



# Context-Specific Nested Effect Models

RECOMB 2018, Paris

**Yuriy Sverchkov**

Elisha Yi-Hsuan Ho

Audrey P. Gash

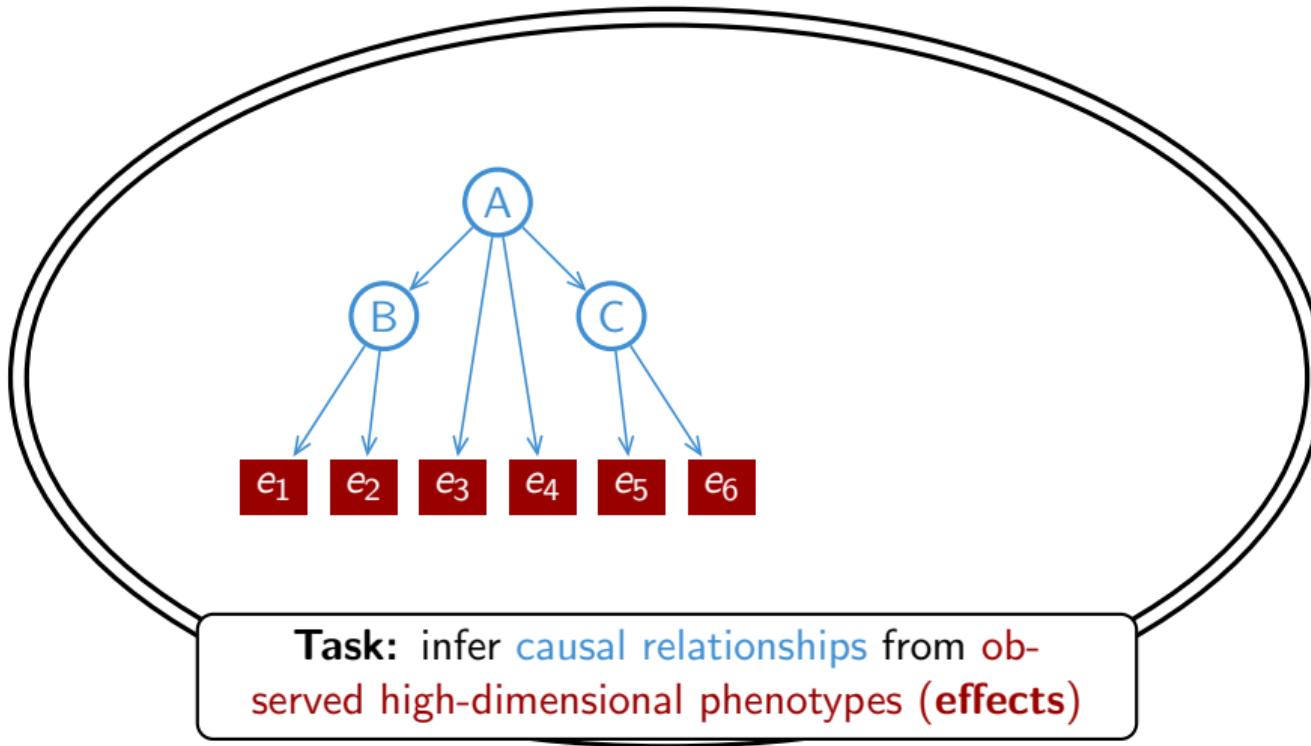
Mark Craven

University of Wisconsin–Madison

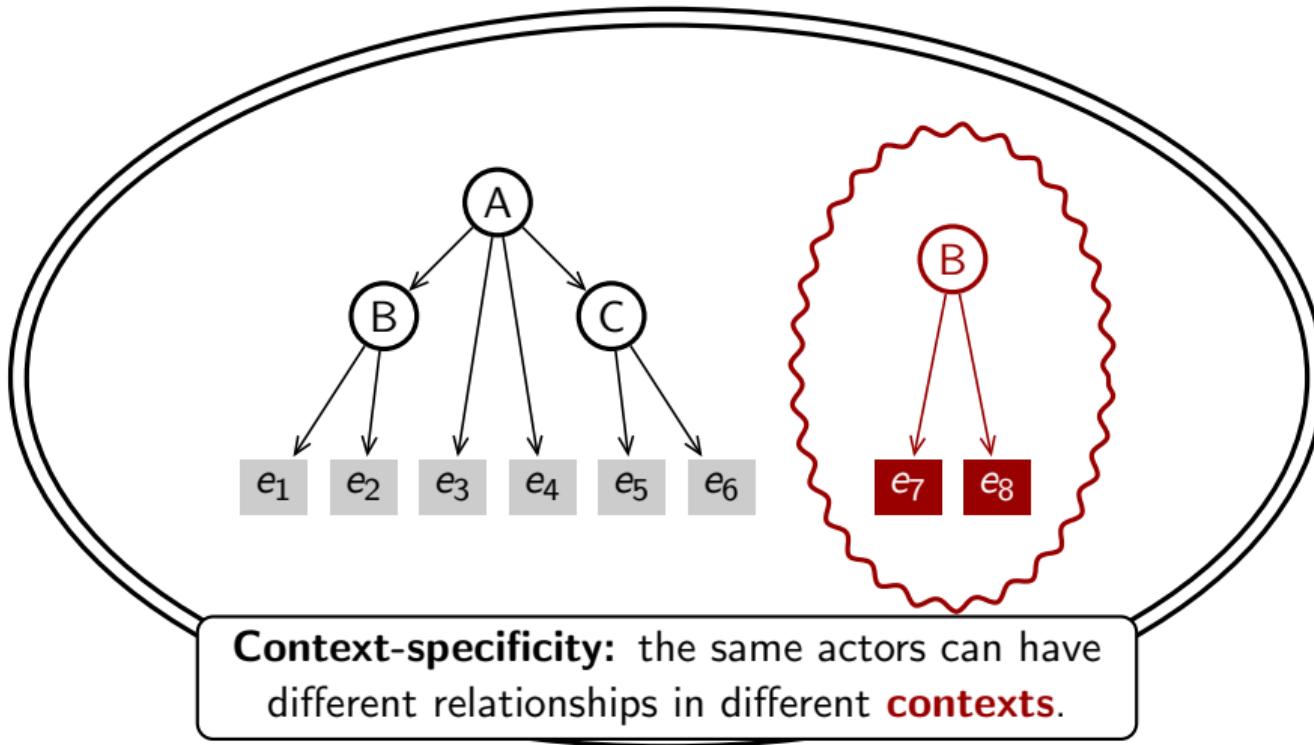


April 24, 2018

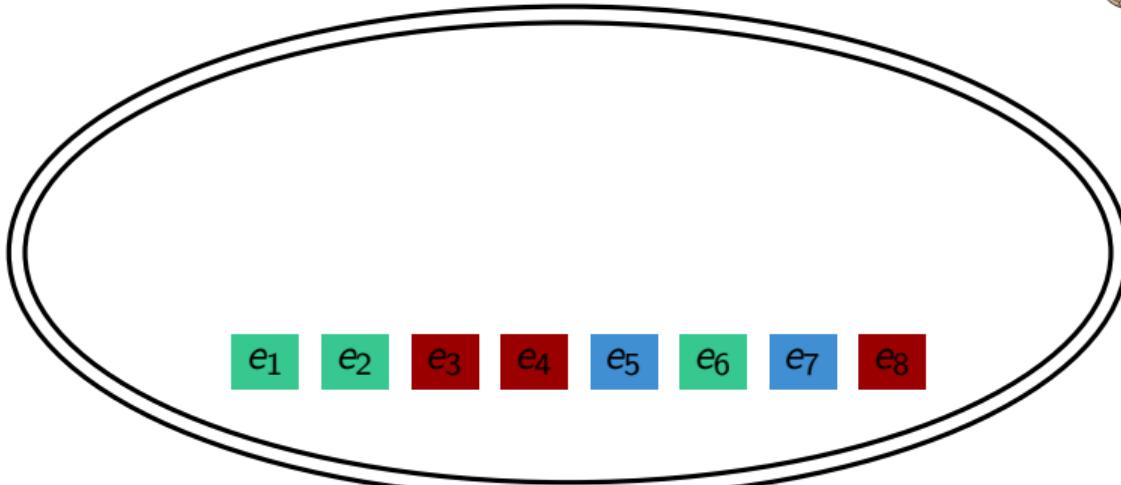
# Context-specific biological networks



# Context-specific biological networks



# Uncovering biological networks

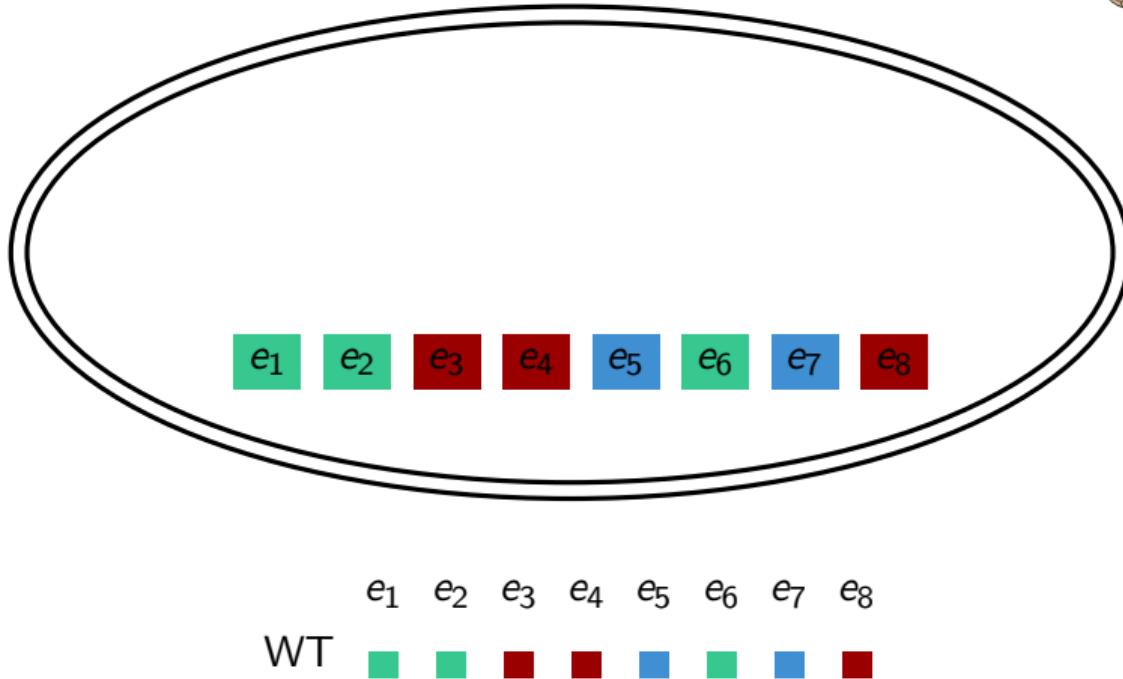


**Effects** are things we can measure: high-dimensional phenotypes, e.g., metabolomics, proteomics, etc.

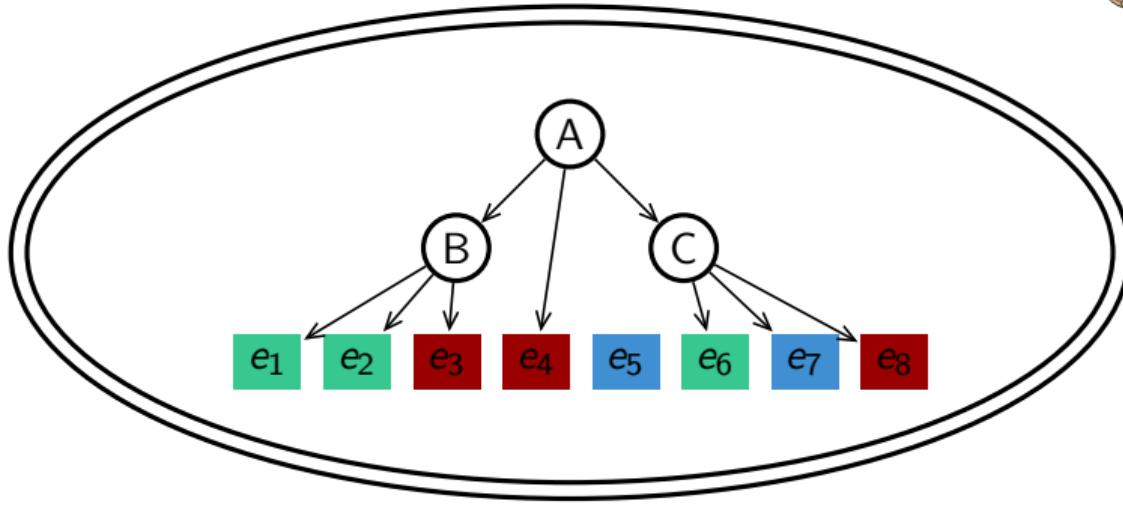
# Uncovering biological networks



School of Medicine  
and Public Health  
UNIVERSITY OF WISCONSIN-MADISON



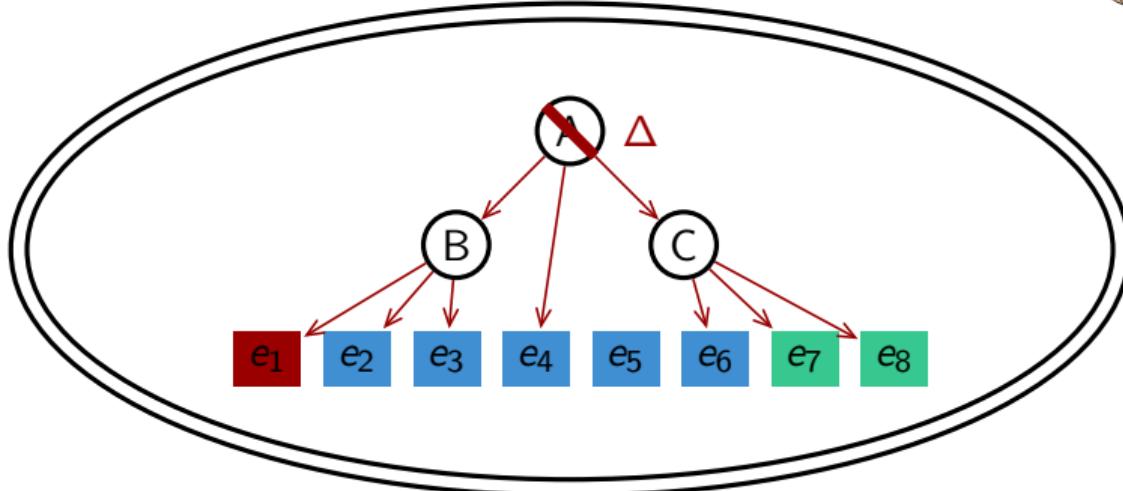
# Uncovering biological networks



$e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \ e_7 \ e_8$

WT

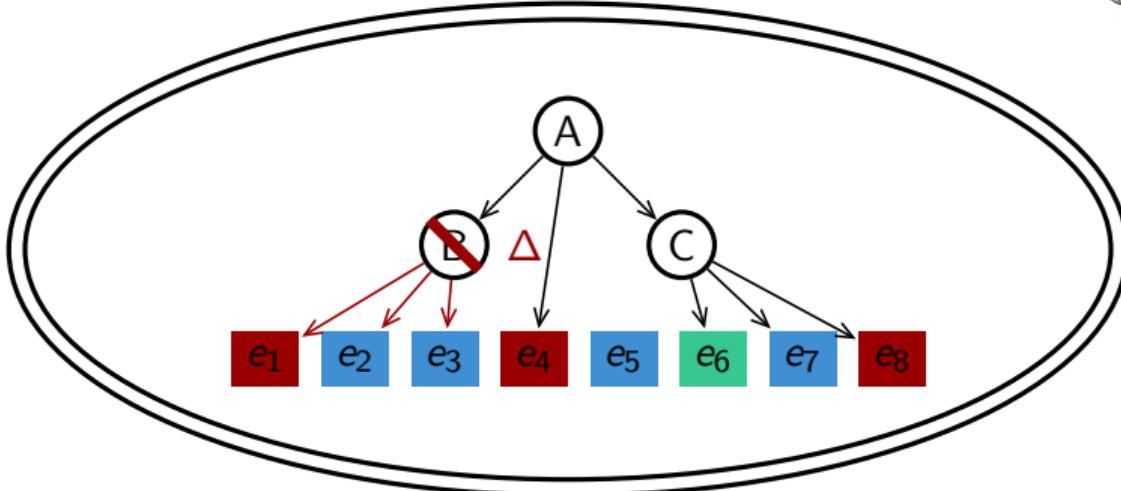
# Uncovering biological networks



$e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \ e_7 \ e_8$

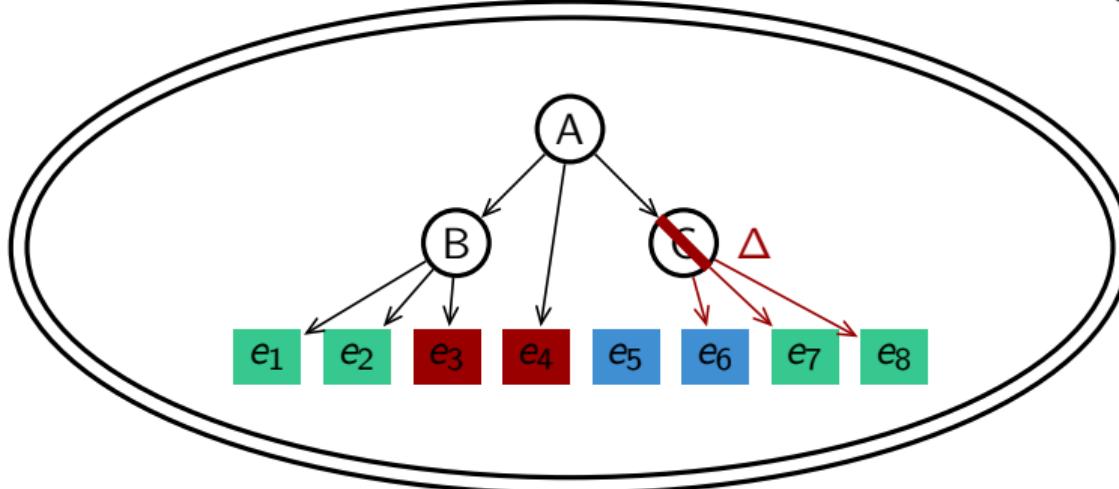
WT

# Uncovering biological networks



	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
WT	green	green	red	red	blue	green	blue	red
$A\Delta$	red	blue	blue	blue	blue	blue	green	green
$B\Delta$	red	blue	blue	red	blue	green	blue	red

# Uncovering biological networks



	e <sub>1</sub>	e <sub>2</sub>	e <sub>3</sub>	e <sub>4</sub>	e <sub>5</sub>	e <sub>6</sub>	e <sub>7</sub>	e <sub>8</sub>
WT	green	green	red	red	blue	green	blue	red
AΔ	red	blue	blue	blue	blue	blue	green	green
BΔ	red	blue	blue	red	blue	green	blue	red
CΔ	green	green	red	red	blue	blue	green	green

# High-dimensional knockout screens

Data ( $D$ )

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
WT	green	green	red	red	blue	green	blue	red
