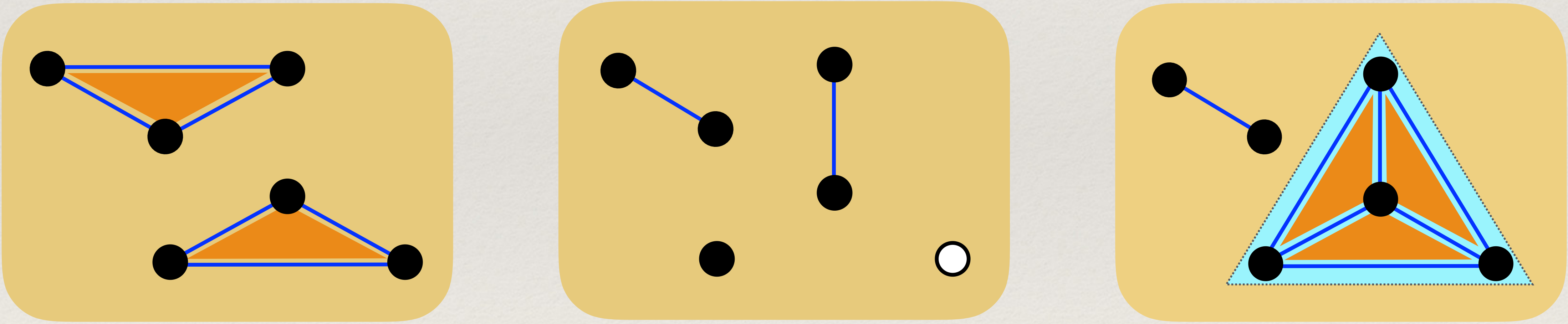
Beyond Pairwise Models for Binary Data

Model Selection with Minimally Complex Model



Clélia de Mulatier, Paolo P. Mazza, Matteo Marsili

APS March Meeting 2020

Modeling Binary Data

Context

Spin Models

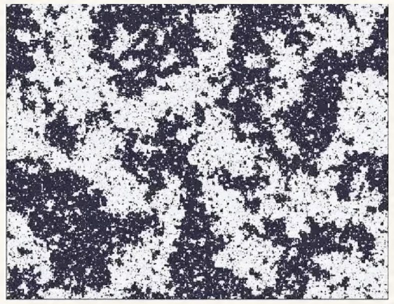
Modeling Data

Model

?

Inverse Problem

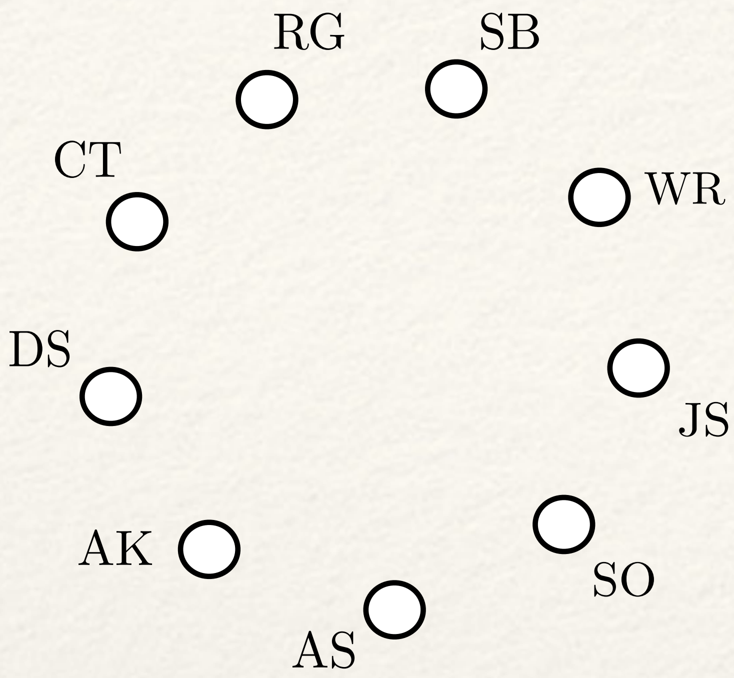
States, observables, properties



Noisy Data: find the model that best captures the patterns hidden within the data…

Modeling Binary Data… with Pairwise Spin Models

How does it work?



Statistical Mechanics of the US Supreme Court

Edward D. Lee, Chase P. Broedersz & William Bialek

Journal of Statistical Physics **160**, 275–301(2015) | Cite this article

2nd Rehnquist Court
(1994-2005)

9 justices, 895 votes

Conservative (1) or Liberal (-1)

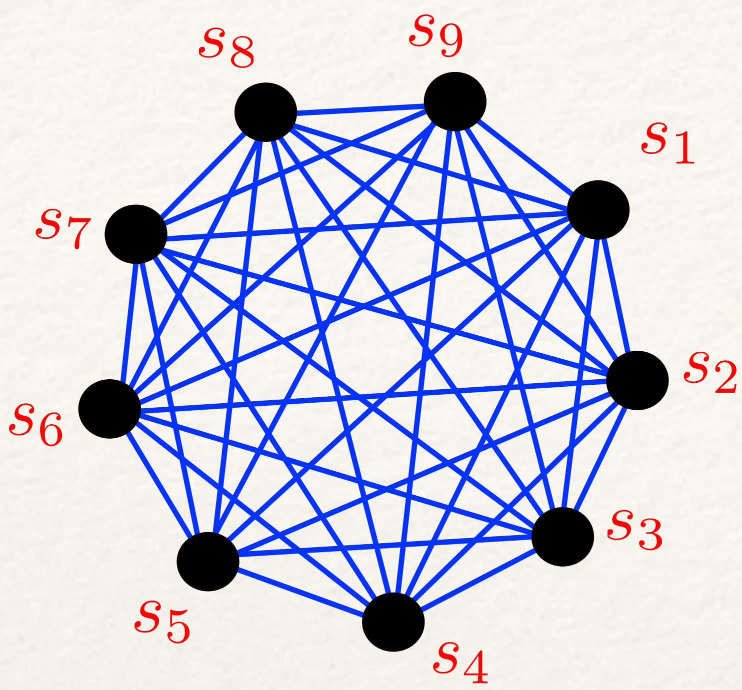
> $s_i \in \{-1, +1\}$ ⟶ ***Spins***

> System is stationary

> Each vote is independently sampled from an underlying probability distribution

Modeling Binary Data… with Pairwise Spin Models

How does it work?



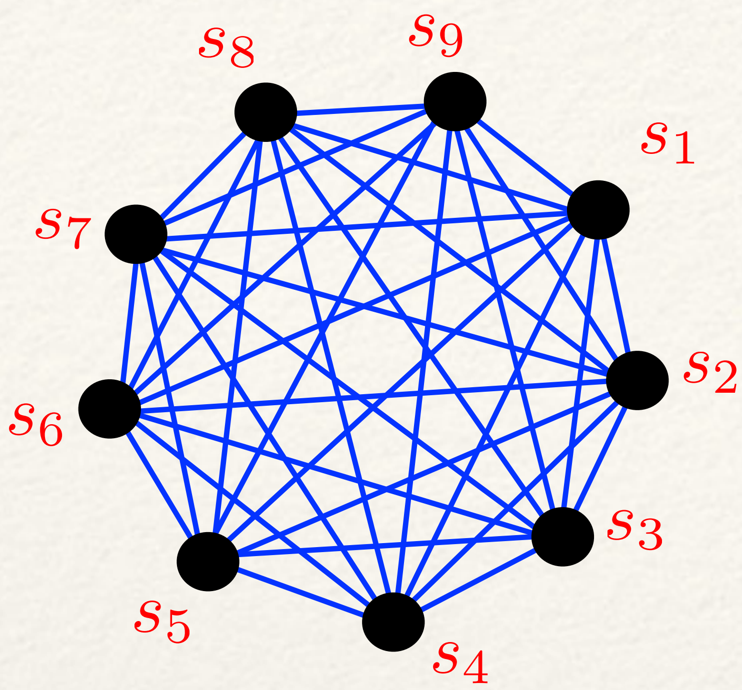
> Underlying distribution has a form:

$$P(\boldsymbol{s}\,|\,\mathcal{M},\boldsymbol{g}) = \frac{1}{Z_{\mathcal{M}}(\boldsymbol{g})}\,\exp\left(\sum_{i\in\mathcal{M}} h_i s_i + \sum_{pair(i,j)\in\mathcal{M}} J_{ij} s_i s_j\right)$$

Parameters to fit

Modeling Binary Data… with Pairwise Spin Models

How does it work?



> Underlying distribution has a form:

$$P(\boldsymbol{s}\,|\,\mathcal{M},\boldsymbol{g}) = \frac{1}{Z_{\mathcal{M}}(\boldsymbol{g})}\,\exp\left(\sum_{i\in\mathcal{M}} h_i s_i + \sum_{pair(i,j)\in\mathcal{M}} J_{ij}\, s_i s_j\right)$$

Parameters to fit

Infer the parameters:

$$\boldsymbol{g}^* = \underset{g}{\mathrm{argmax}}\; P(\hat{s}\,|\,\mathcal{M},\boldsymbol{g})$$

At the maximum:

$$\langle s_i\rangle_{\mathrm{model}} = \langle s_i\rangle_{\mathrm{data}}$$

$$\langle s_i s_j\rangle_{\mathrm{model}} = \langle s_i s_j\rangle_{\mathrm{data}}$$

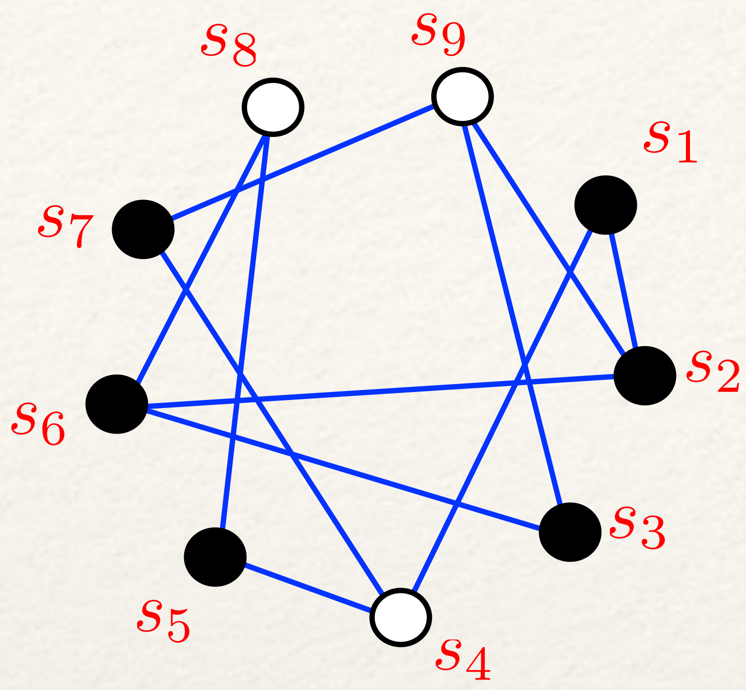
Relevant observables:

$$\langle s_i\rangle \qquad \langle s_i s_j\rangle$$

Model Selection

Do we need all the interactions?

