Decoupled smoothing on graphs

Alex Chin, ***Yatong Chen***, Kristen M. Altenburger, Johan Ugander

Stanford University

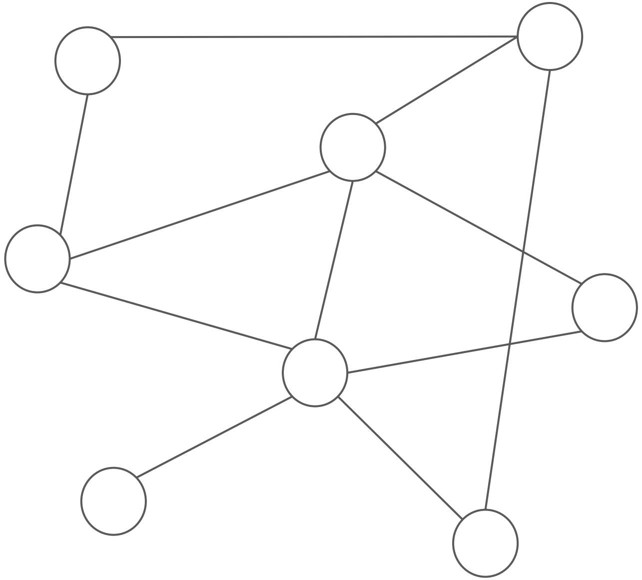
The Web Conference, San Francisco
May 17, 2019

Attribute prediction on graphs

Attribute prediction on graphs

Given:

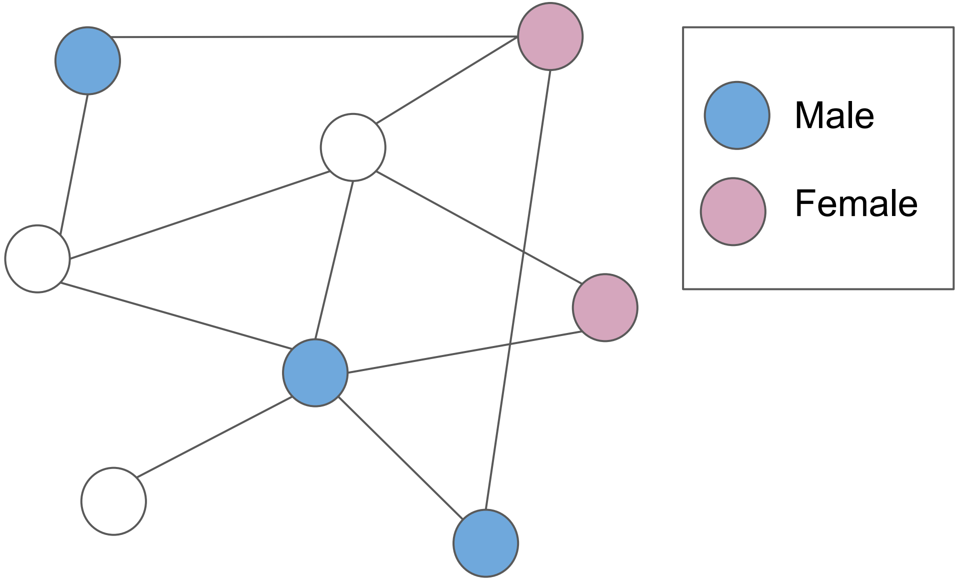
- Social Network $G = (V, E)$



Attribute prediction on graphs

Given:

- Social Network $G = (V, E)$
- Labels for some subset nodes



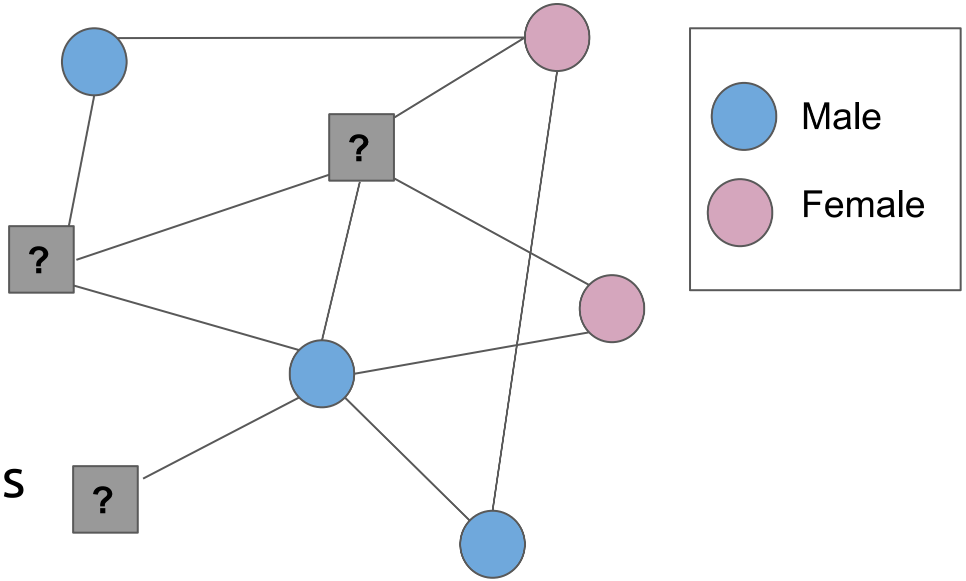
Attribute prediction on graphs

Given:

- Social Network $G = (V, E)$
- Labels for some subset nodes

Goal:

- Infer labels for unlabeled nodes



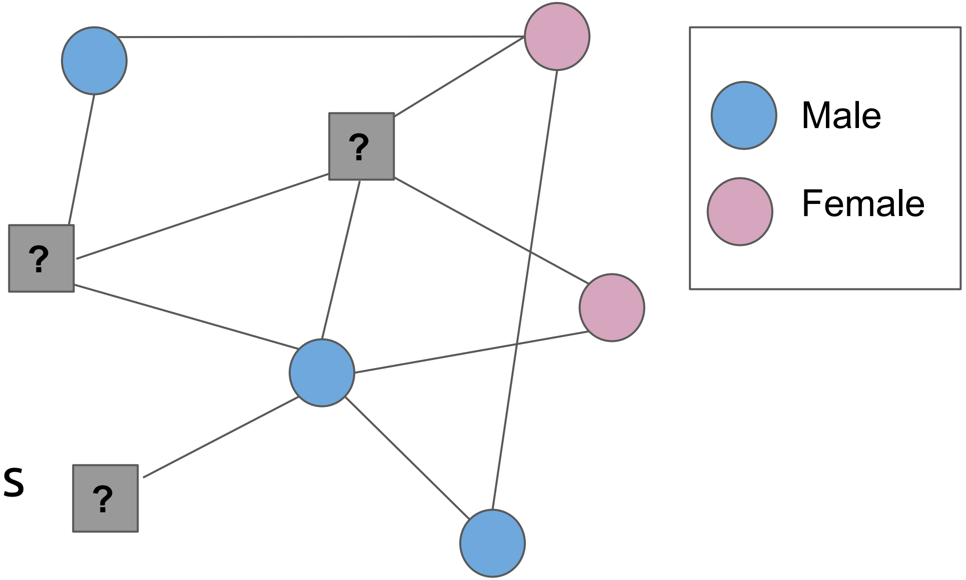
Semi-supervised learning Problem

Given:

- Social Network $G = (V, E)$
- Labels for some subset nodes

Goal:

- Infer labels for unlabeled nodes



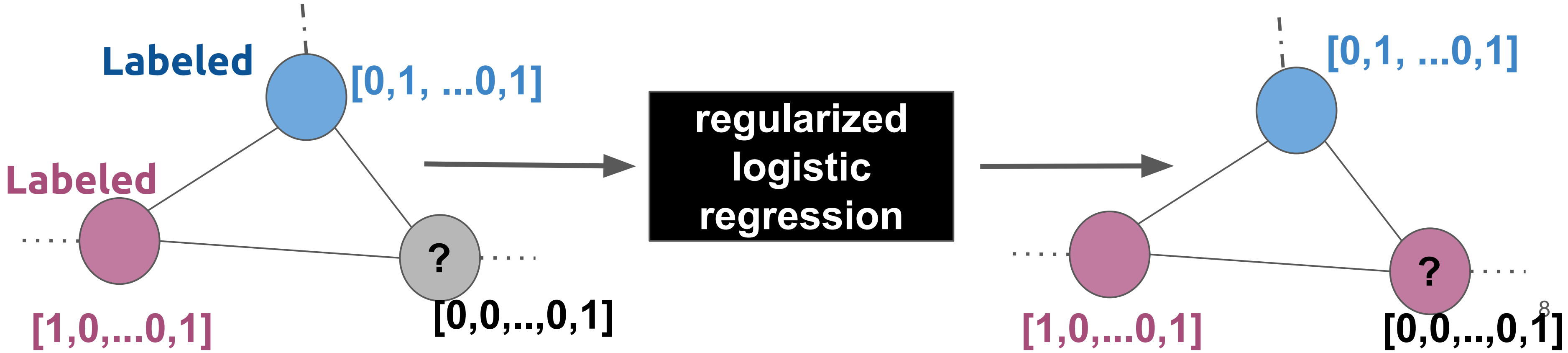
Approaches for attribute prediction

- **Approach 1:** Graph Smoothing based on Gaussian Random Field *[Zhu, Ghahramani, Lafferty 2003]*
  - **Assumption: Gaussian Markov Random Field Prior on true label of all the nodes** $\theta \sim N(0, \tau^2(D - \gamma A)^{-1}) \in R^n$
  - Get the Bayes estimator of θ on unlabeled nodes under the GMRF prior
  - Be referred as **ZGL** later

Approaches for attribute prediction

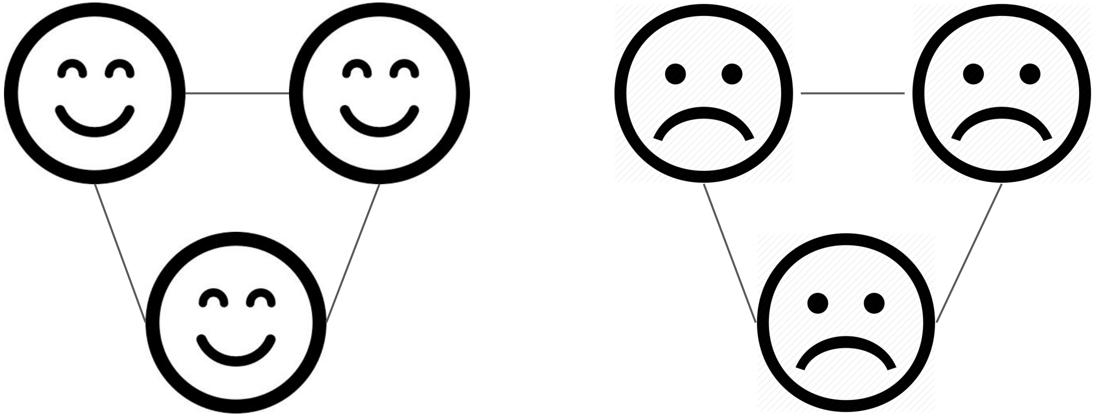
- Approach 2: **LINK** classification *[Lu-Getoor 2003]*
  - Learn a function F:

    F(row i in adjacency matrix) = i's label
  - Example F: regularized logistic regression



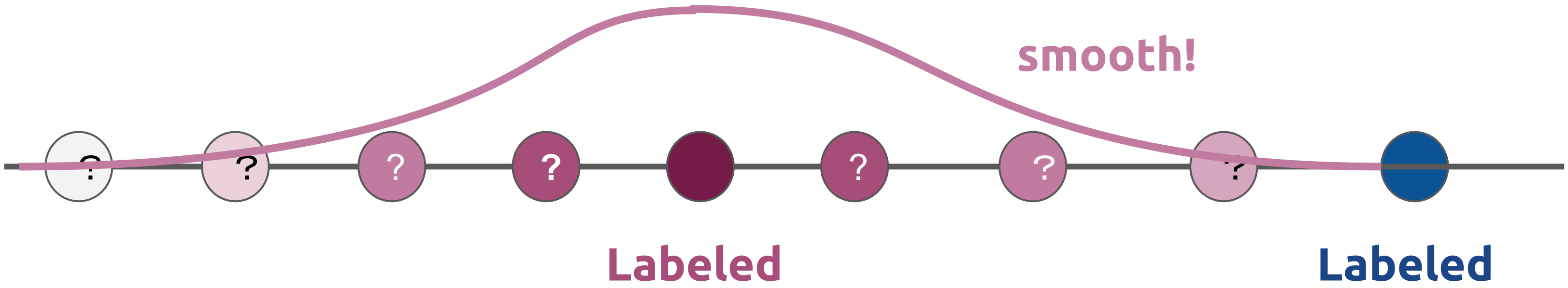
ZGL’s assumption: Homophily

- Homophily [one-hop similarity]:
    - Individuals are similar to their friends



ZGL's assumption: Homophily

- Homophily [one-hop similarity]:
  - ZGL assumes information of a given node decays **smoothly** across the topology of the graph (by imposing the GMRF prior)

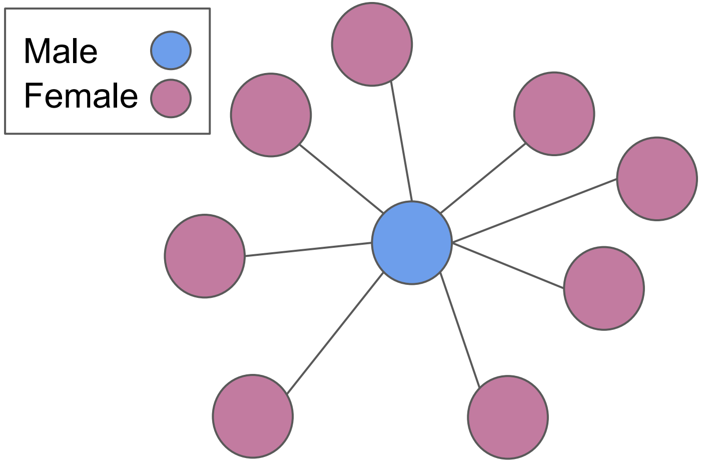


Homophily Assumption: NOT always necessary!

