SAILING LAB

Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Robust Reverse Engineering of Dynamic Gene Networks Under *Sample Size Heterogeneity*

Ankur P. Parikh, Wei Wu, Eric P. Xing
Carnegie Mellon University

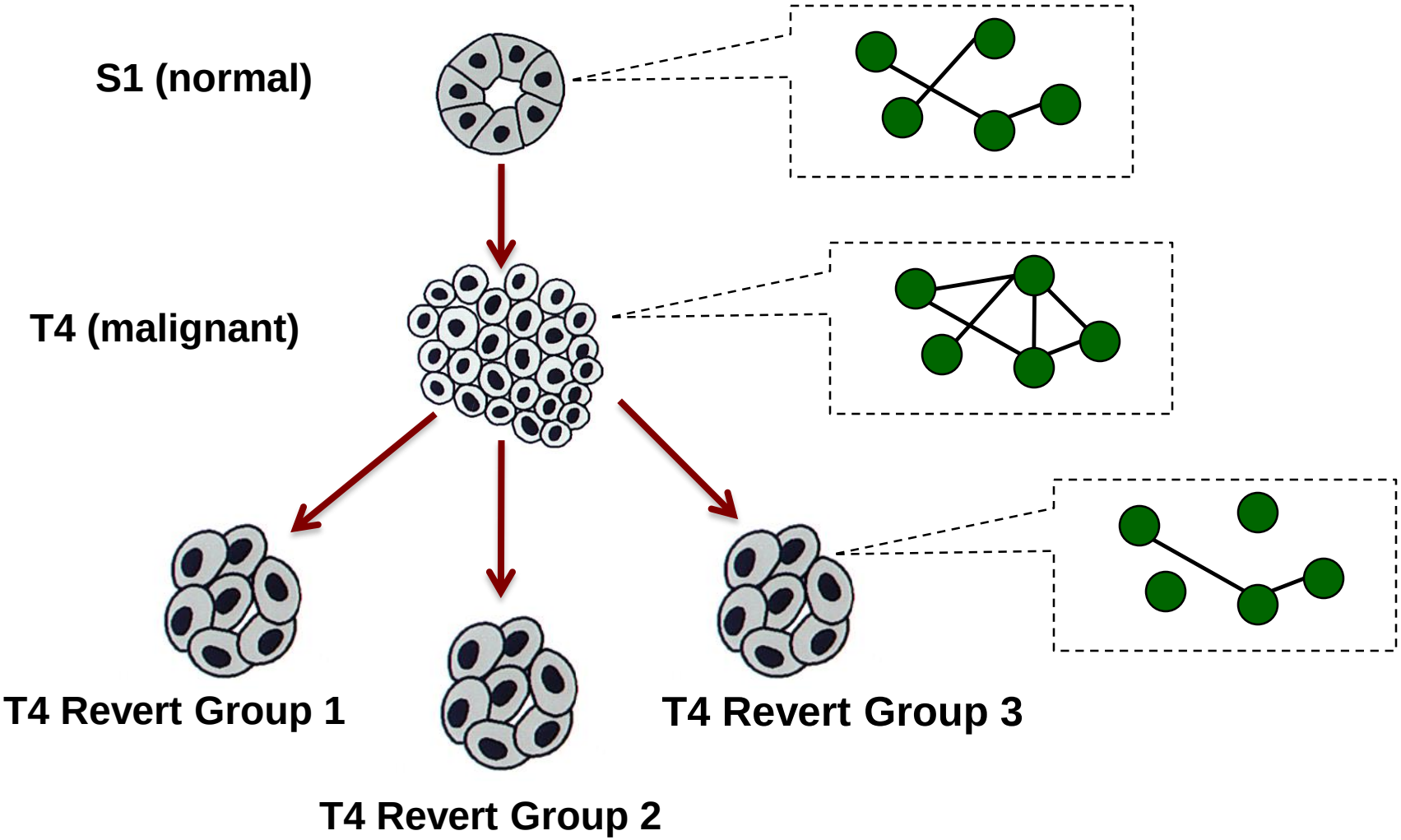
Outline

- Motivation
- Challenge of sample size heterogeneity
- Our solution
- Results
- Conclusion

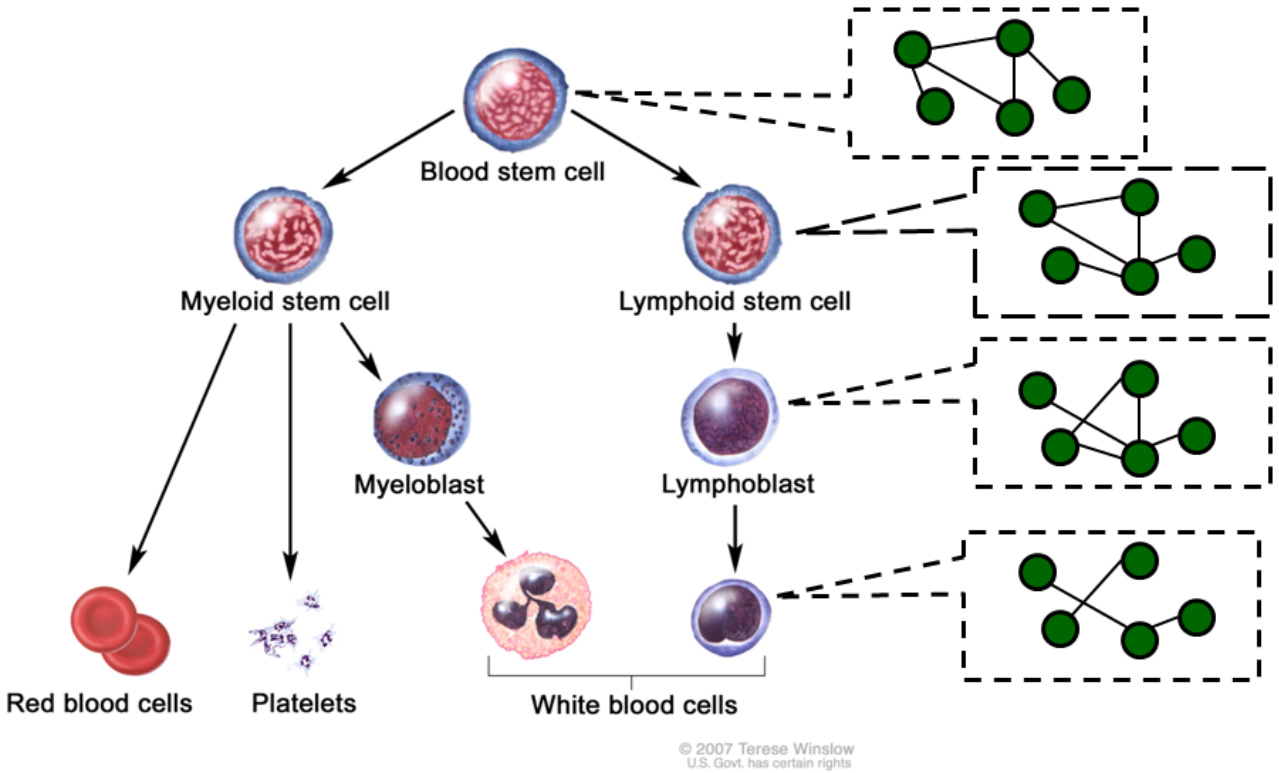
Outline

- Motivation
- Challenge of Sample Size Heterogeneity
- Our solution
- Results
- Conclusion