$A\Delta$	red	blue	blue	blue	blue	blue	green	green
$B\Delta$	red	blue	blue	red	blue	green	blue	red
$C\Delta$	green	green	red	red	blue	blue	green	green

## High-dimensional knockout screens



School of Medicine  
and Public Health  
UNIVERSITY OF WISCONSIN-MADISON

Data ( $D$ )

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
WT	green	green	red	red	blue	green	blue	red
$A\Delta$	red	blue	blue	blue	blue	blue	green	green
$B\Delta$	red	blue	blue	red	blue	green	blue	red
$C\Delta$	green	green	red	red	blue	blue	green	green

## Differential expression

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
$A\Delta$	■	■	■	■		■	■	■
$B\Delta$	■	■	■					
$C\Delta$					■	■	■	

# High-dimensional knockout screens



Data ( $D$ )

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
WT	green	green	red	red	blue	green	blue	red
$A\Delta$	red	blue	blue	blue	blue	blue	green	green
$B\Delta$	red	blue	blue	red	blue	green	blue	red
$C\Delta$	green	green	red	red	blue	blue	green	green

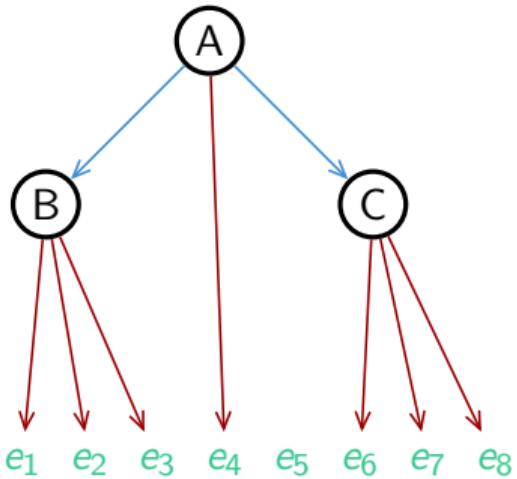
Differential expression

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
$A\Delta$	black							
$B\Delta$	black							
$C\Delta$	black							

Differential expression log likelihood ratio  $\left( R : R_{ae} = \log \frac{\mathbb{P}(D_{ae}|H_1)}{\mathbb{P}(D_{ae}|H_0)} \right)$



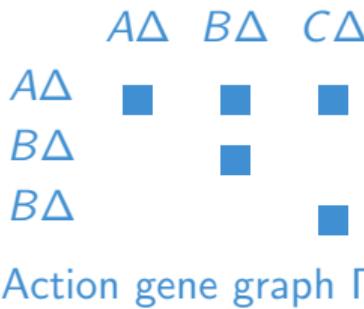
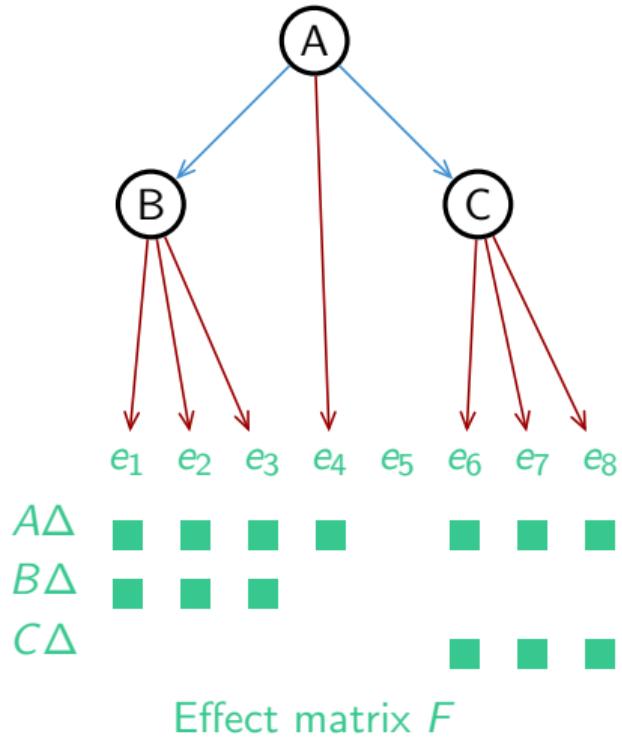
# Nested effect model (NEM)



$A\Delta$	[■]	[■]	[■]	[■]	[■]	[■]	[■]
$B\Delta$	[■]	[■]	[■]				
$C\Delta$				[■]	[■]	[■]	

Effect matrix  $F$

# Nested effect model (NEM)



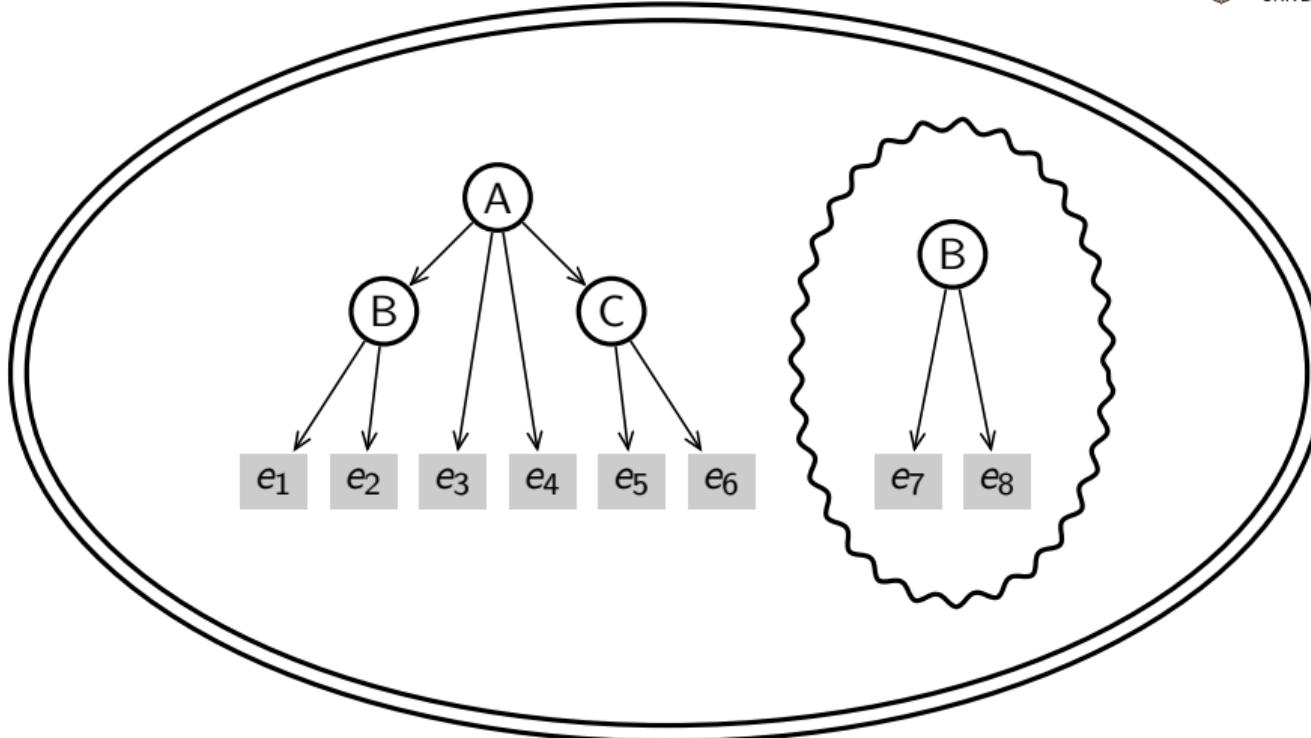
The Attachments  $\Theta$  is a 3x8 grid of red squares. The columns are labeled  $e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8$  and the rows are labeled  $A\Delta, B\Delta, C\Delta$ .

