For distribution purposes, copyright figures have been removed.

Networks from correlation matrices: An alternative to thresholding

Naoki Masuda
University of Bristol, Department of Engineering Mathematics
[Moving to State University of New York at Buffalo, Department of Mathematics, mid August]

Threshold Networks
University of Nottingham
22 July 2019

A typical work flow

e.g. schizophrenia

(Lynall, Bassett, Kerwin, McKenna, Kitzbichler, Muller & Bullmore, JNS 2010)

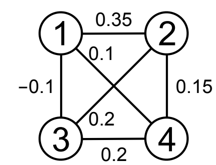
Correlation matrix

- A major form of multivariate data
- Connection between *i* and *j* = strength of correlation between *i* and *j*
- Pearson (or other) correlation
- Measurements: time series, feature vector, questionnaires across human/animal subjects etc.
    - Financial time series, neuroimaging, genetic data, psychology experiments, climate networks etc.
- Conventional approaches
    - Principal component analysis
    - Factor analysis
    - Random matrix theory
    - Markowitz's optimal portfolio

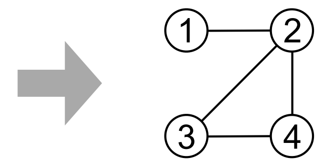
Correlation networks

- To benefit from network analysis tools
- Two conventional approaches
  1. Thresholding
  2. Weighted networks

$$\begin{pmatrix} 1 & 0.35 & -0.1 & 0.1 \\ 0.35 & 1 & 0.2 & 0.15 \\ -0.1 & 0.2 & 1 & 0.2 \\ 0.1 & 0.15 & 0.2 & 1 \end{pmatrix} =$$

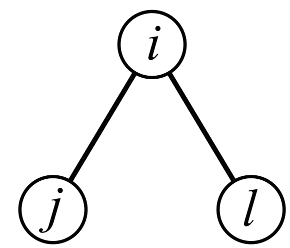


correlation matrix



thresholded network

- Brain networks, financial networks, …
- Problems:
  - False positives (pseudo correlation)
  - Loss of information by thresholding
  - Arbitrariness of the threshold choice
  - Negative correlation in the case of weighted networks



Grand goal

- Analyse correlation matrices as they are, and also exploit various network concepts/algorithms
  - Without thresholding or discarding negatively correlated node pairs

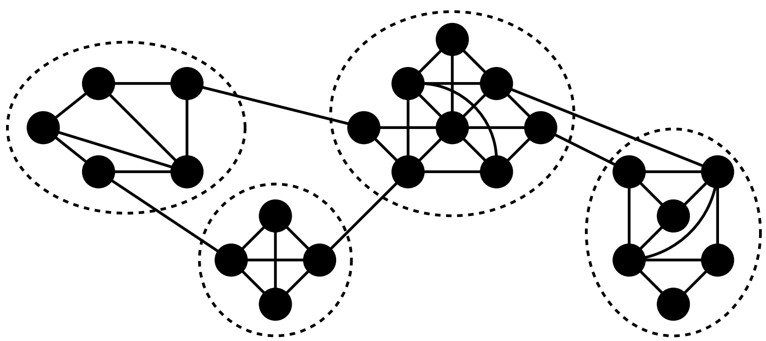
PHYSICAL REVIEW X **5,** 021006 (2015)

Community Detection for Correlation Matrices

Mel MacMahon[1] and Diego Garlaschelli[1,2]

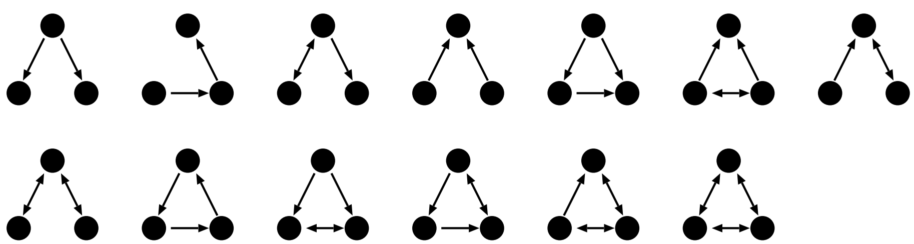
- Null model = identity matrix, or random matrices based on random matrix theory

- However, we usually say that a network has a property *α* relative to the “configuration model” (degree-preserving random graph)

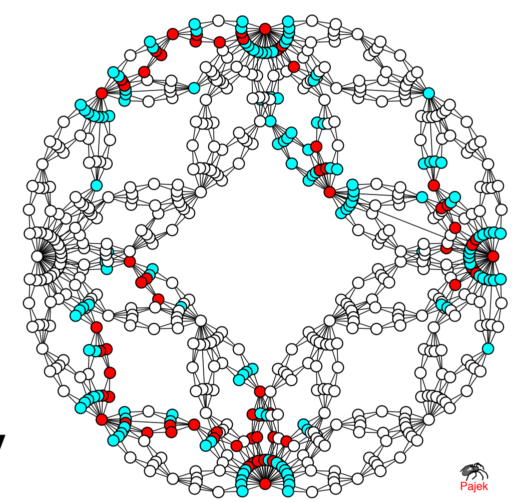


community structure

“rich clubs” Colizza et al. (2006)