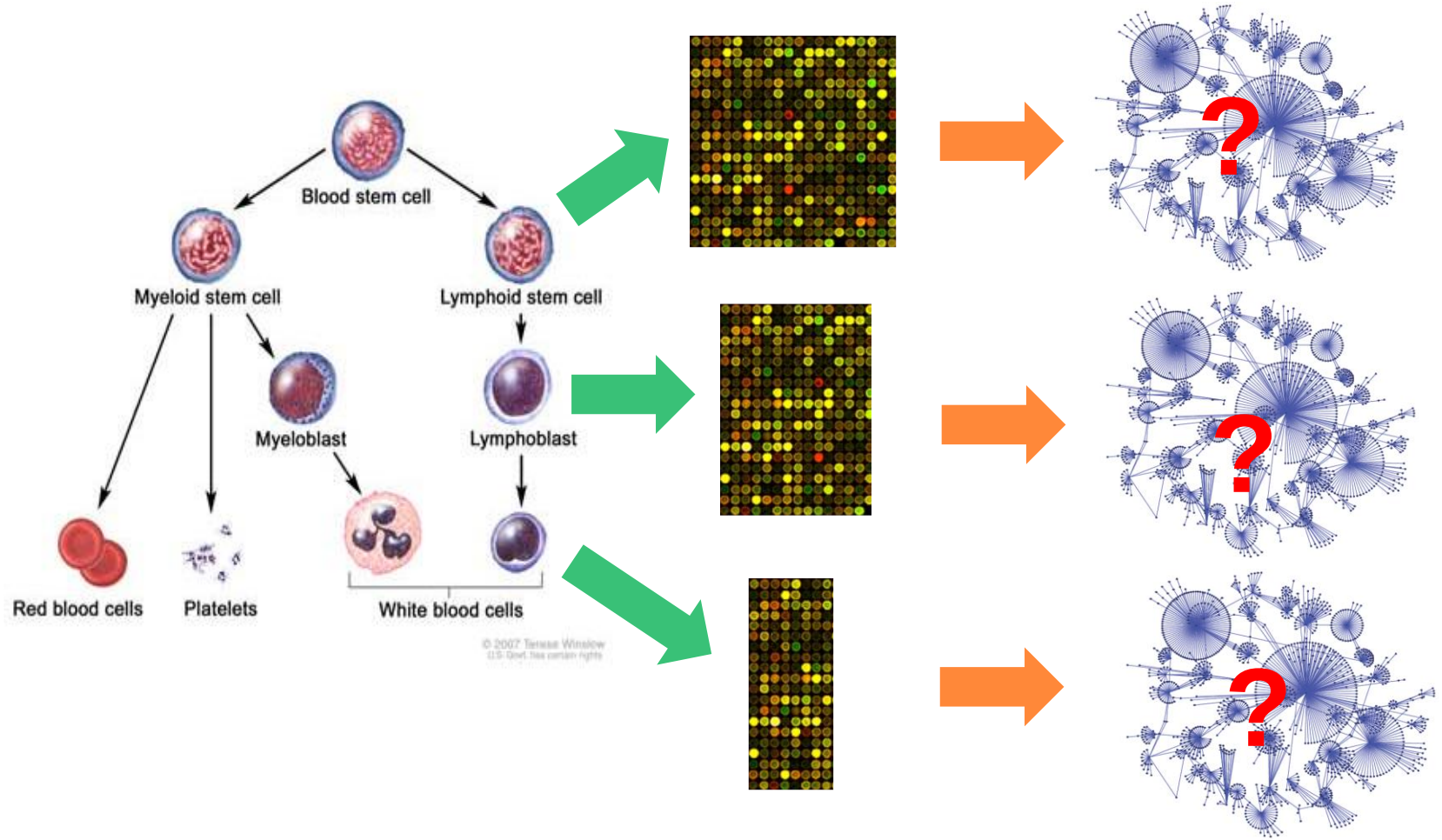
Multiple Network Estimation for Progression and Reversion of Breast Cancer cells



Stem Cell Differentiation



Use Microarray Data to Infer Network Structure



A Data Integration Problem

- Small numbers of samples from many related cell states. Poses opportunities and challenges.

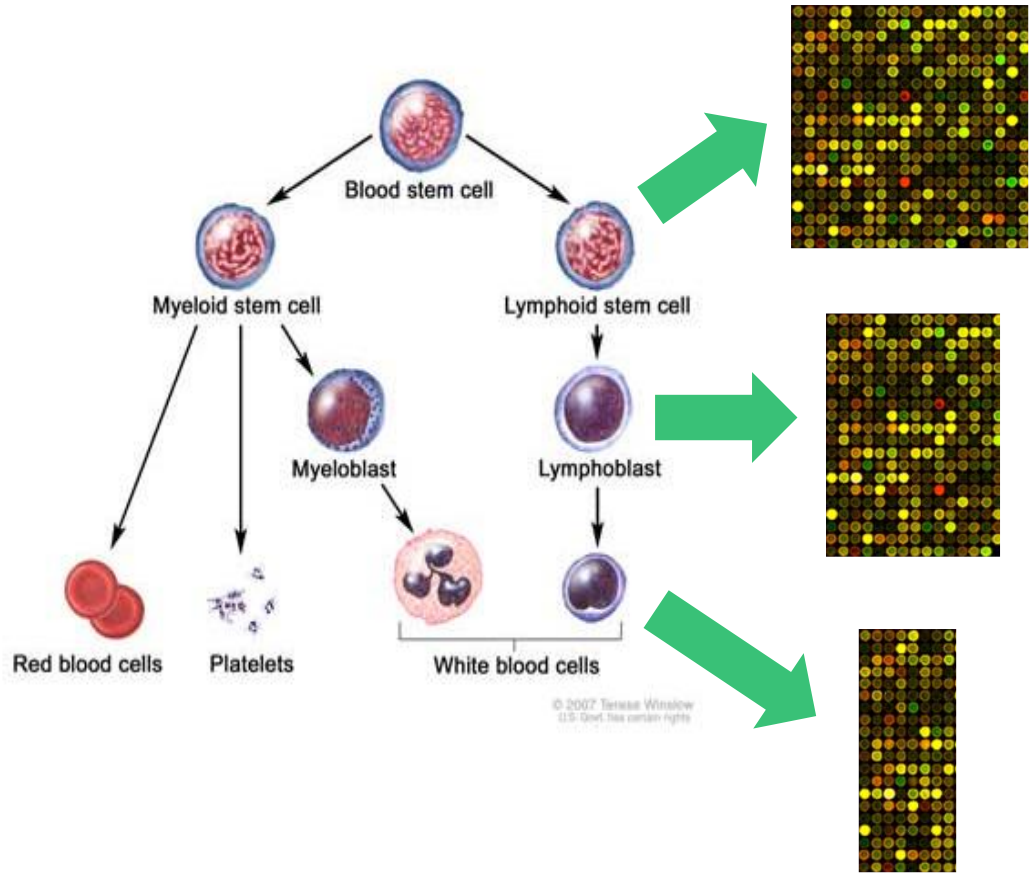
- **Challenges**

  - How to leverage similarities among cell states for more accurate network estimation? *(some of our previous work)*

  - How to integrate data in a “correct” way that does not compromise validity of biological conclusions? ***(a focus of this work)***

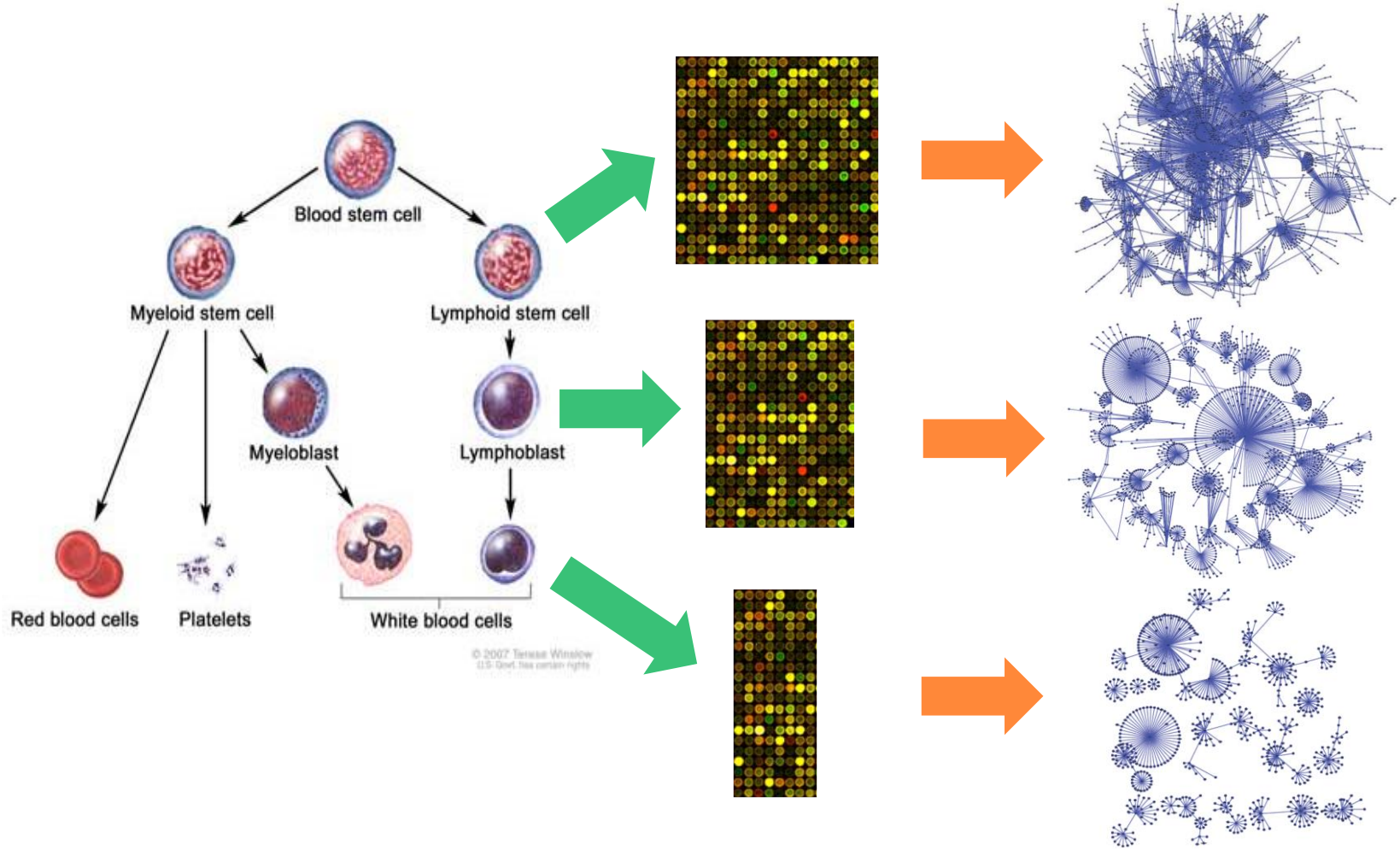
Challenge: Sample Size Heterogeneity for network estimation

