Effect of Network Topology on The Performance of ADMM-based SVMs

Shirin Tavara, University of Borås

Alexander Schliep, University of Gothenburg (https://schlieplab.org/)

UNIVERSITY OF BORÅS

UNIVERSITY OF GOTHENBURG

PROBLEM

- Immense amount of data
  - Limited available memory
  - Slow training/predicting time
  - Deteriorated accuracy
  - Slow convergence
  - Overheads
  - Interaction between nodes

SOLUTION

Distributed optimization

- A large problem is divided into several smaller sub-problems
- Efficient network topology

Research Questions

- How much does the communication between the nodes affect the convergence of a distributed algorithm?
- Which network topology is preferable?

To Answer

- A particular distributed optimization algorithm for solving SVM problems
- A particular network topology with high connectivity, expander graphs

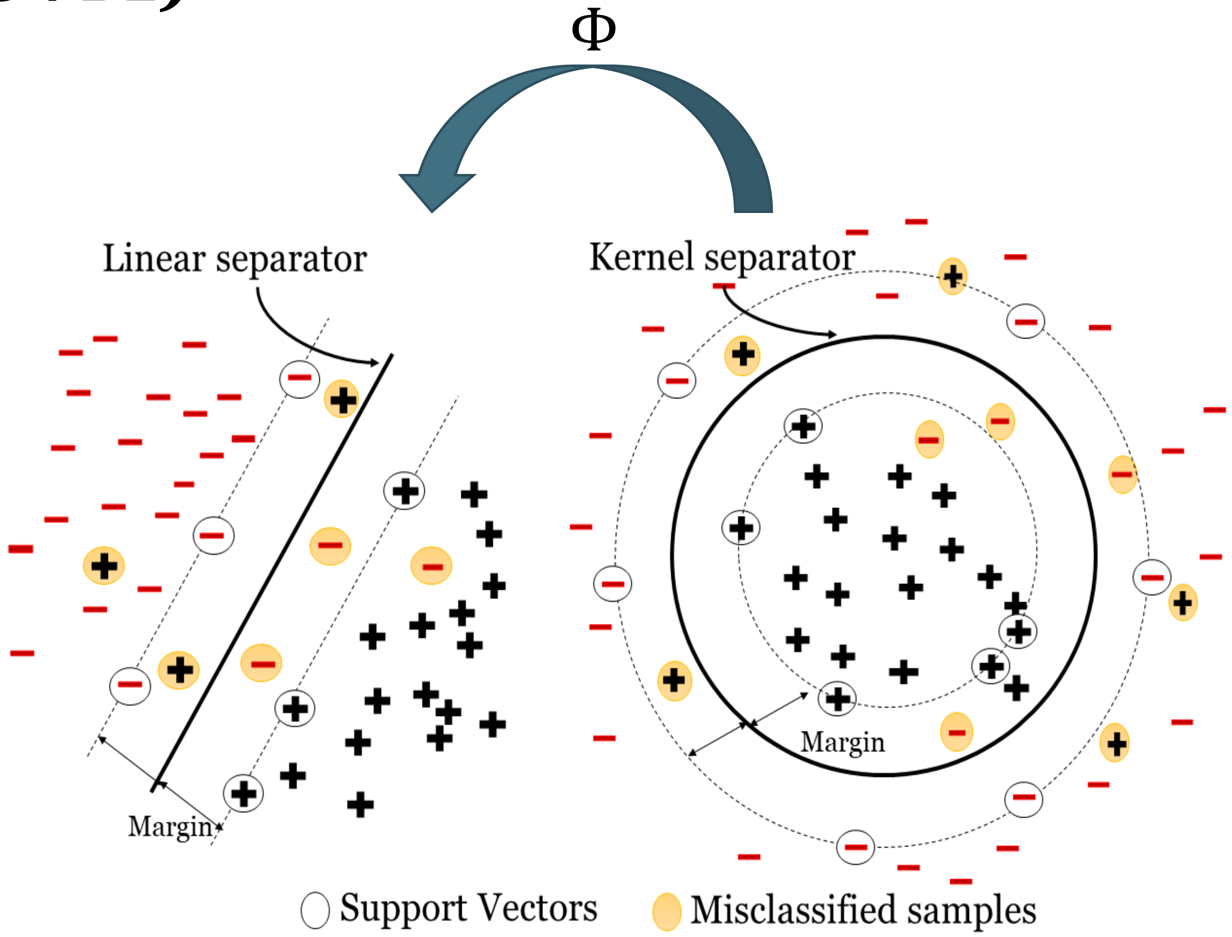
Support Vector Machines (SVM)

