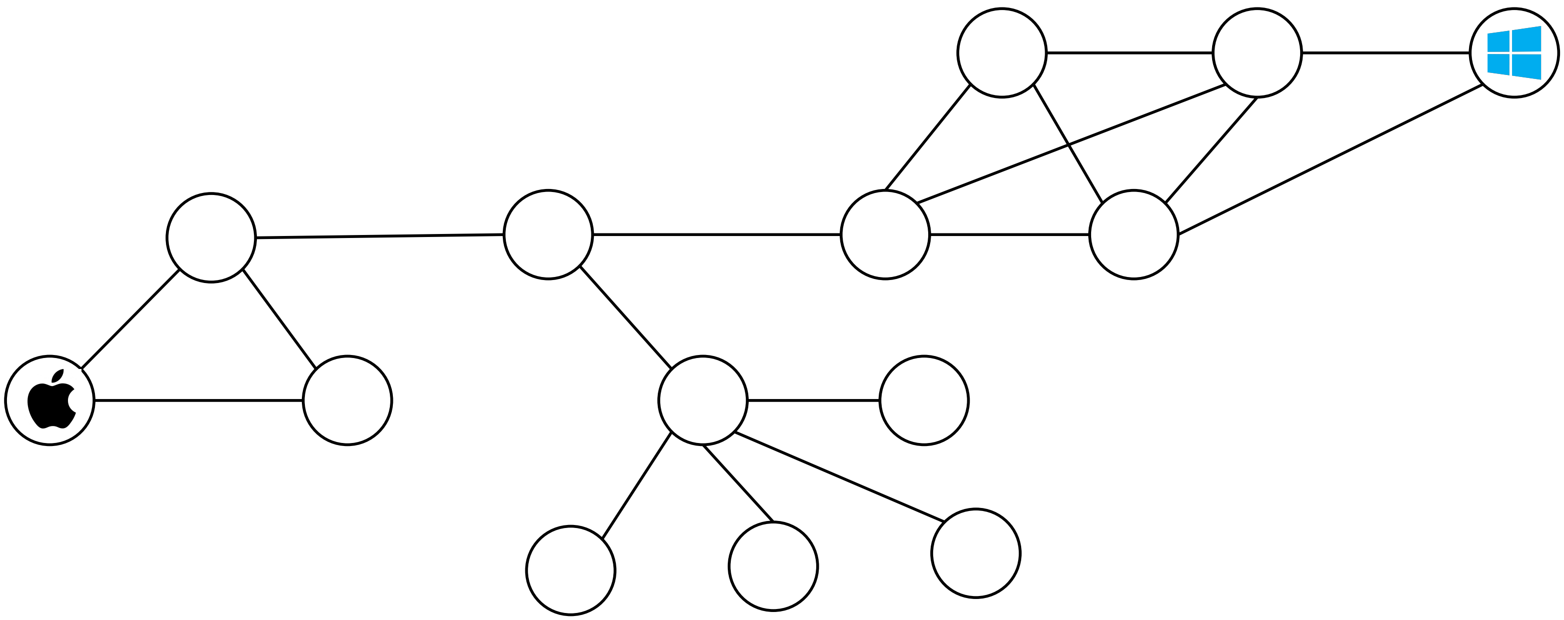
Kwang-Sung Jun and Robert Nowak

Wisconsin Institutes for Discovery, UW-Madison

12/09/2016

Graph-Based Learning

- Task: Predict Mac/PC users in a friendship network.



- Semi-supervised classification
  - A seminal work: [Zhu03a]
- Active setting: pick a few nodes and query their labels.
  - Can we choose judiciously which nodes to query?

