Learning by Fusing Heterogeneous Data

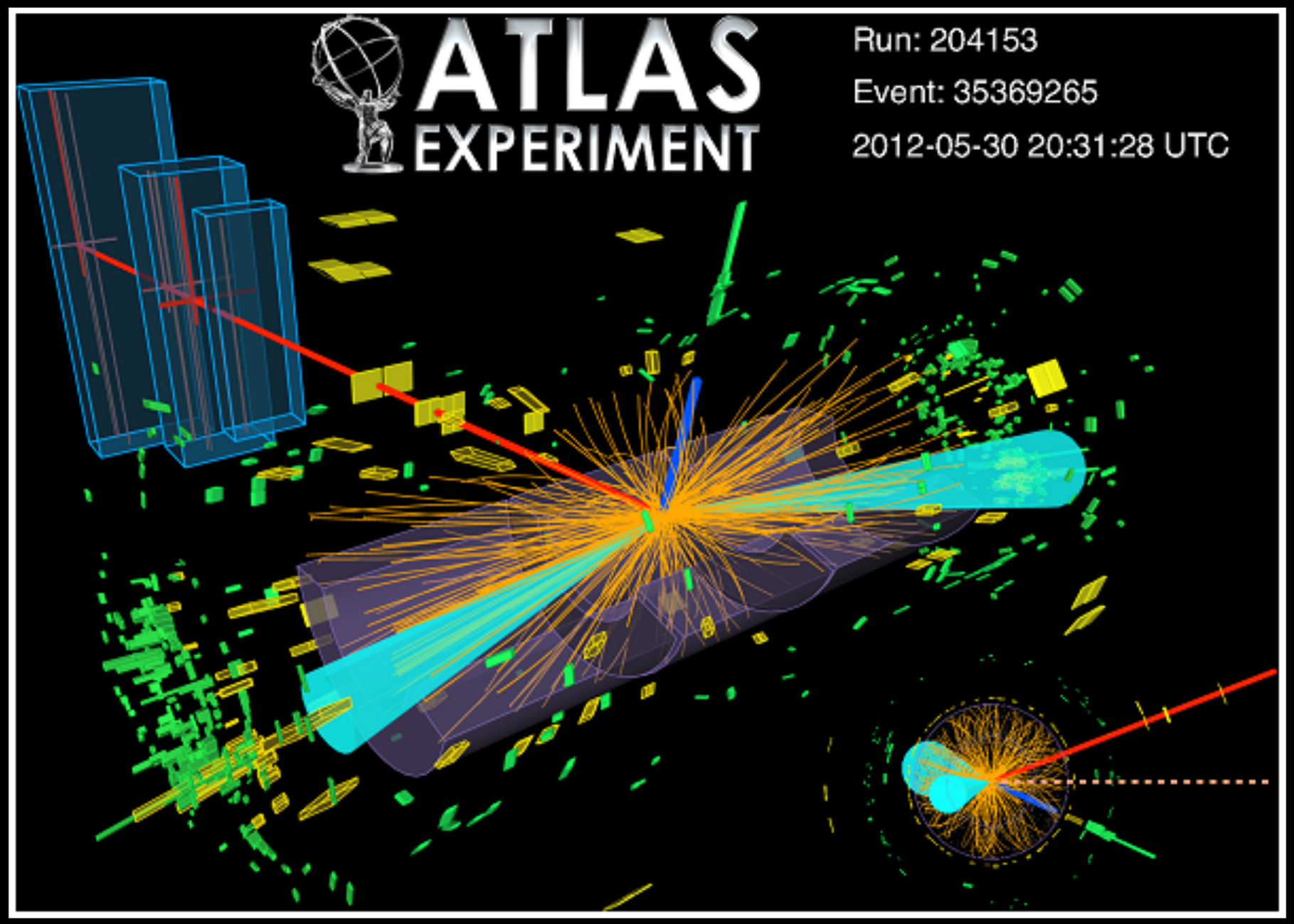
Marinka Zitnik

Thesis Defense, October 22 2015

Motivation

Large Heterogeneous Data Compendia

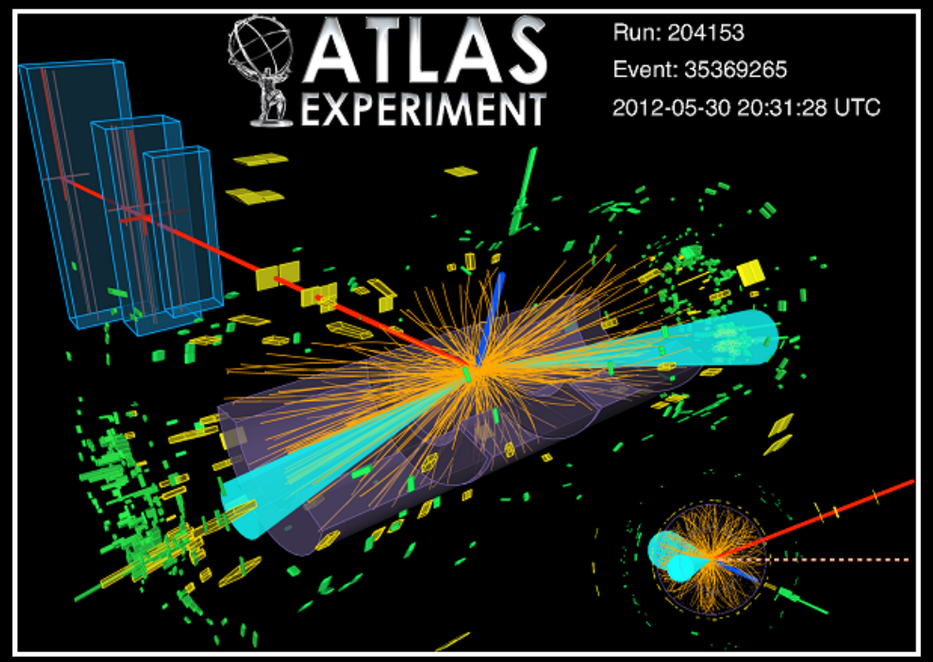
Large Heterogeneous Data Compendia



Large-scale physics experiments

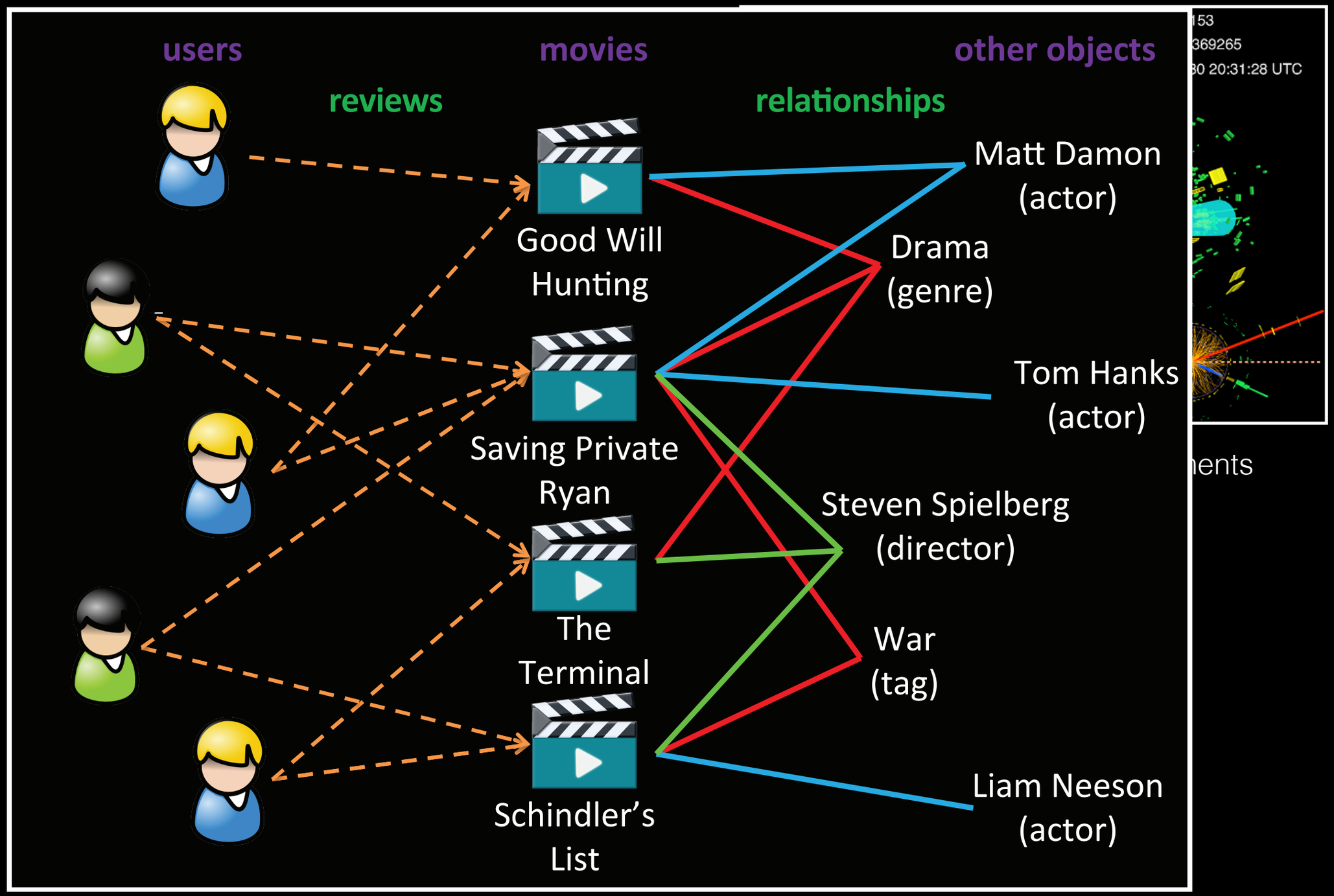
Marinka Zitnik - PhD Thesis

Large Heterogeneous Data Compendia



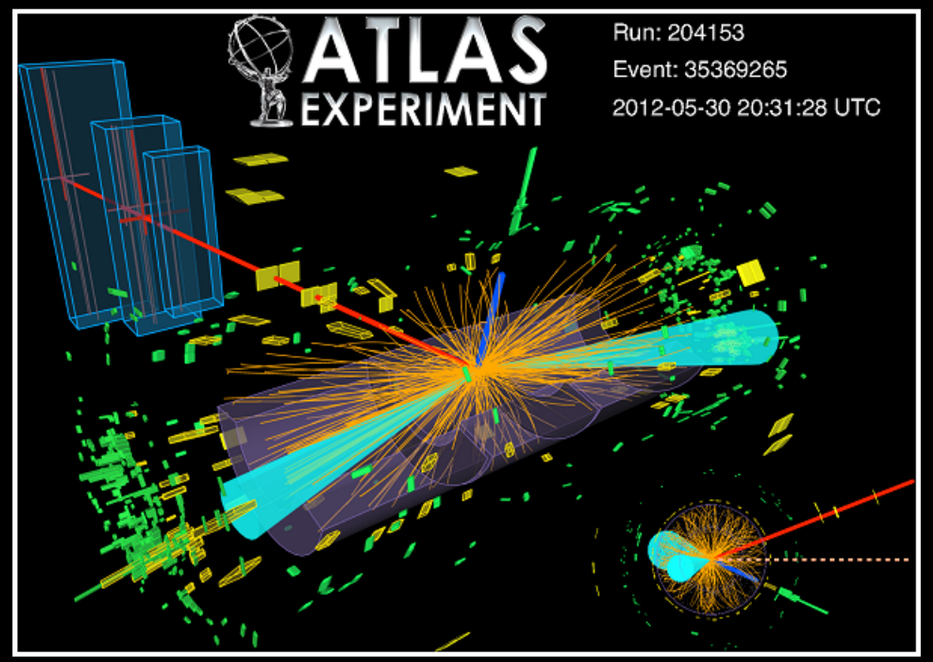
Large-scale physics experiments

Large Heterogeneous Data Compendia

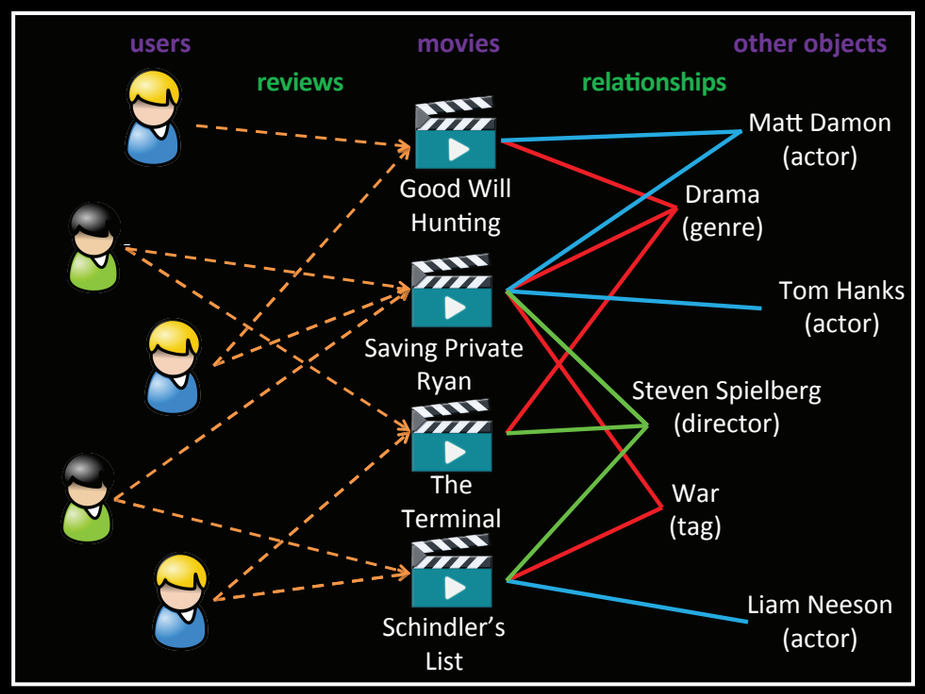


Social networks, recommender systems

Large Heterogeneous Data Compendia



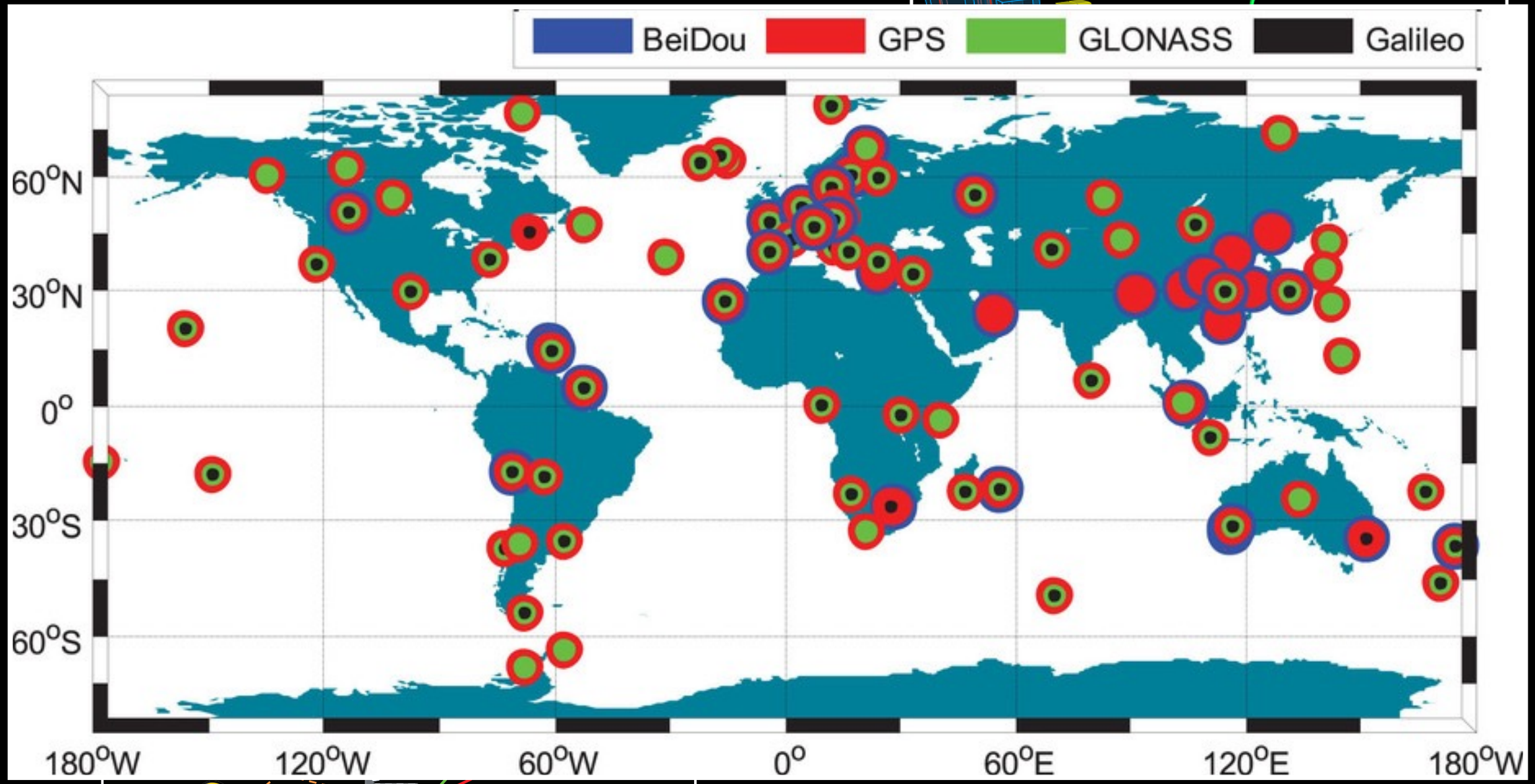
Large-scale physics experiments



Social networks, recommender systems

Marinka Zitnik - PhD Thesis

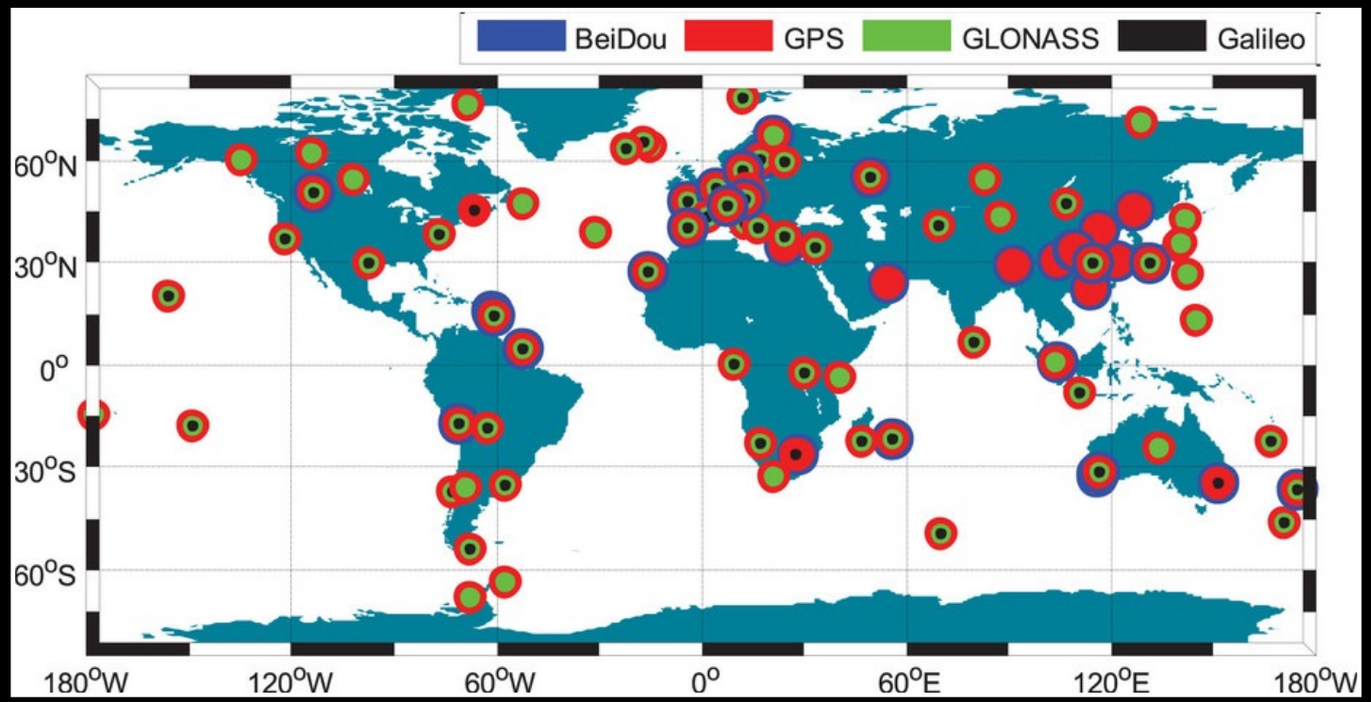
Large Heterogeneous Data Compendia



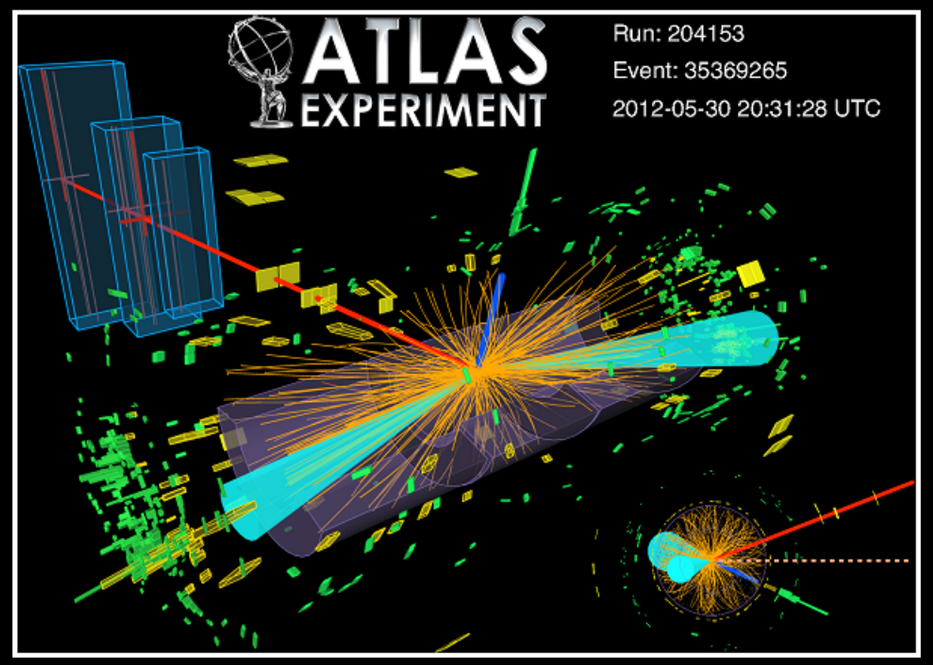
Global navigation satellite systems

Social networks, recommender systems

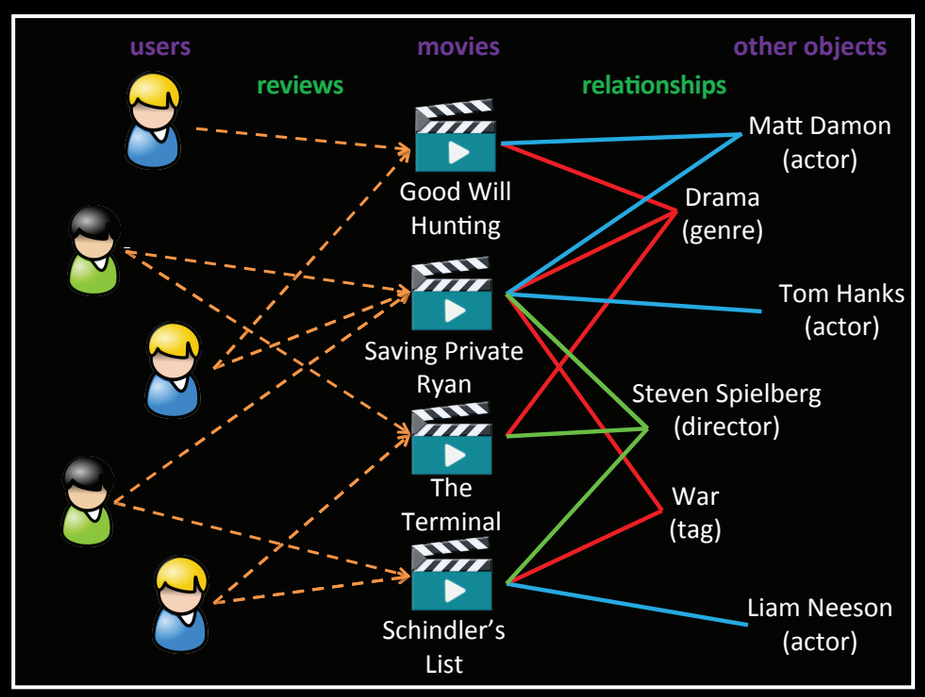
Large Heterogeneous Data Compendia



Global navigation satellite systems



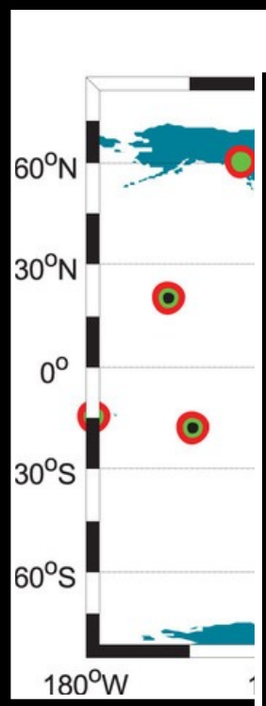
Large-scale physics experiments

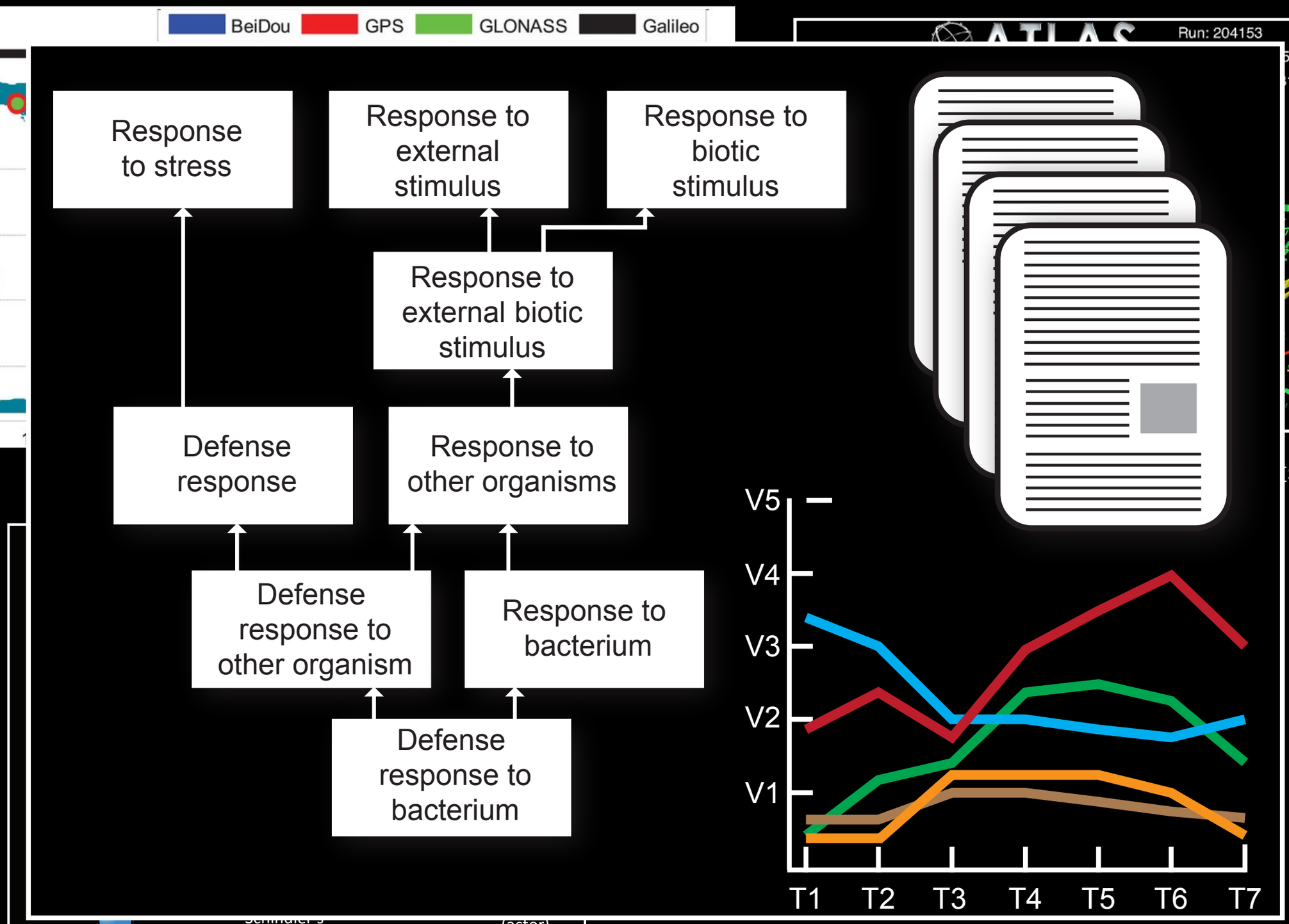


Social networks, recommender systems

Marinka Zitnik - PhD Thesis

Large Heterogeneous Data Compendia



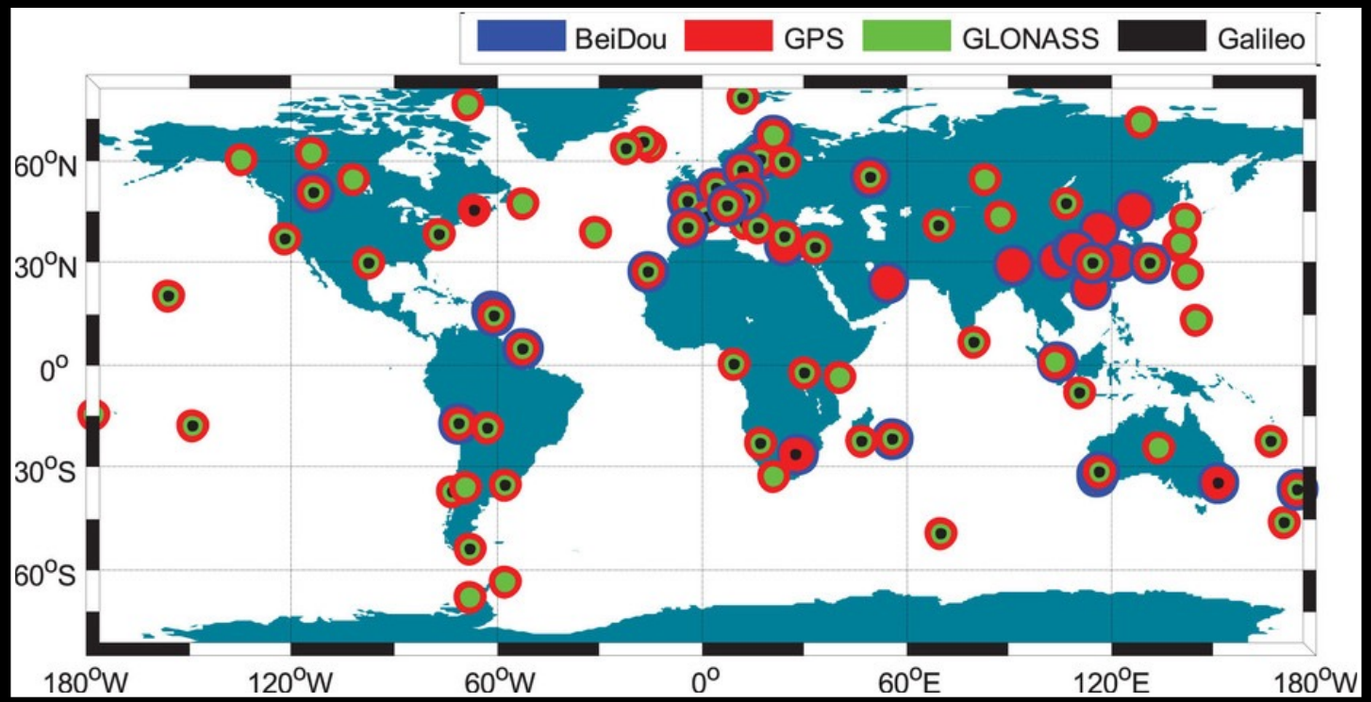


Social networks, recommender systems

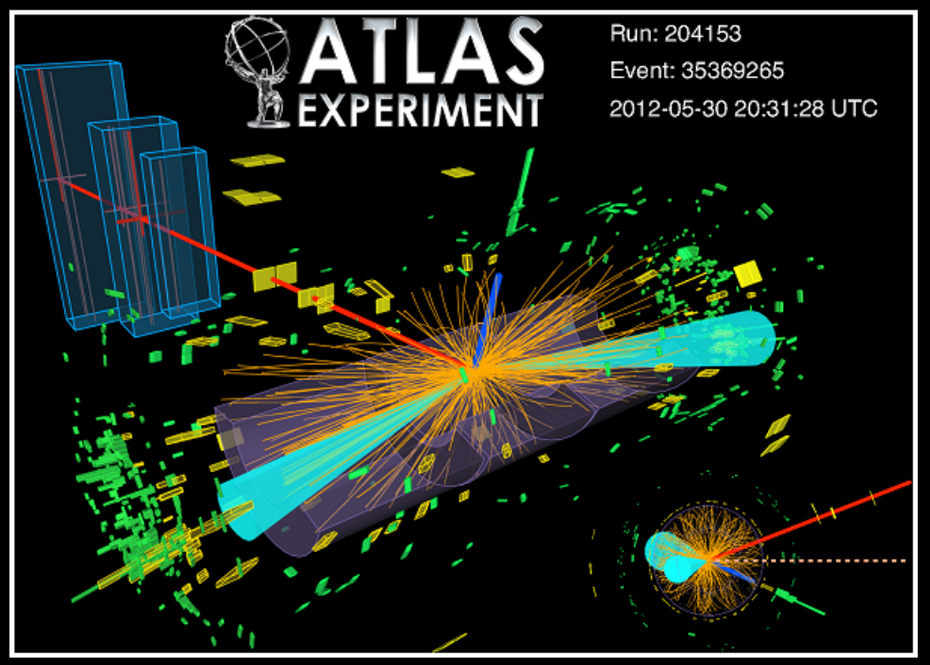
Molecular biology

Marinka Zitnik - PhD Thesis

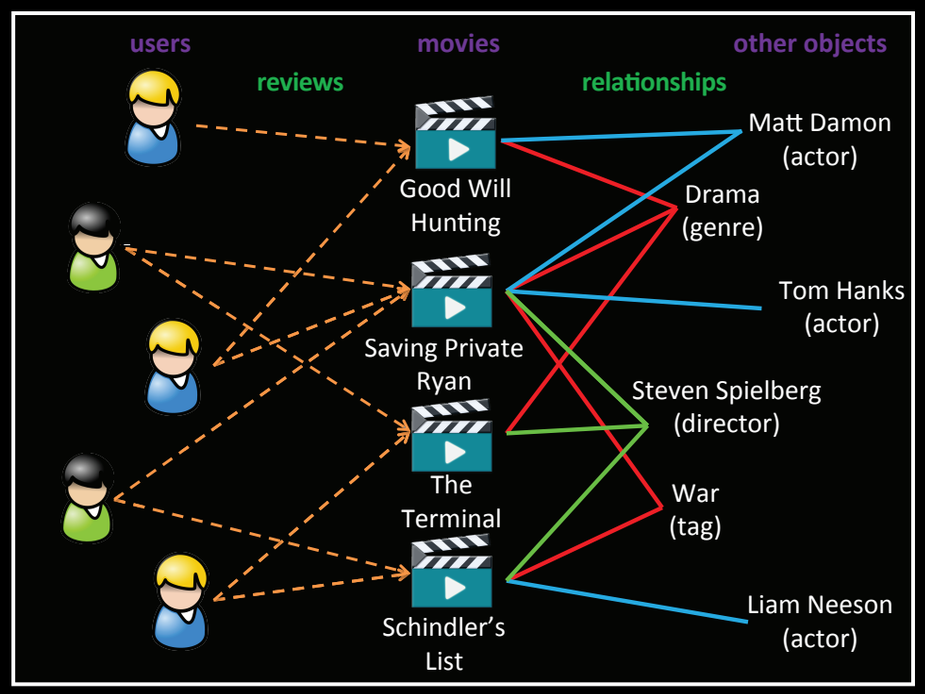
Large Heterogeneous Data Compendia



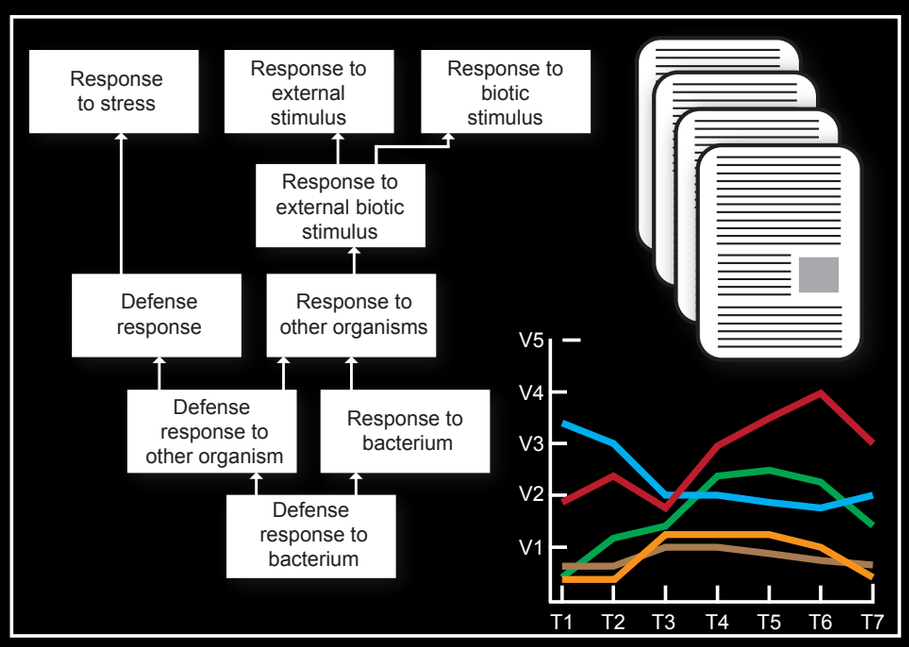
Global navigation satellite systems



Large-scale physics experiments



Social networks, recommender systems



Molecular biology

Complex relationships

Complex relationships