- Classifies with maximum margin
- Real-world data is not always linearly solvable

$$Q_{ij} = y_i y_j \Phi(x_i)\Phi(x_j)$$

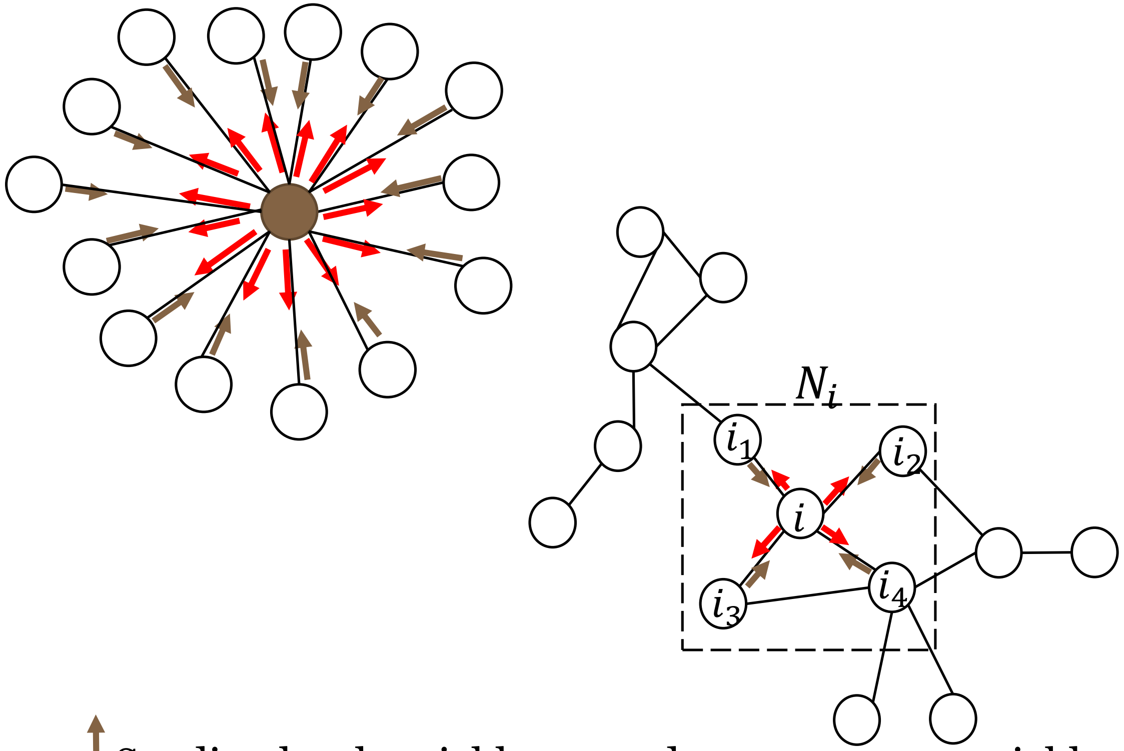
- Kernel trick

$$K(x_i, x_j) = \Phi(x_i)\Phi(x_j)$$



DISTRIBUTED CONSENSUS-BASED SVM

- SVM problems can be formulated as a distributed optimization problem
- Treated as a consensus problem
  - Centralized; global consensus
  - Decentralized; local consensus

