Mapping the modular organization of complex networks

A. Arenas, J. Borge-Holthoefer, P. Erola, S. Gómez and G. Zamora

UNIVERSITAT ROVIRA I VIRGILI

Universität Potsdam

Oxford 2010

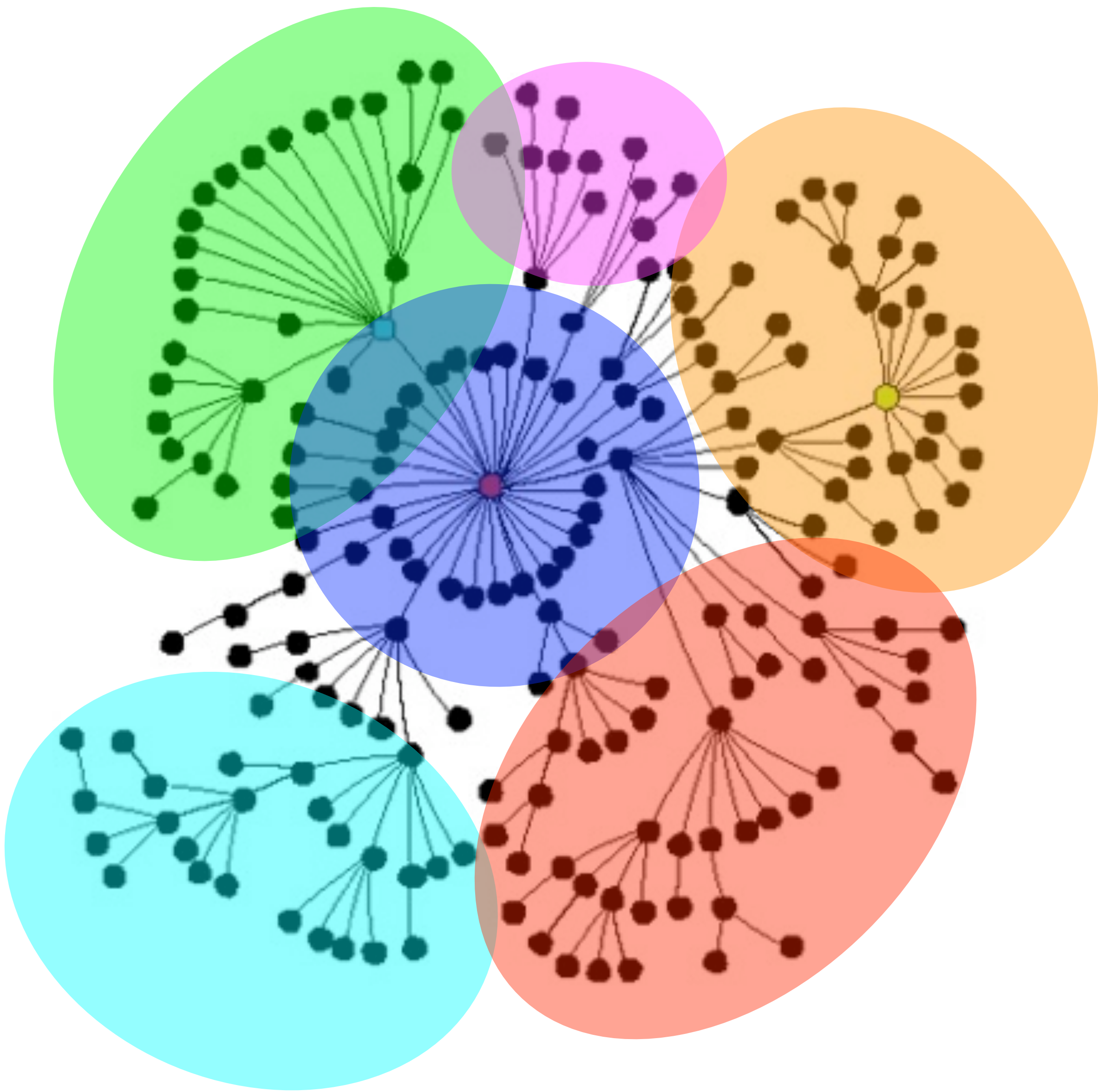
Introduction

Detecting modules, or communities, in real complex network is an important open issue

Modules affect physical processes on networks: synchronization, information or virus spreading, etc.

Great effort to propose modules detection algorithms: Fortunato, Phys. Rep. 486, 75-174, 2010