- *[Altenburger-Ugander 2018]:*
  - **LINK** does well even **without** assuming homophily
  - Homophily assumption is **not necessary** for inference to succeed
- ...but **ZGL** and a lot of other graph smoothing methods all assumes homophily. Can we do graph smoothing without it?
  - Yes (this talk)

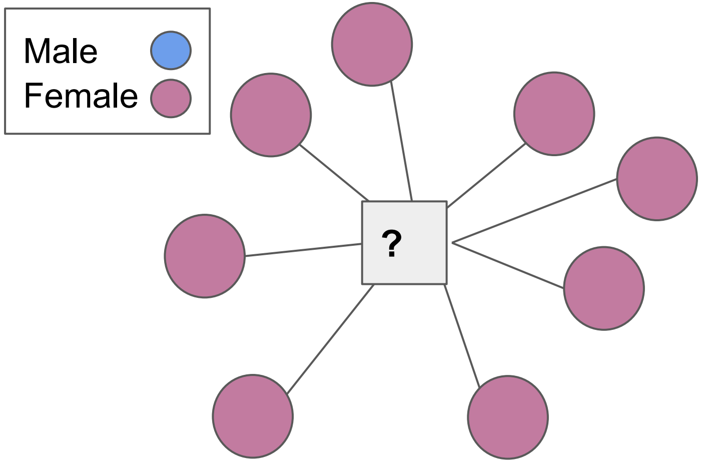
A situation where Homophily assumption fails

- Gender example:



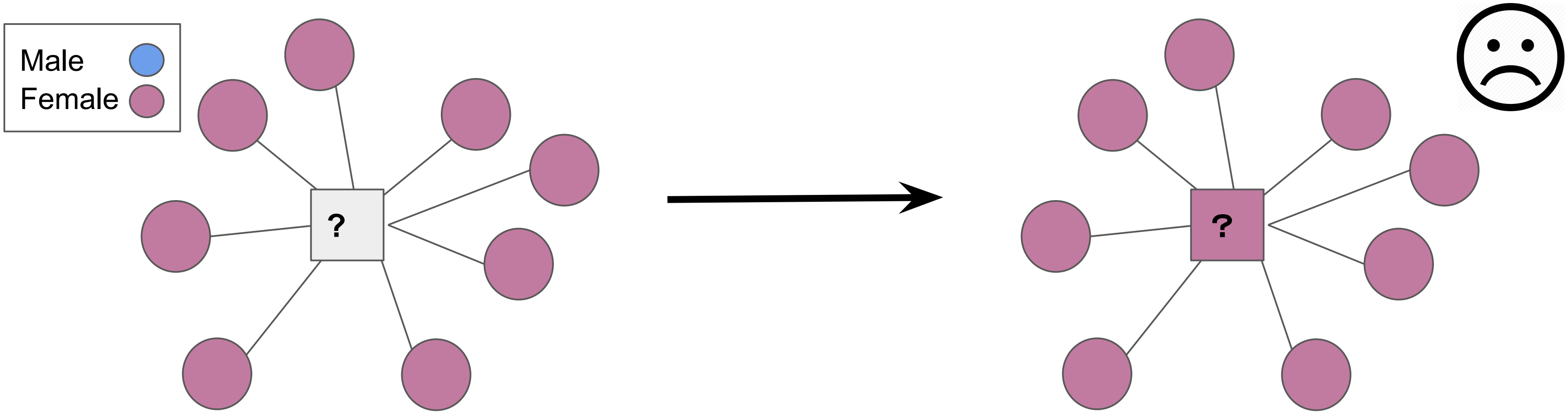
A situation where Homophily assumption fails

- **Gender example:**
  - Want to predict the gender of the center node



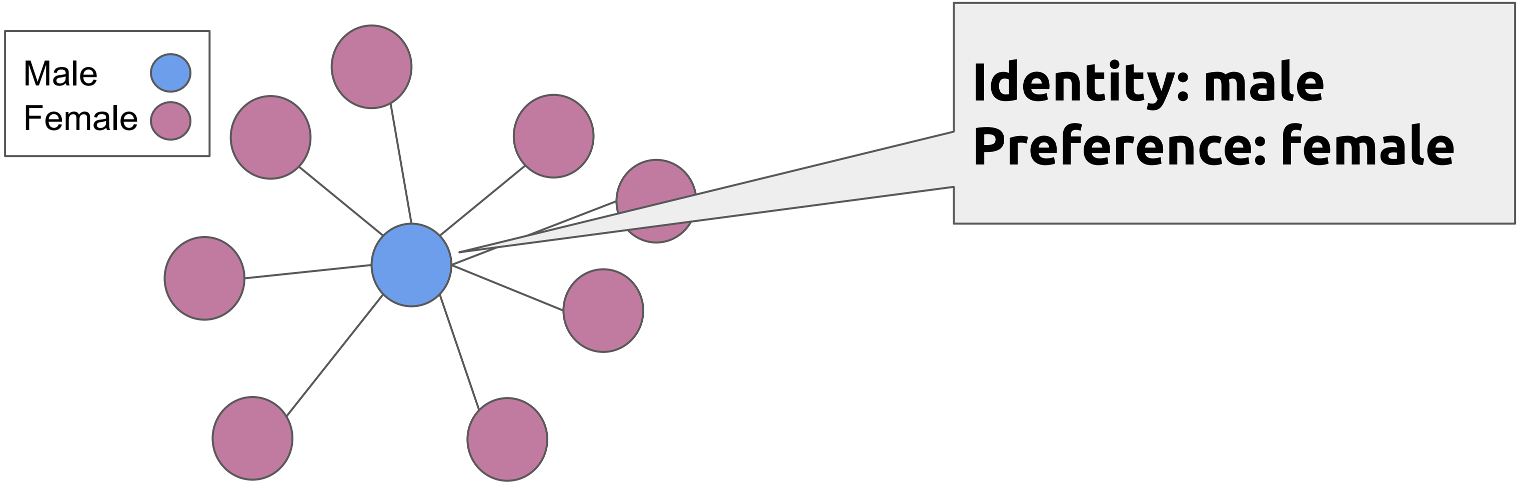
A situation where Homophily assumption fails

- **Gender example:**
  - Want to predict the gender of the center node:
    - Assume Homophily (1-hop majority vote): false



