Cambridge 13 September 2016

Parsimonious modeling with Information Filtering Networks:

construction of predictive graphical models from large numbers of heterogeneous data

Tomaso Aste

&

Tiziana Di Matteo

UCL, Computer Science, Financial Computing & Analytics Group

UCL

FINANCIAL COMPUTING

Prediction

Human Behaviour

Predictive modeling

Observation

Model

Parsimony

Prediction

Predictive modeling

Prediction is the estimation of the probability of a (future) event given the available information about other (past) events

$$p(X_B \mid X_A)$$

We must estimate from data the most likely probability distribution of the system of events

$$p(X_B \mid X_A) = \frac{p(X_A, X_B)}{p(X_A)}$$

Bayes’ formula

$$p(X_A, X_B)$$

joint probability

High dimensional problem! *(especially for big data)*

Predictive modeling

Prediction is not only about the future,

from

$$p(X_B \mid X_A)$$

we can predict the values of the variables X_B for any kind of scenario of the variables X_A

We can estimate the effects of events in X_A on X_B

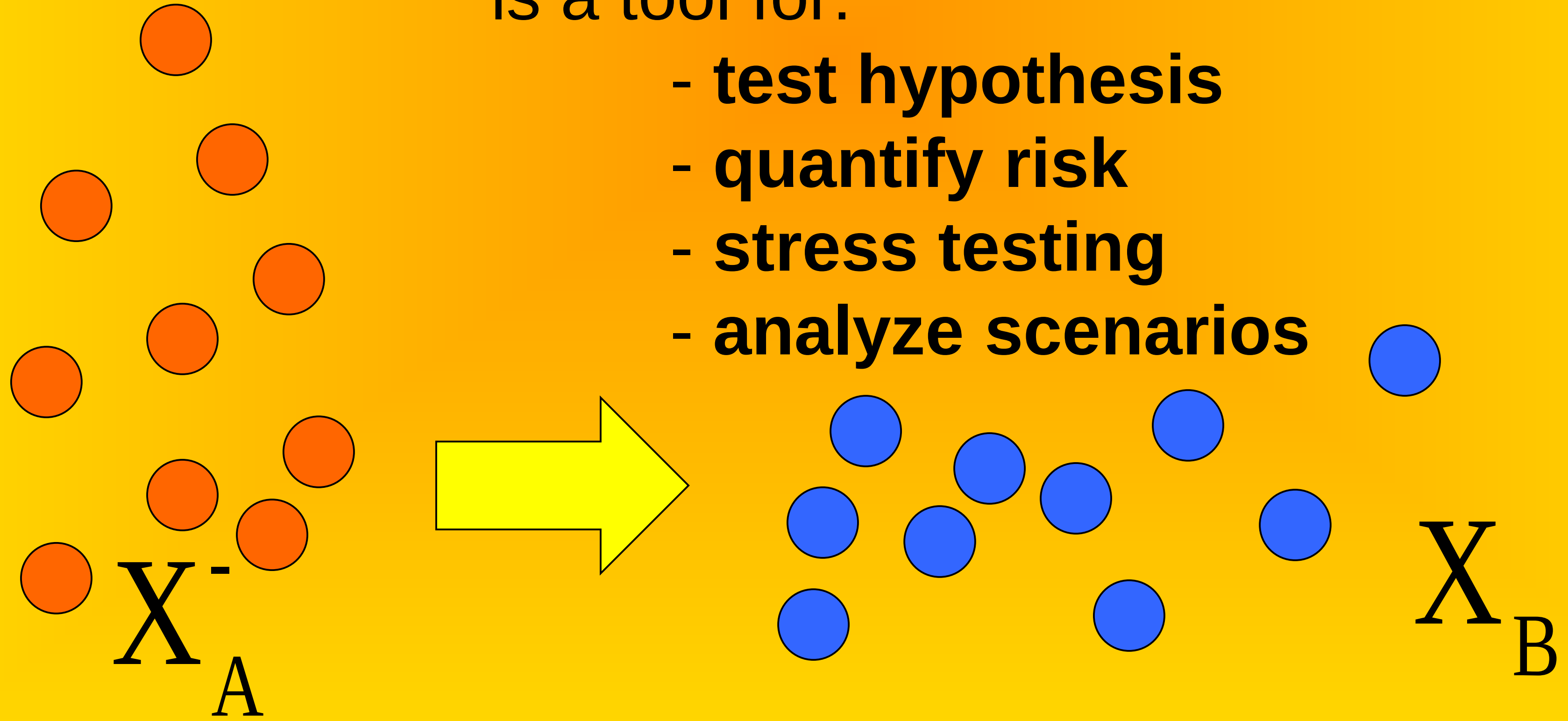
Predictive modeling

The conditional probability

$$p(X_B | X_A)$$

is a tool for:

- **test hypothesis**
- **quantify risk**
- **stress testing**
- **analyze scenarios**



Predictive modeling

Predicted future values of variables X_B given past values of X^-_A are the expectation values

$$E[X_B \mid X^-_A] = \sum_{X_B} X_B \, p(X_B \mid X^-_A)$$

This is the regression and for linear models (multivariate Gaussian) this is the linear regression formula

Uncertainty about the future given the past is quantified by the conditional entropy

$$H(X_B \mid X^-_A) = -\sum_{X^-_A, X_B} p(X_B, X^-_A) \log p(X_B \mid X^-_A)$$

Predictive modeling

The **reduction of uncertainty** on variables X_B given the knowledge of the past of variables X^-_A discounting for their past X^-_B is

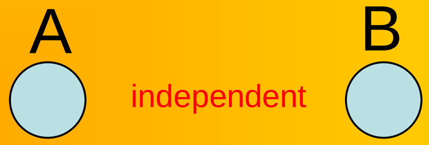
$$H(X_B \mid X^-_B) - H(X_B \mid X^-_A, X^-_B) = TE(X_A \rightarrow X_B)$$

*Thisis the **tranfer entropy** that for liner models (multivariate Gaussians) coincides with **Granger causality***

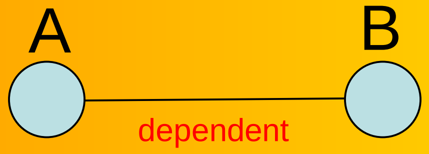
Graphical models

To construct the joint multivariate distribution we make use of the structure of conditional dependency

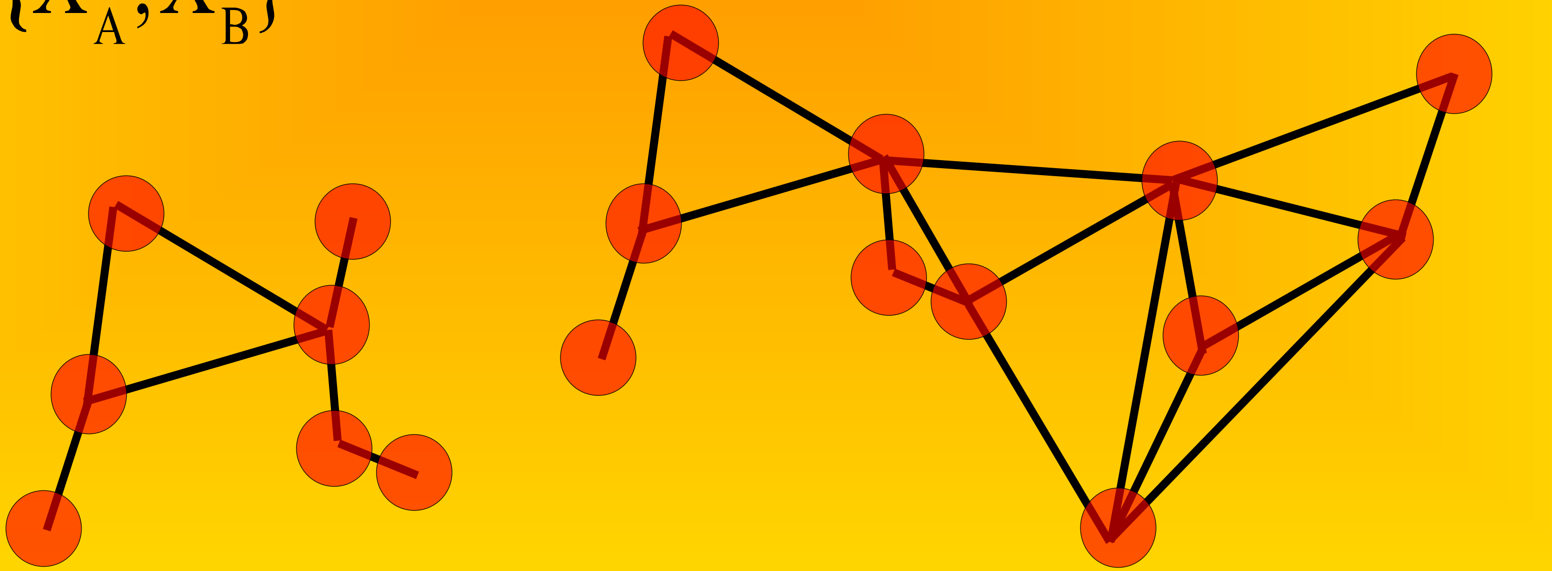
$$p(X_A, X_B \mid \tilde{\mathbf{X}}) = p(X_A \mid \tilde{\mathbf{X}})\, p(X_B \mid \tilde{\mathbf{X}})$$



$$p(X_A, X_B \mid \tilde{\mathbf{X}}) \neq p(X_A \mid \tilde{\mathbf{X}})\, p(X_B \mid \tilde{\mathbf{X}})$$



$$\tilde{\mathbf{X}} = \mathbf{X} \setminus \{X_A, X_B\}$$

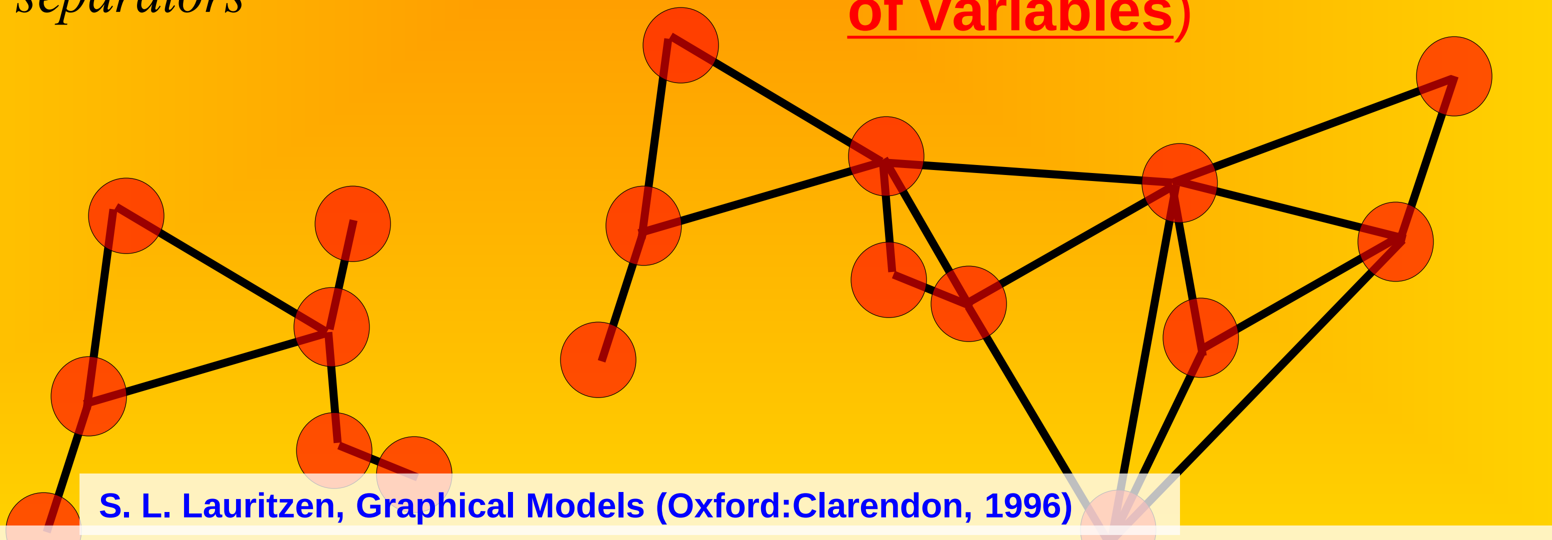


Graphical models

If these inference networks are <u>chordal</u> (or decomposable) we then have

$$p(\mathbf{X}) = \frac{\underset{cliques}{\text{Õ}} \, p(\mathbf{X}_{cliques})}{\underset{separators}{\text{Õ}} \, p(\mathbf{X}_{separators})^{k_s - 1}}$$

The joint probability distribution of the entire system (**<u>large number of variables</u>**) can be estimated form the probability distributions of cliques and separators (**<u>small number of variables</u>**)