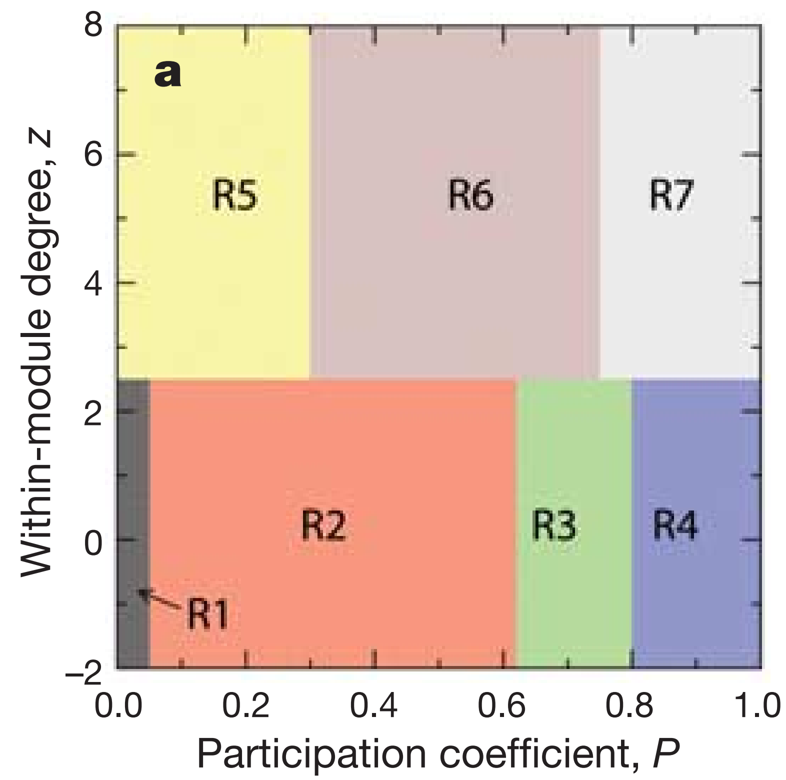
Once modules are found, what can be said about them?



Introduction

First attempt: Guimerà et al. Nature 433 (2005)

Goal: Find the roles of individual nodes in the network
Idea: nodes with the same role should have similar topological properties, with respect to a mesoscopic description in terms of modules



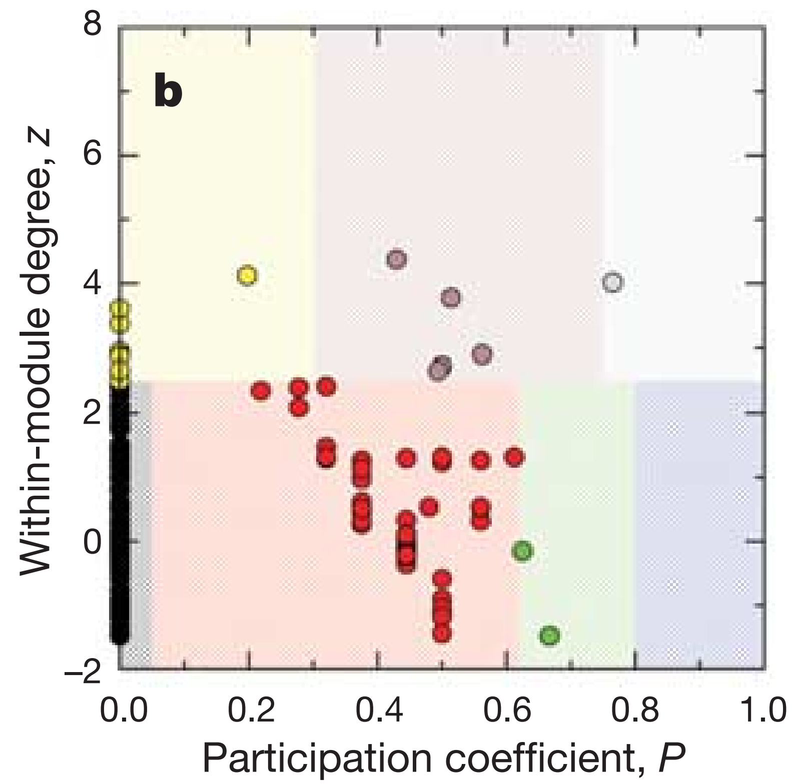
Define:

Within modules degree z-score

$$z_i = \frac{\kappa_i - \bar{\kappa}_{s_i}}{\sigma_{\kappa_{s_i}}}$$

Participation ratio

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{\kappa_{is}}{k_i} \right)^2$$