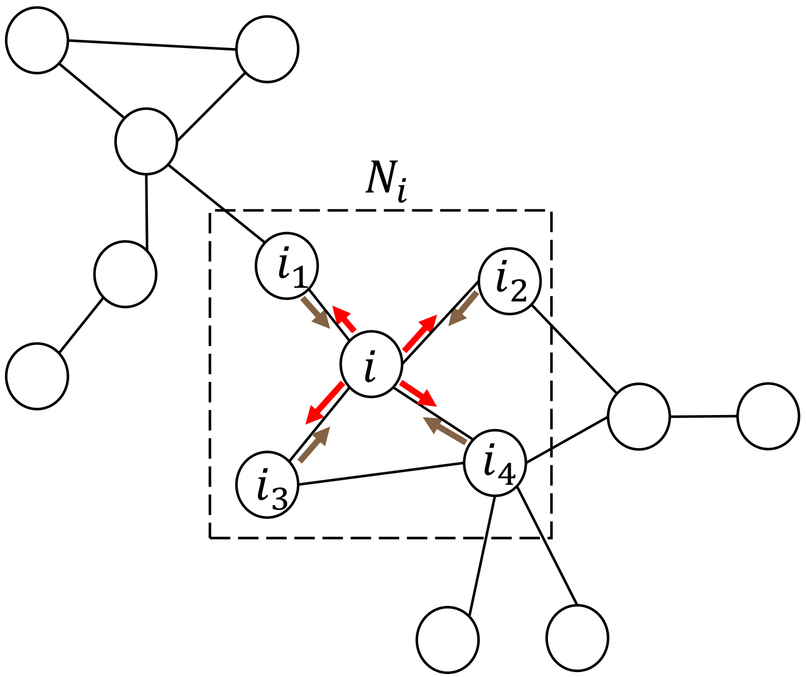


Sending local variables to update consensus variables

Broadcasting consensus variables for calculating local variables

ALTERNATING DIRECTION METHOD OF MULTIPLIERS (ADMM)

- Well-suited in decentralized settings
- Differentiability of objective function not needed
- Guarantees convergence for convex functions
- Solves the problem using iterative updates in decentralized settings



Sending local variables to update consensus variables

Broadcasting consensus variables for calculating local variables

Network Topology

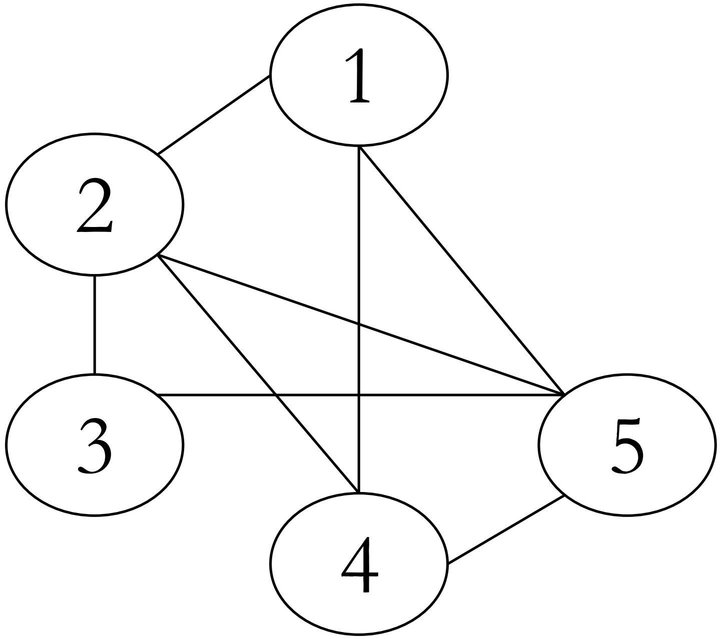
- Adjacency Matrix

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

- Laplacian Matrix

$L(A) = D - A(G)$

$$L(A) = \begin{bmatrix} 3 & -1 & 0 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 2 & 0 & -1 \\ -1 & -1 & 0 & 3 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{bmatrix}$$



