

# Active Semi-supervised Learning Using Sampling Theory for Graph Signals

Akshay Gadde, Aamir Anis and Antonio Ortega

University of Southern California, Los Angeles

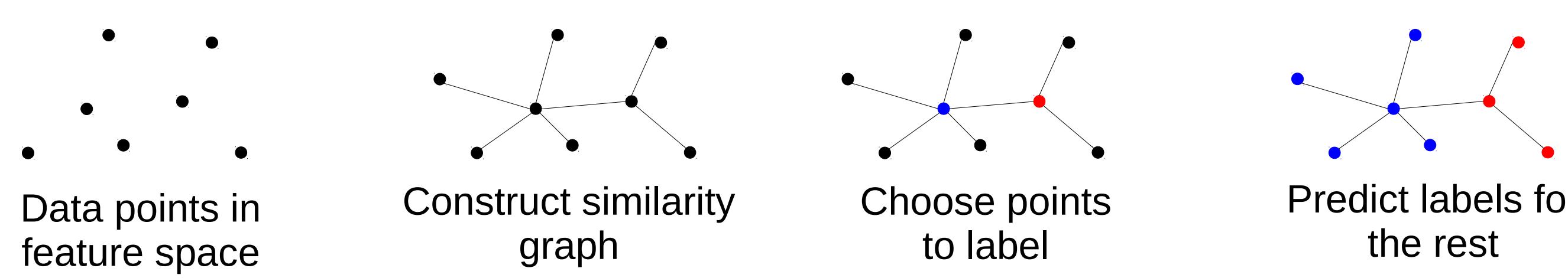
{agadde, aanis}@usc.edu, ortega@sipi.usc.edu



USC University of  
Southern California

## Motivation and Problem Definition

- Unlabeled data is abundant. Labeled data is expensive and scarce.
- **Problem setting:** Pool-based, batch-mode active SSL via graphs.



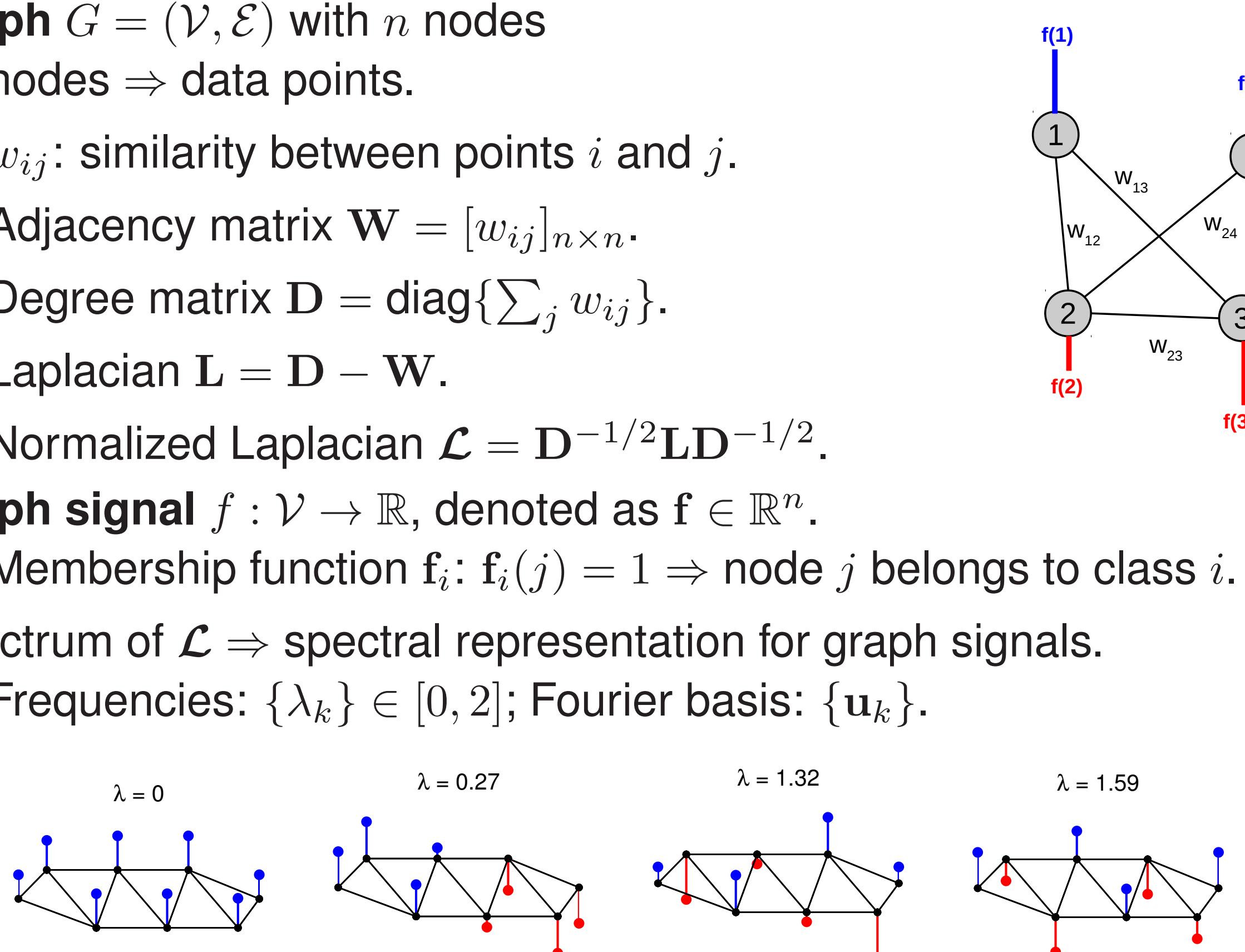
1. How to predict unknown labels from known labels?
2. What is the optimal set of nodes to label, given the learning algorithm?