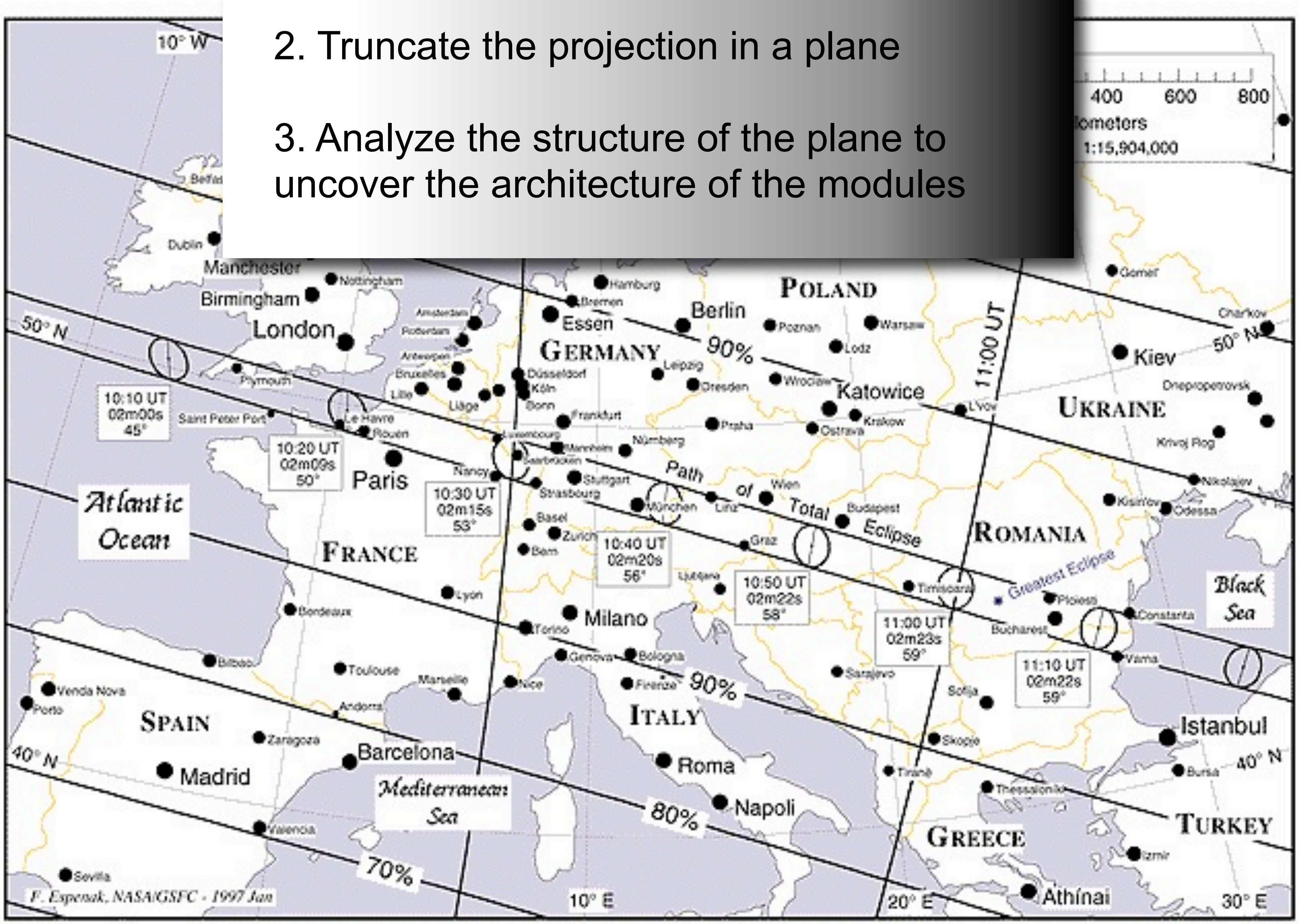


Our approach:

1. Find the best linear projection of the modular structure of a network

2. Truncate the projection in a plane

3. Analyze the structure of the plane to uncover the architecture of the modules



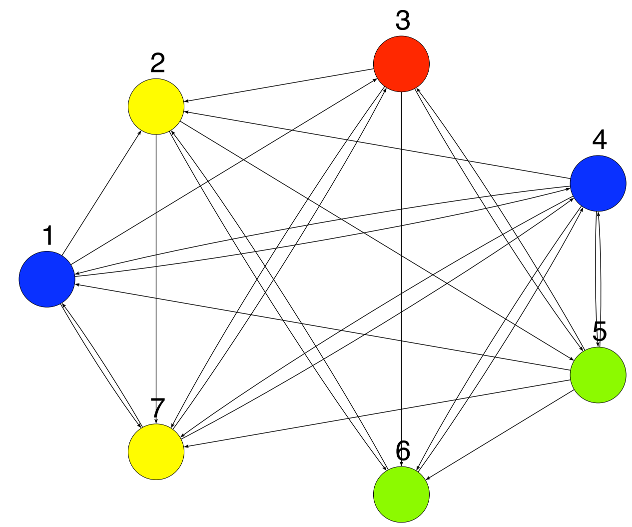
Contribution matrix

$N \times N$

0	1	4	2	0	0	1
0	0	0	0	1	1	2
0	2	0	1	2	3	1
1	1	0	0	2	4	5
2	0	2	1	0	3	4
0	1	0	2	0	0	0
3	0	1	4	0	0	0

W

(weights matrix)



$$C_{i\alpha} = \sum_{j=1}^{N} W_{ij} S_{j\alpha}$$

$N \times M$

0	0	0	1
1	0	0	0
0	0	1	0
0	0	0	1
0	1	0	0
0	1	0	0
1	0	0	0

S

(partition matrix)

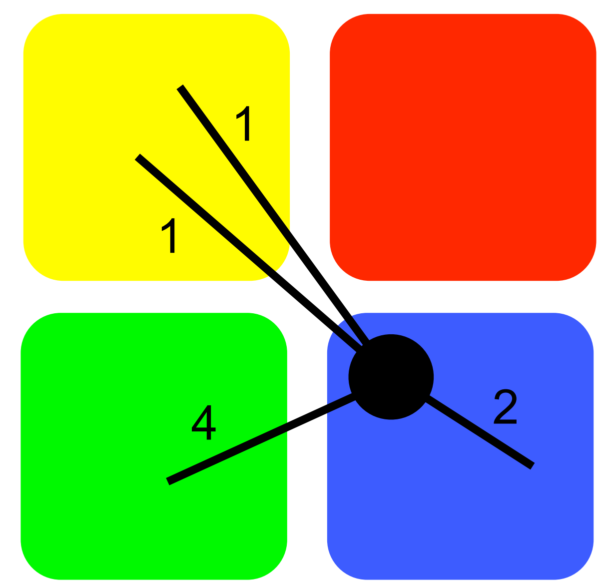
(the partition is given)

$N \times M$

2	0	4	2
2	2	0	0
3	5	0	1
6	6	0	1
4	3	2	3
1	0	0	2
0	0	1	7

C

(contribution matrix)