A situation where Homophily assumption fails

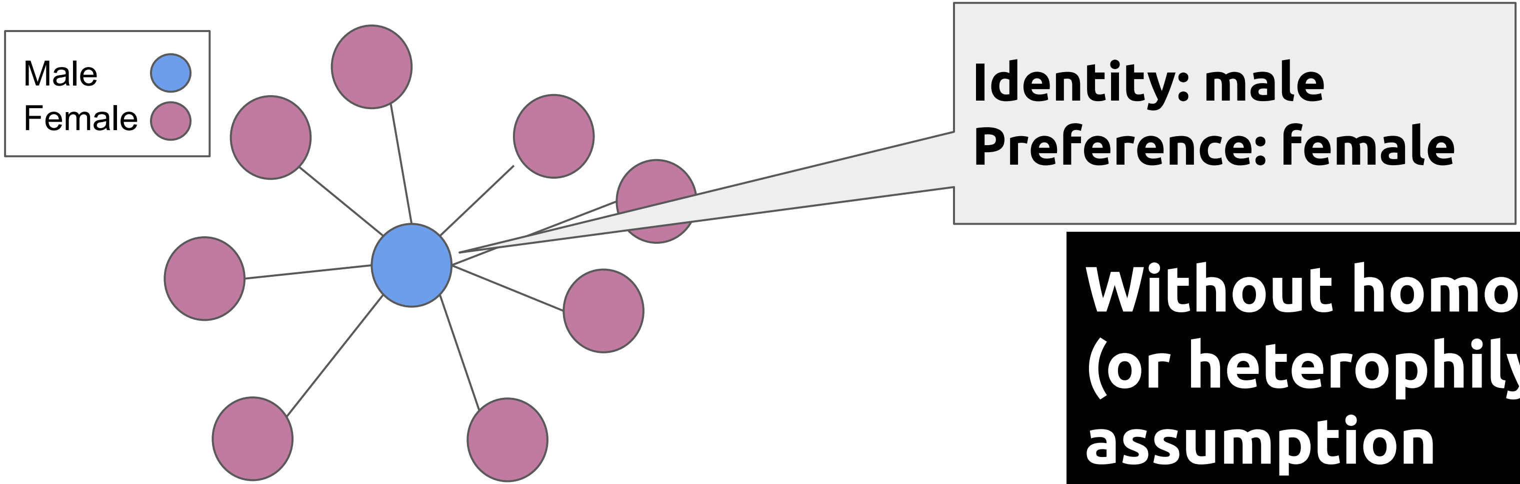
- **Observation:** there are difference between one’s identity and preference



Decoupled smoothing Method

Decoupled smoothing: idea

- **Idea:** ***decoupling*** one's "identity" θ and "preference" ϕ
- Use separate parameters to model them accordingly!

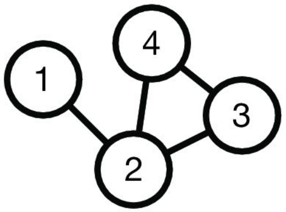


Without homophily (or heterophily) assumption

Decoupled smoothing: idea

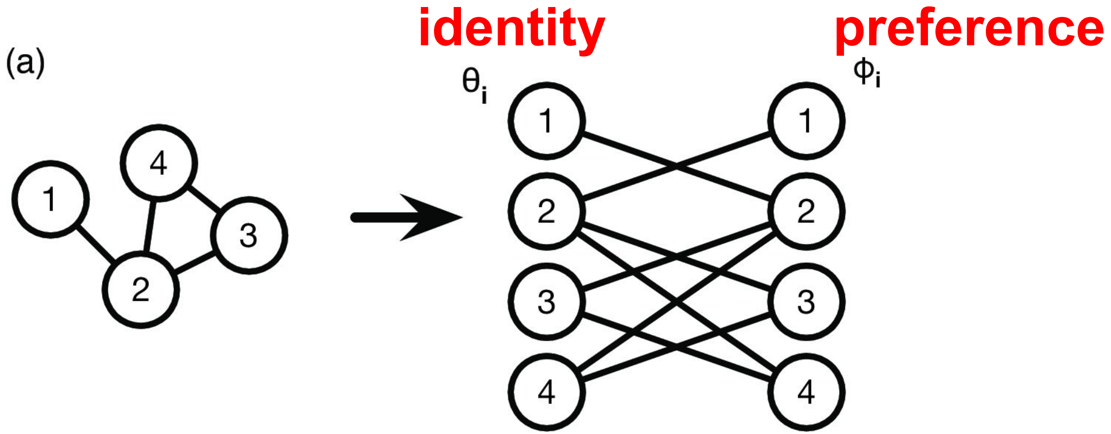
- Idea: *decoupling* one's "identity" θ and "preference" ϕ

(a)



Decoupled smoothing: idea

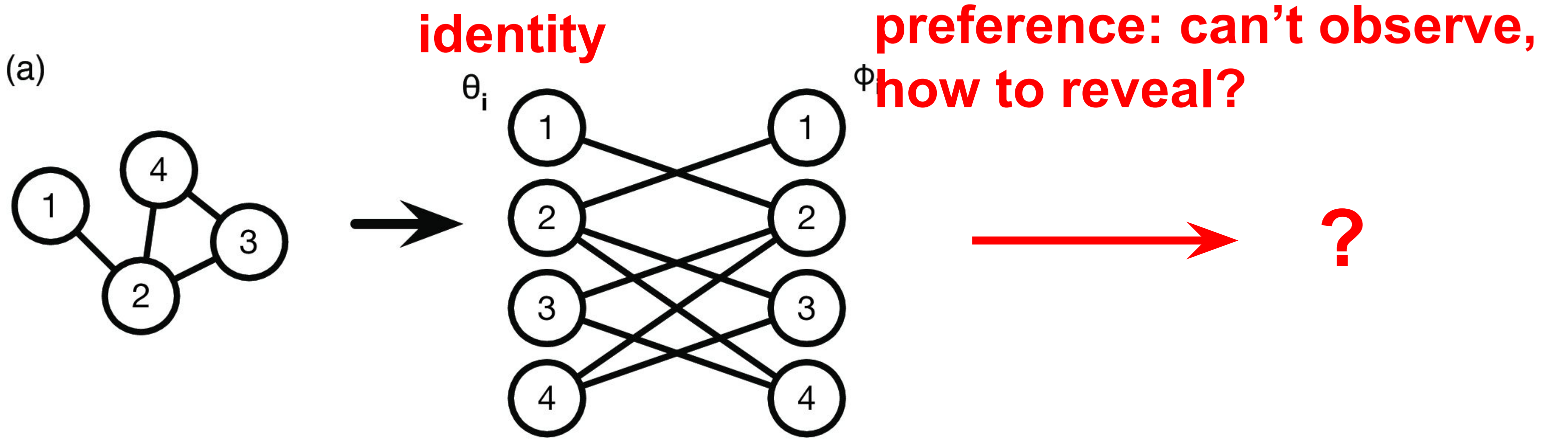
- Idea: *decoupling* one's "identity" θ and "preference" ϕ



- **Intuition:** a person's identity will reveal information about their friend's preference, and vice versa