- Different cell states have different sample sizes



Biases Resulting Networks

- Cell states with more samples have more edges



Sample Heterogeneity Makes Network Analysis Problematic

- Unclear which network differences are meaningful and which are an artifact of sample size heterogeneity
  - Comparing Macro-Topology Features
    - **Network Density**
    - **Centrality**
  - Comparing Gene Level Interactions
    - **Gene neighborhoods**
    - **Local modules**

Our Contribution

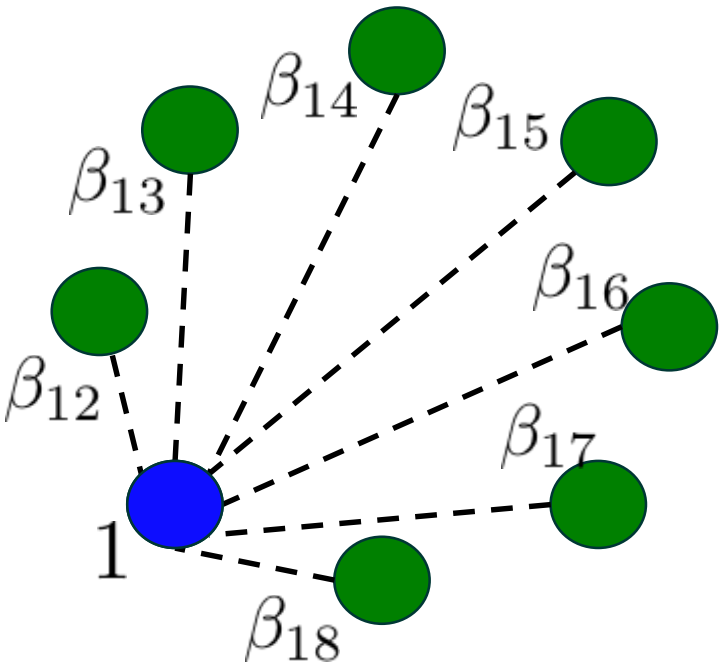
- Identify the novel problem of sample size heterogeneity in network reconstruction
- We focus on sparse regression methods for network reconstruction *[Meinshausen and Buhlmann 2006, Friedman et al. 2008. Song et al. 2009, Parikh et al. 2011]*
- We add a regularization constraint to make these methods robust to sample size heterogeneity

Outline

- Motivation
- Challenge of sample size heterogeneity
- Our solution
  - Background
  - Increasing robustness to sample size heterogeneity
  - Sharing information across states
- Results
- Conclusion

Background on Sparse Regression Network Reconstruction [Meinshausen and Buhlmann 2006]

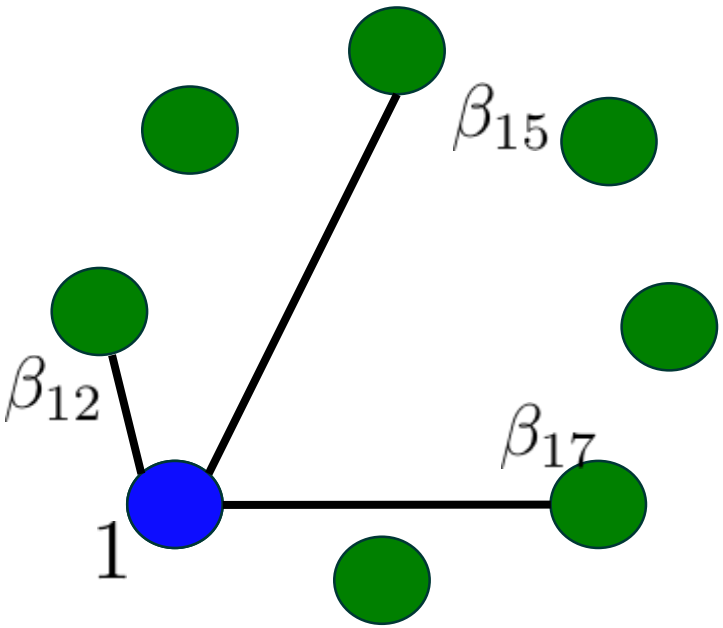
- Network Learning with the Graphical LASSO [Meinshausen and Buhlmann 2006]



Learn each neighborhood separately

Background on Sparse Regression Network Reconstruction [Meinshausen and Buhlmann 2006]

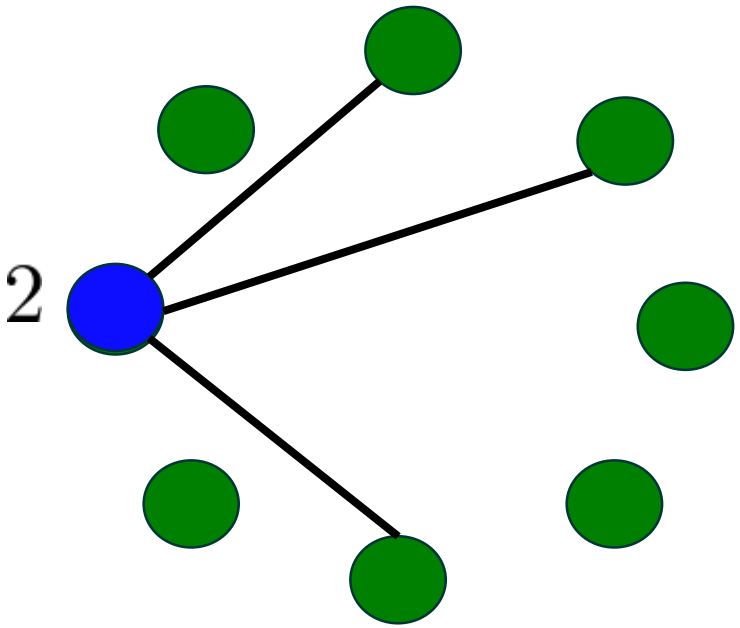
- Use the LASSO to select a sparse set of only the **relevant** edges



Selects sparse set of edges

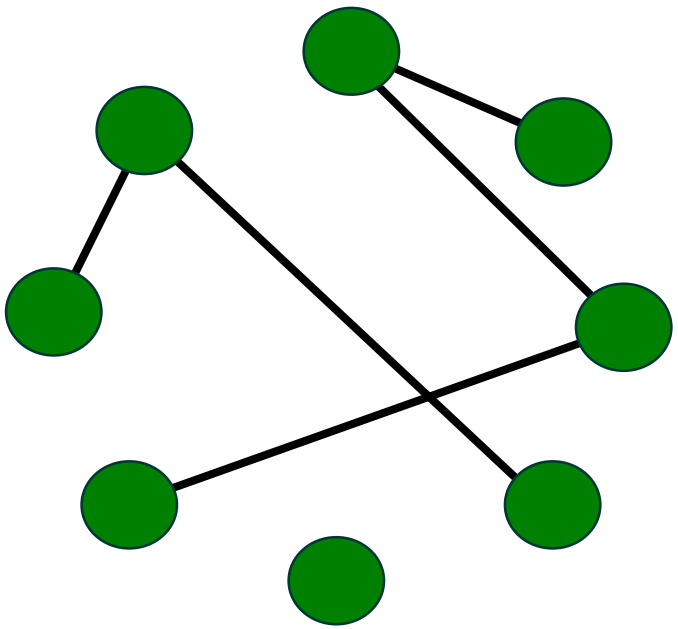
Background on Sparse Regression Network Reconstruction [Meinshausen and Buhlmann 2006]

- Repeat this for every node



Background on Sparse Regression Network Reconstruction [Meinshausen and Buhlmann 2006]

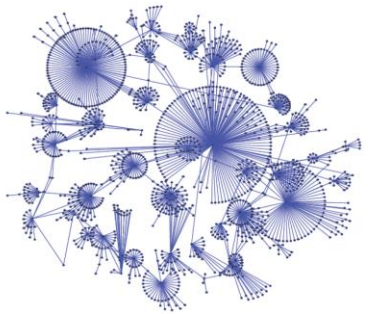
- Combine all the neighborhoods to form a network



Learning Multiple Networks

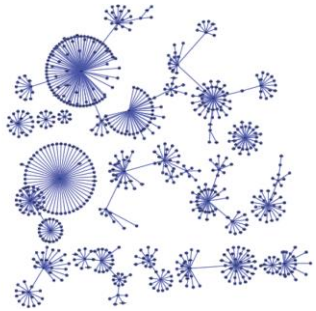
- (Naïve way) run the graphical lasso for each!