Objects of different types

Complex relationships

Objects of different types

Different points in time, space and scale

Complex relationships

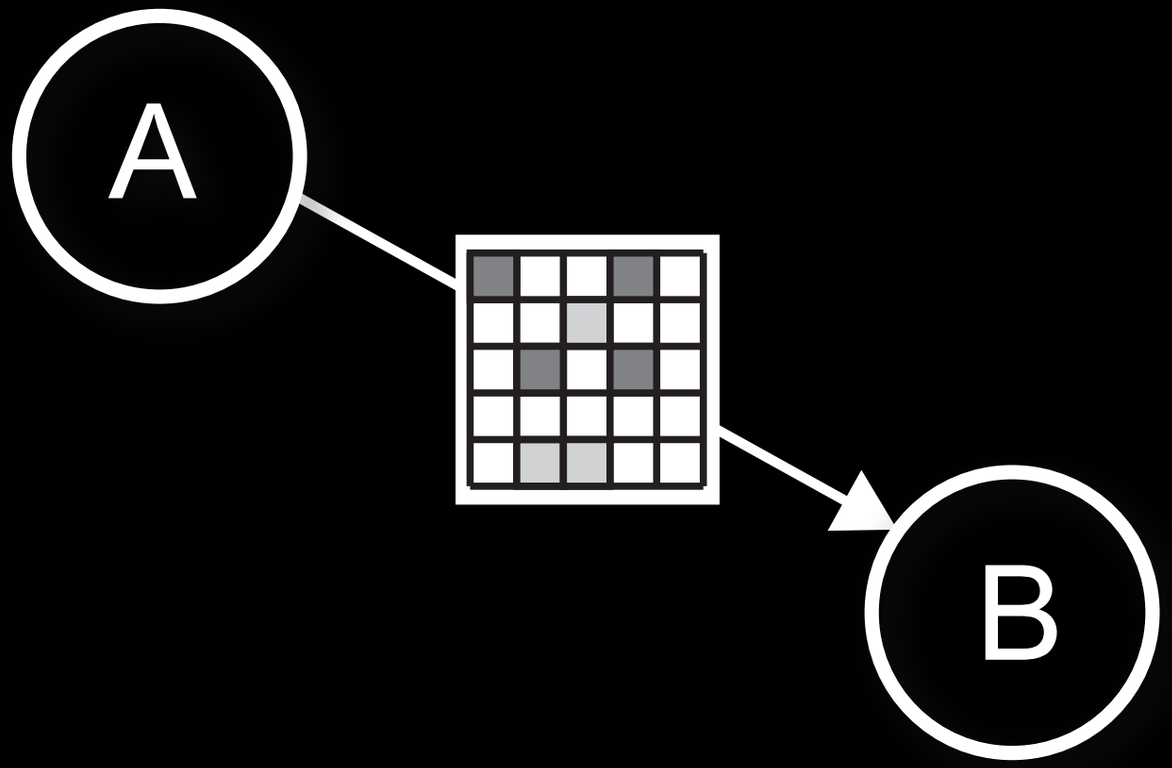
Objects of different types

Different points in time, space and scale

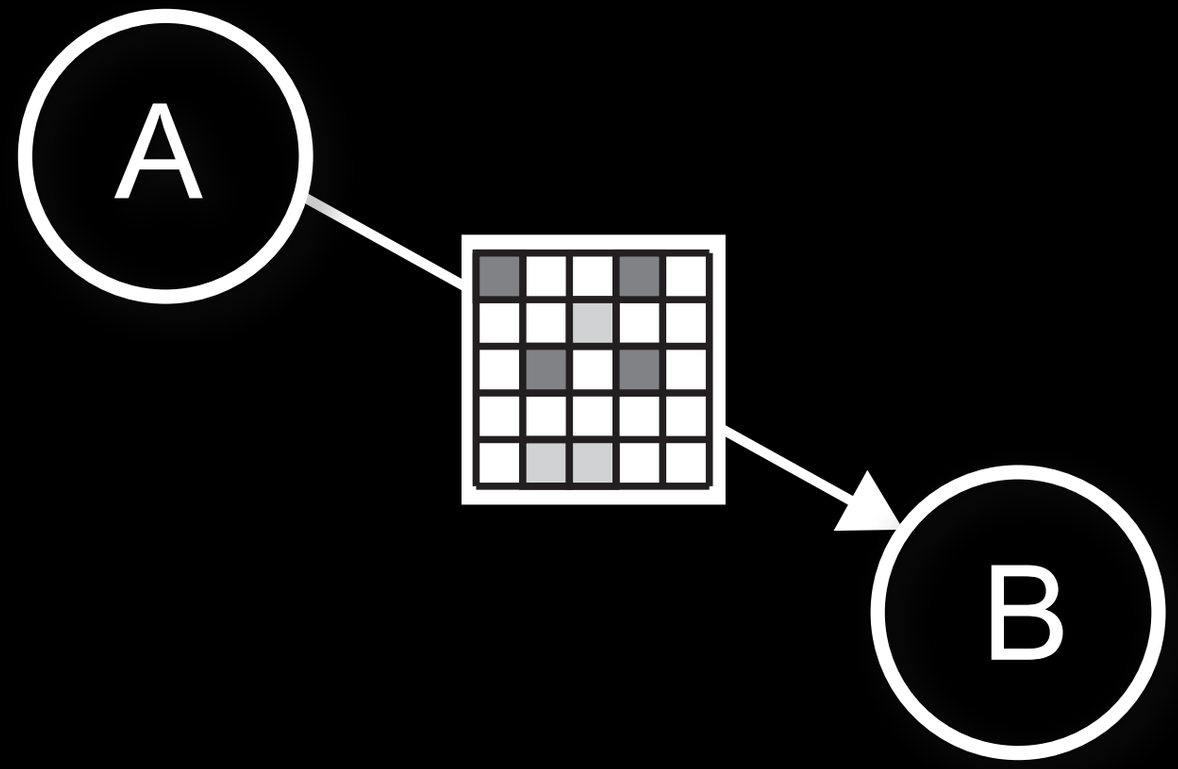
Different perspectives

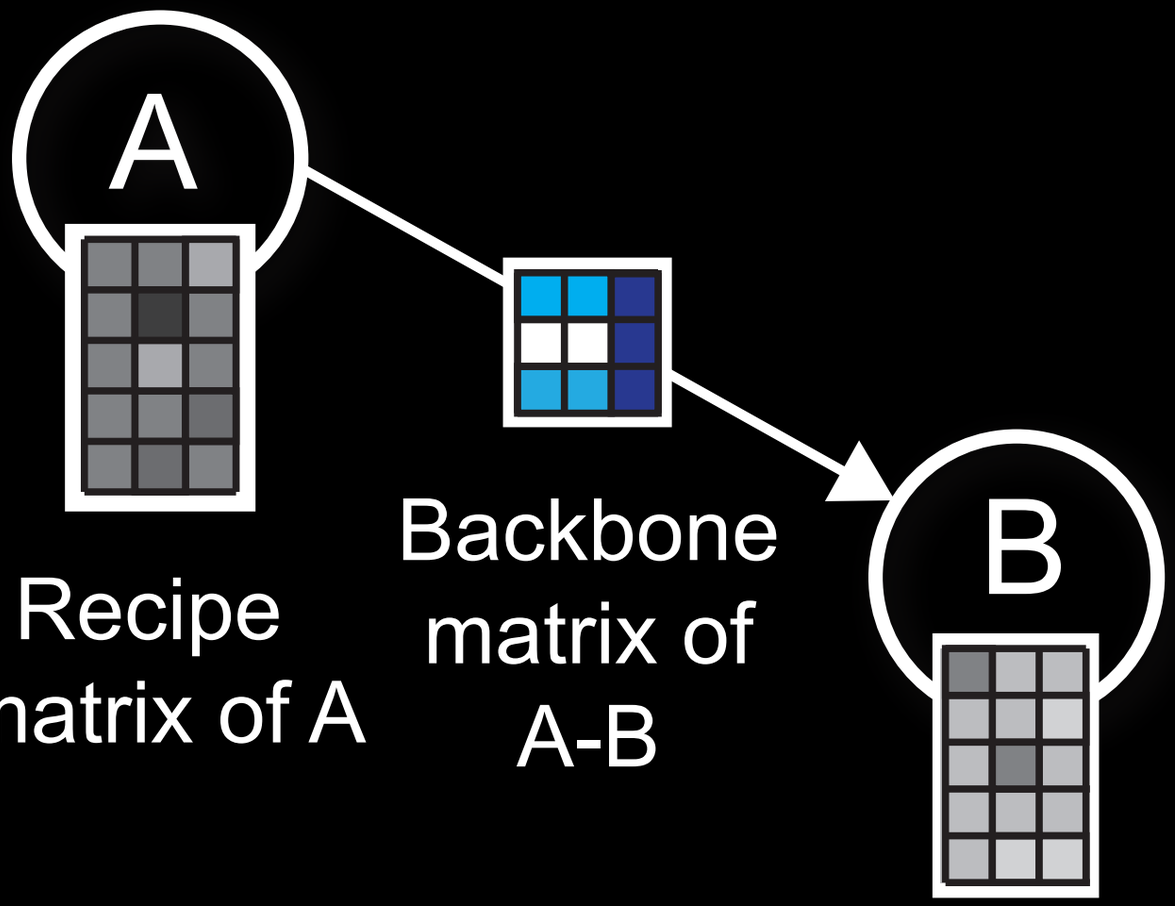
Warming-Up

One Data Matrix

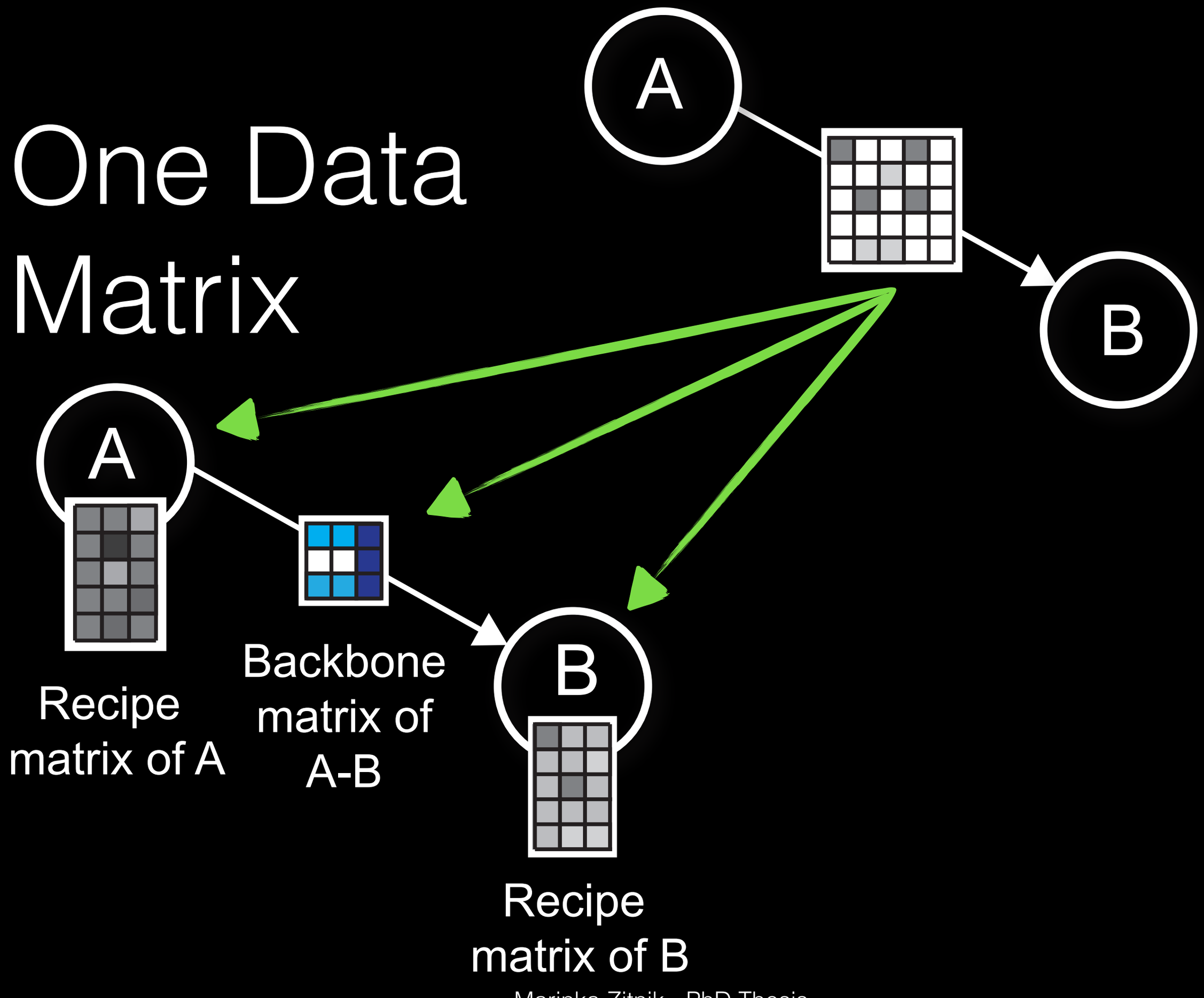


One Data Matrix

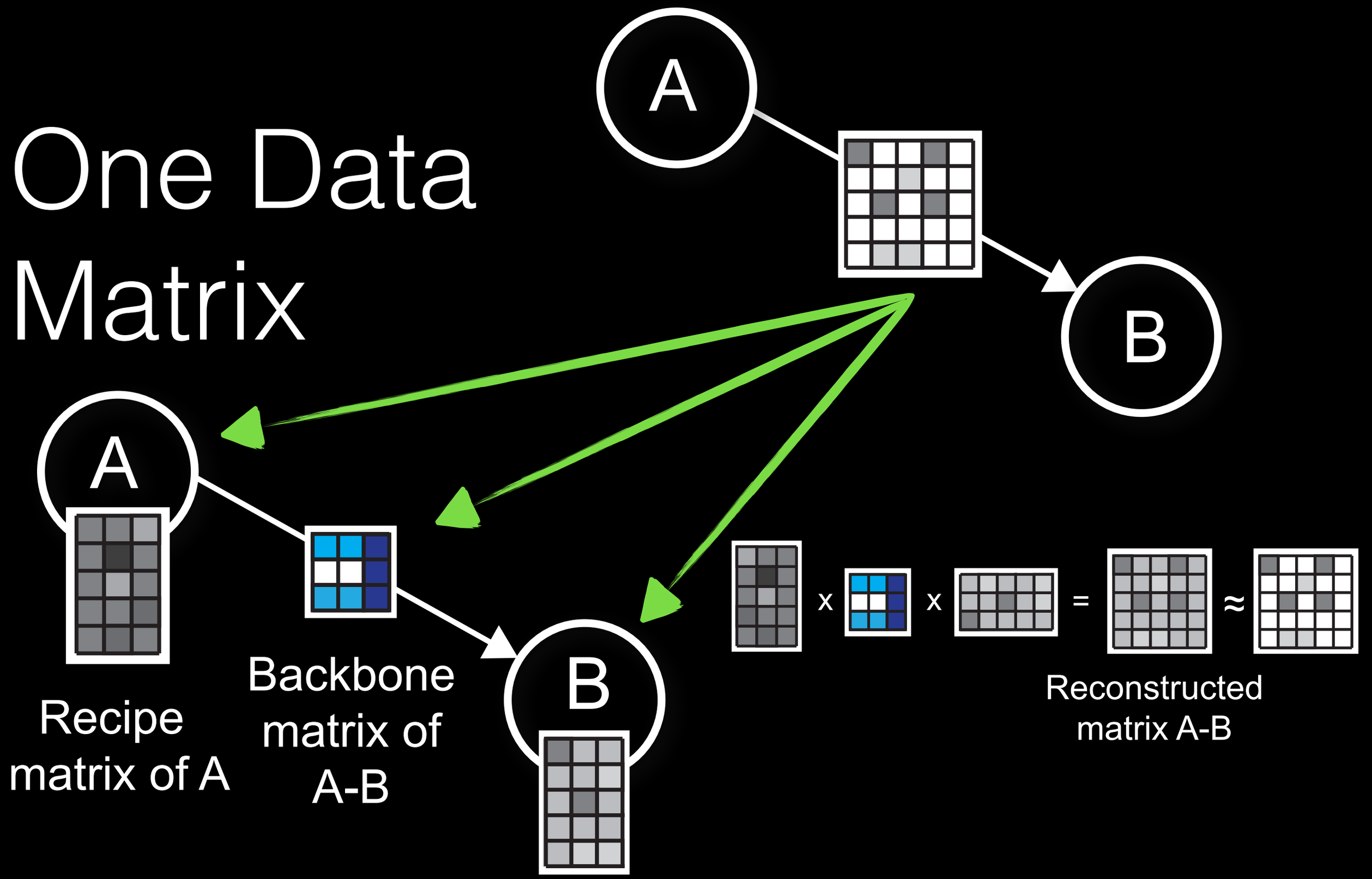




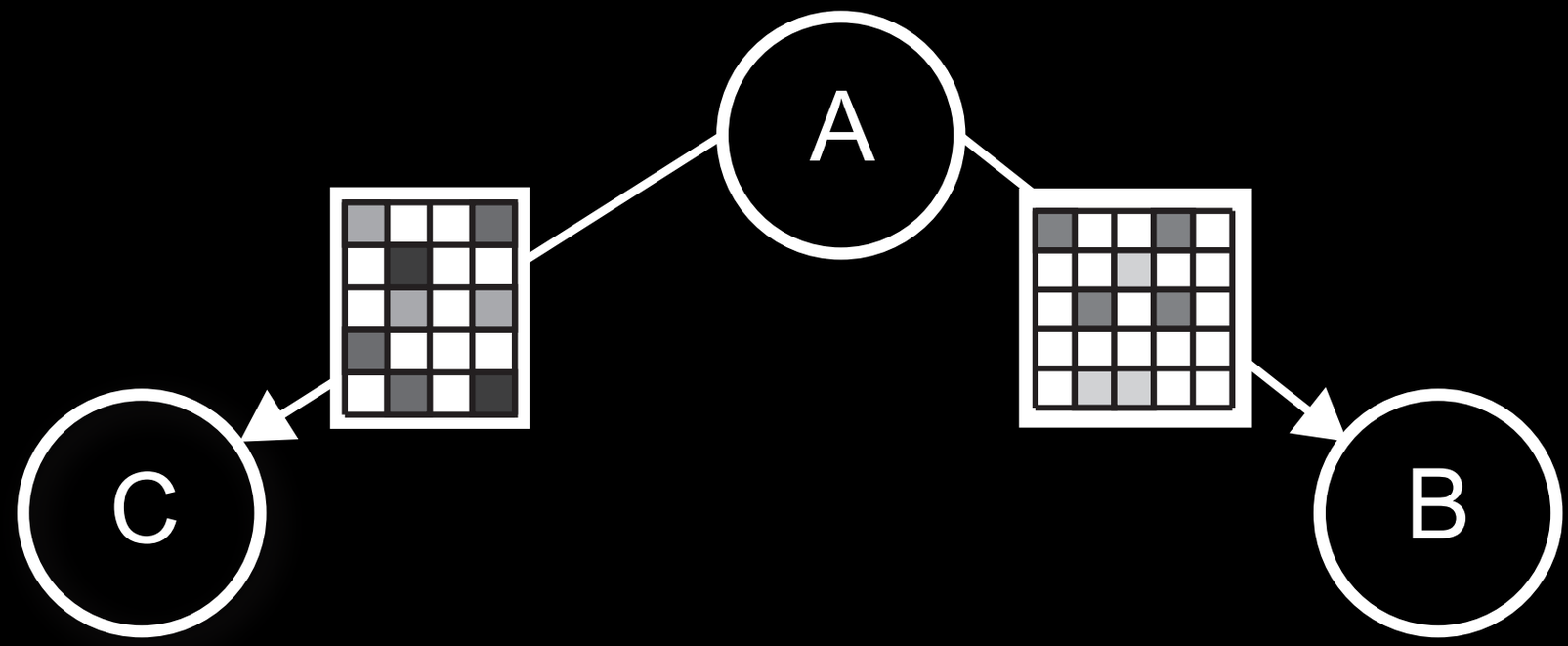
One Data Matrix



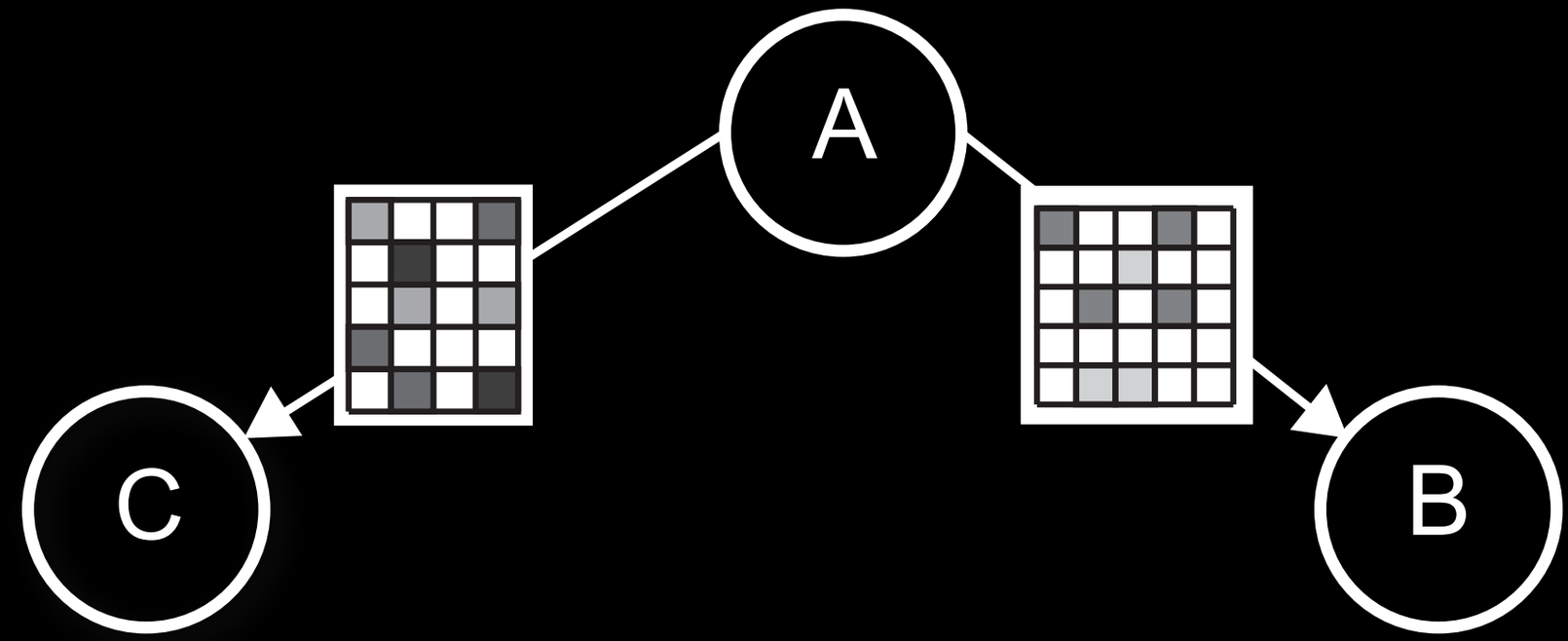
One Data Matrix

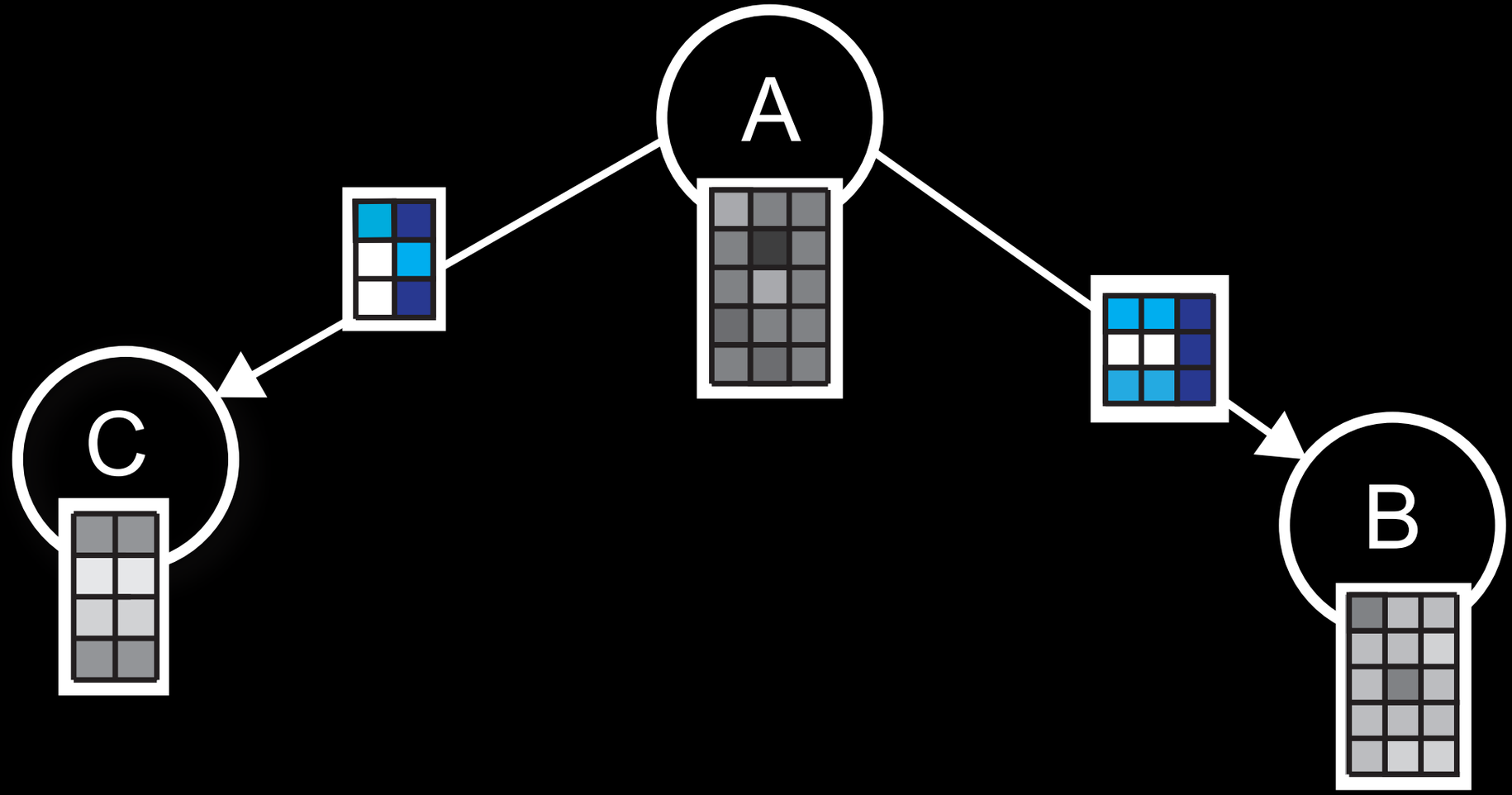


Two Data Matrices

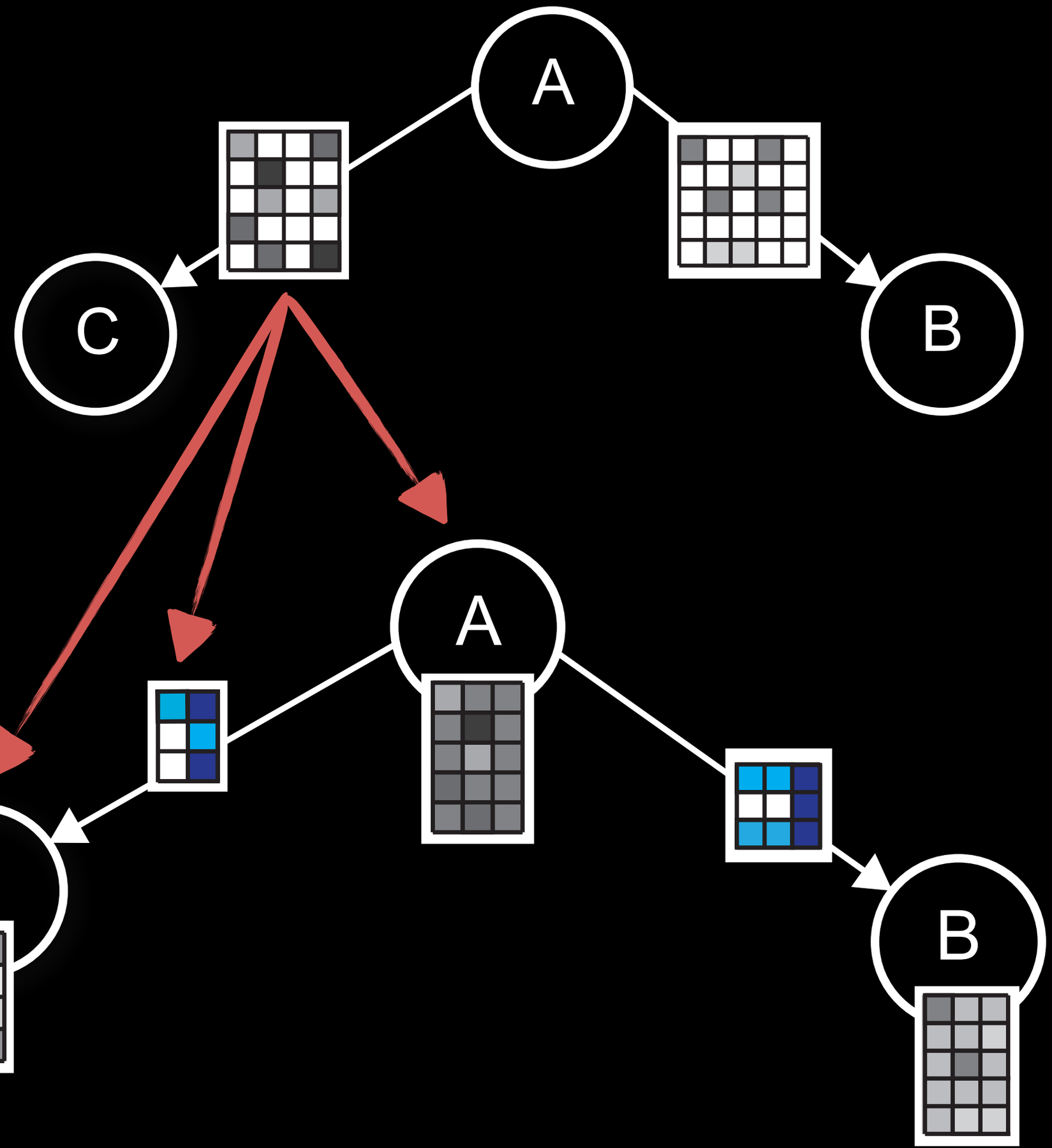


Two Data Matrices

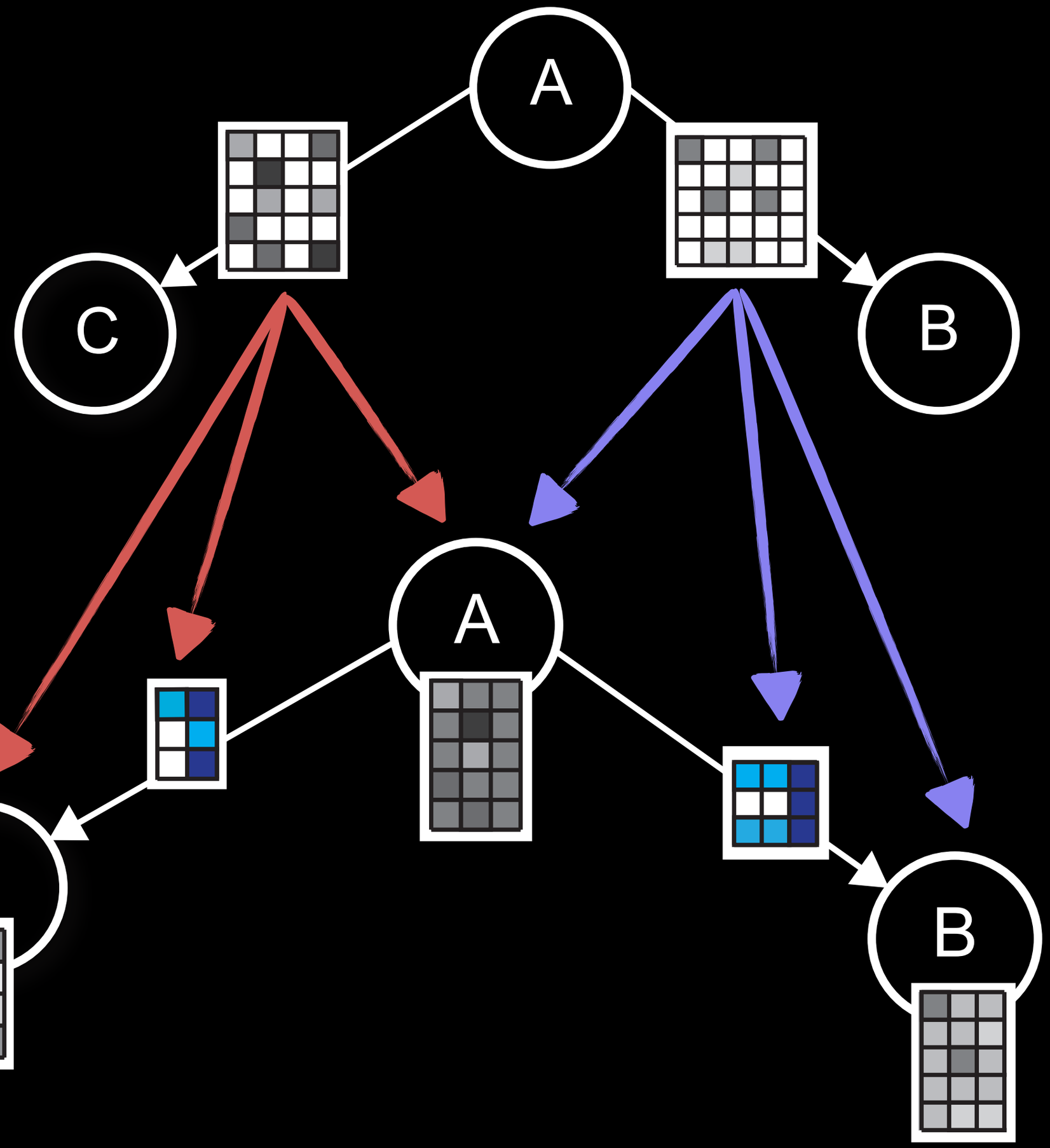




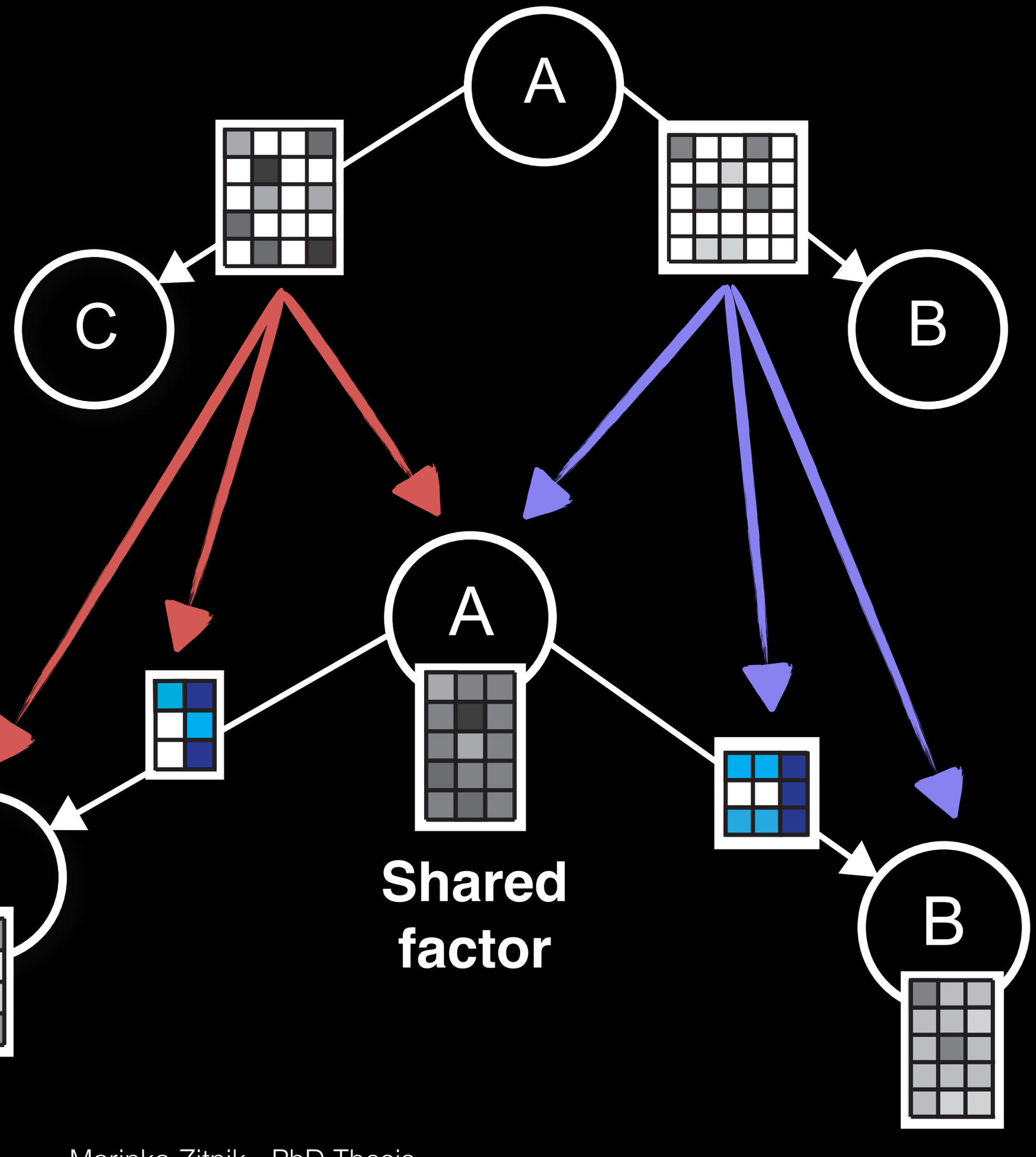
Two Data Matrices



Two Data Matrices

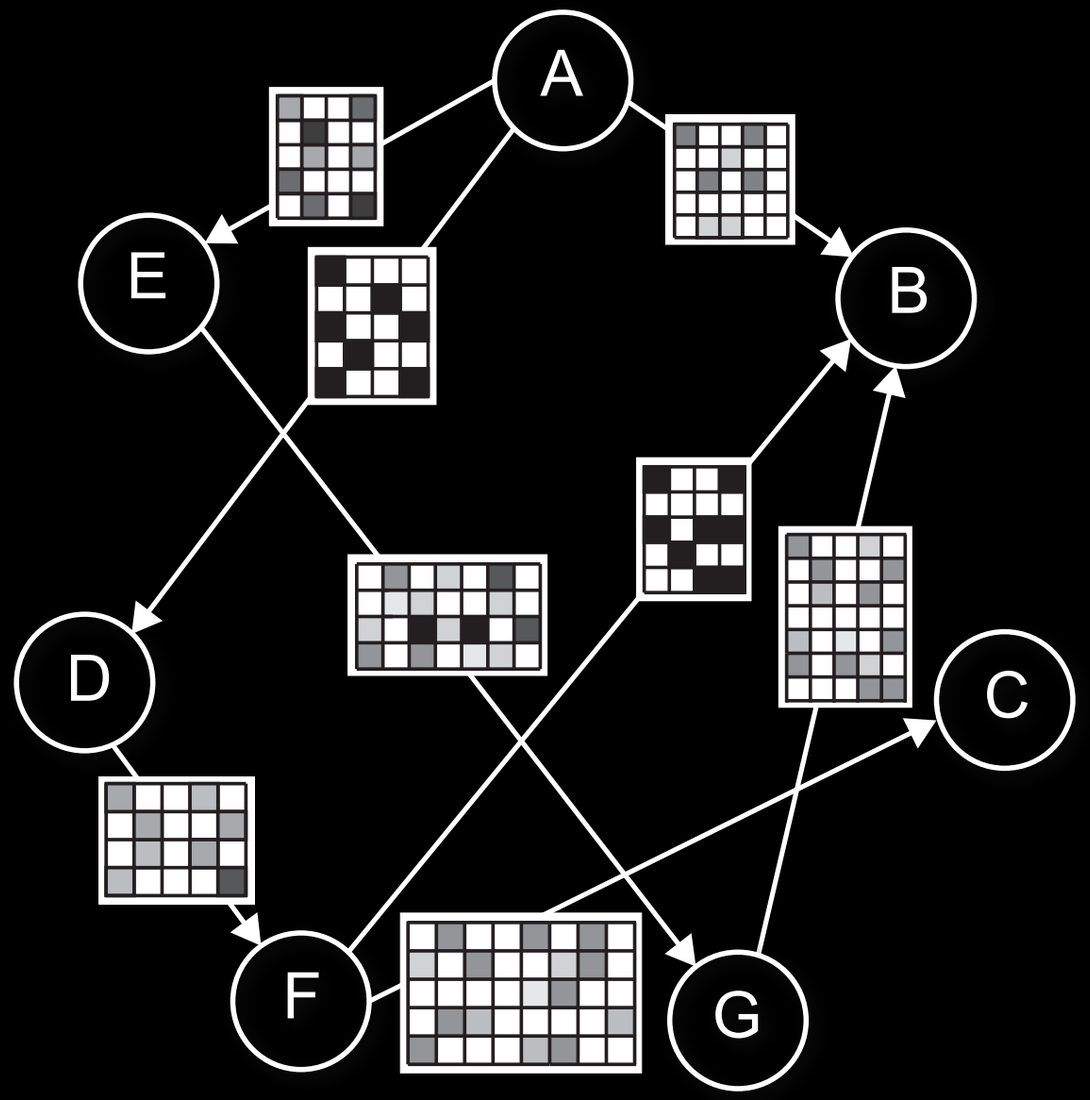


Two Data Matrices

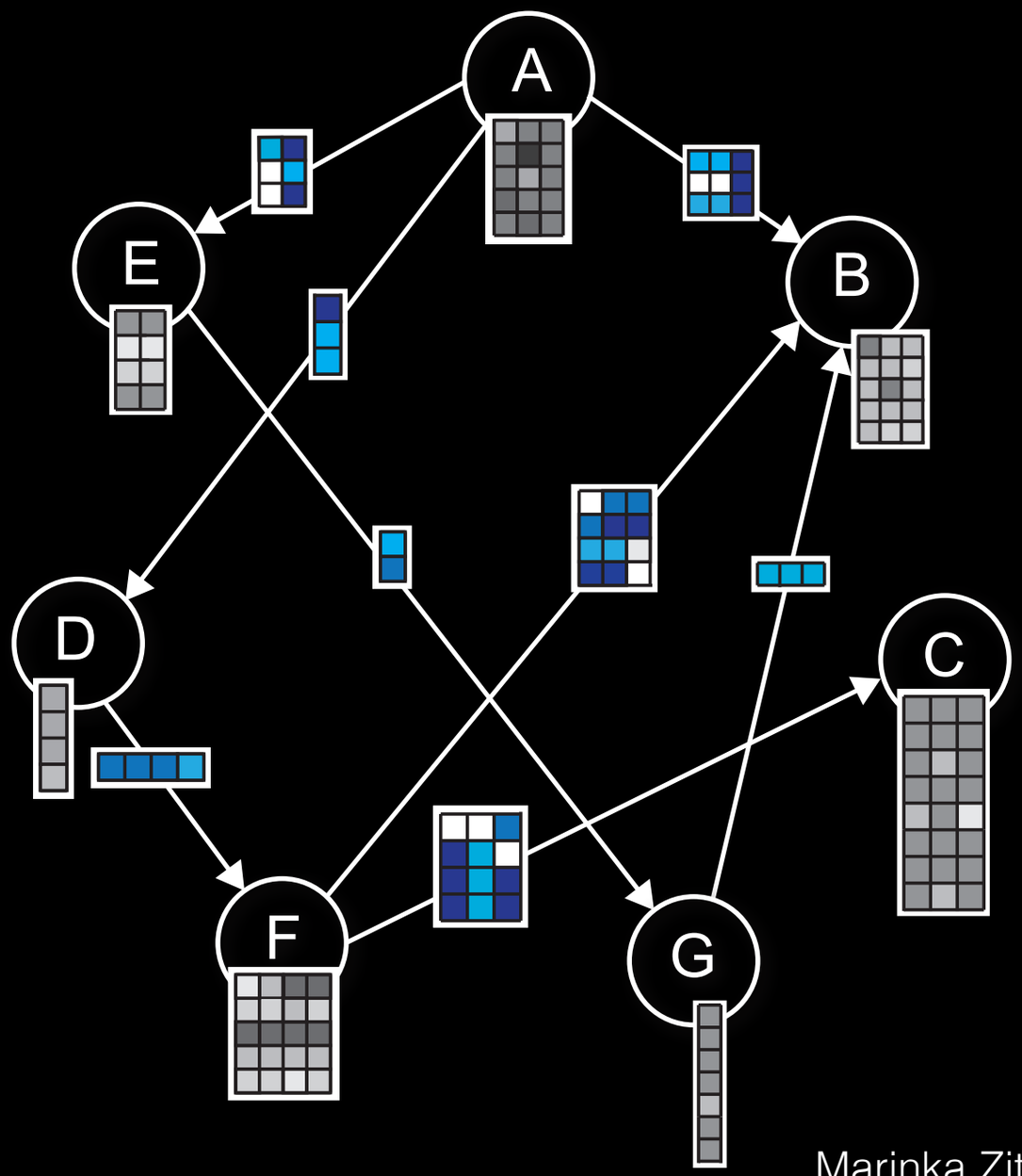


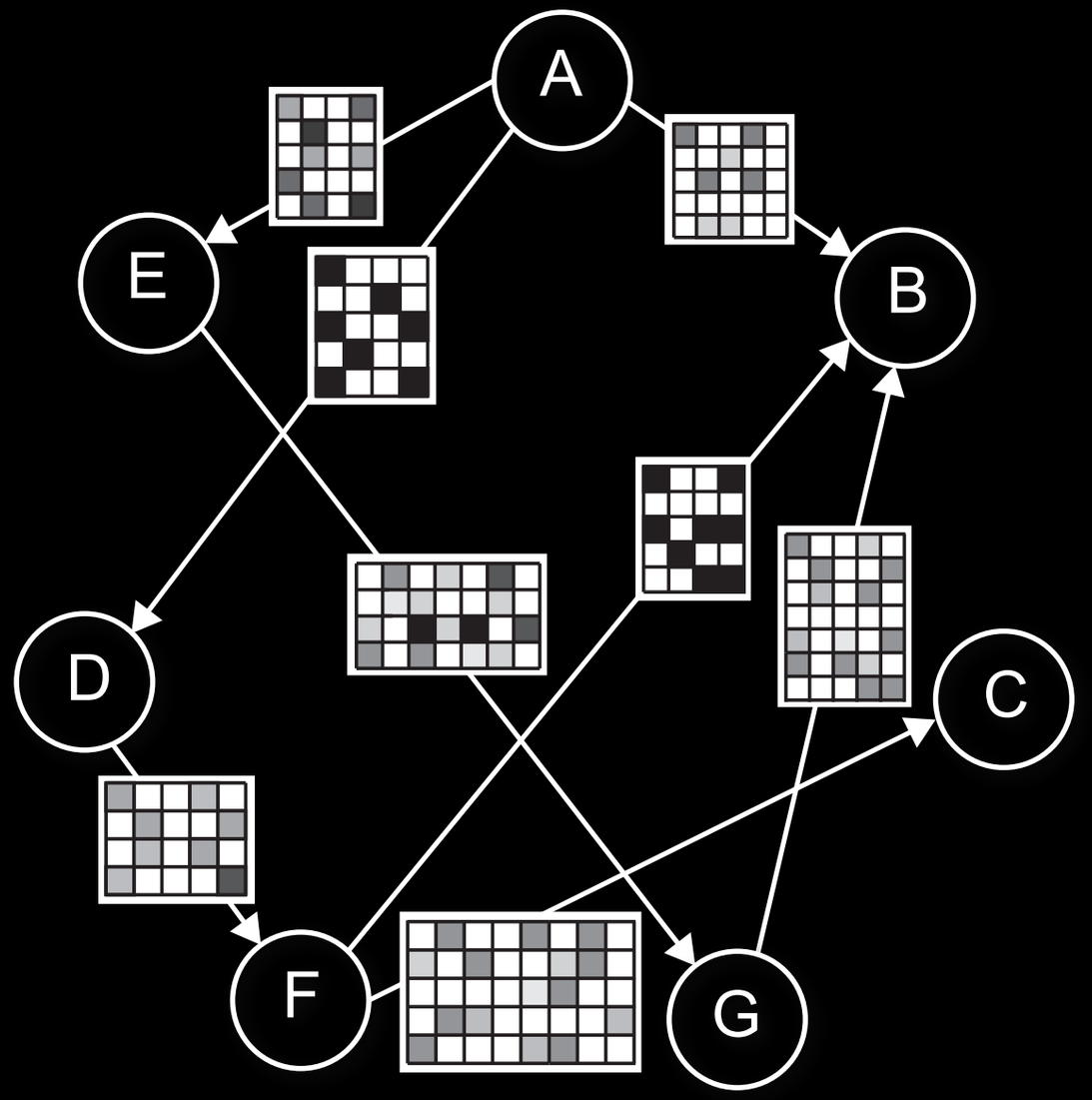
Data Fusion by Collective Matrix Factorization

Many Data Matrices

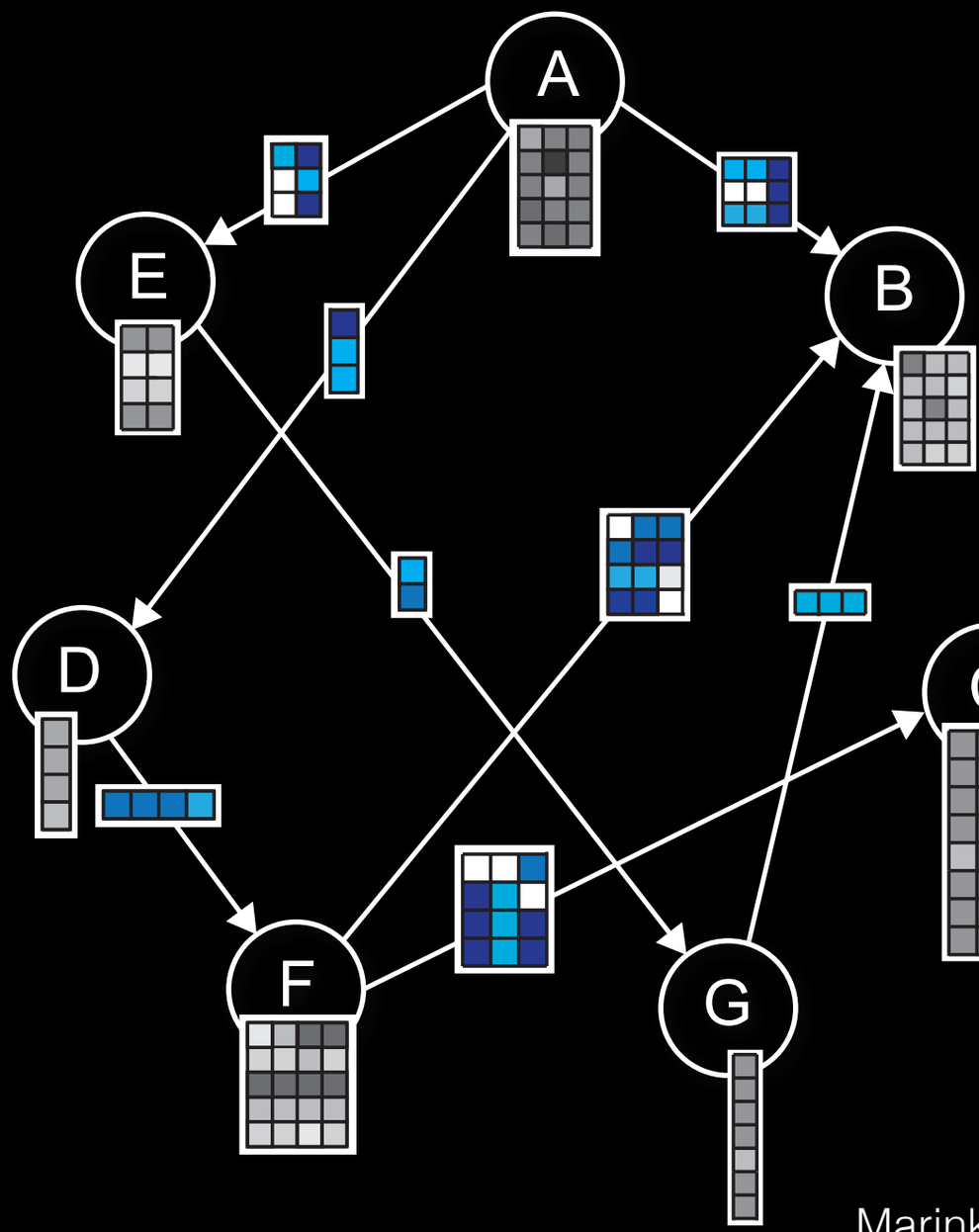


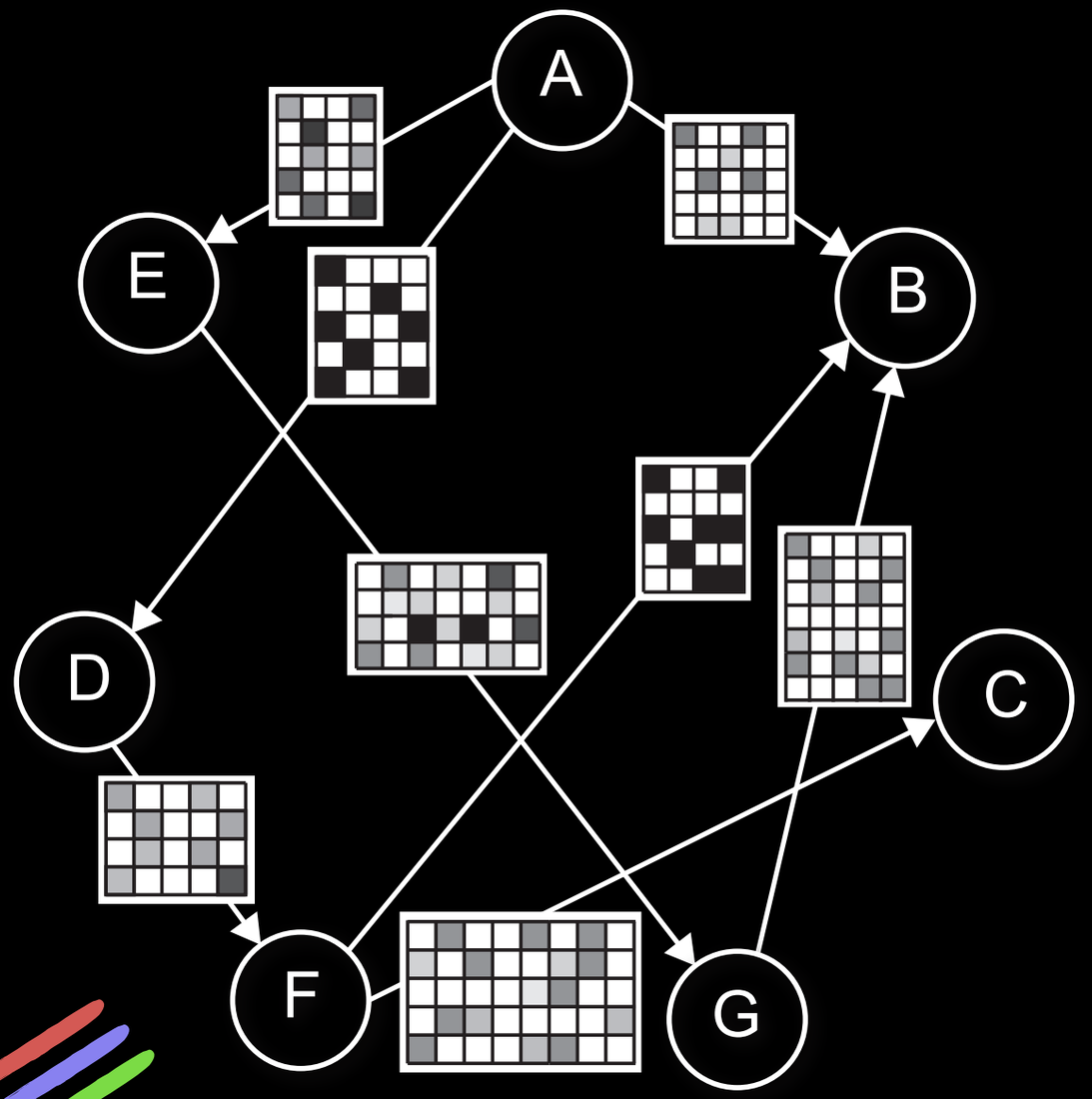
Many Data Matrices





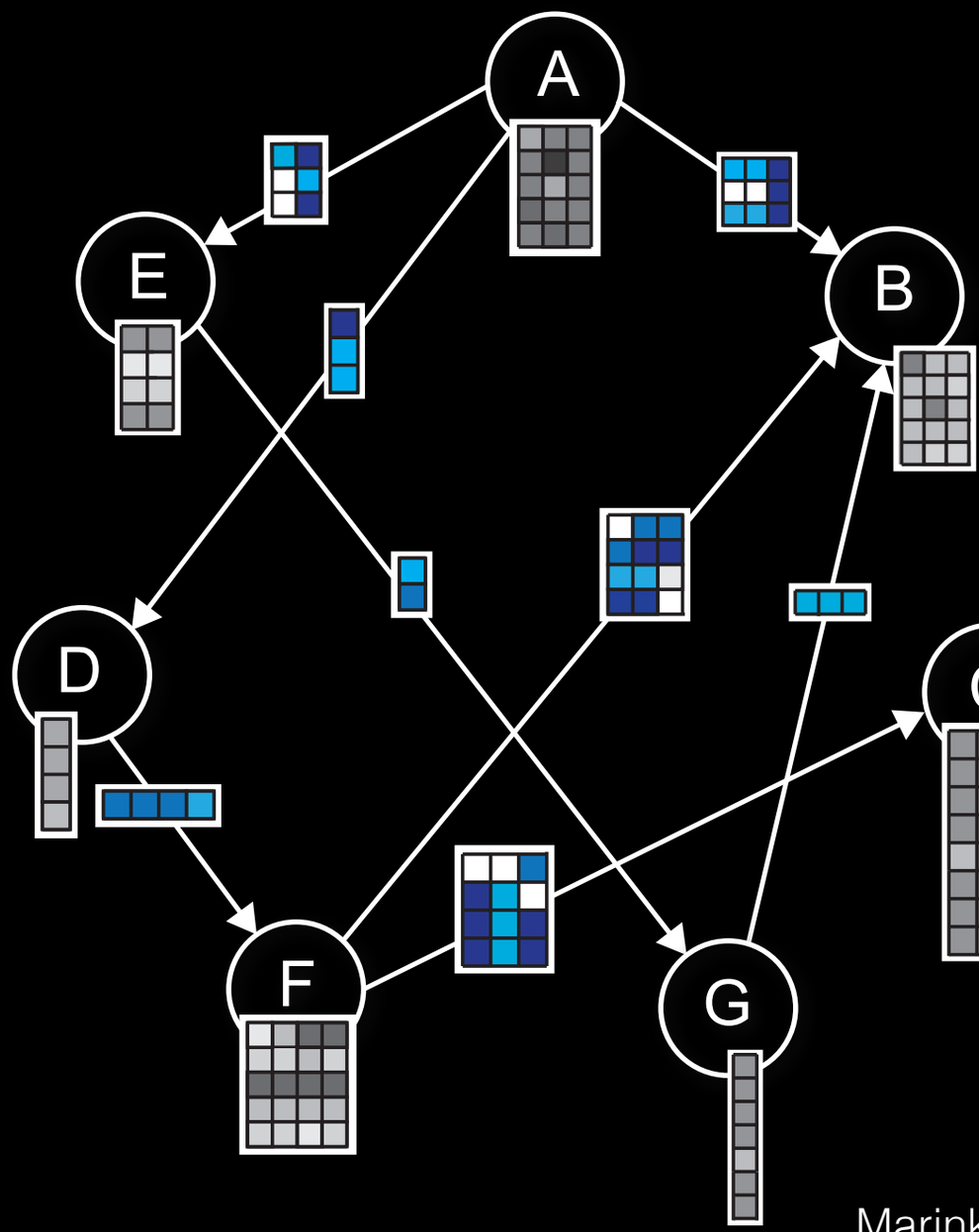
Many Data Matrices

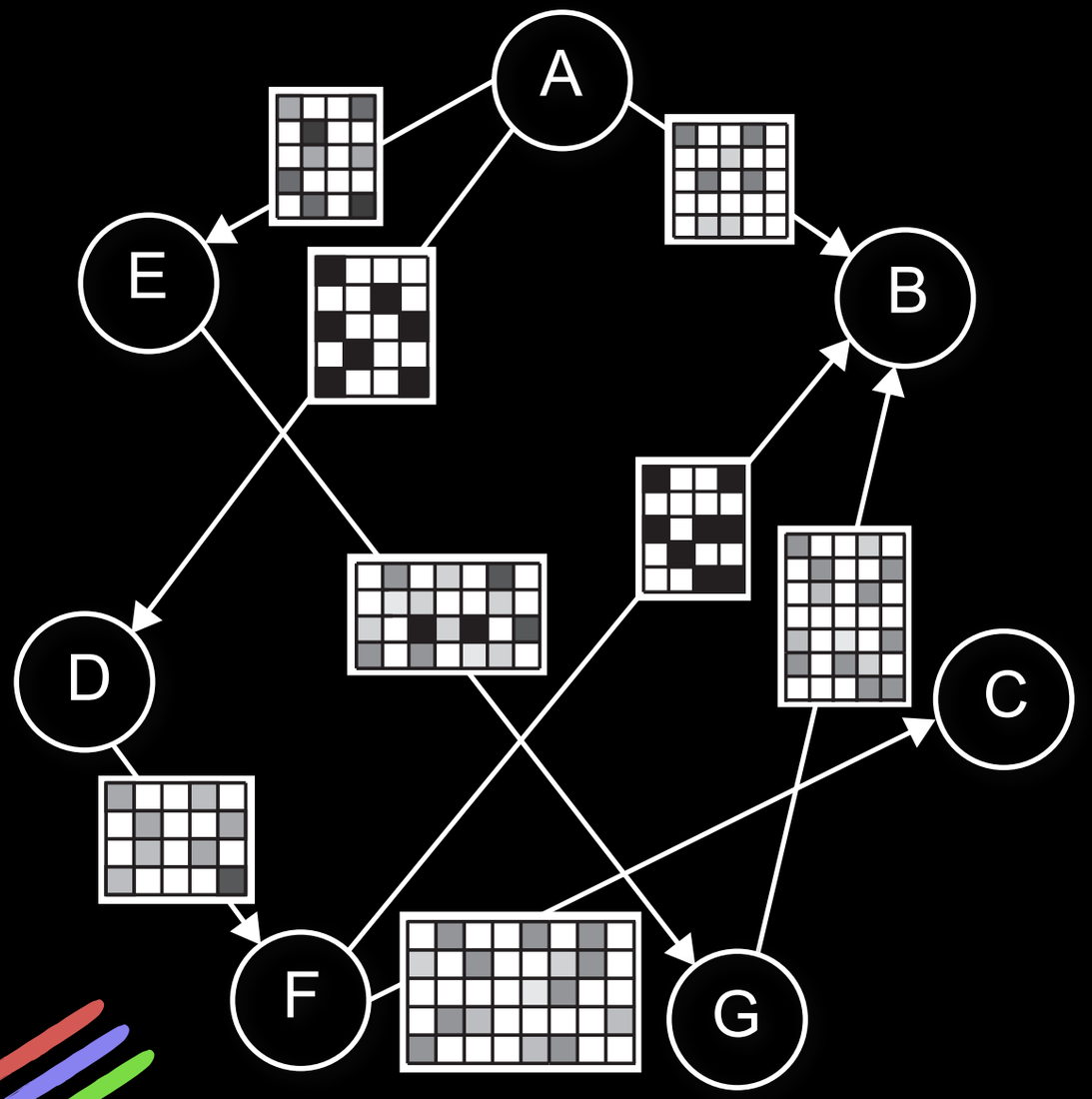




Many shared factors

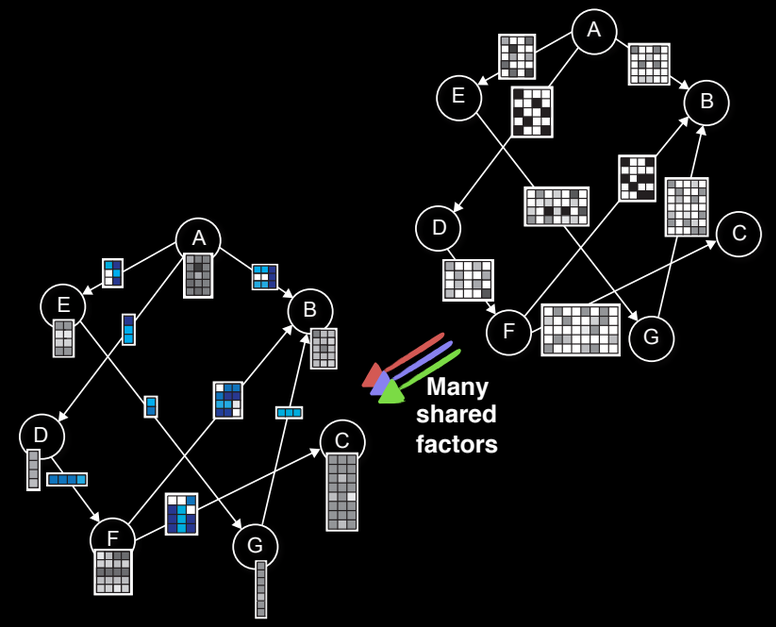
Many Data Matrices





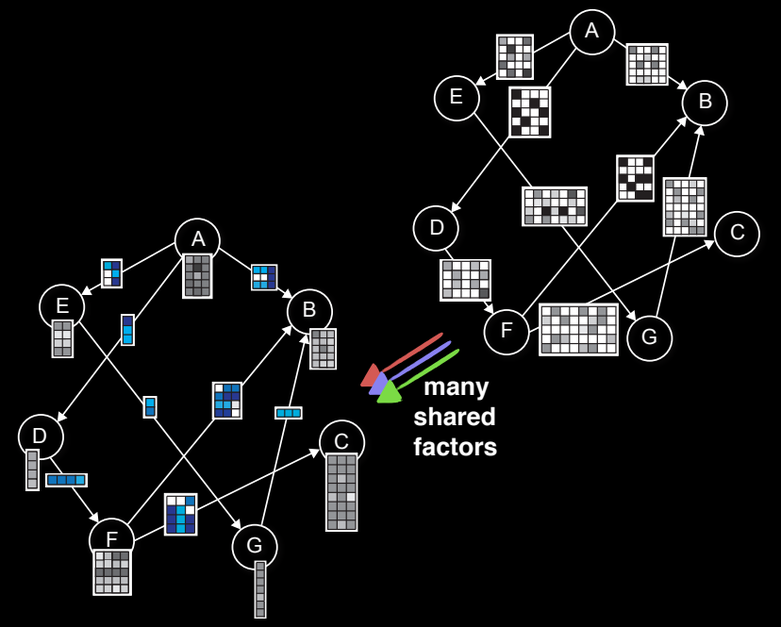
Many shared factors

Many Data Matrices



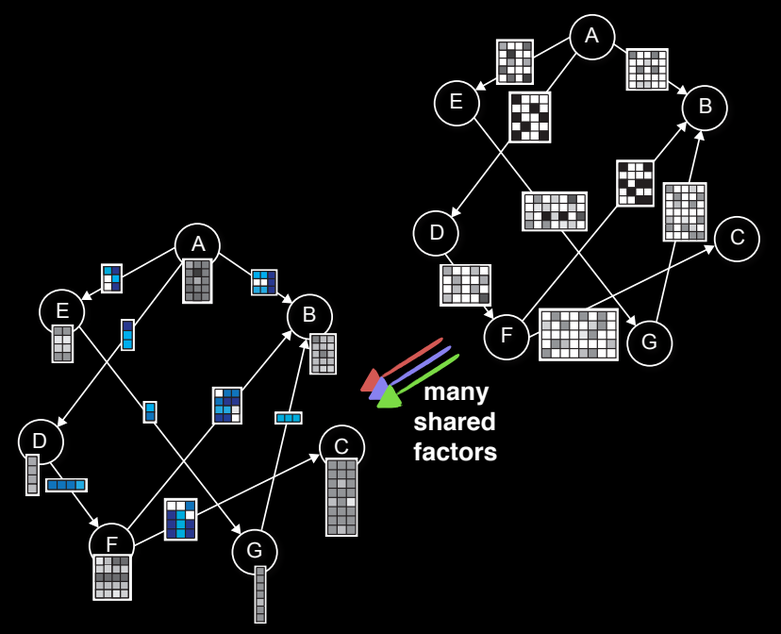
Many Data Matrices

Optimization Problem



Many Data Matrices

Optimization Problem



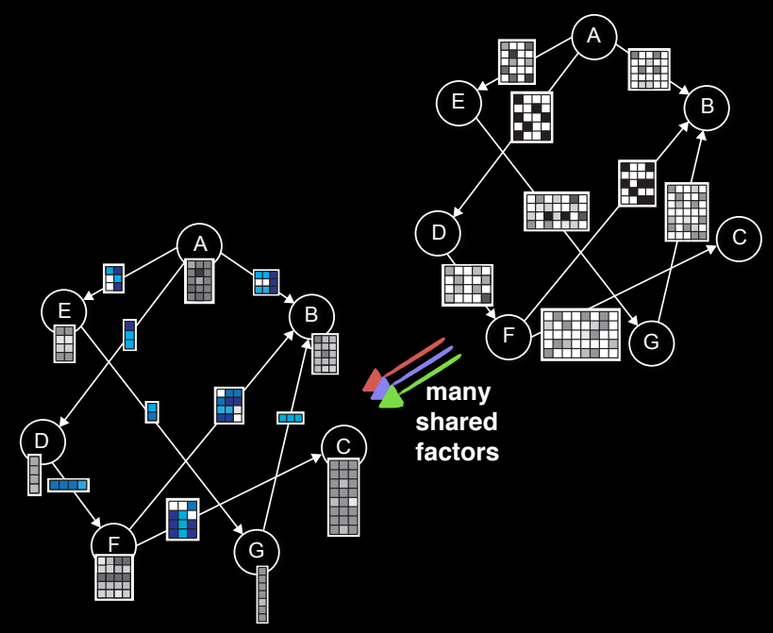
Given

$$\mathcal{R} = \{\mathbf{R}_{ij};\ i \text{ and } j \text{ are object types}\}$$

$$\mathcal{C} = \{\mathbf{\Theta}_i^l;\ l = 1, 2, \ldots, l_i,\ i \text{ is an object type}\}$$