Can we reproduce the correlation patterns with less interactions?



Maybe, can we figure out who is actually connected to who?

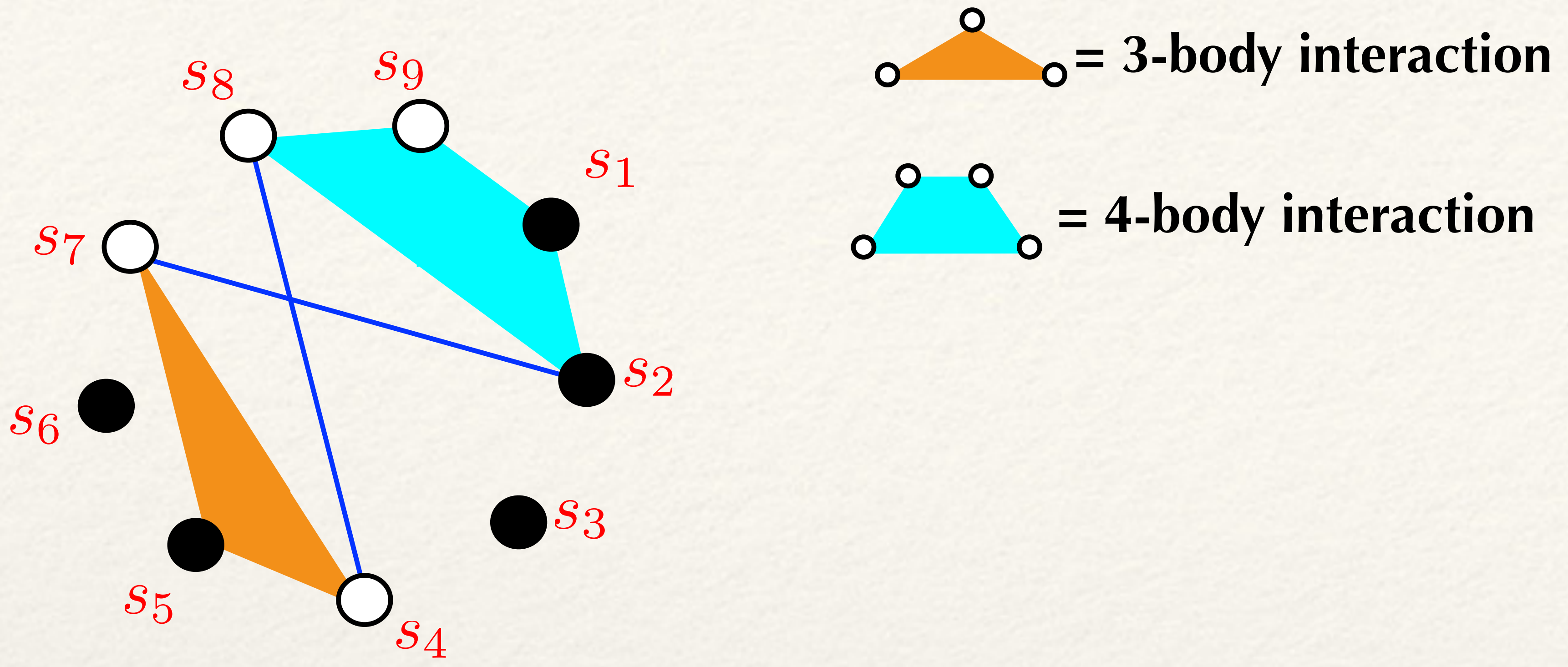
Model Selection

Are the $\langle s_i \rangle$ and $\langle s_i s_j \rangle$ sufficient?

to capture the relevant patterns of the data?

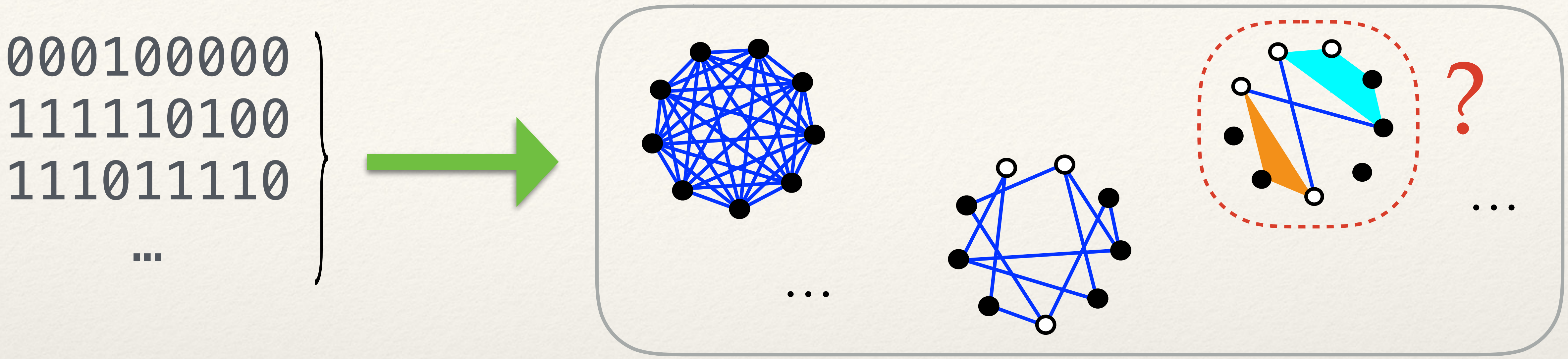
$\langle s_4 s_5 s_7 \rangle$

$\langle s_1 s_2 s_8 s_9 \rangle$



Could it be relevant higher order patterns in the systems?

Which model to select?



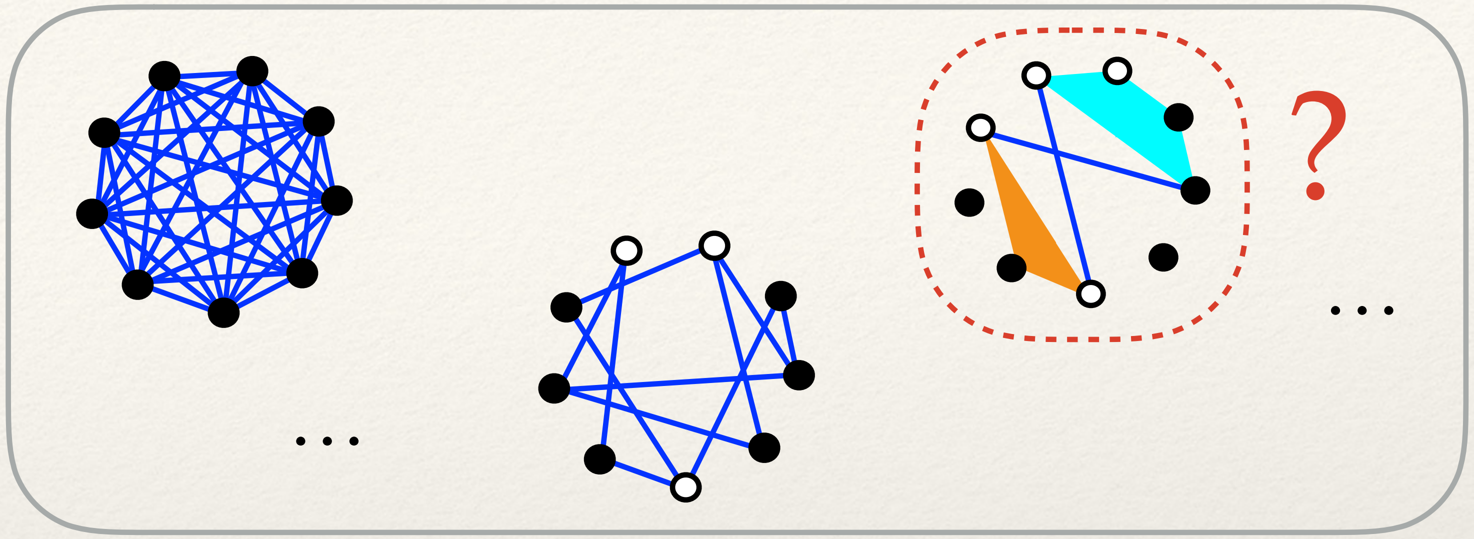
Ideally, we would like the model to be:

not too simple to be able to fit well the data;

not too complex to capture the main patterns of the data and not noise.