network motifs



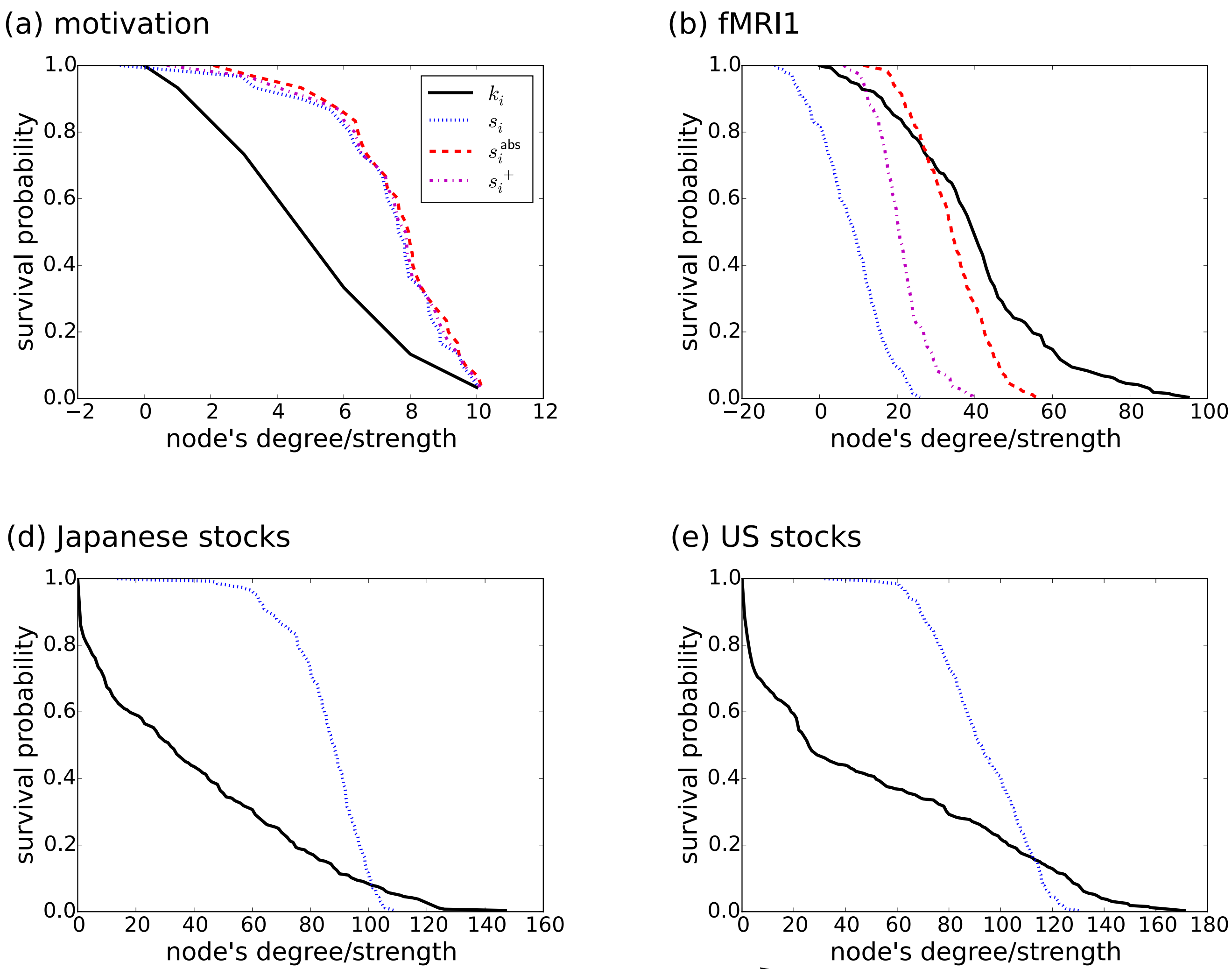
fractality

Aim

- Propose a configuration model for correlation matrices
- Strategy: maximum entropy
- Then, use it for community detection, measurement of clustering coefficients, network filtering etc., hopefully to enrich analyses of correlation matrix data

Methods in short

- Distribution of correlation/covariance matrices
- Boltzmann distribution (exponential family)
  - cf. exponential random graph models
- Constraints
  - Weighted (and signed) degree of each node is conserved.



$$s_i = \sum_{j=1;j\neq i}^{N} \mathrm{Cor}_{ij}$$

$$s_i^{\mathrm{abs}} = \sum_{j=1;j\neq i}^{N} |\mathrm{Cor}_{ij}|$$

$$s_i^{+} = \sum_{j=1;j\neq i;\mathrm{Cor}_{ij}>0}^{N} \mathrm{Cor}_{ij}$$

$$s_i \approx s_i^{\mathrm{abs}} \approx s_i^{+}$$

because $\mathrm{Cor}_{ij}^{\mathrm{org}} \geq 0$ for all but three pairs of nodes

✔ Not long-tailed but moderately heterogeneous

- Work on covariance (as opposed to correlation) matrices basically, as that is mathematically easier.

$$\mathrm{Cor}_{ij} = \frac{\mathrm{Cov}_{ij}}{\sqrt{\mathrm{Cov}_{ii}\mathrm{Cov}_{jj}}}$$

$$\mathrm{Cov}^{\mathrm{con}} = (\mathrm{Cov}^{\mathrm{con}}_{ij}) = \frac{1}{L} X X^{\top}$$

Equivalently

$$\mathrm{Cov}^{\mathrm{con}}_{ij} = \frac{1}{L} \sum_{\ell=1}^{L} x_{i\ell} x_{j\ell}$$

And consider the distribution of X

Rationales:

1. Realistic
2. Mathematically feasible

Maxent preserving node strength

$$H(X) \equiv -\int p(X)\ln p(X)\mathrm{d}X + \sum_{i=1}^{N}\alpha_i\left[\int \mathrm{Cov}_{ii}^{\mathrm{con}}p(X)\mathrm{d}X - \mathrm{Cov}_{ii}^{\mathrm{org}}\right]$$

$$+\sum_{i=1}^{N}\beta_i\left[\int \sum_{j=1;j\neq i}^{N}\mathrm{Cov}_{ij}^{\mathrm{con}}p(X)\mathrm{d}X - \sum_{j=1;j\neq i}^{N}\mathrm{Cov}_{ij}^{\mathrm{org}}\right],$$

$$p(X) \propto \exp\left[\frac{1}{L}\sum_{\ell=1}^{L}\left(\sum_{i=1}^{N}\alpha_i x_{i\ell}^2 + \sum_{i=1}^{N}\beta_i\sum_{j=1;j\neq i}^{N}x_{i\ell}x_{j\ell}\right)\right]$$

$$= \prod_{\ell=1}^{L}\exp\left[\frac{1}{L}\left(\sum_{i=1}^{N}\alpha_i x_{i\ell}^2 + \sum_{i=1}^{N}\beta_i\sum_{j=1;j\neq i}^{N}x_{i\ell}x_{j\ell}\right)\right]$$

$$= \prod_{\ell=1}^{L}\exp\left[-\frac{1}{2}\boldsymbol{x}_\ell^\top\Sigma^{-1}\boldsymbol{x}_\ell\right],$$

Reminder: $\mathrm{Cov}^{\mathrm{con}} = (\mathrm{Cov}_{ij}^{\mathrm{con}}) = \frac{1}{L}XX^\top$

$$\mathrm{Cov}_{ij}^{\mathrm{con}} = \frac{1}{L}\sum_{\ell=1}^{L}x_{i\ell}x_{j\ell}$$

$$p(X) \propto \prod_{\ell=1}^{L} \exp\left[-\frac{1}{2}\boldsymbol{x}_\ell^\top \Sigma^{-1} \boldsymbol{x}_\ell\right]$$

where $\boldsymbol{x}_\ell = (x_{1\ell}, \ldots, x_{N,\ell})^\top$ and

$$\Sigma^{-1} = -\frac{1}{L}\begin{pmatrix} 2\alpha_1 & \beta_1+\beta_2 & \beta_1+\beta_3 & \cdots & \beta_1+\beta_{N-1} & \beta_1+\beta_N \\ \beta_2+\beta_1 & 2\alpha_2 & \beta_2+\beta_3 & \cdots & \beta_2+\beta_{N-1} & \beta_2+\beta_N \\ & & \vdots & & & \\ \beta_N+\beta_1 & \beta_N+\beta_2 & \cdots & \cdots & \beta_N+\beta_{N-1} & 2\alpha_N \end{pmatrix}$$

So multivariate normal distribution:

$$p(X) = \prod_{\ell=1}^{L} \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \exp\left[-\frac{1}{2}\boldsymbol{x}_\ell^\top \Sigma^{-1} \boldsymbol{x}_\ell\right]$$