Decoupled smoothing: idea

- Idea: *decoupling* one's "identity" θ and "preference" ϕ

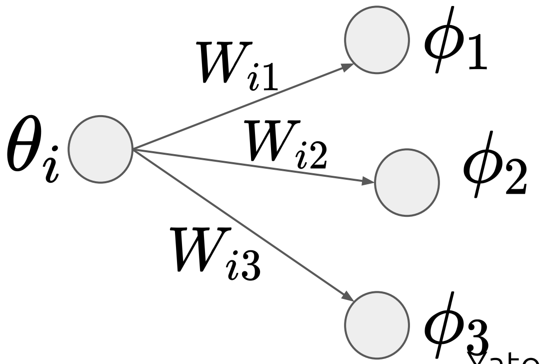


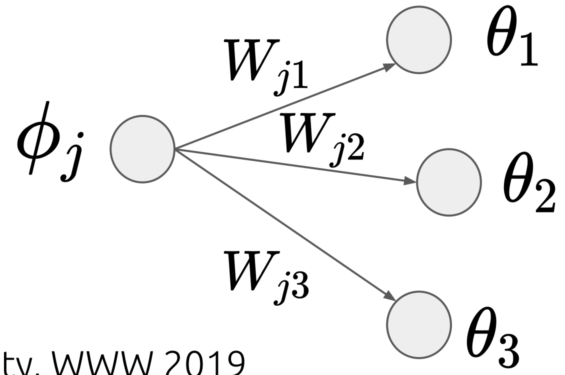
- **Intuition:** a person's identity will reveal information about their friend's preference, and vice versa

Decoupled smoothing: model

- **Intuition:** a person's identity will reveal information about their friend's preference, and vice versa
- **Assumption:**

$$\theta_i|\phi \approx N(\textstyle\sum_{j=1}^{n} W_{ij}\phi_j, \sigma_i^2) \qquad \phi_j|\theta \approx N(\textstyle\sum_{i=1}^{n} W_{ij}\theta_i, \sigma_j^2)$$





Decoupled smoothing: model

- **Intuition:** a person's identity will reveal information about their friend's preference, and vice versa
- **Assumption:**

$$\theta_i|\phi \approx N(\sum_{j=1}^{n} W_{ij}\phi_j, \sigma_i^2) \qquad \phi_j|\theta \approx N(\sum_{i=1}^{n} W_{ij}\theta_i, \sigma_j^2)$$

- **Goal:** to obtain predictions for the identity θ
  - *the preference* ϕ *is nuisance!*
- **Get the marginal prior for** θ **:**
  - $\boxed{\theta \sim N(0, \tau^2(Z - \gamma^2 W Z'^{-1} W^T)^{-1})}$

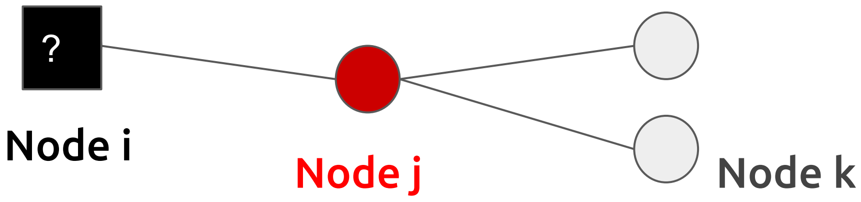
Decoupled smoothing: impose a prior

$$\theta \sim N(0, \tau^2 (Z - \gamma^2 W Z'^{-1} W^T)^{-1})$$

How to estimate W?

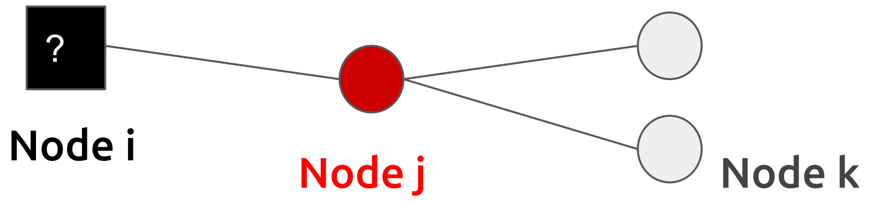
How to estimate W?

- **Intuition:**
  - Node j'th **preference** will imply a **2-hop similarity** between node i and node k's **identities**



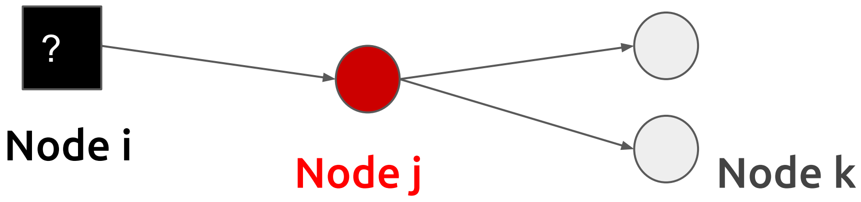
How to estimate W?

- **Intuition:**
  - Node j'th **preference** will imply a **2-hop similarity** between node i and node k's **identities**
  - **Make use of the information of k when predicting i**



How to estimate W?

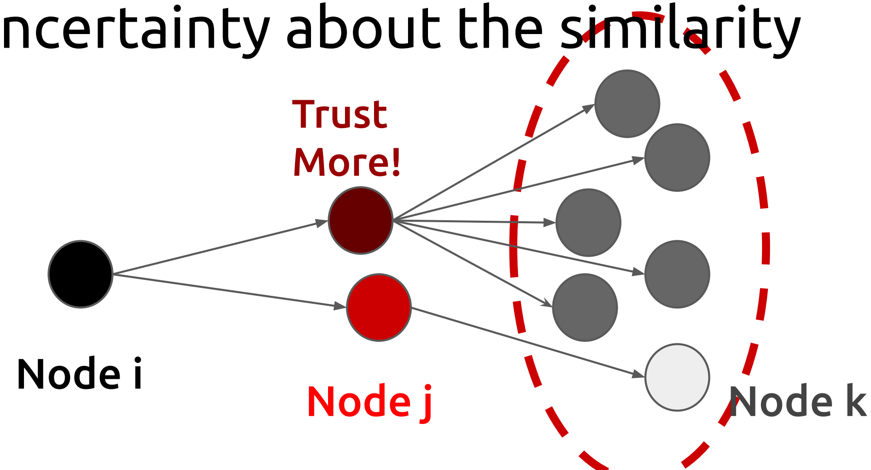
- **Intuition:**
  - Node j'th **preference** will imply a **2-hop similarity** between node i and node k's **identities**
  - **Make use of the information of k when predicting i**
- **Assumption:**
  - i's 2-hop friend k has the distribution: $\theta_k \sim N(\theta_i, \sigma_j^2)$



How to estimate W?

- **Why θ_i as mean:**
  - Similarity among i and k
- **Why σ_j as variance?**
  - The more friends j have$\longrightarrow$ the better its preference being revealed $\rightarrow$ the less uncertainty about the similarity between i and k

$$\theta_k \sim N(\theta_i, \sigma_j^2)$$



How to estimate W?

- **Assumption:**
  - 2-hop friend k has the distribution $\theta_k \sim N(\theta_i, \sigma_j^2)$
  - Homogeneous standard error $\sigma_j^2 = \sigma / d_j^2$

- Then W can be reduced to $W_{ij} = A_{ij} / \sigma_j^2$

We get W!