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
$A\Delta$						■		
$B\Delta$	■	■	■					
$C\Delta$							■	■

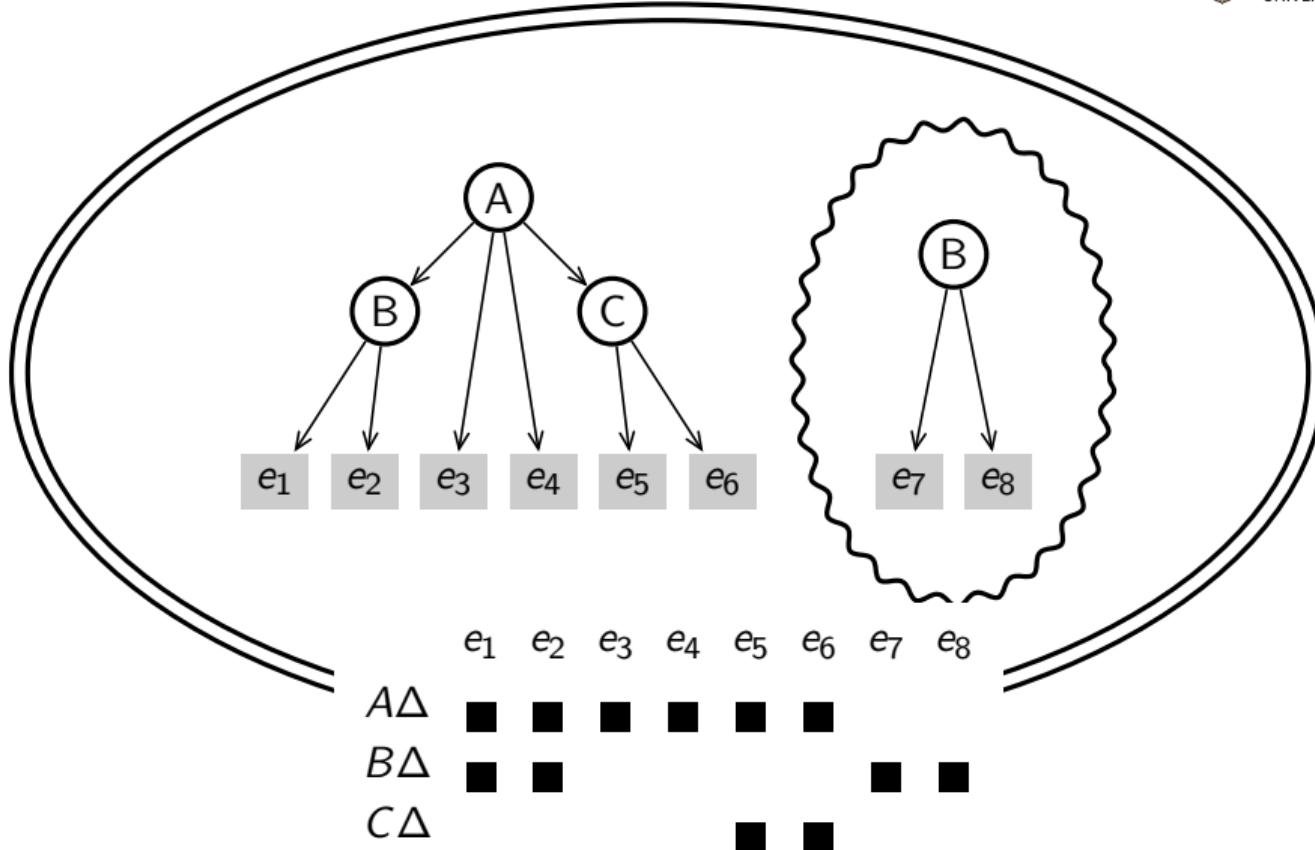
Attachments  $\Theta$

$$F = \Gamma \Theta$$

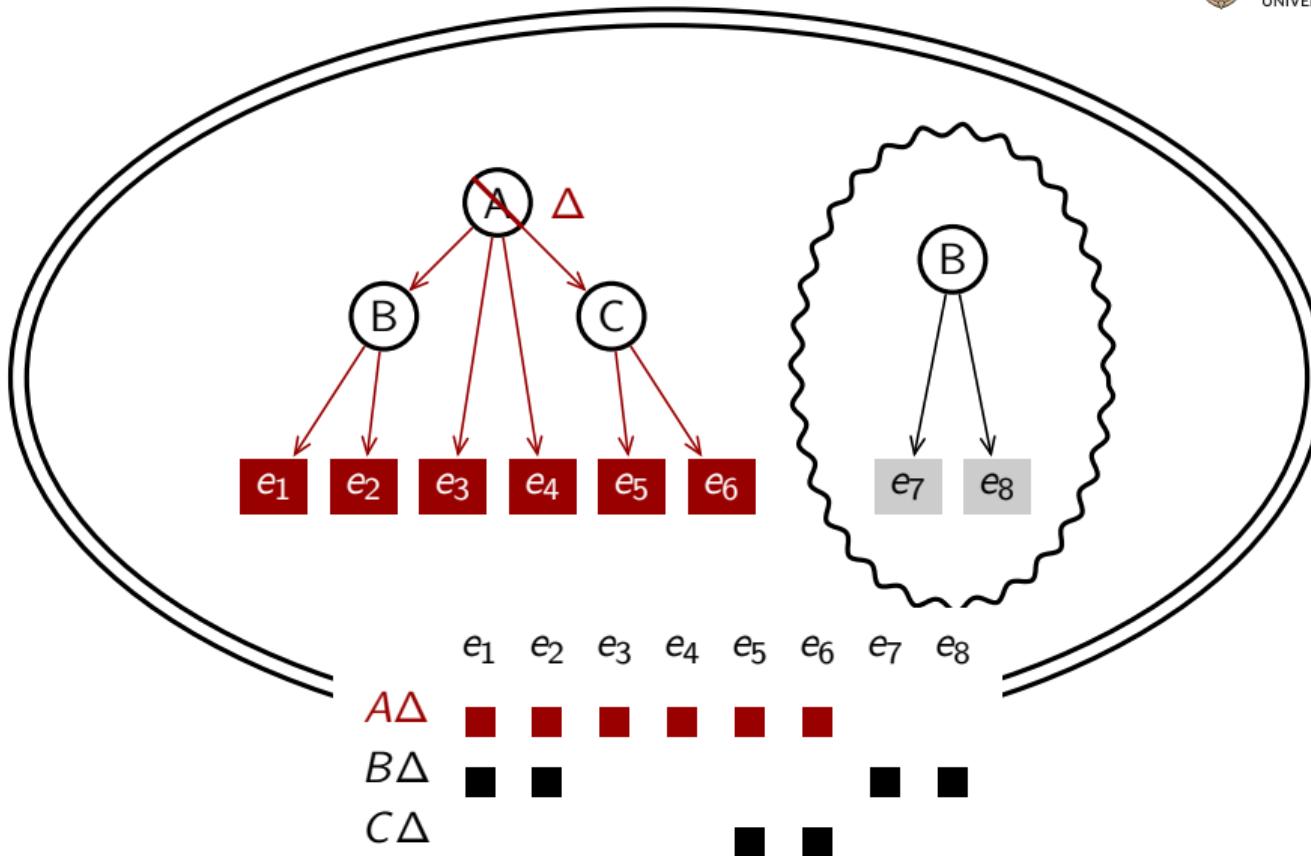
# Context-specific nested effect model (CSNEM)



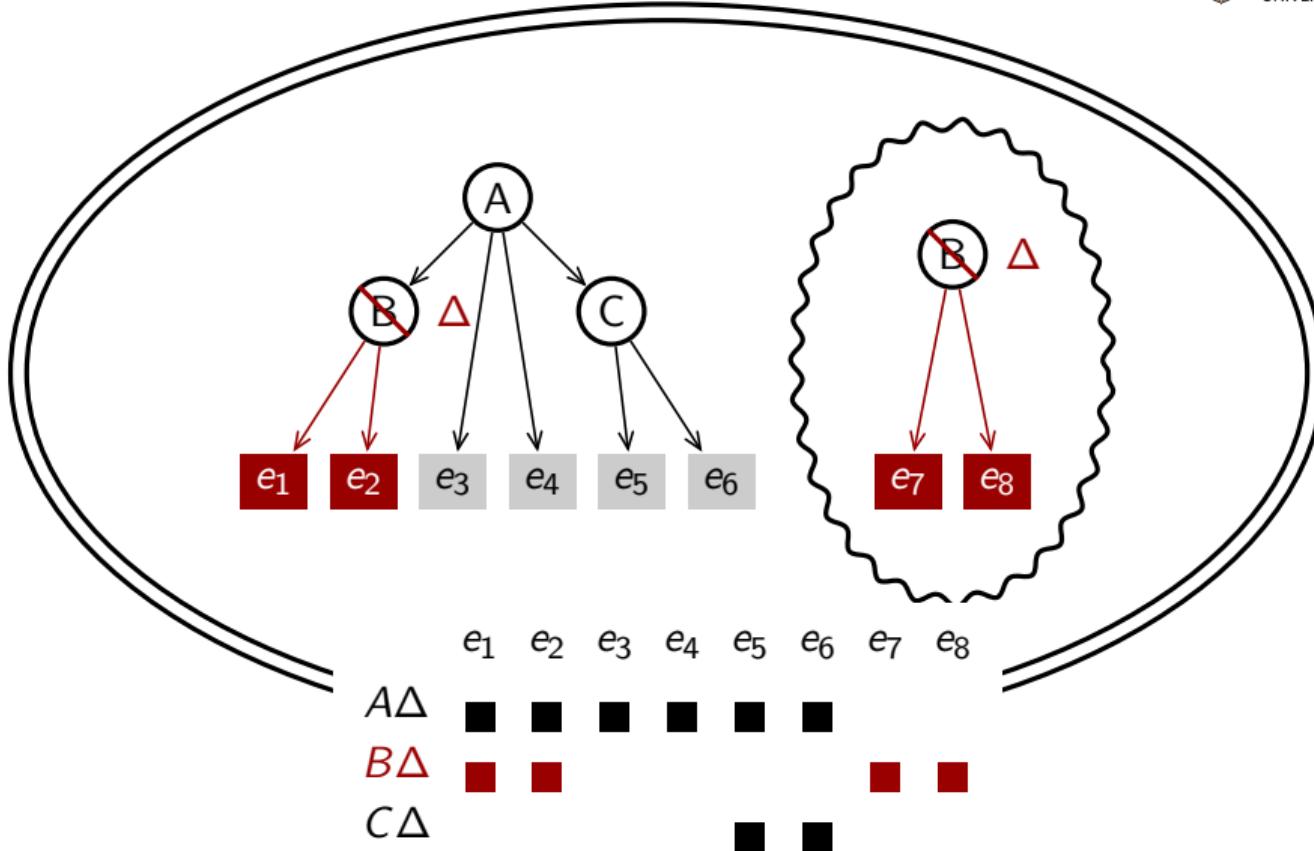
# Context-specific nested effect model (CSNEM)



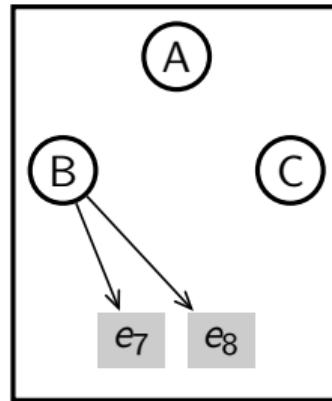
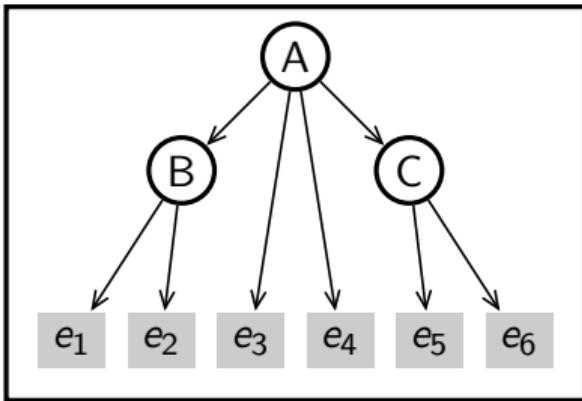
## Context-specific nested effect model (CSNEM)



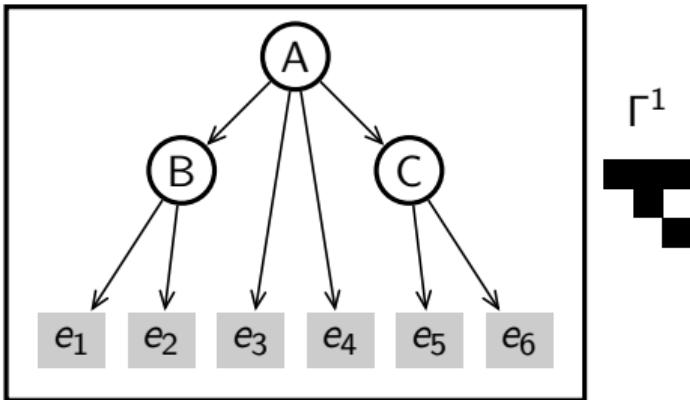
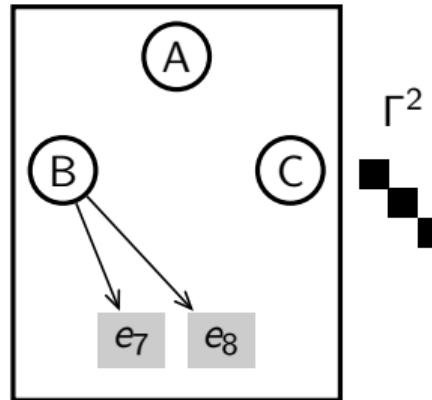
# Context-specific nested effect model (CSNEM)