To infer parameters

- Gradient descent mehtod

$$\alpha_i^{\text{new}} = \alpha_i^{\text{old}} + \epsilon \left(\Sigma_{ii} - \text{Cov}_{ii}^{\text{org}} \right),$$

$$\beta_i^{\text{new}} = \beta_i^{\text{old}} + 2\epsilon \left(\sum_{j=1}^{N} \Sigma_{ij} - \sum_{j=1}^{N} \text{Cov}_{ij}^{\text{org}} \right),$$

- Convex optimisation
    - Faster and more accurate
    - Documentation and code on Github (not in the first PRE paper but in the second paper on arXiv)

Sadamori Kojaku

Numerical simulations: Aim

- Strength (i.e. weighted degree) of each node really preserved?
- Speed
- Comparison with other null models for correlation matrices (e.g. MacMahon & Garlaschelli, PRX 2015)
- Deployment to algorithms or statistical tests that need a null model

Data

- Psychological questionnaires; *N*=30 questions, *L*=686 participants
- Functional MRI from healthy participants; *N*=264 locations, *L*=4760 time points
- Stocks, Japan
  - *N*=264 stocks belonging to the first section of the Tokyo Stock Exchange provided by Nikkei NEEDS
  - from 12 March 1996 to 29 February 2016
  - *L*=4904 trading days (daily observations)
- Stocks, US
  - *N*=325 stocks from Standard & Poor's 500 index using MATHEMATICA FinancialData package
  - from 3 January 1996 to 24 February 2017
  - *L*=5324 trading days

Yukie Sano

H-Q-S null model

- Hirschberger-Qu-Steuer, Eur J Oper Res, 177, 1610 (2007)
- For covariance matrices
- Preserves
  - mean of the on-diagonal entries
  - mean of the off-diagonal entries
  - variance of the off-diagonal entries
- Corresponds to Erdős-Rényi random graph

Null models by MacMahon & Garlaschelli, PRX 2015

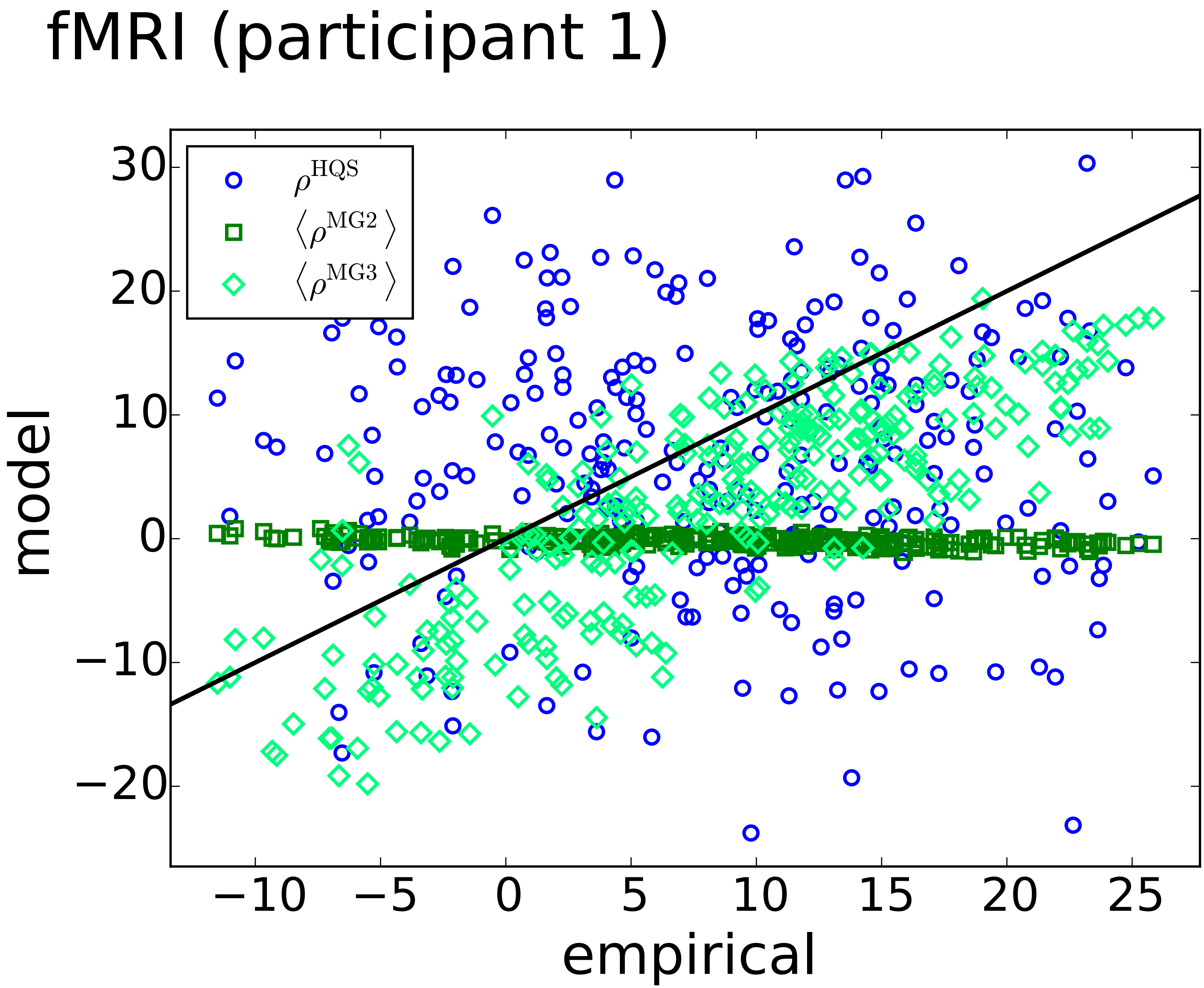
Marcenko-Pastur (also called Senguta-Mitra) distribution:

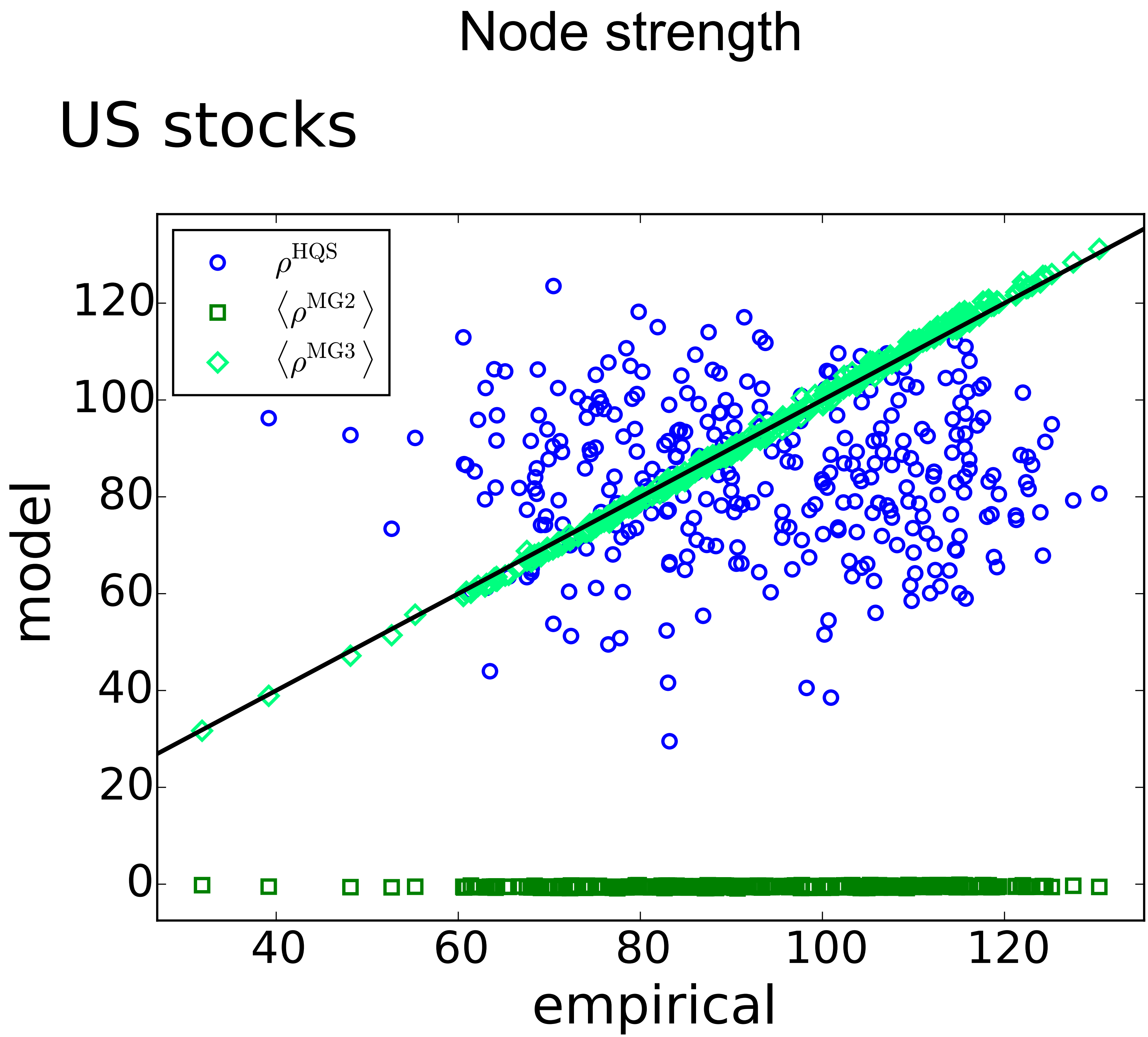
$$\rho(\lambda) = \begin{cases} \frac{L}{N}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda} & (\lambda_- \leq \lambda \leq \lambda_+), \\ 0 & (\text{otherwise}), \end{cases}$$

$$\text{where } \lambda_\pm = \left(1 \pm \sqrt{N/L}\right)^2$$

- Model MG2: Preserves modes whose eigenvalues $\lambda \in [\lambda_-, \lambda_+]$
- Model MG3: Preserves modes with $\lambda \in [\lambda_-, \lambda_+]$ & λ_{max}
- Not generating (random) correlation *matrices*

Node strength





Node strength
US stocks
ρ^{HQS}
$\langle \rho^{\mathrm{MG2}} \rangle$
$\langle \rho^{\mathrm{MG3}} \rangle$
model
empirical
0
20
40
60
80
100
120
40
60
80
100
120

Clustering coefficient

Onnela et al. (Phys Rev E 2005) for weighted networks

Masuda et al. (Frontiers in Neuroinformatics 2018) for correlation matrices

	$C^{\text{wei,O}}$			$C^{\text{cor,M}}$		
Null model	C	Z	P	C	Z	P
Motivation	0.284 (empirical)			0.031 (empirical)		
Configuration	0.503 ± 0.024	-9.23	$<10^{-3}$	0.022 ± 0.001	6.20	$<10^{-3}$
H-Q-S	0.335 ± 0.034	-1.51	0.130	0.029 ± 0.003	0.59	0.557
White-noise	0.111 ± 0.015	11.26	$<10^{-3}$	0.001 ± 0.000	869.28	$<10^{-3}$
fMRI1	0.096 (empirical)			0.013 (empirical)		
Configuration	0.138 ± 0.004	-10.11	$<10^{-3}$	0.003 ± 0.000	85.90	$<10^{-3}$
H-Q-S	0.127 ± 0.005	-6.29	$<10^{-3}$	0.023 ± 0.000	-59.50	$<10^{-3}$
White-noise	0.078 ± 0.005	3.66	$<10^{-3}$	0.000 ± 0.000	23346.41	$<10^{-3}$
Japan	0.413 (empirical)			0.027 (empirical)		
Configuration	0.613 ± 0.008	-25.38	$<10^{-3}$	0.026 ± 0.001	1.32	0.188
H-Q-S	0.425 ± 0.014	-0.89	0.376	0.024 ± 0.001	3.77	$<10^{-3}$
White-noise	0.077 ± 0.005	62.77	$<10^{-3}$	0.000 ± 0.000	48473.21	$<10^{-3}$
US	0.328 (empirical)			0.024 (empirical)		
Configuration	0.508 ± 0.007	-25.96	$<10^{-3}$	0.023 ± 0.000	2.80	0.005
H-Q-S	0.362 ± 0.012	-2.79	0.005	0.022 ± 0.001	4.82	$<10^{-3}$
White-noise	0.075 ± 0.005	51.94	$<10^{-3}$	0.000 ± 0.000	59738.94	$<10^{-3}$

Community detection

Max.