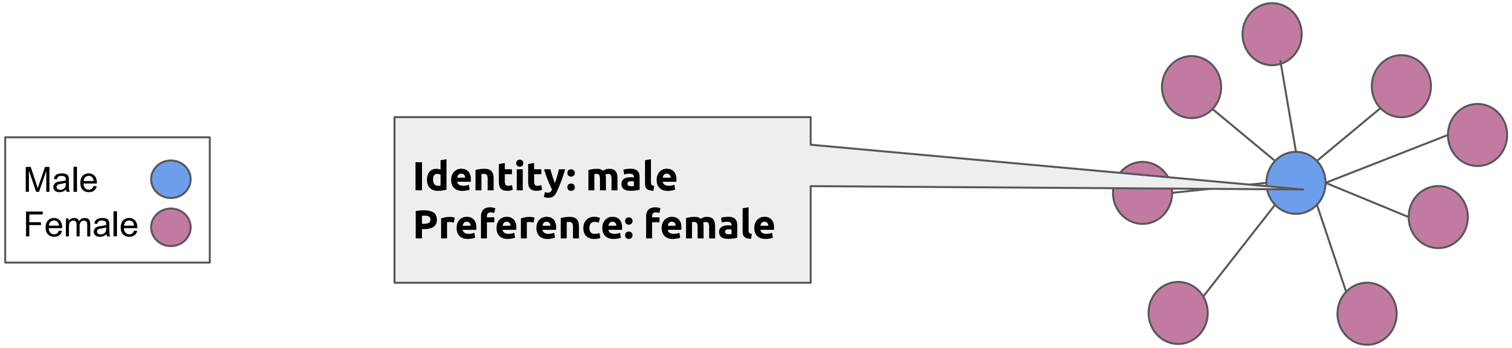
Decoupled smoothing: model

- Now we know everything about the marginal prior for θ:
  - $\theta \sim N(0, \tau^2(Z - \gamma^2 W Z'^{-1} W^T))$
- Next step:
  - Compute the Bayes estimator of θ for unlabeled node and then make the prediction (recall ZGL)
- Done!

Relationship between Decoupled smoothing and some phenomenon/concept/method that are related to it

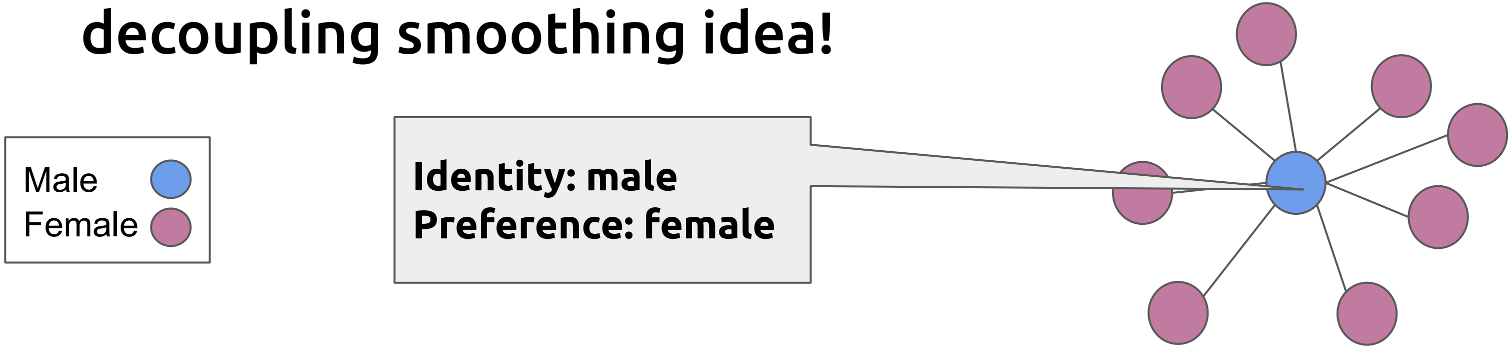
Decoupled smoothing and Monophily

- The phenomenon of **Monophily** *[Altenburger-Ugander 2018]*
  - Two-hop similarity: individuals are similar to their friends' friends
  - Innovative concept compared to Homophily



Decoupled smoothing implies Monophily

- The phenomenon of **Monophily** *[Altenburger-Ugander 2018]*
  - Two-hop similarity:
    - similarity among the friends of a person is the result of personal preference
  - **The 2 hop similarity phenomenon is implied by our decoupling smoothing idea!**



Decoupled smoothing and 2-hop MV

- **Decoupled smoothing reduces to iterative 2-hop majority vote** (under homogeneous standard error):

$$\tilde{A}_{ij} = \sum_k A_{ik} A_{jk} / (d_k \sigma_k^2)$$

$$\hat{\theta}_i^t = (Z^{-1} \tilde{A} \hat{\theta}^{t-1})_i = \frac{1}{z_i} \sum_{j=1}^n \tilde{A}_{ij} \hat{\theta}_j^{t-1}$$

$$= \frac{1}{\sum_{\ell \in N_i} \sigma_\ell^{-2}} \sum_{k \in N_i} \frac{1}{d_k \sigma_k^2} \sum_{j \in N_k} \hat{\theta}_j^{t-1}.$$

$$\boxed{\hat{\theta}_i^t = \frac{1}{\sum_{\ell \in N_i} d_\ell} \sum_{k \in N_i} \sum_{j \in N_k} \hat{\theta}_j^{t-1}}$$

2 hop Majority Vote (MV): Average over the labeled nodes in 2-hop friend sets

Decoupled smoothing and ZGL

- ZGL:
  - Assume Homophily
  - Prior:
    $$\theta \sim N(0, \tau^2 (D - \gamma A)^{-1})$$
  - Matrix A: adjacency matrix
  - Reduce to iteratively 1-hop majority vote update method!
- Decoupled smoothing:
  - Don't assume Homophily
  - Prior:
    $$\theta \sim N(0, \tau^2 (Z - \gamma^2 W Z'^{-1} W^T)^{-1})$$
  - **Auxiliary matrix**:
    $$\tilde{A} = W Z'^{-1} W^T$$
  - Reduce to iteratively 2-hop majority vote update method!