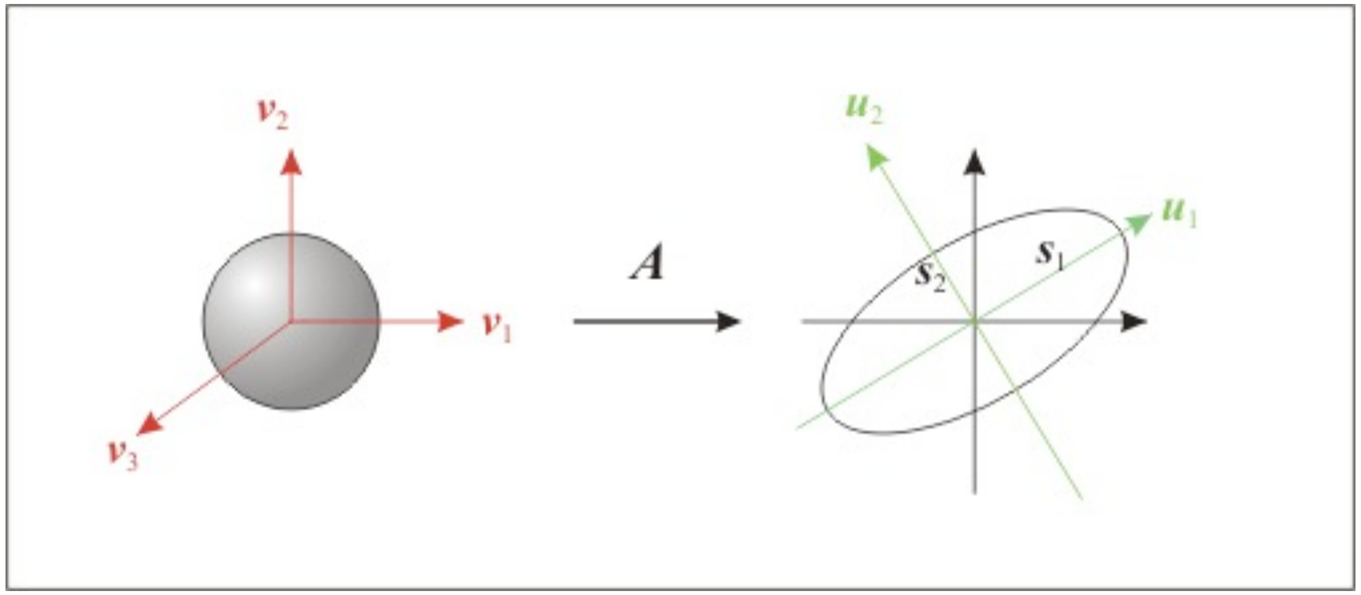


Linear projection: Singular Value Decomposition

Suppose M is an m-by-n real (or complex) matrix. Then there exists a factorization of the form

M = U Σ V*

where U is an m-by-m unitary matrix, the matrix Σ is m-by-n diagonal matrix with nonnegative real numbers on the diagonal, and V* denotes the conjugate transpose of V, an n-by-n unitary matrix. This is called a singular-value decomposition of M.

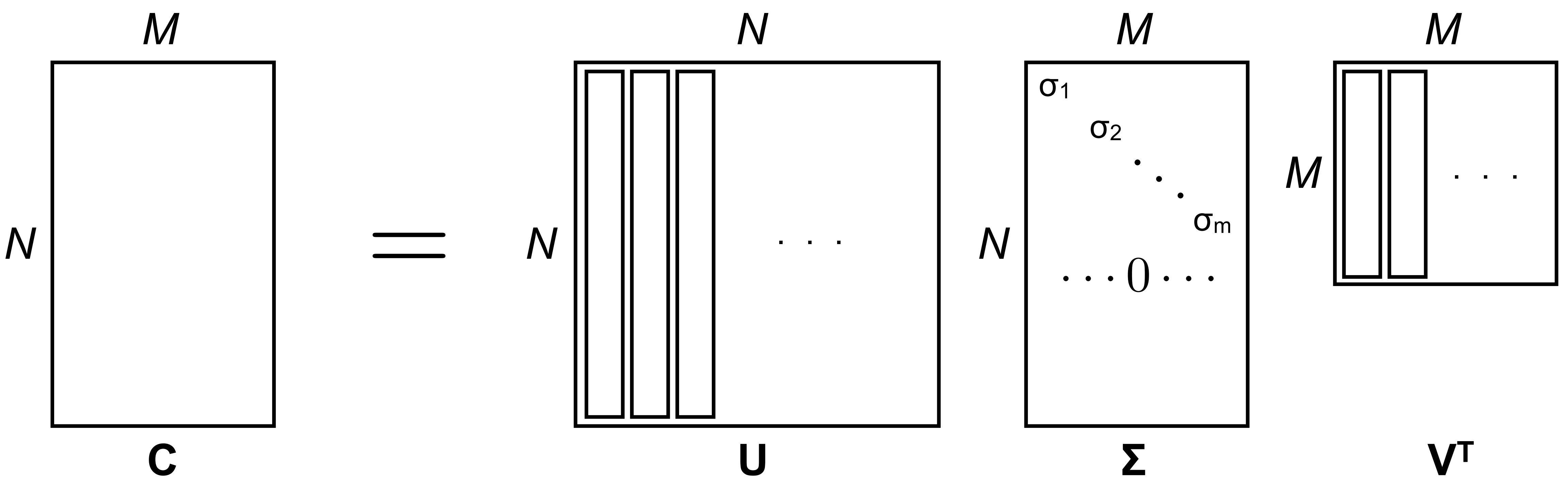


- The columns of V form a set of orthonormal "input" or "analysing" basis vector directions for M. (These are the eigenvectors of M^*M.)