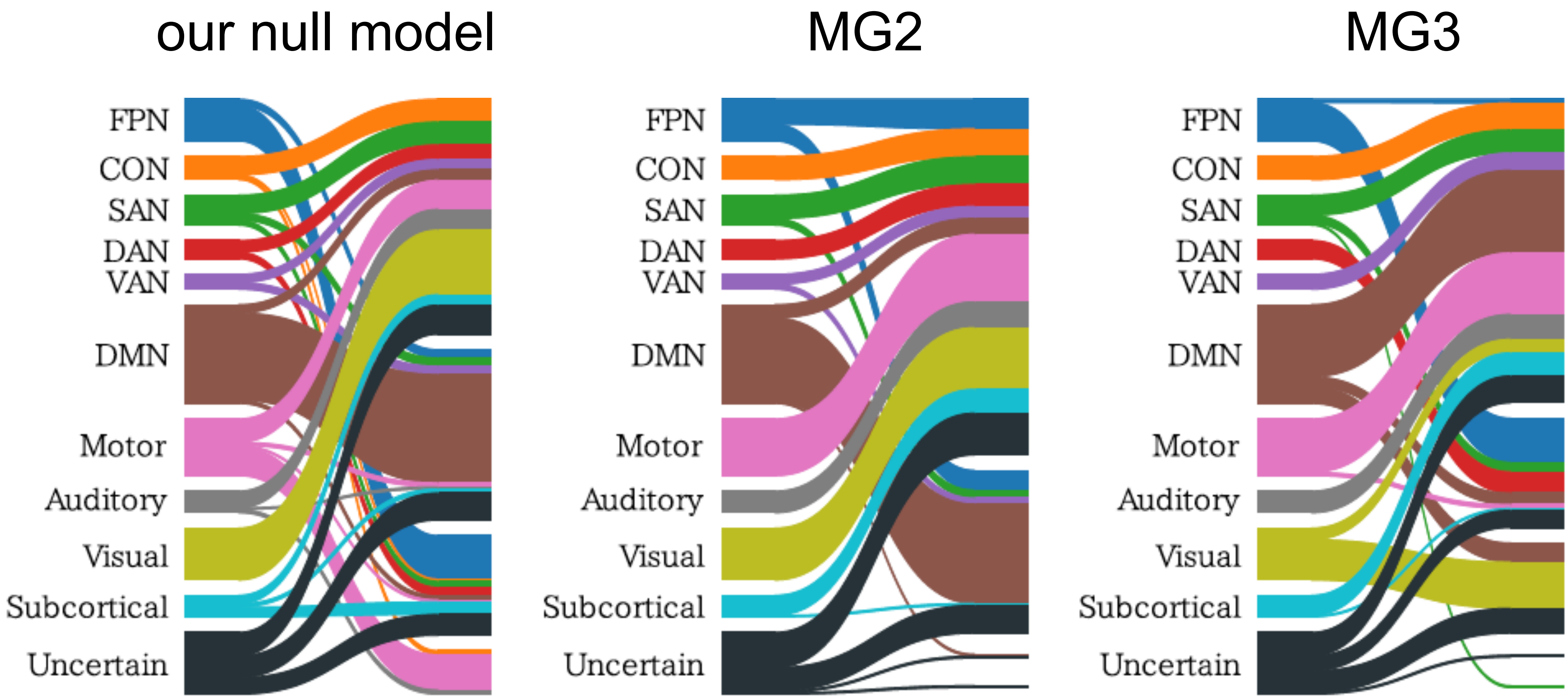
$$Q = \frac{1}{C_{\text{norm}}} \sum_{i,j=1}^{N} (\rho_{ij} - \langle \rho_{ij} \rangle)\delta(g_i, g_j)$$

When the configuration model is fed as input,

Null model	Q Original	Q Random	Z	P
Motivation				
$\langle \rho^{\mathrm{HQS}} \rangle$	0.15	0.09 ± 0.02	2.72	0.007
$\langle \rho^{\mathrm{MG1}} \rangle$	0.88	0.88 ± 0.01	0.85	0.204
$\langle \rho^{\mathrm{MG2}} \rangle$	0.98	0.97 ± 0.00	3.18	$<10^{-3}$
$\langle \rho^{\mathrm{MG3}} \rangle$	0.12	0.22 ± 0.01	−7.00	1.000
fMRI1				
$\langle \rho^{\mathrm{HQS}} \rangle$	0.61	0.17 ± 0.01	35.37	$<10^{-3}$
$\langle \rho^{\mathrm{MG1}} \rangle$	1.10	0.89 ± 0.00	54.62	$<10^{-3}$
$\langle \rho^{\mathrm{MG2}} \rangle$	0.11	1.02 ± 0.00	−966.06	1.000
$\langle \rho^{\mathrm{MG3}} \rangle$	0.11	0.50 ± 0.01	−36.92	1.000
Japan				
$\langle \rho^{\mathrm{HQS}} \rangle$	0.08	0.04 ± 0.01	7.76	$<10^{-3}$
$\langle \rho^{\mathrm{MG1}} \rangle$	0.99	0.99 ± 0.00	−0.19	0.582
$\langle \rho^{\mathrm{MG2}} \rangle$	0.01	1.00 ± 0.00	−25328.61	1.000
$\langle \rho^{\mathrm{MG3}} \rangle$	0.01	0.09 ± 0.00	−17.07	1.000
US				
$\langle \rho^{\mathrm{HQS}} \rangle$	0.10	0.05 ± 0.01	9.29	$<10^{-3}$
$\langle \rho^{\mathrm{MG1}} \rangle$	0.99	0.99 ± 0.00	−0.00	0.482
$\langle \rho^{\mathrm{MG2}} \rangle$	0.01	1.00 ± 0.00	−27404.80	1.000
$\langle \rho^{\mathrm{MG3}} \rangle$	0.01	0.11 ± 0.00	−19.89	1.000

Null model	Q Original	Q Random	Z	P
Motivation				
$\langle\rho^{\mathrm{con}}\rangle$	0.20	0.03 ± 0.02	10.09	$<10^{-3}$
$\langle\rho^{\mathrm{MG2}}\rangle$	0.99	0.98 ± 0.00	1.58	0.044
$\langle\rho^{\mathrm{MG3}}\rangle$	0.22	0.30 ± 0.02	-3.38	1.000
fMRI1				
$\langle\rho^{\mathrm{con}}\rangle$	0.95	0.05 ± 0.01	83.31	$<10^{-3}$
$\langle\rho^{\mathrm{MG2}}\rangle$	1.82	1.84 ± 0.23	-0.10	0.493
$\langle\rho^{\mathrm{MG3}}\rangle$	1.04	1.71 ± 0.22	-3.01	1.000
Japan				
$\langle\rho^{\mathrm{con}}\rangle$	0.03	0.01 ± 0.01	3.95	0.002
$\langle\rho^{\mathrm{MG2}}\rangle$	1.00	1.00 ± 0.00	4.69	$<10^{-3}$
$\langle\rho^{\mathrm{MG3}}\rangle$	0.04	0.08 ± 0.01	-6.75	1.000
US				
$\langle\rho^{\mathrm{con}}\rangle$	0.05	0.01 ± 0.01	5.13	$<10^{-3}$
$\langle\rho^{\mathrm{MG2}}\rangle$	1.00	1.00 ± 0.00	4.10	$<10^{-3}$
$\langle\rho^{\mathrm{MG3}}\rangle$	0.05	0.11 ± 0.01	-7.91	1.000

functional MRI (participant 1)



Comparison with biologically determined communities

Null model	Data	P^{emp}	P^{rand}	$P^{\text{emp}} - P^{\text{rand}}$	$P^{\text{emp}} / P^{\text{rand}}$
All nodes					
$\langle \rho^{\text{con}} \rangle$	fMRI1	0.623	0.316	0.306	1.968
	fMRI2	0.673	0.336	0.337	2.004
$\langle \rho^{\text{MG2}} \rangle$	fMRI1	0.725	0.555	0.170	1.307
	fMRI2	0.726	0.520	0.206	1.397
$\langle \rho^{\text{MG3}} \rangle$	fMRI1	0.605	0.434	0.171	1.393
	fMRI2	0.529	0.362	0.167	1.460