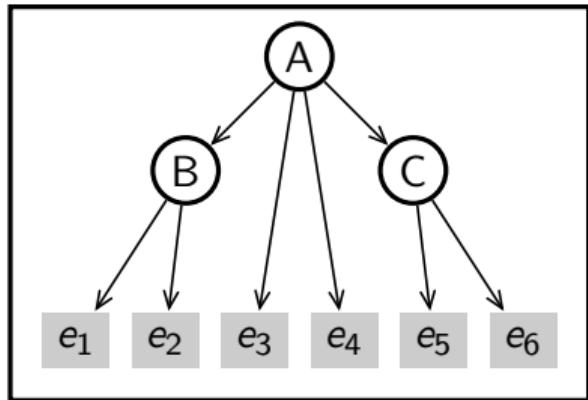
# CSNEM as a mixture of NEMs



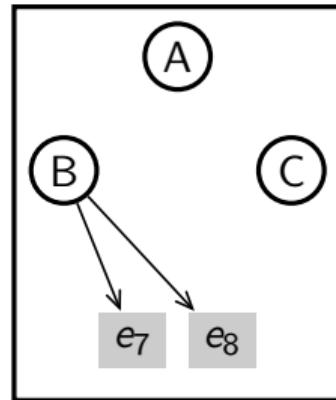
# CSNEM as a mixture of NEMs

 $\Gamma^1$  $\Gamma^2$

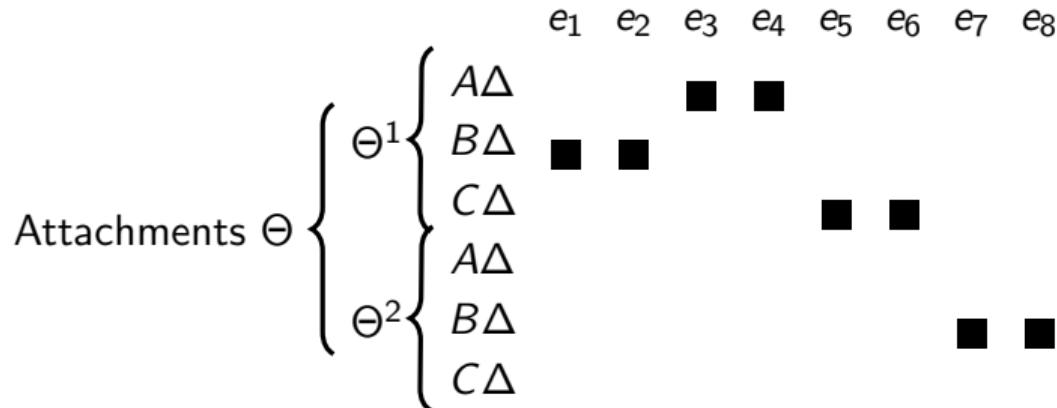
# CSNEM as a mixture of NEMs



$\Gamma^1$



$\Gamma^2$



# Likelihood formulation

$$\underbrace{L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k)}_{\text{CSNEM Likelihood}} = \prod_{i=1}^k \underbrace{L_{\text{NEM}}(\Gamma^i, \Theta^i)}_{\text{NEM Likelihood}}$$

# Likelihood formulation

$$\underbrace{L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k)}_{\text{CSNEM Likelihood}} = \prod_{i=1}^k \underbrace{L_{\text{NEM}}(\Gamma^i, \Theta^i)}_{\text{NEM Likelihood}}$$

Tresch and Markowetz (2008):

$$\log L_{\text{NEM}}(\Gamma^i, \Theta^i) = \sum_{(a,e) \in \mathcal{A} \times \mathcal{E}} \log \mathbb{P}(D_{ae} | (\Gamma^i \Theta^i)_{ae})$$

# Likelihood formulation

$$\underbrace{L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k)}_{\text{CSNEM Likelihood}} = \prod_{i=1}^k \underbrace{L_{\text{NEM}}(\Gamma^i, \Theta^i)}_{\text{NEM Likelihood}}$$

Tresch and Markowetz (2008):

$$\log L_{\text{NEM}}(\Gamma^i, \Theta^i) = \sum_{(a,e) \in \mathcal{A} \times \mathcal{E}} \log \mathbb{P}(D_{ae} | (\Gamma^i \Theta^i)_{ae}) = \text{tr}(\Gamma^i \Theta^i R^T) + c$$

$$R = \text{log likelihood ratio matrix: } R_{ae} = \log \frac{\mathbb{P}(D_{ae} | H_1)}{\mathbb{P}(D_{ae} | H_0)}$$

# Likelihood formulation

$$\underbrace{L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k)}_{\text{CSNEM Likelihood}} = \prod_{i=1}^k \underbrace{L_{\text{NEM}}(\Gamma^i, \Theta^i)}_{\text{NEM Likelihood}}$$

Tresch and Markowetz (2008):

$$\log L_{\text{NEM}}(\Gamma^i, \Theta^i) = \sum_{(a,e) \in \mathcal{A} \times \mathcal{E}} \log \mathbb{P}(D_{ae} | (\Gamma^i \Theta^i)_{ae}) = \text{tr}(\Gamma^i \Theta^i R^T) + c$$

$$R = \text{log likelihood ratio matrix: } R_{ae} = \log \frac{\mathbb{P}(D_{ae} | H_1)}{\mathbb{P}(D_{ae} | H_0)}$$

⋮

$$\log L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k) = \text{tr}(\Gamma \Theta R^{*T}) + c^*$$

# CSNEM Log Likelihood

$$\log L_{\text{CSNEM}}(\Gamma^1, \dots, \Gamma^k, \Theta^1, \dots, \Theta^k) =$$

$$\text{tr} \left( \underbrace{\begin{bmatrix} \Gamma^1 & 0 & \cdots & 0 \\ 0 & \Gamma^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Gamma^k \end{bmatrix}}_{\text{Block diagonal } \Gamma} \underbrace{\begin{bmatrix} \Theta^1 \\ \Theta^2 \\ \vdots \\ \Theta^k \end{bmatrix}}_{\text{"Big" } \Theta; (|\mathcal{A}|k \times |\mathcal{E}|)} \underbrace{\begin{bmatrix} R^T & R^T & \cdots & R^T \end{bmatrix}}_{k \text{ copies}} \right) + c$$

## Learning approach

Use MC-EMiNEM (Niederberger et al. 2012) to solve the modified NEM learning problem.

$$\log L(\Gamma\Theta) + \underbrace{\sum_{(a,b) \in \mathcal{A} \times \mathcal{A}} \log \mathbb{P}(\Gamma_{a,b})}_{\text{edge-wise prior}} + \log \mathbb{P}(\Theta)$$

Bioconductor R package: `nem` (Fröhlich et al. 2008)