## Key Ideas

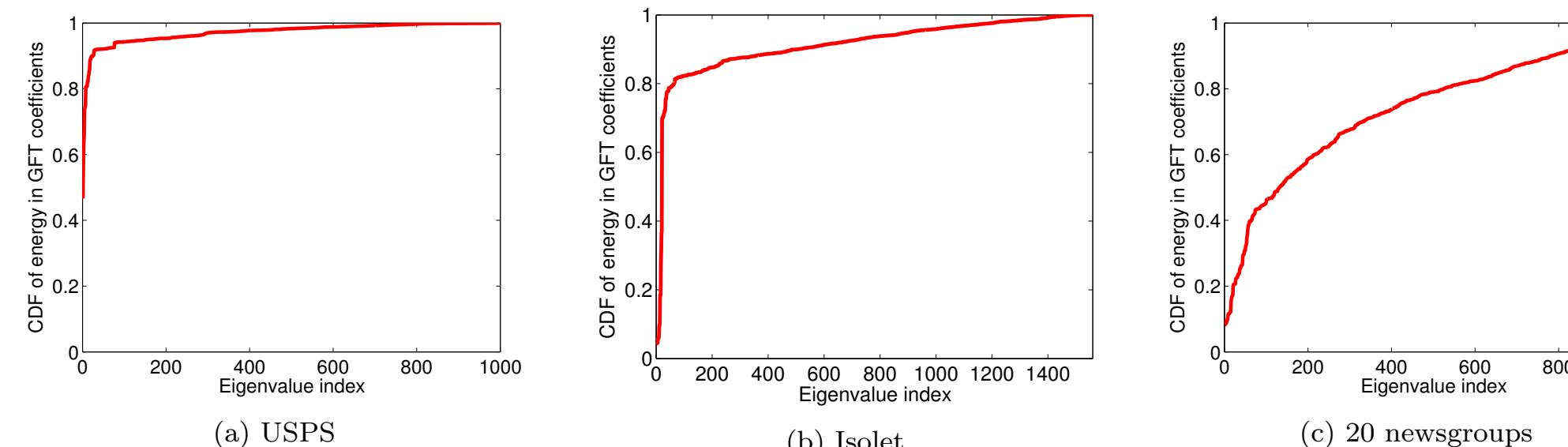
- Class labels  $\Rightarrow$  Smooth or bandlimited graph signals.
- Choosing nodes to label  $\Rightarrow$  Best sampling set selection.
- Predicting unknown labels  $\Rightarrow$  Signal reconstruction from samples.