S. L. Lauritzen, Graphical Models (Oxford:Clarendon, 1996)

Alexander Denev Probabilistic Graphical Models: A New Way of Thinking in Financial Modelling (Risk Books, 2015)

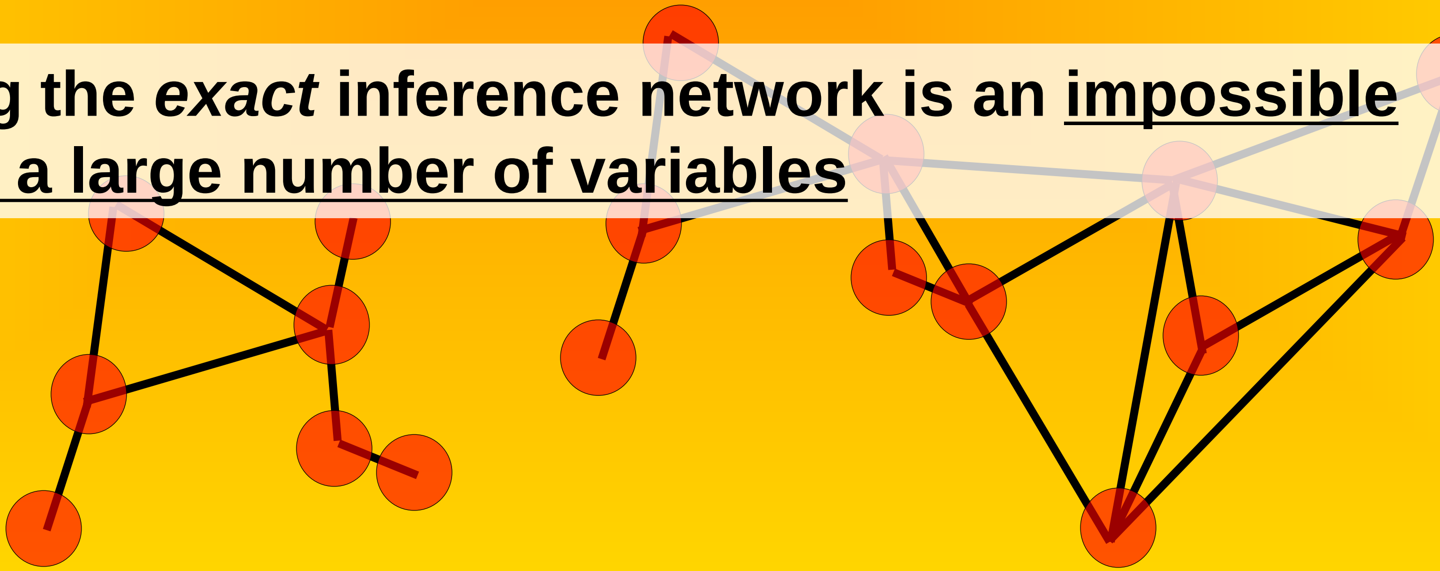
Graphical models

This is great… however to establish conditional dependency

$$p(\mathrm{X}_{\mathrm{A}},\mathrm{X}_{\mathrm{B}}\,|\,\tilde{\mathbf{X}}) \neq p(\mathrm{X}_{\mathrm{A}}\,|\,\tilde{\mathbf{X}})\,p(\mathrm{X}_{\mathrm{B}}\,|\,\tilde{\mathbf{X}})$$

is very hard… actually it is as hard as computing the entiere joint distribution function!

Building the *exact* inference network is an impossible task for a large number of variables

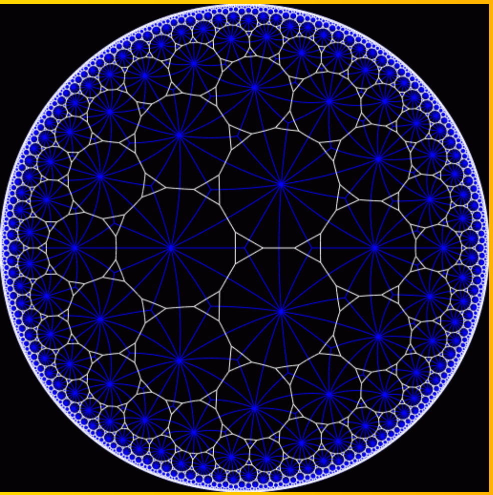


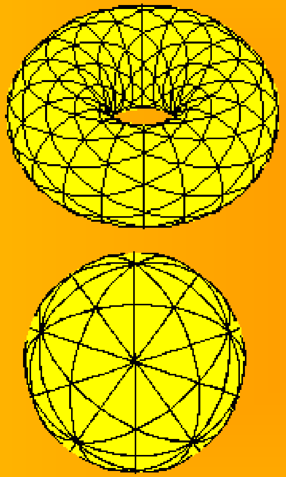
Information filtering networks

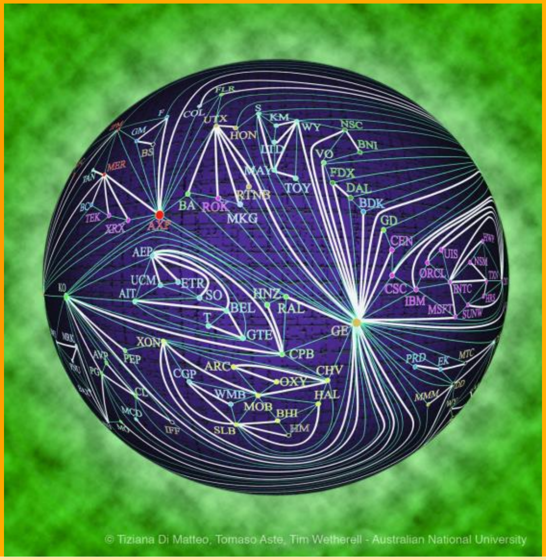
To solve this problem we propose to build the inference structure for the graphical model as an

Information filtering network

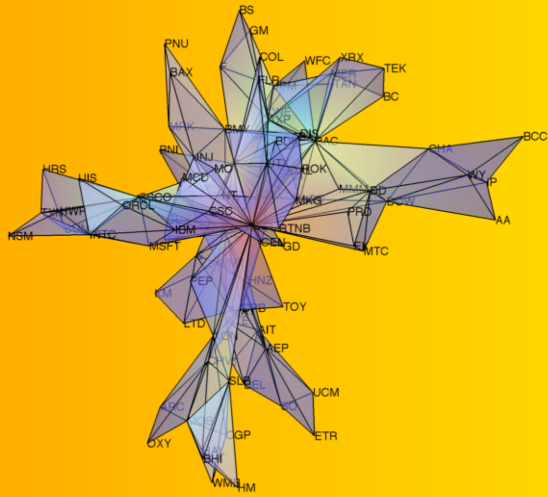
TA, T. Di Matteo and S. T. Hyde, *Complex networks on hyperbolic surfaces* Physica *A* 346 (2005) 20-26.







© Tiziana Di Matteo, Tomaso Aste, Tim Wetherell - Australian National University



- Massara, Guido Previde, Tiziana Di Matteo, and TA. "Network Filtering for Big Data: Triangulated Maximally Filtered Graph" Journal of Comlex Networks (2016) arXiv preprint arXiv:1505.02445 (2015).
- Nicoló Musmeci,, Tomaso Aste, and Tiziana Di Matteo. "Relation between financial market structure and the real economy: comparison between clustering methods." PloS one 10.3 (2015): e0116201.
- F. Pozzi, T. Di Matteo, and TA , "Spread of risk across financial markets: better to invest in the peripheries", Scientific Reports 3 (2013) 1665.
- W.M. Song, T. Di Matteo and T. Aste, "Hierarchical information clustering by means of topologically embedded graphs", *PLoS ONE*, 7 (2012) e31929
- M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems" Proceedings of the National Academy of Sciences of the United States of America 102, 10421 (2005).

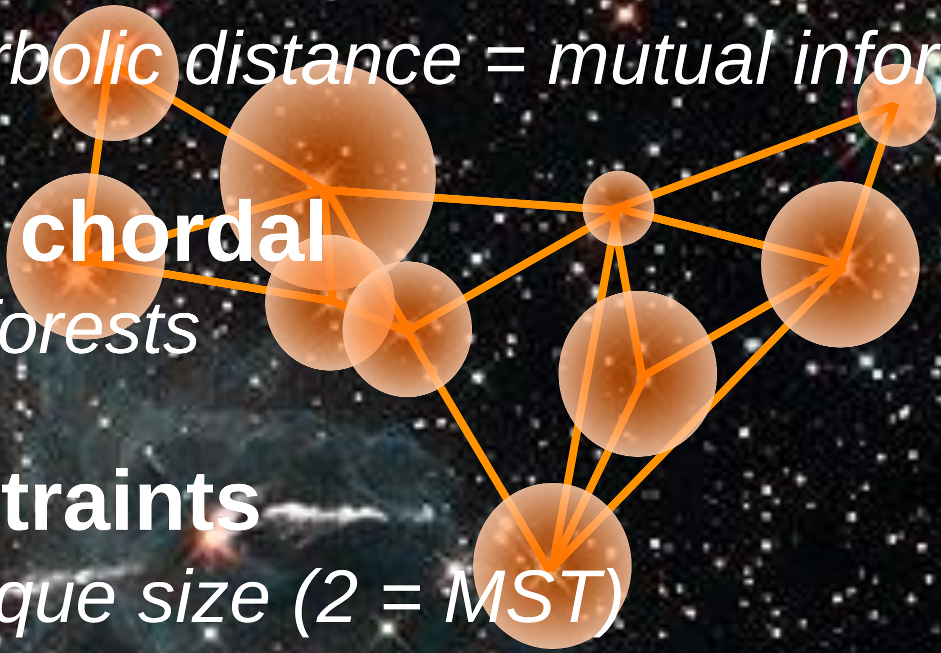
Information filtering networks

Information filtering networks

Connect the nearest vertices

eucleadean distance = most correlated

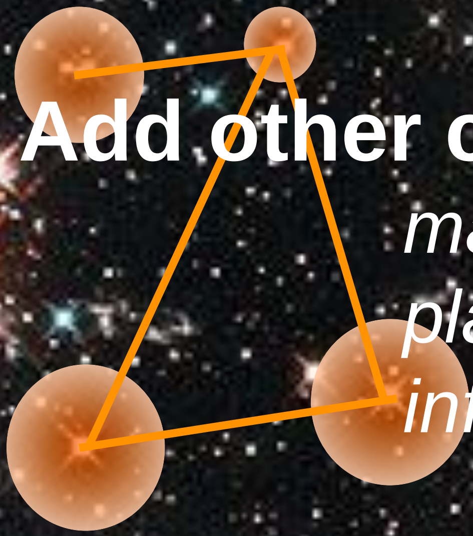
hyperbolic distance = mutual information



Keep the graph chordal

clique forests

Add other constraints



max clique size (2 = MST)

planarity (TMFG)

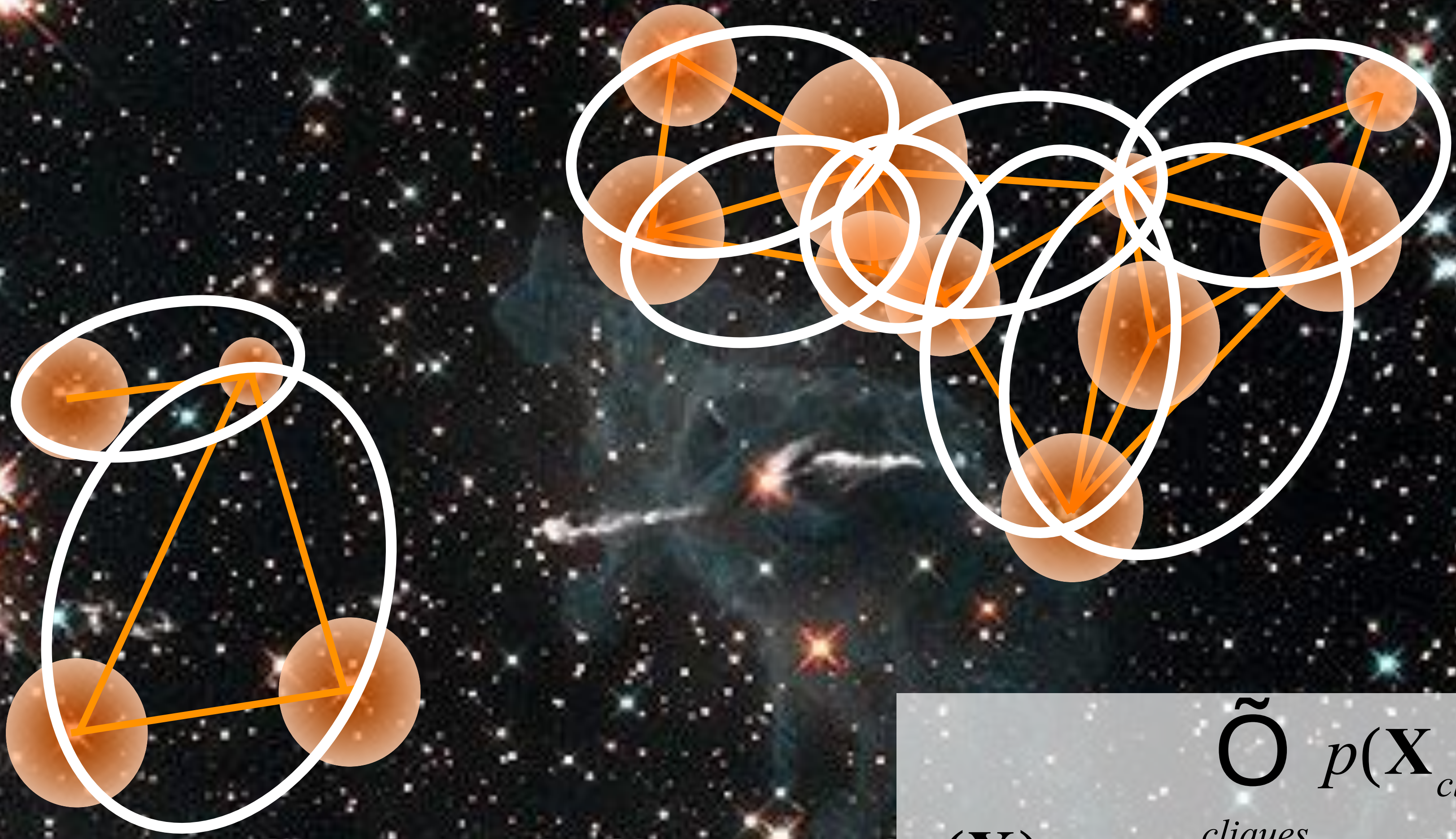
information criteria (e.g. Akaike)