# Evaluation on simulated data



School of Medicine  
and Public Health

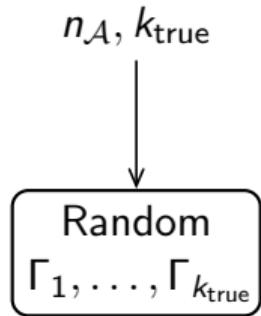
UNIVERSITY OF WISCONSIN-MADISON

# Evaluation on simulated data

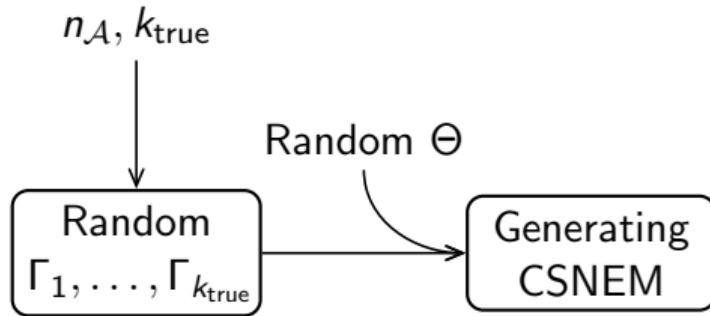


$n_{\mathcal{A}}, k_{\text{true}}$

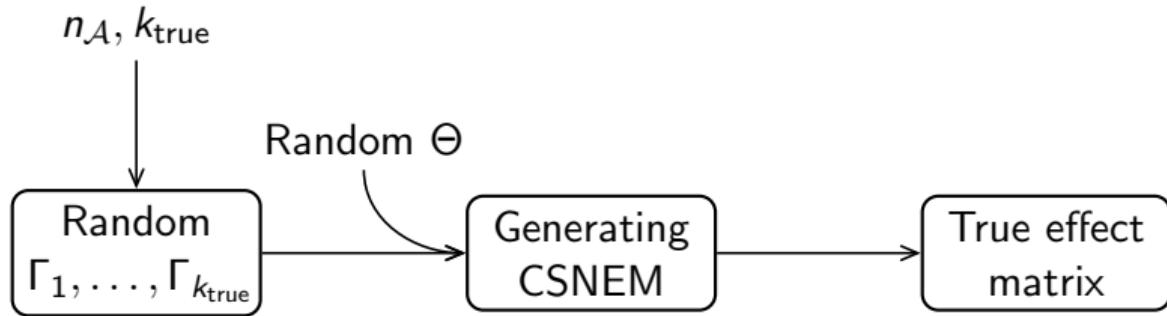
# Evaluation on simulated data



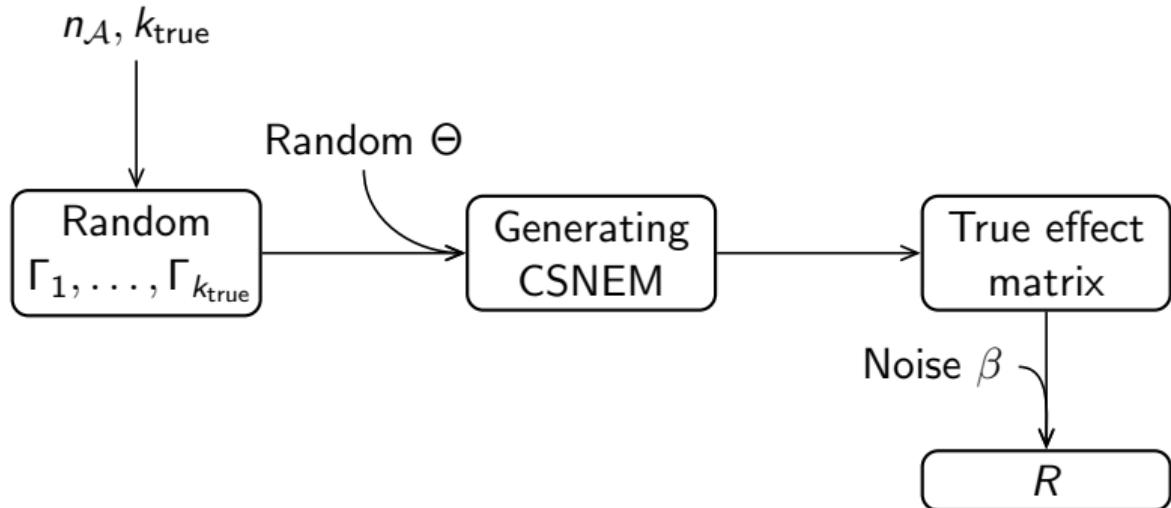
## Evaluation on simulated data



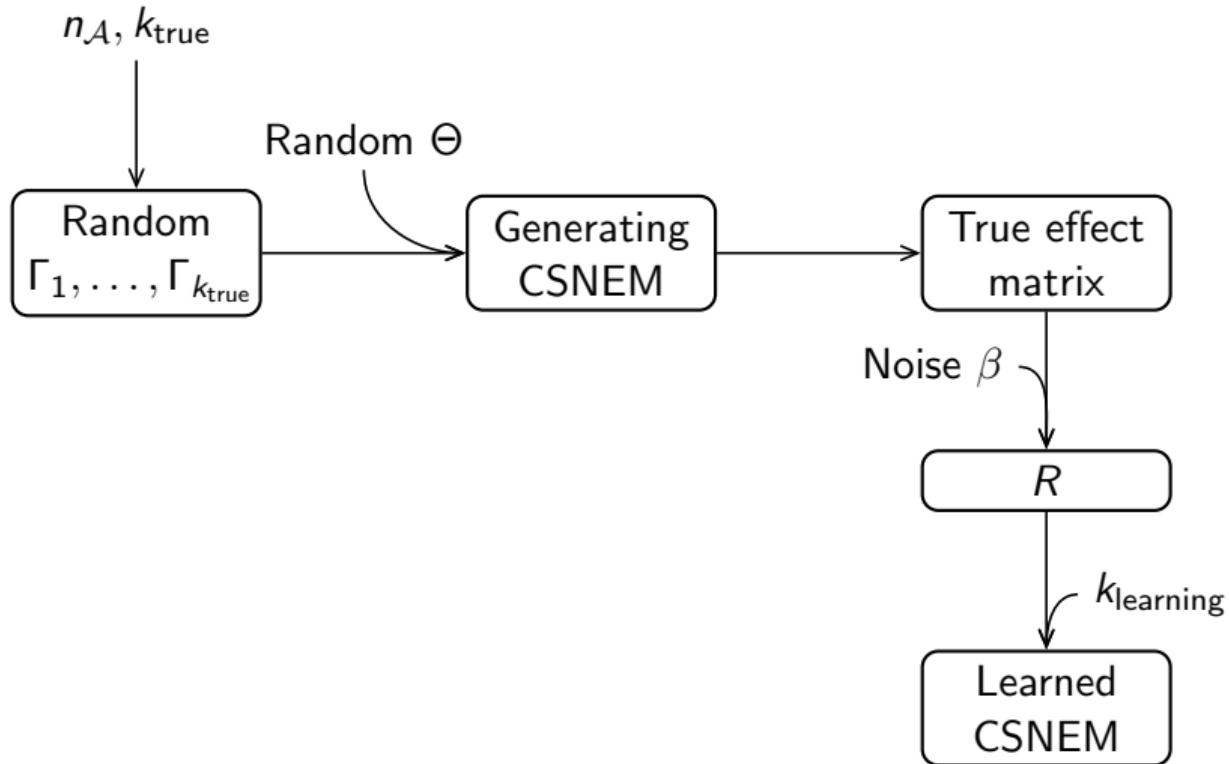
# Evaluation on simulated data



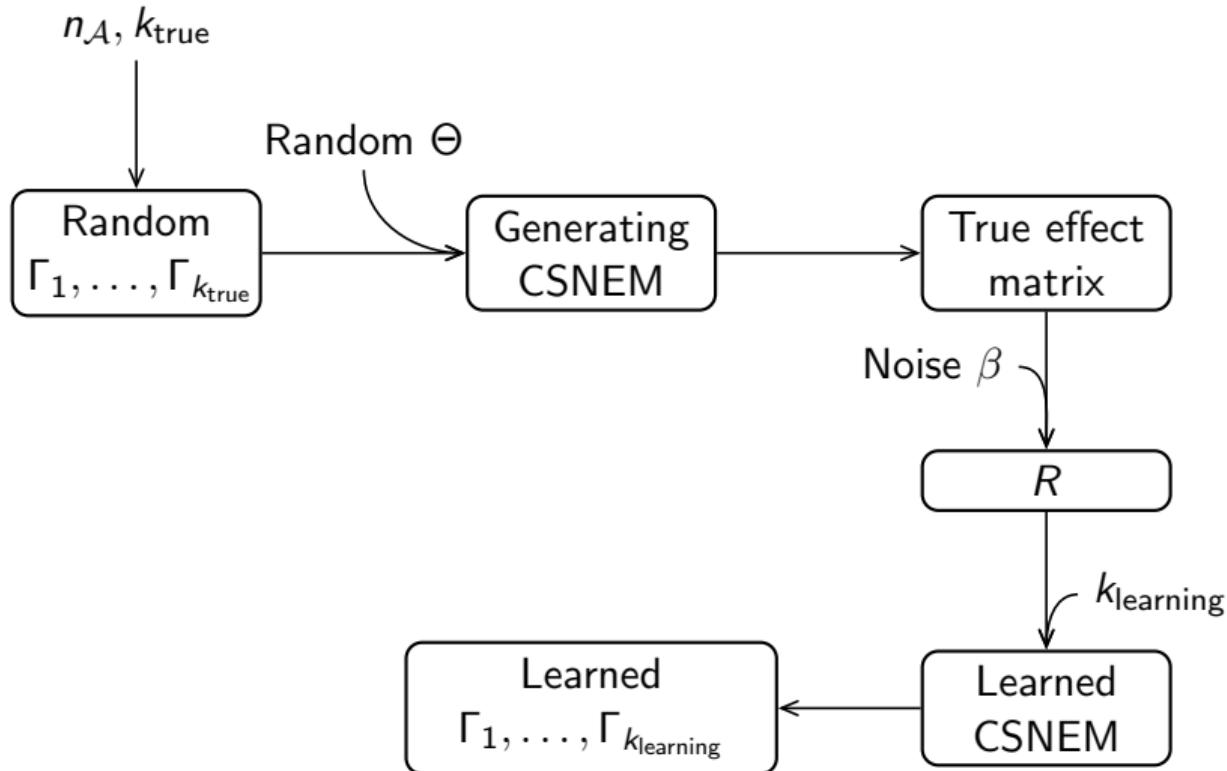
# Evaluation on simulated data



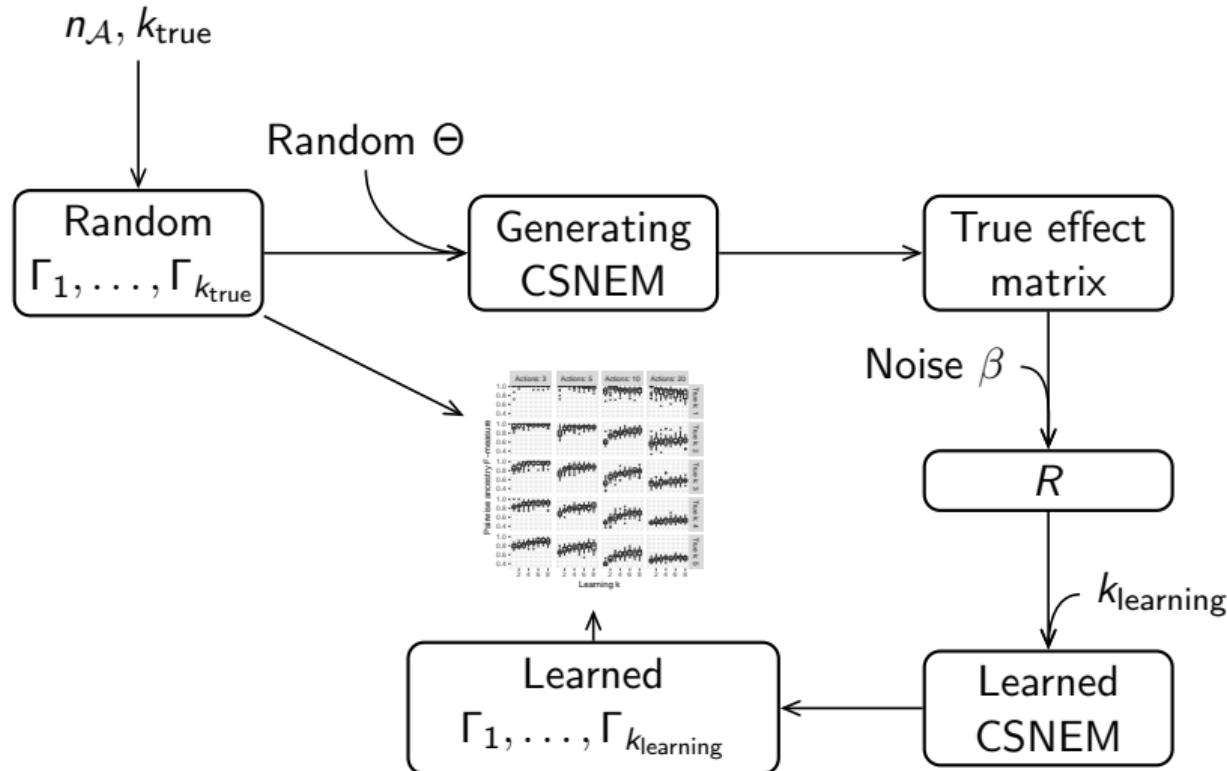
# Evaluation on simulated data



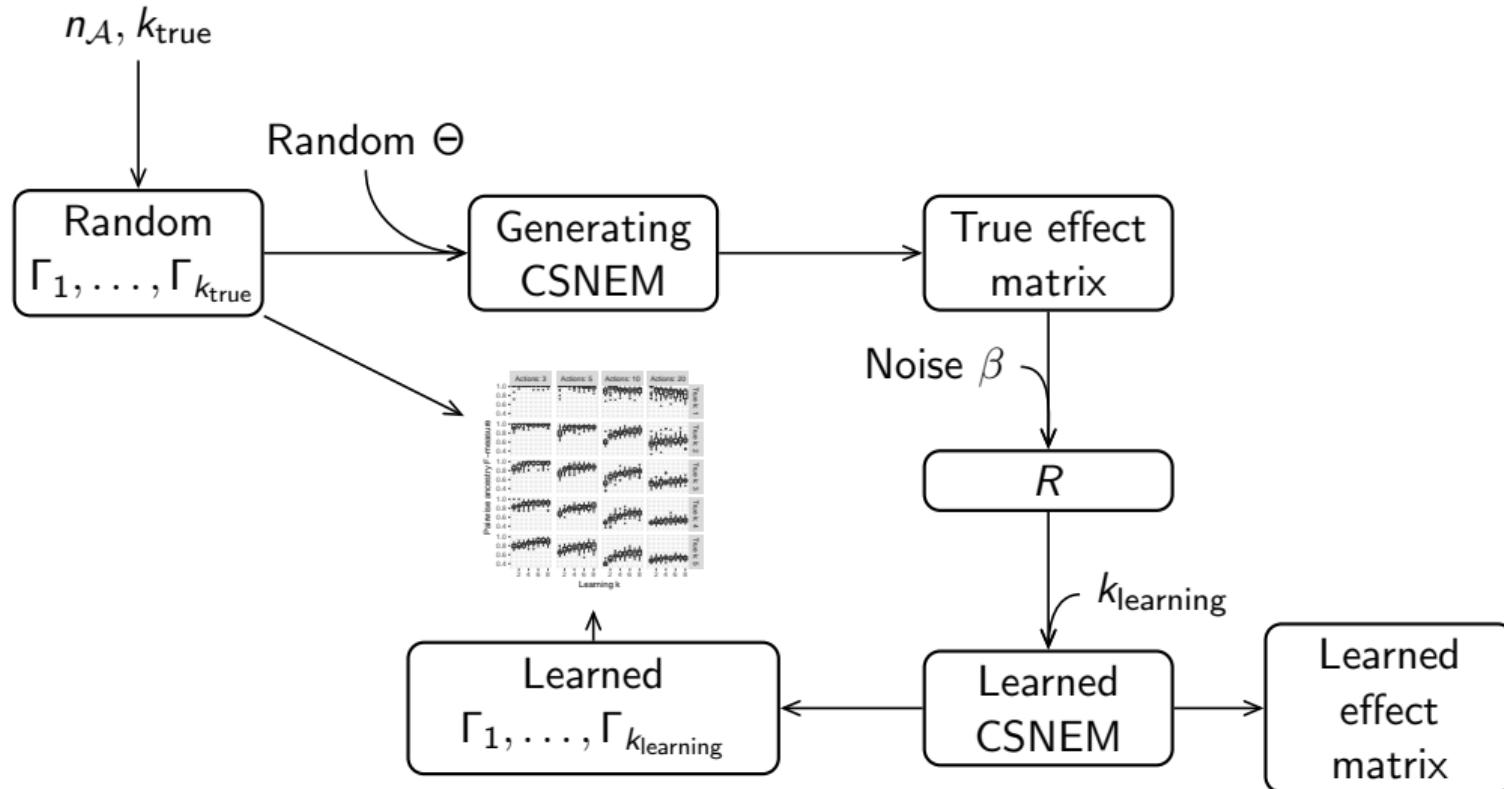
# Evaluation on simulated data



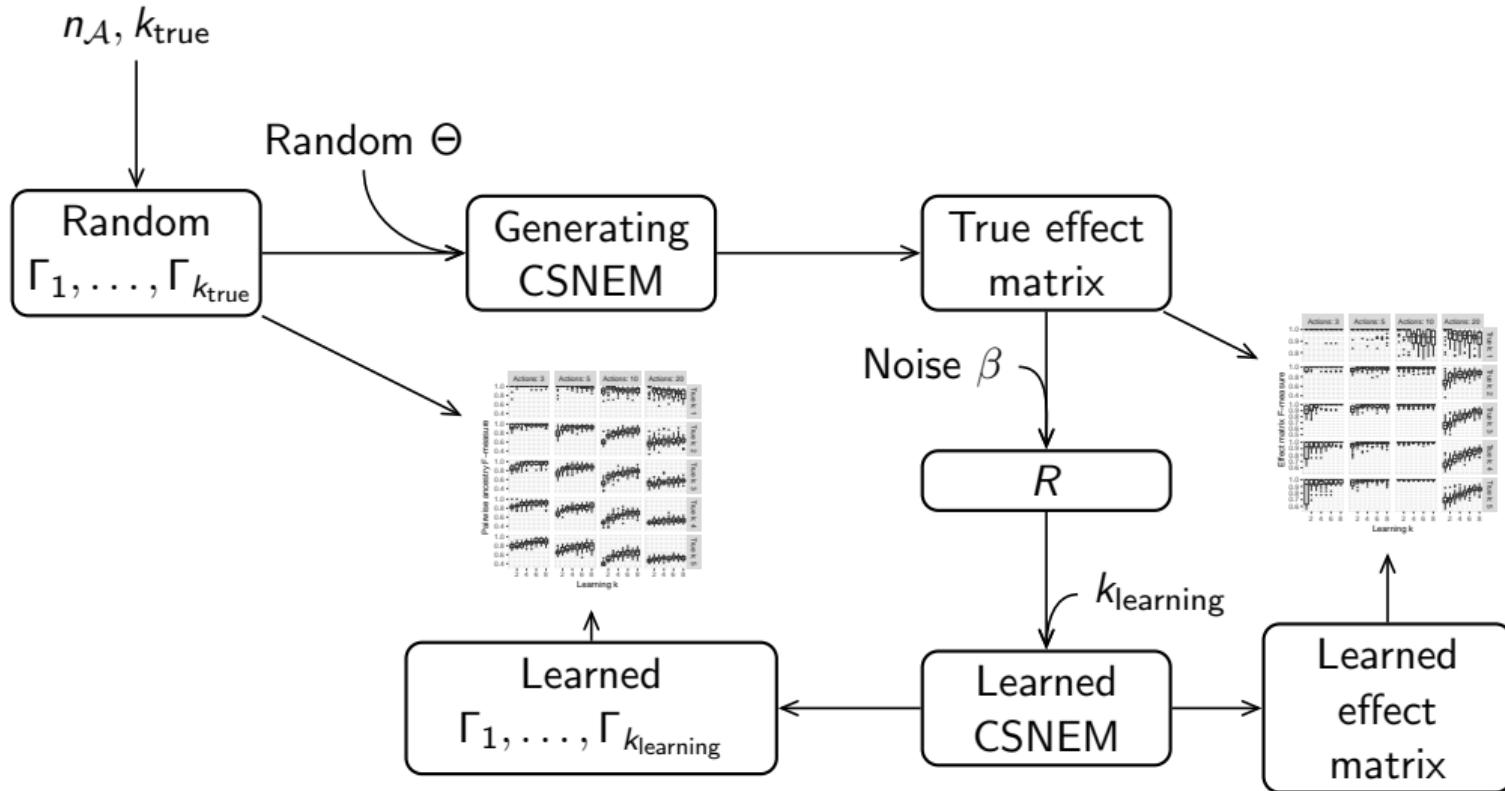
# Evaluation on simulated data



# Evaluation on simulated data

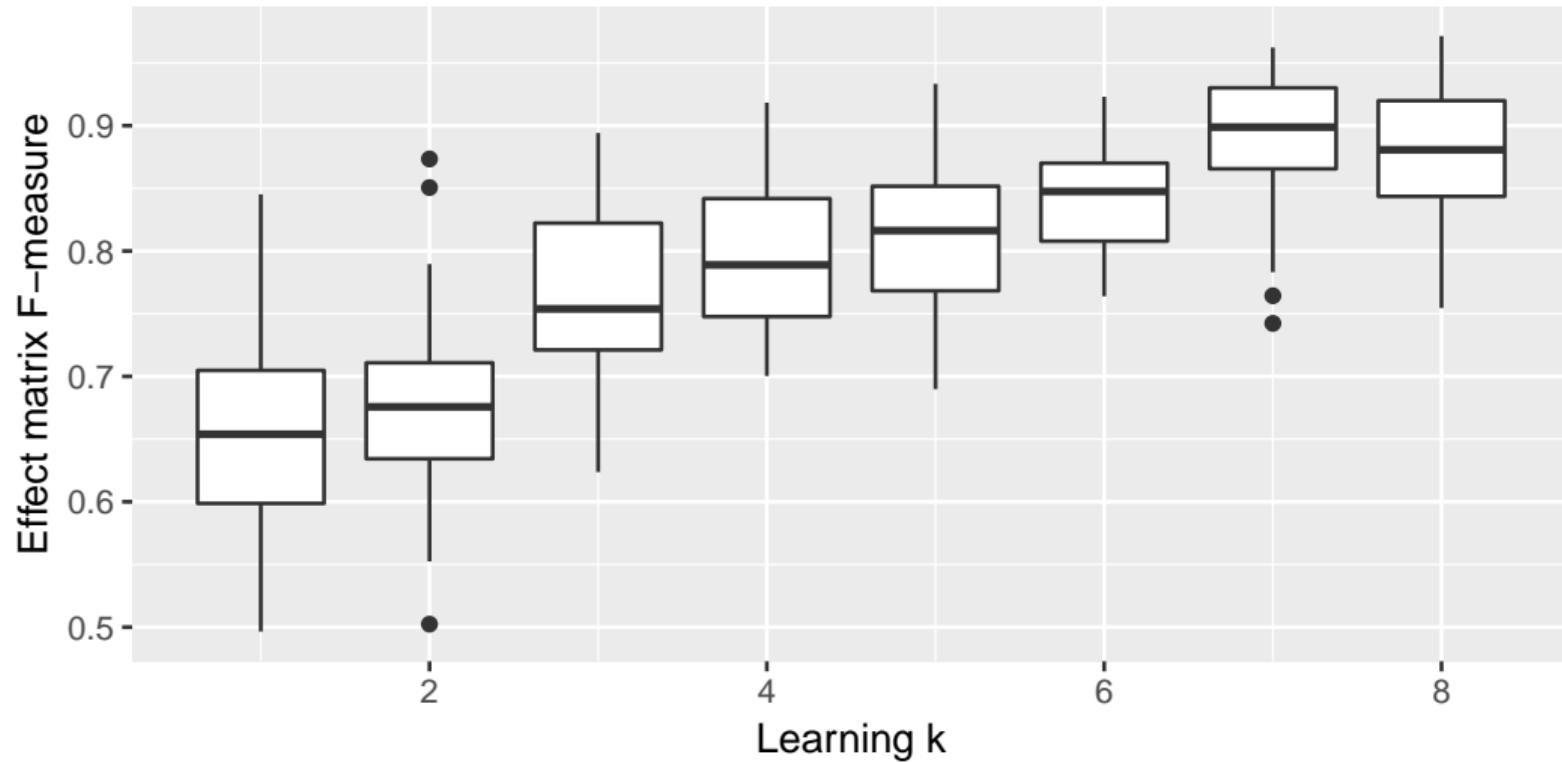


# Evaluation on simulated data

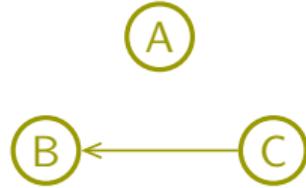
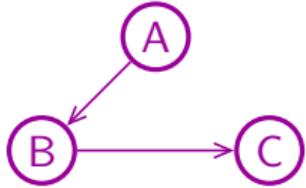
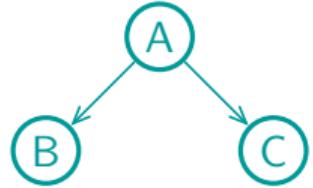