Graphical lasso for network 1 →



network 1

Graphical lasso for network 2 →

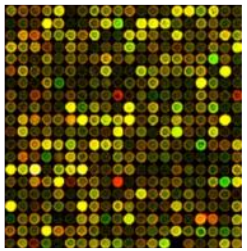


network 2

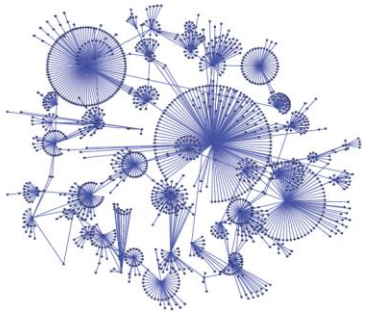
- More sophisticated techniques to share information exist (Song et al. 2009, Ahmed and Xing 2009, Parikh et al. 2011)

Sample Size Heterogeneity

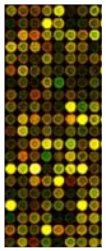
- The graphical lasso will encourage networks with more samples to have more edges.



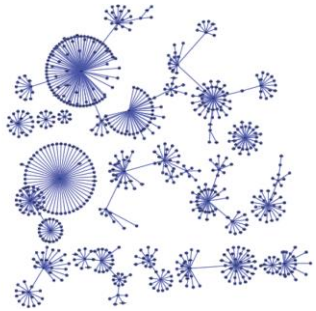
Graphical lasso for network 1



network 1



Graphical lasso for network 2

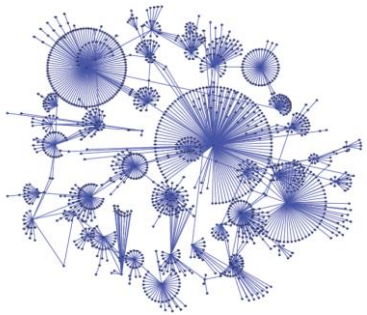


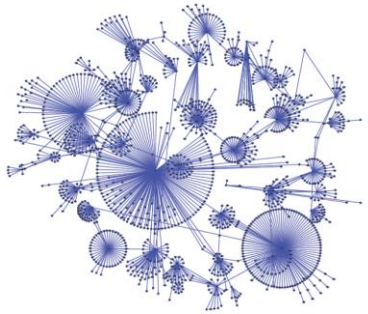
network 2

- Simply tricks such as scaling parameters only work in theory but not in practice.

Intuition Of Our Solution

- Make **"normalizing"** assumption that absolute sum of edge weights for all networks is equal

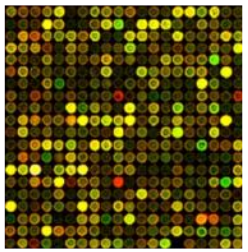




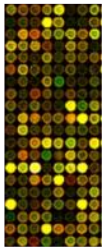
$$sum\ of\ (absolute) edge\ weights\ of\ network\ 1 = sum\ of\ (absolute) edge\ weights\ of\ network\ 2 = C$$

- Incorporate this assumption into graphical lasso procedure

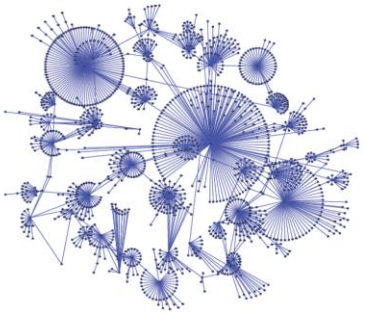
Intuition Of Our Solution



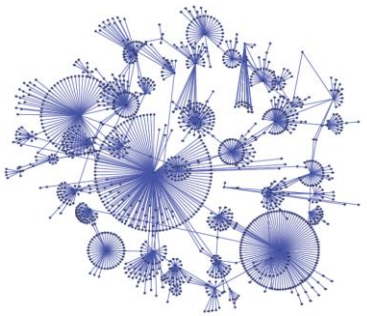
Graphical lasso for network 1
with constraint



Graphical lasso for network 2
with constraint

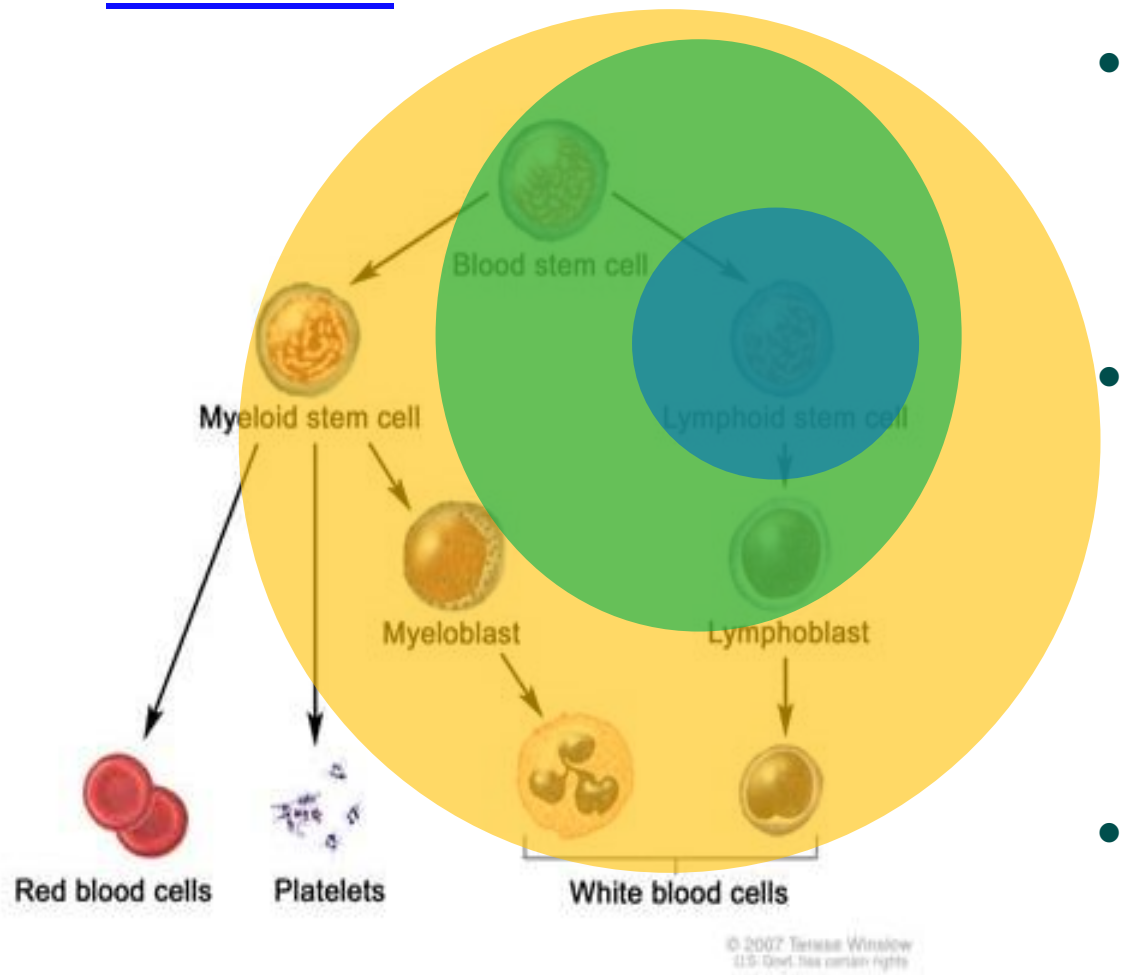


network 1



network 2

Sharing Information Among Related States



- Often number of samples per state is **very small.**
- "Share" samples among related states
  - *Song et al. 2009*
  - *Ahmed and Xing 2009*
  - *Parikh et al. 2011*
- These strategies can be integrated into our robust framework