These are fast algorithms $O(N^2)$

(topological & homological measures, betty numbers, cycles and cliques retrieved from construction)

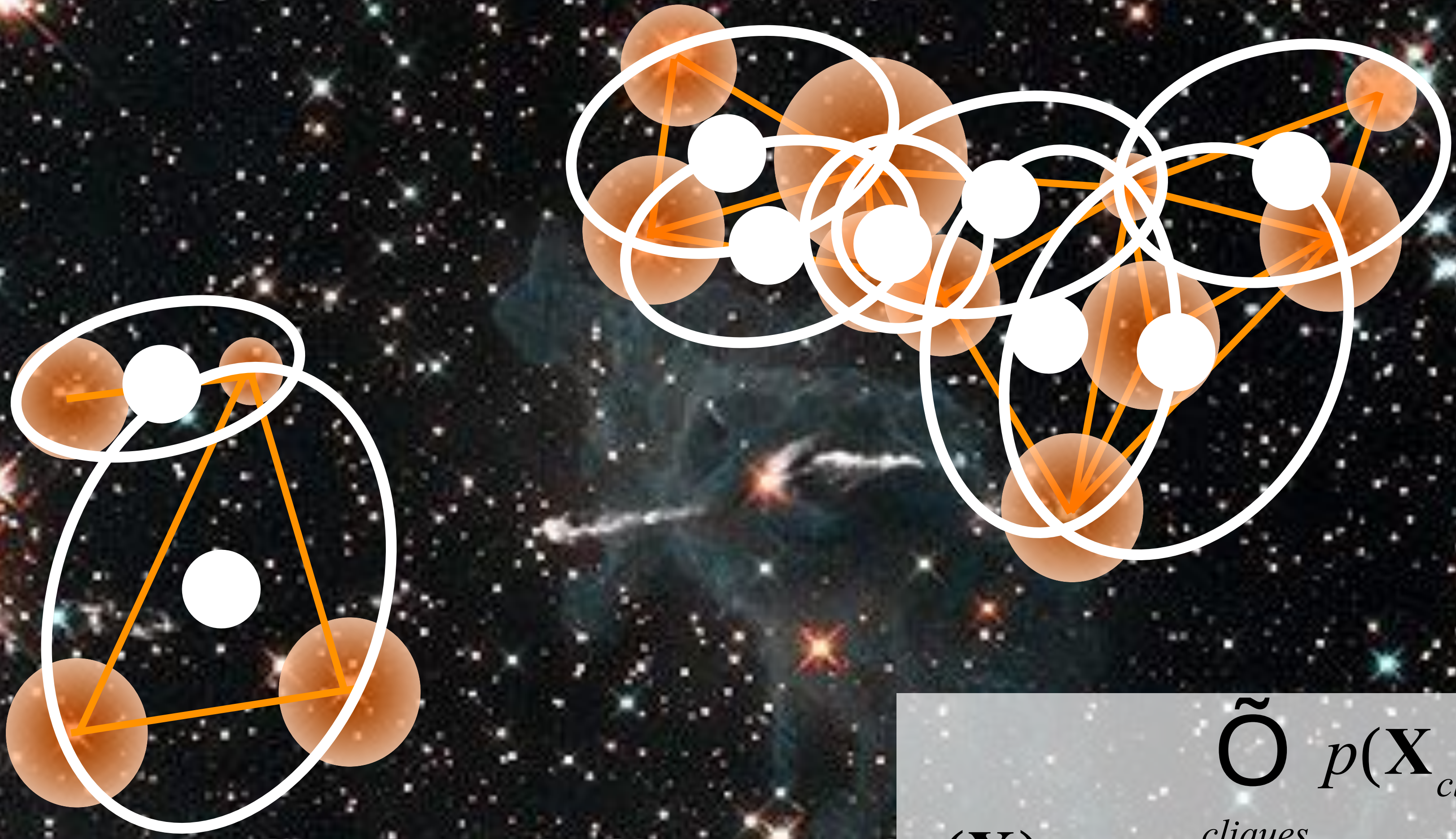
Massara, Guido Previde, Tiziana Di Matteo, and TA. "Network Filtering for Big Data: Triangulated Maximally Filtered Graph" Journal of Comlex Networks (2016) arXiv preprint arXiv:1505.02445 (2015).

Clique forest



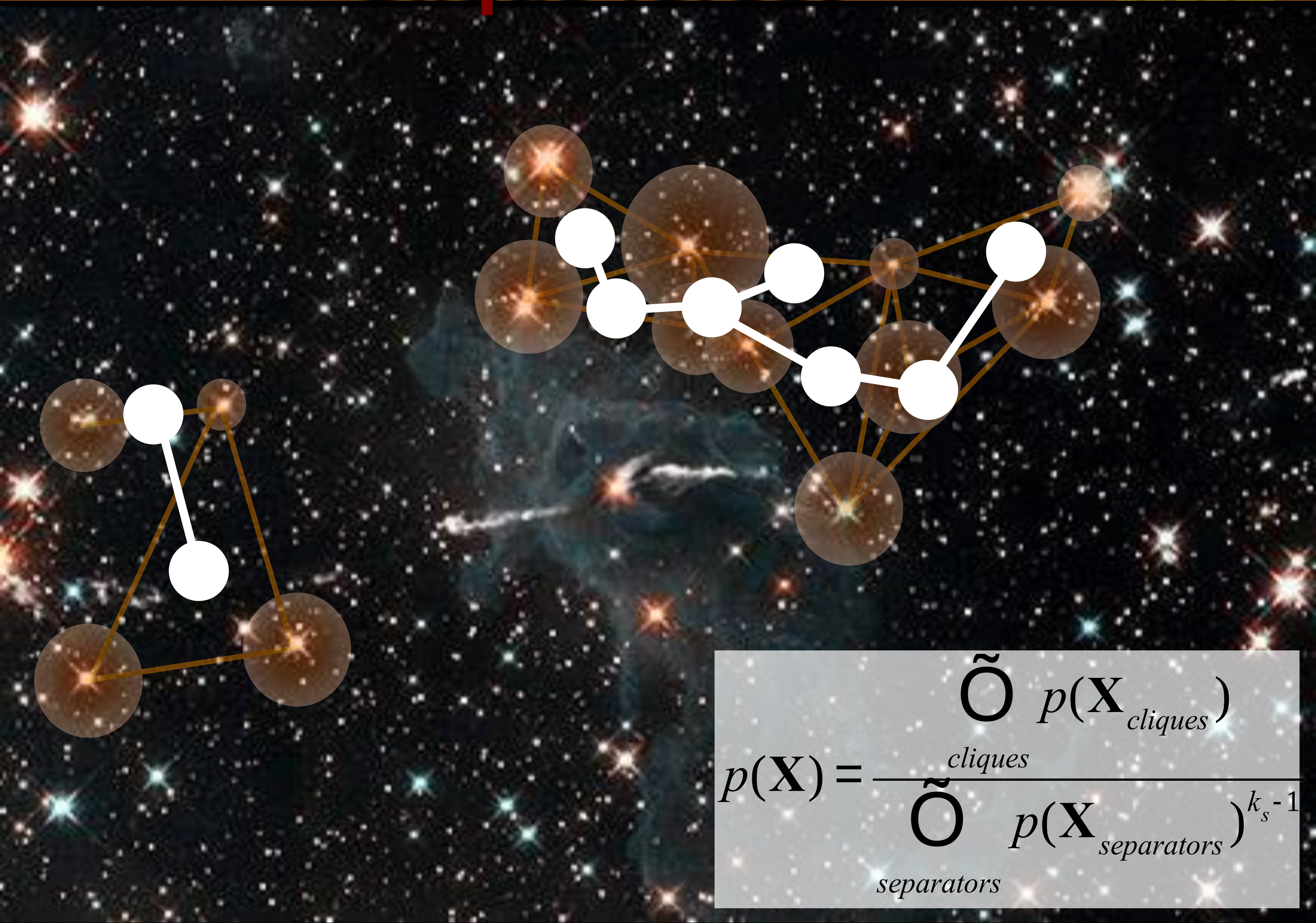
$$p(\mathbf{X}) = \frac{\tilde{\mathrm{O}}_{cliques}\; p(\mathbf{X}_{cliques})}{\tilde{\mathrm{O}}_{separators}\; p(\mathbf{X}_{separators})^{k_s - 1}}$$

Clique forest



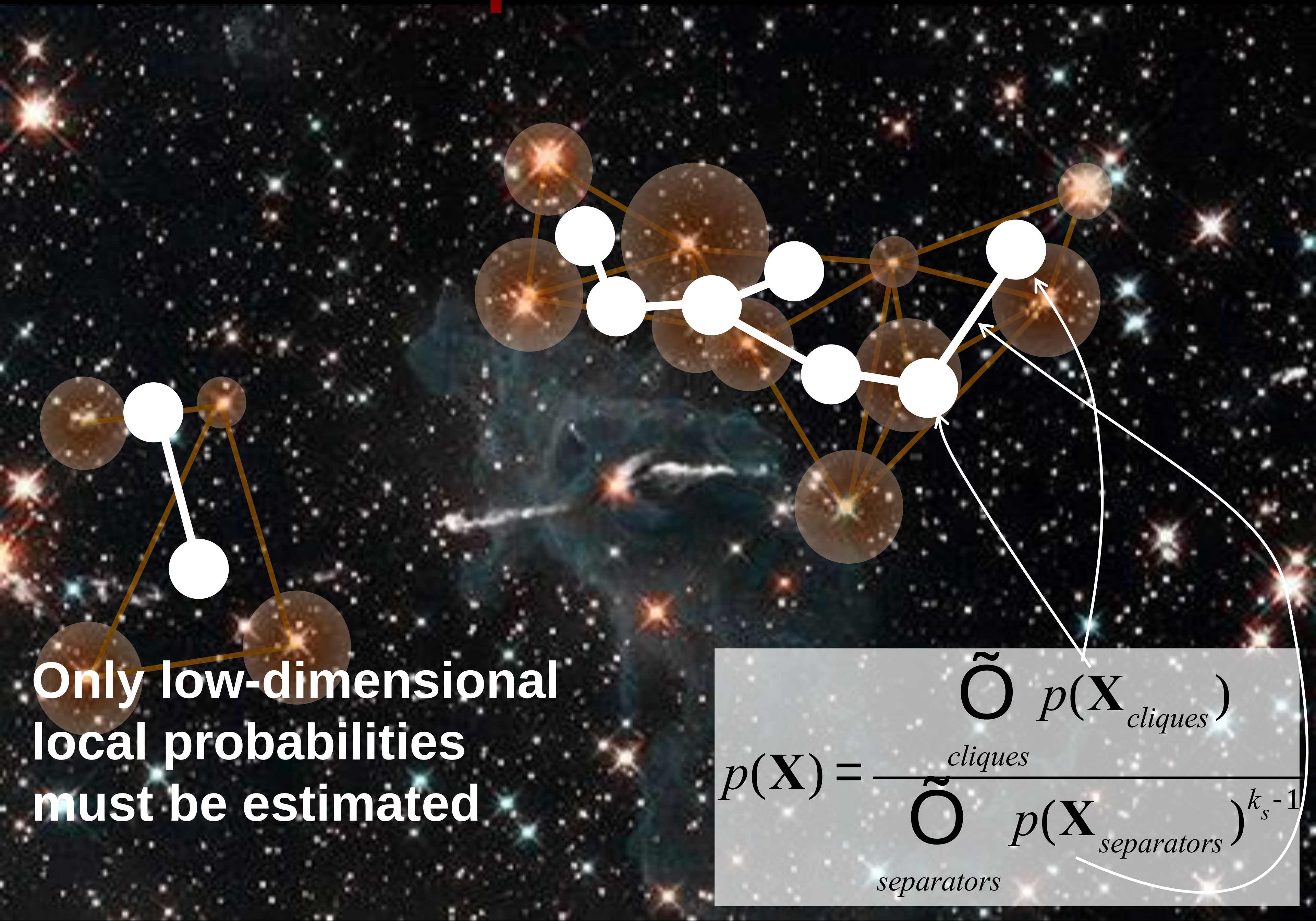
$$p(\mathbf{X}) = \frac{\underset{cliques}{\tilde{\mathrm{O}}} \; p(\mathbf{X}_{cliques})}{\underset{separators}{\tilde{\mathrm{O}}} \; p(\mathbf{X}_{separators})^{k_s - 1}}$$