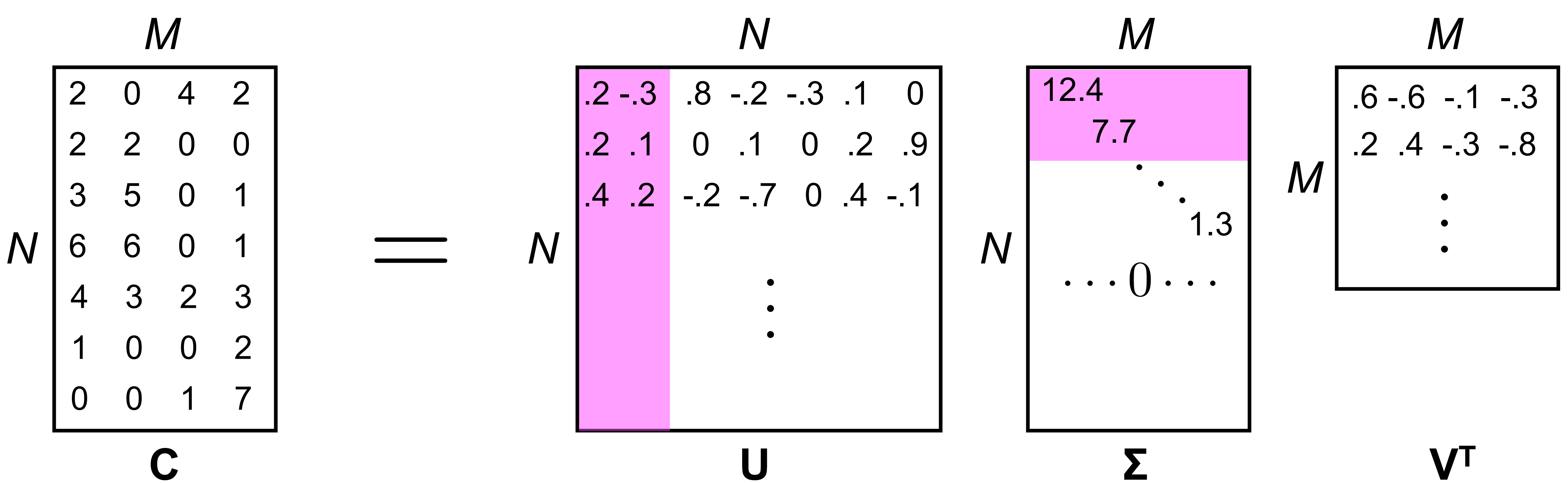
- The columns of U form a set of orthonormal "output" basis vector directions for M. (These are the eigenvectors of MM^*.)

- The diagonal values in matrix Σ are the singular values, which can be thought of as scalar "gain controls" by which each corresponding input is multiplied to give a corresponding output. (These are the square roots of the eigenvalues of MM^* and M^*M that correspond with the same columns in U and V.)

Singular Value Decomposition (SVD)