Graph-Based Active Learning

- G=(N,E), n := |N|, w_{ij} = non-negative edge weight (undirected)
- $Y_i \in \{1,-1\}$: label of node i
- $\boldsymbol{\ell}$: the set of labeled nodes, $\boldsymbol{u}$:= N \ $\boldsymbol{\ell}$

- For t= 1…T
  - **(Predict)** Predict $\hat{Y}_i$ for i $\in \boldsymbol{u}$. The algorithm suffers prediction error $\frac{1}{n}\sum_{i=1}^{n} 1\{\hat{Y}_i = Y_i\}$, which is unknown to the algorithm.
  - **(Query)** choose a node q $\in \boldsymbol{u}$ and request its label Y_q.
    Set $\boldsymbol{\ell} \leftarrow \boldsymbol{\ell}$ \ {q}.

- For **Predict**, [Zhu03a] is the de facto standard.

Various Approaches

- Theoretical approach (learning theory community)
  - Assumption: adversarial labels.
  - [Cesa-Bianchi10]: analysis on tree graphs only.
  - [Dasarathy15]: a weak form of guarantee; "when can we perform a perfect prediction?".

- Graph sampling theory (signal processing community)
  - [Gadde14, Chen15]: assume the categorical labels are bandlimited signals (real-valued).

- Probabilistic approach (**OURS**) (machine learning community)
  - Bayesian, model-based approach. Categorical labels.
  - Assumption: the labels are generated by a distribution $P(Y_{1:n})$.
  - Then, we can compute the expected error!

Binary Markov Random Field (BMRF)

- **BMRF**: $\mathbb{P}(\mathbf{Y}_{1:n} = \mathbf{y}_{1:n}) = \frac{1}{Z} \exp\left(-\frac{\beta}{2} \sum_{i<j} w_{ij} (y_i - y_j)^2 \right)$, where $\beta > 0$
  - Encourages the same labels along the edges.
  - Different from Ising. BMRF is for nonnegative w_{ij}, not restricted to lattice.

Expected Error Minimization (EEM)

- obs := $\{Y_i = y_i\}_{i \in \boldsymbol{\ell}}$
- Lookahead risk of query q

"expected error after Yq is revealed"

$$R^{+q}(\text{obs}) := \mathbb{E}_{Y_q} \boxed{\mathbb{E}_{\mathbf{Y}_{\mathbf{u} \setminus \{q\}}} \left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\widehat{Y}_i \neq Y_i\} \,\middle|\, Y_q, \text{obs} \right]}$$

- Query $q^* := \arg\min_{q \in \boldsymbol{u}} R^{+q}(\text{obs})$

⇒ One step optimal!

EEM is hard

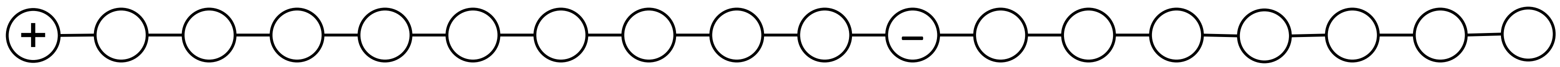
Q: Can't you just compute $q^* := \arg\min_{q \in \mathcal{U}} R^{+q}(obs)$?

A: No. There is no known polynomial time algorithm.

- This is where a lot of efforts were put.
  - ZLG [Zhu03b]: naïve approximation of the marginal distribution
  - VOpt [Ji12]: continuous relaxation of $Y_1, \ldots, Y_n$
  - SOpt [Ma13]: continuous relaxation with an alternative error criterion

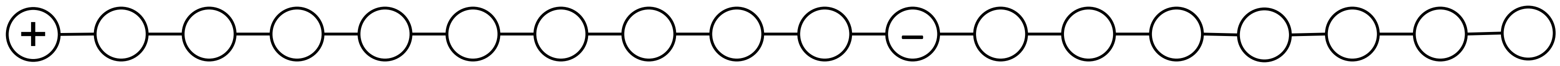
Claim: None of the above is satisfactory.

Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	**0.50**
ZLG					cut region										uncertain region				
SOpt																			
Exact																			
TSA																			

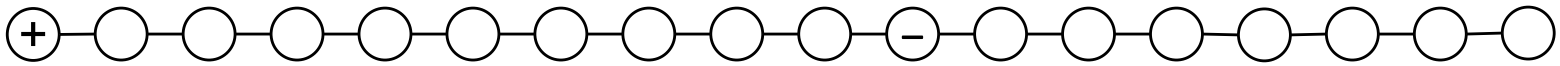
Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	0.50
ZLG	+	+	+	+	+	1	+	2	3	4	−	−	−	−	−	−	−	−	0.33
SOpt																			
Exact																			
TSA																			

- ZLG lacks exploration queries.

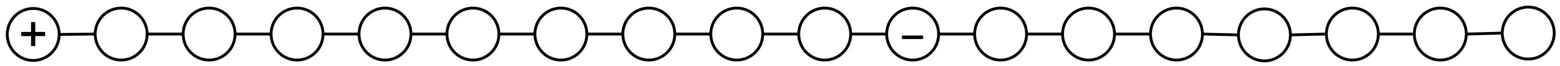
Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	0.50
ZLG	+	+	+	+	+	1	+	2	3	4	−	−	−	−	−	−	−	−	0.33
SOpt	+	+	+	+	+	1	+	+	−	−	−	−	−	+	+	2	+	+	0.06
Exact																			
TSA																			

- ZLG lacks exploration queries.
- SOpt lacks exploitation queries (non-adaptive).

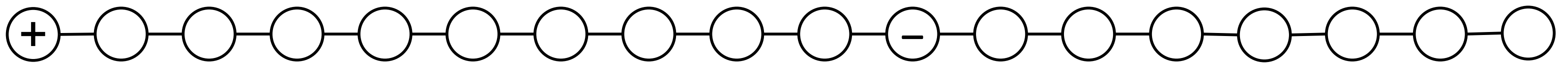
Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	–	–	–	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	**0.50**
ZLG	+	+	+	+	+	**1**	+	**2**	**3**	**4**	–	–	–	–	–	–	–	–	**0.33**
SOpt	+	+	**3**	+	+	**1**	+	+	–	–	–	–	**4**	+	+	**2**	+	+	**0.06**
Exact																			
TSA																			

- ZLG lacks exploration queries.
- SOpt lacks exploitation queries (non-adaptive).

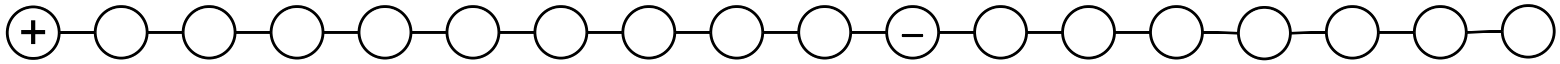
Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	–	–	–	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	0.50
ZLG	+	+	+	+	+	1	+	2	3	4	–	–	–	–	–	–	–	–	0.33
SOpt	+	+	3	+	+	1	+	+	–	–	–	–	4	+	+	2	+	+	0.06
Exact	+	+	+	+	+	1	+	+	–	–	–	–	–	+	+	2	+	+	0.00
TSA																			

- ZLG lacks exploration queries.
- SOpt lacks exploitation queries (non-adaptive).
- Exact computation balances between exploration and exploitation.

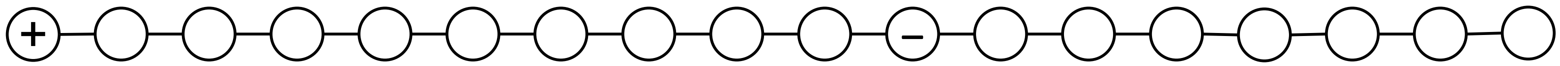
Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	–	–	–	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	0.50
ZLG	+	+	+	+	+	1	+	2	3	4	–	–	–	–	–	–	–	–	0.33
SOpt	+	+	3	+	+	1	+	+	–	–	–	–	4	+	+	2	+	+	0.06
Exact	+	+	+	+	+	1	+	3	+	–	–	–	4	+	+	2	+	+	0.00
TSA																			

- ZLG lacks exploration queries.
- SOpt lacks exploitation queries (non-adaptive).
- Exact computation balances between exploration and exploitation.