Many Data Matrices

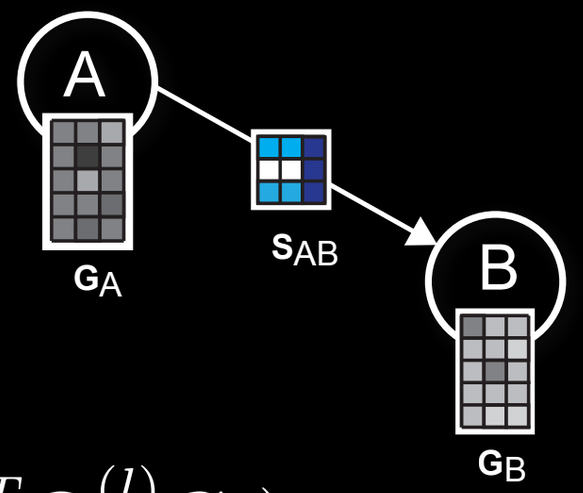
Optimization Problem



Given

$\mathcal{R} = \{\mathbf{R}_{ij};\ i \text{ and } j \text{ are object types}\}$

$\mathcal{C} = \{\mathbf{\Theta}_i^l;\ l = 1, 2, \ldots, l_i,\ i \text{ is an object type}\}$

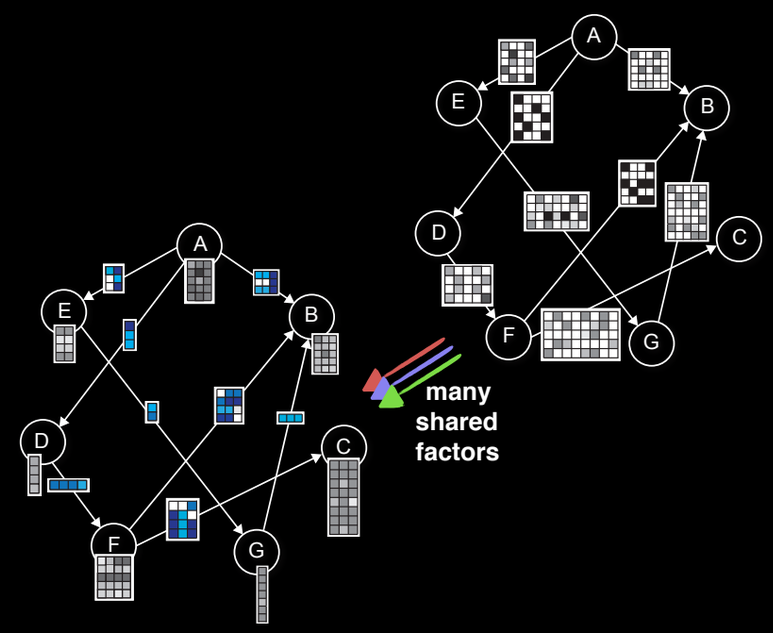


Find latent matrices $\mathbf{G}_i$ and $\mathbf{S}_{ij}$ that minimize

$$\min_{\mathbf{G}_i \geq 0, \mathbf{S}_{ij}} \sum_{\mathbf{R}_{ij} \in \mathcal{R}} \|\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\|_{\text{Fro}}^2 + \sum_{\mathbf{\Theta_i} \in \mathcal{C}} \sum_{l=1}^{l_i} \text{tr}(\mathbf{G}_i^T \mathbf{\Theta}_i^{(l)} \mathbf{G}_i)$$

Many Data Matrices

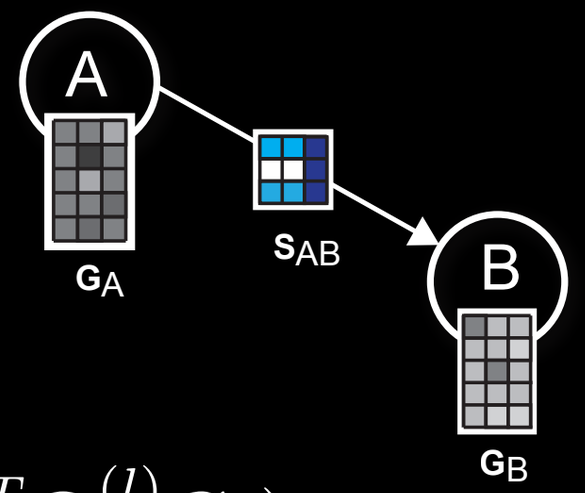
Optimization Problem



Given

$\mathcal{R} = \{\mathbf{R}_{ij};\ i \text{ and } j \text{ are object types}\}$

$\mathcal{C} = \{\mathbf{\Theta}_i^l;\ l = 1, 2, \ldots, l_i,\ i \text{ is an object type}\}$



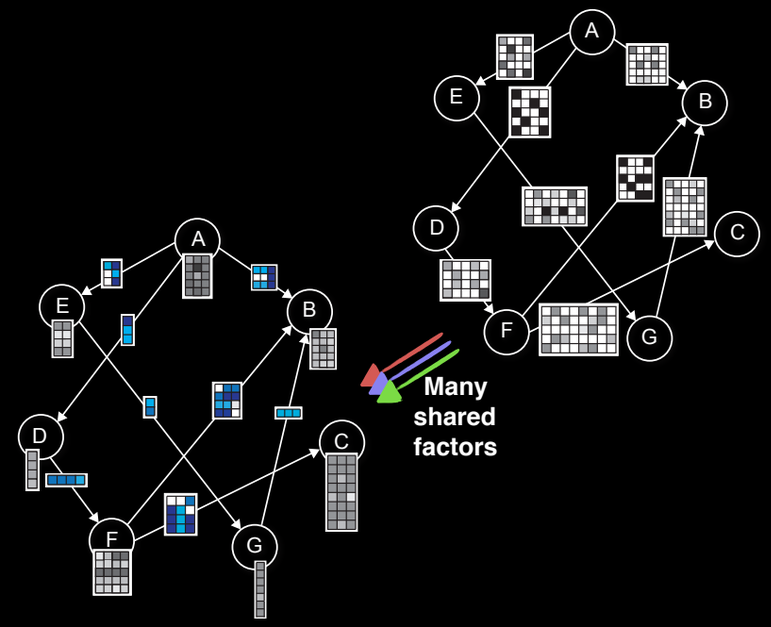
Find latent matrices $\mathbf{G}_i$ and $\mathbf{S}_{ij}$ that minimize

$$\min_{\mathbf{G}_i \geq 0, \mathbf{S}_{ij}} \sum_{\mathbf{R}_{ij} \in \mathcal{R}} \|\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\|_{\text{Fro}}^2 + \sum_{\mathbf{\Theta_i} \in \mathcal{C}} \sum_{l=1}^{l_i} \text{tr}(\mathbf{G}_i^T \mathbf{\Theta}_i^{(l)} \mathbf{G}_i)$$

The problem is non-convex. The global optimum is unknown

Many Data Matrices

Solution: DFMF Algorithm



Many Matric

Solution: DF

Input: A set $\mathcal{R}$ of relation matrices $\mathbf{R}_{ij}$; constraint matrices $\mathbf{\Theta}^{(t)}$ for $t \in \{1, 2, \ldots, \max_i t_i\}$; ranks $k_1, k_2, \ldots, k_r$ $(i, j \in [r])$.
Output: Matrix factors $\mathbf{S}$ and $\mathbf{G}$.