Clique forest



$$p(\mathbf{X}) = \frac{\prod_{cliques} p(\mathbf{X}_{cliques})}{\prod_{separators} p(\mathbf{X}_{separators})^{k_s - 1}}$$

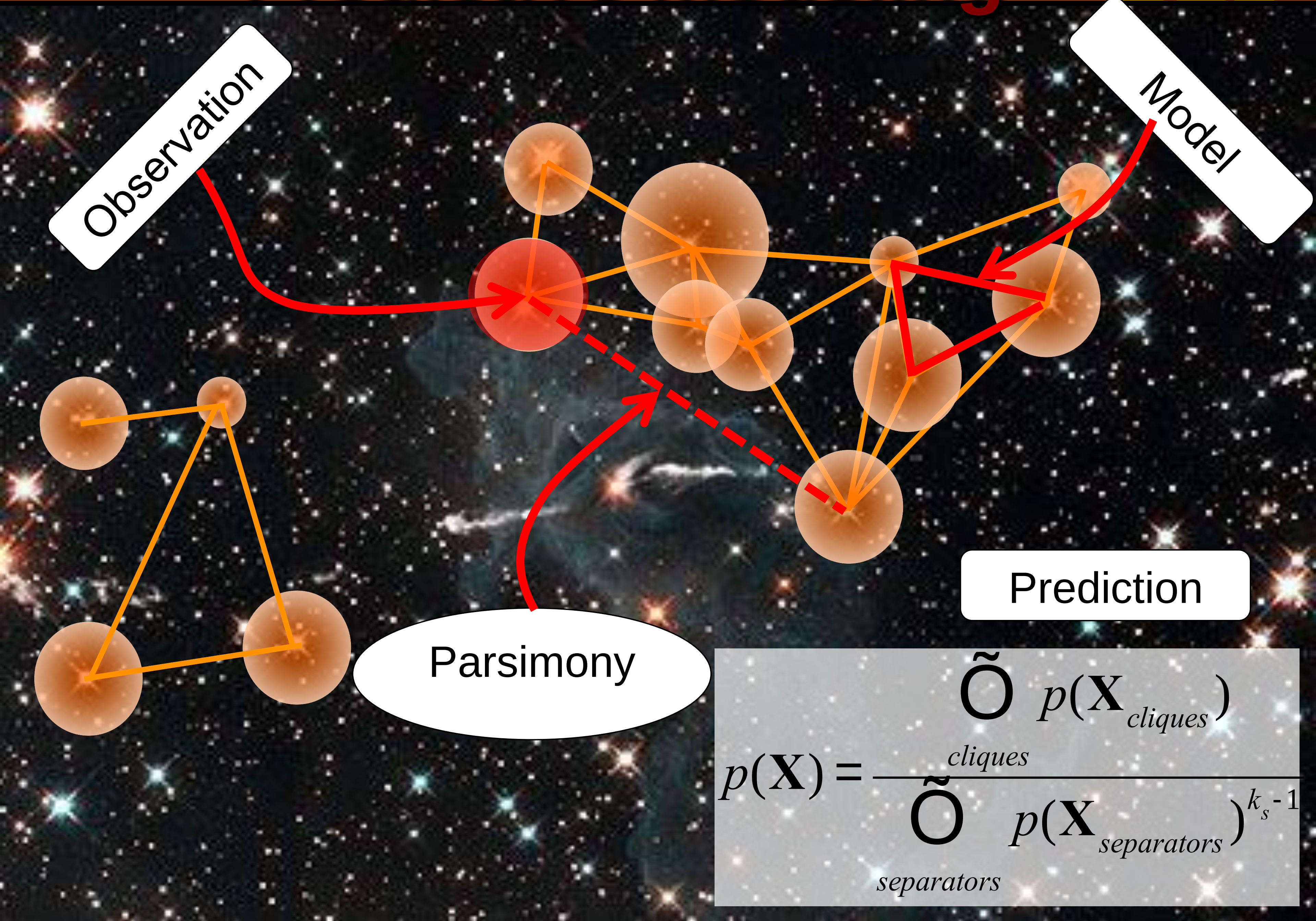
Graphical Model



Only low-dimensional local probabilities must be estimated

$$p(\mathbf{X}) = \frac{\prod_{cliques} p(\mathbf{X}_{cliques})}{\prod_{separators} p(\mathbf{X}_{separators})^{k_s - 1}}$$

Predictive modeling



Observation

Model

Prediction

Parsimony

$$p(\mathbf{X}) = \frac{\underset{cliques}{\tilde{\mathrm{O}}}\; p(\mathbf{X}_{cliques})}{\underset{separators}{\tilde{\mathrm{O}}}\; p(\mathbf{X}_{separators})^{k_s - 1}}$$

LoGo

By constraining the model to reproduce observed moments while maximizing Shannon-Gibbs entropy (maximum Entropy method), at the second order, we have that the model must be a multivariate Gaussian:

$$p(X_1,\ldots X_N) = \frac{1}{Z}\exp\Big(-\sum_{i,j} X_i \mathbf{J}_{\mathbf{i,j}} X_j\Big)$$

We keep only the significant interactions and set to zero (Max Ent.) the uncertain ones: $\mathbf{J}_{i,j} = 0$ iff X_i , X_j conditionally independent

$\mathbf{J}_{i,j}$ is sparse and it has the structure given by the information filtering network

LoGo

J$_{i,j}$ is computed form local inversion of the covariance matrix over the clique forest

$$\mathbf{J}_{\mathbf{i,j}} = \sum_{\mathit{Cliques}\ \mathrm{C}} \Sigma(C)^{-1} - \sum_{\mathit{Separators}\ \mathrm{S}} (k_S - 1)\Sigma(S)^{-1}$$

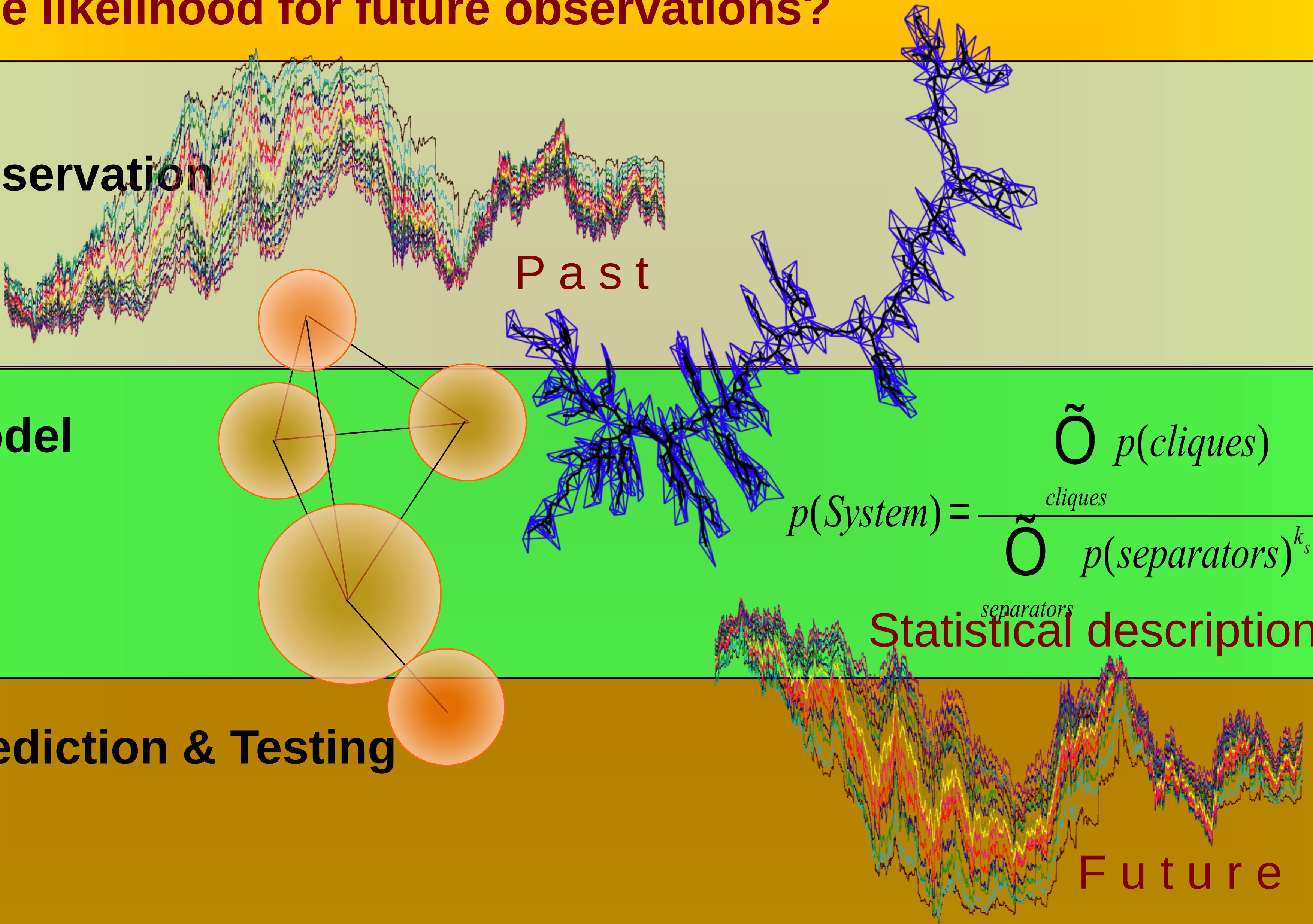
We obtain a sparse inverse covariance (our graphical model) by doing local inversion only

Super-fast algorithm O(N) *even O(logN) if parallelized*

W. Barfuss, GP Massara, T Di Matteo & TA "Parsimonious modeling with Information Filtering Networks" arXiv preprint arXiv:1602.07349 (2016).

Test:

is our model, built form past observations, associated with a large likelihood for future observations?