## Evaluation on simulated data

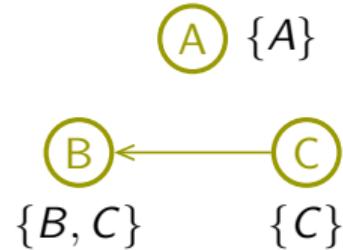
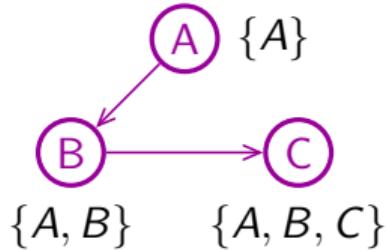
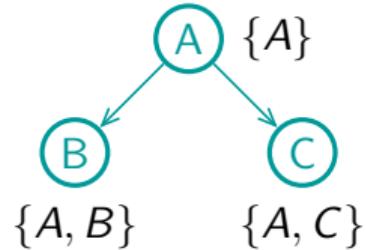
$n_A = 20$ ;  $k_{\text{true}} = 3$ ; low noise ( $\beta = 10$ )



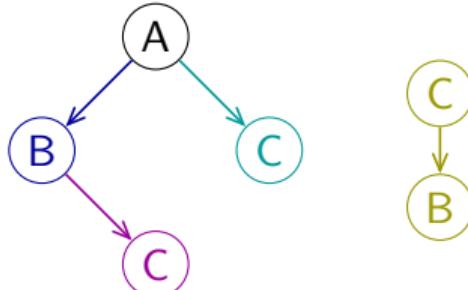
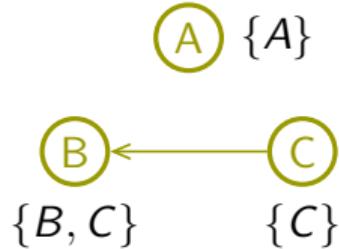
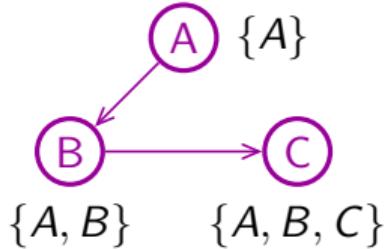
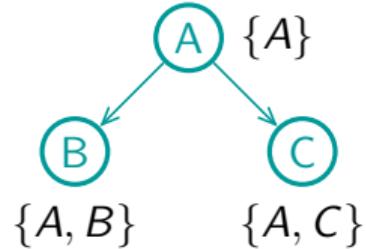
# Synthesis



# Synthesis



# Synthesis



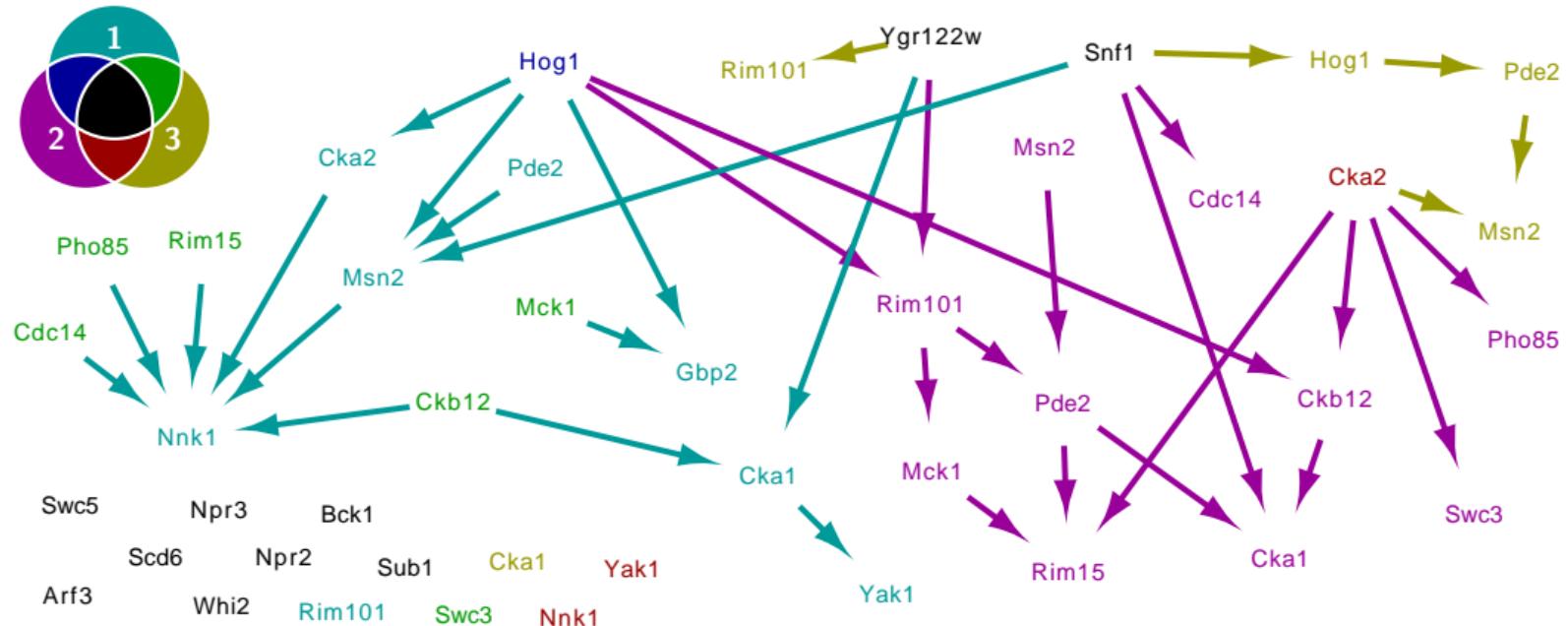
# *Saccharomyces cerevisiae* salt stress data

- ▶ Wild type strains and 28 single-gene knockout strains
- ▶ Gene expression measured by microarray
  - ▶ Measurement without salt treatment
  - ▶ Measurement 30 minutes after 0.7 M NaCl treatment
- ▶ Change in response:

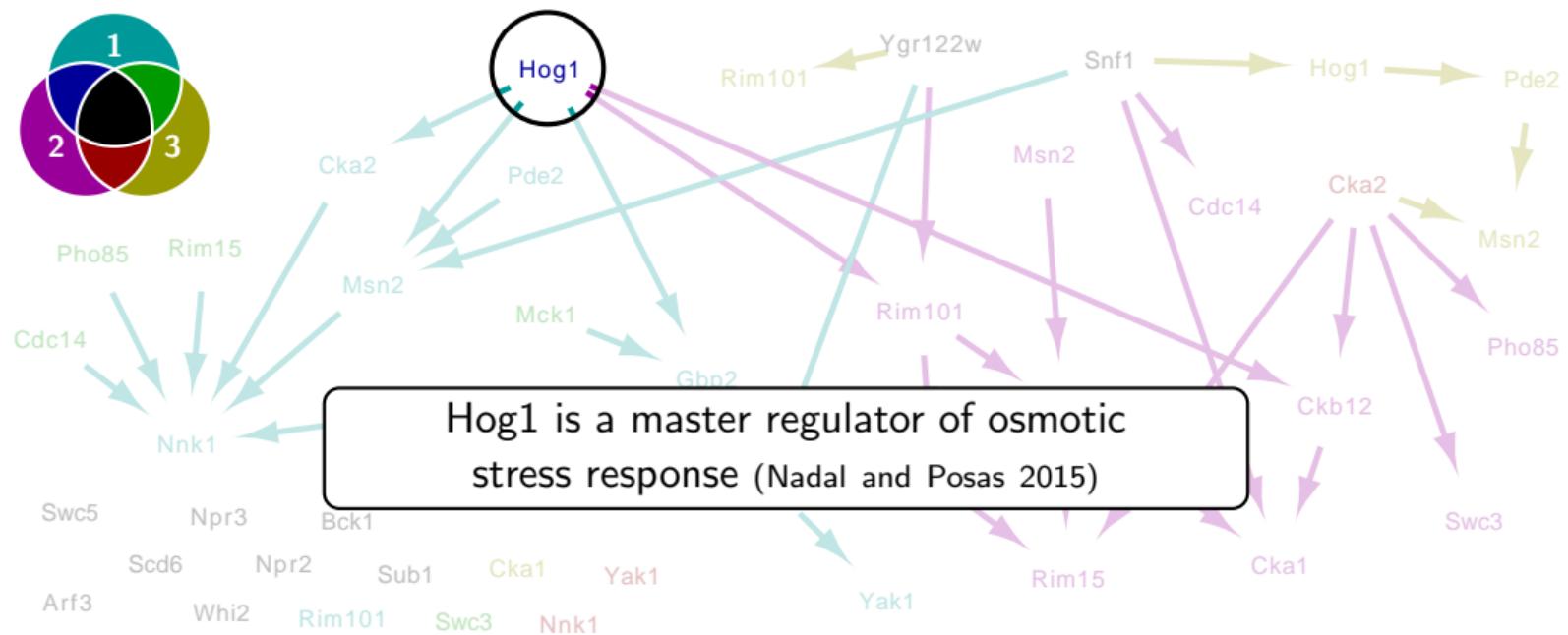
$$\log_2 \left( \frac{\Delta \text{ strain treated}}{\Delta \text{ strain untreated}} \middle/ \frac{\text{WT strain treated}}{\text{WT strain untreated}} \right)$$

(Berry and Gasch 2008; Lee et al. 2011)