1) Initialize $\mathbf{G}_i$ for $i = 1, 2, \ldots, r$.
2) Repeat until convergence:

- Construct $\mathbf{R}$ and $\mathbf{G}$ using their definitions in Eq. (1) and Eq. (3).
- Update $\mathbf{S}$ using:

$$\mathbf{S} \leftarrow (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}.$$

- Set $\mathbf{G}_i^{(e)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
- Set $\mathbf{G}_i^{(d)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
- For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &\mathrel{+}= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^+ + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^- \\
\mathbf{G}_i^{(d)} &\mathrel{+}= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^- + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^+ \\
\mathbf{G}_j^{(e)} &\mathrel{+}= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^+ + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^- \\
\mathbf{G}_j^{(d)} &\mathrel{+}= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^- + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^+ \quad (10)
\end{aligned}$$

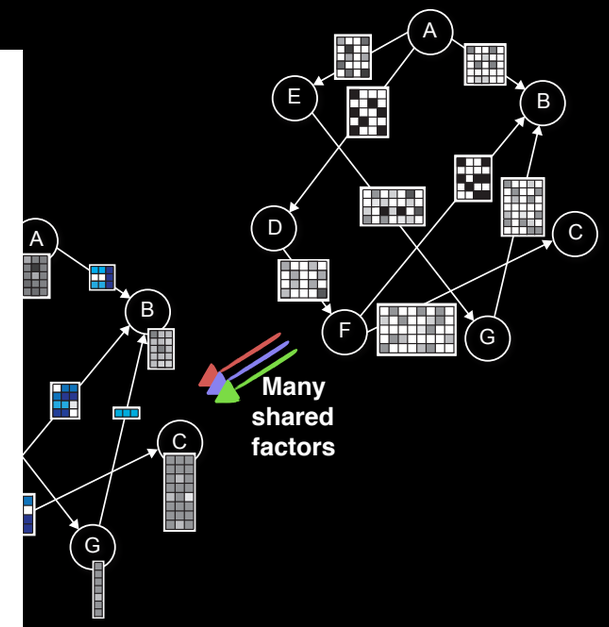
- For $t = 1, 2, \ldots, \max_i t_i$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &\mathrel{+}= [\mathbf{\Theta}_i^{(t)}]^-\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \\
\mathbf{G}_i^{(d)} &\mathrel{+}= [\mathbf{\Theta}_i^{(t)}]^+\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \quad (11)
\end{aligned}$$

- Construct $\mathbf{G}$ as:

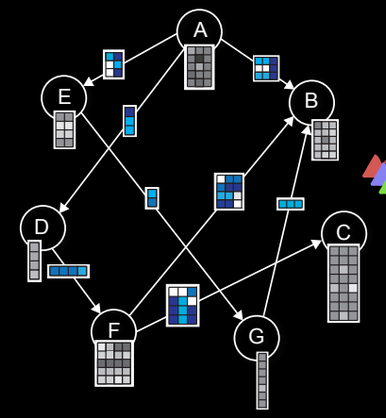
$$\mathbf{G} \leftarrow \mathbf{G} \circ \text{Diag}\left(\sqrt{\frac{\mathbf{G}_1^{(e)}}{\mathbf{G}_1^{(d)}}}, \sqrt{\frac{\mathbf{G}_2^{(e)}}{\mathbf{G}_2^{(d)}}}, \ldots, \sqrt{\frac{\mathbf{G}_r^{(e)}}{\mathbf{G}_r^{(d)}}}\right), \quad (12)$$

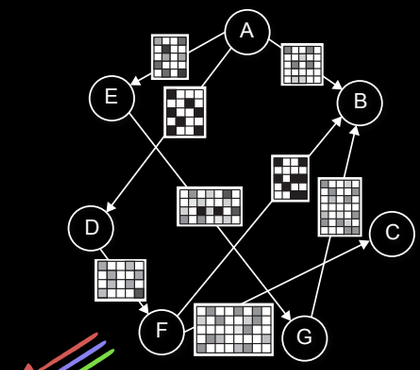
where $\circ$ denotes the Hadamard product. The $\sqrt{\cdot}$ and $\frac{\cdot}{\cdot}$ are entry-wise operations.



Many Data Matrices

Solution: DFMF Algorithm





Many shared factors

Input: A set $\mathcal{R}$ of relation matrices $\mathbf{R}_{ij}$; constraint matrices $\mathbf{\Theta}^{(t)}$ for $t \in \{1, 2, \ldots, \max_i t_i\}$; ranks $k_1, k_2, \ldots, k_r$ $(i, j \in [r])$.
Output: Matrix factors $\mathbf{S}$ and $\mathbf{G}$.

1) Initialize $\mathbf{G}_i$ for $i = 1, 2, \ldots, r$.
2) Repeat until convergence:
   - Construct $\mathbf{R}$ and $\mathbf{G}$ using their definitions in Eq. (1) and Eq. (3).
   - Update $\mathbf{S}$ using:

$$\mathbf{S} \leftarrow (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}.$$

   - Set $\mathbf{G}_i^{(e)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
   - Set $\mathbf{G}_i^{(d)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
   - For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^+ + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^- \\
\mathbf{G}_i^{(d)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^- + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^+ \\
\mathbf{G}_j^{(e)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^+ + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^- \\
\mathbf{G}_j^{(d)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^- + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^+ \quad (10)
\end{aligned}$$

   - For $t = 1, 2, \ldots, \max_i t_i$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= [\mathbf{\Theta}_i^{(t)}]^-\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \\
\mathbf{G}_i^{(d)} &+= [\mathbf{\Theta}_i^{(t)}]^+\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \quad (11)
\end{aligned}$$

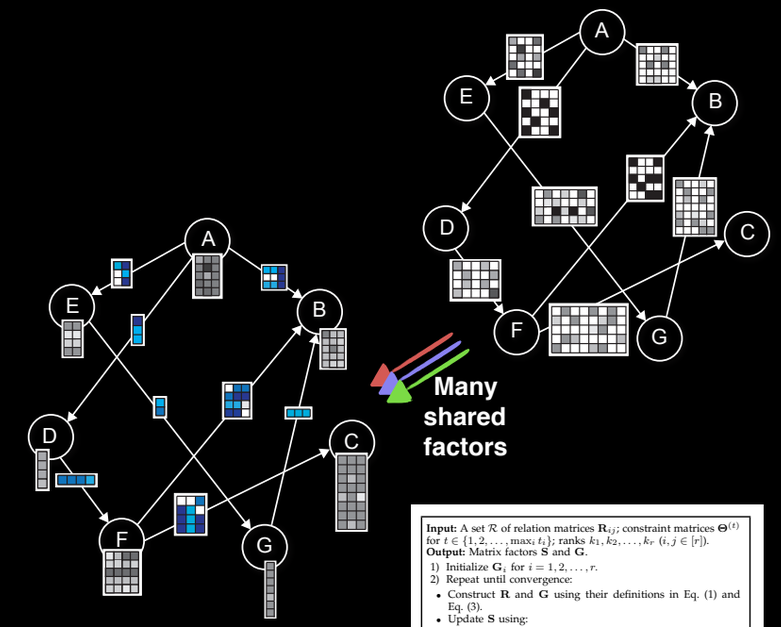
   - Construct $\mathbf{G}$ as:

$$\mathbf{G} \leftarrow \mathbf{G} \circ \mathrm{Diag}\left(\sqrt{\frac{\mathbf{G}_1^{(e)}}{\mathbf{G}_1^{(d)}}}, \sqrt{\frac{\mathbf{G}_2^{(e)}}{\mathbf{G}_2^{(d)}}}, \ldots, \sqrt{\frac{\mathbf{G}_r^{(e)}}{\mathbf{G}_r^{(d)}}}\right), \quad (12)$$

where $\circ$ denotes the Hadamard product. The $\sqrt{\cdot}$ and $\frac{\cdot}{\cdot}$ are entry-wise operations.

Many Data Matrices

Solution: DFMF Algorithm



Input: A set $\mathcal{R}$ of relation matrices $\mathbf{R}_{ij}$; constraint matrices $\mathbf{\Theta}^{(t)}$ for $t \in \{1, 2, \ldots, \max_i t_i\}$; ranks $k_1, k_2, \ldots, k_r$ $(i, j \in [r])$.
Output: Matrix factors $\mathbf{S}$ and $\mathbf{G}$.

1) Initialize $\mathbf{G}_i$ for $i = 1, 2, \ldots, r$.
2) Repeat until convergence:
   - Construct $\mathbf{R}$ and $\mathbf{G}$ using their definitions in Eq. (1) and Eq. (3).
   - Update $\mathbf{S}$ using:

$$\mathbf{S} \leftarrow (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}.$$

   - Set $\mathbf{G}_i^{(e)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
   - Set $\mathbf{G}_i^{(d)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
   - For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^+ + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^- \\
\mathbf{G}_i^{(d)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^- + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^+ \\
\mathbf{G}_j^{(e)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^+ + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^- \\
\mathbf{G}_j^{(d)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^- + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^+ \quad (10)
\end{aligned}$$

   - For $t = 1, 2, \ldots, \max_i t_i$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= [\mathbf{\Theta}_i^{(t)}]^-\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \\
\mathbf{G}_i^{(d)} &+= [\mathbf{\Theta}_i^{(t)}]^+\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \quad (11)
\end{aligned}$$

   - Construct $\mathbf{G}$ as:

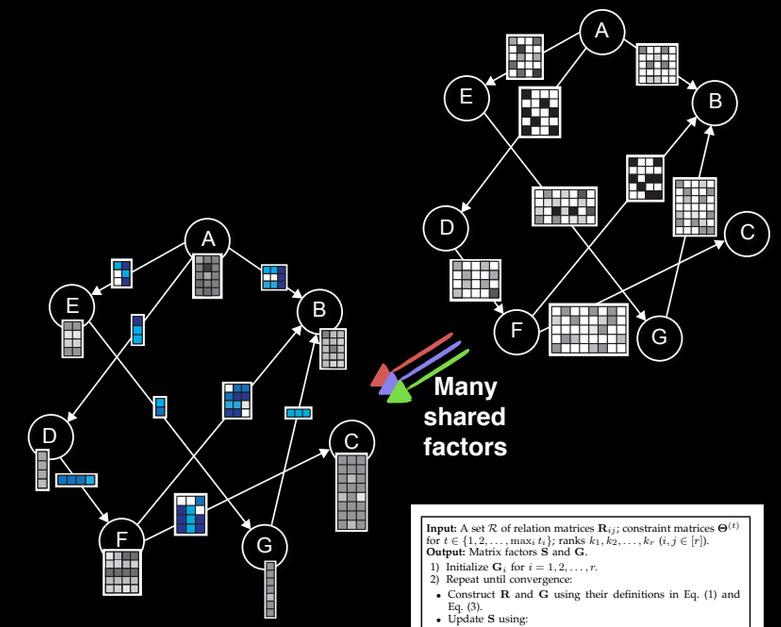
$$\mathbf{G} \leftarrow \mathbf{G} \circ \mathrm{Diag}\left(\sqrt{\frac{\mathbf{G}_1^{(e)}}{\mathbf{G}_1^{(d)}}}, \sqrt{\frac{\mathbf{G}_2^{(e)}}{\mathbf{G}_2^{(d)}}}, \ldots, \sqrt{\frac{\mathbf{G}_r^{(e)}}{\mathbf{G}_r^{(d)}}}\right), \quad (12)$$

where $\circ$ denotes the Hadamard product. The $\sqrt{\cdot}$ and $\frac{\cdot}{\cdot}$ are entry-wise operations.

Theorem 1 (Correctness of DFMF algorithm): If the update rules for matrix factors $\mathbf{G}_i$ and $\mathbf{S}_{ij}$ from the DFMF algorithm converge, then the final solution satisfies the Karush-Kuhn-Tucker conditions of optimality.

Many Data Matrices

Solution: DFMF Algorithm



Input: A set $\mathcal{R}$ of relation matrices $\mathbf{R}_{ij}$; constraint matrices $\mathbf{\Theta}_i^{(t)}$ for $t \in \{1, 2, \ldots, \max_i t_i\}$; ranks $k_1, k_2, \ldots, k_r$ $(i, j \in [r])$.
Output: Matrix factors $\mathbf{S}$ and $\mathbf{G}$.

1) Initialize $\mathbf{G}_i$ for $i = 1, 2, \ldots, r$.
2) Repeat until convergence:
- Construct $\mathbf{R}$ and $\mathbf{G}$ using their definitions in Eq. (1) and Eq. (3).
- Update $\mathbf{S}$ using:

$$\mathbf{S} \leftarrow (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}.$$

- Set $\mathbf{G}_i^{(e)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
- Set $\mathbf{G}_i^{(d)} \leftarrow \mathbf{0}$ for $i = 1, 2, \ldots, r$.
- For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^+ + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^- \\
\mathbf{G}_i^{(d)} &+= (\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T)^- + \mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T)^+ \\
\mathbf{G}_j^{(e)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^+ + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^- \\
\mathbf{G}_j^{(d)} &+= (\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij})^- + \mathbf{G}_j(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij})^+
\end{aligned} \quad (10)$$

- For $t = 1, 2, \ldots, \max_i t_i$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} &+= [\mathbf{\Theta}_i^{(t)}]^-\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r \\
\mathbf{G}_i^{(d)} &+= [\mathbf{\Theta}_i^{(t)}]^+\mathbf{G}_i \quad \text{for } i = 1, 2, \ldots, r
\end{aligned} \quad (11)$$

- Construct $\mathbf{G}$ as:

$$\mathbf{G} \leftarrow \mathbf{G} \circ \mathrm{Diag}(\sqrt{\frac{\mathbf{G}_1^{(e)}}{\mathbf{G}_1^{(d)}}}, \sqrt{\frac{\mathbf{G}_2^{(e)}}{\mathbf{G}_2^{(d)}}}, \ldots, \sqrt{\frac{\mathbf{G}_r^{(e)}}{\mathbf{G}_r^{(d)}}}), \quad (12)$$

where $\circ$ denotes the Hadamard product. The $\sqrt{\cdot}$ and $\frac{\cdot}{\cdot}$ are entry-wise operations.

Theorem 1 (Correctness of DFMF algorithm): If the update rules for matrix factors $\mathbf{G}_i$ and $\mathbf{S}_{ij}$ from the DFMF algorithm converge, then the final solution satisfies the Karush-Kuhn-Tucker conditions of optimality.

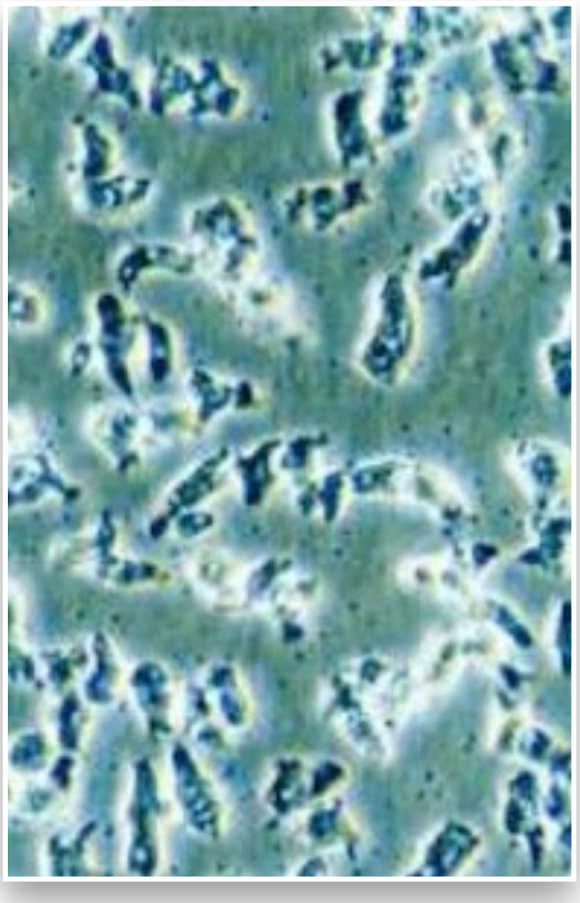
Theorem 2 (Convergence of DFMF algorithm): The objective function:

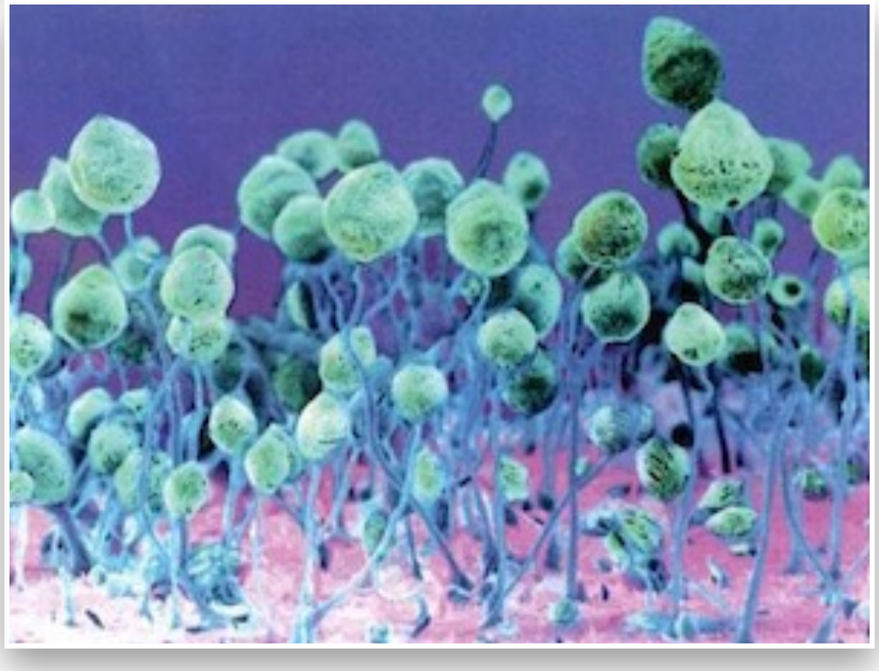
$$\min_{\mathbf{G}_i \geq 0, \mathbf{S}_{ij}} \sum_{\mathbf{R}_{ij} \in \mathcal{R}} \|\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\|_{\mathrm{Fro}}^2 + \sum_{\mathbf{\Theta}_i \in \mathcal{C}} \sum_{l=1}^{l_i} \mathrm{tr}(\mathbf{G}_i^T\mathbf{\Theta}_i^{(l)}\mathbf{G}_i)$$

is nonincreasing under the updating rules for matrix factors $\mathbf{G}_i$ and $\mathbf{S}_{ij}$ given by DFMF algorithm.