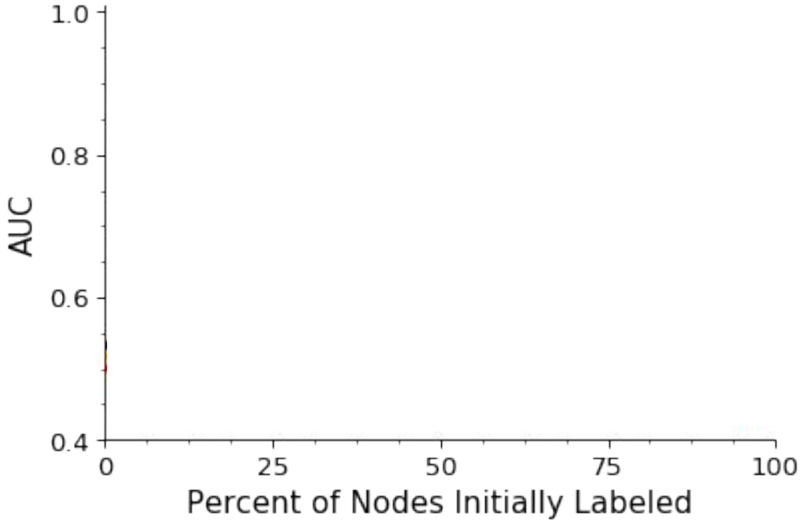
Empirical Results

Dataset

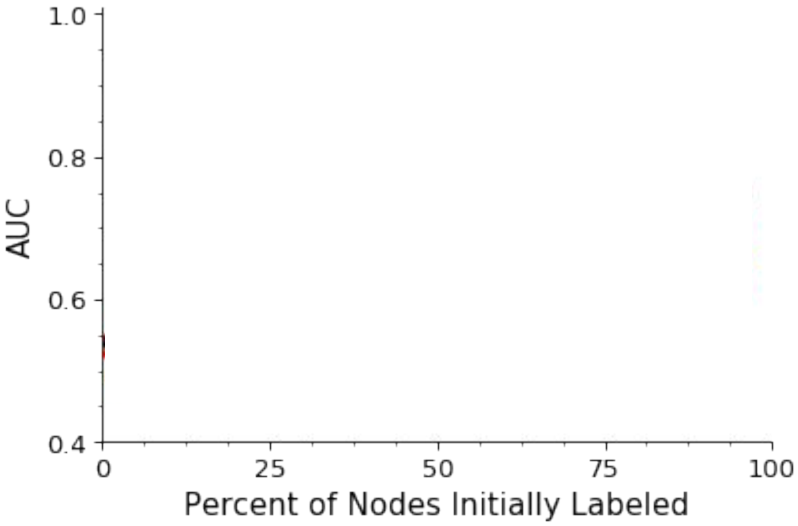
- **Facebook 100:** *Single-day snapshots of Facebook in September 2005.*
- **Goal:** gender prediction

School Name	Number of Nodes	Number of Edges
Amherst	*2032*	*78733*
Reed	*962*	*18812*
Haverford	*1350*	*53904*
Swarthmore	*1517*	*53725*