Summary of Method

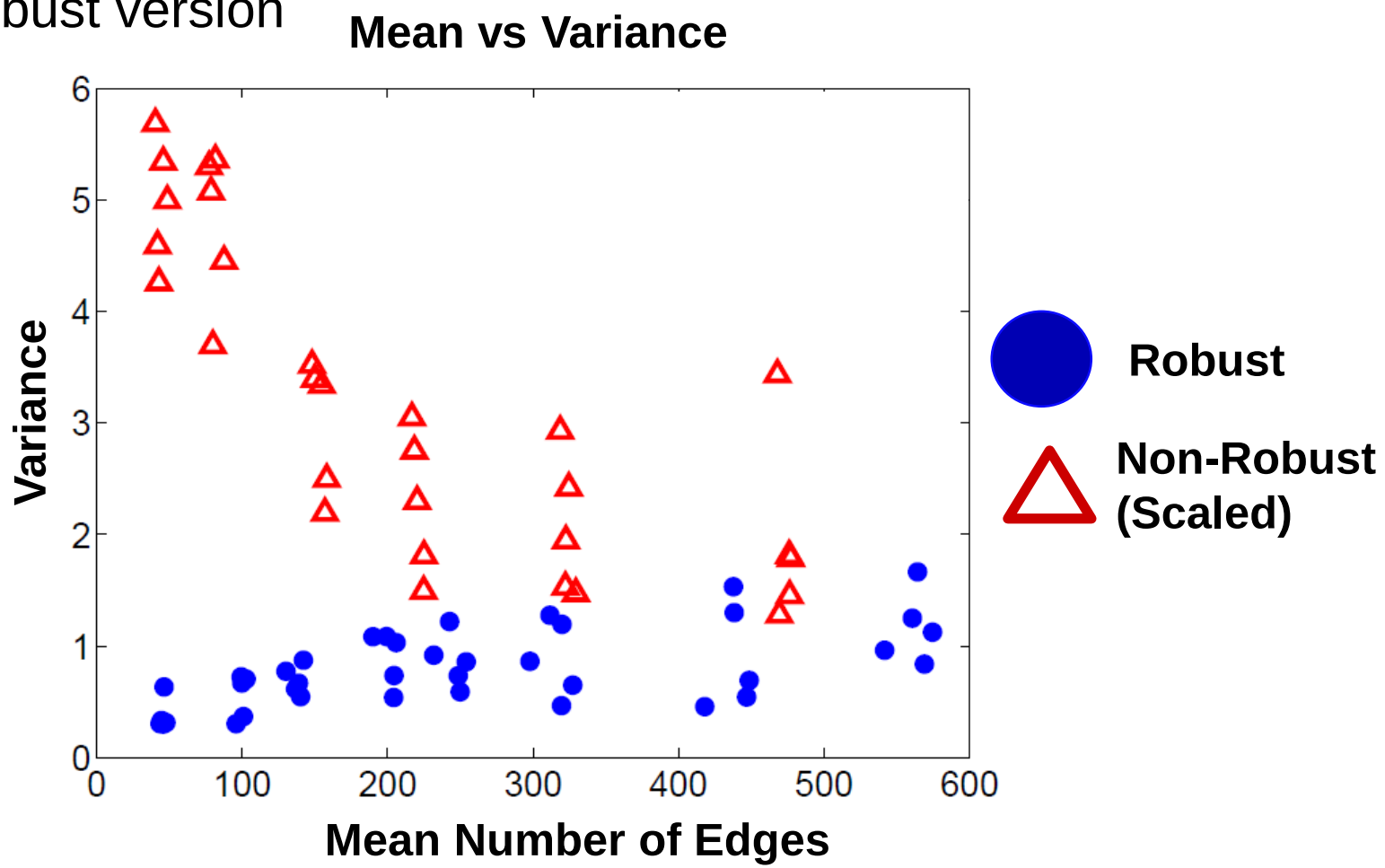
- Based on the graphical lasso [Meinshausen and Buhlmann 2006]
- Add regularization constraint so that networks are calibrated
- Share information among related cell states by taking weighted average of related samples

Outline

- Motivation
- Challenge of sample size heterogeneity
- Our solution
- Results
- Conclusion

Results on Synthetic Data

- Robust algorithm produces more calibrated networks than non robust version

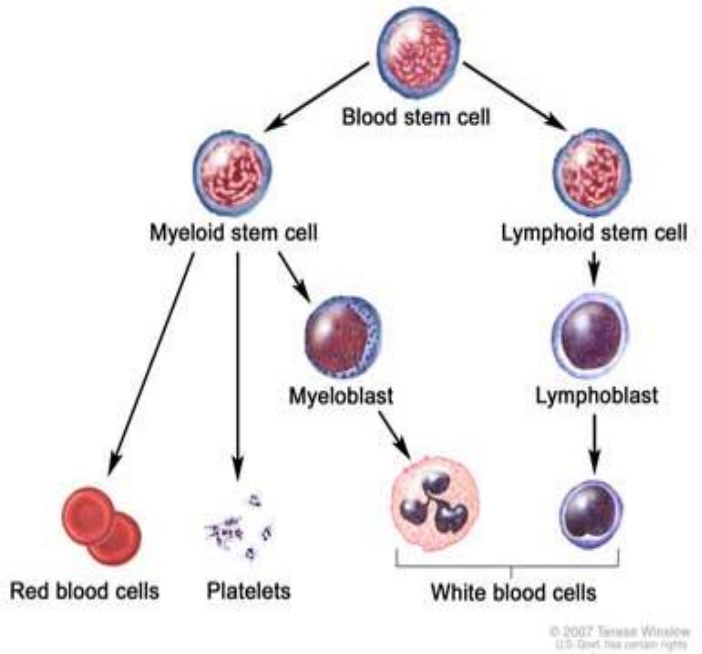


- Both methods produce comparable F1 score

Stem Cell Data

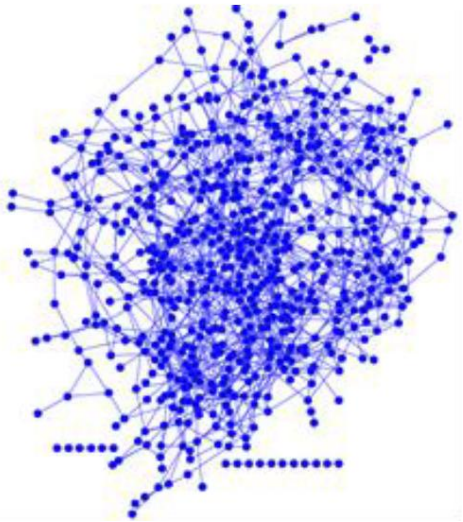
- Hematopoietic stem cell data from Novershtern et al. 2011