Which model to select?

```
000100000
111110100
111011110
...
```



Bayesian Model Selection:

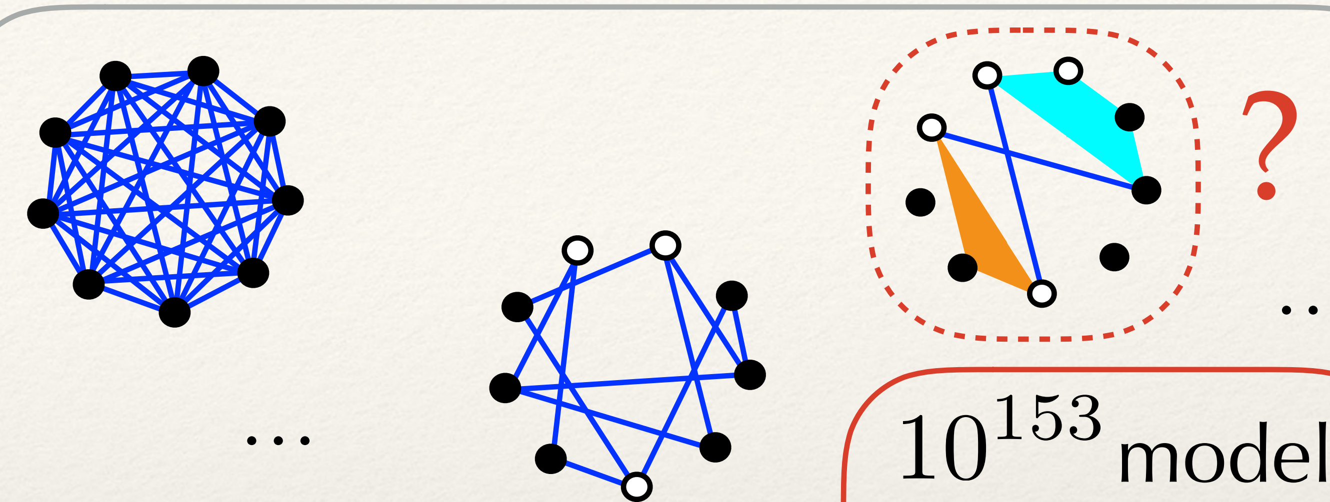
Maximize $P(\hat{s} \,|\, \mathcal{M})$

Minimum Description Length principle:

Minimize $$L(\hat{s} \,|\, \mathcal{M}) = \underbrace{-\log P(\hat{s} \,|\, \mathcal{M}, \boldsymbol{g}^*)}_{\text{How good is the fit?}} + \underbrace{\mathrm{COMP}(\mathcal{M})}_{\text{How complex is the model?}}$$

Which model to select?

```
000100000
111110100
111011110
...
```



10^{153} models

2^{2^n} models!

Bayesian Model Selection:

Maximize $\boxed{P(\hat{s}\,|\,\mathcal{M})}$ $\longrightarrow$ **Hard to compute…**

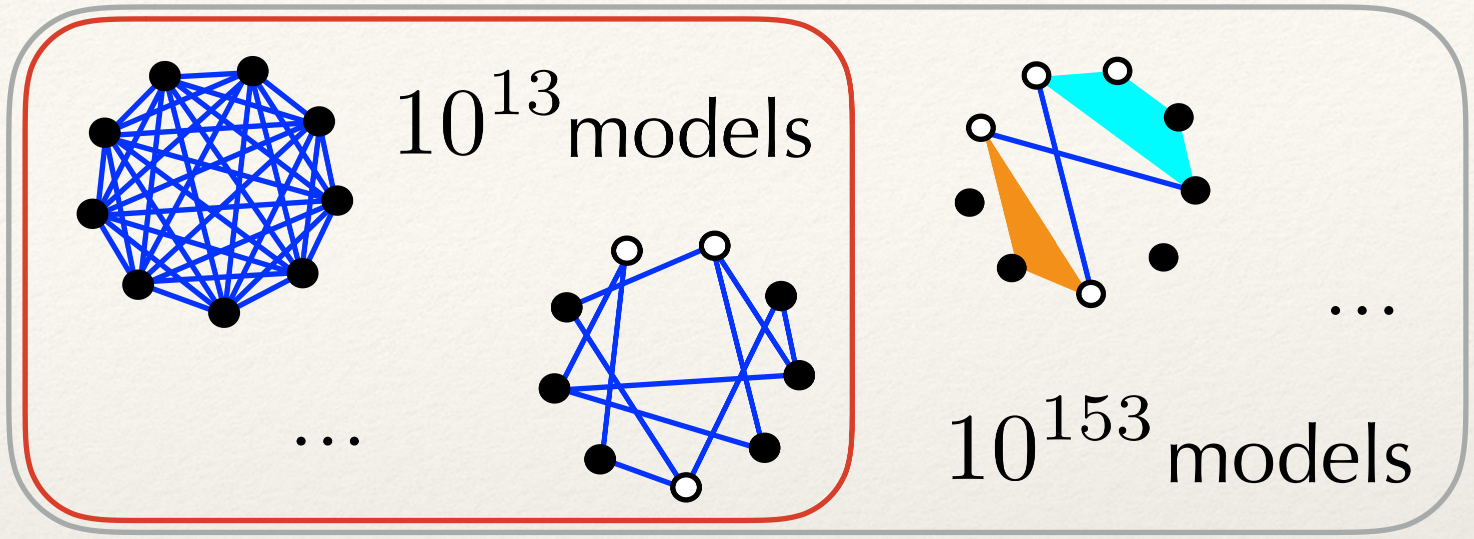
Minimum Description Length principle:

Minimize $L(\hat{s}\,|\,\mathcal{M}) = \boxed{-\log P(\hat{s}\,|\,\mathcal{M}, \boldsymbol{g}^*)} + \boxed{\mathrm{COMP}(\mathcal{M})}$

Pairwise models

Less models:

$2^{n^2/2}$ models!



Why we like pairwise models?

pairwise interactions easier to interpret

able to fit broad types of data

good algorithms for pairwise model selection

But: we already perform a selection…

Are there alternatives?

The Complexity of Spin Models

Are Pairwise Models really Simple?

Alberto Beretta, Claudia Battistin,

Clélia de Mulatier, Iacopo Mastromatteo, Matteo Marsili

The Stochastic Complexity of Spin Models: Are Pairwise Models Really Simple?

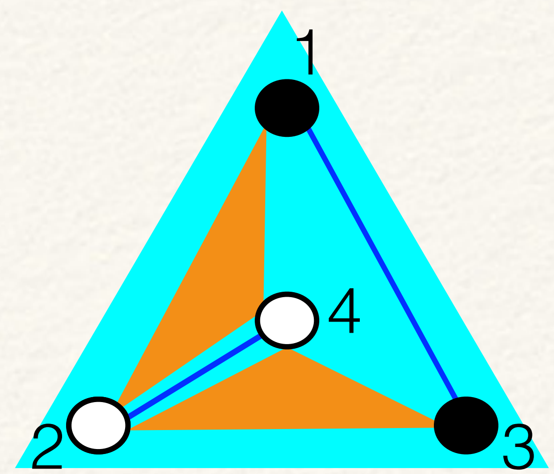
Entropy **2018**, *20*(10), 739

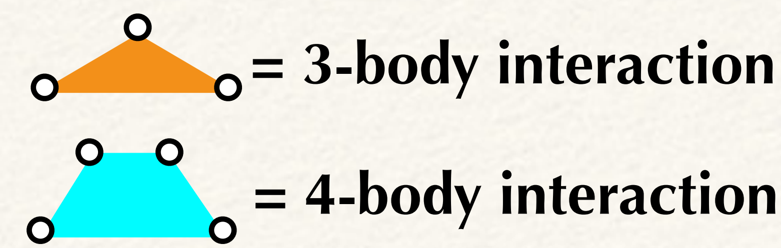
Which Model is the Simplest?

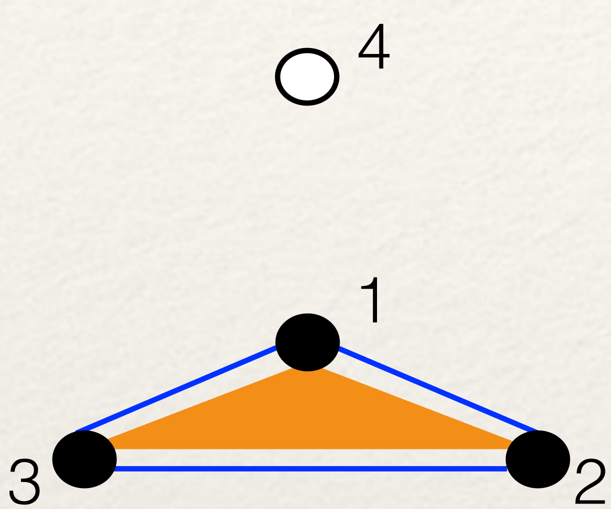
[Model = skeleton]

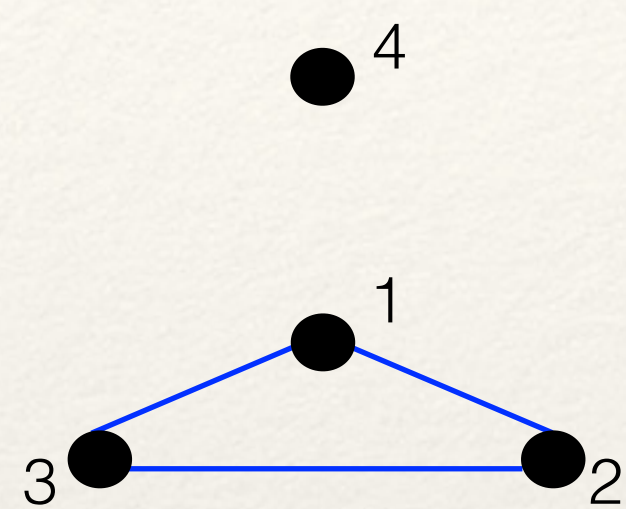
4 spins

7 parameters









$$L(\hat{s}\,|\,\mathcal{M}) = -\log P(\hat{s}\,|\,\mathcal{M}, \boldsymbol{g}^*) + \boxed{\mathrm{COMP}(\mathcal{M})}$$

Thought Experiment

Bob's dataset:

$$\hat{s} = \{\vec{s}^{\,(i)}\}$$

$$\xrightarrow{\mathcal{T}}$$

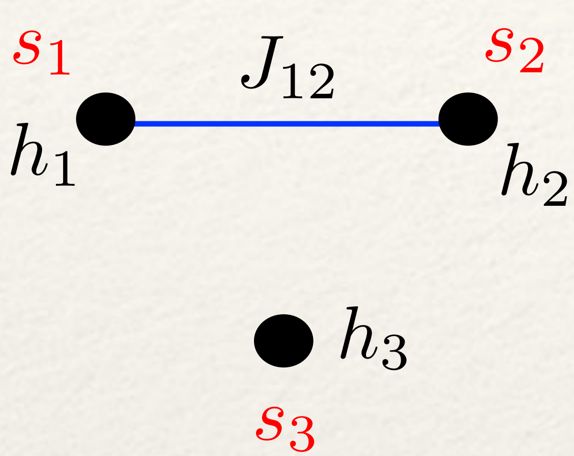
$$\begin{cases} \sigma_1 = s_1 s_2 s_3 \\ \sigma_2 = s_2 \\ \sigma_3 = s_3 \end{cases}$$

Alice's dataset:

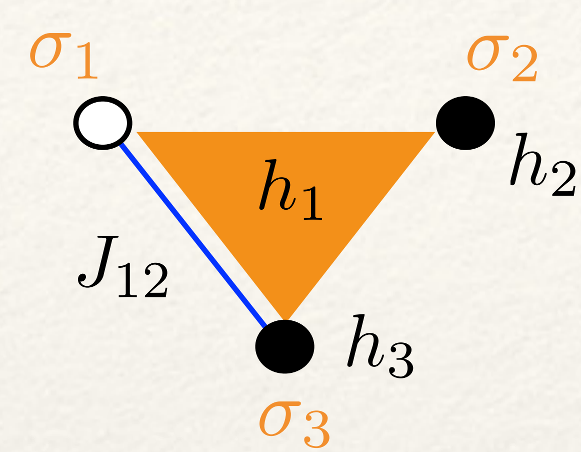
$$\hat{\sigma} = \{\vec{\sigma}^{\,(i)}\}$$

"magic" MS

"magic" MS



$$\xRightarrow{\mathcal{T}}$$



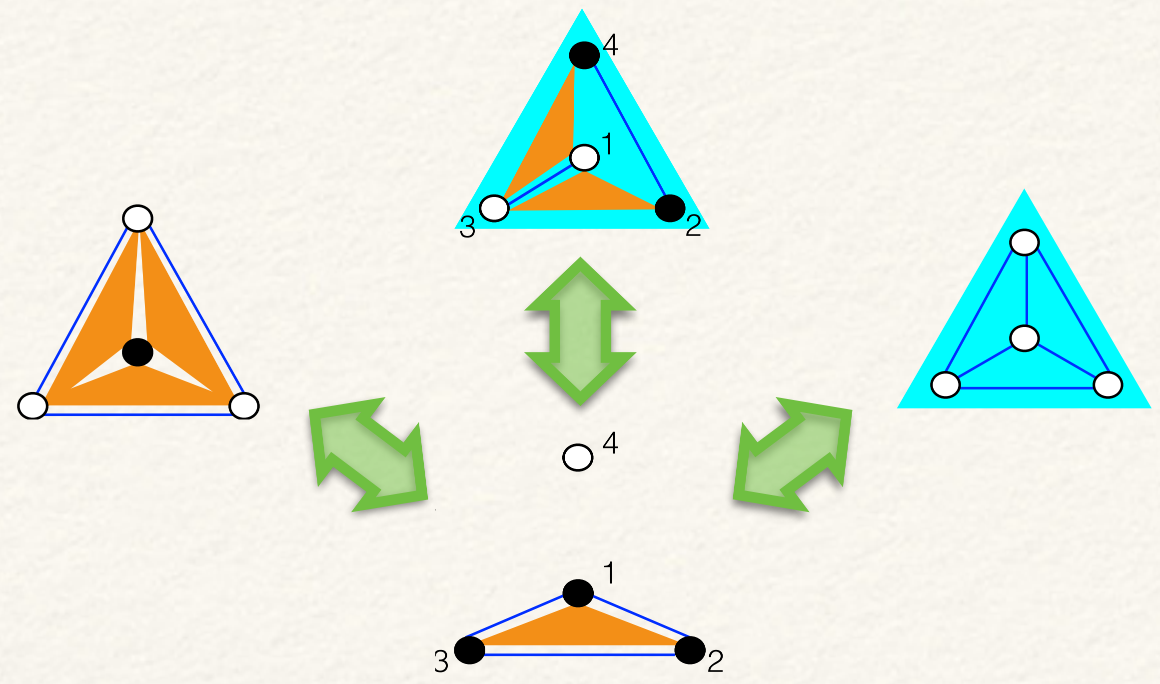
$\longrightarrow$ These 2 models must be **As Complex!!**

Pairwise models are **not necessarily simpler**

Complexity of Spin Models

Complexity **does not dependent** on the order of the interactions

Equivalent classes of models

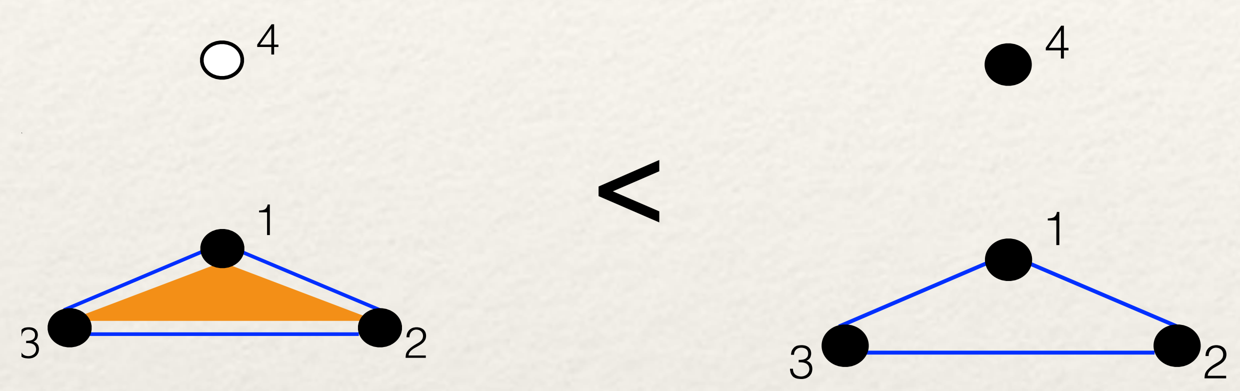


[At fixed K]

Simplest models? = **the most constraints** between the interactions:

— less degrees of freedom;

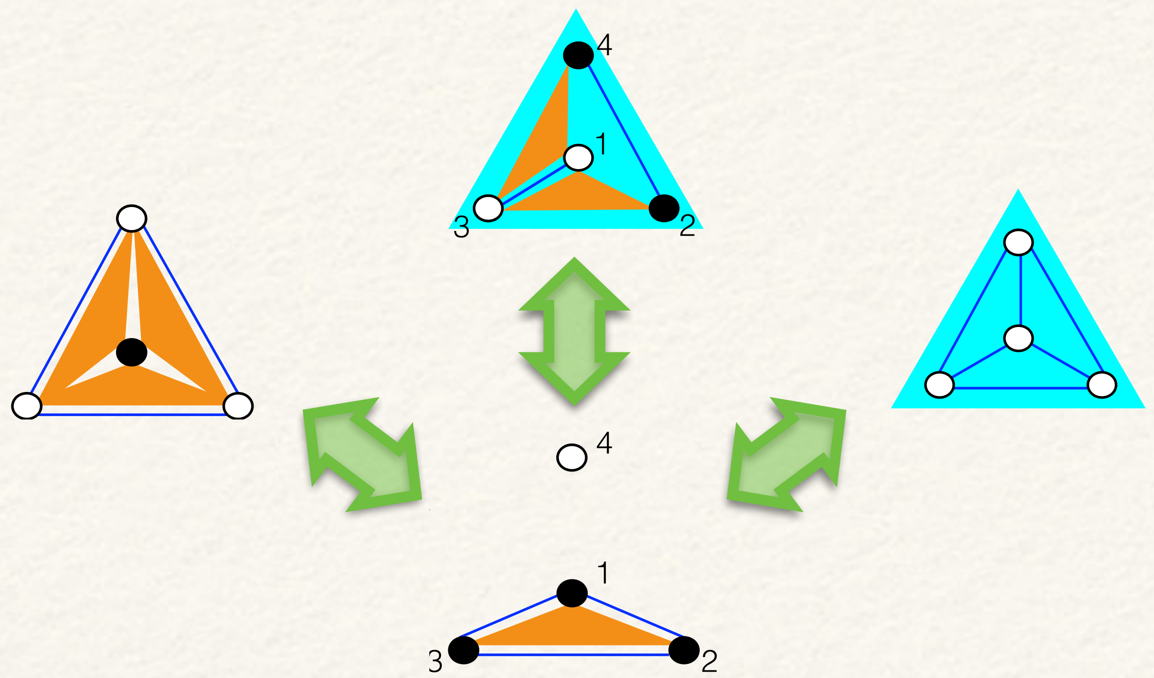
— as compact as possible.



Complexity of Spin Models

Complexity **does not dependent** on the order of the interactions

Equivalent classes of models

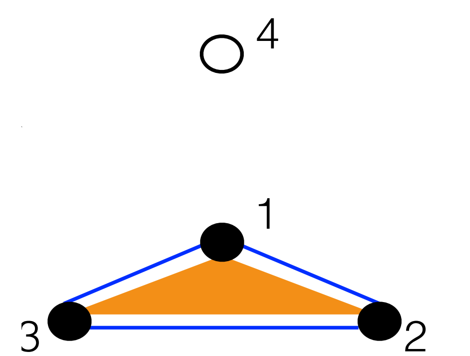


[At fixed K]

Simplest models? = **the most constraints** between the interactions:

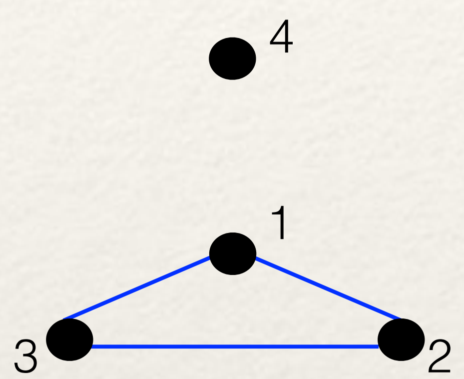
— less degrees of freedom;

— as compact as possible.