Decoupled smoothing: empirical result



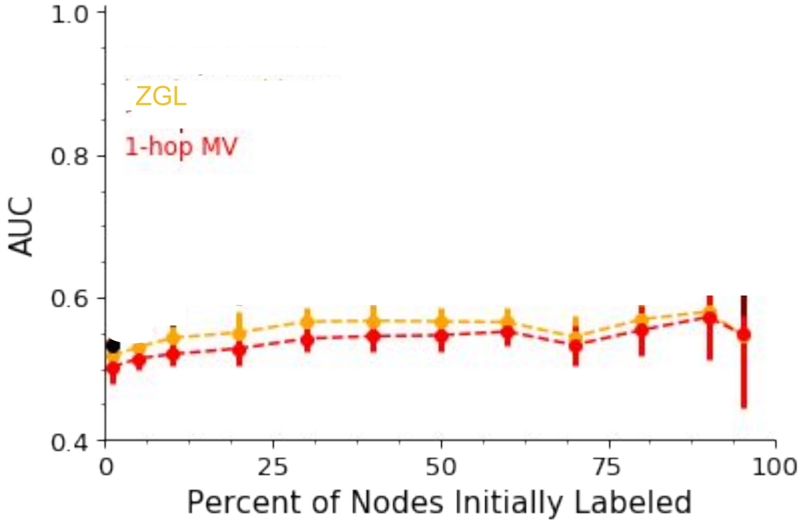
Reed



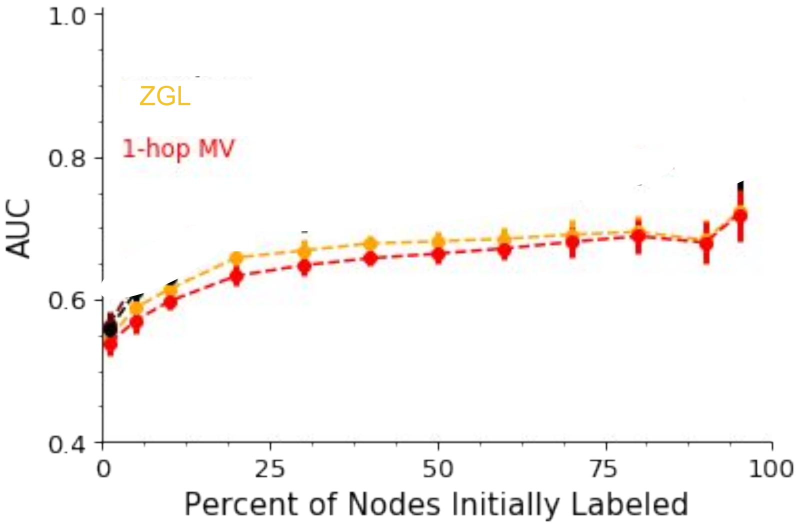
Swarthmore

Yatong Chen, Stanford University, WWW 2019

Decoupled smoothing: empirical result



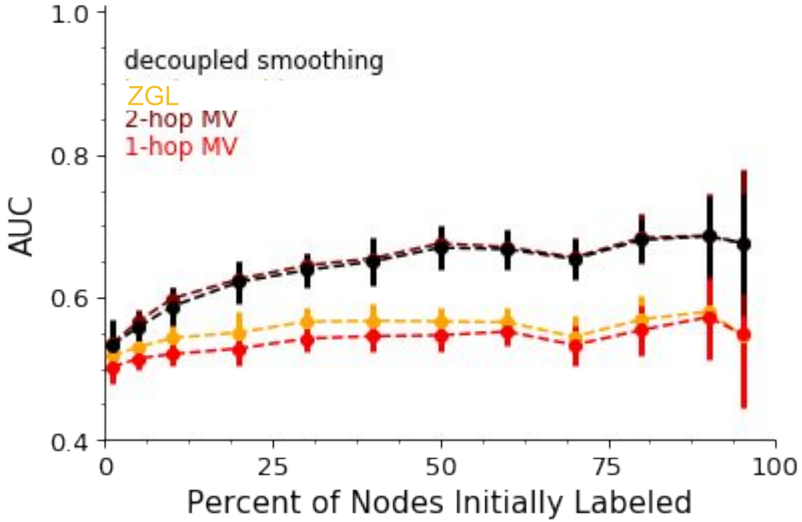
Reed



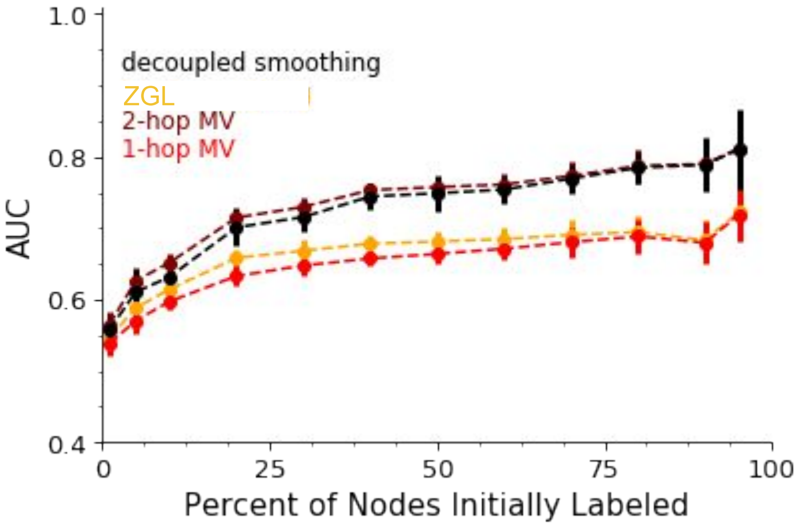
Swarthmore

Yatong Chen, Stanford University, WWW 2019

Decoupled smoothing: empirical result



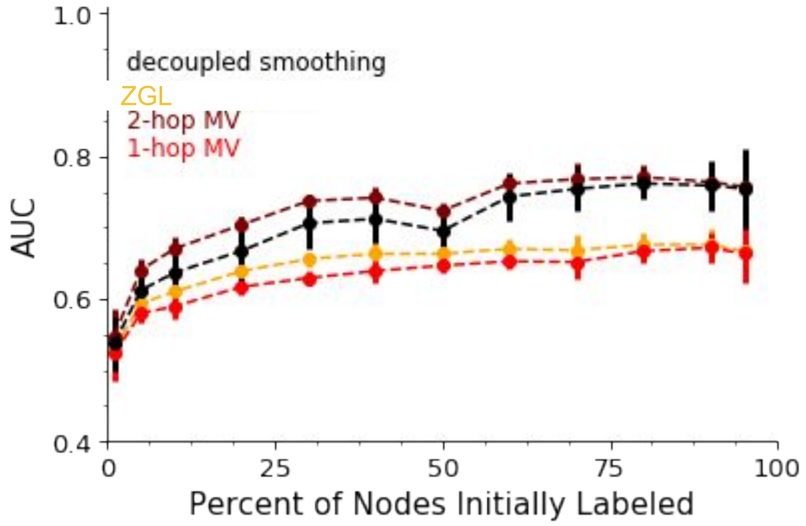
Reed



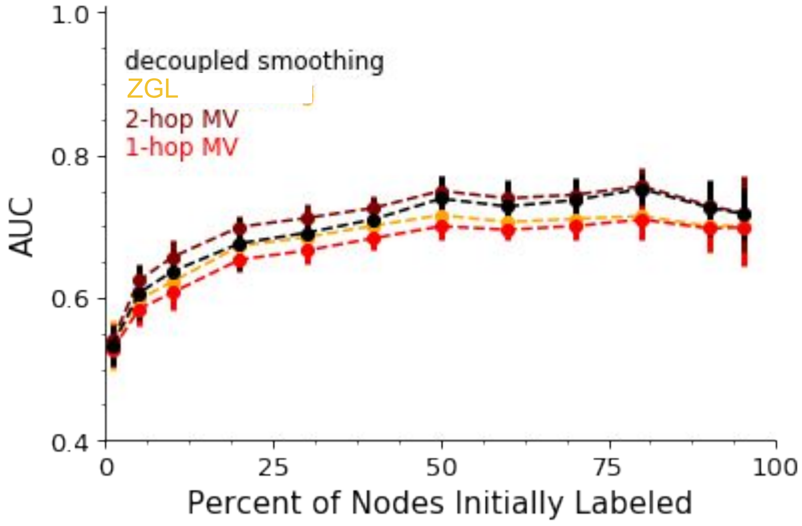
Swarthmore

Yatong Chen, Stanford University, WWW 2019

Decoupled smoothing: empirical result



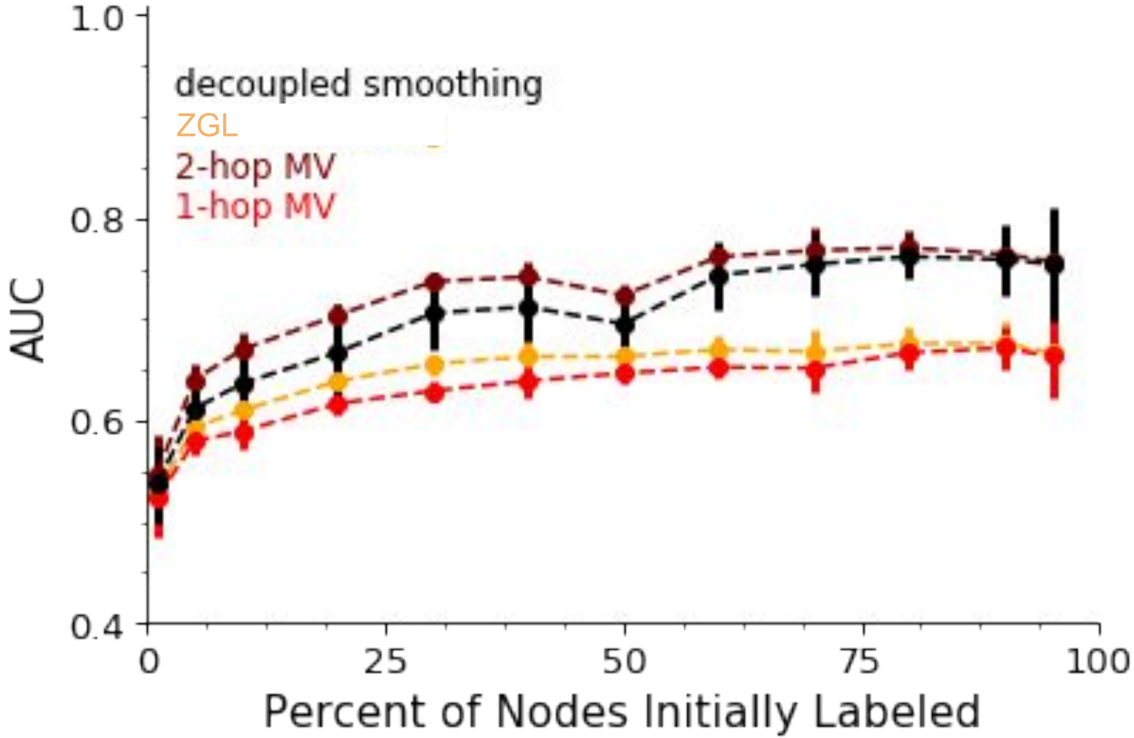
Amherst



Haverford

Yatong Chen, Stanford University, WWW 2019

Why 2-hop Majority Vote beats Decoupled Smoothing?



Amherst

Yatong Chen, Stanford University, WWW 2019

Summary

- **Introduce the idea of *decoupling* one's "identity" and "preference"**
- **Justify/explain the phenomenon of 2-hop similarity without assuming homophily**
- **Open questions:**
    - How to choose the weighted matrix W?
    - Why 2-hop Majority Vote outperforms decoupled smoothing: can you do better?

Questions?

Thank you for your attention!