Two Case Studies of Collective Matrix Factorization

#1: Amoeba

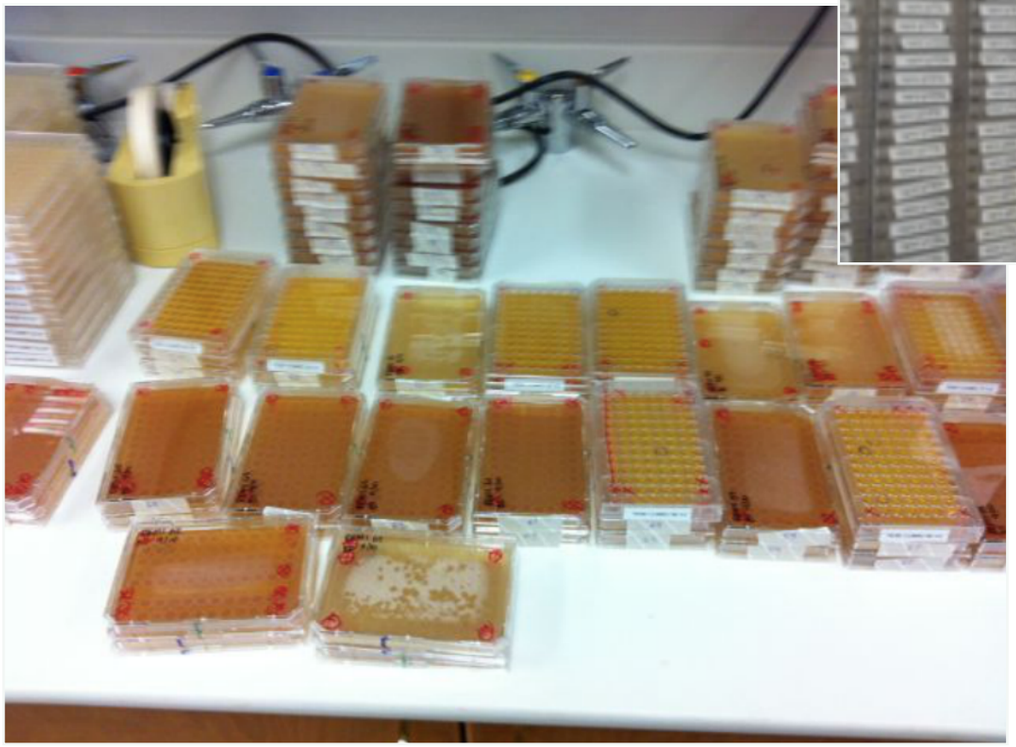




Search for Bacterial Response Genes

50,000 clonal mutants

genetic screen



genome	12,000 genes
found	7 genes
workload	5 years
estimated	~200 genes

Gram+ defective:
swp1, gpi, nagB1

Gram- defective:
clkB, spc3, alyL, nip7

Nasser et al (2013) *Curr Biol*

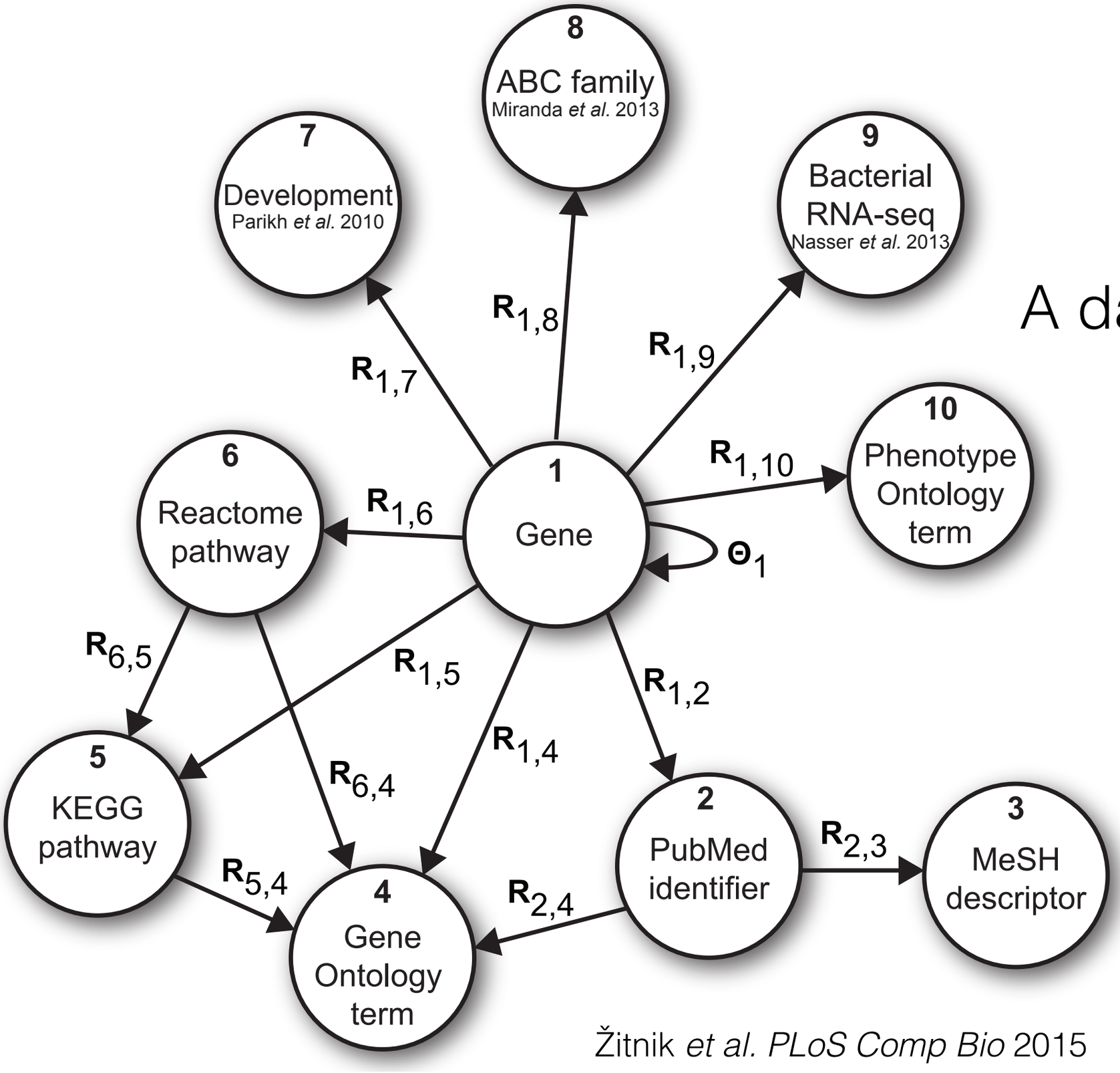
Dictyostelium Bacterial Gene Hunt

A data-driven approach

14 data sources
4 Gram- seed genes
9 candidate genes

Žitnik *et al. PLoS Comp Bio* 2015

Dictyostelium Bacterial Gene Hunt

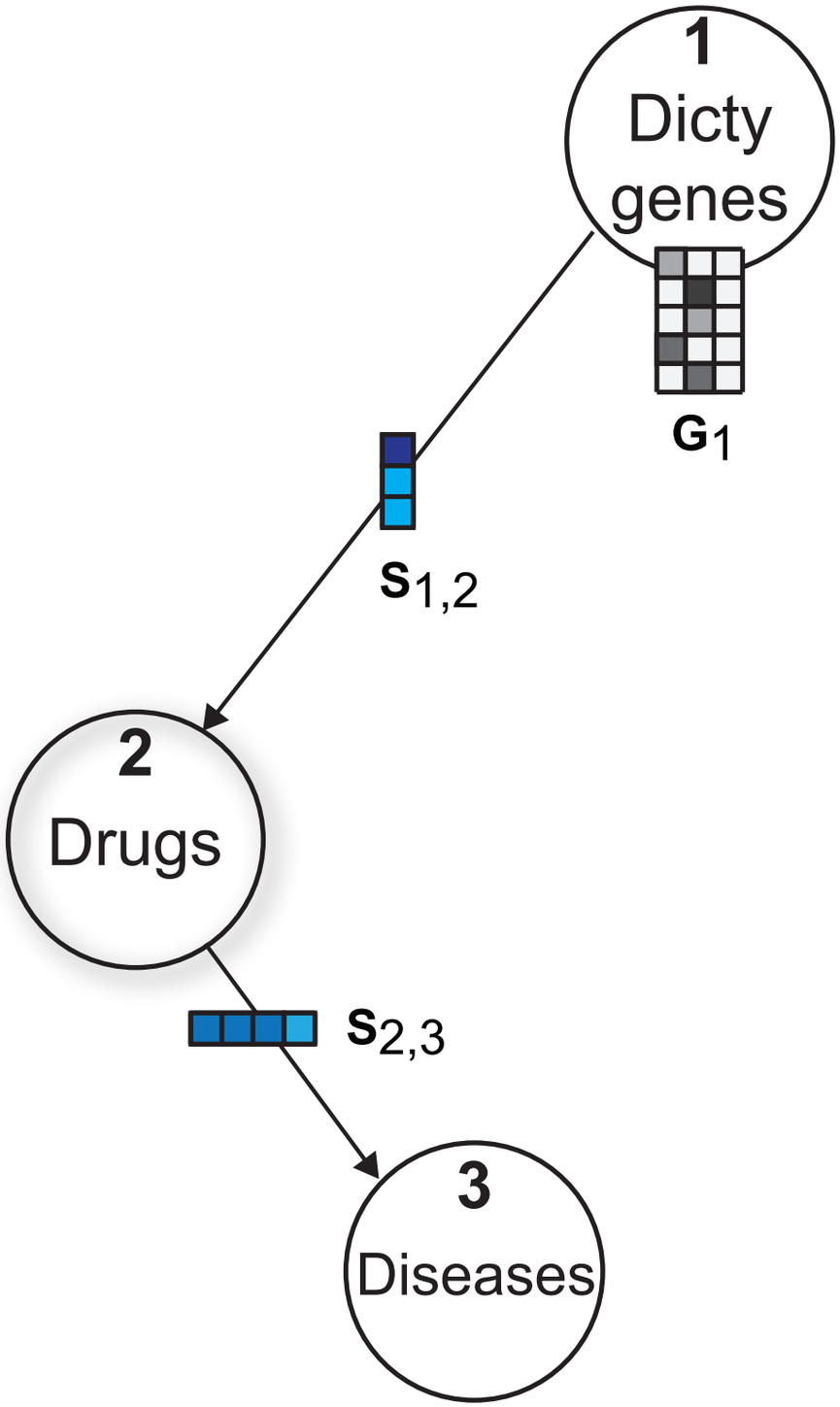


A data-driven approach

14 data sources
4 Gram- seed genes
9 candidate genes

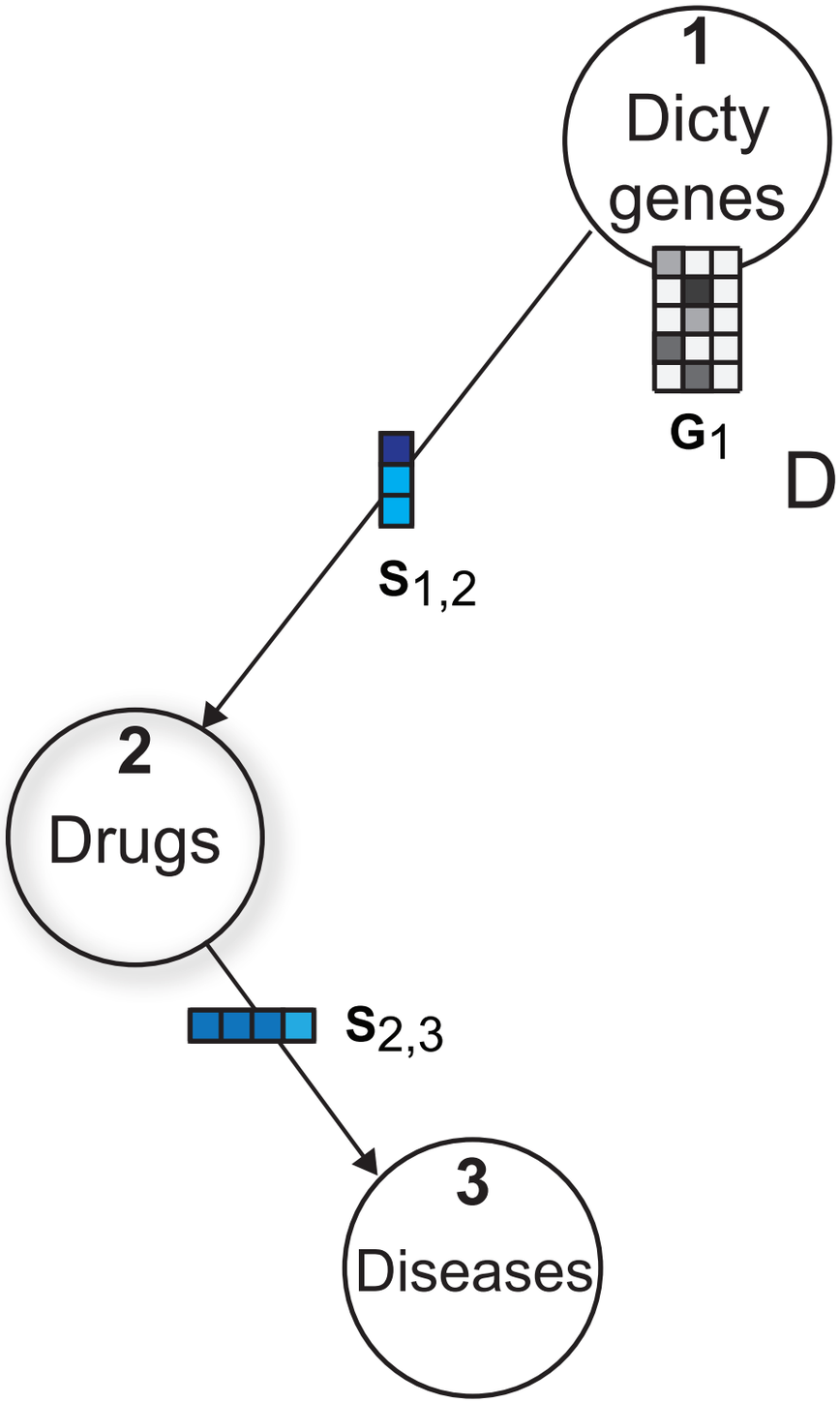
Žitnik *et al. PLoS Comp Bio* 2015

Latent Chaining and Profiling



Žitnik *et al. PLoS Comp Bio* 2015

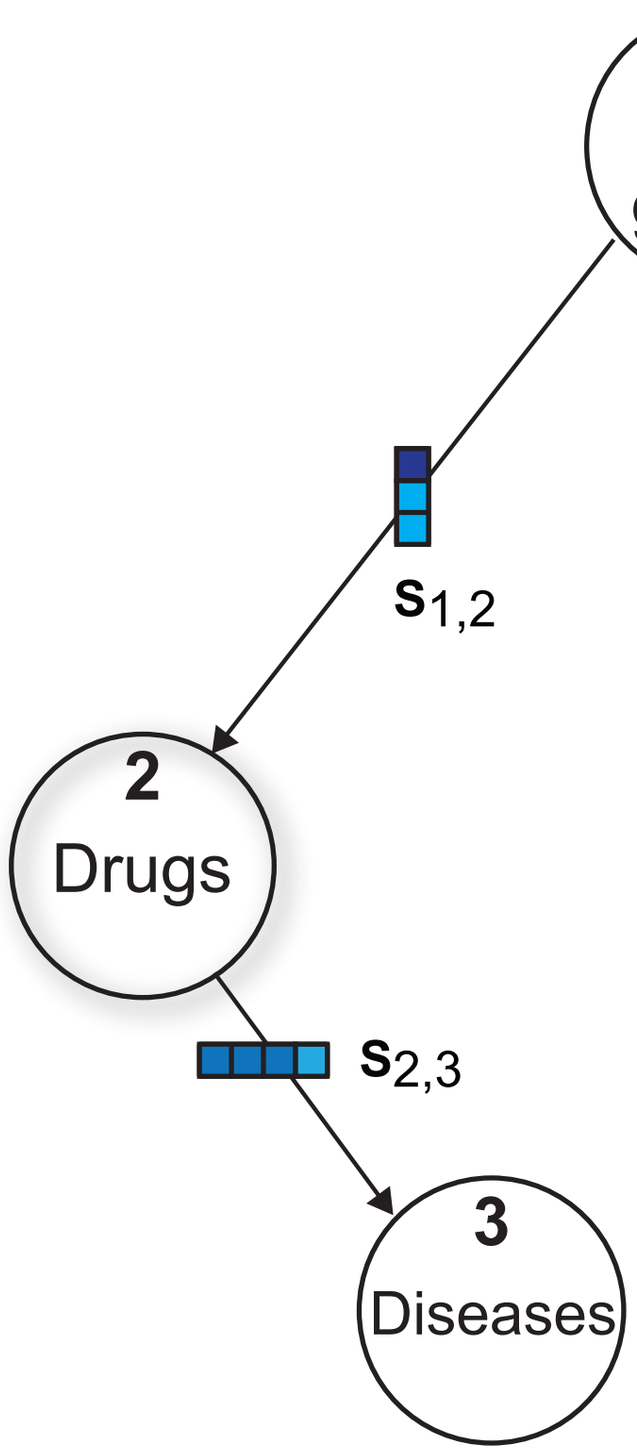
Latent Chaining and Profiling



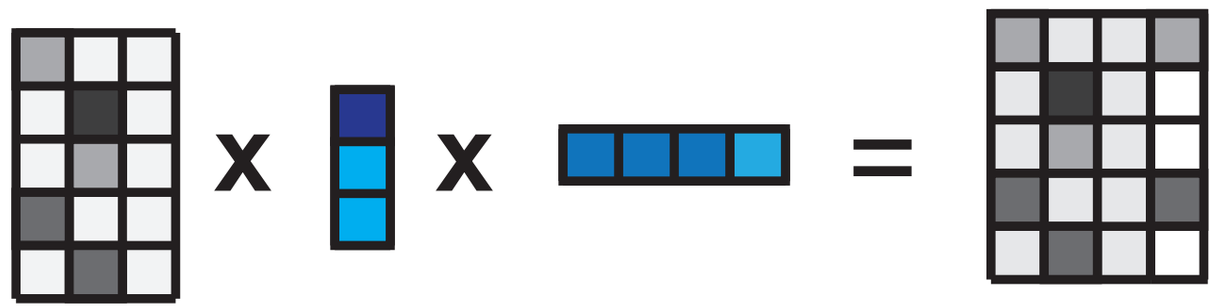
Dicty genes → Drugs → Diseases

Žitnik *et al. PLoS Comp Bio* 2015

Latent Chaining and Profiling



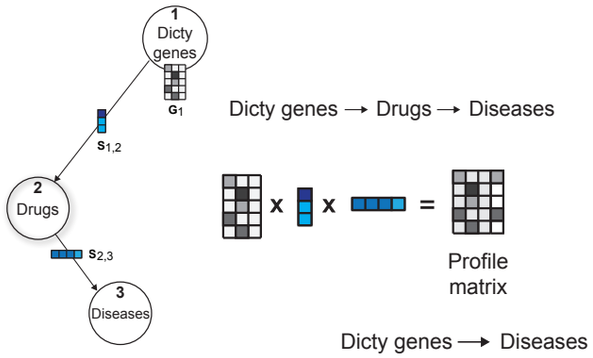
Dicty genes → Drugs → Diseases



Profile matrix

Dicty genes → Diseases

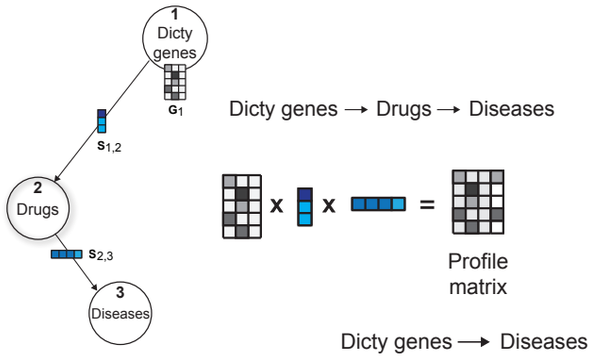
Žitnik *et al. PLoS Comp Bio* 2015



Latent Chaining and Profiling

Žitnik *et al. PLoS Comp Bio* 2015

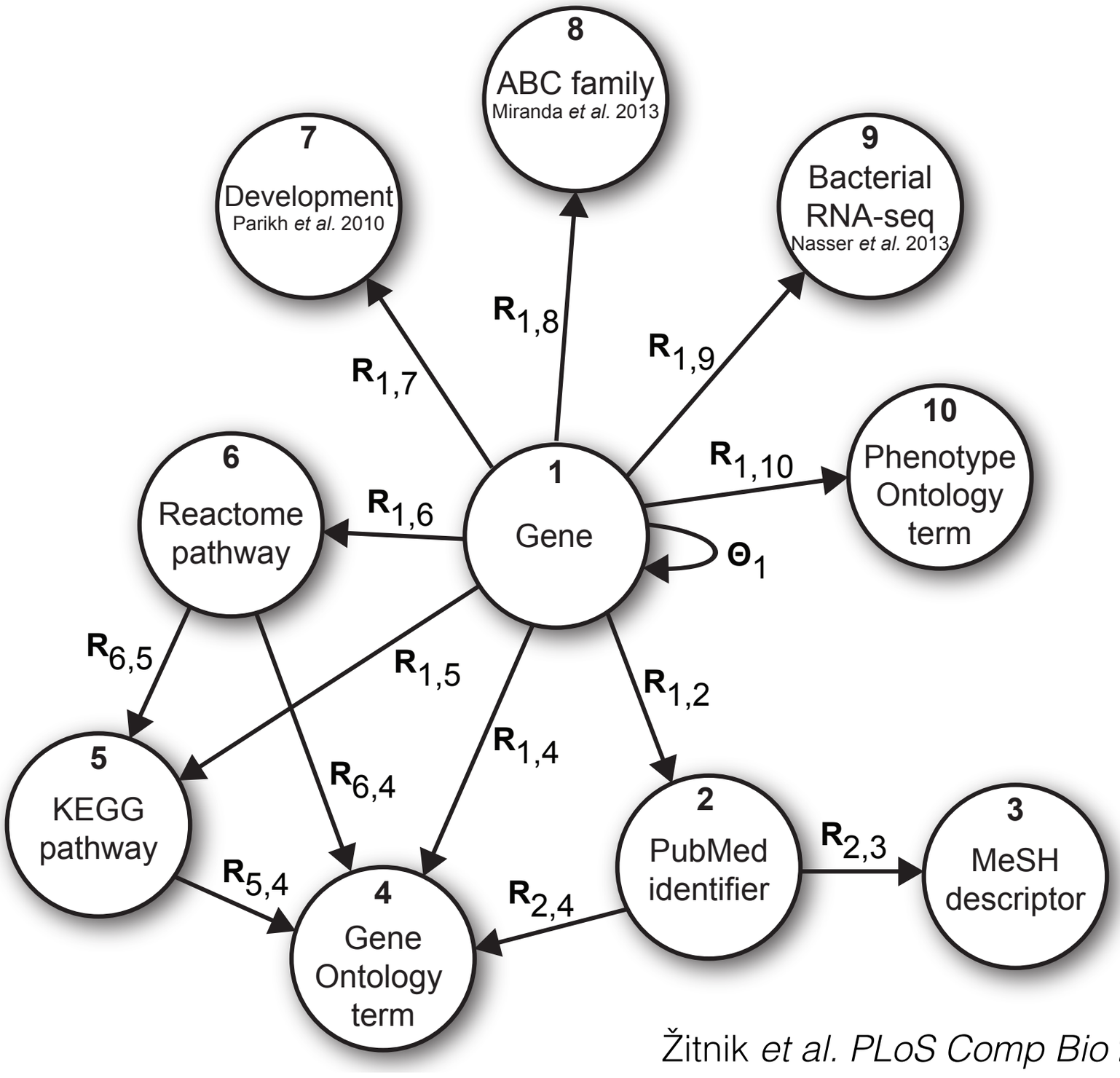
Latent Chaining and Profiling



Dicty genes → Drugs → Diseases

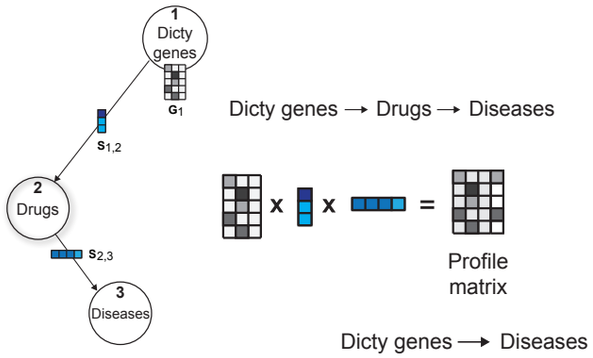
Profile matrix

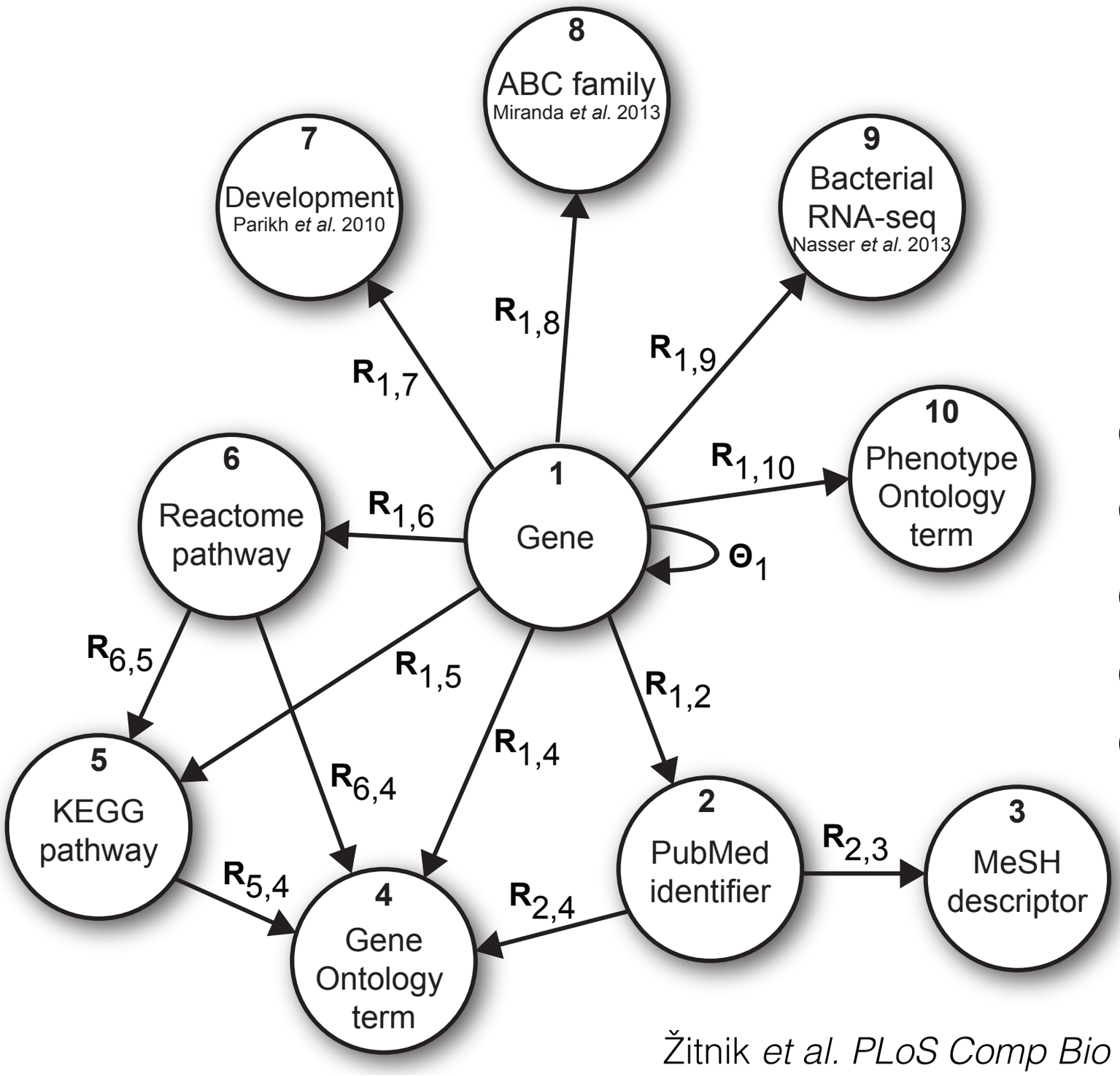
Dicty genes → Diseases



Žitnik *et al. PLoS Comp Bio* 2015

Latent Chaining and Profiling

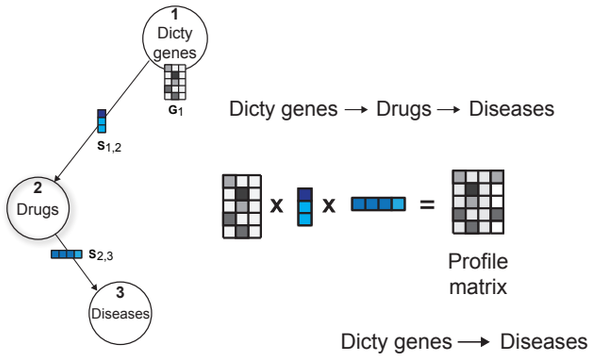


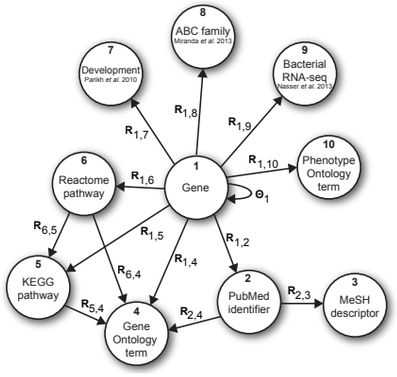


Latent chains

$\mathbf{G}_1$, $\mathbf{G}_1\mathbf{S}_{1,7}$, $\mathbf{G}_1\mathbf{S}_{1,8}$, $\mathbf{G}_1\mathbf{S}_{1,9}$, $\mathbf{G}_1\mathbf{S}_{1,10}$, $\mathbf{G}_1\mathbf{S}_{1,2}$, $\mathbf{G}_1\mathbf{S}_{1,6}$, $\mathbf{G}_1\mathbf{S}_{1,5}$, $\mathbf{G}_1\mathbf{S}_{1,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,3}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,4}$, $\mathbf{G}_1\mathbf{S}_{1,5}\mathbf{S}_{5,4}$ and $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}\mathbf{S}_{5,4}$

Žitnik *et al. PLoS Comp Bio* 2015





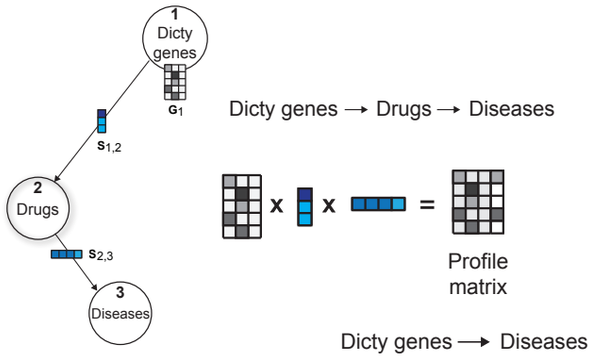
Latent chains

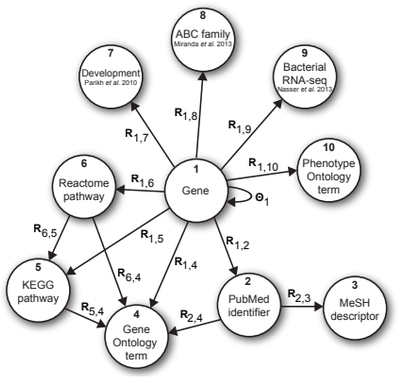
$\mathbf{G}_1$, $\mathbf{G}_1\mathbf{S}_{1,7}$, $\mathbf{G}_1\mathbf{S}_{1,8}$, $\mathbf{G}_1\mathbf{S}_{1,9}$, $\mathbf{G}_1\mathbf{S}_{1,10}$, $\mathbf{G}_1\mathbf{S}_{1,2}$, $\mathbf{G}_1\mathbf{S}_{1,6}$, $\mathbf{G}_1\mathbf{S}_{1,5}$, $\mathbf{G}_1\mathbf{S}_{1,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,3}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,4}$, $\mathbf{G}_1\mathbf{S}_{1,5}\mathbf{S}_{5,4}$ and $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}\mathbf{S}_{5,4}$

Latent Chaining and Profiling

Žitnik *et al. PLoS Comp Bio* 2015

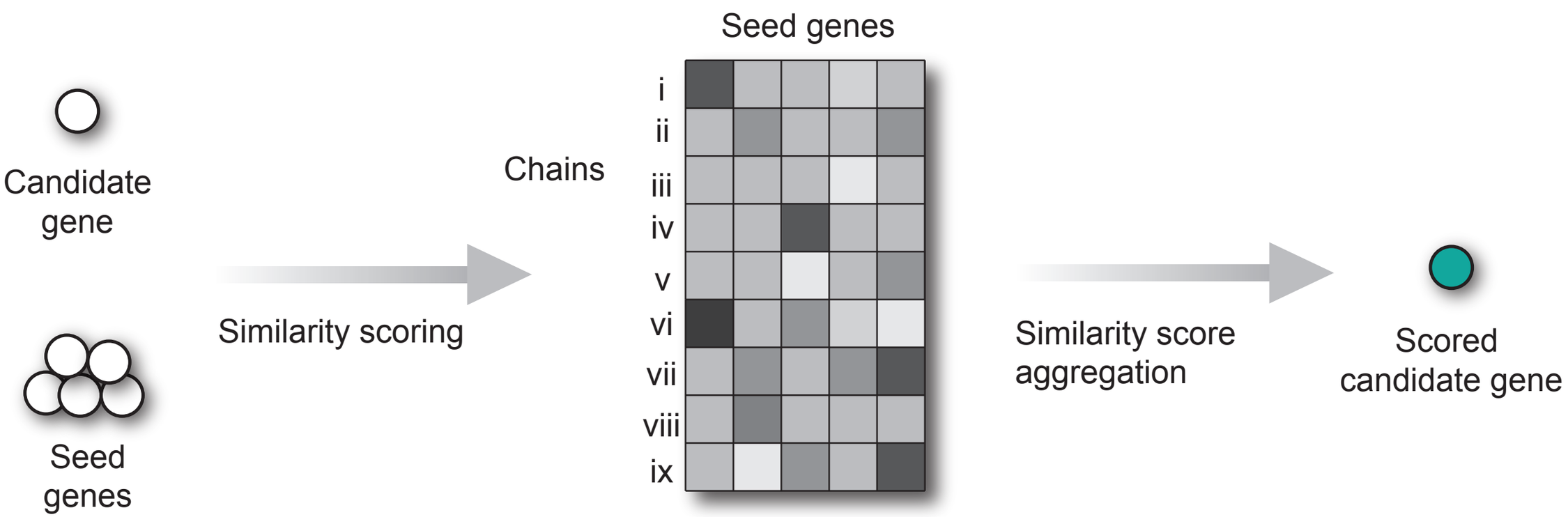
Latent Chaining and Profiling





Latent chains

$\mathbf{G}_1$, $\mathbf{G}_1\mathbf{S}_{1,7}$, $\mathbf{G}_1\mathbf{S}_{1,8}$, $\mathbf{G}_1\mathbf{S}_{1,9}$, $\mathbf{G}_1\mathbf{S}_{1,10}$, $\mathbf{G}_1\mathbf{S}_{1,2}$, $\mathbf{G}_1\mathbf{S}_{1,6}$, $\mathbf{G}_1\mathbf{S}_{1,5}$, $\mathbf{G}_1\mathbf{S}_{1,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,3}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}$, $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,4}$, $\mathbf{G}_1\mathbf{S}_{1,2}\mathbf{S}_{2,4}$, $\mathbf{G}_1\mathbf{S}_{1,5}\mathbf{S}_{5,4}$ and $\mathbf{G}_1\mathbf{S}_{1,6}\mathbf{S}_{6,5}\mathbf{S}_{5,4}$



Žitnik *et al. PLoS Comp Bio* 2015

Dictyostelium Bacterial Gene Hunt

Žitnik *et al. PLoS Comp Bio* 2015

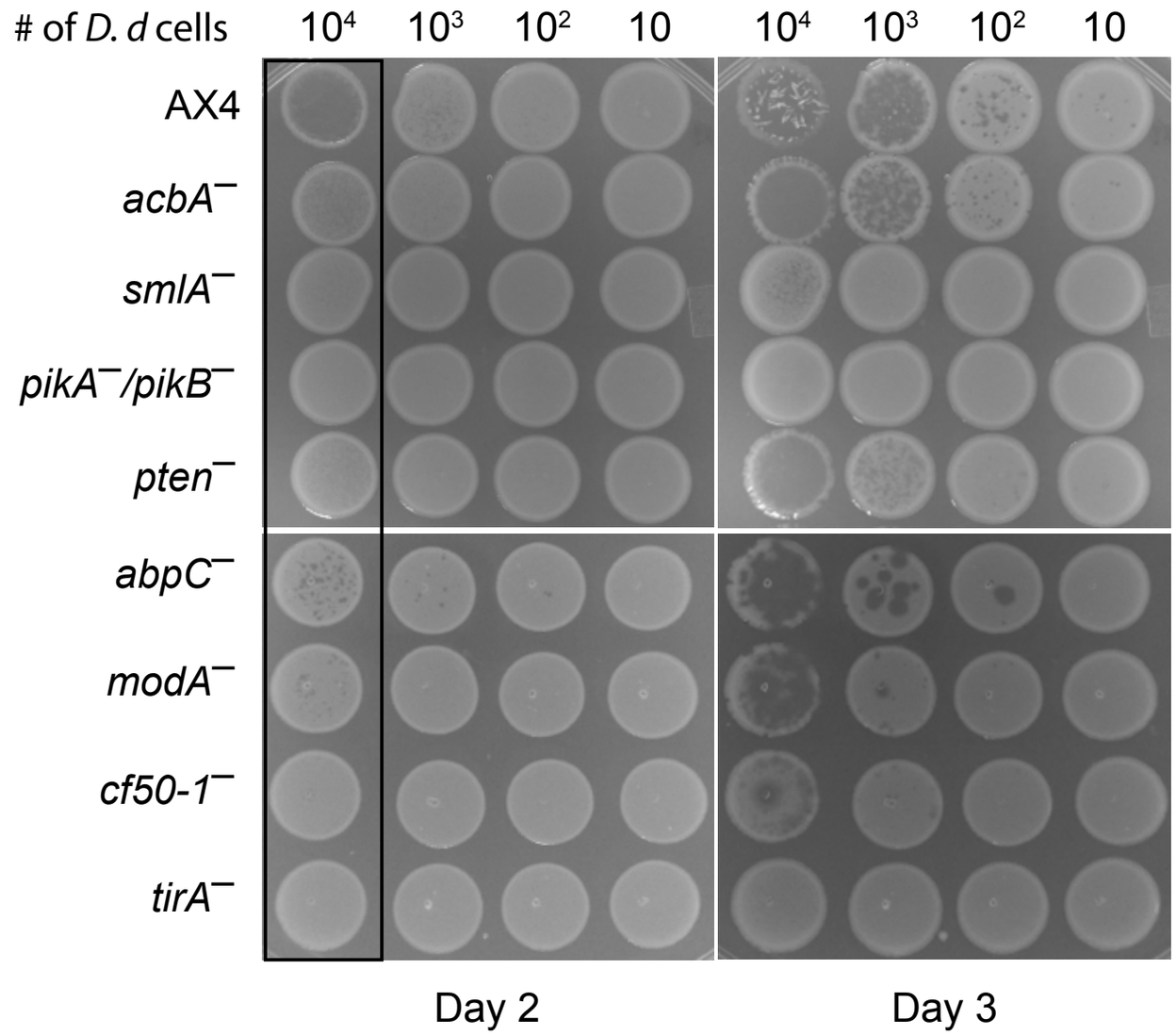
Dictyostelium Bacterial Gene Hunt

cf50-1
smlA
acbA
pirA
rps10
abpC
tirA
DDB_G0272184
pikB
vps46
pikA
swp1
ggtA
DDB_G0288519
pten
DDB_G0288551
tra2
DDB_G0286429
dscA-1
cinC
udpB
sfbA
modA
DDB_G0287399

Žitnik *et al. PLoS Comp Bio* 2015

Dictyostelium Bacterial Gene Hunt

cf50-1
smlA
acbA
pirA
rps10
abpC
tirA
DDB_G0272184
pikB
vps46
pikA
swp1
ggtA
DDB_G0288519
pten
DDB_G0288551
tra2
DDB_G0286429
dscA-1
cinC
udpB
sfbA
modA
DDB_G0287399



8/9 predictions correct!

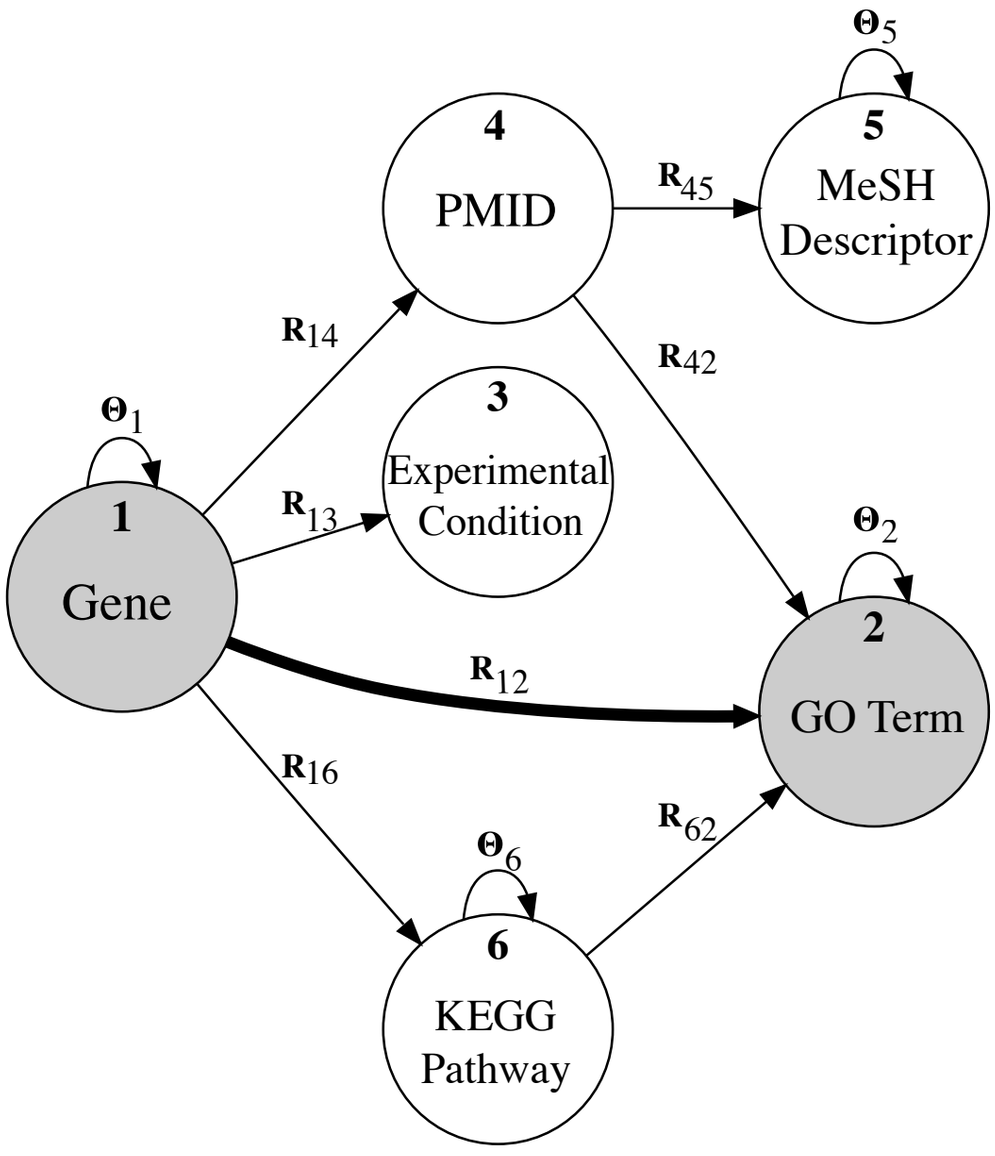
14 data sources
4 Gram- seed genes
9 candidate genes

Žitnik *et al. PLoS Comp Bio* 2015

#2: Functional Genomics

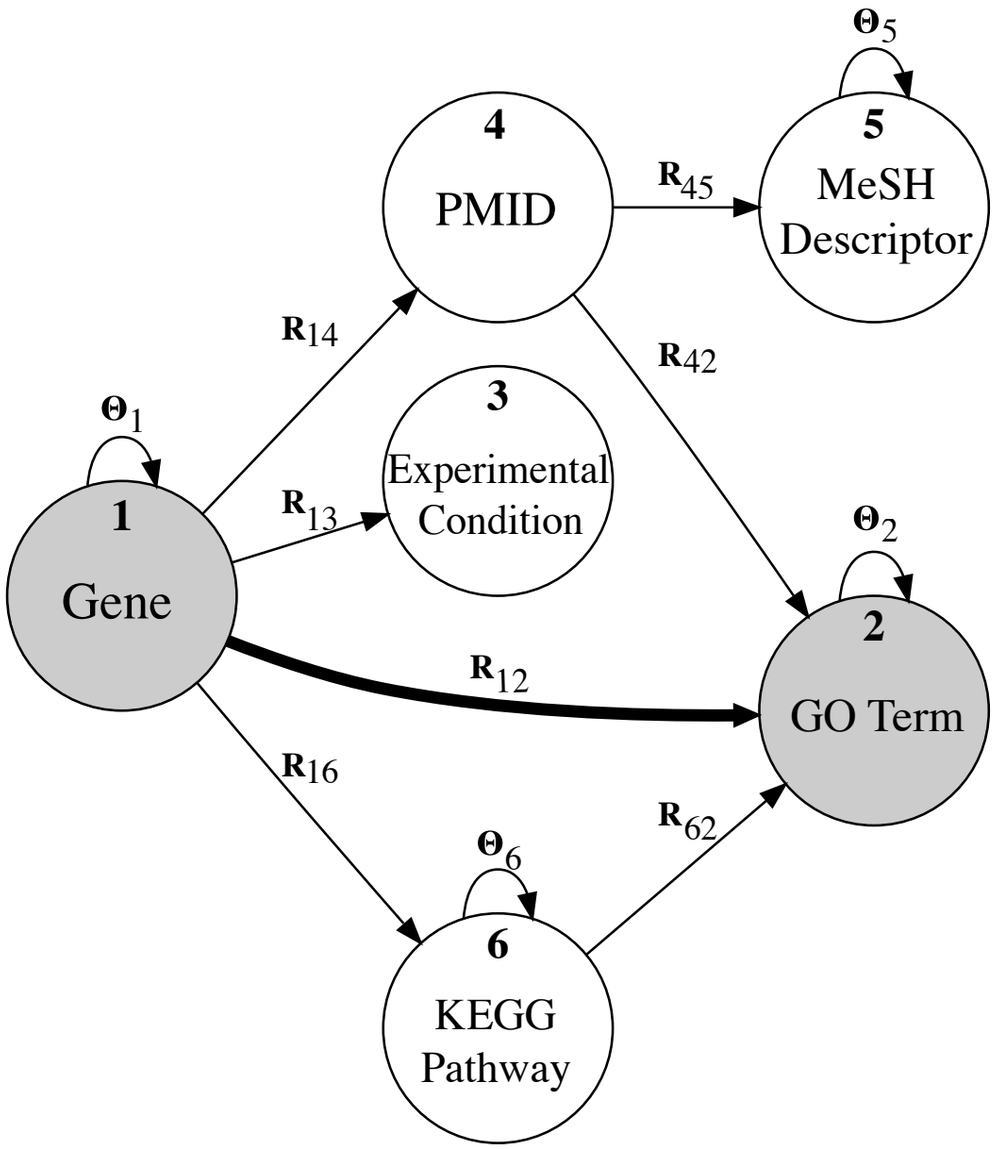
Žitnik & Zupan *IEEE TPAMI* 2015

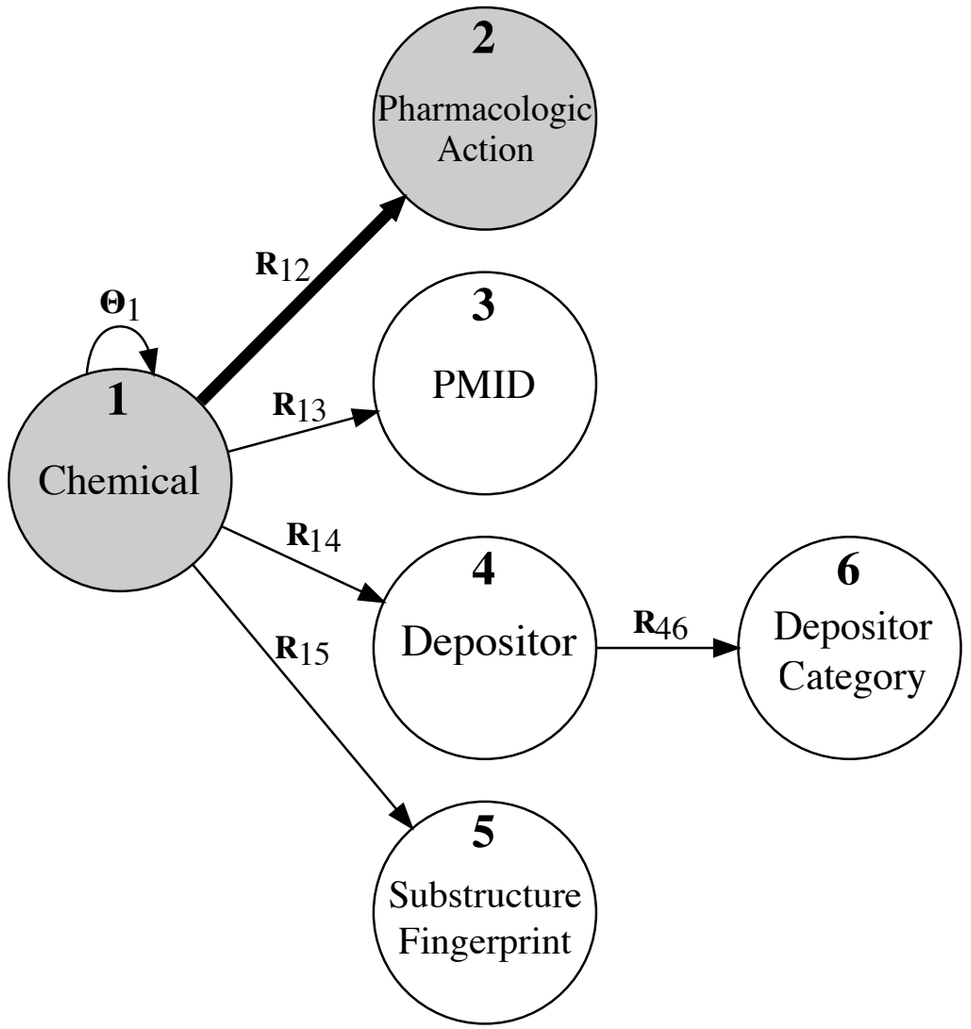
#2: Functional Genomics



Žitnik & Zupan *IEEE TPAMI* 2015

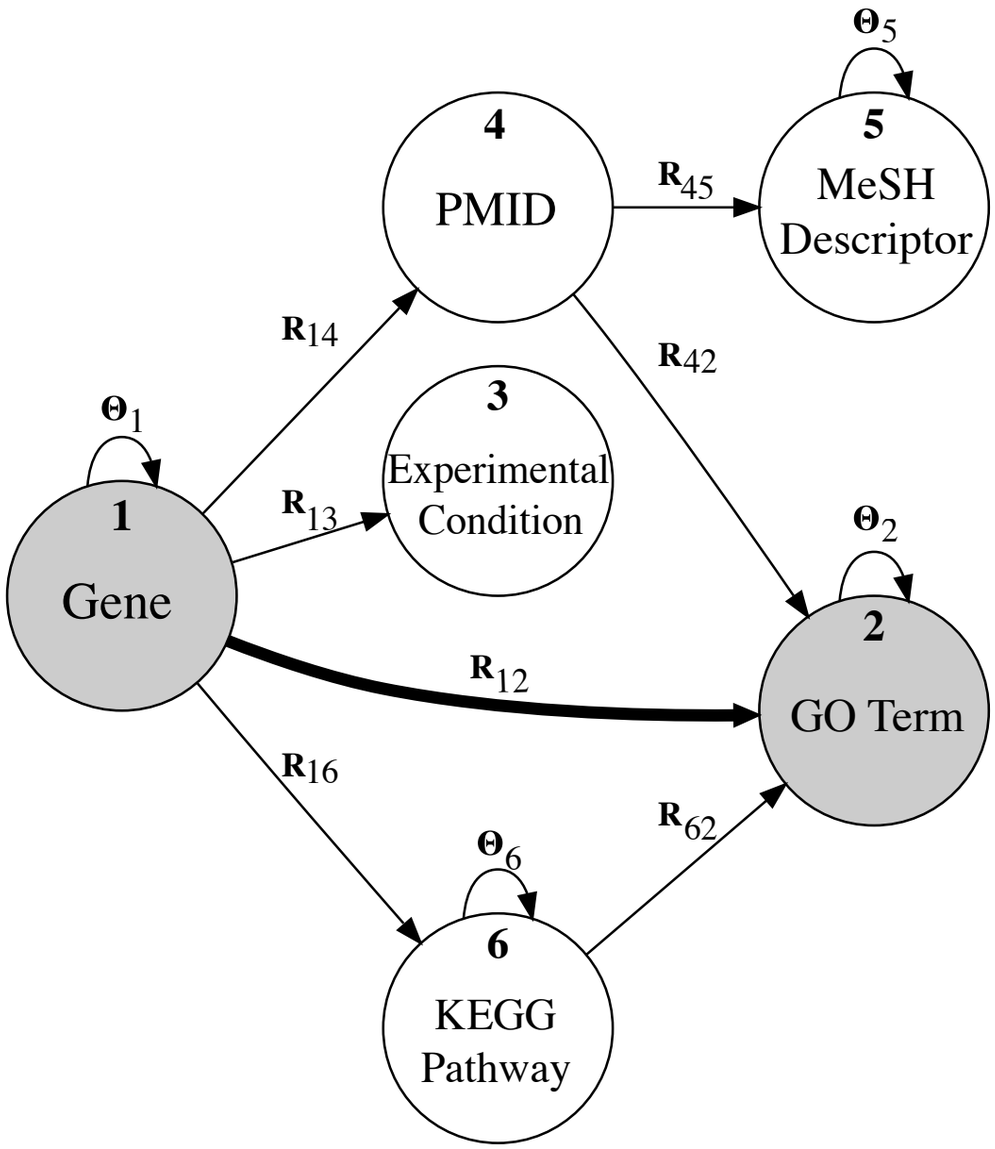
#2: Functional Genomics

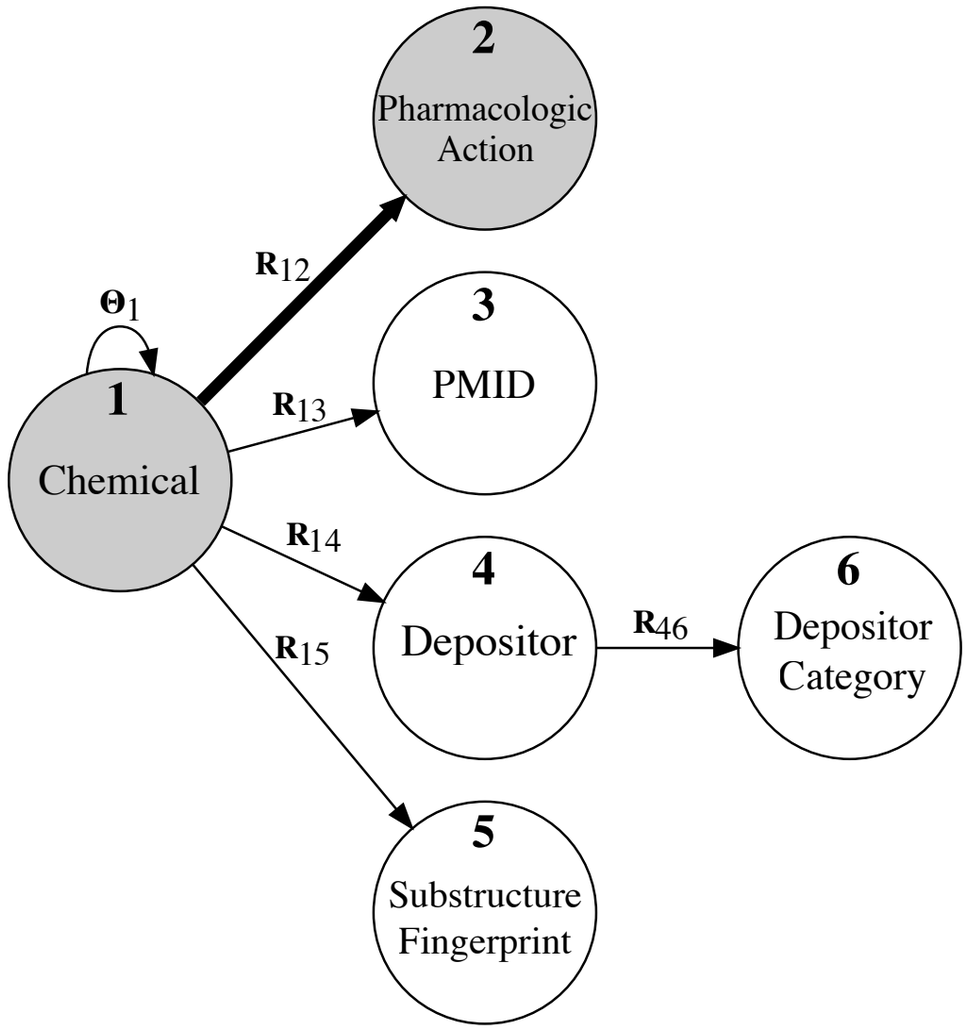




Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics

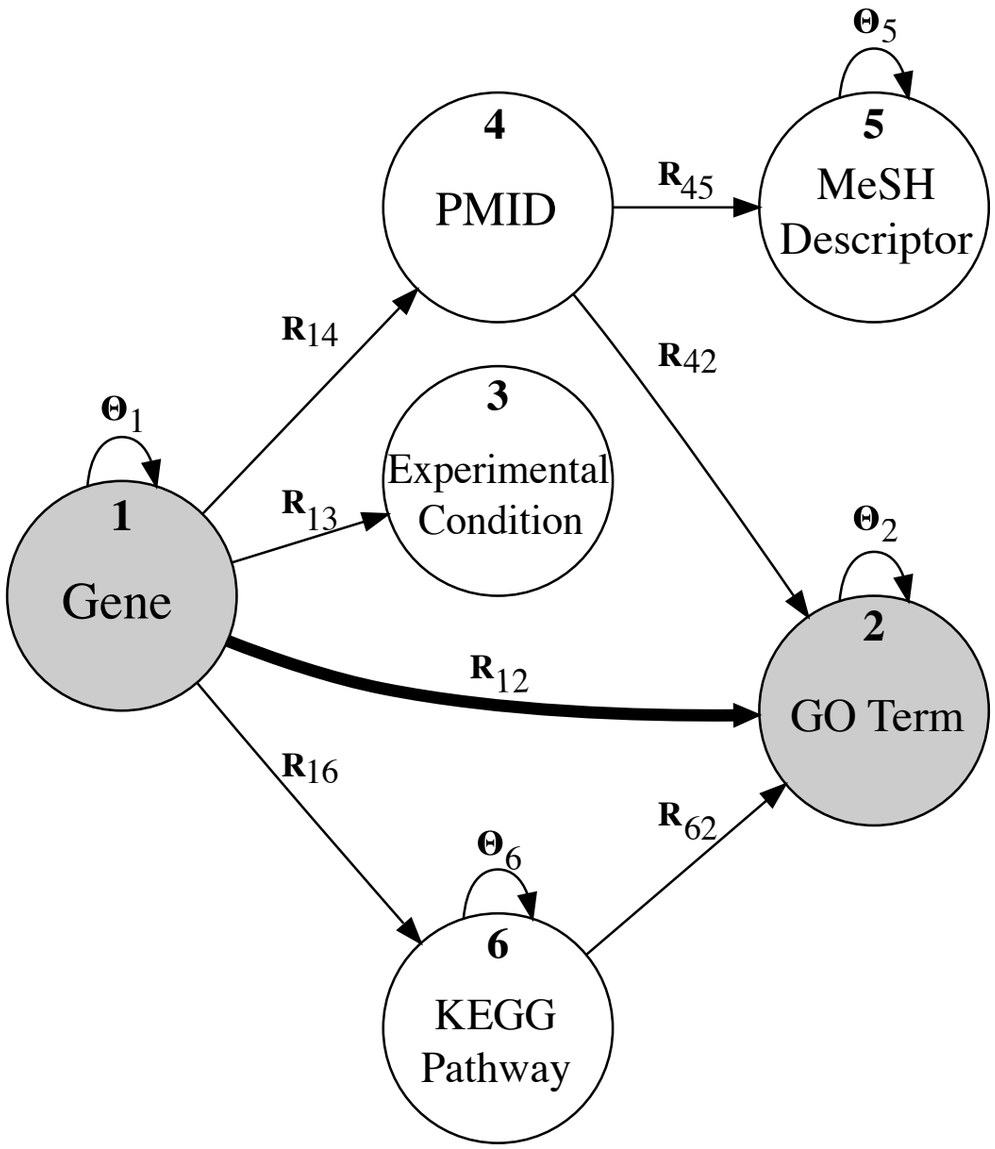


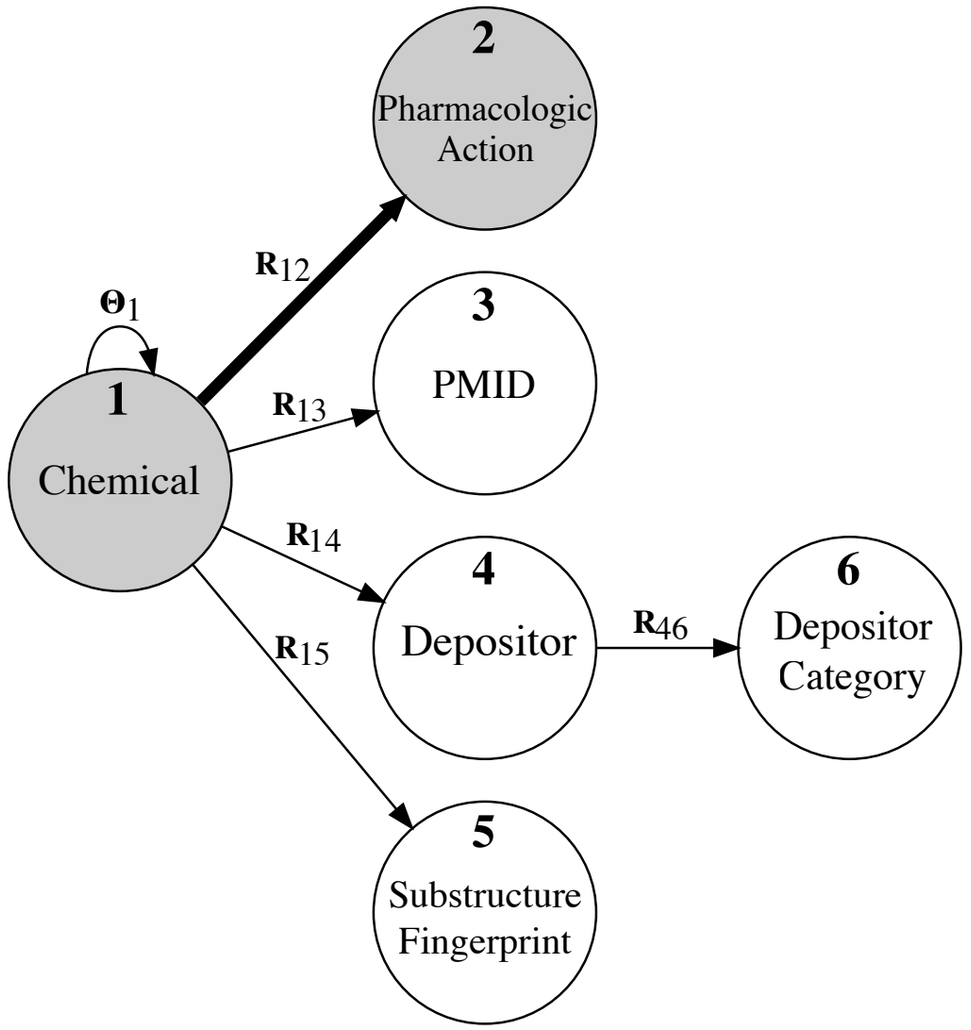


Prediction task	DFMF	
	F_1	AUC
100 *D. discoideum* genes	0.799	0.801
1000 *D. discoideum* genes	0.826	0.823
Whole *D. discoideum* genome	0.831	0.849
Pharmacologic actions	0.663	0.834