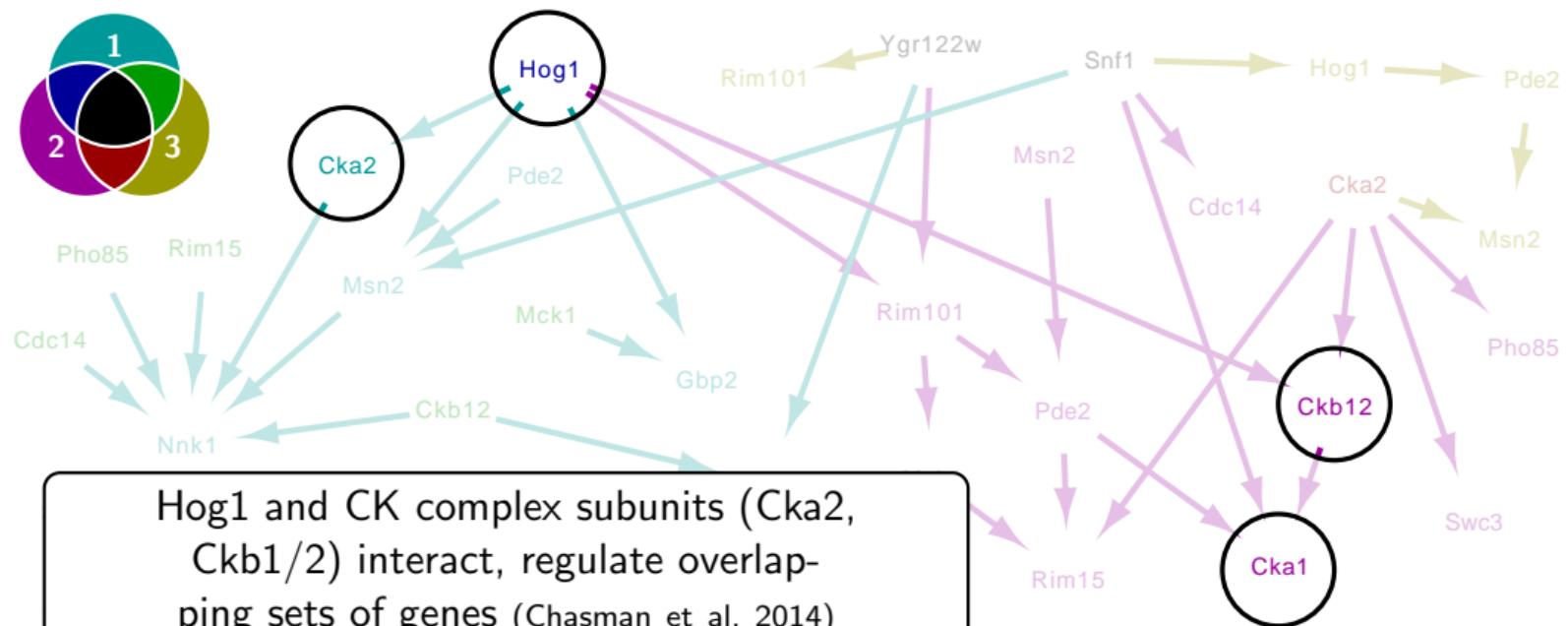
# Network learned from the *S. cerevisiae* data



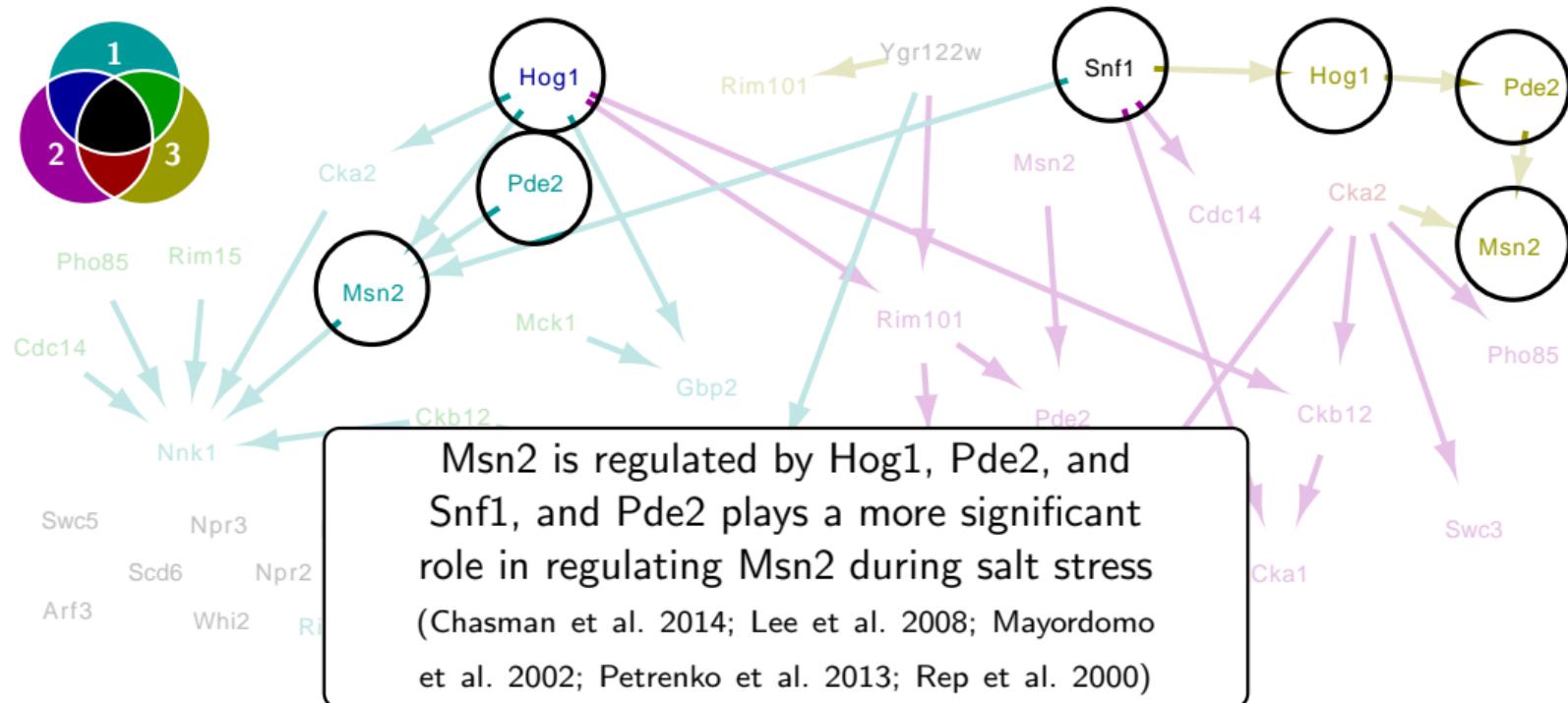
## Network learned from the *S. cerevisiae* data



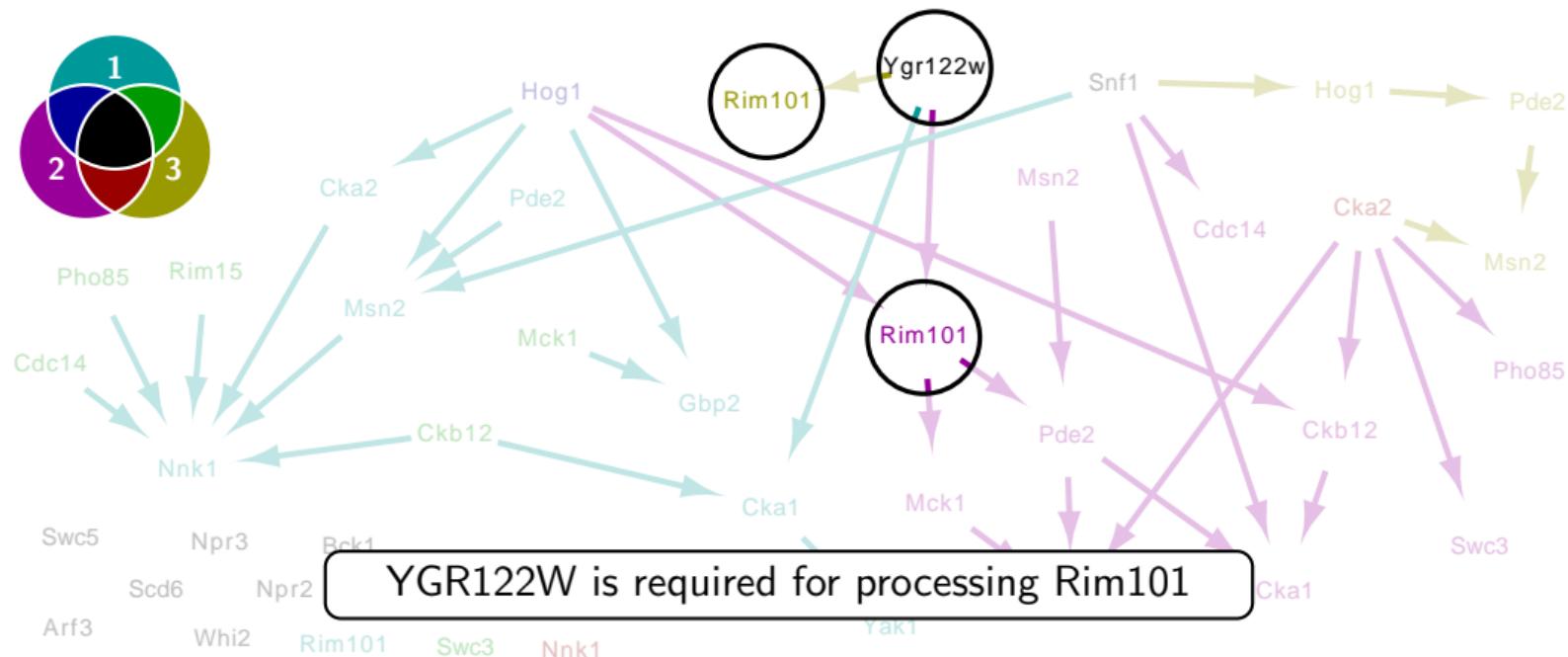
# Network learned from the *S. cerevisiae* data



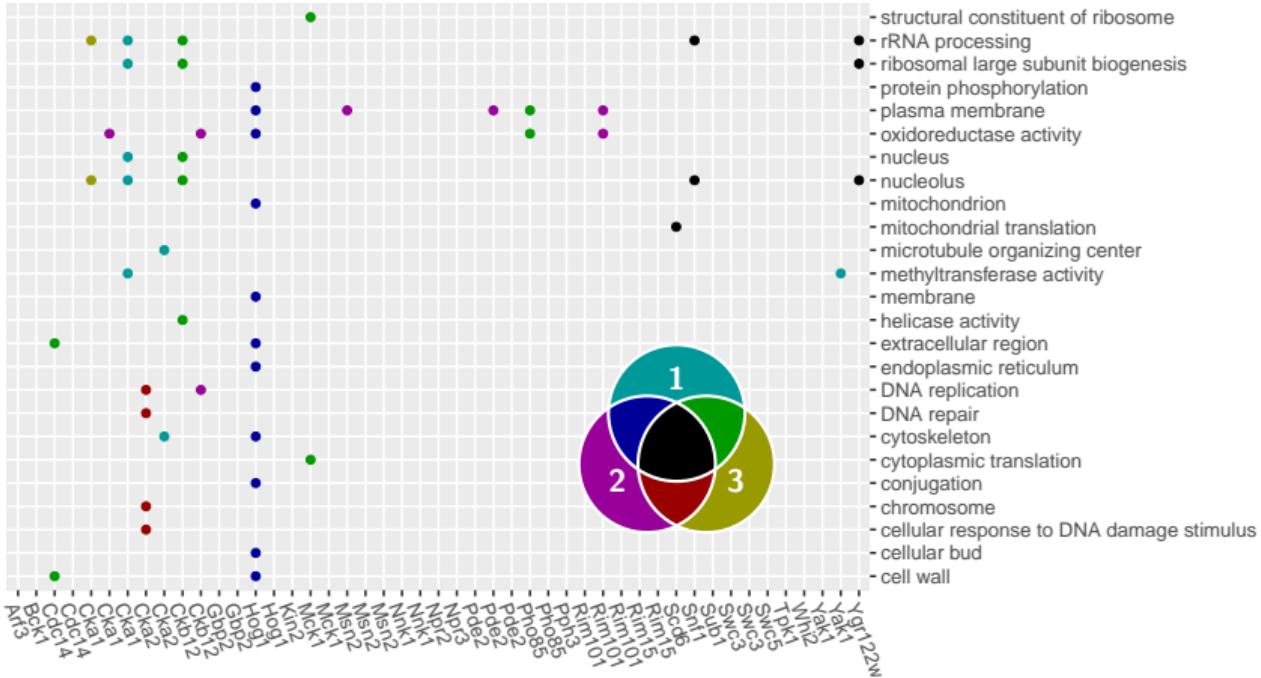
# Network learned from the *S. cerevisiae* data



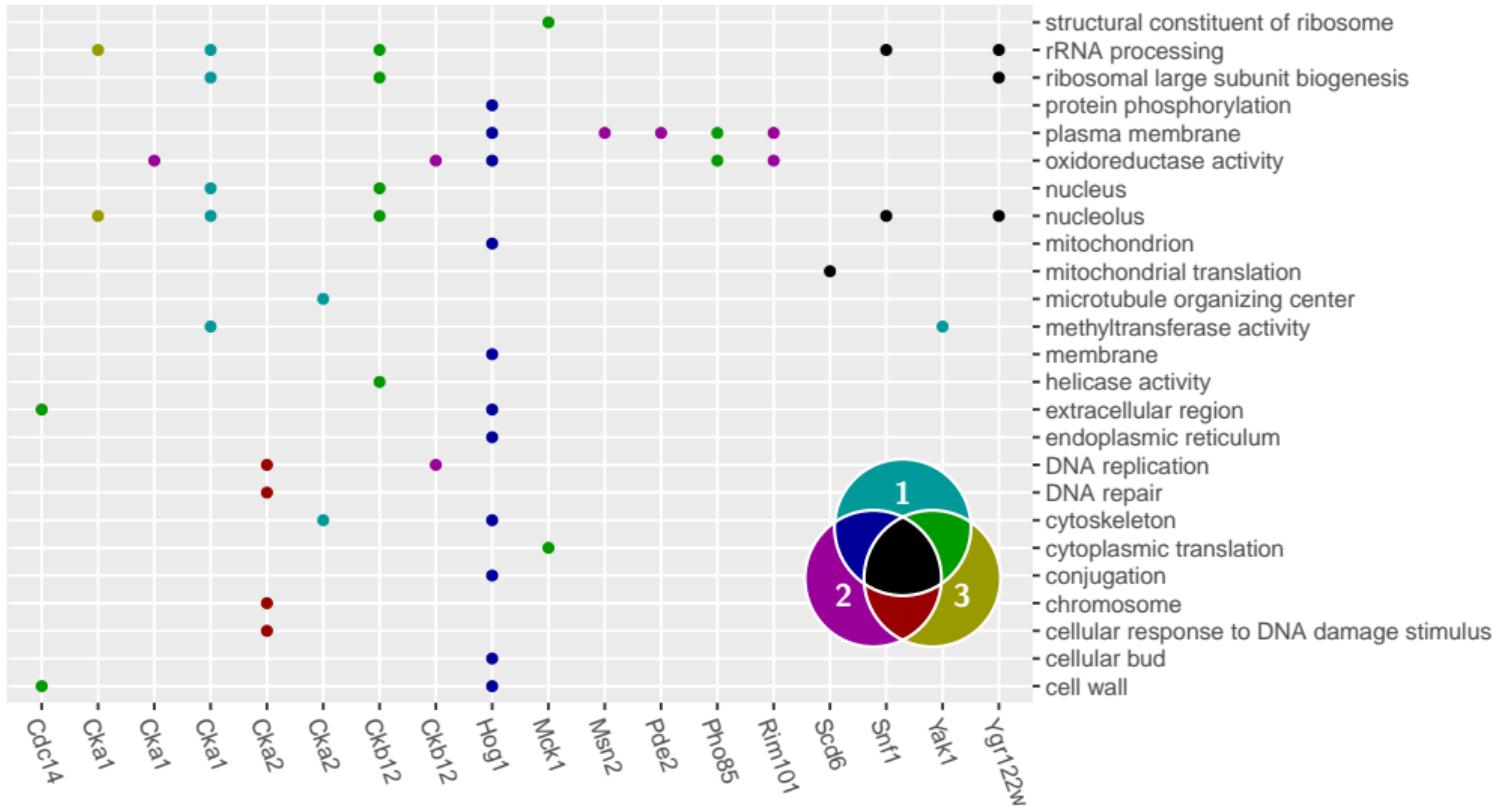
# Network learned from the *S. cerevisiae* data