## Background: Graph Signal Processing

- **Graph**  $G = (\mathcal{V}, \mathcal{E})$  with  $n$  nodes
  - nodes  $\Rightarrow$  data points.
  - $w_{ij}$ : similarity between points  $i$  and  $j$ .
  - Adjacency matrix  $\mathbf{W} = [w_{ij}]_{n \times n}$ .
  - Degree matrix  $\mathbf{D} = \text{diag}\{\sum_j w_{ij}\}$ .
  - Laplacian  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ .
  - Normalized Laplacian  $\mathcal{L} = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2}$ .
- **Graph signal**  $f : \mathcal{V} \rightarrow \mathbb{R}$ , denoted as  $f \in \mathbb{R}^n$ .
  - Membership function  $f_i : f_i(j) = 1 \Rightarrow$  node  $j$  belongs to class  $i$ .
- Spectrum of  $\mathcal{L} \Rightarrow$  spectral representation for graph signals.
  - Frequencies:  $\{\lambda_k\} \in [0, 2]$ ; Fourier basis:  $\{\mathbf{u}_k\}$ .

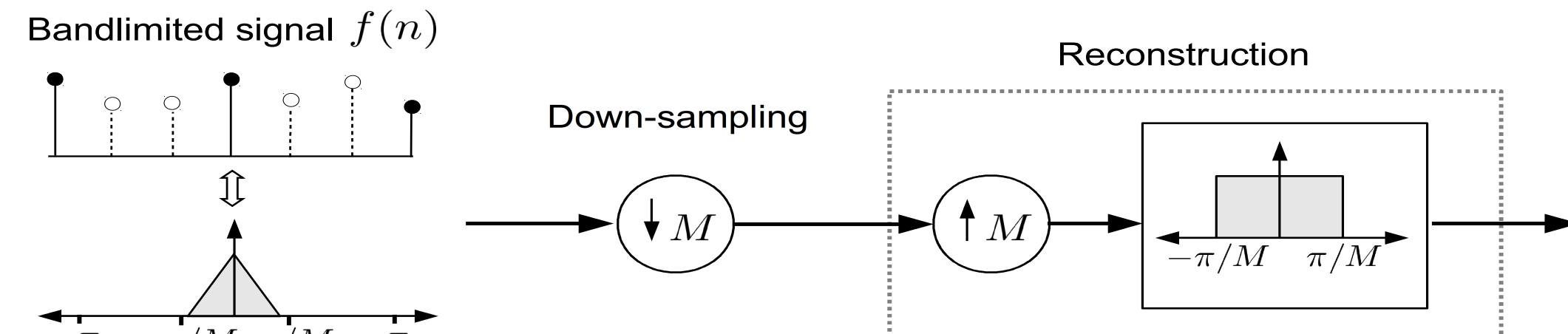


- **Graph Fourier Transform (GFT):**  $\tilde{f}(\lambda_k) = \langle f, \mathbf{u}_k \rangle$  or  $\tilde{f} = \mathbf{U}^\top f$ .
- **PW $_\omega$ (G)** : space of  $\omega$ -bandlimited graph signals
  - Support of GFT =  $[0, \omega]$ , i.e.,  $\tilde{f}(\lambda) = 0 \quad \forall \lambda > \omega$
- **Class membership functions are smooth graph signals.**