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics

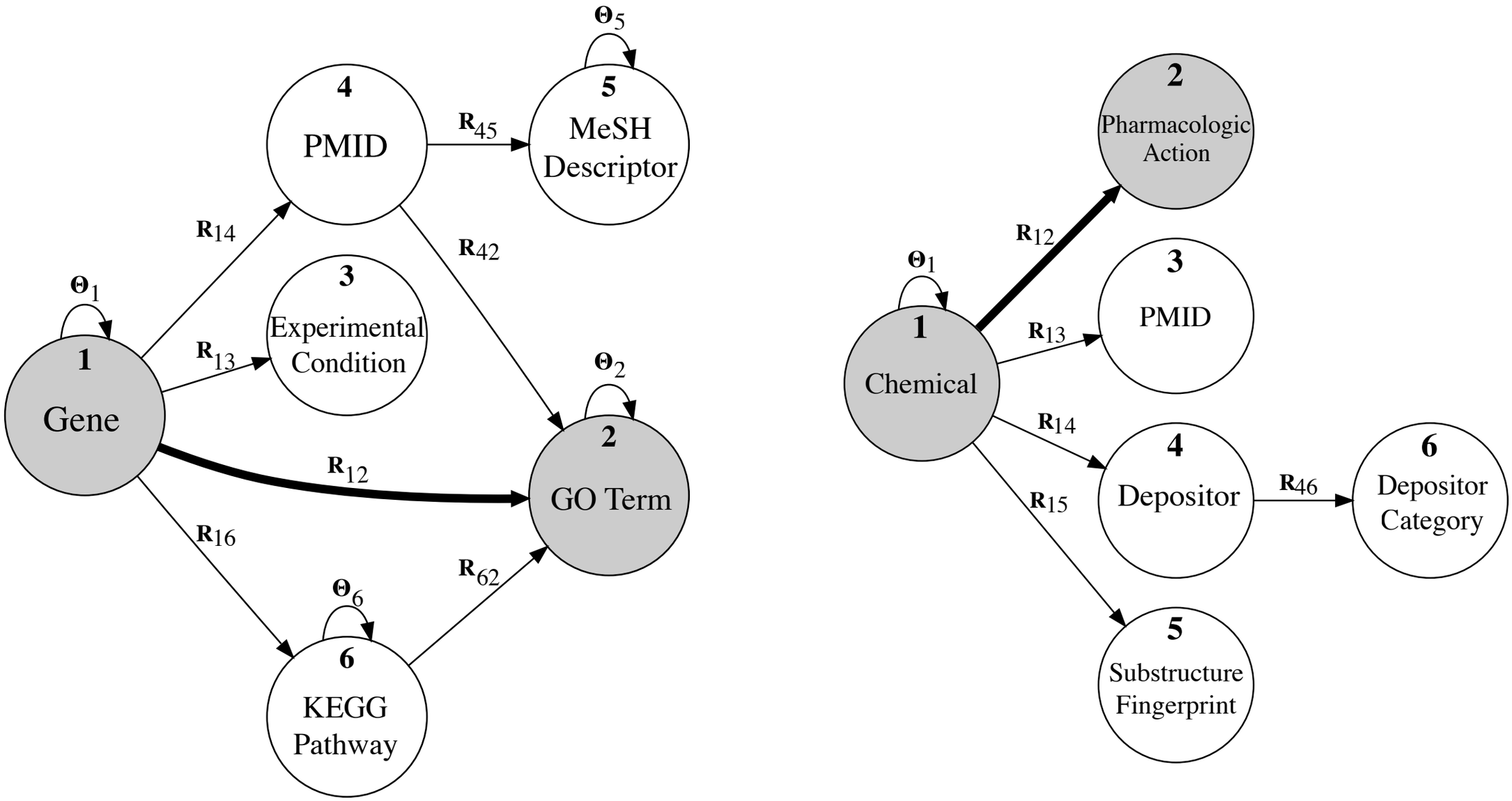




Prediction task	DFMF		MKL	
	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821
Pharmacologic actions	0.663	0.834	0.639	0.811

Žitnik & Zupan *IEEE TPAMI* 2015

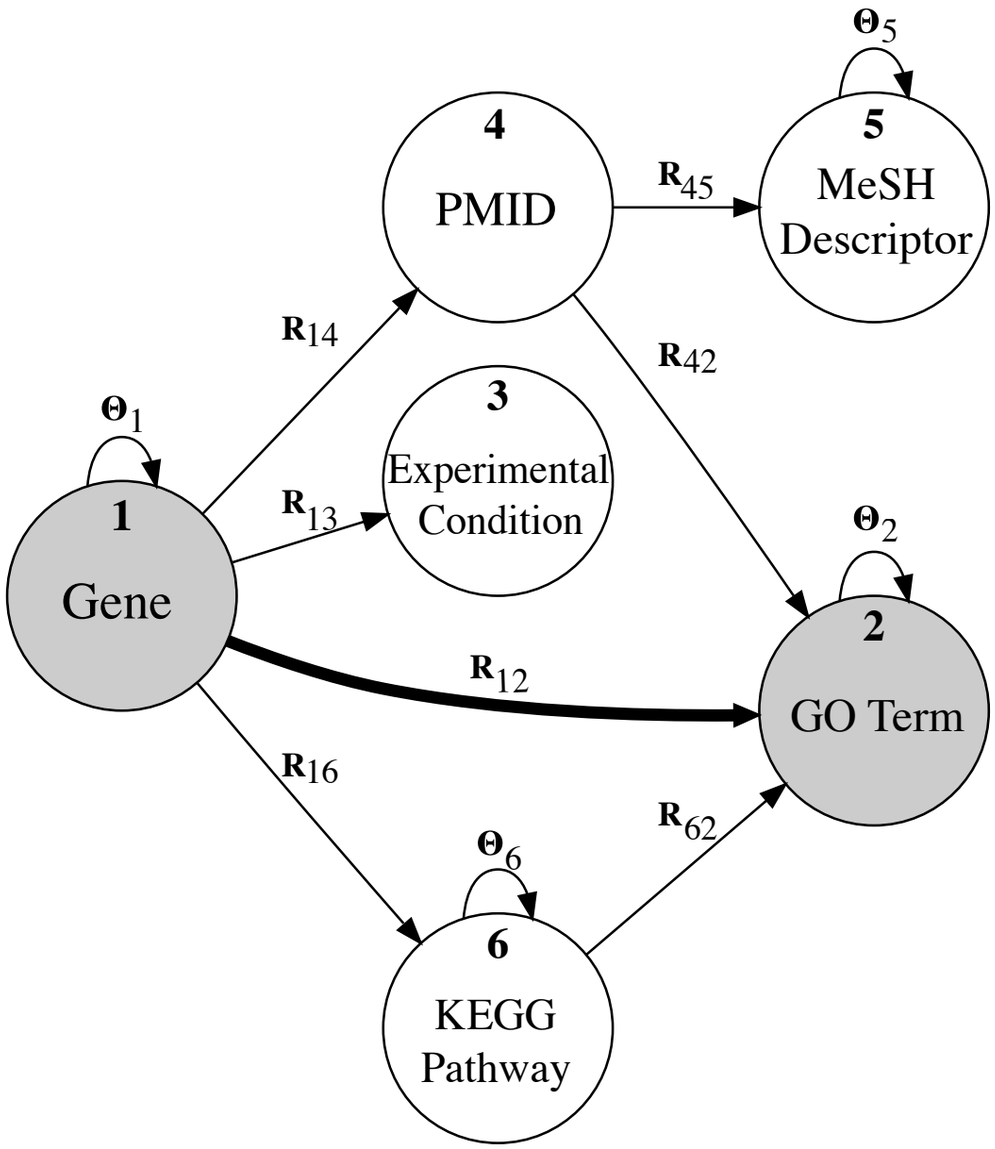
#2: Functional Genomics

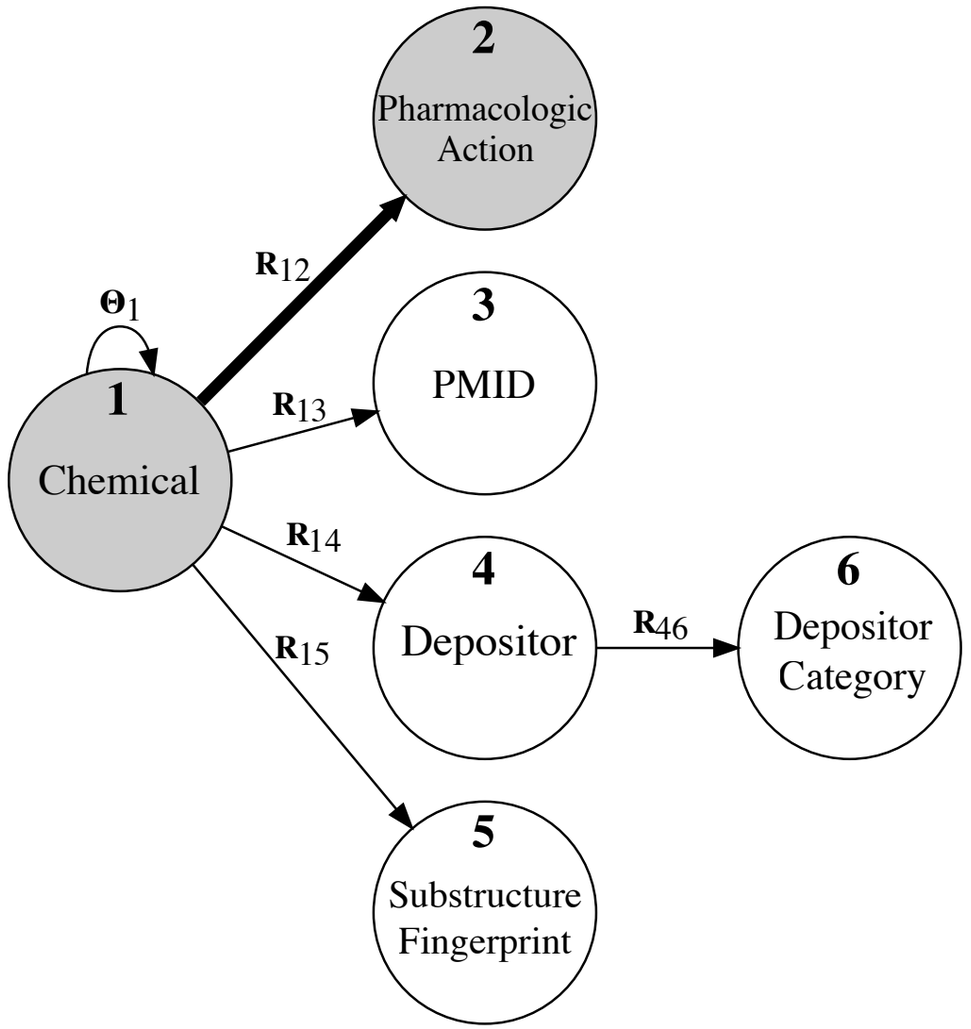


Prediction task	DFMF		MKL		RF	
	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics

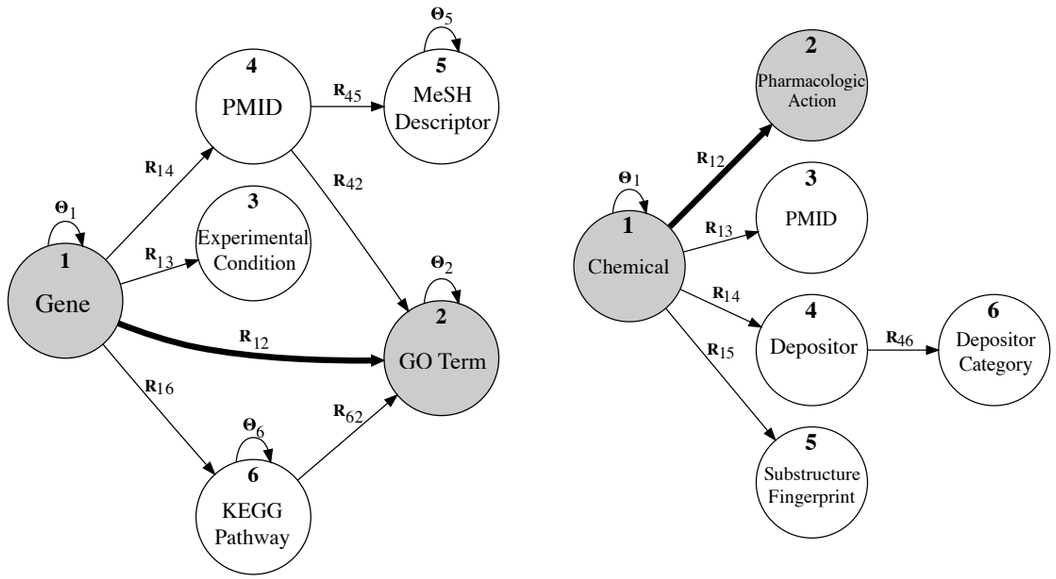




Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

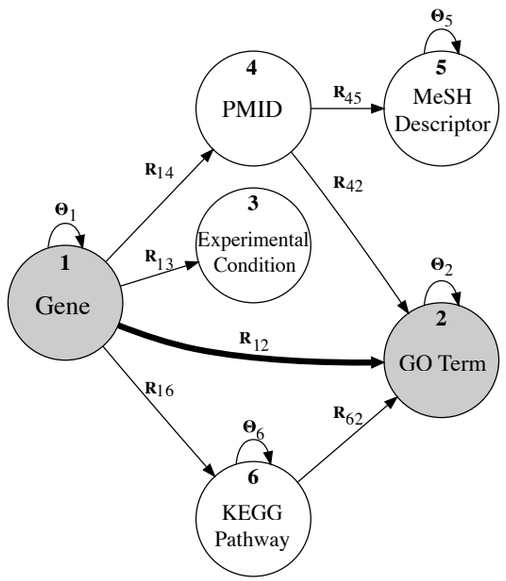
#2: Functional Genomics

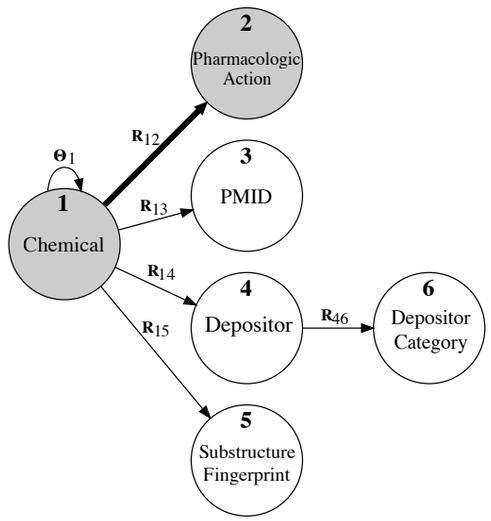


Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics

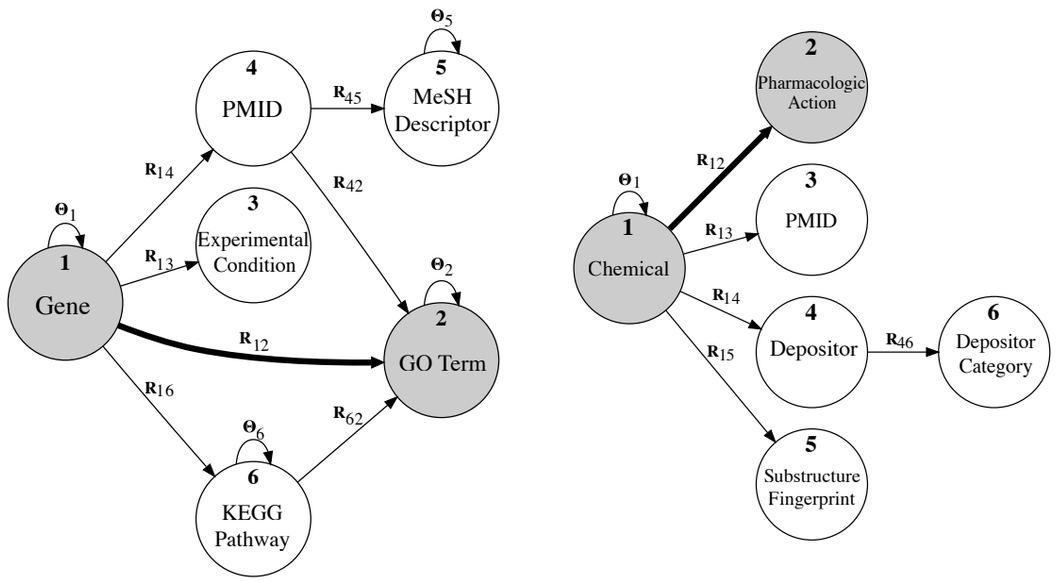




Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics



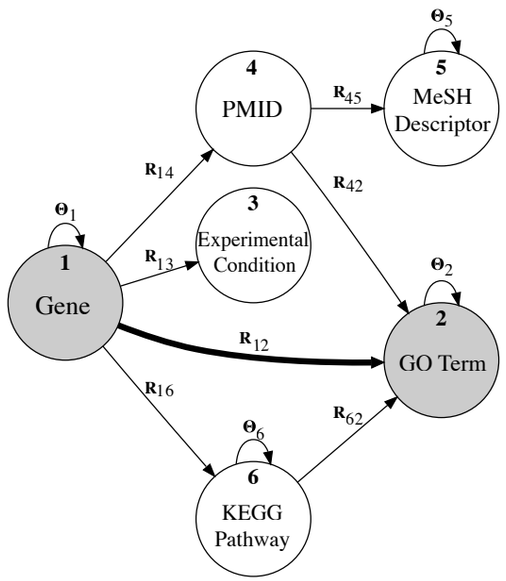
Mining disease associations

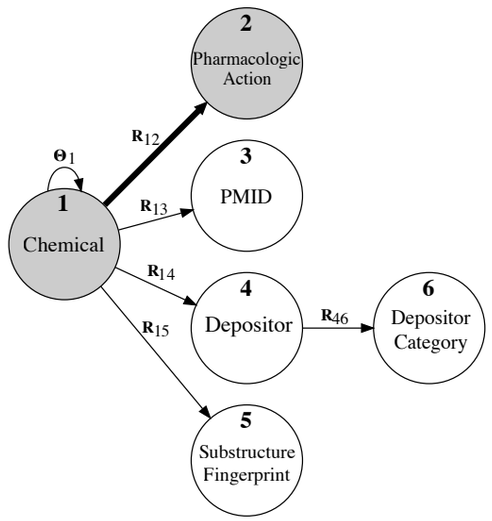
Žitnik *et al Scientific Reports* 2013

Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics





Mining disease associations

Žitnik *et al Scientific Reports* 2013

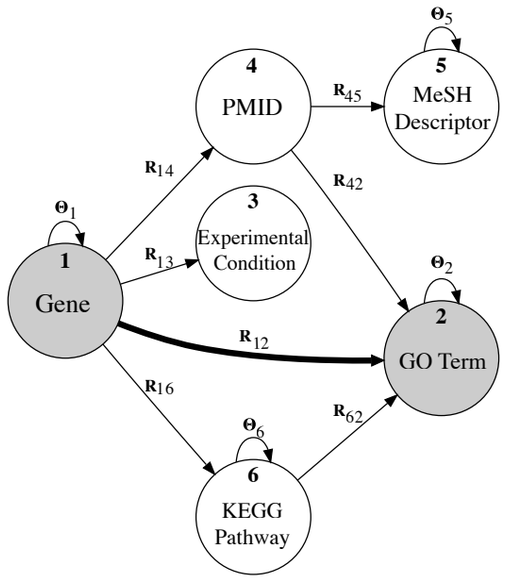
Predicting drug toxicity

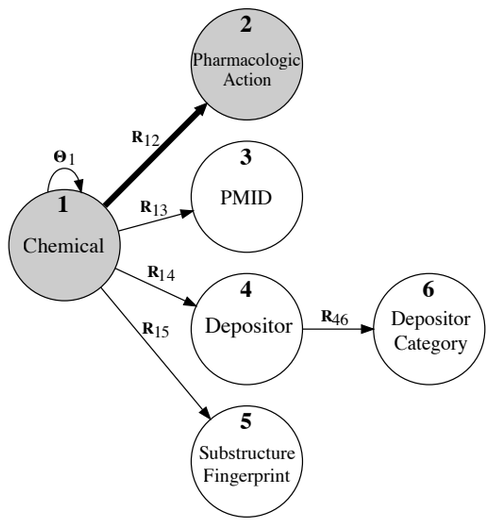
Žitnik & Zupan *Systems Biomedicine* 2014 (CAMDA Award)

Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics





Mining disease associations

Žitnik *et al Scientific Reports* 2013

Predicting drug toxicity

Žitnik & Zupan *Systems Biomedicine* 2014 (CAMDA Award)

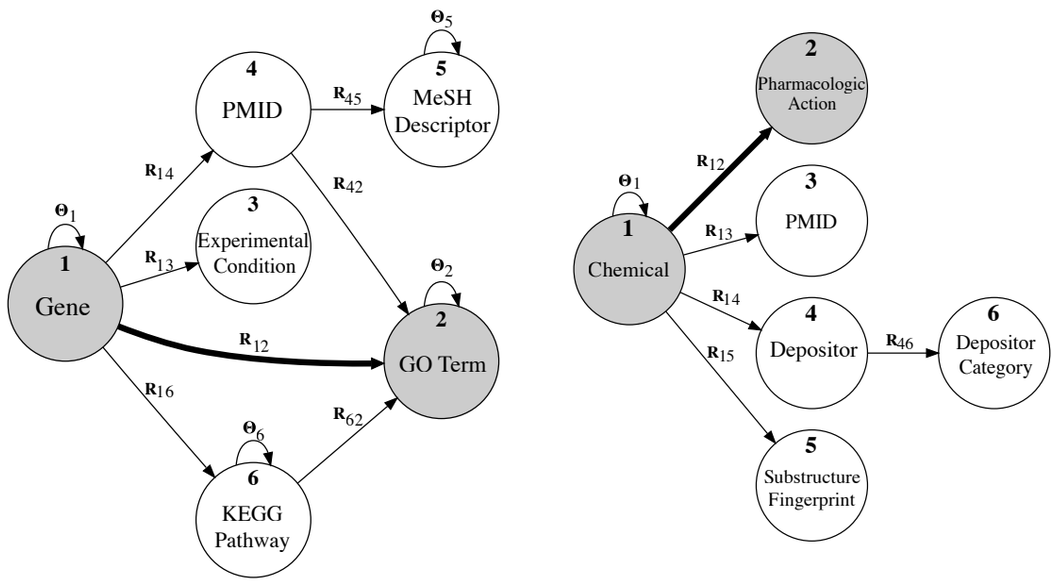
Predicting gene functions

Žitnik & Zupan In PSB 2014

Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

Žitnik & Zupan *IEEE TPAMI* 2015

#2: Functional Genomics



Mining disease associations

Žitnik *et al Scientific Reports* 2013

Predicting drug toxicity

Žitnik & Zupan *Systems Biomedicine* 2014 (CAMDA Award)

Predicting gene functions

Žitnik & Zupan In PSB 2014

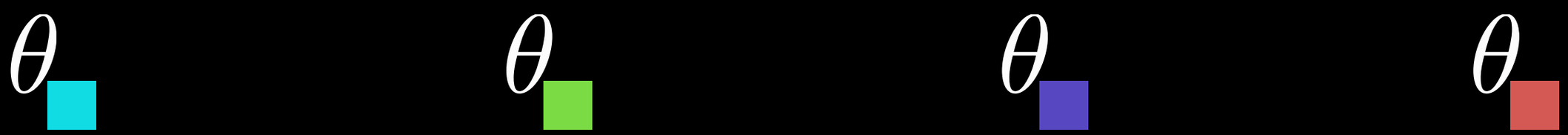
Predicting cancer survival

Žitnik & Zupan *Systems Biomedicine* 2015 (CAMDA Award)

Prediction task	DFMF		MKL		RF		tri-SPMF	
	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC
100 *D. discoideum* genes	0.799	0.801	0.781	0.788	0.761	0.785	0.731	0.724
1000 *D. discoideum* genes	0.826	0.823	0.787	0.798	0.767	0.788	0.756	0.741
Whole *D. discoideum* genome	0.831	0.849	0.800	0.821	0.782	0.801	0.778	0.787
Pharmacologic actions	0.663	0.834	0.639	0.811	0.643	0.819	0.641	0.810

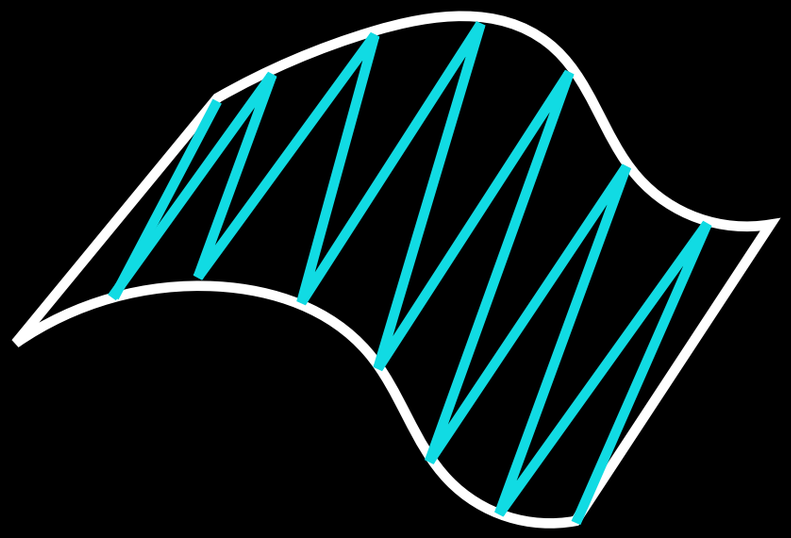
Žitnik & Zupan *IEEE TPAMI* 2015

Key Idea: Transfer of Knowledge



θ Model parameters

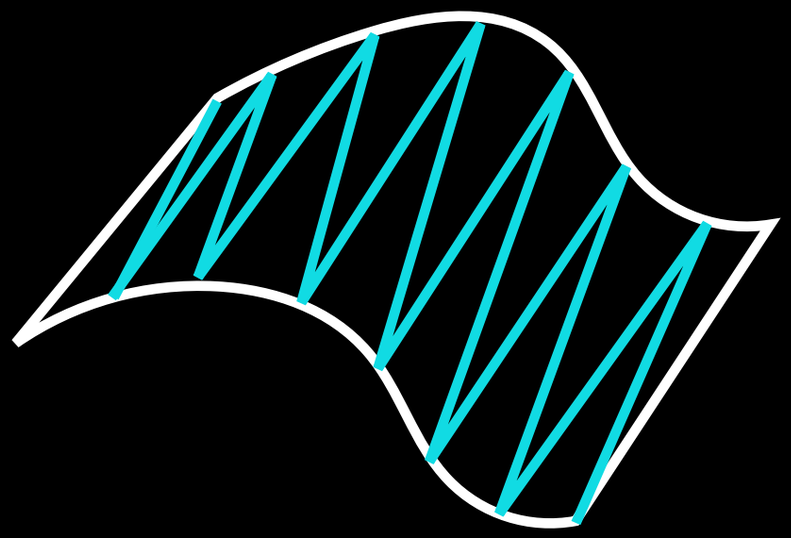
Key Idea: Transfer of Knowledge



θ θ θ θ

θ Model parameters

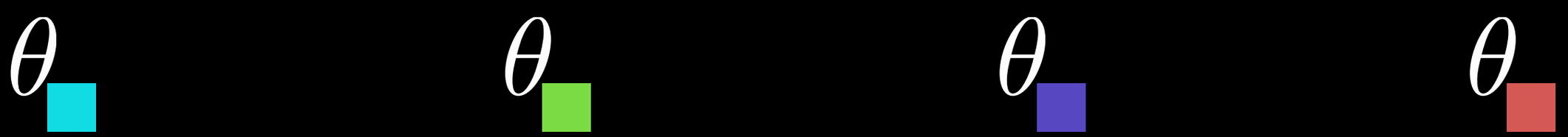
Key Idea: Transfer of Knowledge



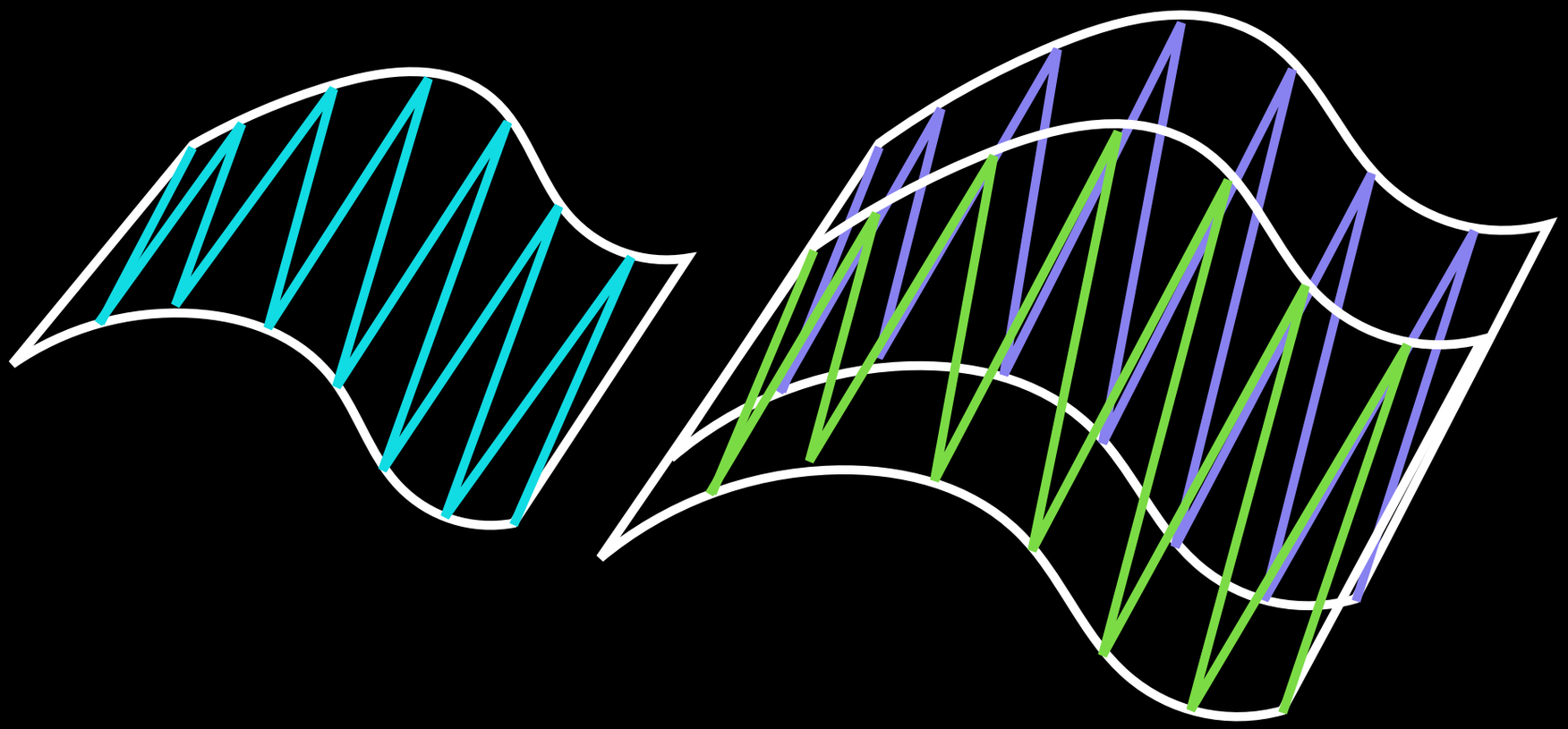
θ θ θ θ

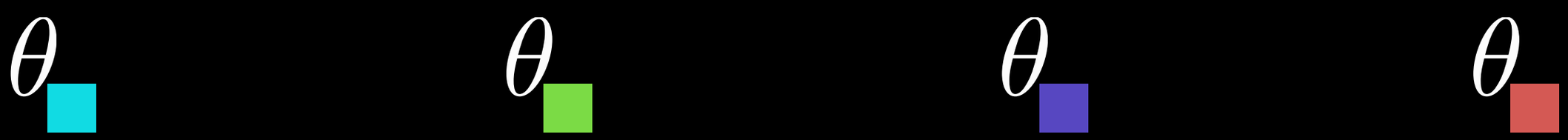
Objects of one type Data view θ Model parameters