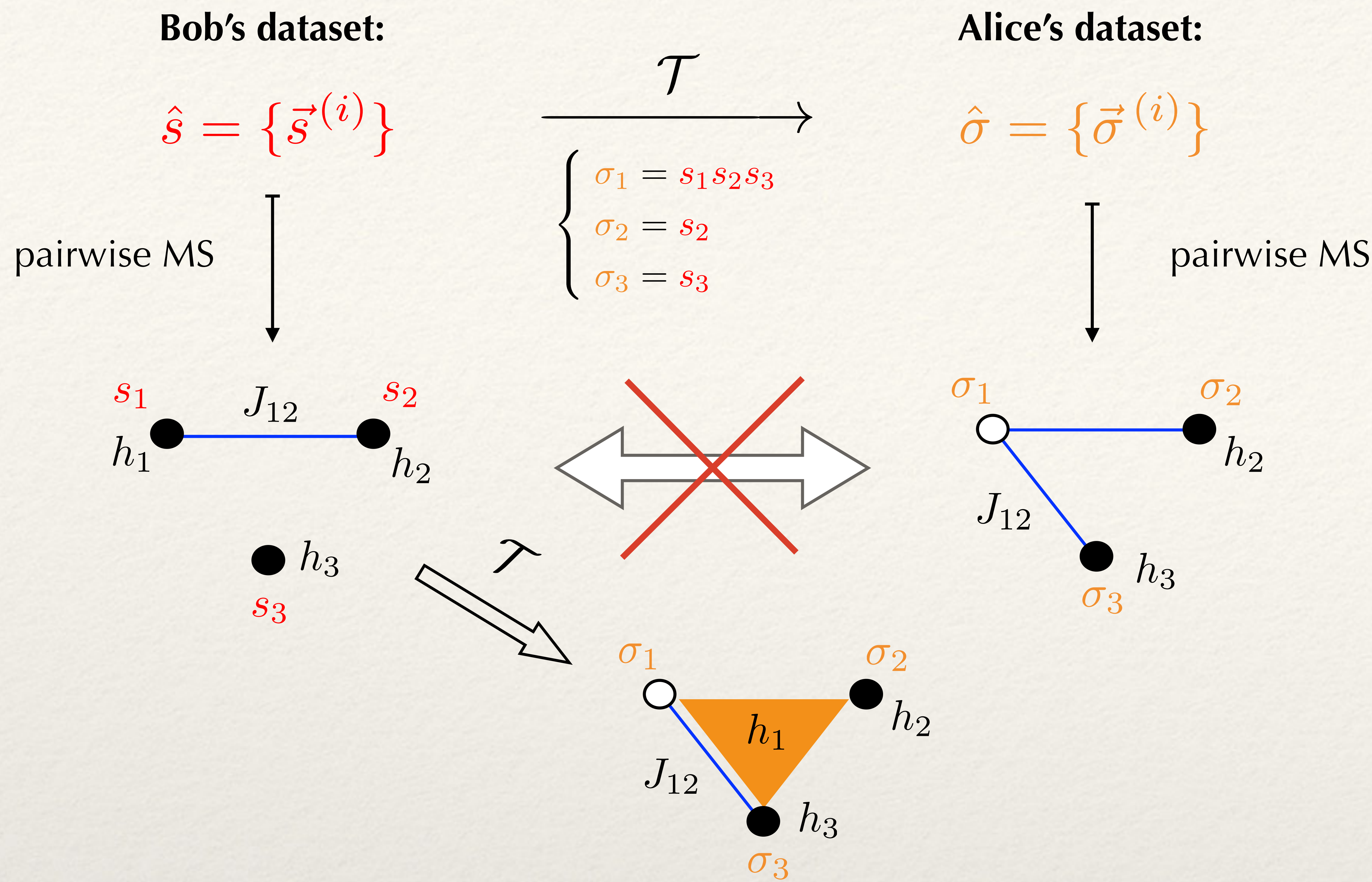


Sub-Complete

<



Pairwise model selection depends on the basis!



Minimally Complex Model Selection

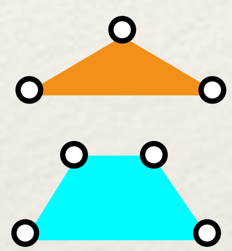
Coming soon on Arxiv….

Clélia de Mulatier, Paolo P. Mazza, Matteo Marsili

Minimally Complex Models (MCM)

Model composed of

Independent Sub-Complete Models



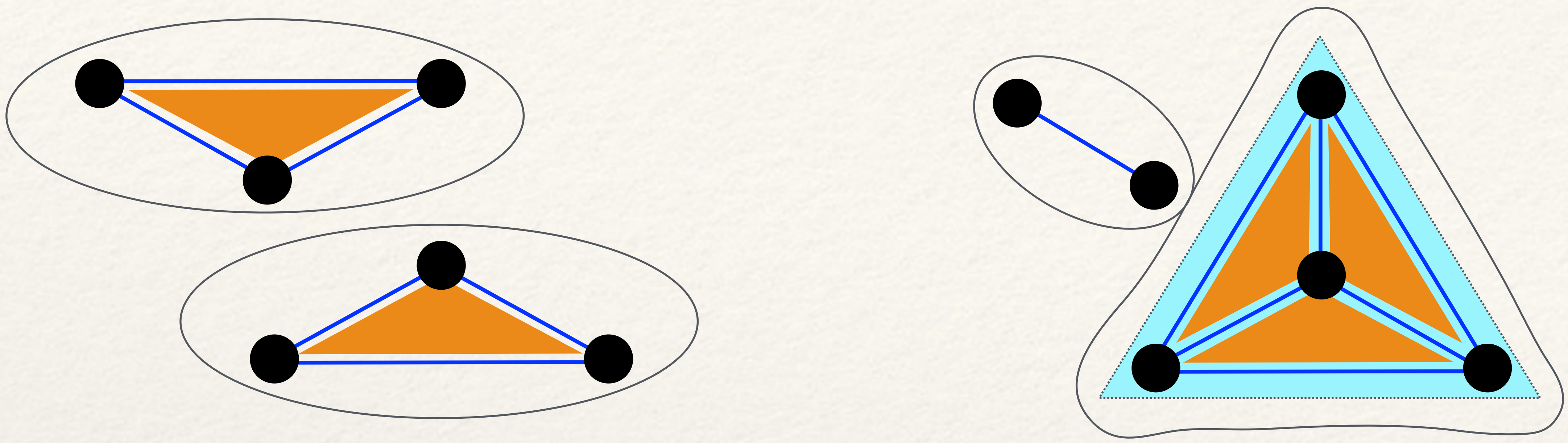
= 3-body interaction

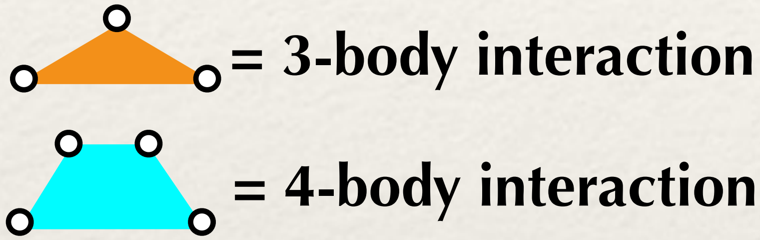
= 4-body interaction

Minimally Complex Models (MCM)