$$G(V, E), V = \{1,2,3,4,5\}$$
$$E = \{(1,2), (1,4), (1,5), (2,3), (2,4), (2,5), (3,5), (4,5)\}$$

CONNECTIVITY OF A GRAPH

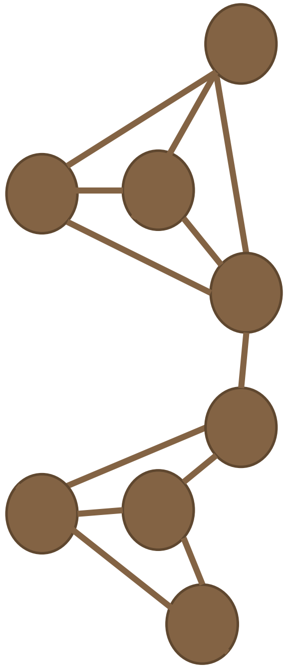
$\mu_2 = \max\{\mu'_2, |\mu'_n|\}$

- Connectivity of a graph relates to eigenvalues of A(G) or L(A)
- A(G) is symmetric and has real eigenvalues; $k_{max} = \mu'_1 \geq \mu'_2 \geq \ldots \geq \mu'_n$
- L(A) has positive eigenvalues; $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \leq 2k_{max}$
- Connectivity can be defined by graph's Spectral Gap (*SG*)
  - The *SG* regarding A(G) is

    $$SG = k_{max} - \mu_2$$

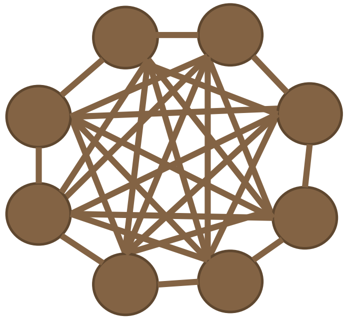
  - The *SG* regarding L(A) is

    $$SG = \lambda_2$$

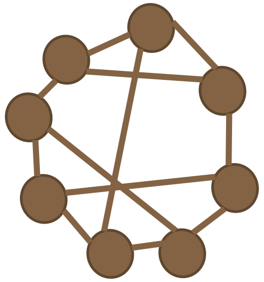


Complete Graphs and Expander Graphs

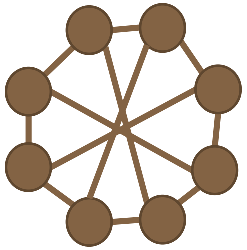
- Each node communicates with all other nodes
- Communication overhead
- Expander Graphs
  - Sparse connected graphs with low diameters
  - Any subset of nodes efficiently connects to many nodes
  - Efficient communication between nodes and high connectivity