Application to network construction

- We may not be interested in (i, j) if the $\text{corr}(X_i, X_j)$ value is "unsurprising".
- Statistical tests on the individual edges (node pairs) face a multiple comparison problem. And $\text{corr}(X_1, X_2)$ and $\text{corr}(X_1, X_3)$ are generally correlated with each other.
- Filtering correlation matrices
  - max likelihood + regularization (to estimate sparse networks)

Our contributions

- Main strategies in the field:
  - Estimate sparse precision matrices (“graphical lasso”)
  - Estimate sparse covariance matrices … seems less popular
- Alternative to “thresholding networks”
- Lacking null model thinking

- Sparse network reconstruction from COrrelational daa with LAsso, “Scola”
  - Correlation matrix/network
  - Use null models

$$\boldsymbol{C} = \boldsymbol{C}^{\mathrm{null}} + \boldsymbol{W}$$

Estimated corr matrix (which may not be equal to the sample corr matrix)

Null model calculated from data

Sparse matrix estimated by lasso

$$\boldsymbol{X} = (x_{\ell i})$$

where $\ell = 1, \ldots, L$; $i = 1, \ldots, N$;
N: number of nodes; L: number of samples (e.g. time points in a time series)

ℓ-th sample: $$\boldsymbol{x}_\ell = [x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell N}]$$

Sample corr matrix: $$\boldsymbol{C}^{\text{sample}} \equiv \boldsymbol{X}^\top \boldsymbol{X} / L$$

Given the data and the calculated null model, find $\boldsymbol{W}$ maximizing

$$P(\boldsymbol{X}|\boldsymbol{C}) \equiv \prod_{\ell=1}^{L} \frac{1}{(2\pi)^{N/2} \det(\boldsymbol{C})^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{x}_\ell \boldsymbol{C}^{-1} \boldsymbol{x}_\ell^\top\right)$$

Reminder: $$\boldsymbol{C} = \boldsymbol{C}^{\text{null}} + \boldsymbol{W}$$

$$P\left(\boldsymbol{X}|\boldsymbol{C}\right) \equiv \prod_{\ell=1}^{L} \frac{1}{(2\pi)^{N/2}\det\left(\boldsymbol{C}\right)^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}_\ell \boldsymbol{C}^{-1}\boldsymbol{x}_\ell^\top\right)$$

Using $\sum_{\ell=1}^{L} \boldsymbol{x}_\ell \boldsymbol{C}^{-1}\boldsymbol{x}_\ell^\top = \mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{C}^{-1}\right)$ we get

$$\ln P\left(\boldsymbol{X}|\boldsymbol{C}\right) = -\frac{L}{2}\ln\det\left(\boldsymbol{C}\right) - \frac{L}{2}\mathrm{tr}\left(\boldsymbol{C}^{\mathrm{sample}}\boldsymbol{C}^{-1}\right) - \frac{NL}{2}\ln(2\pi)$$

Using $\boldsymbol{C} = \boldsymbol{C}^{\mathrm{null}} + \boldsymbol{W}$ we get

$$\mathcal{L}\left(\boldsymbol{W}\right) \equiv -\frac{L}{2}\ln\det\left(\boldsymbol{C}^{\mathrm{null}} + \boldsymbol{W}\right) - \frac{L}{2}\mathrm{tr}\left[\boldsymbol{C}^{\mathrm{sample}}\left(\boldsymbol{C}^{\mathrm{null}} + \boldsymbol{W}\right)^{-1}\right] - \frac{NL}{2}\ln(2\pi)$$

Though (in fact) $\boldsymbol{W}^{\mathrm{MLE}} \equiv \boldsymbol{C}^{\mathrm{sample}} - \boldsymbol{C}^{\mathrm{null}}$

$\boldsymbol{W}^{\mathrm{MLE}}$ overfits the data, resulting in many (spurious?) edges.

So, use lasso to maximize a penalized (log) likelihood:

$$\hat{\mathcal{L}}(\boldsymbol{W}|\boldsymbol{\lambda}) \equiv \mathcal{L}(\boldsymbol{W}) - \frac{L}{2}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\lambda_{ij}|W_{ij}|$$

Not concave w.r.t. ***W***, so use an extension of a previous numerical optimization algorithm.

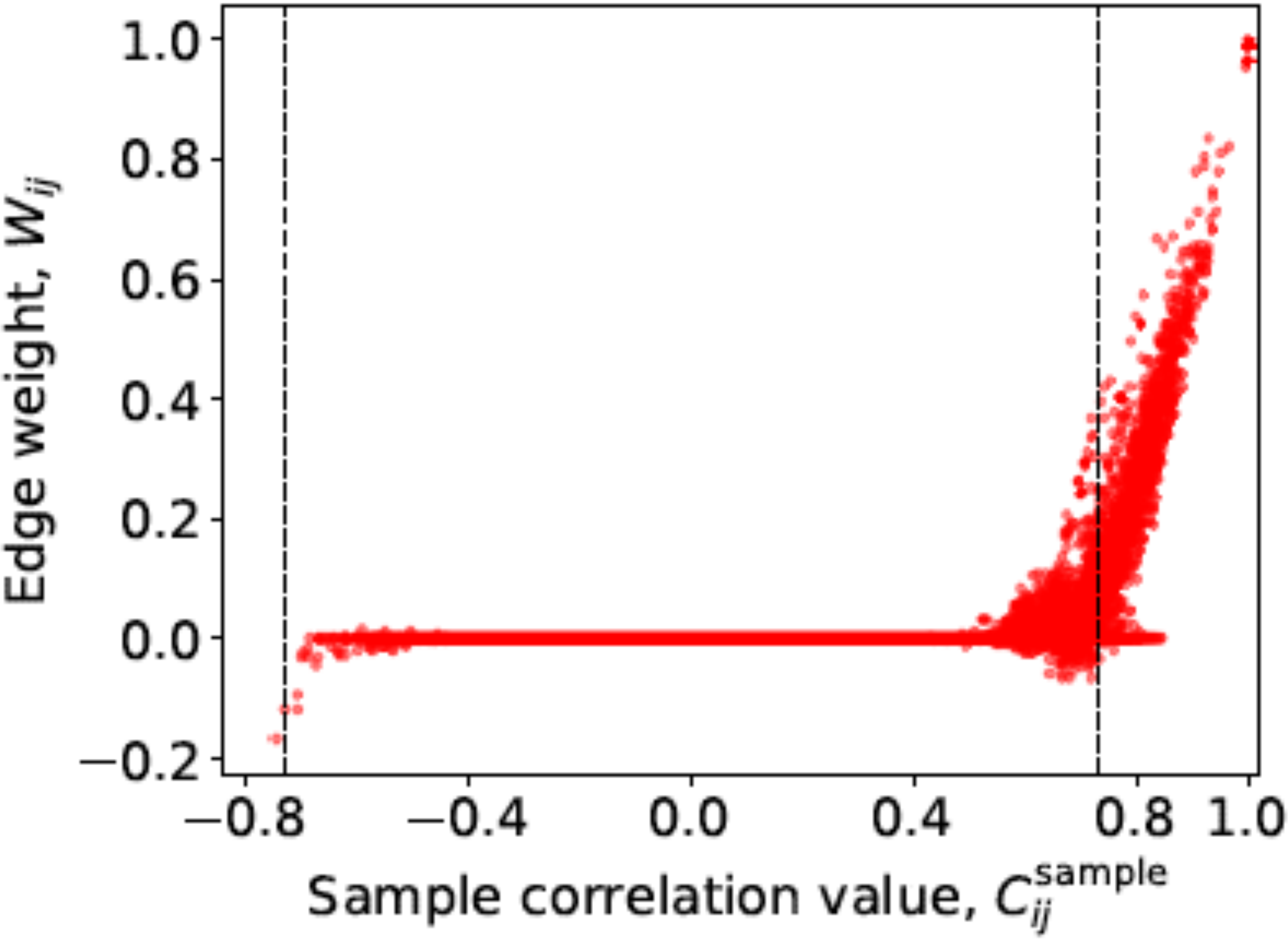
Adaptive Lasso:

$$\lambda_{ij} = \overline{\lambda}|W_{ij}^{\mathrm{MLE}}|^{-\gamma}$$

where $\overline{\lambda} \geq 0$ and $\gamma > 0$ are hyperparameters.

Determined by a model selection criterion.

We use a typical value $\gamma = 2$.



Edge weight, W_{ij}
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.8
−0.4
0.0
0.4
0.8
1.0
Sample correlation value, C_{ij}^{sample}

Model selection

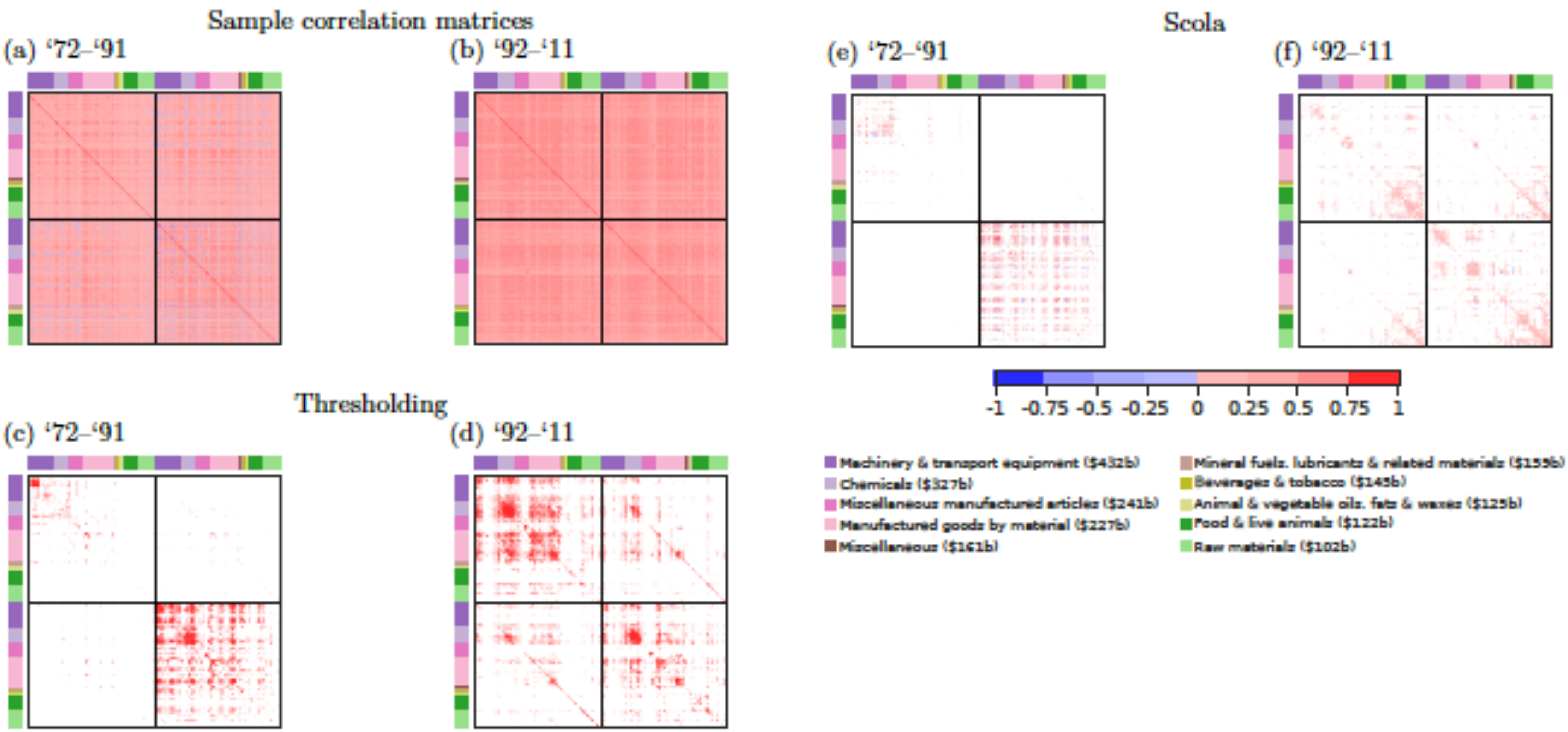
- Extended Bayes Information Criterion (Chen & Chen, Biometrika, 2008; Foygel & Drton, NIPS'10)
  - Smaller the better.
- WN: White noise null model
- HQS: Hirschberger-Qi-Steuer null model (≈ Erdős-Rényi RG)
- con: Our configuration model

Data	Null model					
	Correlation matrix			Precision matrix		
	C^{WN}	C^{HQS}	C^{con}	C^{WN}	C^{HQS}	C^{con}
Product space						
'72–'91	0.226	0.202	0.187	0.232	0.190	0.208
'92–'11	0.300	0.301	0.226	0.304	0.238	0.256
S&P 500						
'00–'07	0.600	0.562	0.553	0.640	0.557	0.587
'08–'15	0.664	0.558	0.524	0.596	0.539	0.607
Nikkei	0.991	0.874	0.861	1.001	0.837	0.882