Approx. EEM: Exploration vs Exploitation



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Error
Truth	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	
Initial	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	0.50
ZLG	+	+	+	+	+	1	+	2	3	4	−	−	−	−	−	−	−	−	0.33
SOpt	+	+	3	+	+	1	+	+	−	−	−	−	4	+	+	2	+	+	0.06
Exact	+	+	+	+	+	1	+	3	+	−	−	−	4	+	+	2	+	+	0.00
TSA	+	+	+	+	+	1	+	3	+	−	−	−	4	+	+	2	+	+	0.00

proposed!

- ZLG lacks exploration queries.
- SOpt lacks exploitation queries (non-adaptive).
- Exact computation balances between exploration and exploitation.
- TSA (ours) resembles BMRF.

Proposed: Two-Step Approximation (TSA)

Key Idea (skipping detail)

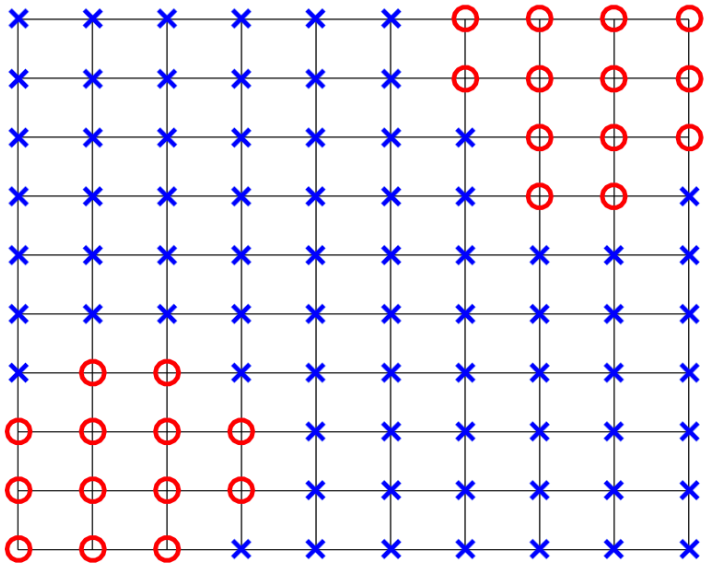
$g(y_{1:k})$: a quadratic function with negative definite Hessian.

$$\log\left(\sum_{y_{1:k}\in\{1,-1\}^k} \exp(g(y_{1:k}))\right) \leq \max_{y_{1:k}\in\{1,-1\}^k} g(y_{1:k}) + k\log(2)$$