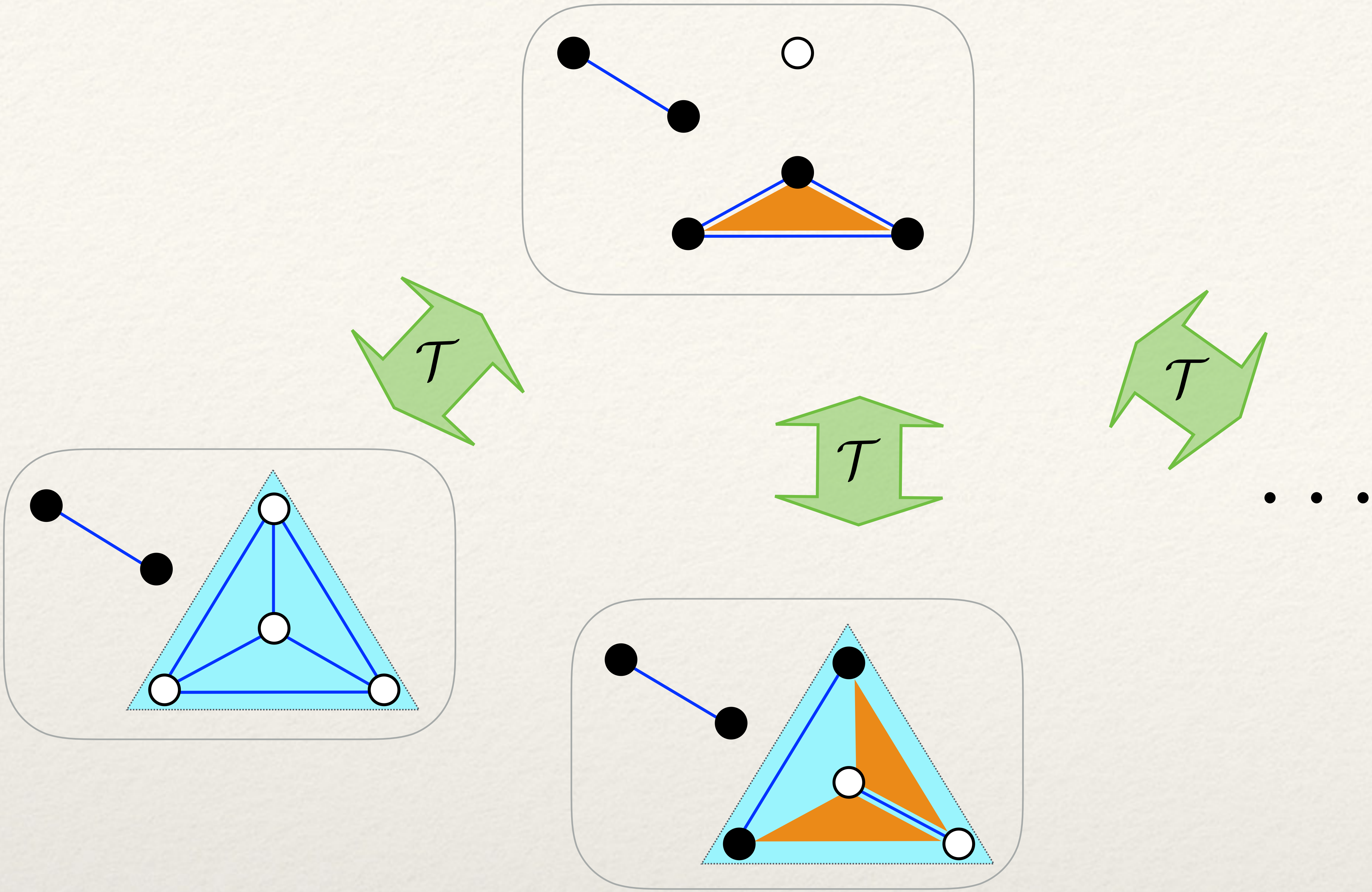
Model composed of

Independent Sub-Complete Models





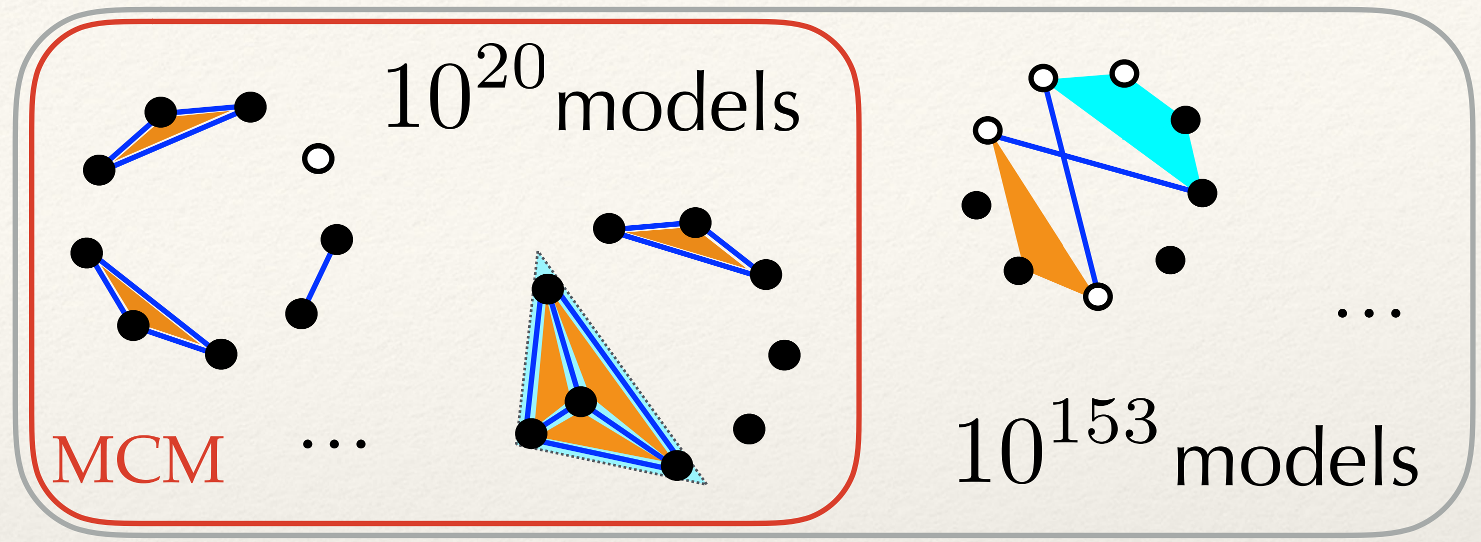
Minimally Complex Models (MCM)



Why we like MCM?

Less models:

2^{n^2} models



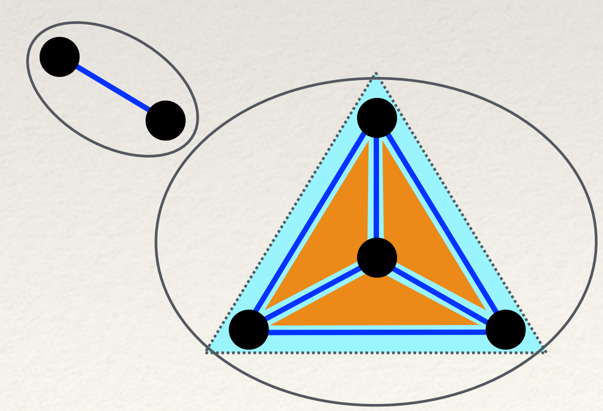
They are *simple* (at fix K and fix degree of freedom)

Among models with the lowest complexity

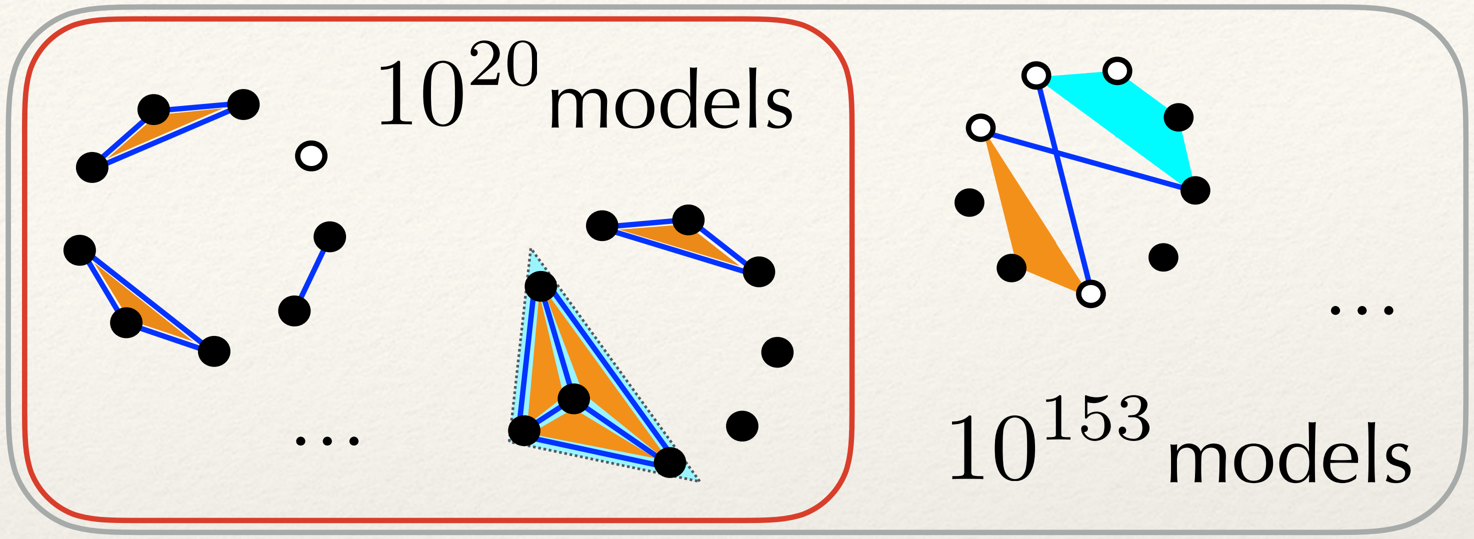
Interpretation Tell us about

dependencies and **independencies** in the system

communities



MCM are easy to compare!



Bayesian Model Selection:

Maximize $\boxed{P(\hat{s}\,|\,\mathcal{M})}$ $\longrightarrow$

Easy to compute

No need to infer parameters!

Minimum Description Length principle:

Minimize $L(\hat{s}\,|\,\mathcal{M}) = \boxed{-\log P(\hat{s}\,|\,\mathcal{M}, \boldsymbol{g}^*)} + \boxed{\mathrm{COMP}(\mathcal{M})}$

Algorithm for finding the best MCM

Find the Best Independent Model:

basis in which the system is **closest to be independent**

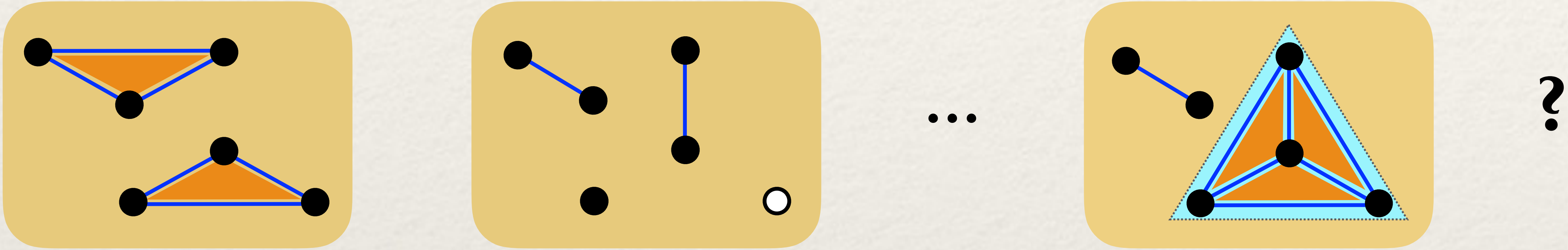
> Most biased independent operators

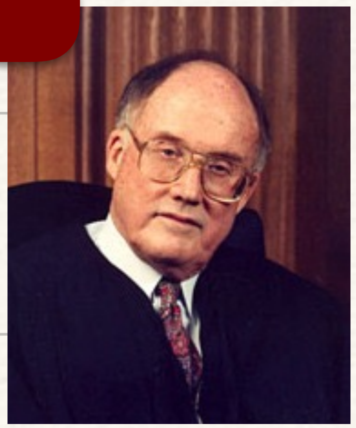
> Decreasing Order of relevance

Reduce the dimension:

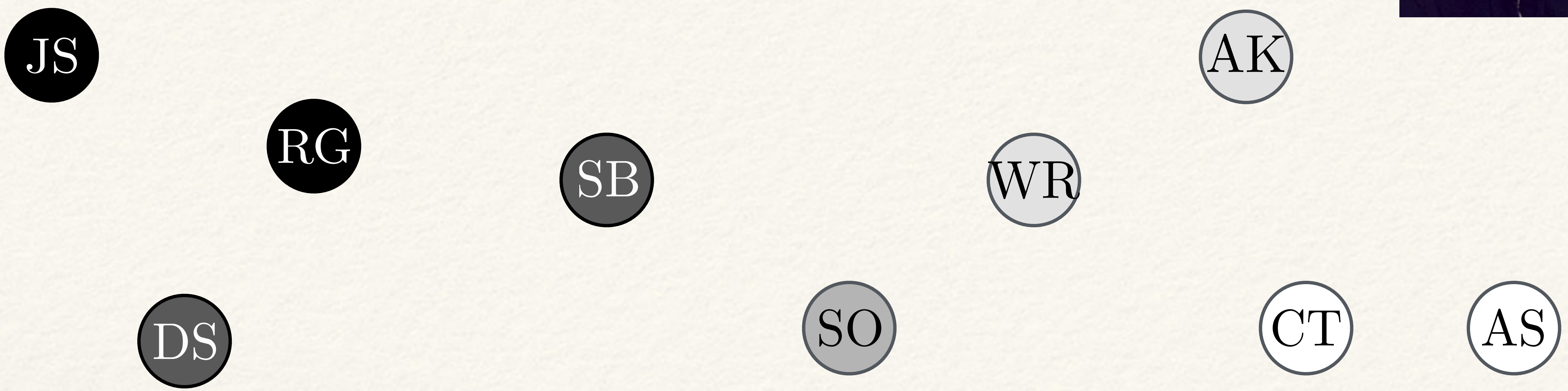
Select only the dimension of the dataset are interesting