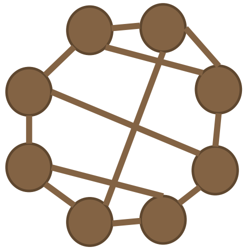
$\lambda_2 = 4$



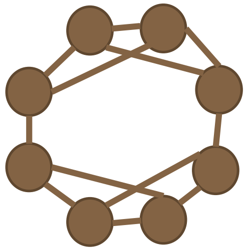
$\lambda_2 = 1.176$



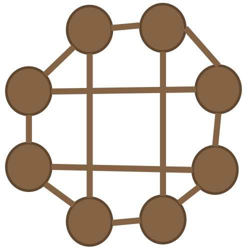
$\lambda_2 = 2$



$\lambda_2 = 1.27$

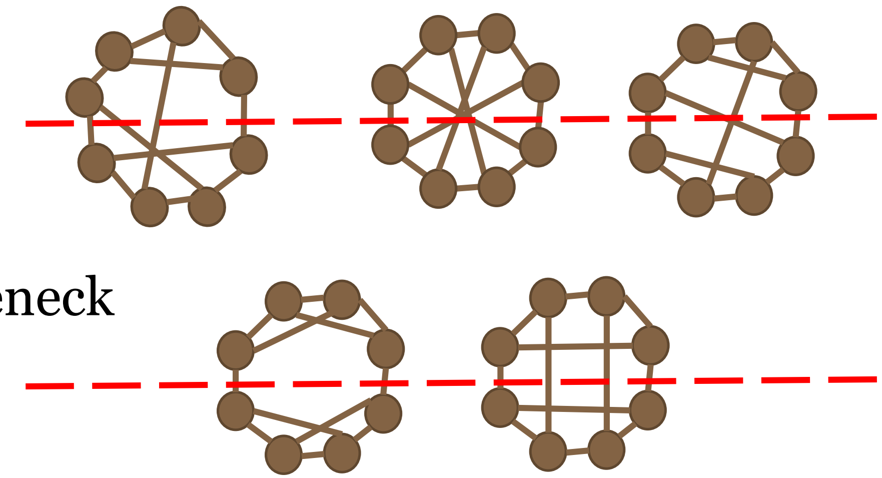


$\lambda_2 = 0.763$



$\lambda_2 = 2$

EXPANDER GRAPHS

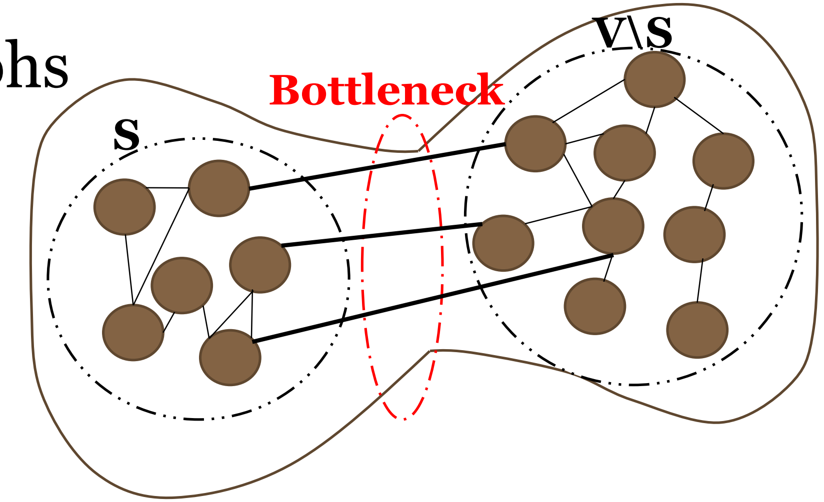


- Expansion property
  - Cheeger constant shows whether a graph has a bottleneck

$$h(G) = \min_{S \subseteq V, |S| \leq \frac{|V|}{2}} \frac{|\partial(S)|}{|S|}$$

- *d*-regular graphs, a good example of expander graphs

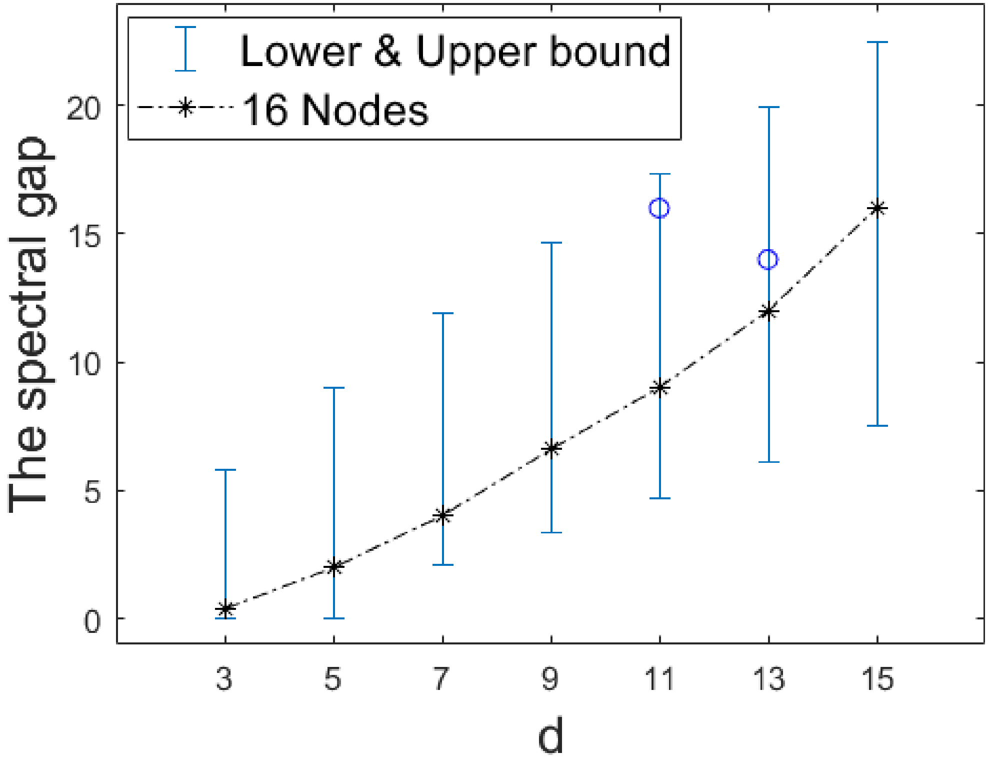
$$SG = (d - \mu_2) \text{ or } SG = \lambda_2$$