- 38 cell states
  - 4-7 samples per cell state
  - 211 total samples

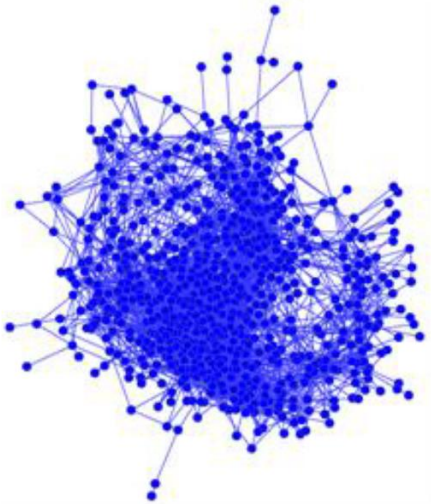


Stem Cell Data Results

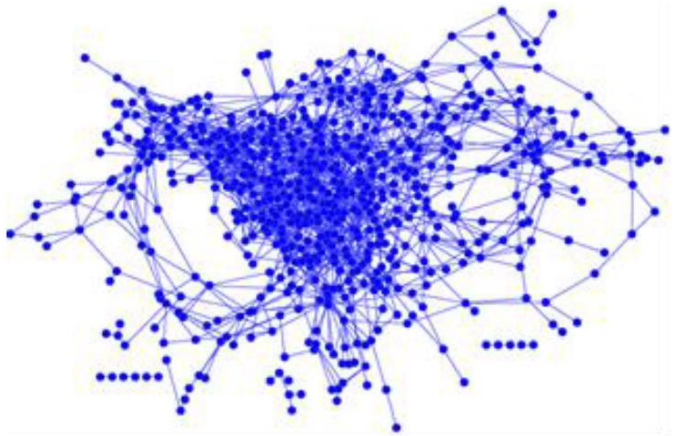
Granulocyte network



Common Myleoid Progenitor Network



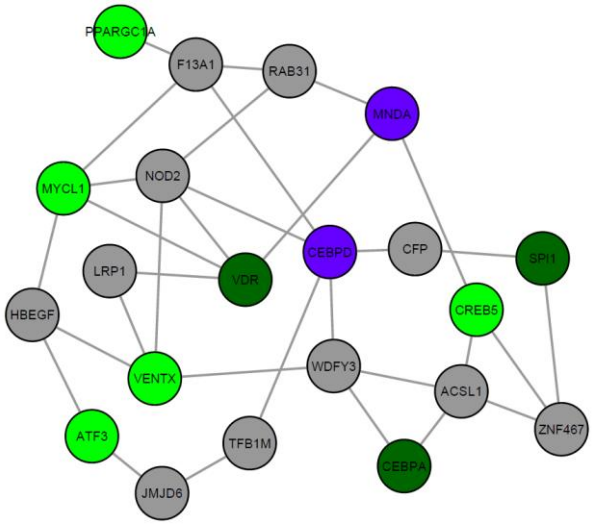
Non-Robust



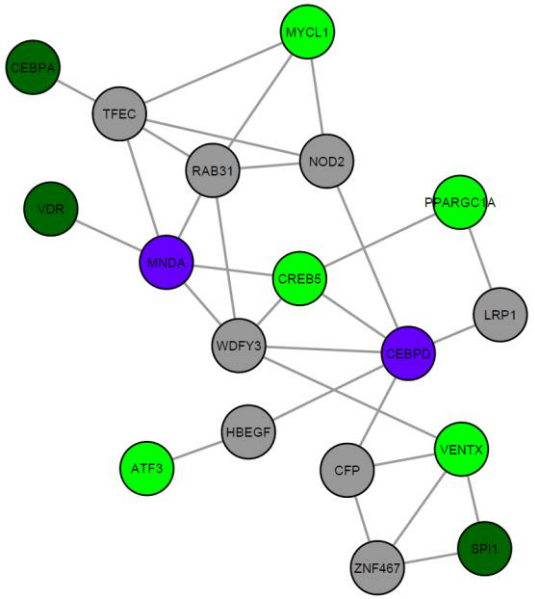
Robust

Stem Cell Data: Biological Analysis

granulocyte-subnetwork



monocyte-subnetwork



Experimentally validated regulators (Novershtern et al.)

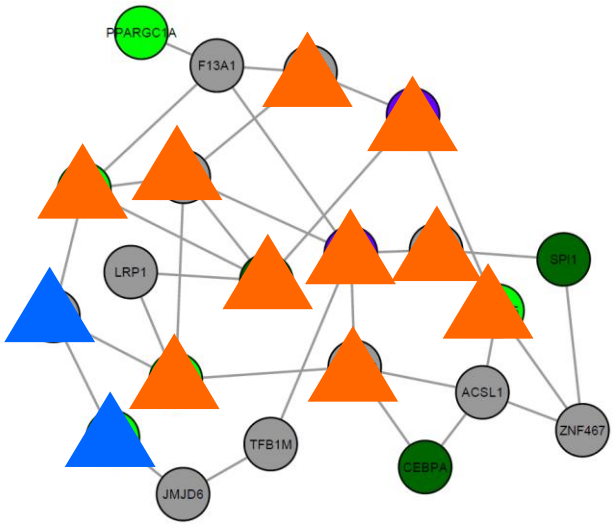
Experimentally validated genes

Non-experimentally validated genes

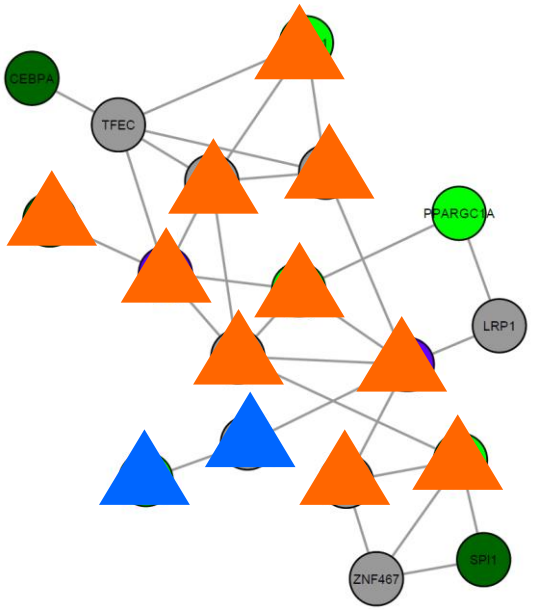
Noveshtern et al. 2011 doesn't give insight into module topology.

Stem Cell Data: Biological Analysis

granulocyte-subnetwork



monocyte-subnetwork



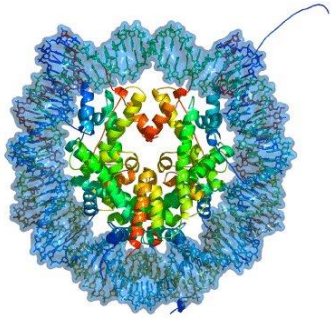
- Two modules with identical gene interaction patterns
  - **large 10-gene module**
  - **small 2-gene module**

- 7 out of these 12 genes experimentally validated by Novershtern et al.

Conclusion

- We addressed the novel problem of sample size heterogeneity in the context of dynamic network reconstruction
- Our solution produces more calibrated networks allowing for more meaningful biological comparison
- We hope this can help efforts in personalized medicine and data integration

Acknowledgements



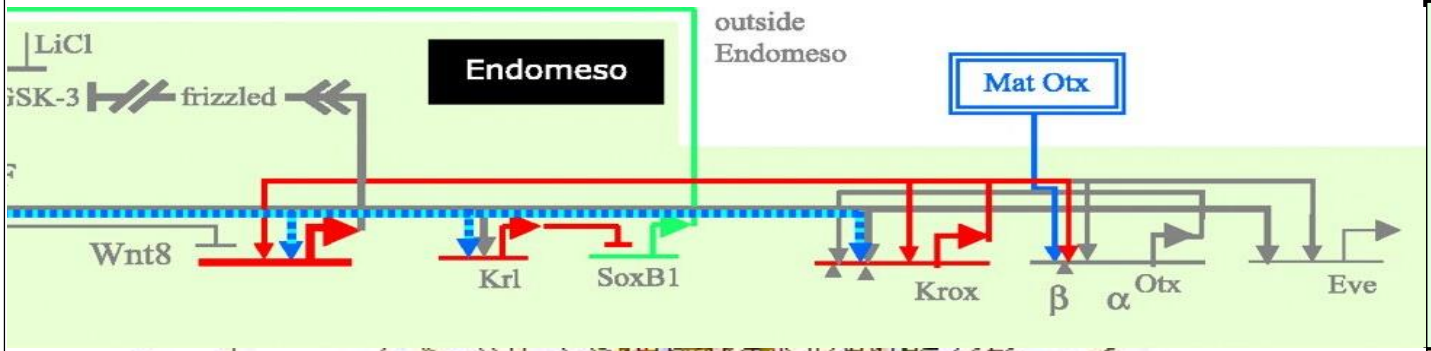
PSB Travel Fellowship 2014

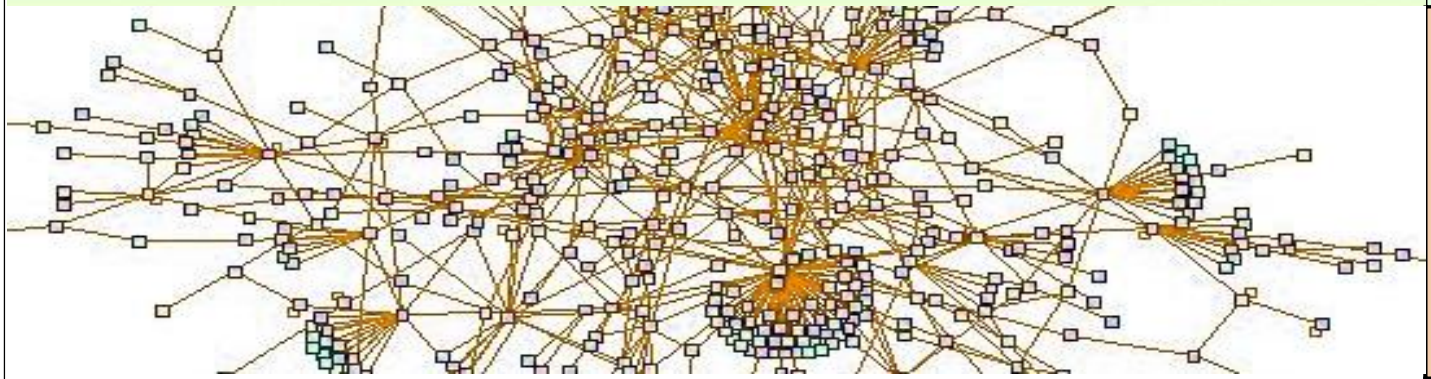
NIH

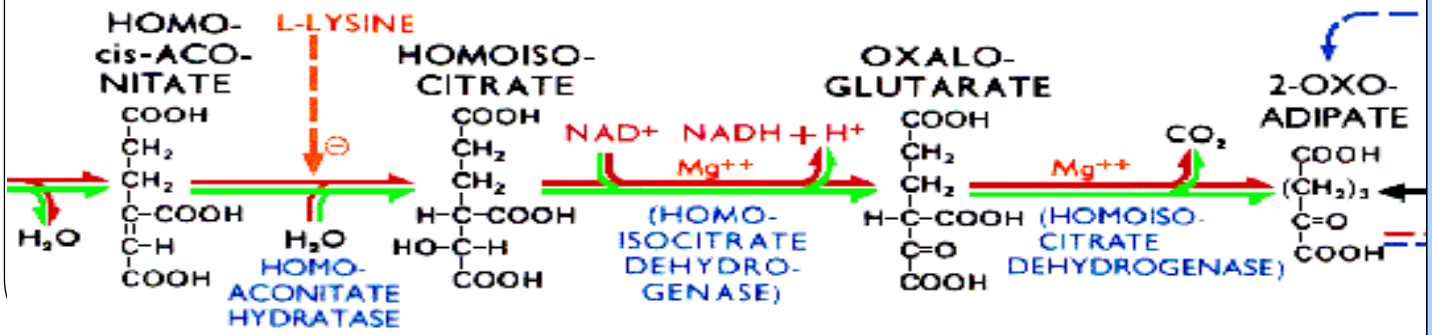
NSF Graduate Fellowship

Thanks!