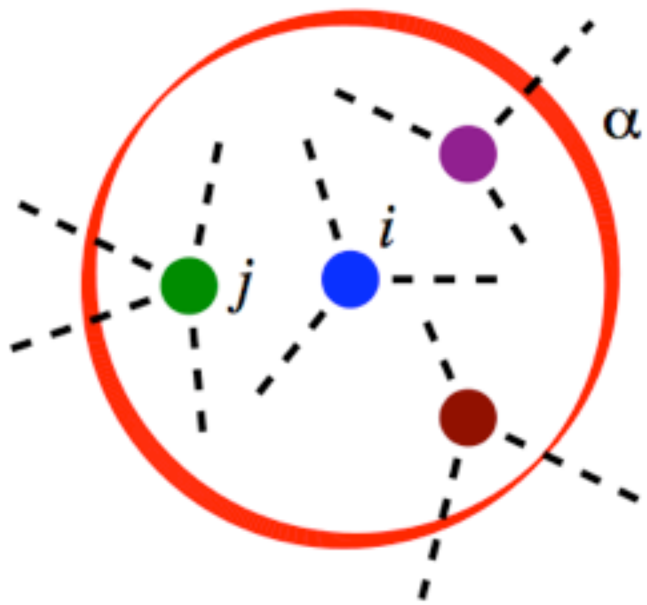


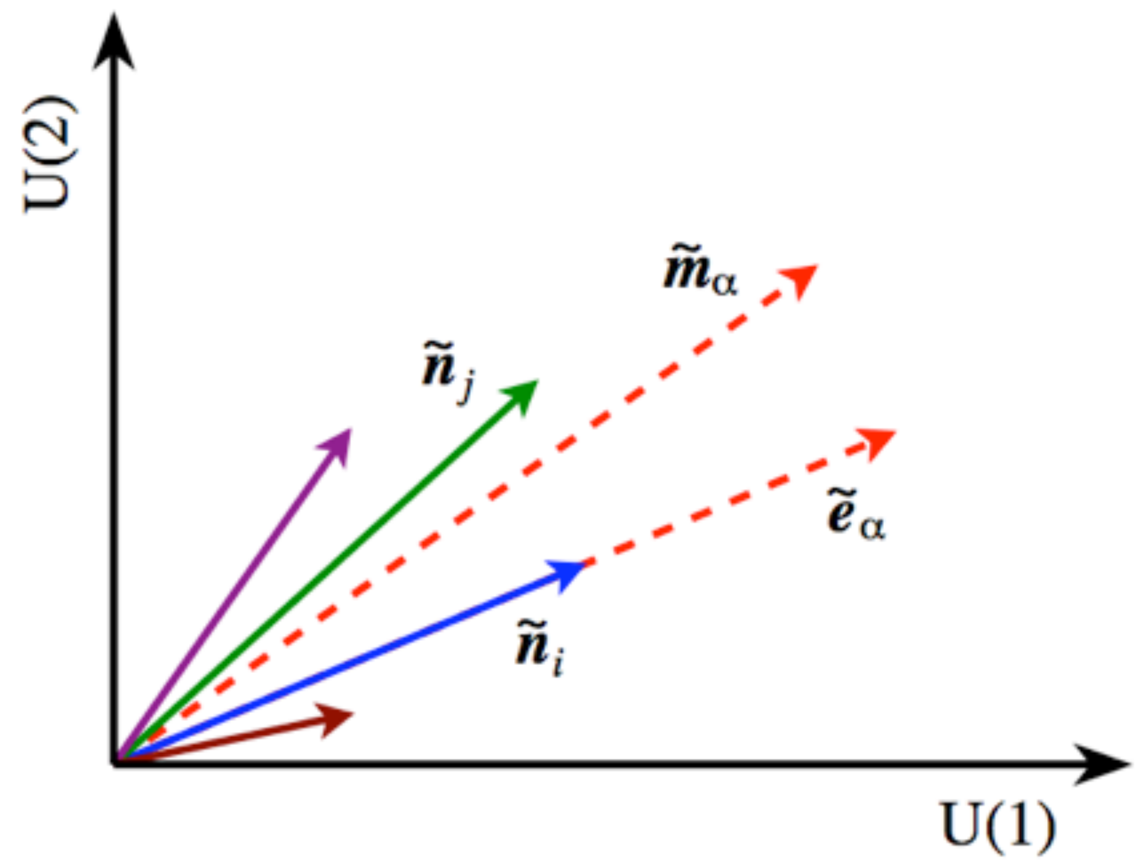
Truncated Singular Value Decomposition (TSVD)



least squares optimal

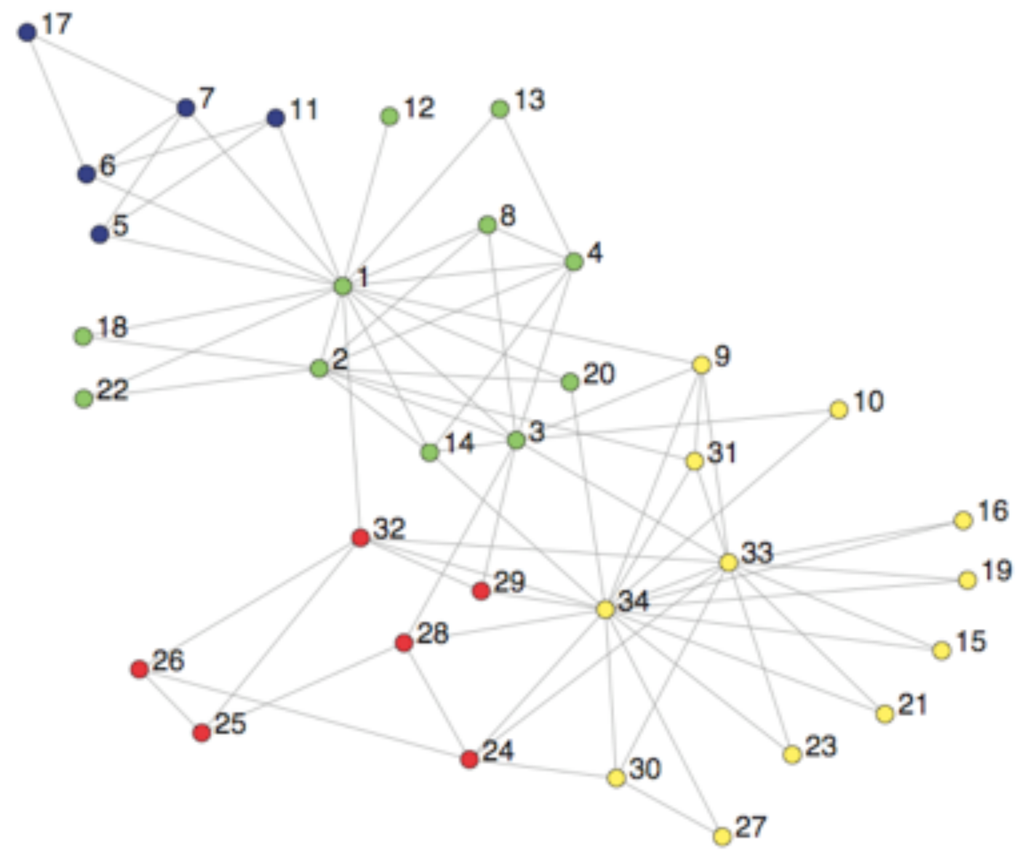
Truncated Singular Value Decomposition (TSVD), r = 2