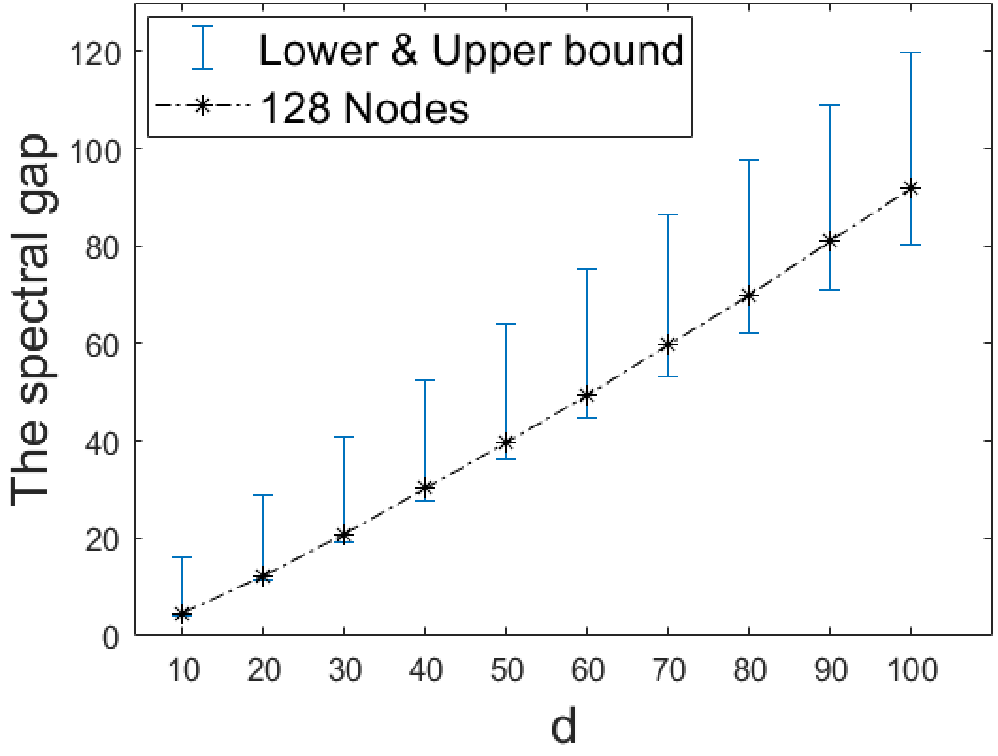


DATASETS

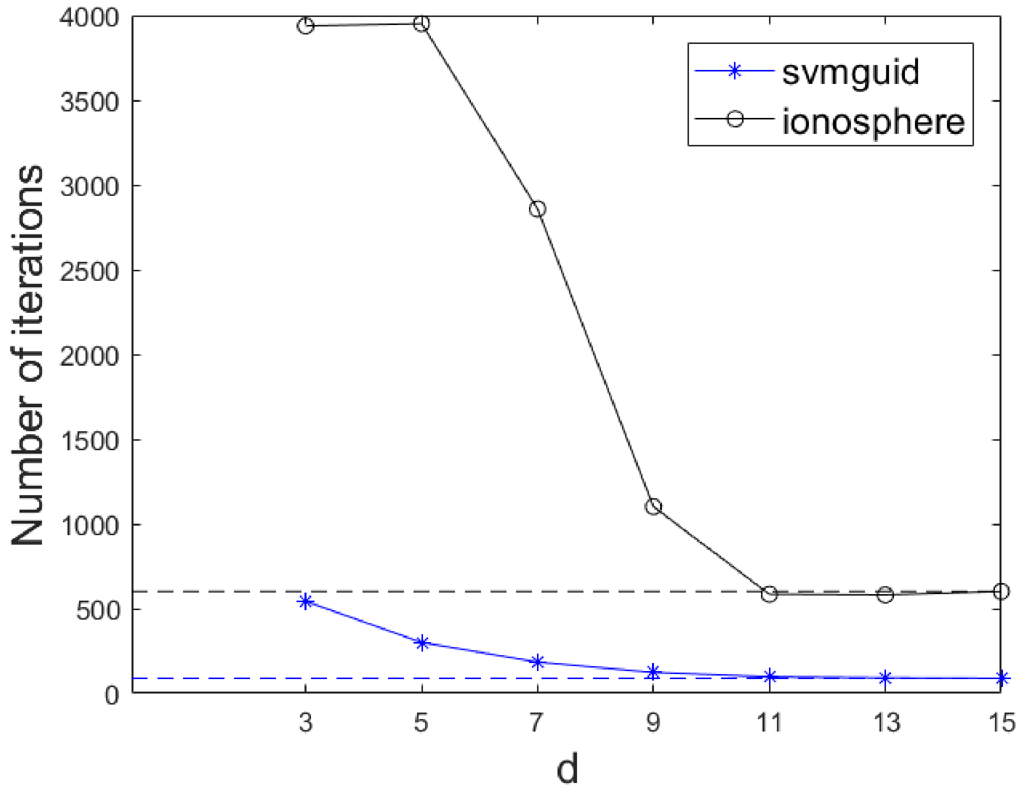
Name	Training Samples	Testing Samples	Features
ionosphere	300	51	34
svmguid	3,089	4,000	4