Observation

Model

Prediction & Testing

Past

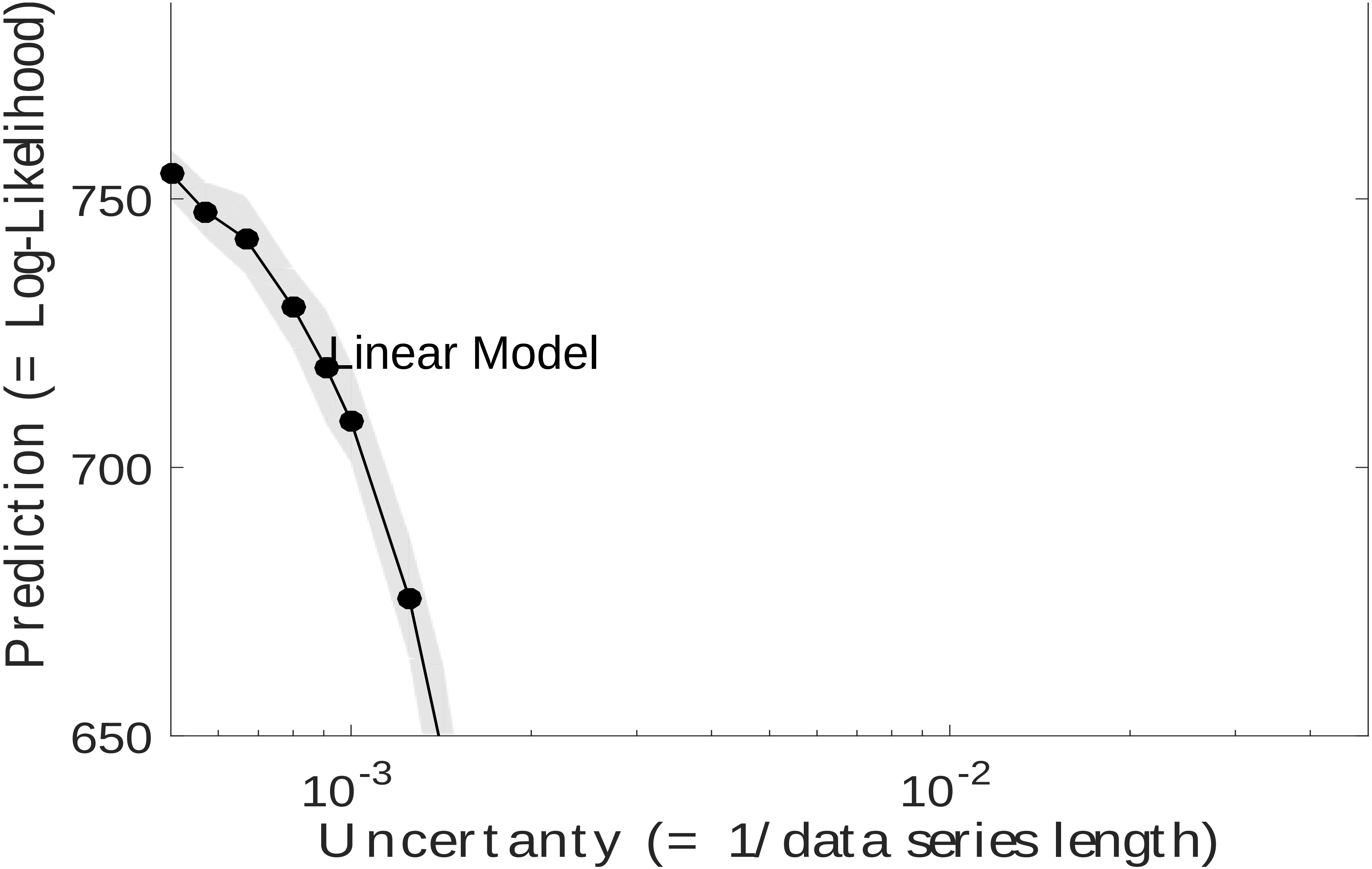
Statistical description

Future

$$p(System) = \frac{\underset{cliques}{\text{Õ}}\ p(cliques)}{\underset{separators}{\text{Õ}}\ p(separators)^{k_s - 1}}$$

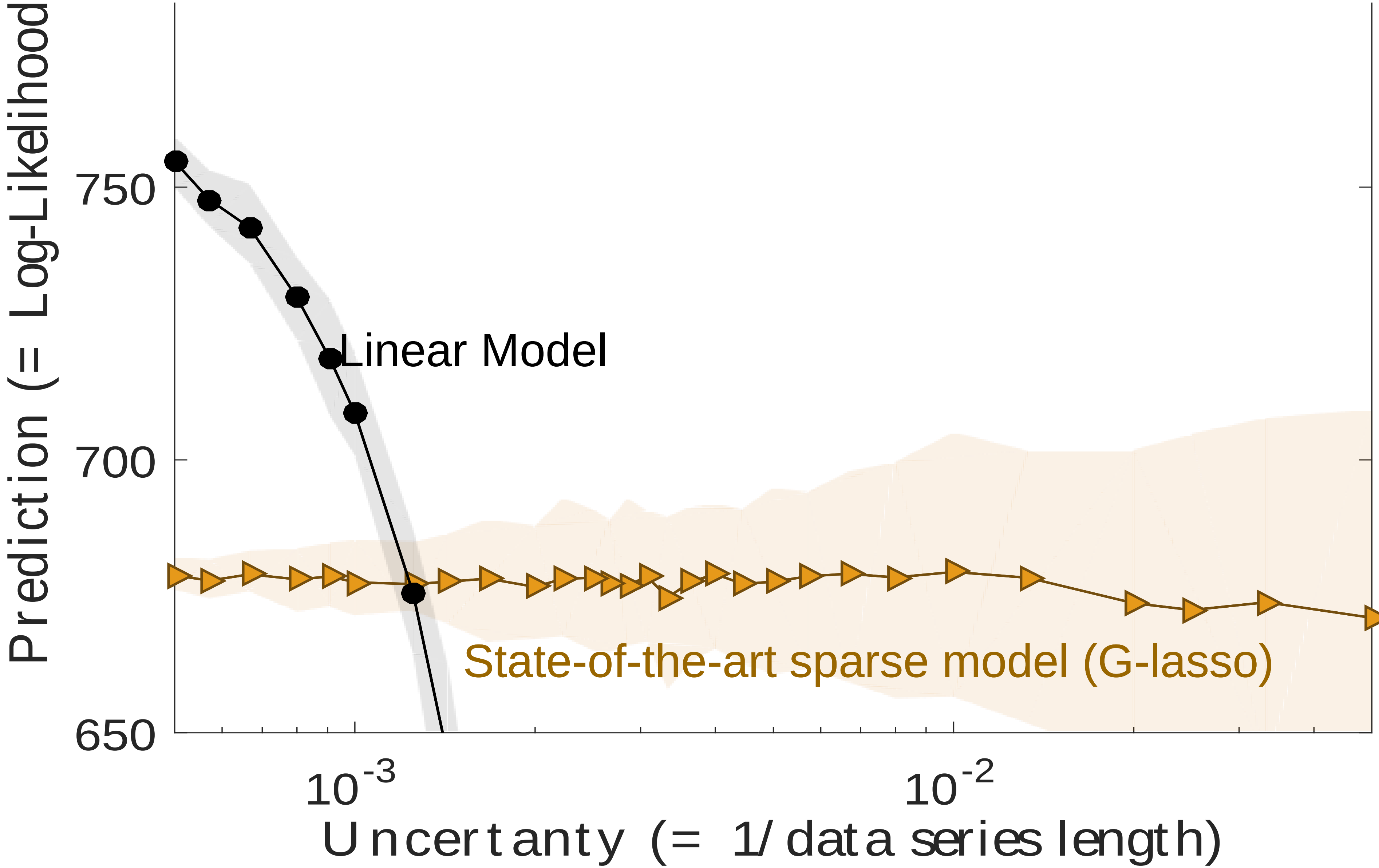
Test:

is our model, built form past observations, associated with a large likelihood for future observations?



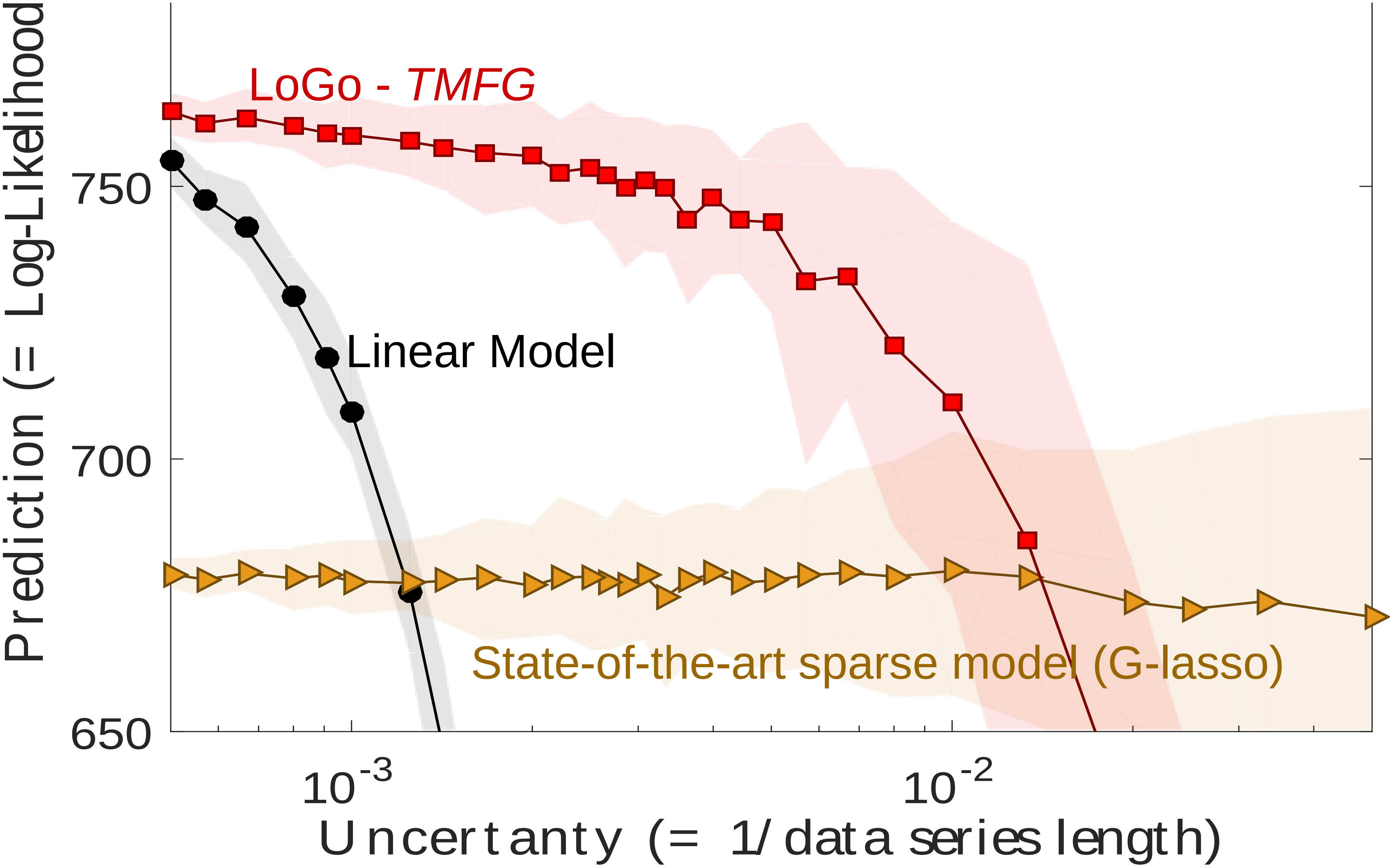
Test:

is our model, built form past observations, associated with a large likelihood for future observations?



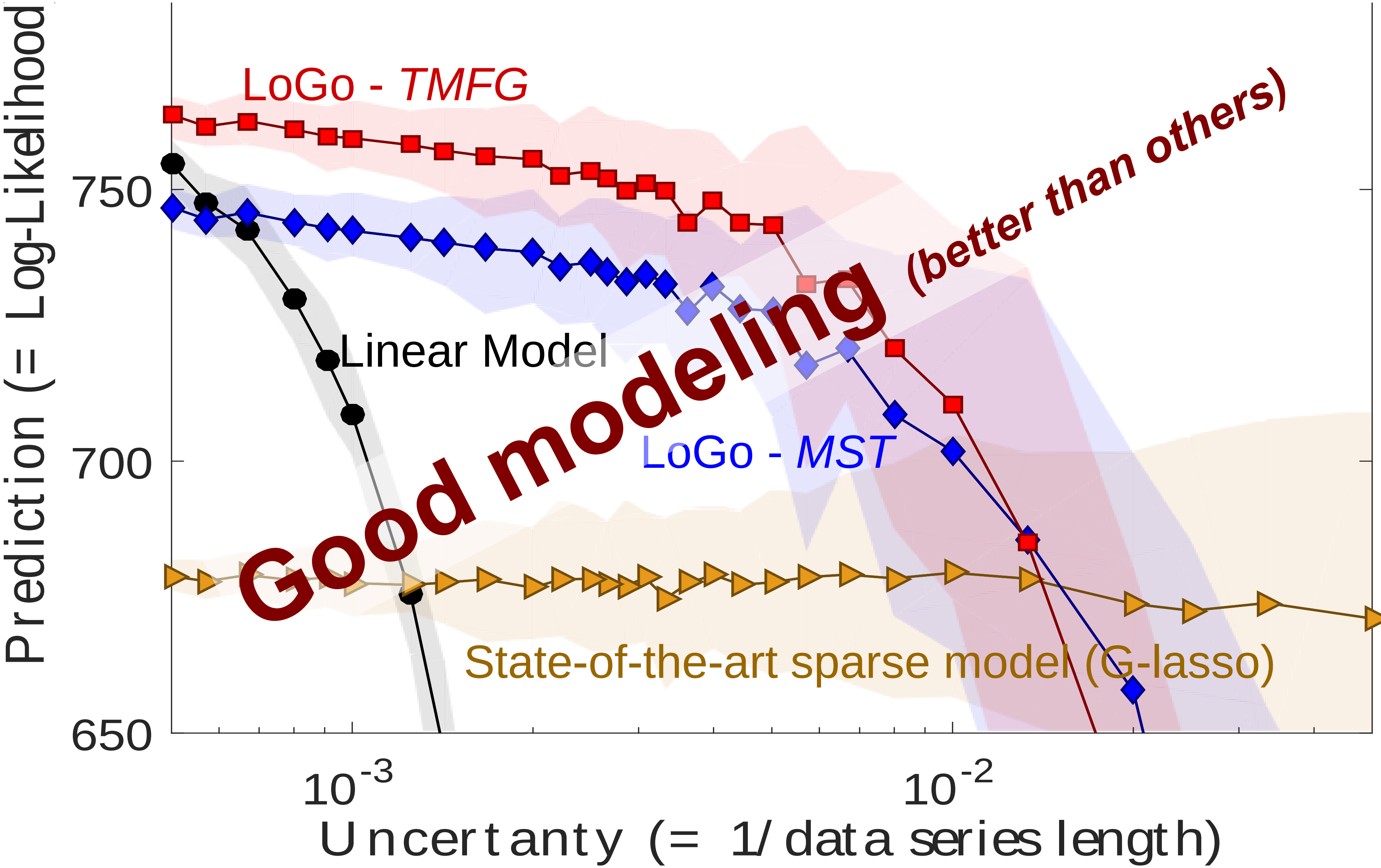
Test:

is our model, built form past observations, associated with a large likelihood for future observations?