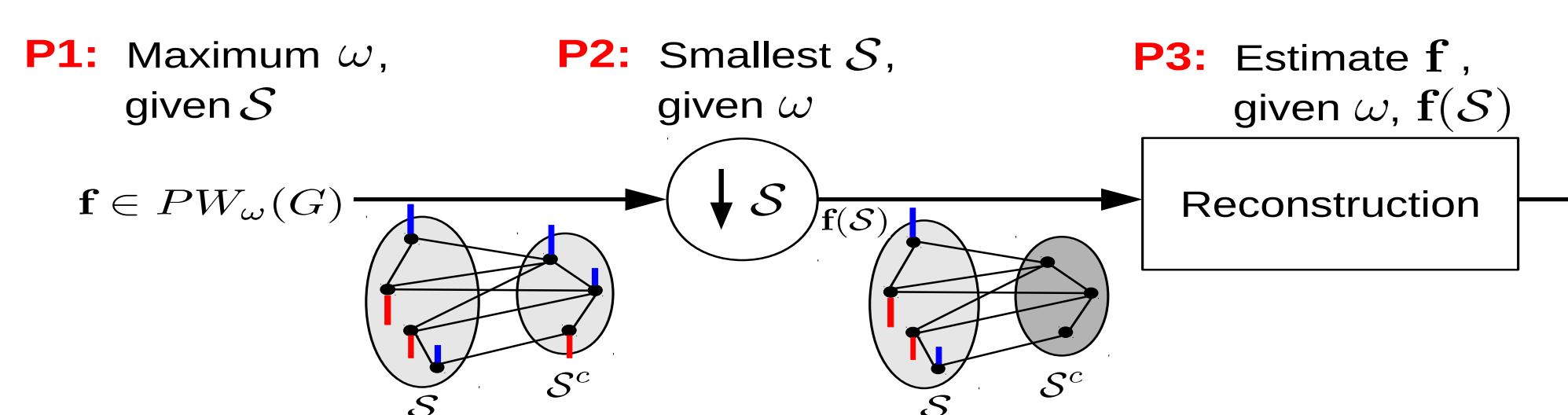


## Sampling Theory for Graph Signals

- Sampling theorem: BW  $\omega \Leftrightarrow$  sampling rate for unique representation

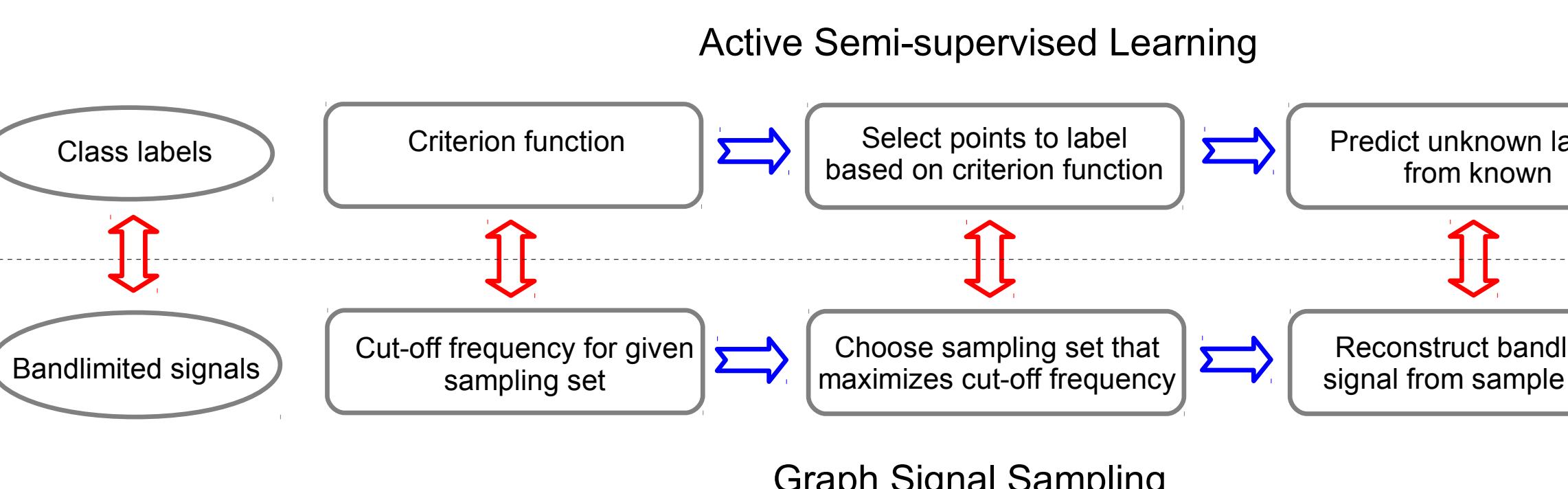


- Sampling theory for graph signals:



## Active Semi-supervised Learning

### Approach



### P1: Cut-off frequency criterion

- $L_2(\mathcal{S}^c) = \{\phi : \phi(\mathcal{S}) = 0\}$ .
- For unique sampling, we need  $PW_\omega(G) \cap L_2(\mathcal{S}^c) = \{0\}$ .
- Cut-off frequency = smallest BW that a  $\phi \in L_2(\mathcal{S}^c)$  can have.
- Estimate by  $\min_{\phi(\mathcal{S})=0} \left( \frac{\phi^\top \mathcal{L}^k \phi}{\phi^\top \phi} \right)^{1/k}$ . Higher  $k \Rightarrow$  better estimate.
- Let  $\{\sigma_{1,k}, \psi_{1,k}\}$  be the smallest eigen-pair of  $(\mathcal{L}^k)_{\mathcal{S}^c}$ .
- Thus, cut-off estimate  $\Omega_k(\mathcal{S}) = (\sigma_{1,k})^{1/k}$  with  $\phi_{\text{opt}}(\mathcal{S}^c) = \psi_{1,k}$ .