Regulation of cell response to stimuli is paramount, but we can usually only measure (or compute) steady-state interactions







Transcriptional interactions

- Chromatin IP
- Microarrays

Protein—protein interactions

- Protein coIP
- Yeast two-hybrid

Biochemical reactions

- Metabolic flux measurements

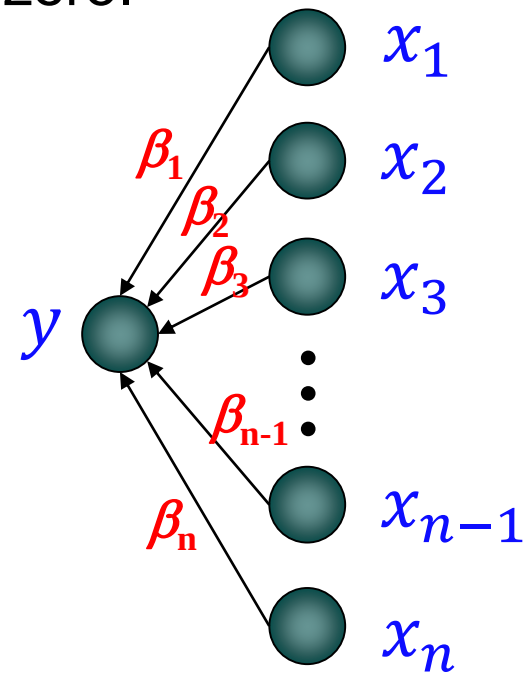
Sparsity: In a mathematical sense

- Consider least squares linear regression problem:
- Sparsity means most of the beta's are zero.

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$$

$$subject\ to$$

$$\sum_{j=1}^{p} I\left[|\beta_j| > 0\right] \leq C$$



- But this is not convex!!! Many local optima, computationally intractable.

L1 Regularization (LASSO) [Tibshirani 1996]

- A convex relaxation.

Constrained Form

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$$

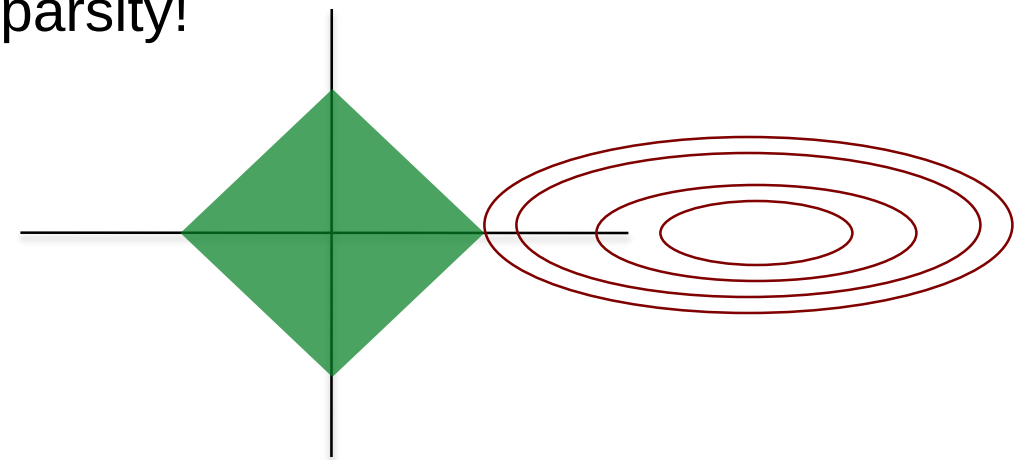
$$subject\ to$$

$$\sum_{j=1}^{p} I\left[|\beta_j| > 0\right] \leq C$$

Lagrangian Form

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

- Still enforces sparsity!



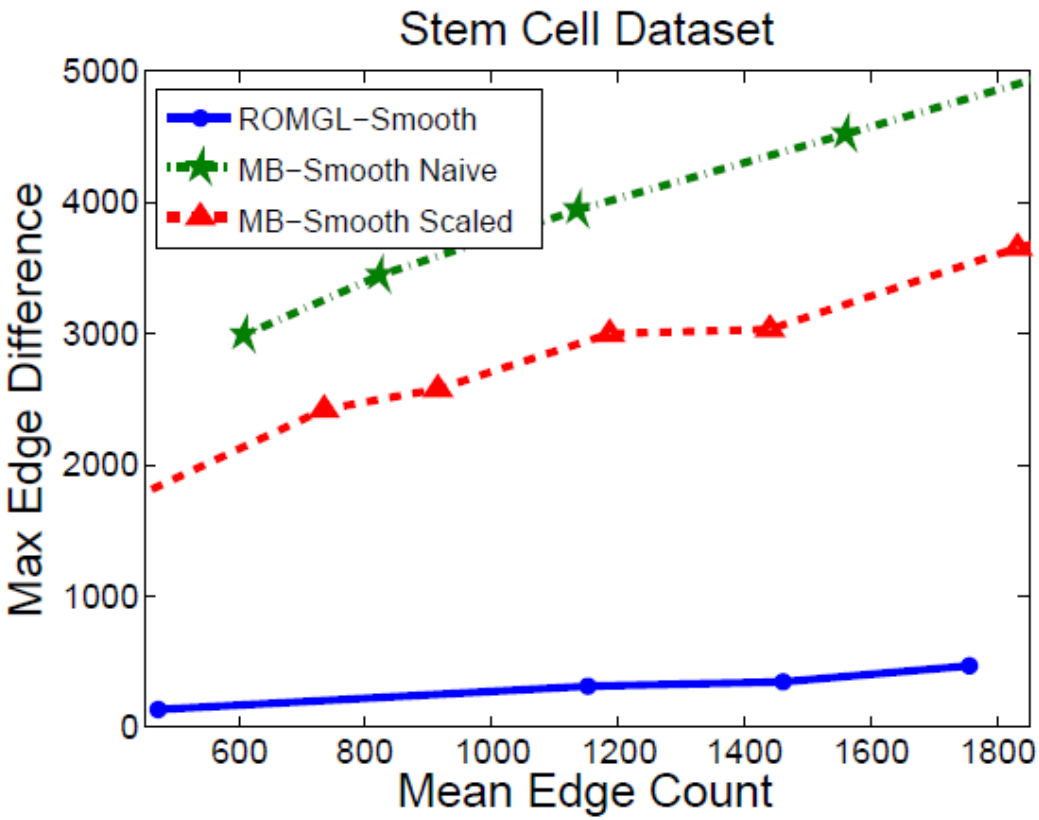
Our Solution – Network Calibration

- **Calibrate** networks to remove this sample size "**bias**"
- Analogous to microarray processing to remove systematic dye bias
- Except our calibration is not a pre/post processing step but rather a fundamental part of the method

Sparsity!