Find the best MCM based on this basis:



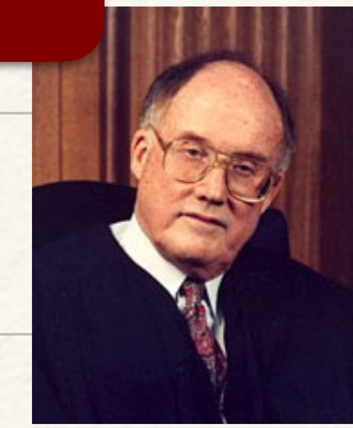


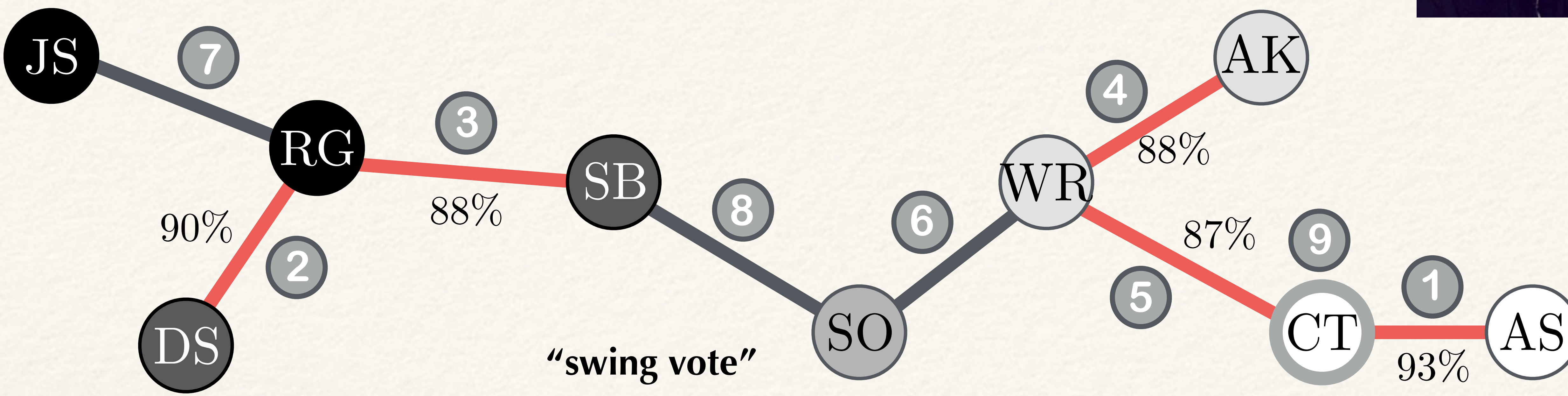
US Supreme Court?



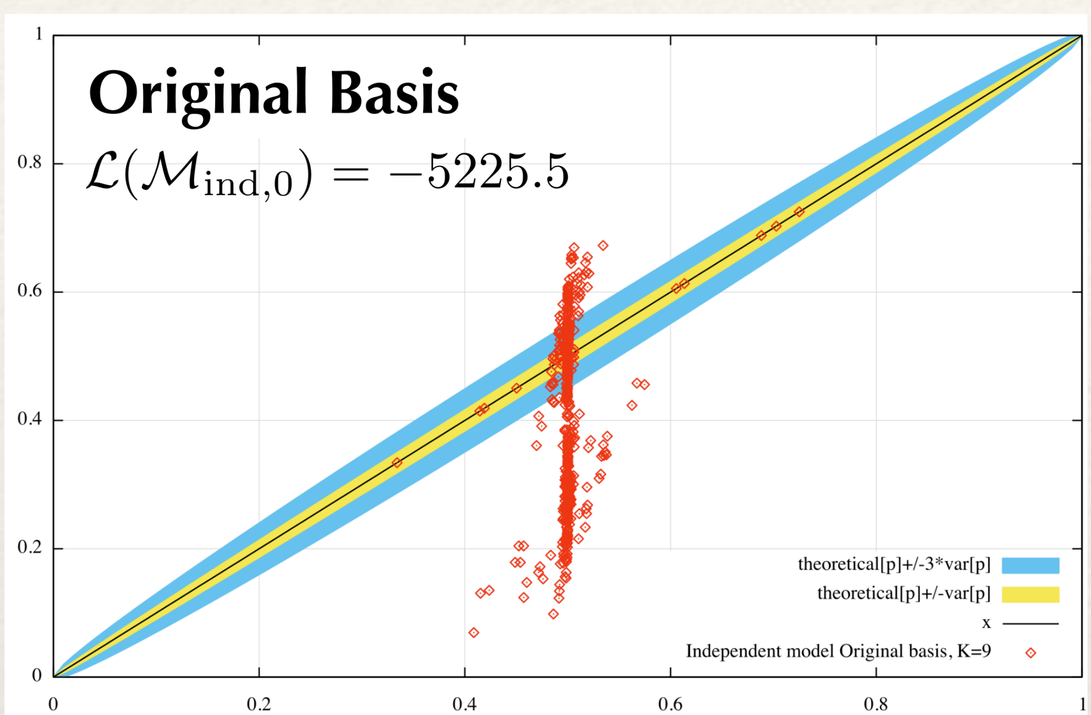
MCM Selection

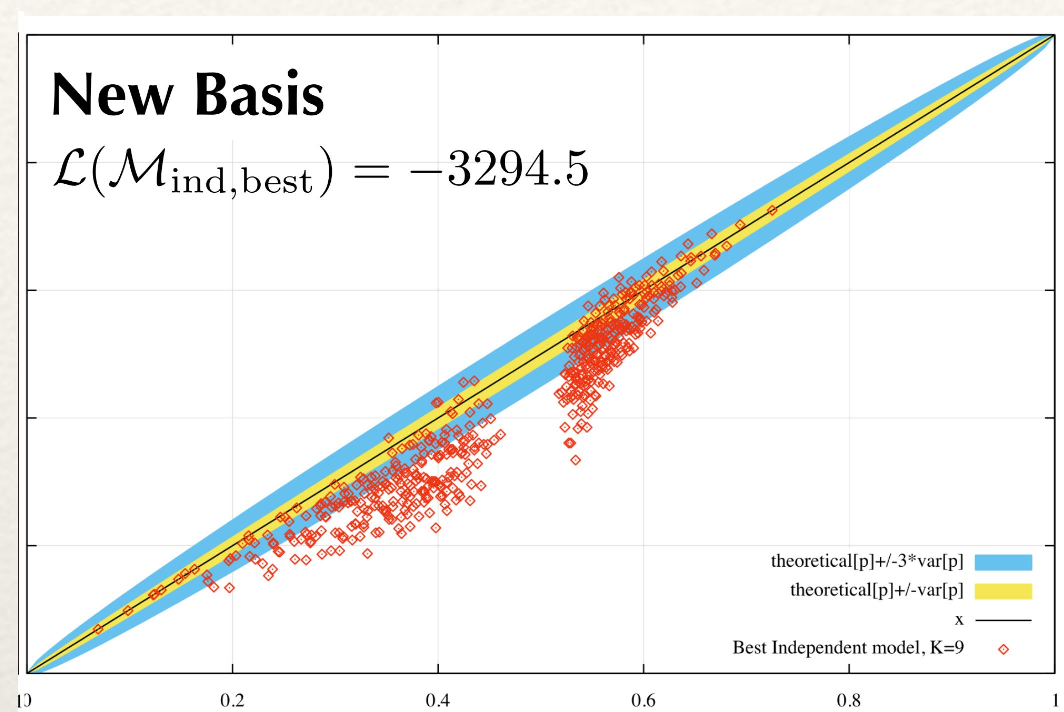
Find Best Independent Model

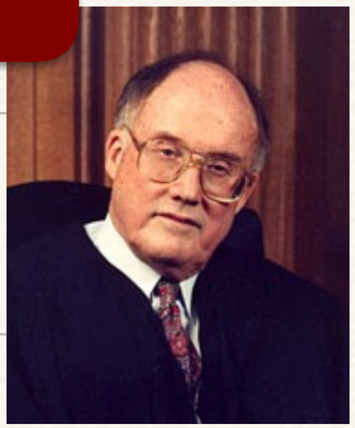




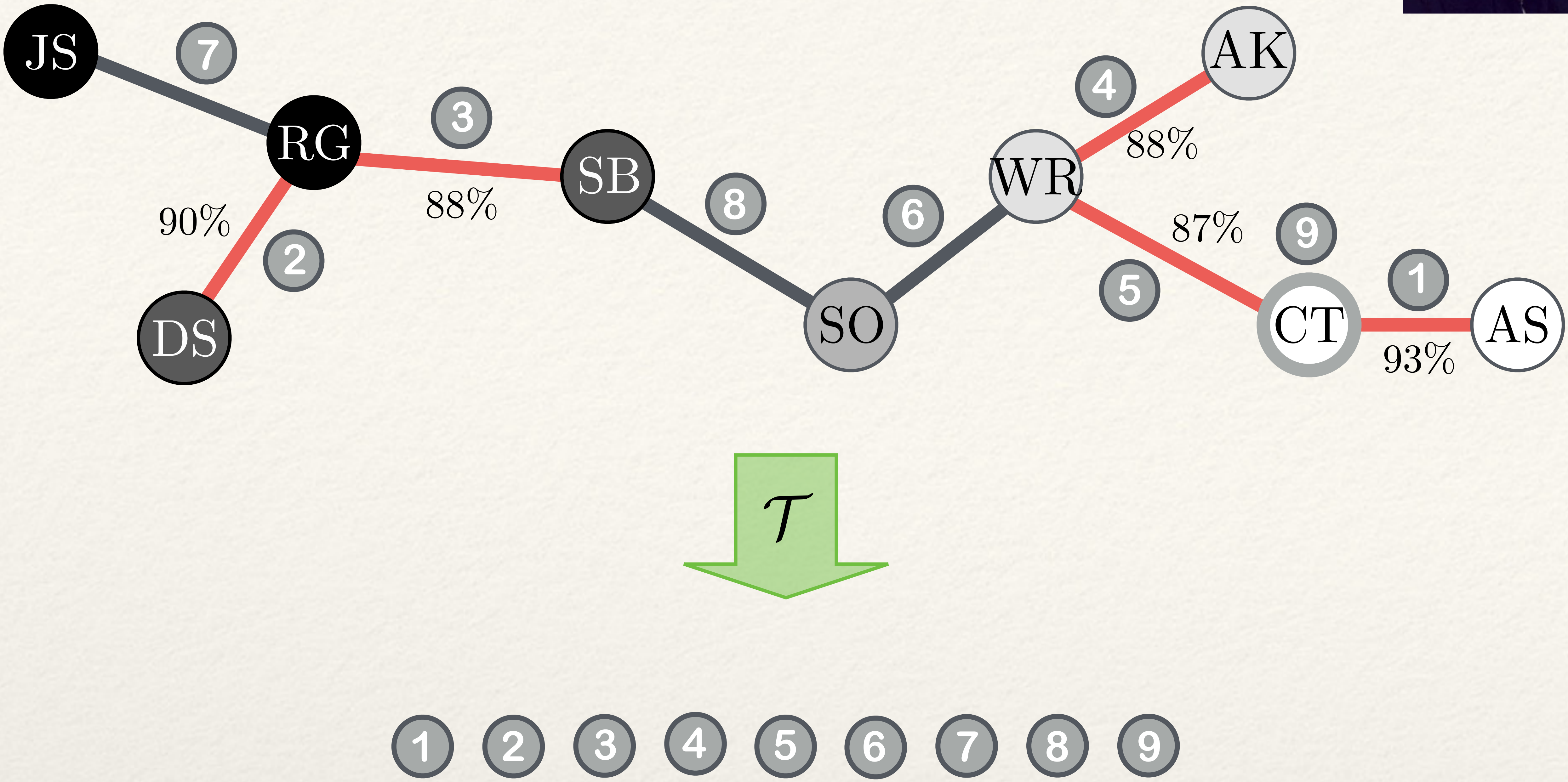
86% of PM !