$$f \in PW_\omega(G) \quad \phi : \phi(\mathcal{S}) = 0 \quad g : g(\mathcal{S}) = f(\mathcal{S})$$

$$\boxed{\sigma_{1,k} + \psi_{1,k} = \sigma_{1,k}}$$

### P2: Sampling set selection

- For given size, sampling set must be able to maximally capture signal information;  $\mathcal{S}_{\text{opt}} = \arg \max_{|\mathcal{S}|=m} \Omega_k(\mathcal{S})$ .

$$(\Omega_k(\mathcal{S}))^k = \min_{\phi(\mathcal{S})=0} \frac{\phi^\top \mathcal{L}^k \phi}{\phi^\top \phi} \approx \min_{\mathbf{x}} \left( \frac{\mathbf{x}^\top \mathcal{L}^k \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} + \alpha \frac{\mathbf{x}^\top \text{diag}(\mathbf{t}) \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \right) \Big|_{\mathbf{t}=1_{\mathcal{S}}} = \lambda_k^\alpha(\mathbf{t})|_{\mathbf{t}=1_{\mathcal{S}}}$$

binary relaxation

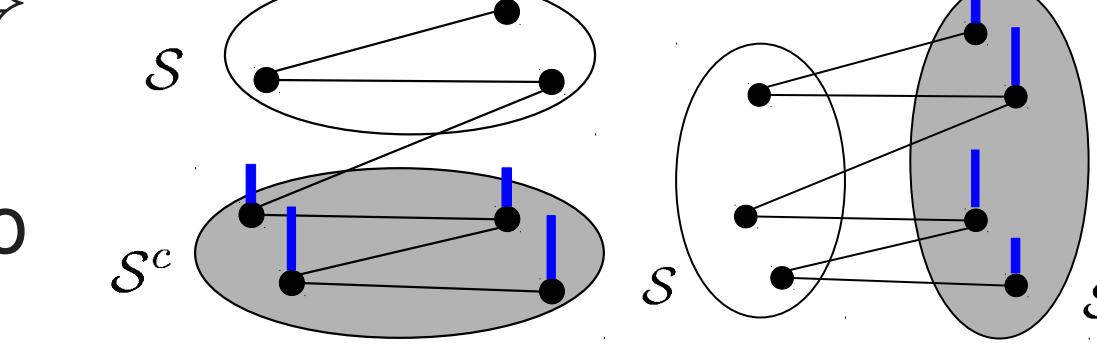
relax the constraint

- $\mathbf{x}_{\text{opt}} \approx \phi_{\text{opt}} \Rightarrow \frac{d\lambda_k^\alpha(\mathbf{t})}{dt(i)} \approx \alpha(\phi_{\text{opt}}(i))^2$ .
- **Greedy algorithm:**  $\mathcal{S} \leftarrow \mathcal{S} \cup v$ , where  $v = \arg \max_j (\phi_{\text{opt}}(j))^2$

## Active Semi-supervised Learning (continued)

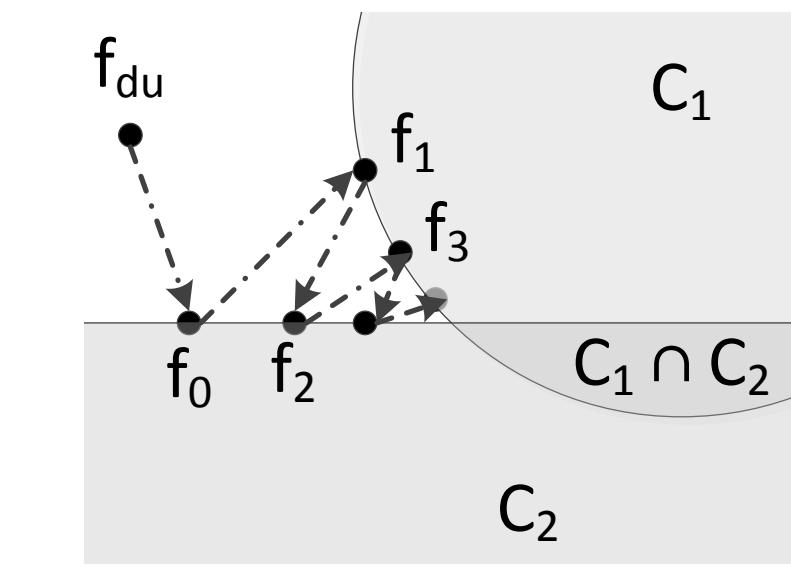
### Maximizing Cut-off Frequency $\Rightarrow$ Active Learning