- One common assumption to make **sparsity.**
- **Makes biological sense:** Genes are only assumed to interface with small groups of other genes.
- **Makes statistical sense:** Learning is now feasible in high dimensions with small sample size

Stem Cell Data: Quantitative Results



Intuition Of Our Solution

$$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}^{(\boldsymbol{z})}, \ldots, \widehat{\boldsymbol{\beta}}_{\boldsymbol{p}}^{(\boldsymbol{z})} = \underset{\boldsymbol{\beta}_{\mathbf{1}}^{(\boldsymbol{z})}, \ldots, \boldsymbol{\beta}_{\boldsymbol{p}}^{(\boldsymbol{z})}}{\operatorname{argmin}} \sum_{j=1}^{p} \left\| \boldsymbol{X}_{\boldsymbol{j}}^{(\boldsymbol{z})} - \boldsymbol{X}_{-\boldsymbol{j}}^{(\mathbf{1})} \boldsymbol{\beta}_{\boldsymbol{j}}^{(\boldsymbol{z})} \right\|_2^2$$

$$subject\ to$$

$$\sum_{j=1}^{p} \left\| \boldsymbol{\beta}_{\boldsymbol{j}}^{(\boldsymbol{z})} \right\|_1 = C$$

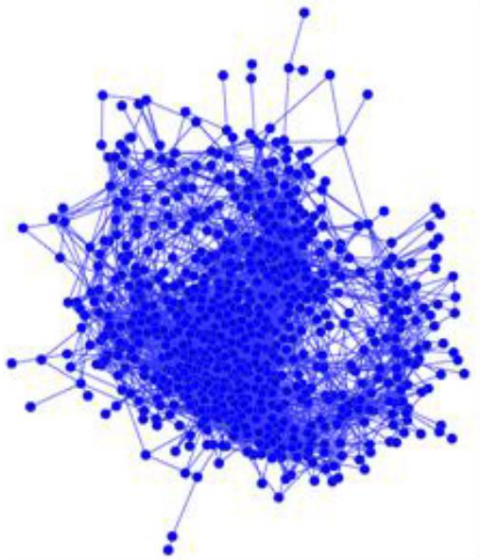
$$for\ all\ cell\ states\ \boldsymbol{z}$$

(something like this is true)

Stem Cell Data Results

28 nodes with degree > 0

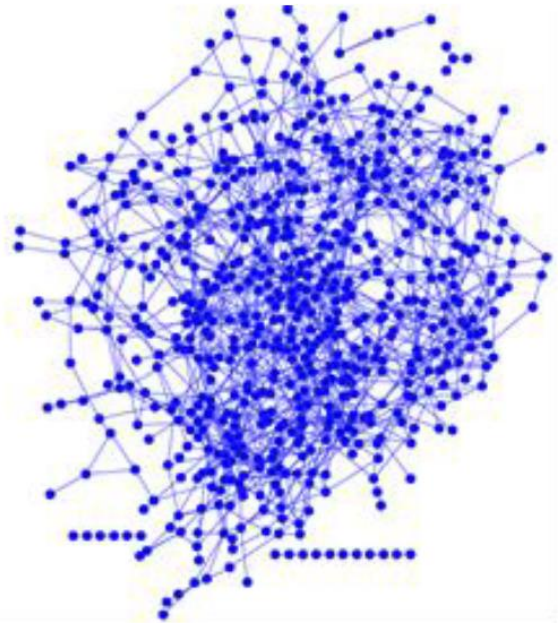
Granulocyte network



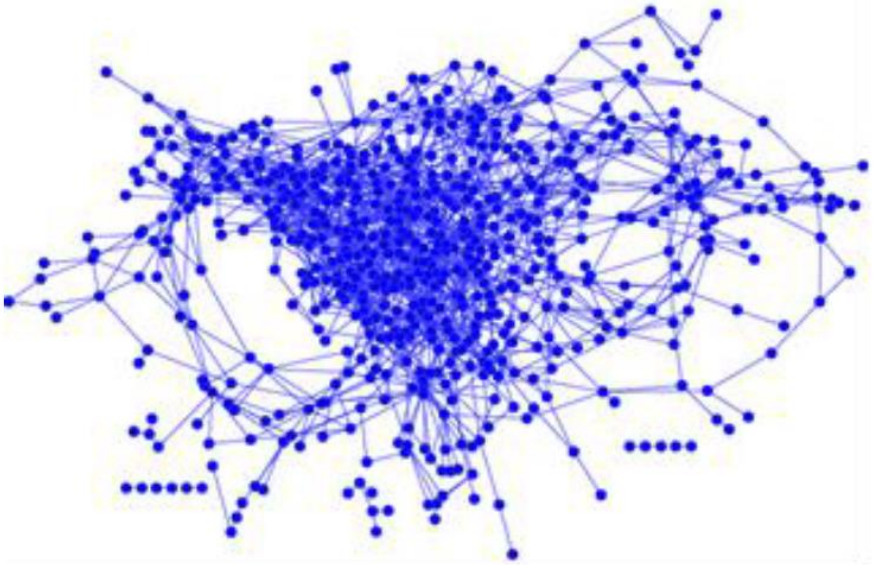
719 nodes with degree > 0

Common Myleoid Progenitor Network

Our Robust Approach

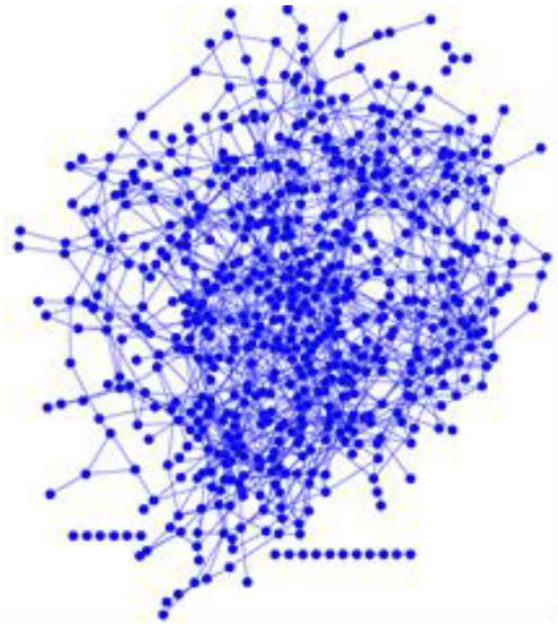


643 nodes with degree > 0

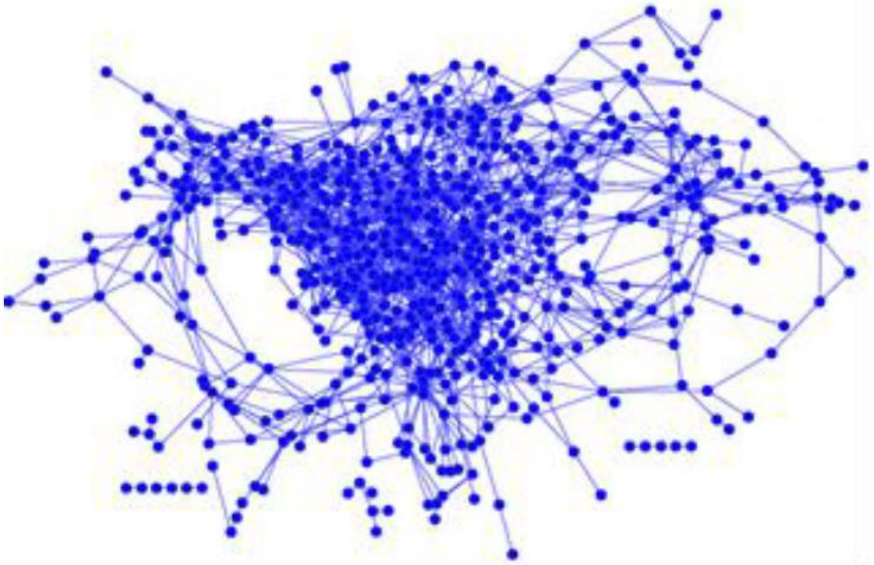


677 nodes with degree > 0

Our Robust Approach



643 nodes with degree > 0



677 nodes with degree > 0

Sample Size Heterogeneity for Network Estimation

$$\widehat{\boldsymbol{\beta}}_1^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}_p^{(1)} = \underset{\boldsymbol{\beta}_1^{(1)}, \ldots, \boldsymbol{\beta}_p^{(1)}}{\operatorname{argmin}} \sum_{j=1}^{p} \left\| \boldsymbol{X}_j^{(1)} - \boldsymbol{X}_{-j}^{(1)} \boldsymbol{\beta}_j^{(1)} \right\|_2^2 + \lambda \sum_{j=1}^{p} \left\| \boldsymbol{\beta}_j^{(1)} \right\|_1$$

Increases as number of samples increases

stays constant as a function of number of samples

Thus applying the same λ will result in cell types with more samples having more edges

What About Scaling?

$$\widehat{\boldsymbol{\beta}}_1^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}_p^{(1)} = \underset{\boldsymbol{\beta}_1^{(1)}, \ldots, \boldsymbol{\beta}_p^{(1)}}{\operatorname{argmin}} \sum_{j=1}^{p} \frac{1}{S^{(1)}} \left\| \boldsymbol{X}_j^{(1)} - \boldsymbol{X}_{-j}^{(1)} \boldsymbol{\beta}_j^{(1)} \right\|_2^2 + \frac{\lambda}{\sqrt{S^{(1)}}} \sum_{j=1}^{p} \left\| \boldsymbol{\beta}_j^{(1)} \right\|_1$$

- **Only works in theory, not in practice.**
  - Requires restrictive modeling assumptions to hold
  - Requires larger sample size to kick in