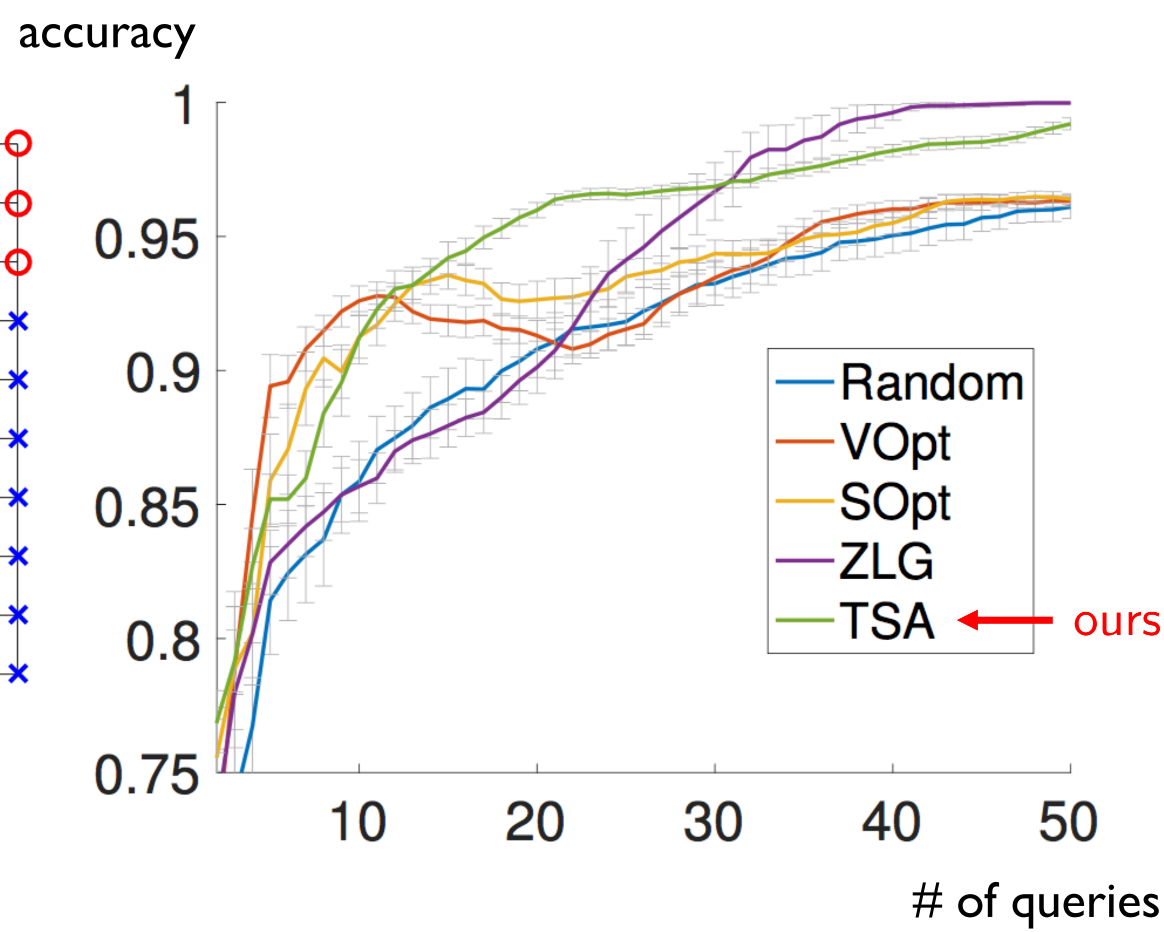
$$\leq \max_{y_{1:k}\in[-1,1]^k} g(y_{1:k}) + k\log(2)$$

A closed-form solution exists!

Experiment 1: Two Boxes(n=15)