- Cut-off  $\Omega_k(\mathcal{S}) \equiv$  variation of smoothest signal  $\phi_{\text{opt}}$  in  $L_2(\mathcal{S}^c)$ .
- Larger cut-off  $\Rightarrow$  more variation in  $\phi_{\text{opt}} \Rightarrow$  more cross-links.
- Unlabeled nodes are strongly connected to labeled nodes!



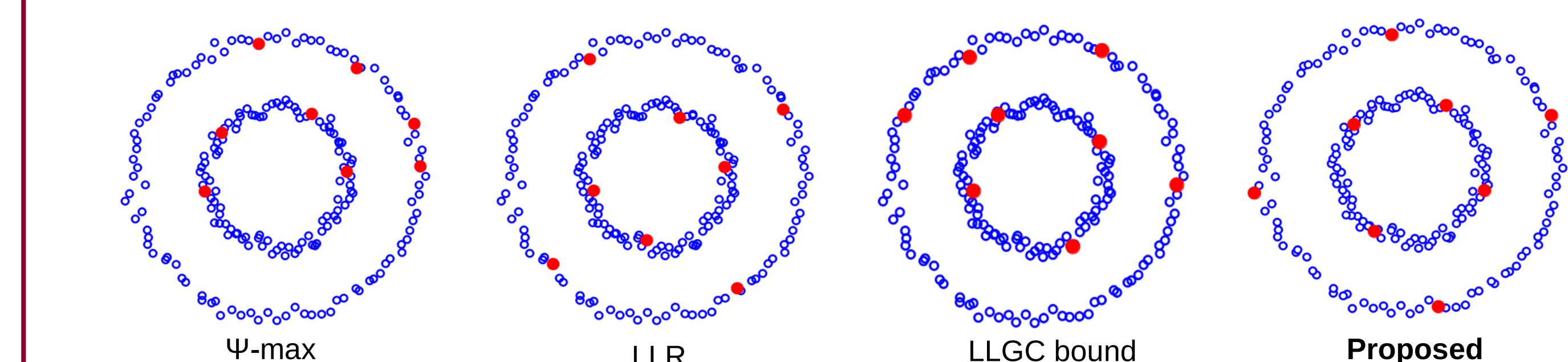
### P3: Label prediction by reconstruction

- $\mathcal{C}_1 = \{x : x(\mathcal{S}) = f(\mathcal{S})\}$  and  $\mathcal{C}_2 = PW_\omega(G)$ .
- Solution:  $f \in \mathcal{C}_1 \cap \mathcal{C}_2$ , sampling theory  $\Rightarrow$  unique  $f$ .
- POCS:  $f_{i+1} = \mathbf{P}_{\mathcal{C}_2} \mathbf{P}_{\mathcal{C}_1} f_i$ , where  $f_0 = [f(\mathcal{S})^\top, 0]^\top$ .
  - $\mathbf{P}_{\mathcal{C}_1}$  resets the samples on  $\mathcal{S}$  to  $f(\mathcal{S})$ .
  - $\mathbf{P}_{\mathcal{C}_2} = \mathbf{U}h(\Lambda)\mathbf{U}^\top$  sets  $\tilde{f}(\lambda) = 0$  if  $\lambda > \omega$ .
- $\mathbf{P}_{\mathcal{C}_2} \approx \sum_{i=1}^n \left( \sum_{j=0}^p a_j \lambda_i^j \right) \mathbf{u}_i \mathbf{u}_i^\top = \sum_{j=0}^p a_j \mathcal{L}^j \rightarrow p$ -hop localized