phishing	11,055	1,655	68
a9a	32,561	16,281	123
ijcnn	35,000	14,990	22
skinNonskin	37,492	5,000	3

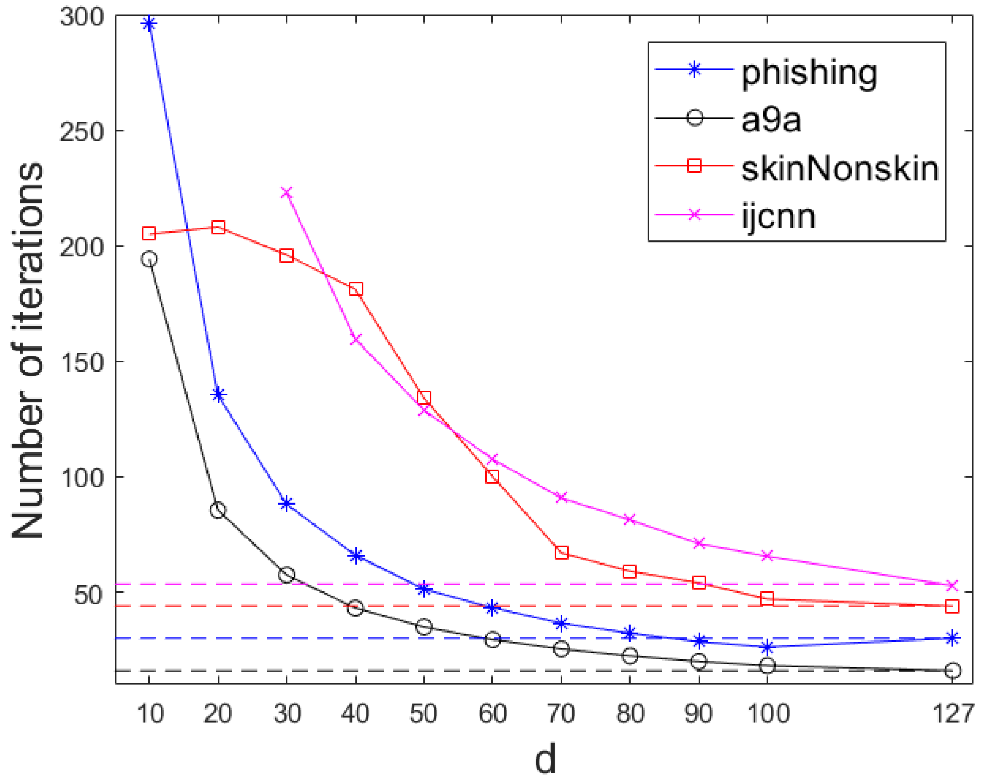
RESULTS: SPECTRAL BOUNDS FOR d-REGULAR GRAPHS