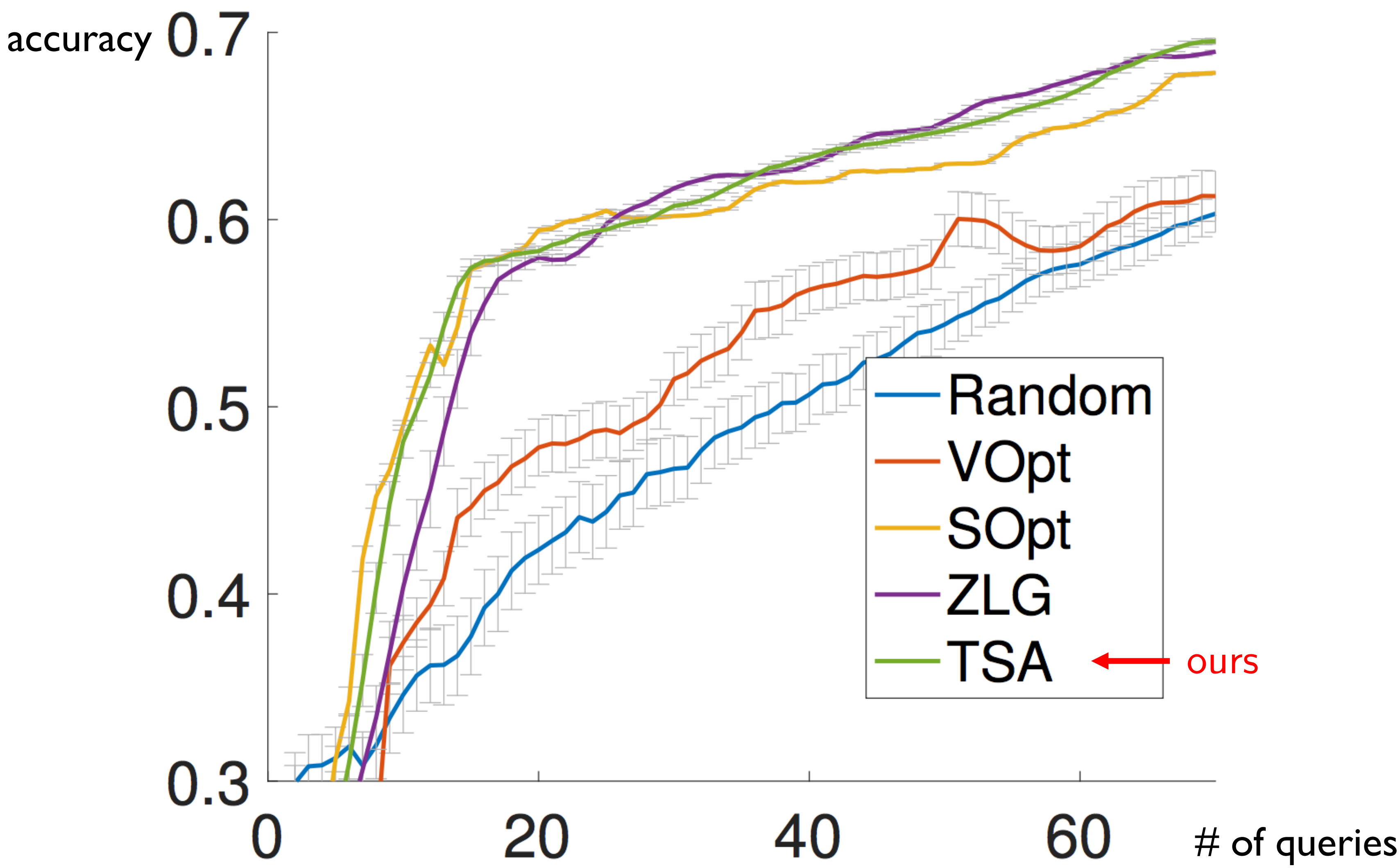


Ground truth



Experiment 2: DBLP

- Citation network, 4 classes, n=1,711



Discussion

- A close approximation of EEM that balances between exploration and exploitation.

- Future work
  - Theory on adversarial labels: are there convincing theoretical reason to prefer balancing exploration-exploitation?

  - Active Search: Find as many positive nodes as possible
    - E.g., find Mac users

  - Active Survey: Find the proportion of positive nodes as accurate as possible
    - E.g., Clinton vs Trump

Q&A

- Thank you!

References

- [Zhu03a] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. "Semi-supervised learning using Gaussian fields and harmonic functions." ICML, 2003
- [Zhu03b] Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. "Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions." ICML workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining, 2003
- [Cesa-Bianchi10] Nicolo Cesa-Bianchi, Claudio Gentile, Fabio Vitale and Giovanni Zappella. "Active Learning on Graphs via Spanning Trees." COLT, 2010.
- [Dasarathy15] Gautam Dasarathy, Robert Nowak, Xiaojin Zhu. "S2: An efficient graph based active learning algorithm with application to nonparametric classification." COLT, 2015
- [Ji12] Ming Ji and Jiawei Han, "A Variance Minimization Criterion to Active Learning on Graphs." AISTATS, 2012
- [Ma13] Yifei Ma, Roman Garnett, and Jeff Schneider, "Sigma-Optimality in Active Learning on Gaussian Random Fields." NIPS, 2013.
- [Gadde14] Akshay Gadde, Aamir Anis, and Antonio Ortega. "Active semi-supervised learning using sampling theory for graph signals." KDD, 2014.
- [Chen15] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling Theory." IEEE Trans. Signal Process., 2015.