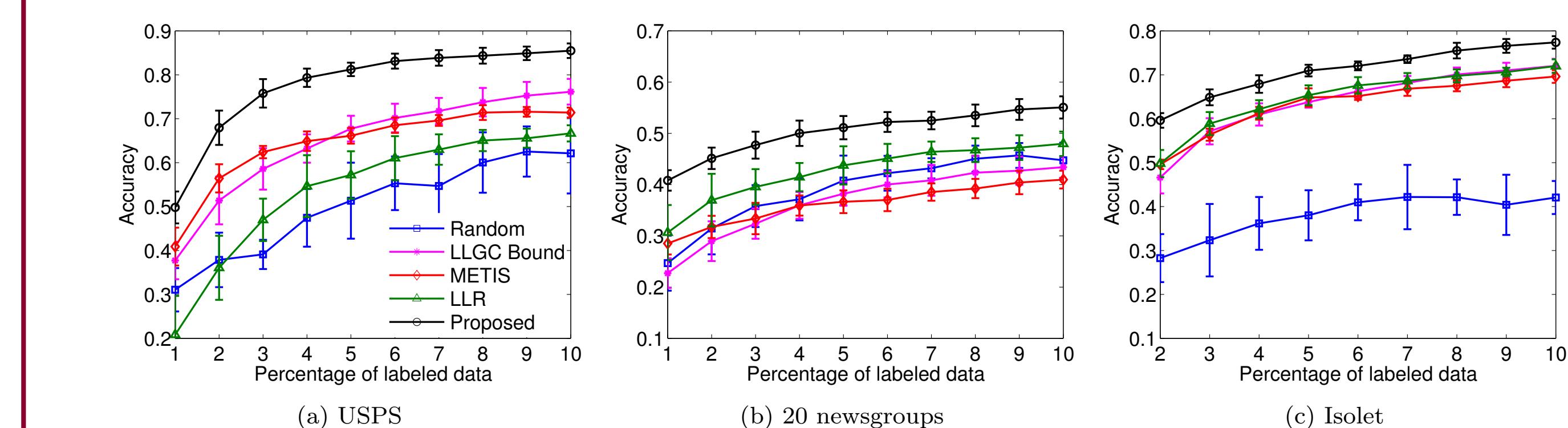


## Results

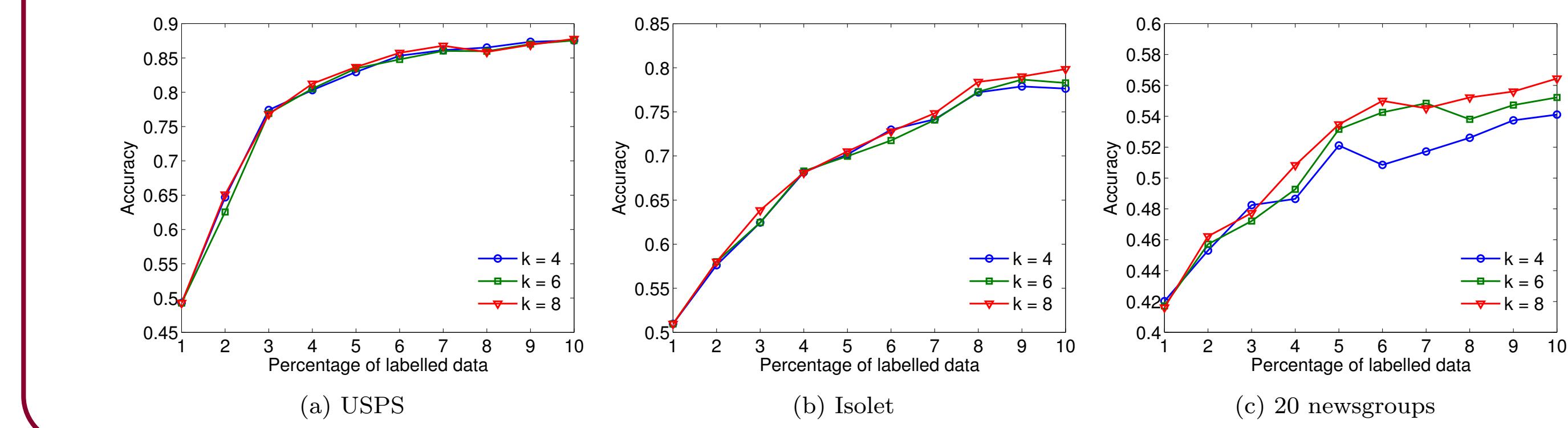
### Toy example:



### Real datasets:



### Effect of k:



## References