Key Idea: Transfer of Knowledge

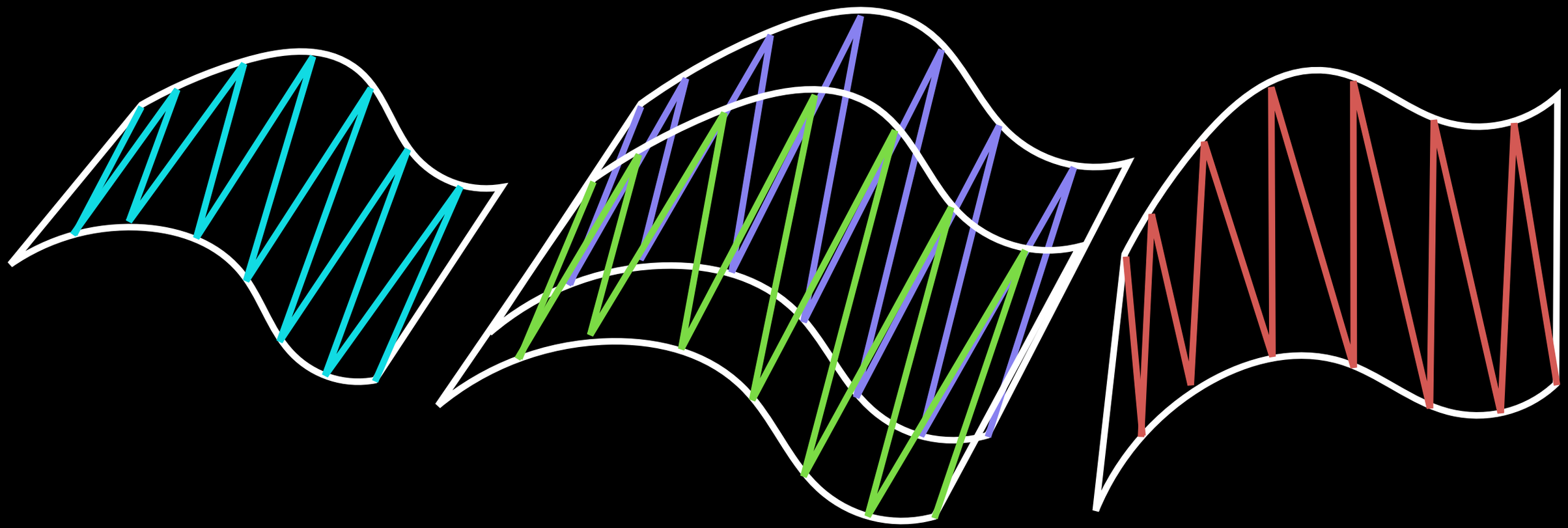


Key Idea: Transfer of Knowledge





Key Idea: Transfer of Knowledge

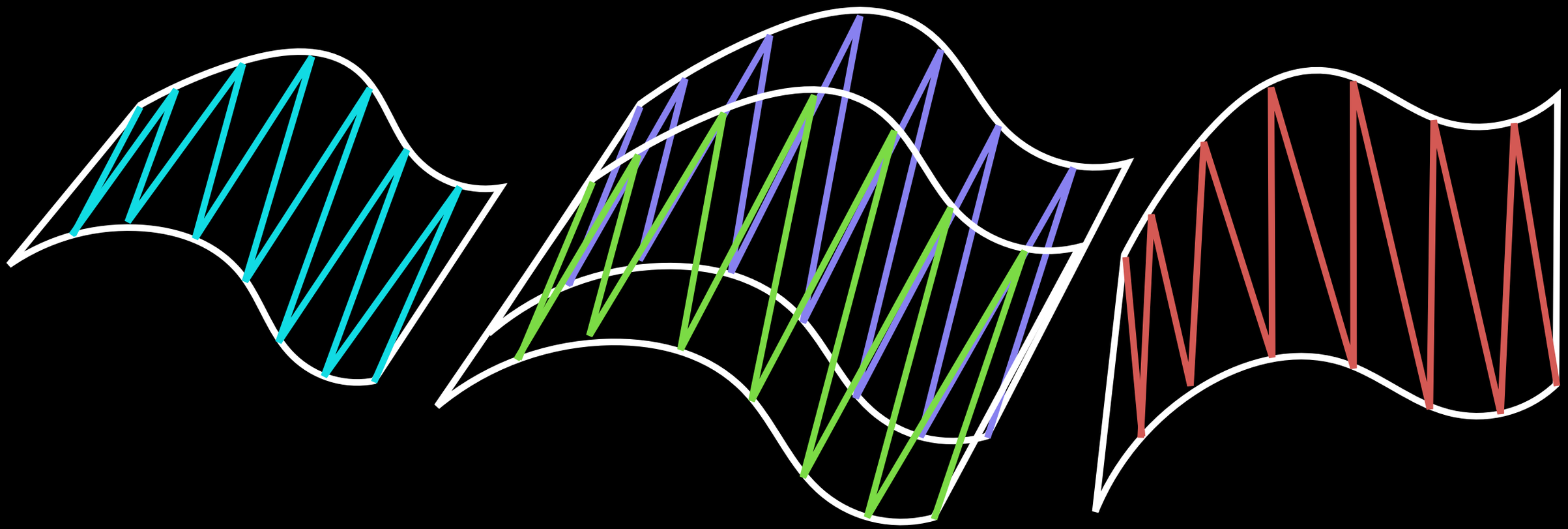


θ θ θ θ

Objects of one type Data view θ Model parameters

Key Idea: Transfer of Knowledge

Heterogeneous data domain space

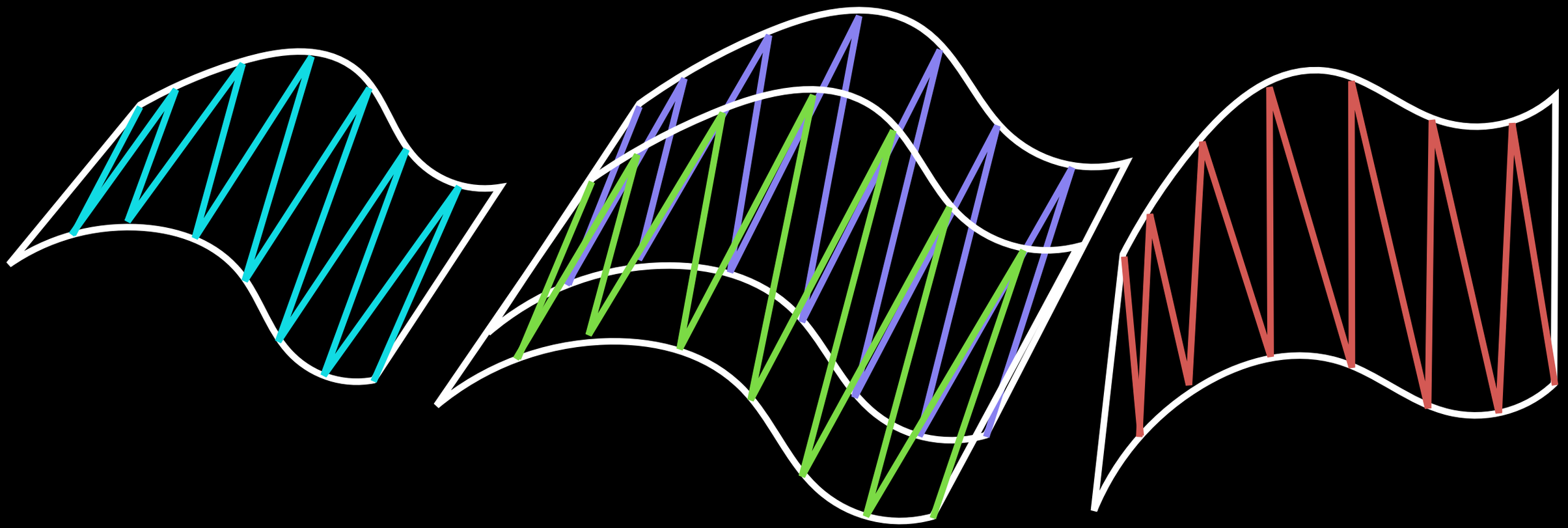


θ θ θ θ

Objects of one type Data view θ Model parameters

Key Idea: Transfer of Knowledge

Heterogeneous data domain space

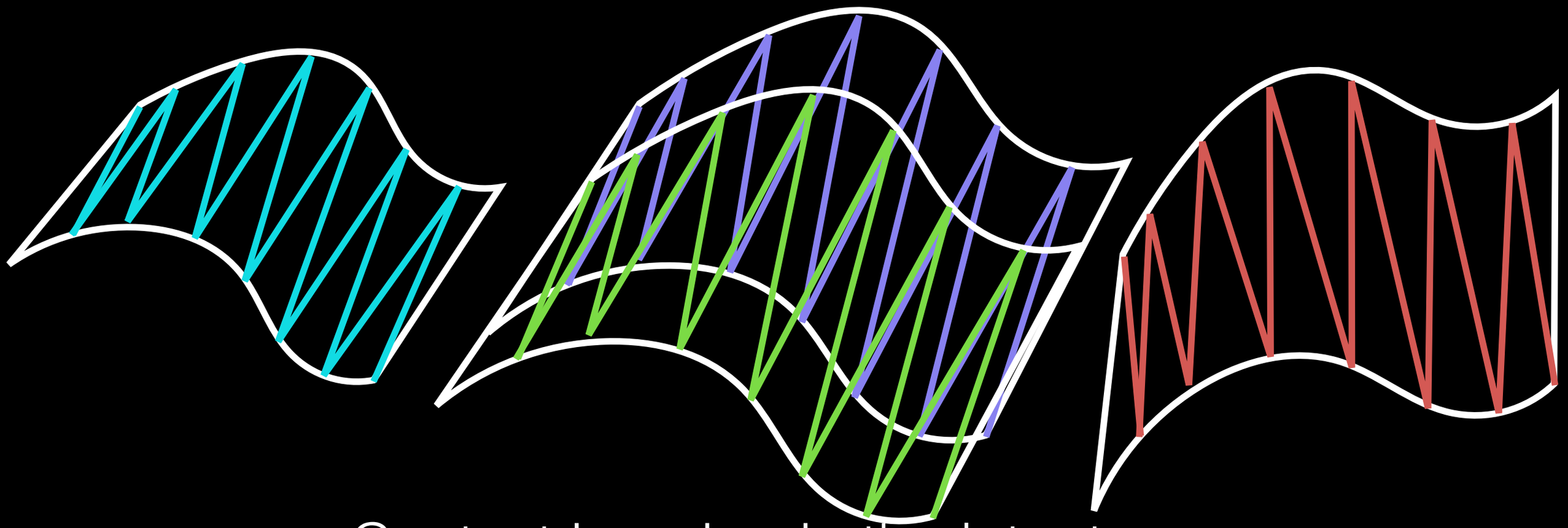


Objects of one type Data view θ Model parameters

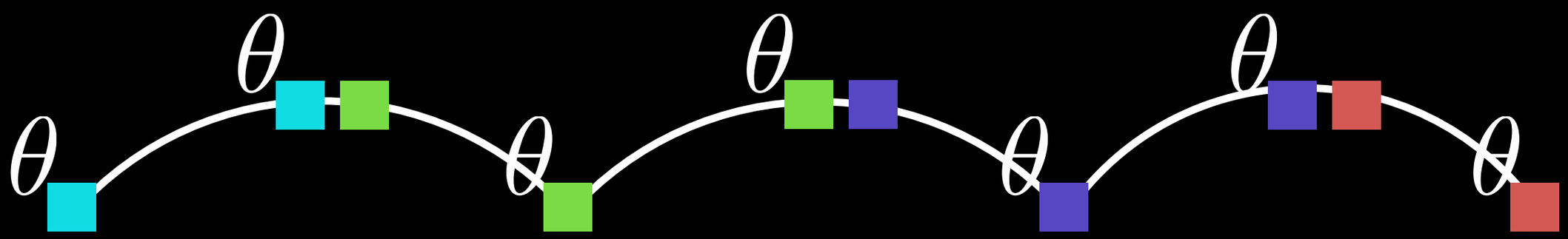
Marinka Zitnik - PhD Thesis

Key Idea: Transfer of Knowledge

Heterogeneous data domain space



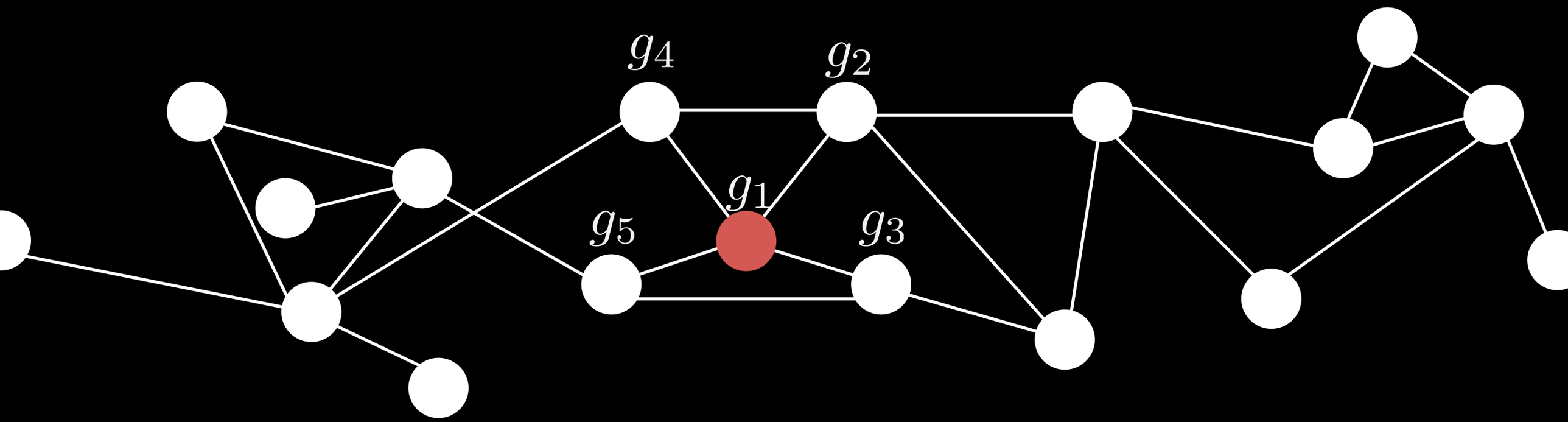
Context jumping in the latent space

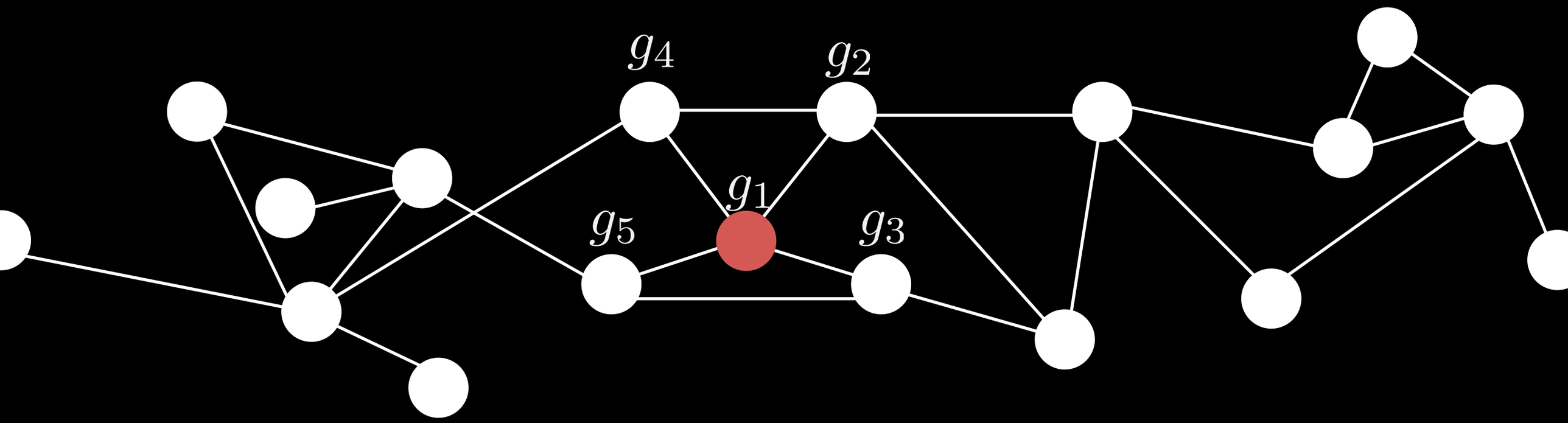


Objects of one type Data view θ Model parameters

Marinka Zitnik - PhD Thesis

Transfer of Knowledge:
Another Example
Network Inference from Mixed Data

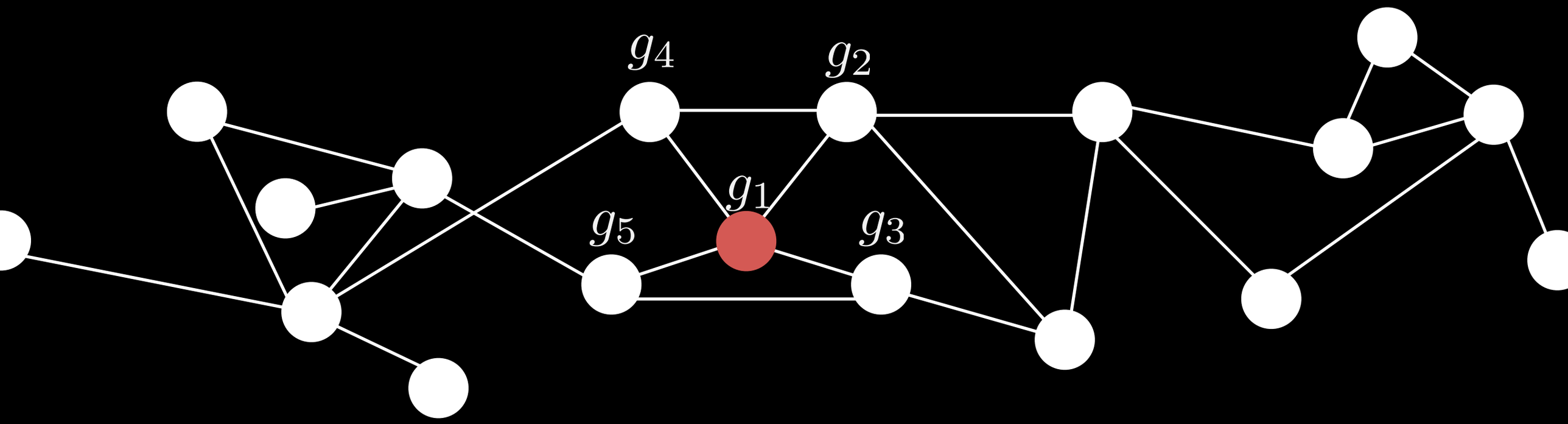

g_4
g_2
g_1
g_5
g_3



Direct inference

$$\mathcal{N}(g_1) = \{g_i \in V \setminus \{g_1\} : \mathrm{sim}(g_1, g_i) \geq T\}$$

threshold value



Direct inference

$$\mathcal{N}(g_1) = \{g_i \in V \setminus \{g_1\} : \mathrm{sim}(g_1, g_i) \geq T\}$$

threshold value

Model-based inference

$$g_1 = \theta_2 g_2 + \theta_3 g_3 + \theta_4 g_4 + \theta_5 g_5 + \cdots + \theta_n g_n$$

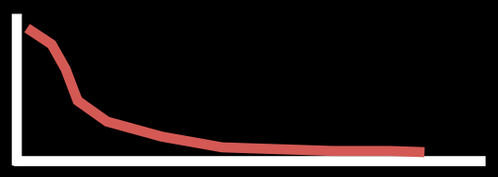
$$\mathcal{N}(g_1) = \{g_i \in V \setminus \{g_1\} : \theta_i \neq 0\}$$

model parameters

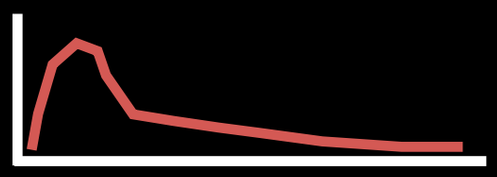
Mixed Data

Mixed Data

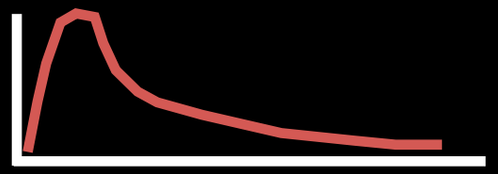
RNA-seq count data



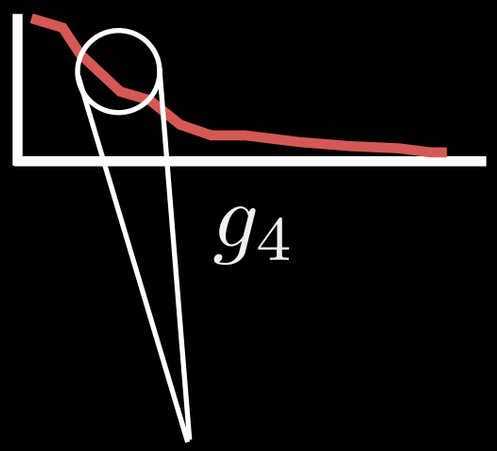
g_1



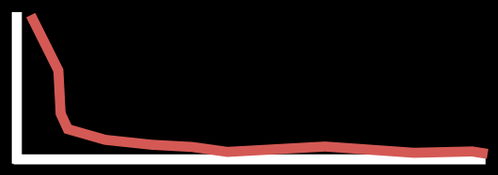
g_2



g_3



g_4

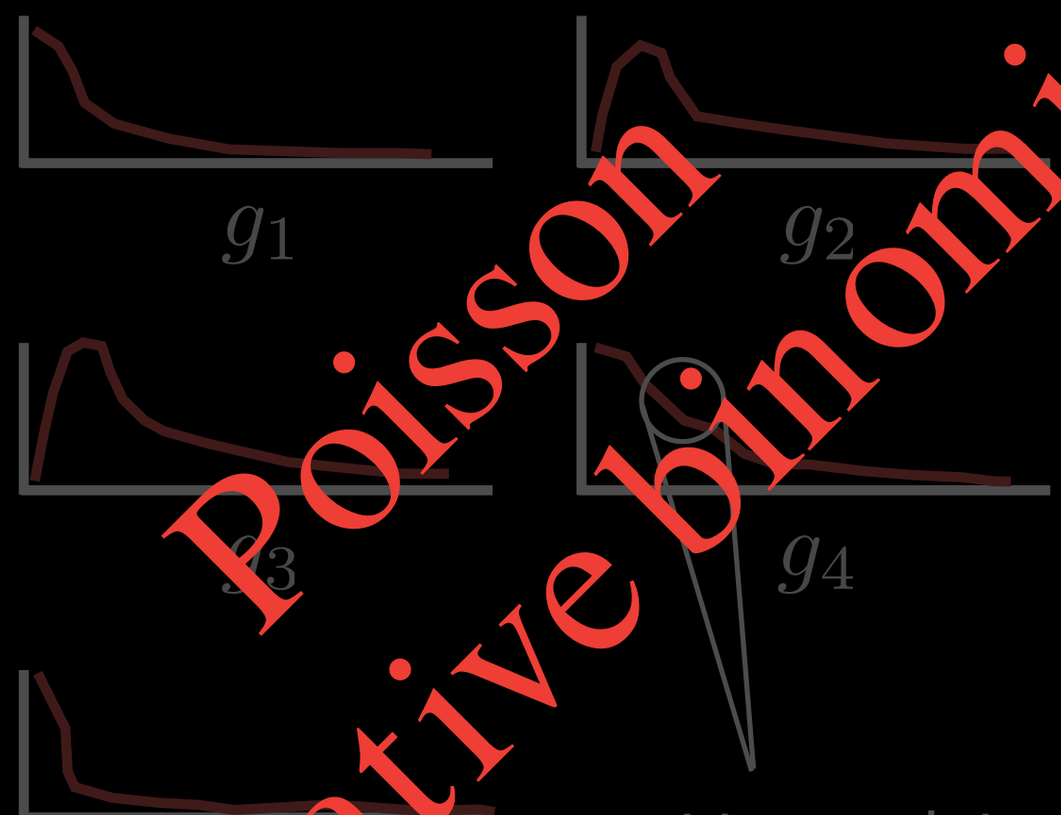


g_5

count transcripts mapped to genomic locations

Mixed Data

RNA-seq count data



count transcripts mapped to genomic locations

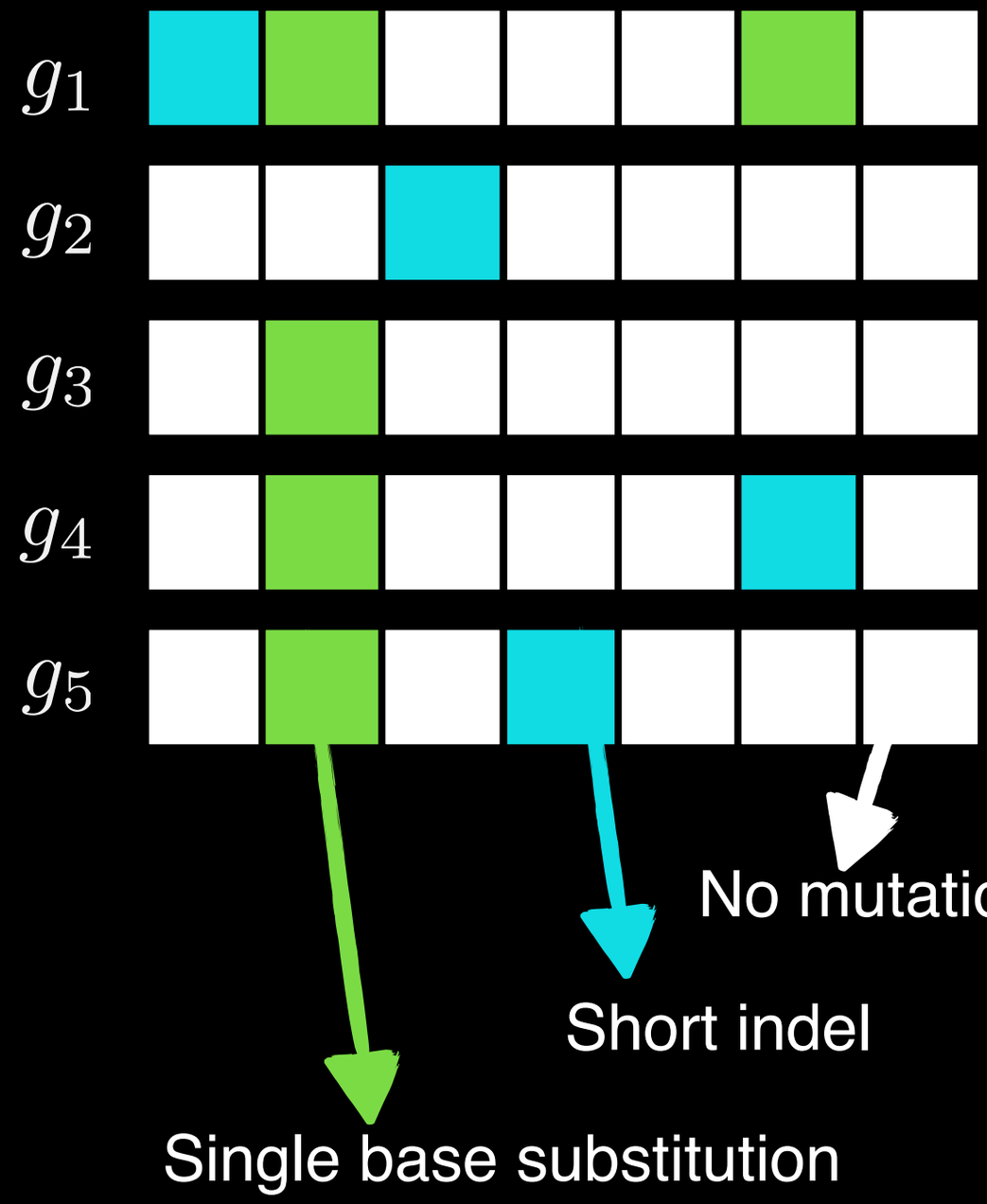
Mixed Data

RNA-seq count data

Poisson
Negative binomial

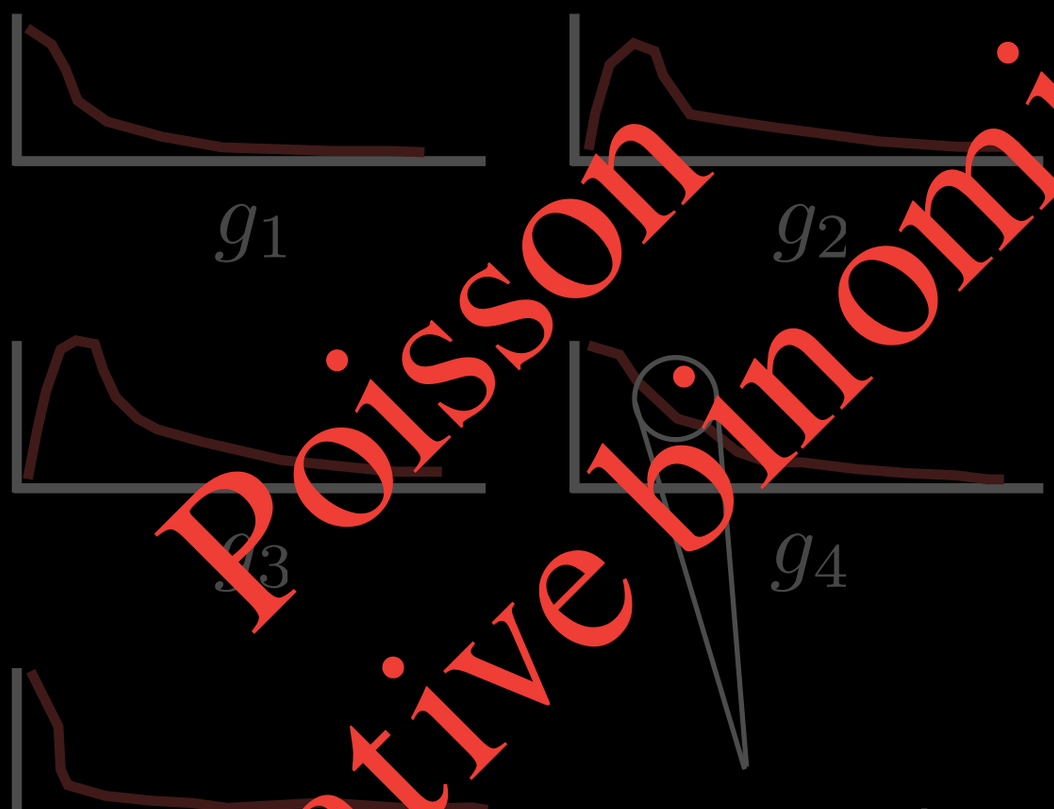
count transcripts mapped to genomic locations

Somatic mutations



Mixed Data

RNA-seq count data



Poisson

Negative binomial

count transcripts mapped to genomic locations

Somatic mutations

Multinomial

No mutation

Short indel

Single base substitution

Network Inference from Mixed Data

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

Network Inference from Mixed Data

$X = (X_1, X_2, \ldots, X_n), X_i$ is an object of interest

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

Network Inference from Mixed Data

$X = (X_1, X_2, \ldots, X_n), X_i$ is an object of interest

$$P_\Theta(X) \propto \exp\Big(\underbrace{\sum_{g \in V} \theta_g \phi_g(X_g)}_{\text{Nodes}} + \underbrace{\sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)}_{\text{Edges}}\Big)$$

Network Inference from Mixed Data

$X = (X_1, X_2, \ldots, X_n), X_i$ is an object of interest

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

Nodes

Edges

$\theta_g = \mathbf{U}_g$

Object weights

Network Inference from Mixed Data

$X = (X_1, X_2, \ldots, X_n), X_i$ is an object of interest

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

Nodes

Edges

$\theta_g = \mathbf{U}_g$

Object weights

$\theta_{gh} = \mathbf{U}_g^T \mathbf{W}^T \mathbf{W} \mathbf{U}_h$

Object-object interactions

Network Inference from Mixed Data

Objective function

Network Inference from Mixed Data

Objective function

$$\min_{\mathbf{U},\mathbf{W}_x,\mathbf{W}_y} \sum_{g \in V} \ell_{g;P_x}(\mathbf{U}, \mathbf{W}_x; \mathbf{X})$$

Data $\mathbf{X}$ following distribution P_x

Network Inference from Mixed Data

Objective function

$$\min_{\mathbf{U},\mathbf{W}_x,\mathbf{W}_y} \sum_{g\in V} \ell_{g;P_x}(\mathbf{U},\mathbf{W}_x;\mathbf{X}) + \ell_{g;P_y}(\mathbf{U},\mathbf{W}_y;\mathbf{Y}) + \text{reg. param.}$$

Data $\mathbf{X}$ following distribution P_x

Data $\mathbf{Y}$ following distribution P_y

Network Inference from Mixed Data

Objective function

Latent factor reuse

$$\min_{\mathbf{U},\mathbf{W}_x,\mathbf{W}_y} \sum_{g\in V} \ell_{g;P_x}(\mathbf{U},\mathbf{W}_x;\mathbf{X}) + \ell_{g;P_y}(\mathbf{U},\mathbf{W}_y;\mathbf{Y}) + \text{reg. param.}$$

Data $\mathbf{X}$ following distribution P_x

Data $\mathbf{Y}$ following distribution P_y

Network Inference from Mixed Data

$$\min_{\mathbf{U},\mathbf{W}_x,\mathbf{W}_y} \sum_{g\in V} \ell_{g;P_x}(\mathbf{U},\mathbf{W}_x;\mathbf{X}) + \ell_{g;P_y}(\mathbf{U},\mathbf{W}_y;\mathbf{Y}) + \text{reg. param.}$$

Network Inference from Mixed Data

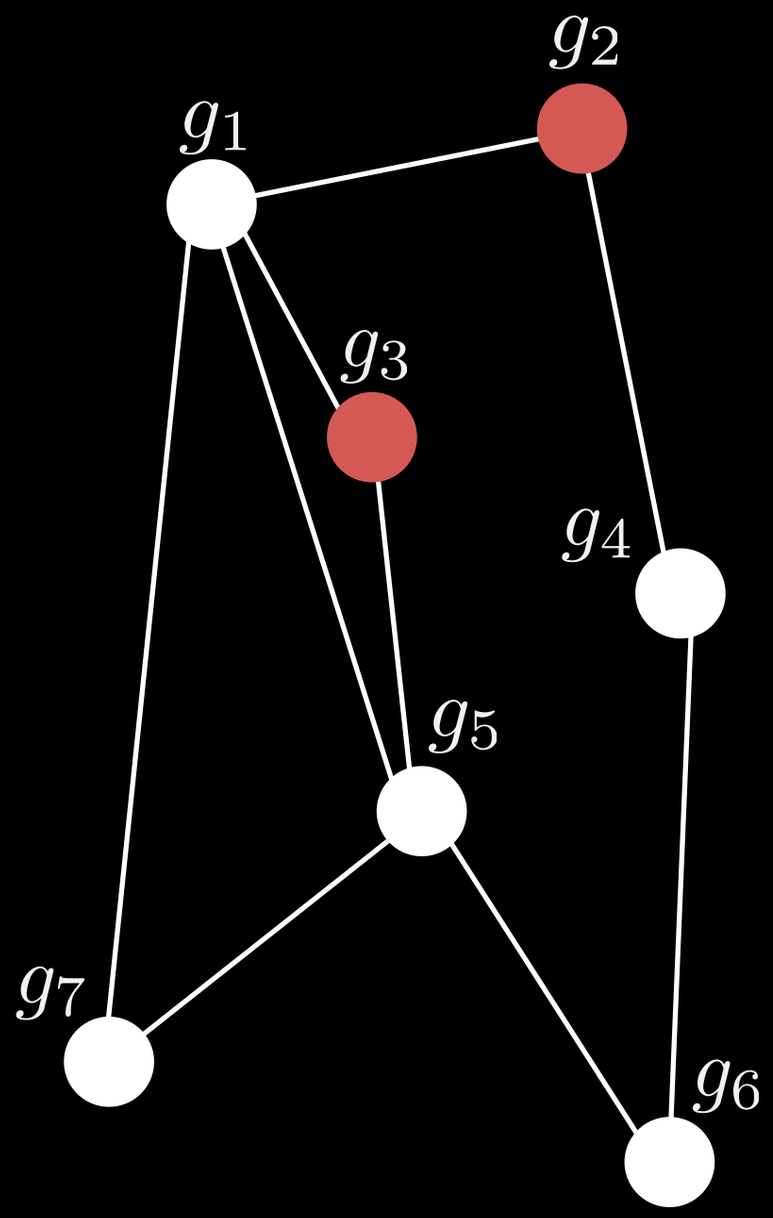
$$\min_{\mathbf{U},\mathbf{W}_x,\mathbf{W}_y} \sum_{g \in V} \ell_{g;P_x}(\mathbf{U},\mathbf{W}_x;\mathbf{X}) + \ell_{g;P_y}(\mathbf{U},\mathbf{W}_y;\mathbf{Y}) + \text{reg. param.}$$

$$\widehat{\mathcal{N}}(g) = \{h \in V \setminus \{g\} \ : \ \mathbf{U}_g^T \mathbf{W}_x^T \mathbf{W}_x \mathbf{U}_h \neq 0 \vee \mathbf{U}_g^T \mathbf{W}_y^T \mathbf{W}_y \mathbf{U}_h \neq 0\}$$

Data $\mathbf{X}$

Data $\mathbf{Y}$

Network Inference from Mixed Data

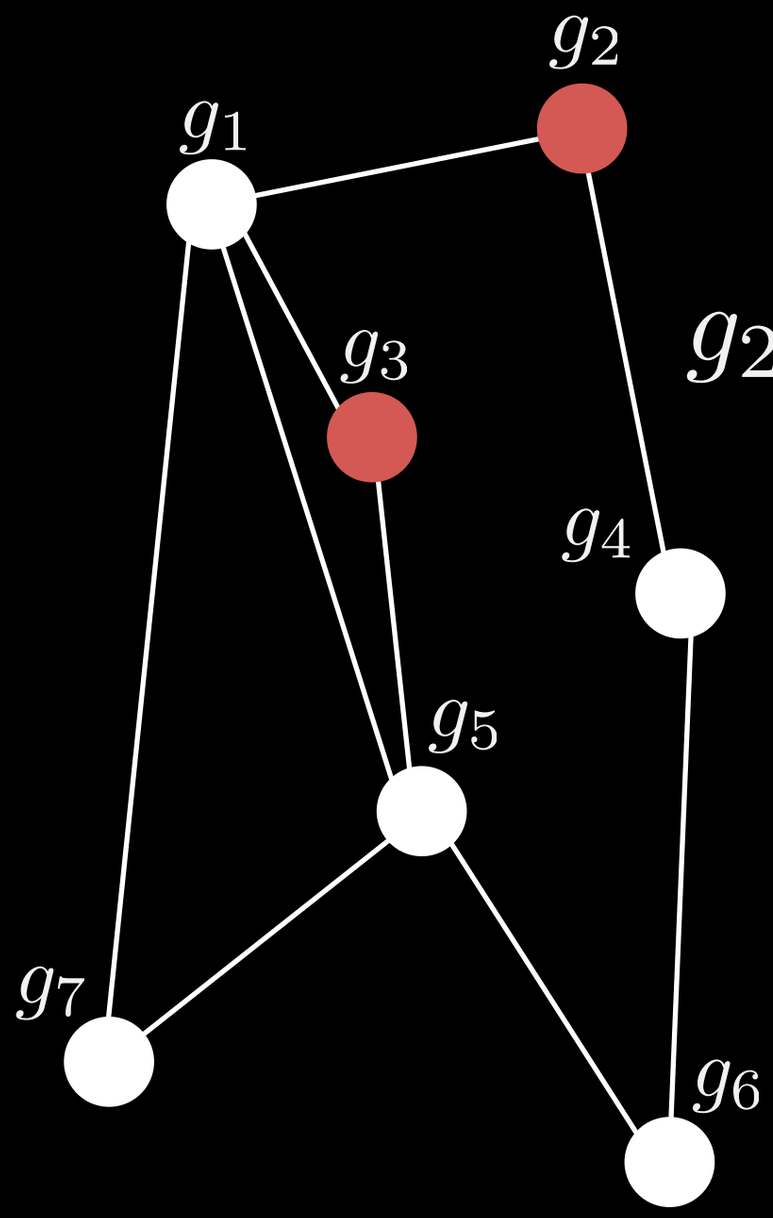


$$\widehat{\mathcal{N}}(g) = \{h \in V \setminus \{g\} \; : \; \mathbf{U}_g^T \mathbf{W}_x^T \mathbf{W}_x \mathbf{U}_h \neq 0 \vee \mathbf{U}_g^T \mathbf{W}_y^T \mathbf{W}_y \mathbf{U}_h \neq 0\}$$