Test:

is our model, built form past observations, associated with a large likelihood for future observations?



Prediction

In which sense we predict?

With $p(X_{future}|X_{past})$ we can predict the future

This is the same as (linear) regression

$$E[X_B | X_A] = \sum_{X_B} X_B p(X_B | X_A) = -\mathbf{J}_{\mathbf{BB}}^{-1} \mathbf{J}_{\mathbf{BA}} X_A$$

and also Granger causality (2x Transfer entropy)

$$G(X_A \to X_B) = \log|\mathbf{J}_{\mathbf{B^-B^-}}| - \log|\mathbf{J}_{\mathbf{B^-B^-}} - \mathbf{J}_{\mathbf{B^-A^-}} \mathbf{J}_{\mathbf{A^-A^-}}^{-1} \mathbf{J}_{\mathbf{A^-B^-}}|$$

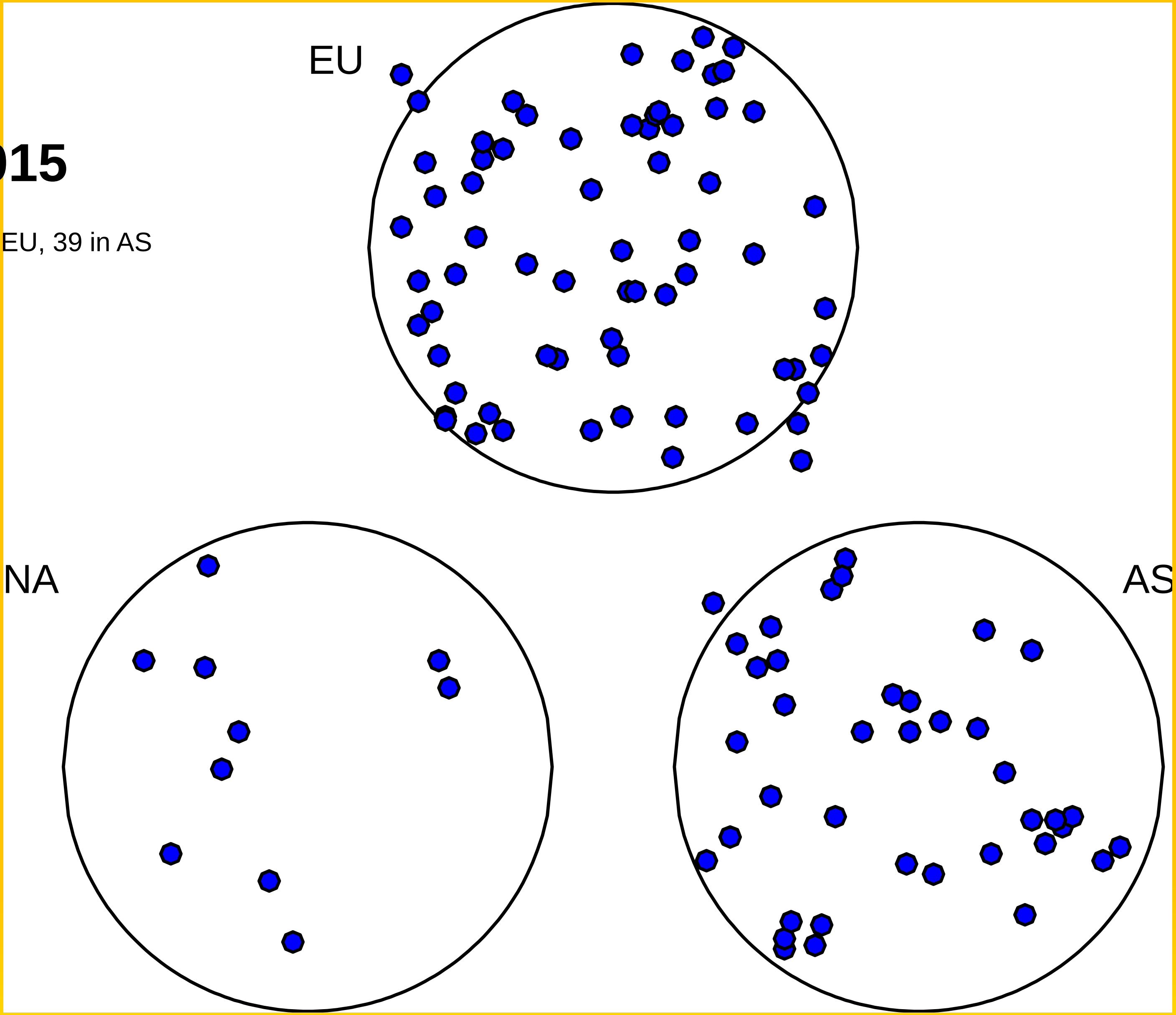
The advantage is that we have a **sparse model** computed in a very efficient way applicable to big-data predictive analytics

Test:

Uncertainty spillover across regions in banking system

2005-2015

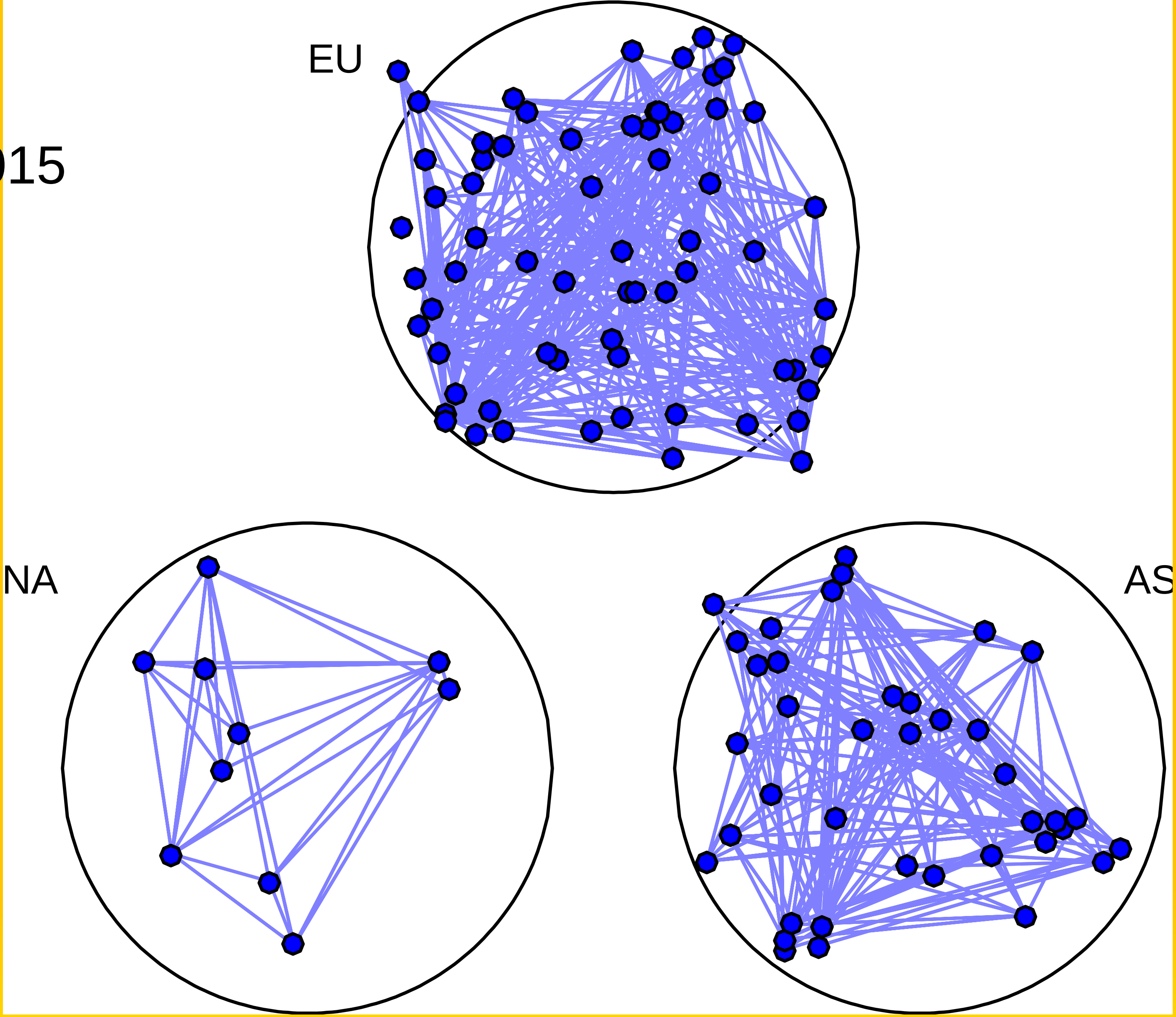
115 banks:
10 in NA , 66 in EU, 39 in AS



Test:

Uncertainty spillover across regions in banking system

2005-2015

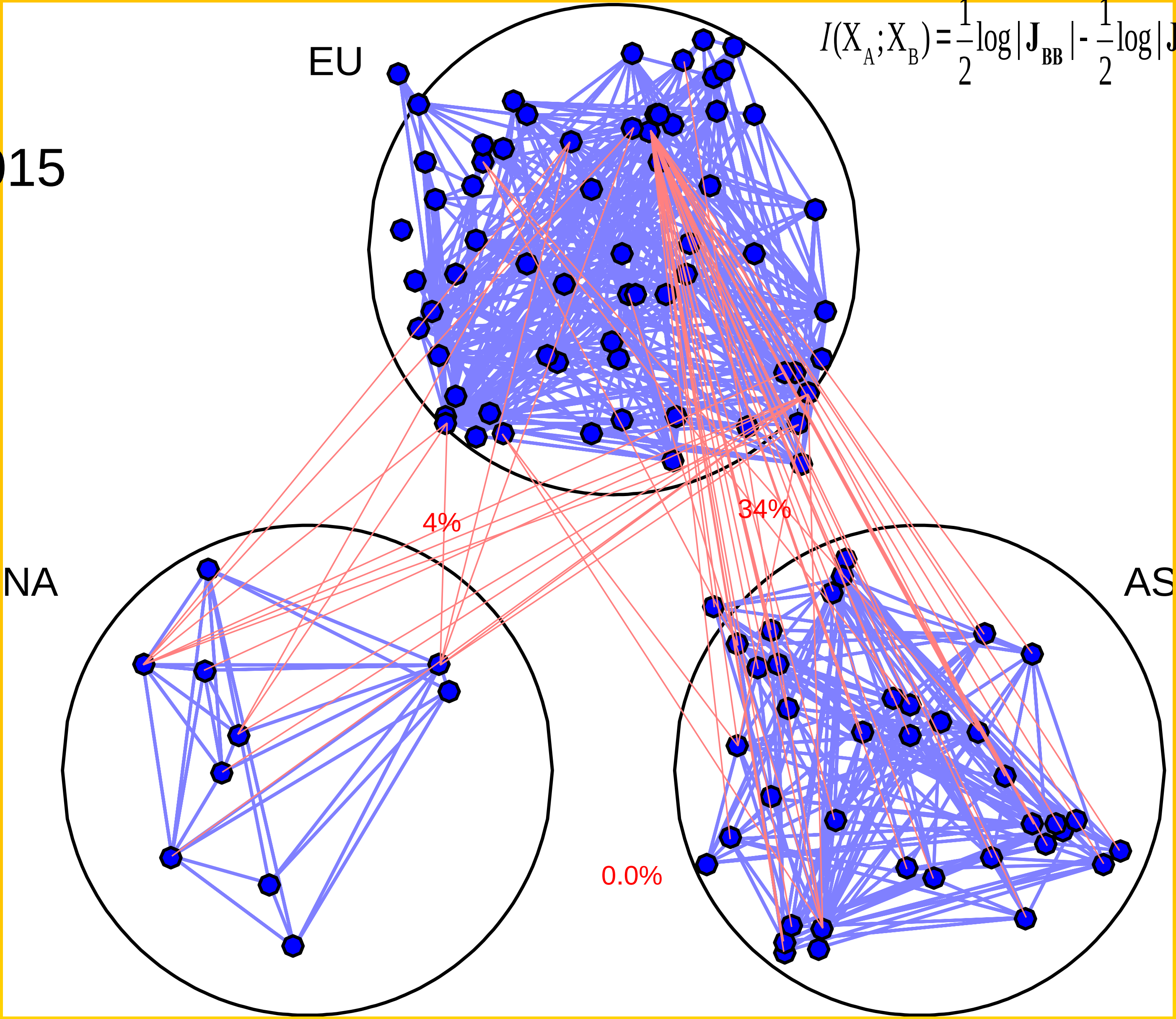


Test:

Uncertainty spillover across regions in banking system

$$I(X_A;X_B)=\frac{1}{2}\log|\mathbf{J}_{BB}|-\frac{1}{2}\log|\mathbf{J}_{BB}-\mathbf{J}_{BA}\mathbf{J}_{AA}^{-1}\mathbf{J}_{AB}|$$

2005-2015



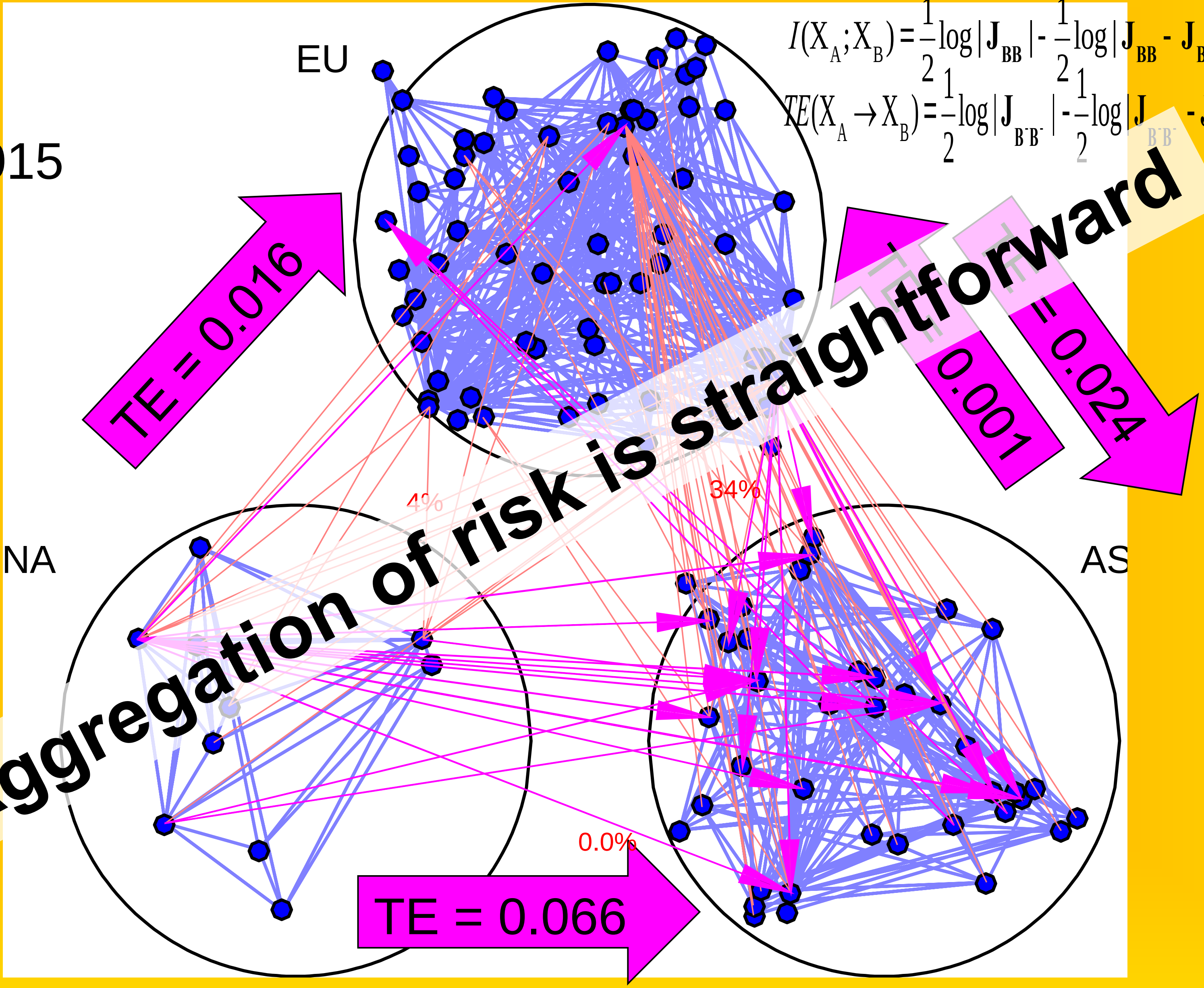
Test:

Uncertainty spillover across regions in banking system

2005-2015

$$I(X_A;X_B)=\frac{1}{2}\log|\mathbf{J}_{BB}|-\frac{1}{2}\log|\mathbf{J}_{BB}-\mathbf{J}_{BA}\mathbf{J}_{AA}^{-1}\mathbf{J}_{AB}|$$

$$TE(X_A\rightarrow X_B)=\frac{1}{2}\log|\mathbf{J}_{B^-B^-}|-\frac{1}{2}\log|\mathbf{J}_{B^-B^-}-\mathbf{J}_{B^-A^-}\mathbf{J}_{A^-A^-}^{-1}\mathbf{J}_{A^-B^-}|$$



Aggregation of risk is straightforward

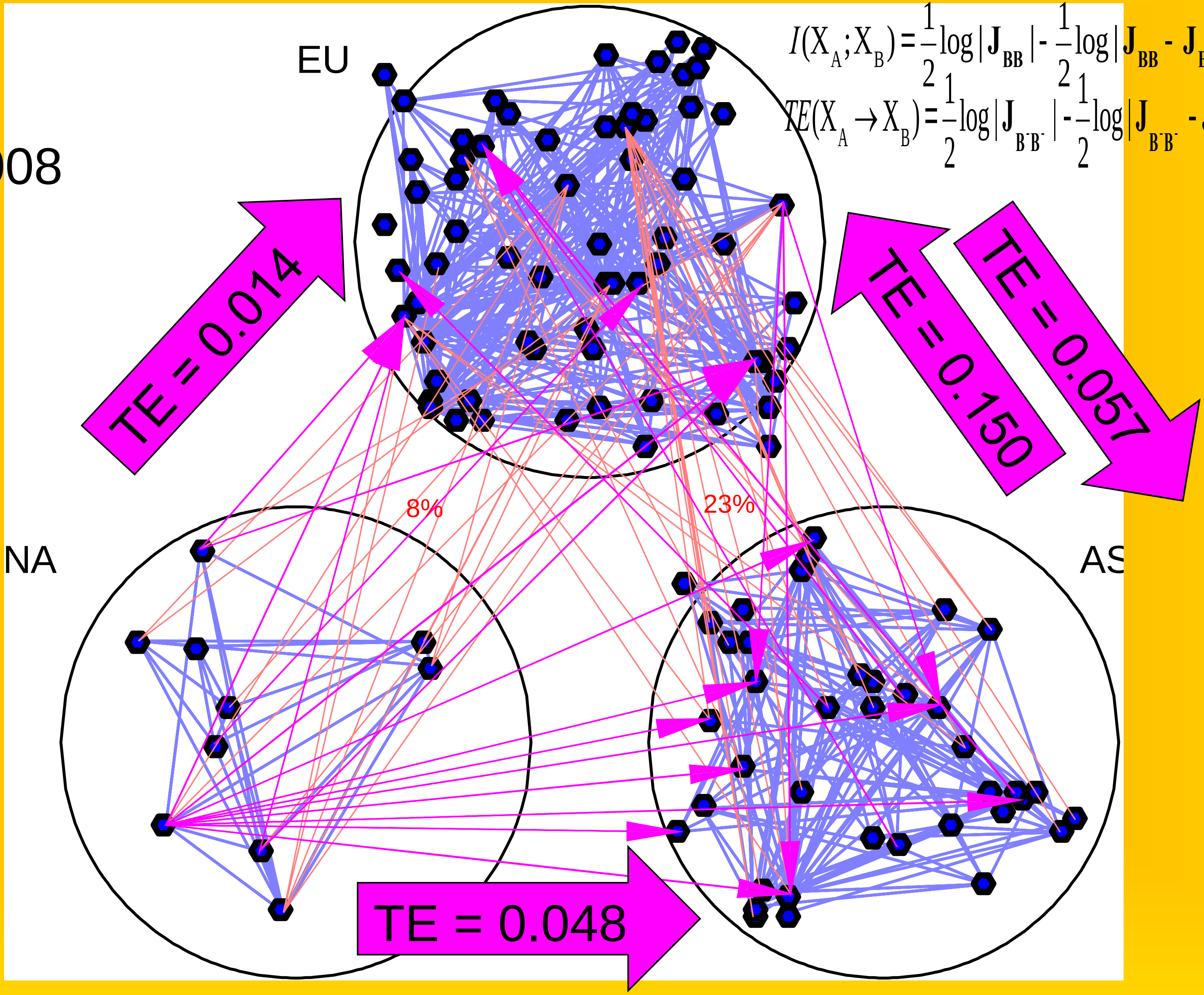
Test:

Uncertainty spillover across regions in banking system

$$I(X_A;X_B)=\frac{1}{2}\log|\mathbf{J}_{BB}|-\frac{1}{2}\log|\mathbf{J}_{BB}-\mathbf{J}_{BA}\mathbf{J}^{-1}_{AA}\mathbf{J}_{AB}|$$

$$TE(X_A\rightarrow X_B)=\frac{1}{2}\log|\mathbf{J}_{B^-B^-}|-\frac{1}{2}\log|\mathbf{J}_{B^-B^-}-\mathbf{J}_{B^-A^-}\mathbf{J}^{-1}_{A^-A^-}\mathbf{J}_{A^-B^-}|$$

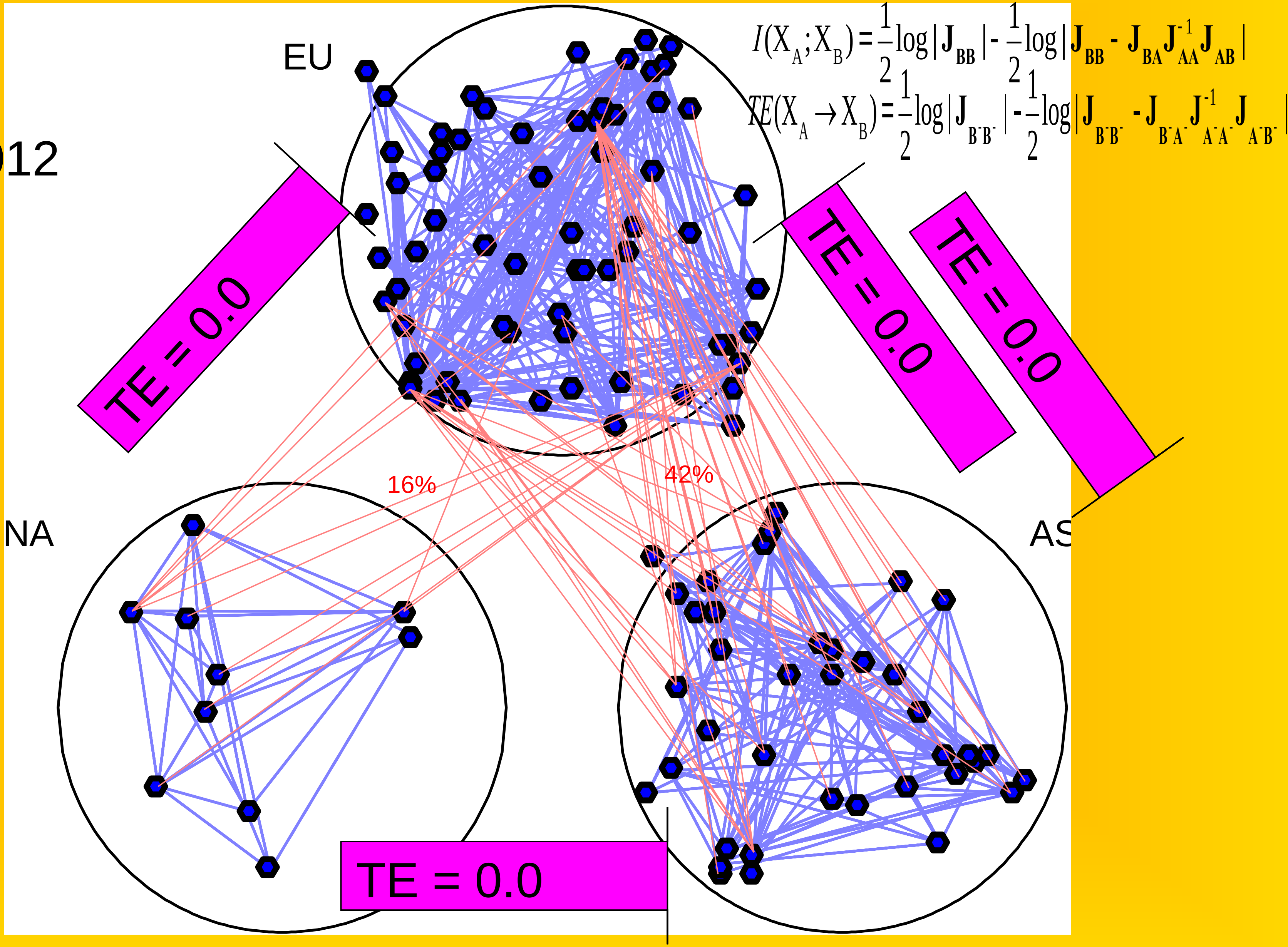
2005-2008



Test:

Uncertainty spillover across regions in banking system

2008-2012



$$I(X_A;X_B) = \frac{1}{2}\log|\mathbf{J}_{BB}| - \frac{1}{2}\log|\mathbf{J}_{BB} - \mathbf{J}_{BA}\mathbf{J}_{AA}^{-1}\mathbf{J}_{AB}|$$

$$TE(X_A \rightarrow X_B) = \frac{1}{2}\log|\mathbf{J}_{B^{-}B^{-}}| - \frac{1}{2}\log|\mathbf{J}_{B^{-}B^{-}} - \mathbf{J}_{B^{-}A^{-}}\mathbf{J}_{A^{-}A^{-}}^{-1}\mathbf{J}_{A^{-}B^{-}}|$$

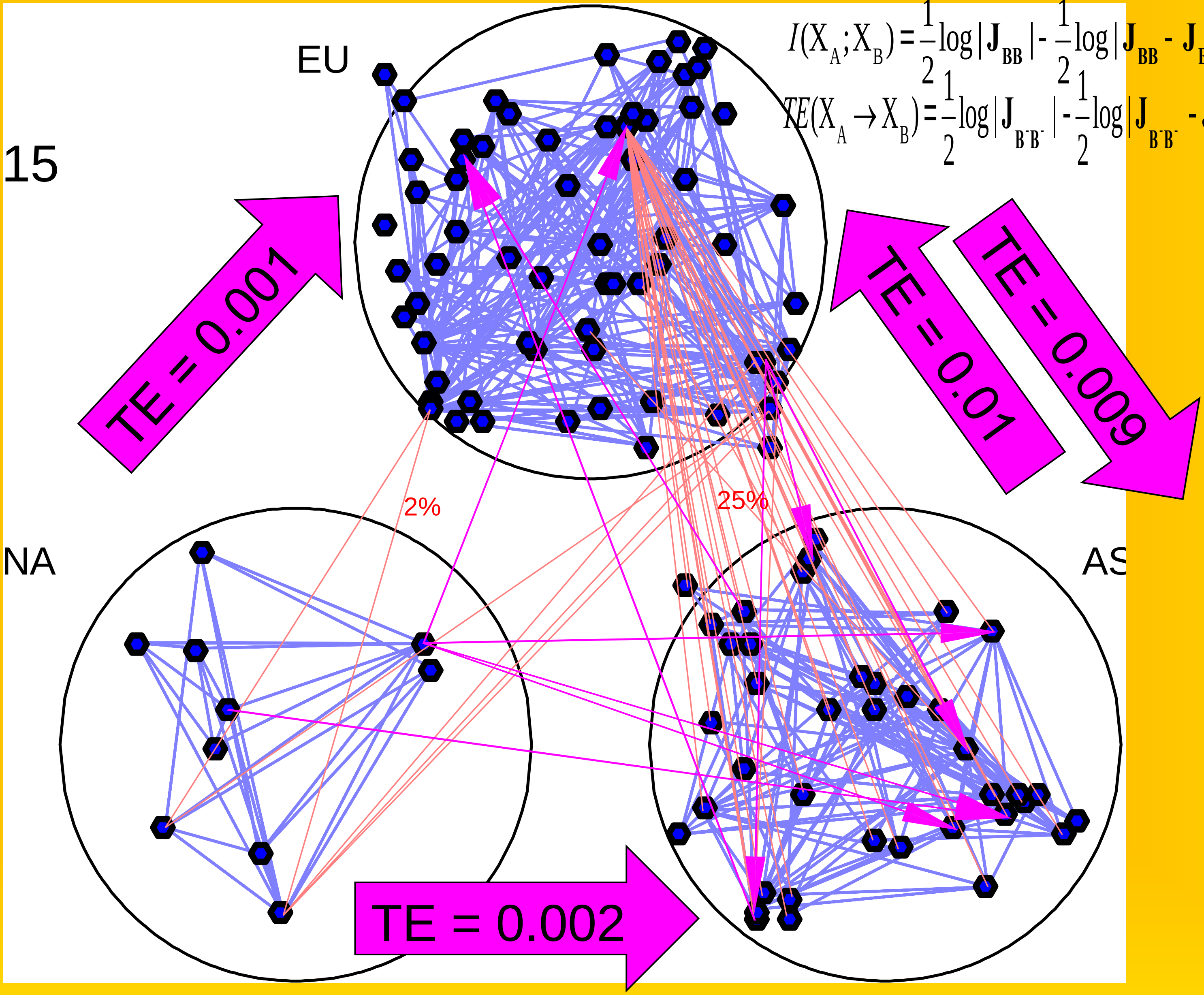
Test:

Uncertainty spillover across regions in banking system

2012-2015

$$I(X_A;X_B)=\frac{1}{2}\log|\mathbf{J}_{BB}|-\frac{1}{2}\log|\mathbf{J}_{BB}-\mathbf{J}_{BA}\mathbf{J}_{AA}^{-1}\mathbf{J}_{AB}|$$

$$TE(X_A\rightarrow X_B)=\frac{1}{2}\log|\mathbf{J}_{B^-B^-}|-\frac{1}{2}\log|\mathbf{J}_{B^-B^-}-\mathbf{J}_{B^-A^-}\mathbf{J}_{A^-A^-}^{-1}\mathbf{J}_{A^-B^-}|$$



Networks for Prediction

With $p(X_B|X_{A-})$ we can quantify probability of future events

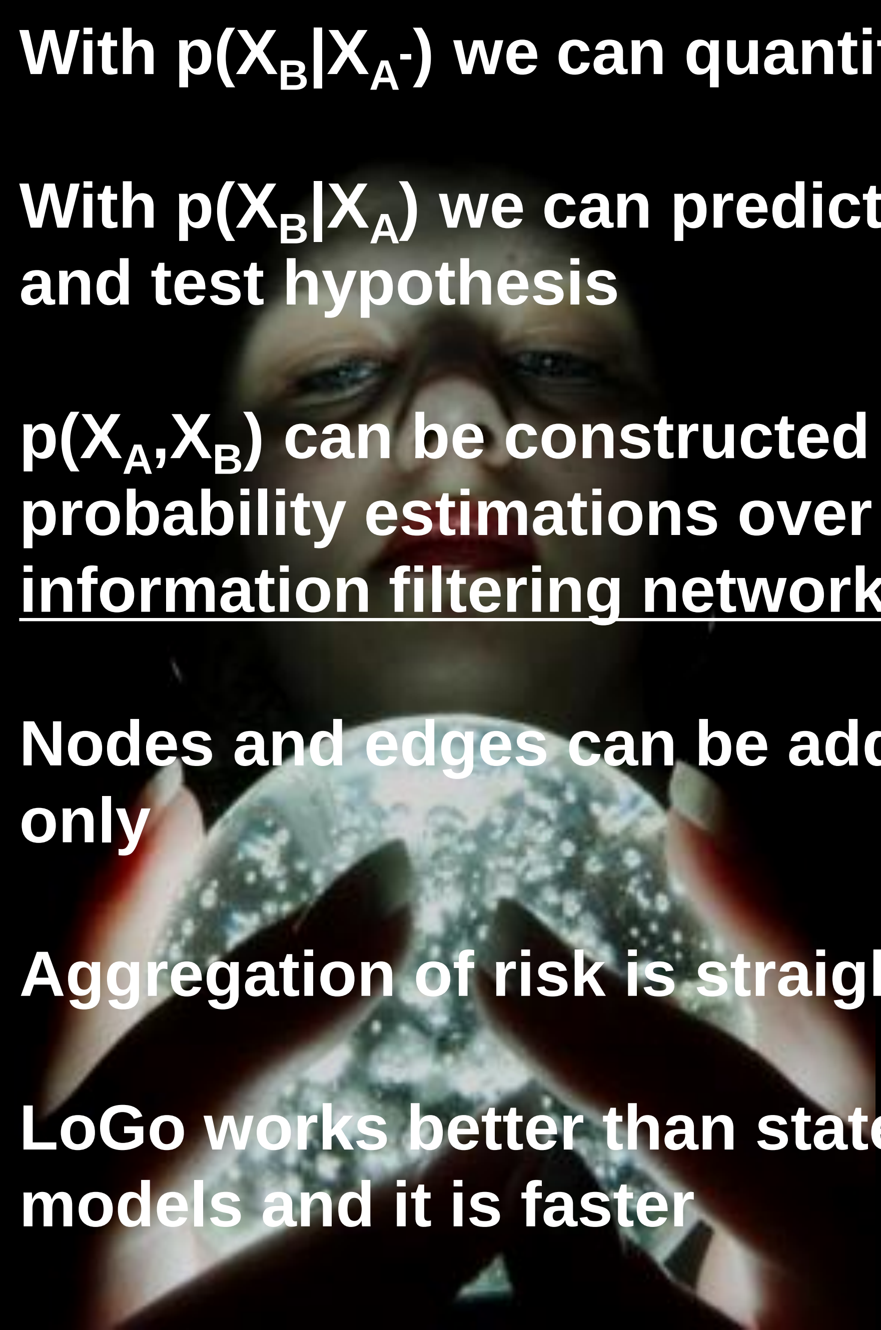
With $p(X_B|X_A)$ we can predict impact of unobserved scenarios and test hypothesis

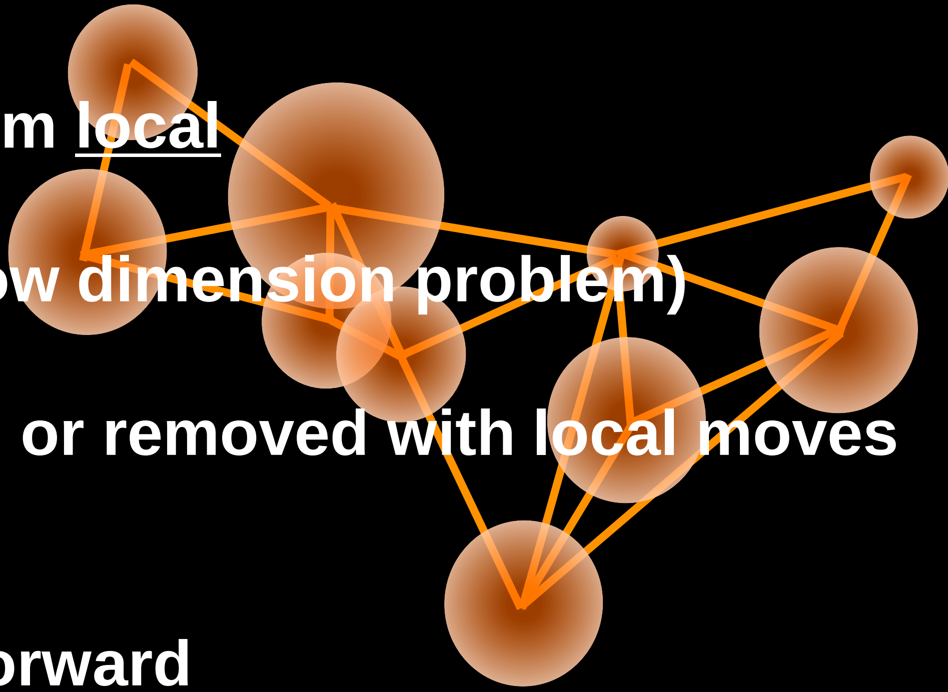
$p(X_A,X_B)$ can be constructed from local probability estimations over an information filtering network (low dimension problem)

Nodes and edges can be added or removed with local moves only

Aggregation of risk is straightforward

LoGo works better than state-of-the-art sparse graphical models and it is faster





Thank YOU!

W. Barfuss, GP Massara, T Di Matteo & TA “Parsimonious modeling with Information Filtering Networks” arXiv preprint arXiv:1602.07349 (2016).

Massara, Guido Previde, Tiziana Di Matteo, and TA. "Network Filtering for Big Data: Triangulated Maximally Filtered Graph" Journal of Comlex Networks (2016) arXiv preprint arXiv:1505.02445 (2015).

http://www.cs.ucl.ac.uk/staff/tomaso_aste/

http://fincomp.cs.ucl.ac.uk/

http://blockchain.cs.ucl.ac.uk/

Si l’ordre satisfait la raison, le désordre fait les délices de l’imagination
Paul Claudel