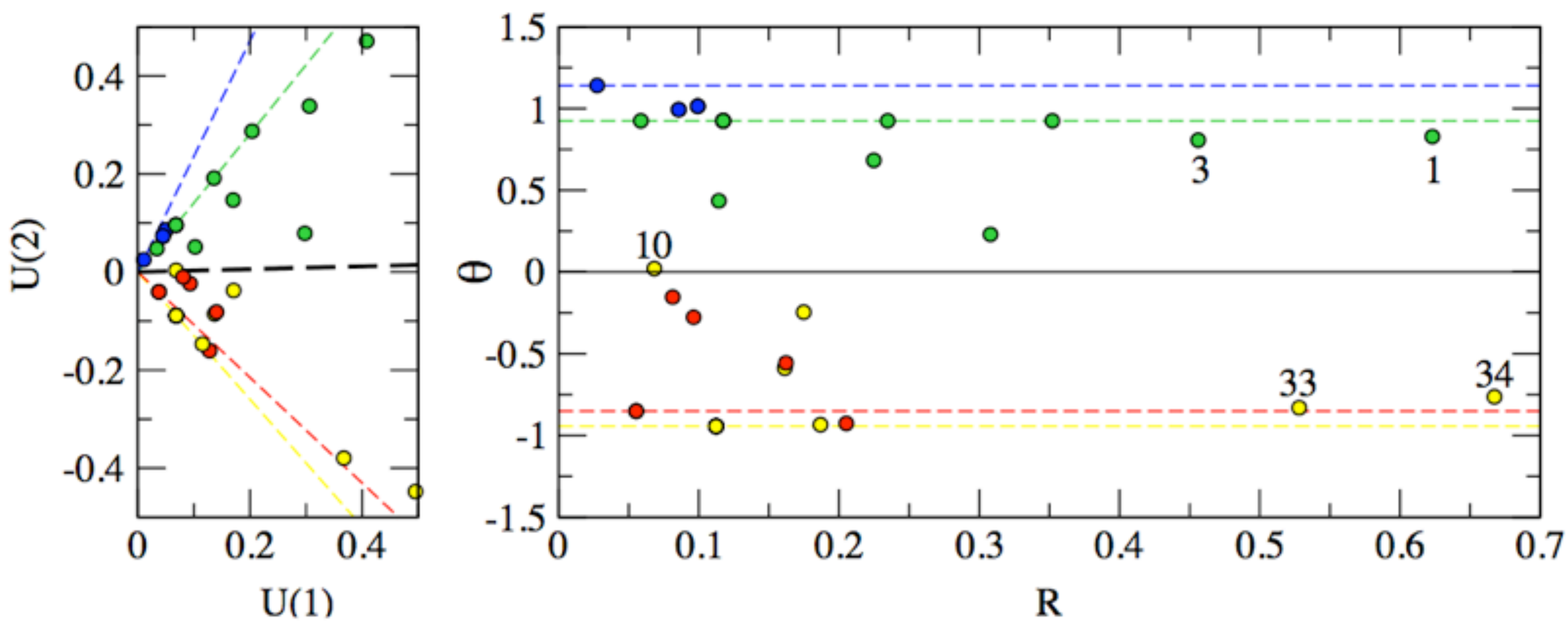


$\tilde{n}_i$	Node *i* contribution projection
$\tilde{e}_\alpha$	Intramodular projection of α
$\tilde{m}_\alpha$	Modular projection of α