Data $\mathbf{X}$

Data $\mathbf{Y}$

Network Inference from Mixed Data



$$g_2 \perp g_3 \mid \{g_1, g_4\}$$

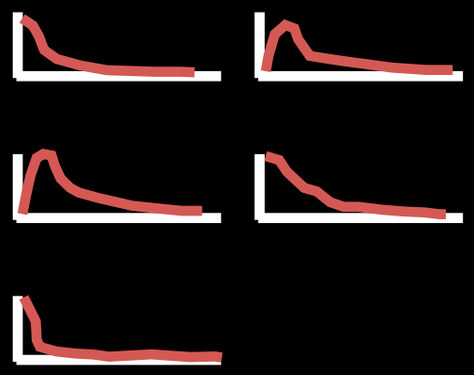
$$\widehat{\mathcal{N}}(g) = \{h \in V \setminus \{g\} \; : \; \mathbf{U}_g^T \mathbf{W}_x^T \mathbf{W}_x \mathbf{U}_h \neq 0 \vee \mathbf{U}_g^T \mathbf{W}_y^T \mathbf{W}_y \mathbf{U}_h \neq 0\}$$

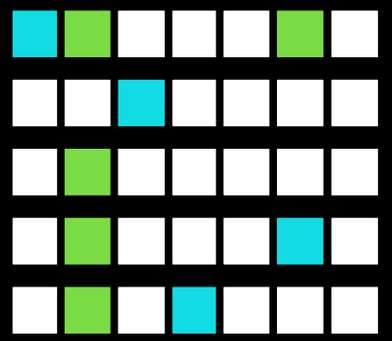
Data $\mathbf{X}$

Data $\mathbf{Y}$

FuseNet

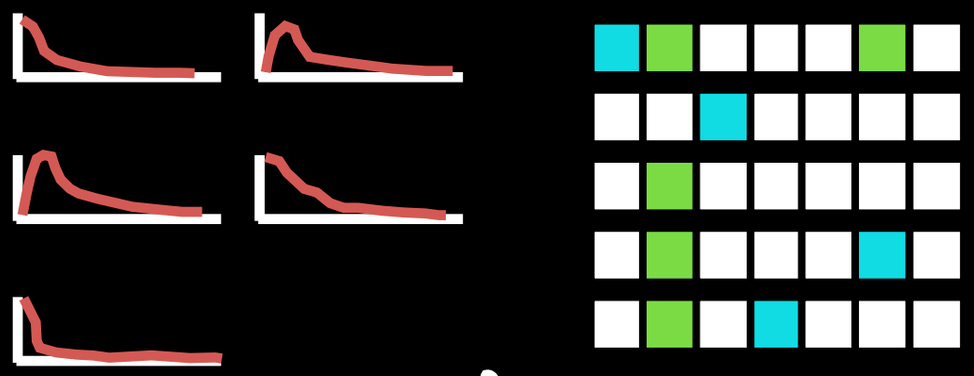
Data





FuseNet

Data

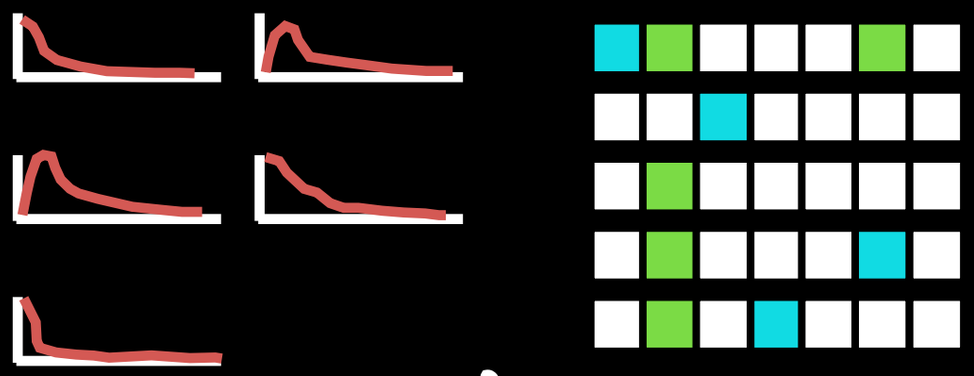


Model

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

FuseNet

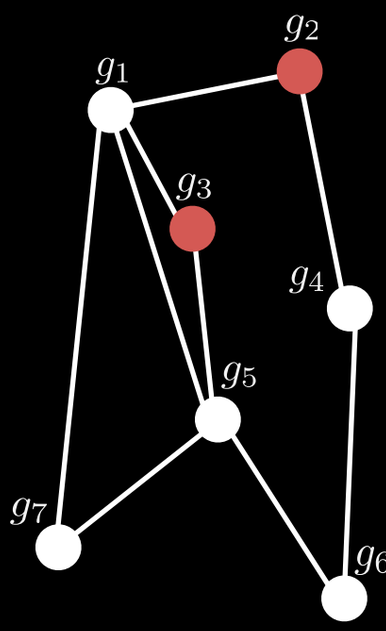
Data

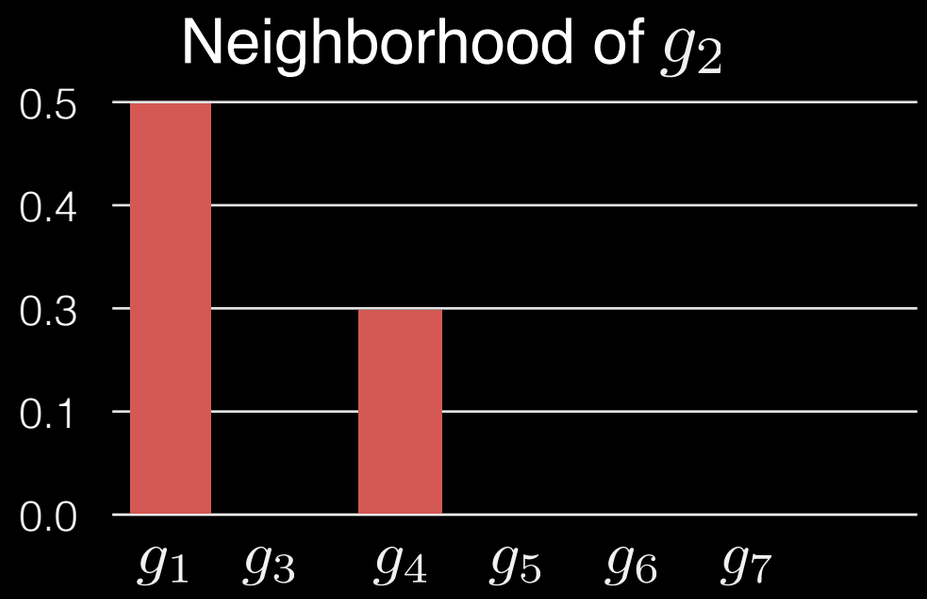


Model

$$P_\Theta(X) \propto \exp\Big(\sum_{g \in V} \theta_g \phi_g(X_g) + \sum_{(g,h) \in E} \theta_{gh} \phi_{gh}(X_g, X_h)\Big)$$

Network





Poisson Data

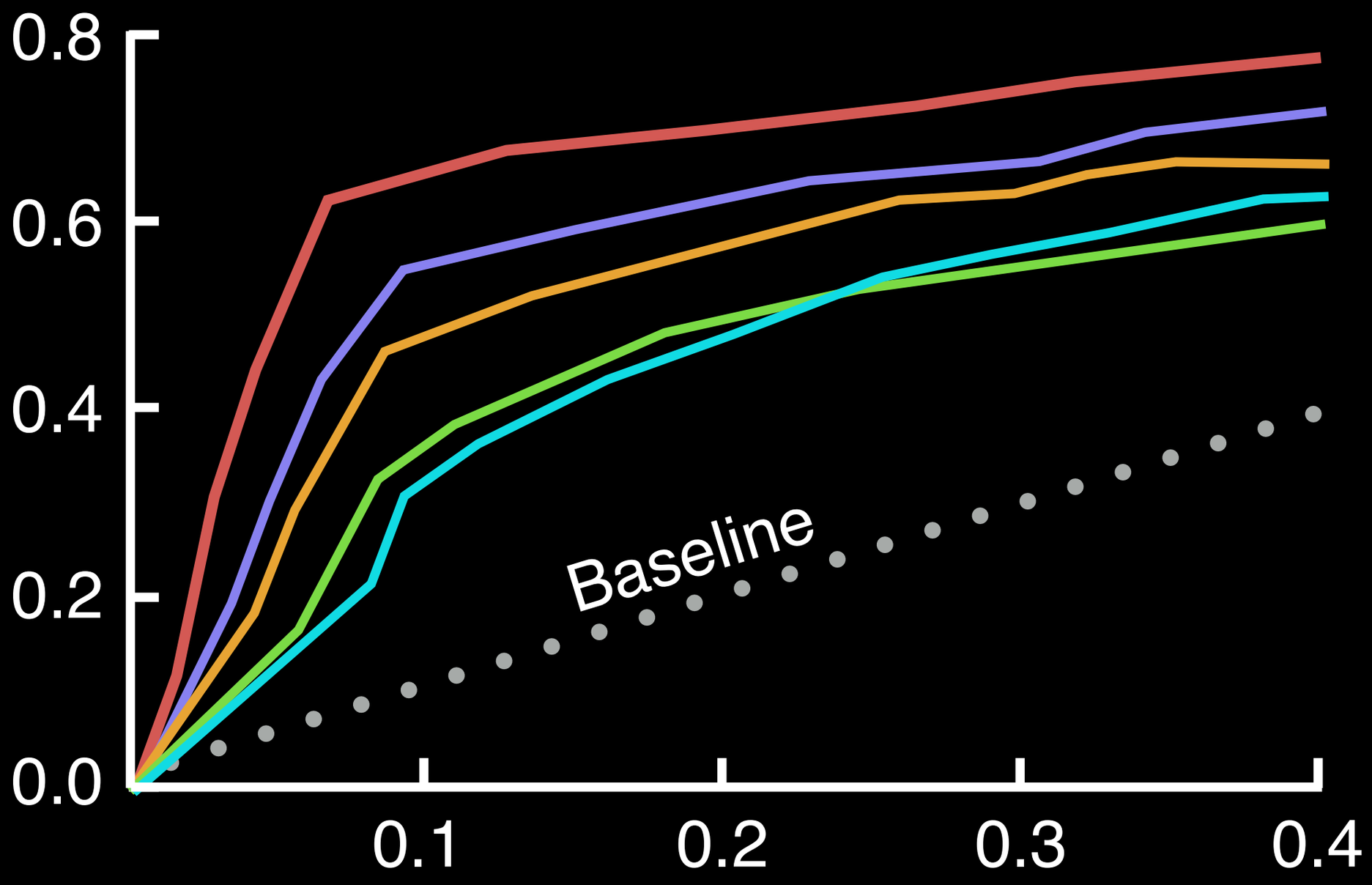
	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
Sample 1	452	872	495	348	2	297	348	982
Sample 2	482	124	726	132	872	29	77	144
Sample 3	719	2	198	376	193	287	173	346
Sample 4	56	24	99	0	239	928	376	660

Poisson Data

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
Sample 1	452	872	495	348	2	297	348	982
Sample 2	482	124	726	132	872	29	77	144
Sample 3	719	2	198	376	193	287	173	346
Sample 4	56	24	99	0	239	928	376	660

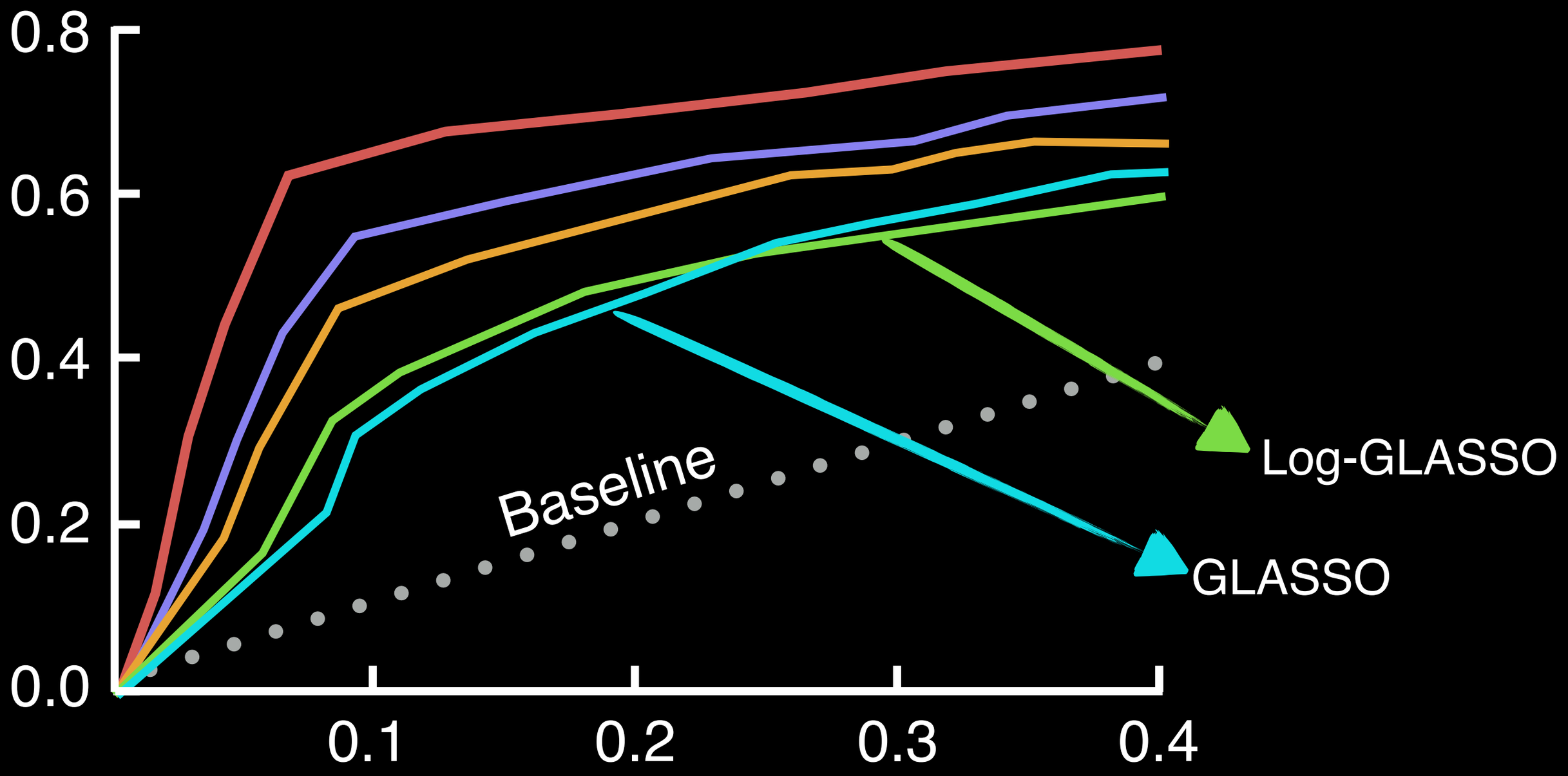
Poisson distribution

Recovery of Poisson Networks



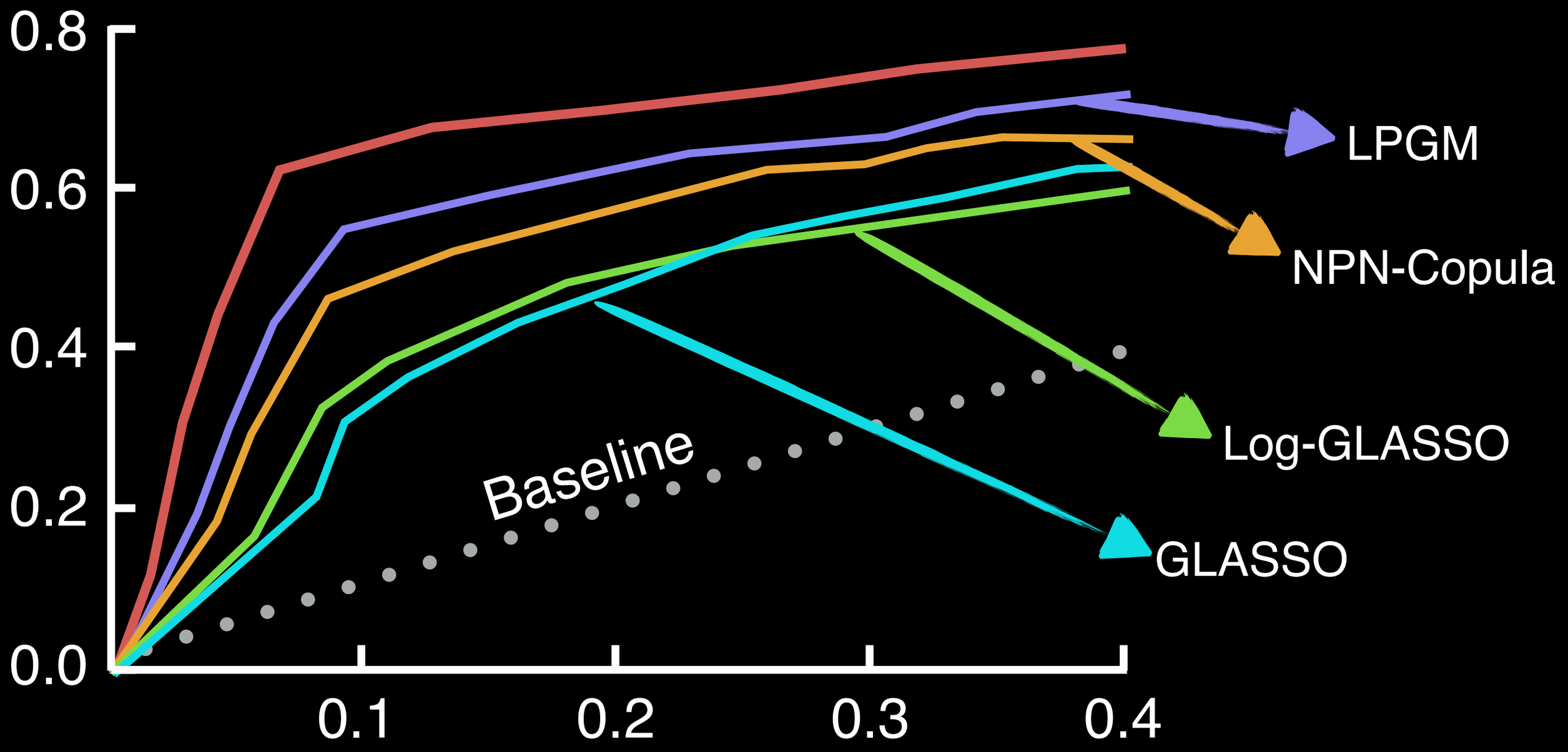
FuseNet - Our method; LPGM - Allen & Liu 2014; NPN-Copula - Liu et al. 2009;
log-GLASSO - Gallopin et al 2013; GLASSO - Friedman et al 2007

Recovery of Poisson Networks



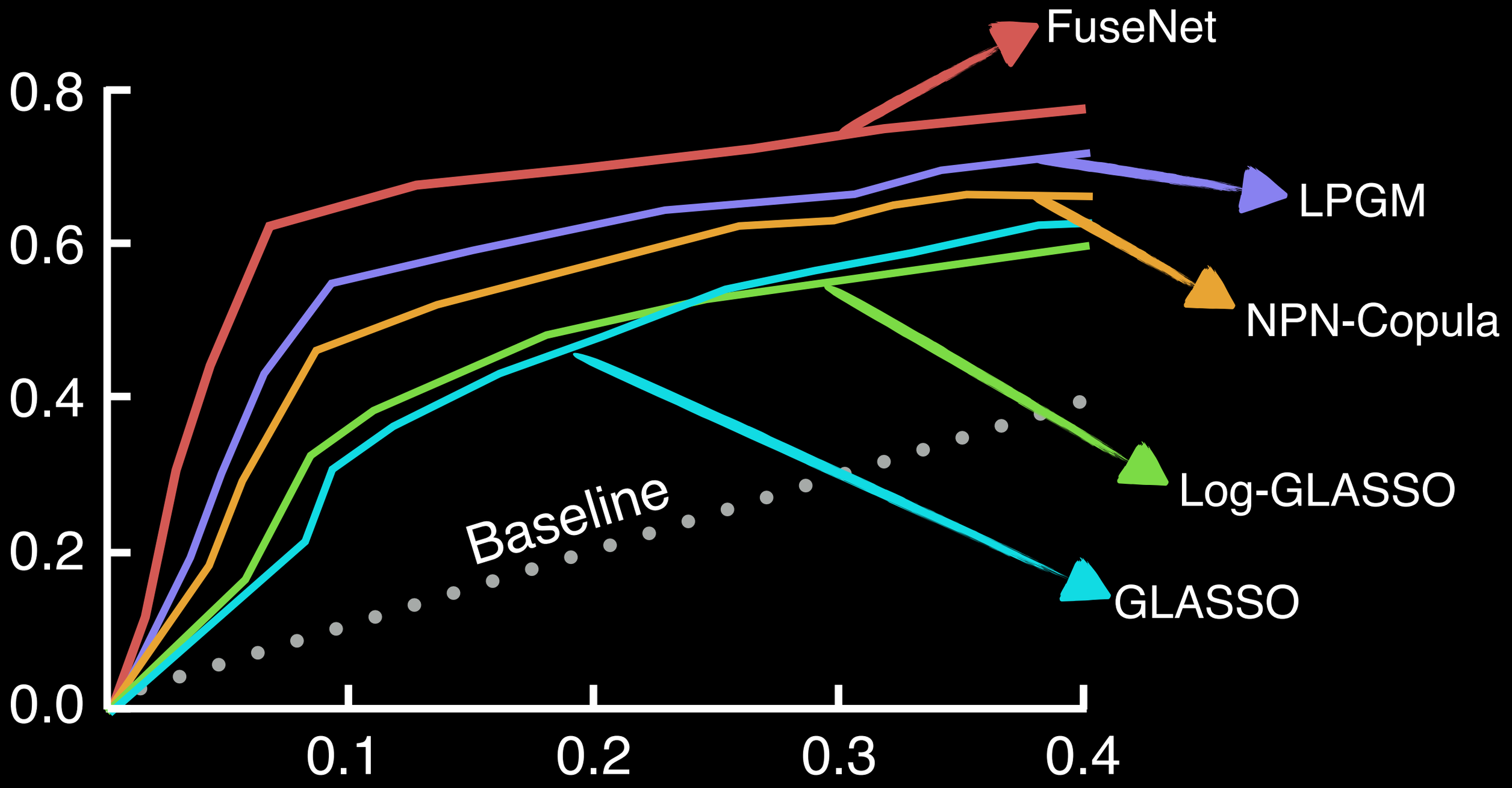
FuseNet - Our method; LPGM - Allen & Liu 2014; NPN-Copula - Liu et al. 2009;
log-GLASSO - Gallopin et al 2013; GLASSO - Friedman et al 2007

Recovery of Poisson Networks



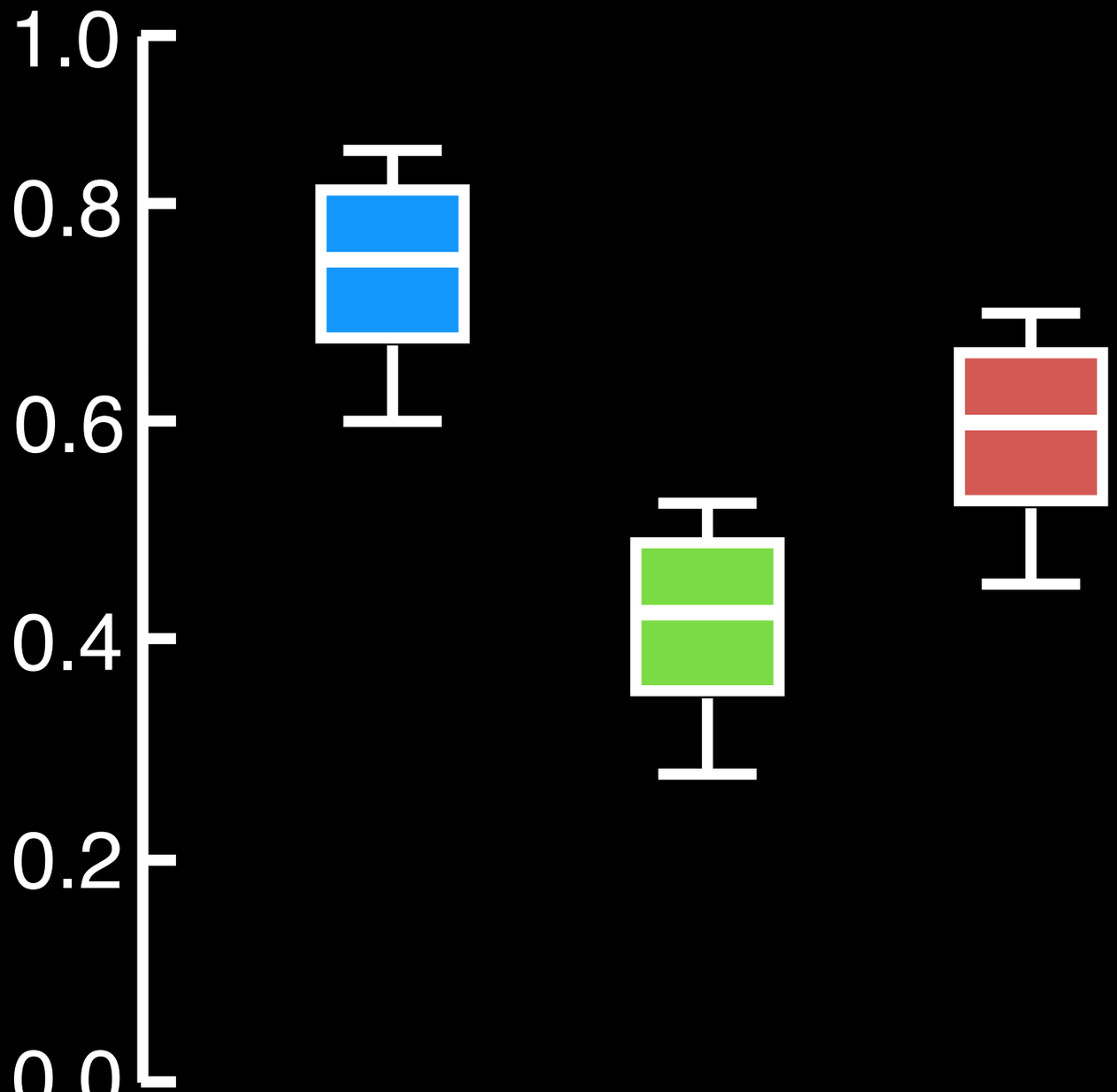
FuseNet - Our method; LPGM - Allen & Liu 2014; NPN-Copula - Liu et al. 2009;
log-GLASSO - Gallopin et al 2013; GLASSO - Friedman et al 2007

Recovery of Poisson Networks



FuseNet - Our method; LPGM - Allen & Liu 2014; NPN-Copula - Liu et al. 2009;
log-GLASSO - Gallopin et al 2013; GLASSO - Friedman et al 2007

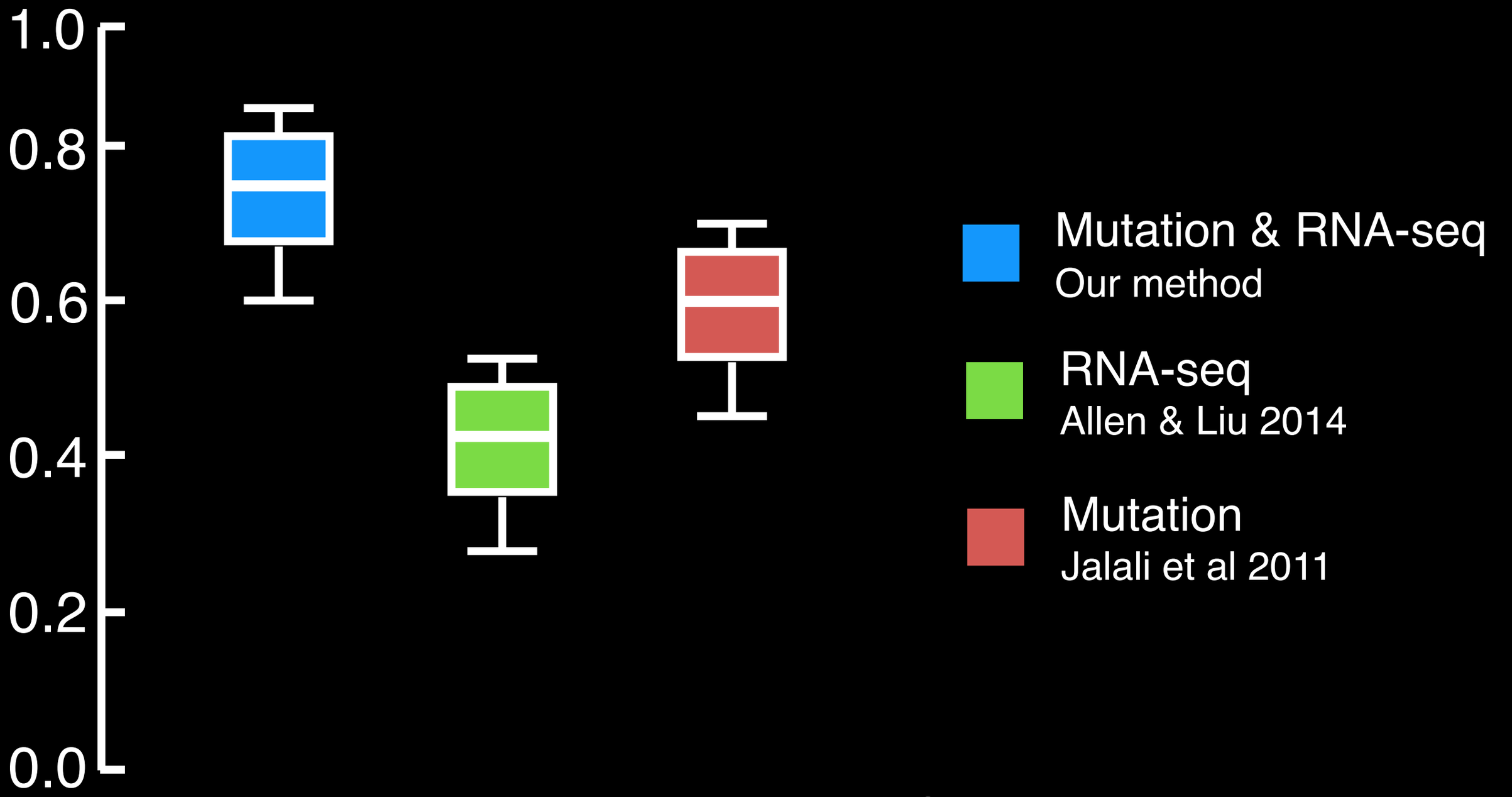
Functional Content of Inferred Cancer Networks



Higher score indicates a more informative network

Data from International Cancer Genome Consortium, BRCA

Functional Content of Inferred Cancer Networks



Higher score indicates a more informative network

Data from International Cancer Genome Consortium, BRCA

Summary of Contributions

Relation Heterogeneity

Markov network inference for mixed data

Epistasis network inference

Collective pairwise classification for multi-way data

Z & Z. *JMLR* 2012;
Z & Z. *Bioinformatics* 2014 (in ISMB 2014);
Z & Z. *Bioinformatics* 2015 (in ISMB 2015);
Z & Z. In PSB 2016

Relation Heterogeneity

Markov network inference for mixed data

Epistasis network inference

Collective pairwise classification for multi-way data

Z & Z. *JMLR* 2012;
Z & Z. *Bioinformatics* 2014 (in ISMB 2014);
Z & Z. *Bioinformatics* 2015 (in ISMB 2015);
Z & Z. In PSB 2016

Object Heterogeneity

Latent profile chaining

Z et al. PLOS Comp Bio 2015

Relation Heterogeneity

Markov network inference for mixed data

Epistasis network inference

Collective pairwise classification for multi-way data

Z & Z. *JMLR* 2012;
Z & Z. *Bioinformatics* 2014 (in ISMB 2014);
Z & Z. *Bioinformatics* 2015 (in ISMB 2015);
Z & Z. In PSB 2016

Object Heterogeneity

Latent profile chaining

Z et al. PLOS Comp Bio 2015

Dual Heterogeneity

Network guided matrix completion

Survival regression by data fusion

Z & Z. *Systems Biomedicine* 2015;
Z & Z. In RECOMB 2014;
Z & Z. *Journal of Comp Bio* 2015

Relation Heterogeneity

Markov network inference for mixed data

Epistasis network inference

Collective pairwise classification for multi-way data

Z & Z. *JMLR* 2012;
Z & Z. *Bioinformatics* 2014 (in ISMB 2014);
Z & Z. *Bioinformatics* 2015 (in ISMB 2015);
Z & Z. In PSB 2016

Object Heterogeneity

Latent profile chaining

Z et al. PLOS Comp Bio 2015

Dual Heterogeneity

Network guided matrix completion

Survival regression by data fusion

Z & Z. *Systems Biomedicine* 2015;
Z & Z. In RECOMB 2014;
Z & Z. *Journal of Comp Bio* 2015

Triple Heterogeneity

collective matrix factorization

Z et al. Scientific Reports 2013;
Z & Z. *Systems Biomedicine* 2014;
Z & Z. In PSB 2014;
Z & Z. *IEEE TPAMI* 2015;

Relation Heterogeneity

Markov network inference for mixed data

Epistasis network inference

Collective pairwise classification for multi-way data

Z & Z. *JMLR* 2012;
Z & Z. *Bioinformatics* 2014 (in ISMB 2014);
Z & Z. *Bioinformatics* 2015 (in ISMB 2015);
Z & Z. In PSB 2016

Object Heterogeneity

Latent profile chaining

Z et al. PLOS Comp Bio 2015

Dual Heterogeneity

Network guided matrix completion

Survival regression by data fusion

Z & Z. *Systems Biomedicine* 2015;
Z & Z. In RECOMB 2014;
Z & Z. *Journal of Comp Bio* 2015

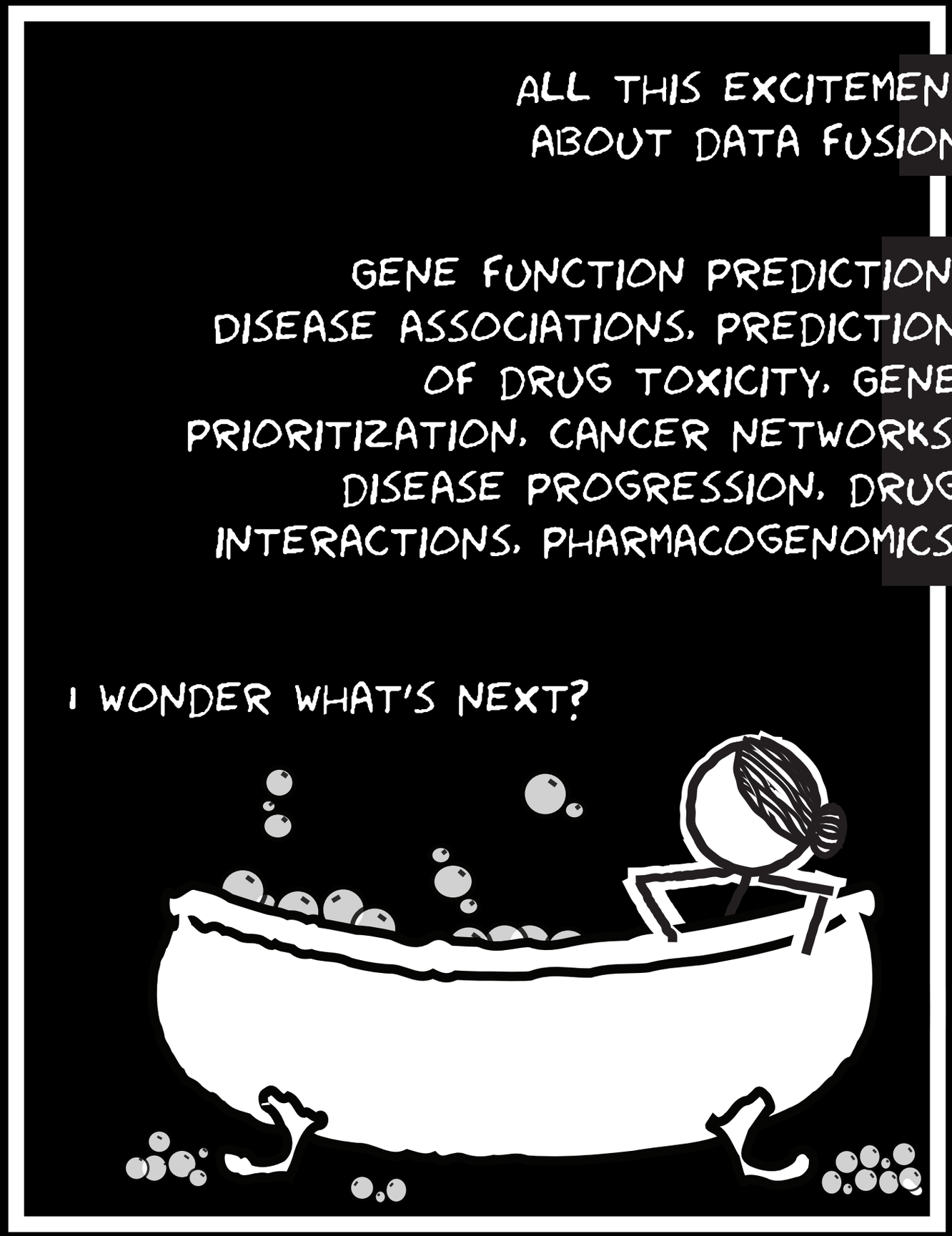
Triple Heterogeneity

collective matrix factorization

Z et al. Scientific Reports 2013;
Z & Z. *Systems Biomedicine* 2014;
Z & Z. In PSB 2014;
Z & Z. *IEEE TPAMI* 2015;

Exploring Heterogeneity

Sensitivity estimation using Frechet derivatives



Best poster awards at BC^2 2015 (Basel, Switzerland); RECOMB 2014 (Pittsburgh, PA, USA)

Marinka Zitnik - PhD Thesis

biolab

Blaz Zupan

Stanford University

Jure Leskovec

Adam Kuspa
Edward Nam
Chris Dinh

IJS

Uroš Petrovic
Petra Kaferle

Karolinska Institutet

SciLifeLab

Thomas Helleday
Jordi C. Puigvert

Baylor College of Medicine

Charles Boone
Mojca M. Usaj

UNIVERSITY OF TORONTO

Natasa Przulj
Vuk Janjic

Imperial College London

Gad Shaulsky
Rafael Rosengarten
Mariko Kurasawa
Balaji Santhanam

Marinka Zitnik - PhD Thesis