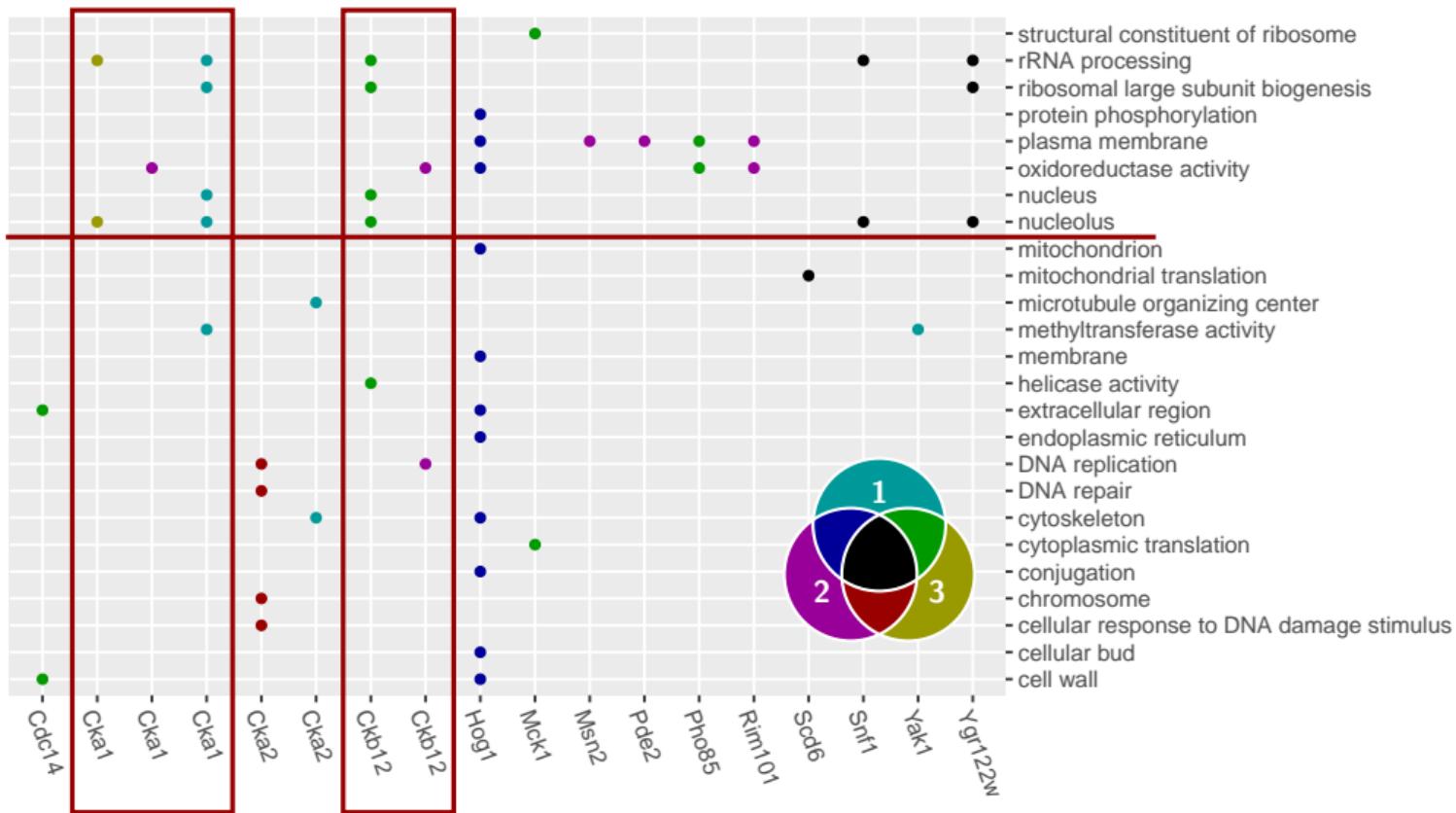
# Effect group GO term enrichments



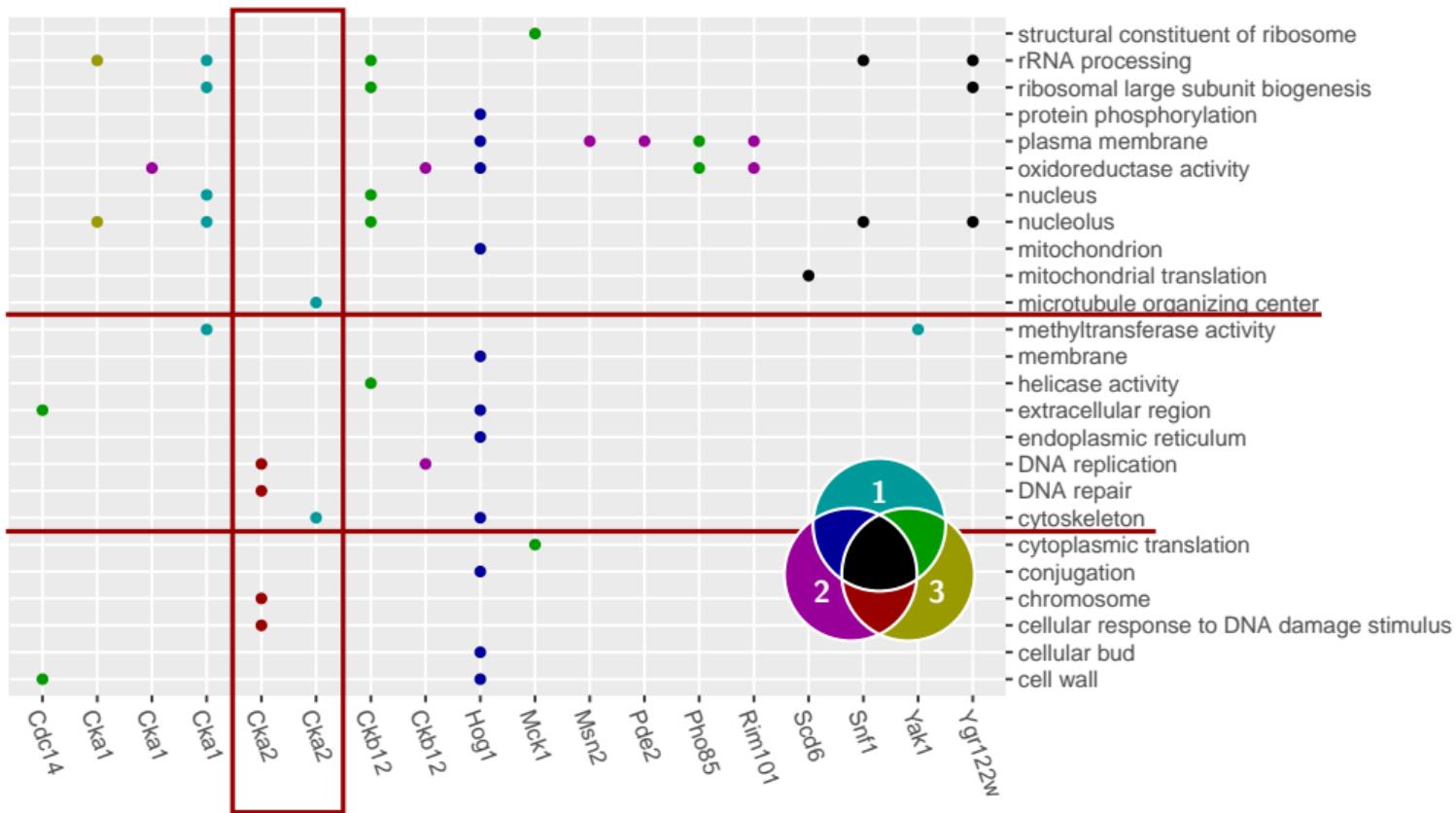
# Effect group GO term enrichments



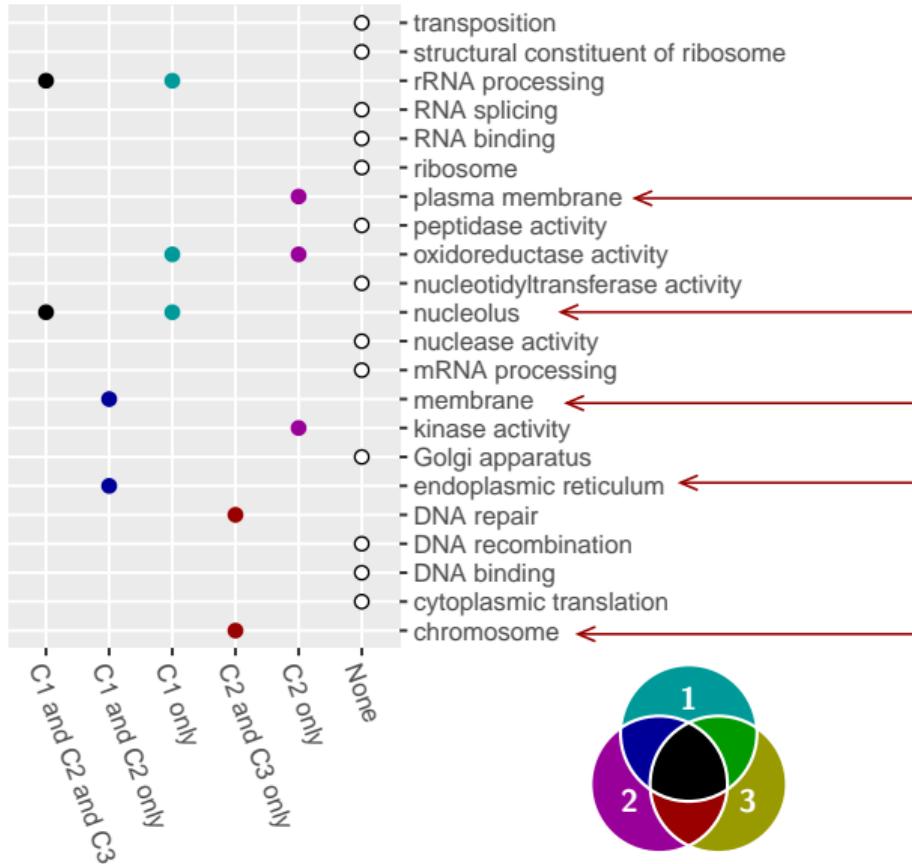
# Effect group GO term enrichments



# Effect group GO term enrichments



# Context-wise GO term enrichments



Location GO terms



# Conclusions

- ▶ We introduce a novel approach to modeling context-specific biological interactions.

# Conclusions

- ▶ We introduce a novel approach to modeling context-specific biological interactions.
- ▶ Simulations show that
  - ▶ context-specificity is recoverable, and
  - ▶ a context-specific model is necessary when the data-generating process is context-specific.
- ▶ Application to salt-stress data shows
  - ▶ context-specific effect groups are enriched for meaningful GO categories, and
  - ▶ the learned CSNEM aligns with known biology.

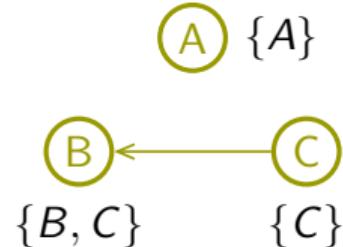
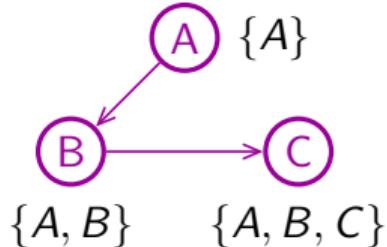
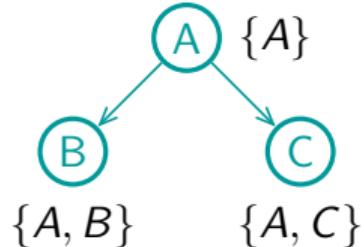
# Conclusions

- ▶ We introduce a novel approach to modeling context-specific biological interactions.
- ▶ Simulations show that
  - ▶ context-specificity is recoverable, and
  - ▶ a context-specific model is necessary when the data-generating process is context-specific.
- ▶ Application to salt-stress data shows
  - ▶ context-specific effect groups are enriched for meaningful GO categories, and
  - ▶ the learned CSNEM aligns with known biology.

## Acknowledgments

NIH/NLM grant T15 LM0007359, NIH/NIAID grant U54 AI117954, and NIH/NIGMS grant R01 GM083989.

# Identifiability



- ▶ Set of ancestries:  $\{\{A\}, \{C\}, \{A, B\}, \{A, C\}, \{B, C\}, \{A, B, C\}\}$
- ▶ CSNEMs with identical sets of actor ancestries can express the same effect patterns.