Kojaku & Masuda, arXiv:1903.10805

Application to product space data/networks

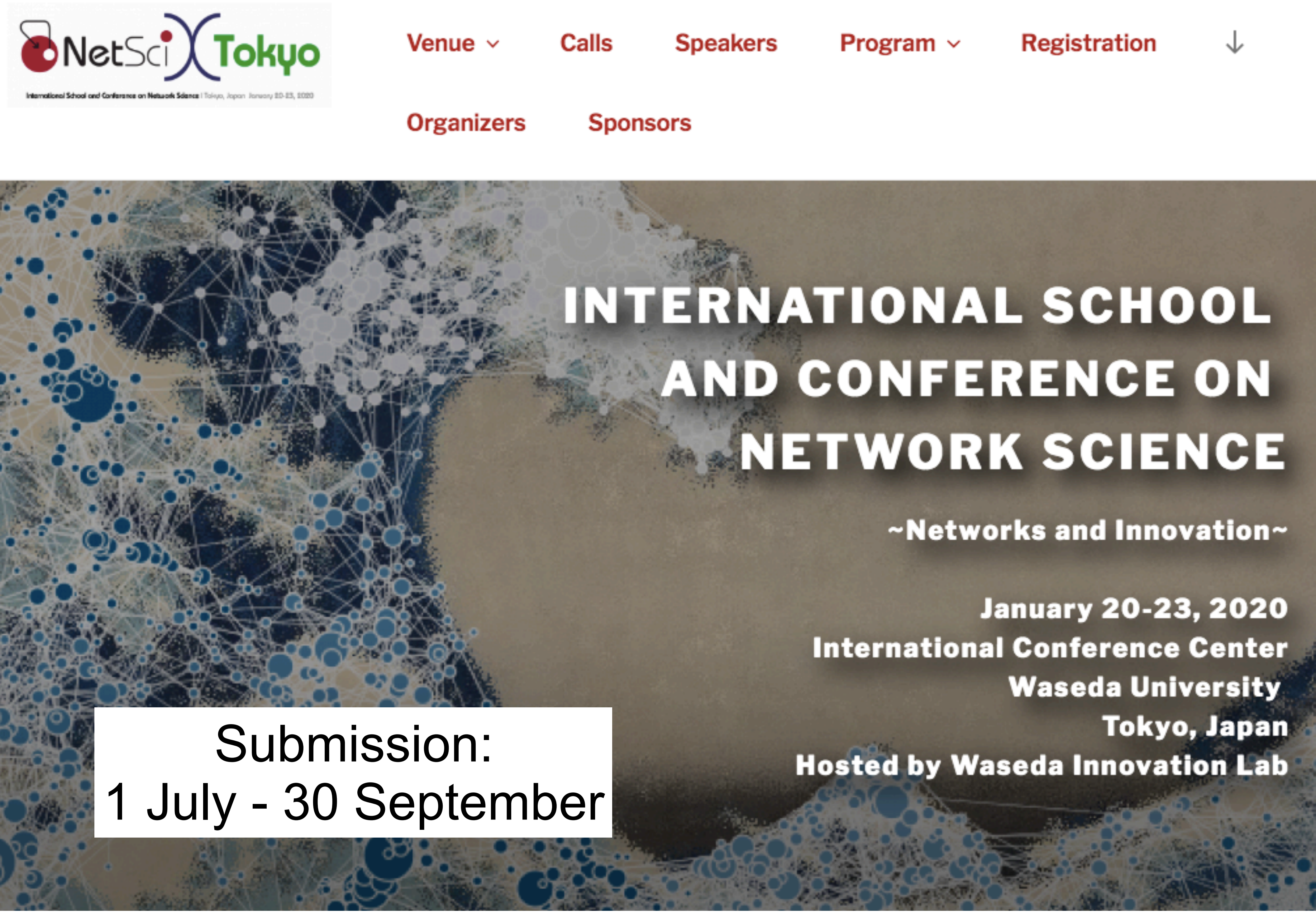
- Prediction



- → Poster by Sadamori Kojaku

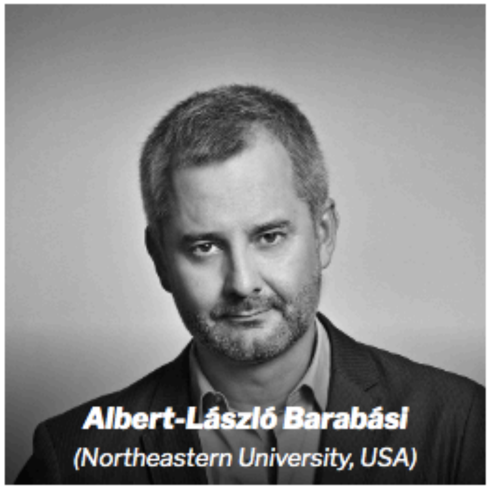
Conclusions

- Configuration model for correlation/covariance matrices
    - Maximum entropy
    - Node strength (i.e. weighted degree) preserved
    - Illustration with clustering coefficients and community detection
- Filtering as an alternative to thresholding
- Code/algorithmic documentation on Github (in python and MATLAB)
- Future problems: more examples, wherever null models are needed, …

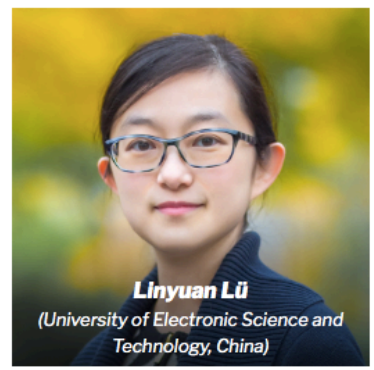


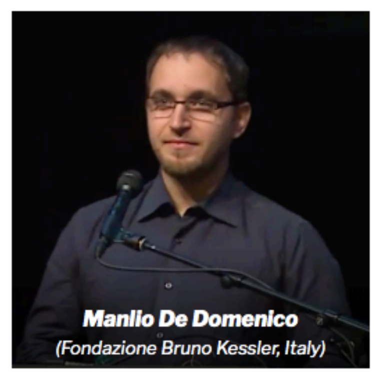
Submission:
1 July - 30 September

Keynotes:



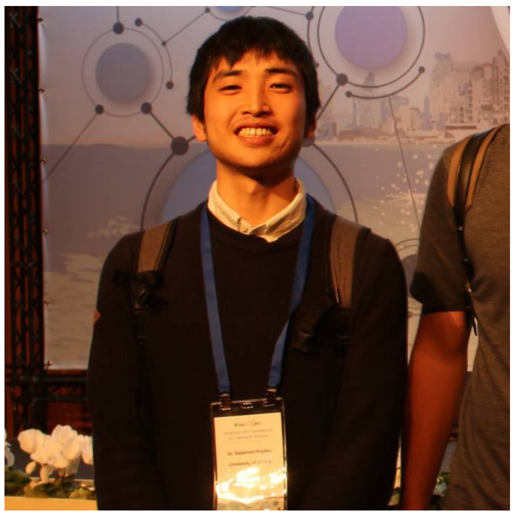
Invited:





Acknowledgments

Sadamori Kojaku
(Kobe University; former postdoc at University of Bristol)



Yukie Sano
(Tsukuba University)

References

- Masuda, Kojaku & Sano, Physical Review E, 98, 012312 (2018).
- Kojaku & Masuda, arXiv:1903.10805