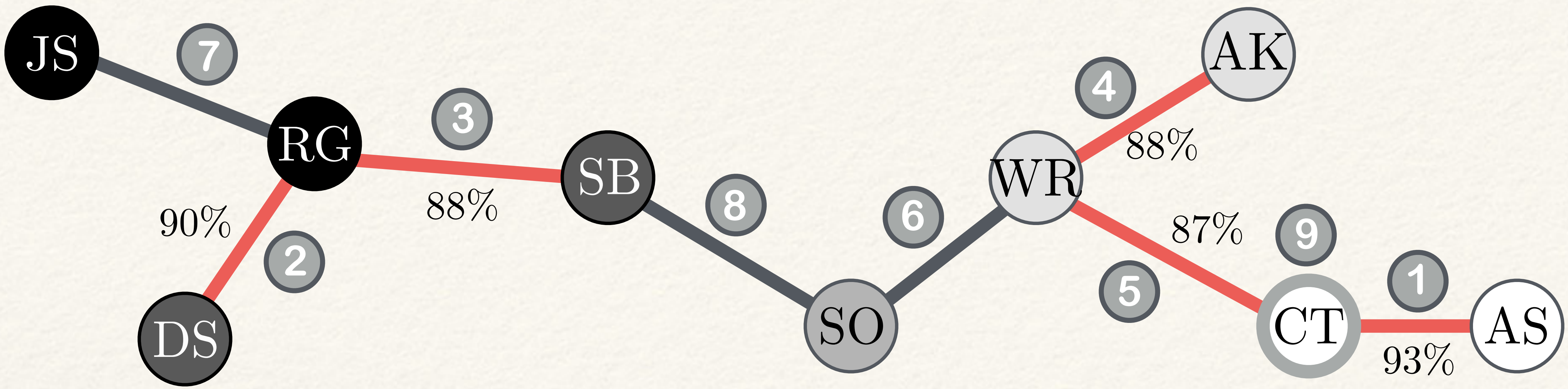




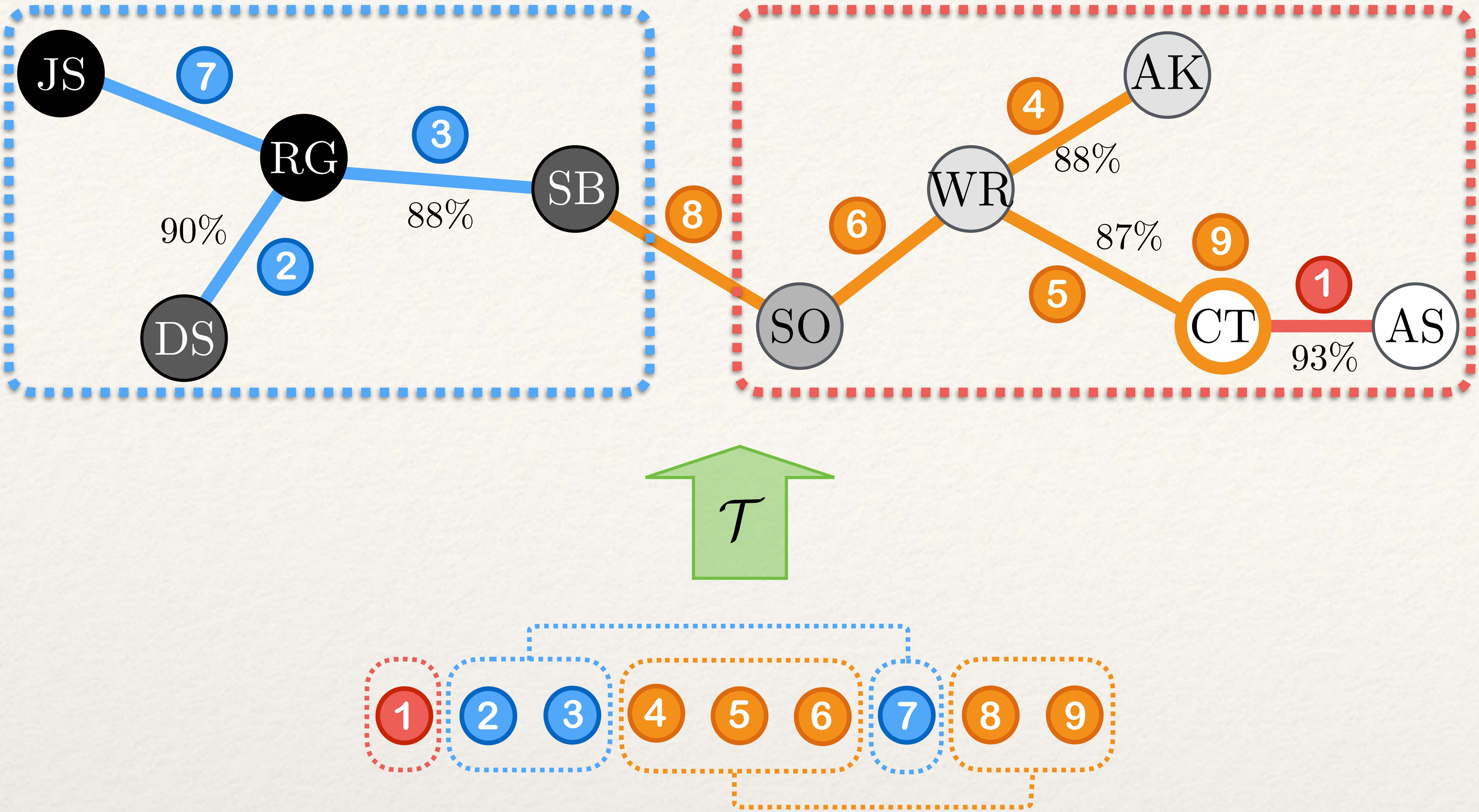
Change basis and reduce dimension



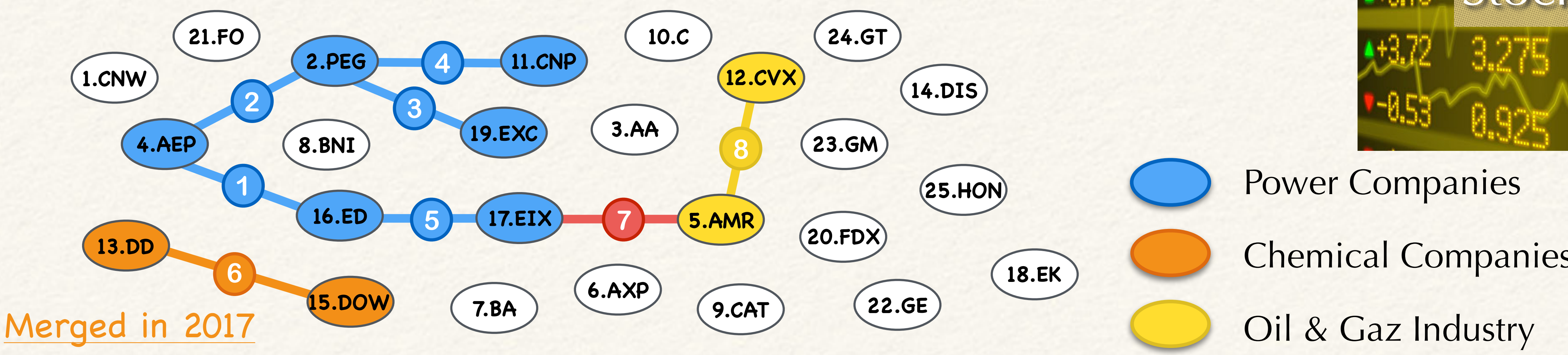
Find Best Minimally Complex Model



Find Best Minimally Complex Model



Other applications



Searching for communities in Bird song data

Eve Armstrong (NYIT) Marc Schmidt Vijay Balasubramanian

David White (Wilfrid Laurier University, CA)

Primary Auditory Cortex: Search for coordinated neuronal ensembles

Taku Banno	**Yale Cohen**	**Vijay Balasubramanian**	
Lalitta Suriya-Arunroj	Jean-Hugues Lestang	Gregory Forkin	*Cassius and Domo*
Ron DiTullio	Jaejin Lee	Songhan Zhang	

Minimally Complex Models

Tell us about Dependencies / Independencies in the system

Communities

Easy to compare No fitting required!

Bayesian approach and MDL principle approach straightforward

Many models but Simple operations —> GPU!

Independent of the basis in which the data are recorded!

Conclusion

Explored new possibilities

There is not yet a perfect model selection.

All techniques are complementary and tells us a part of the story.

Search for simple representation

rather than

simple interpretation