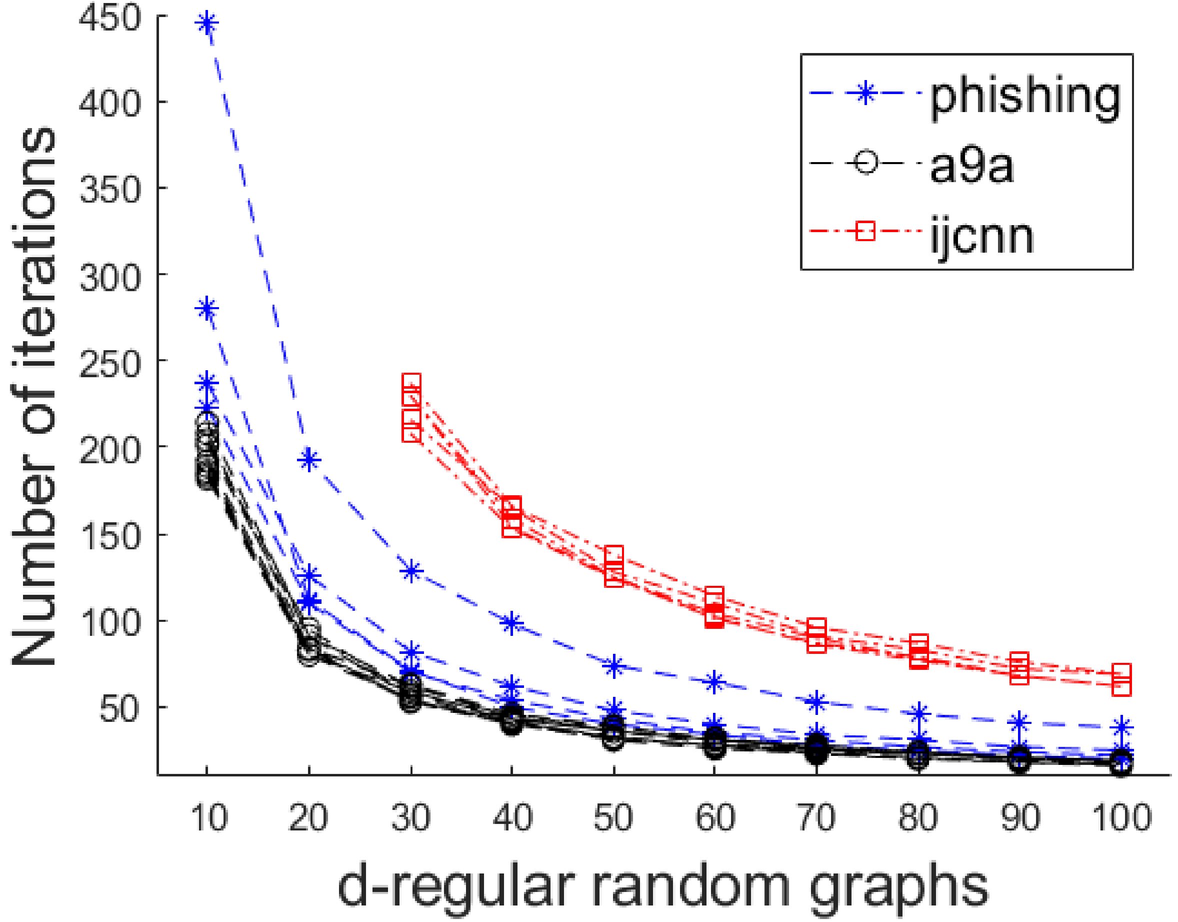


RESULTS: IMPACT OF d-REGULAR EXPANDER GRAPHS





RESULTS: SHUFFLING DATA



Ongoing Work

- Distributed parallel settings
- Application to real-world data
- Comparison of expander graphs with other types of graphs

Thanks for your attention