The output of TSVD



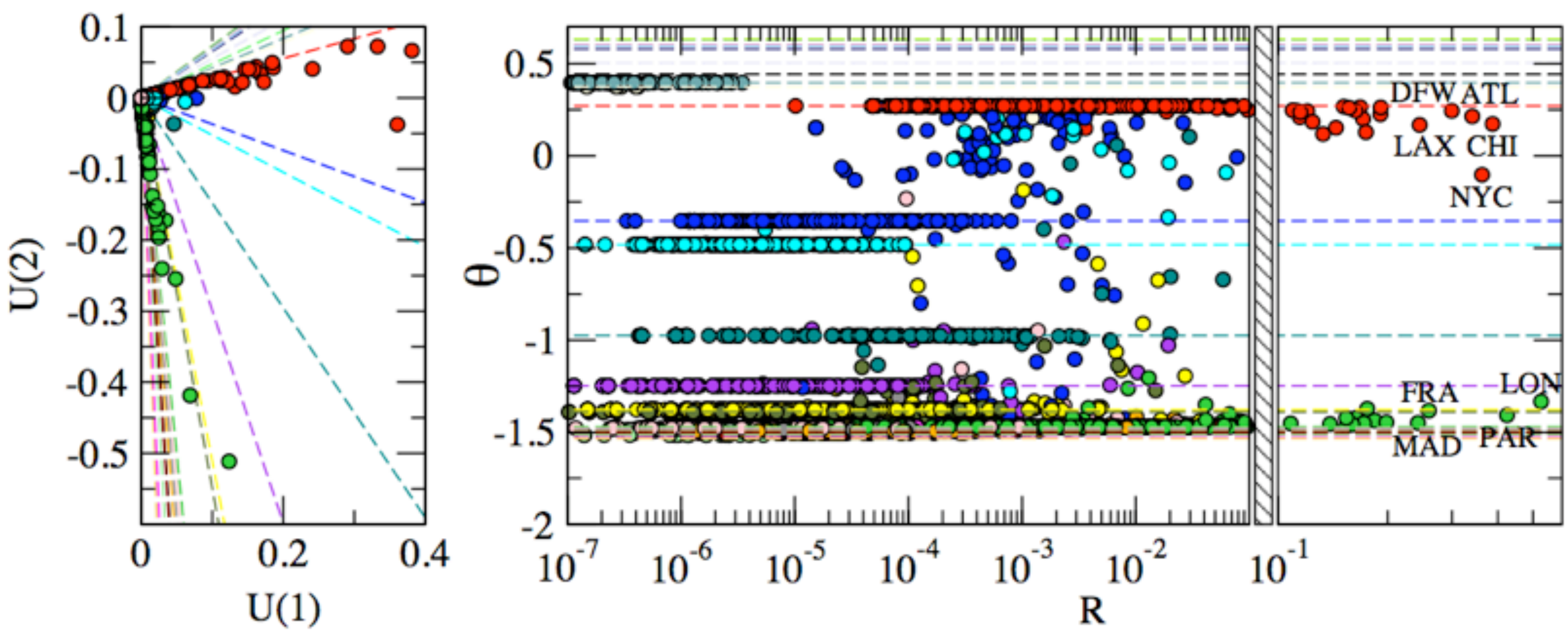
N = 34
M = 4



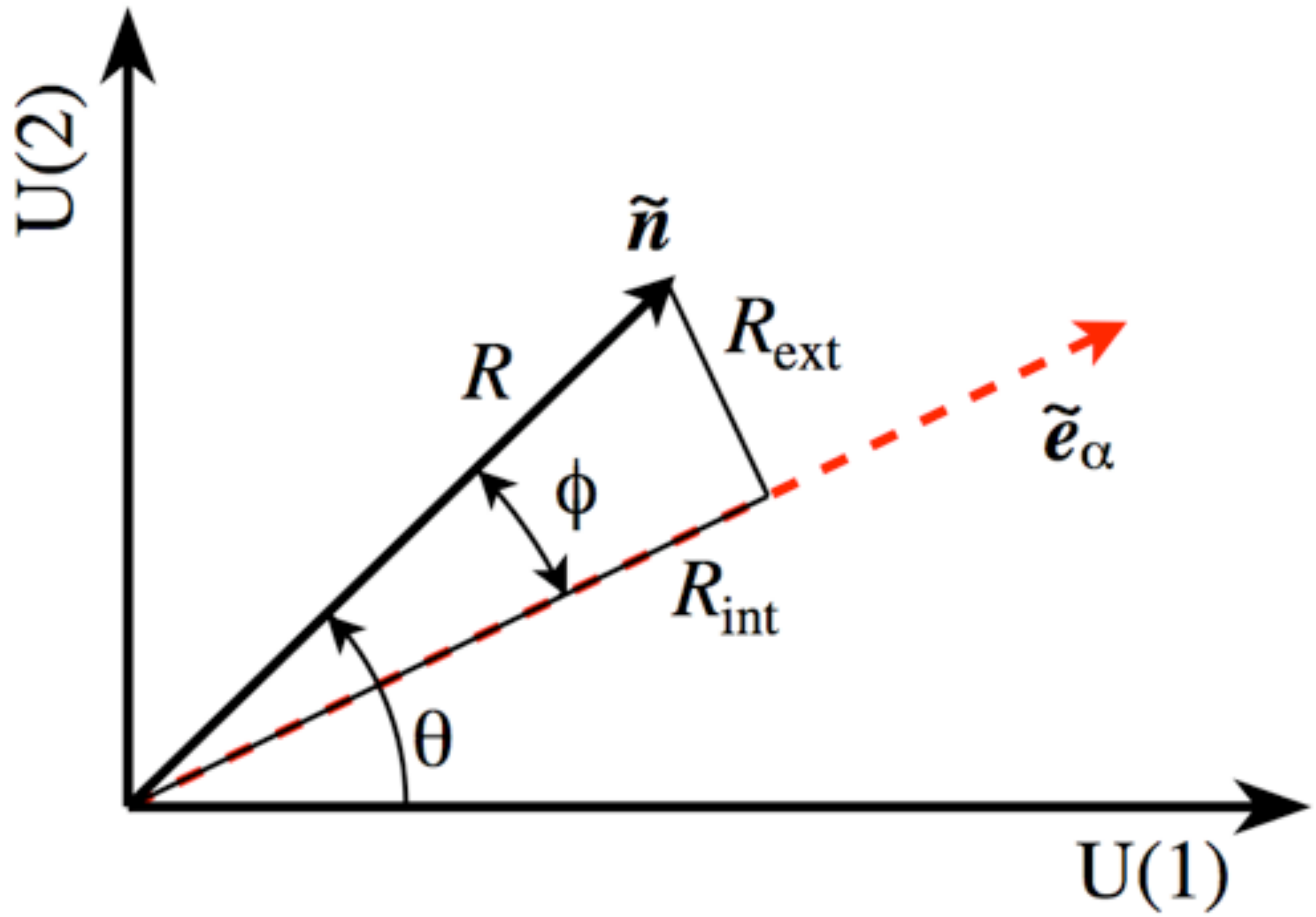
The output of TSVD

N = 3618

M = 26



Interpreting TSVD: the structure of individual modules

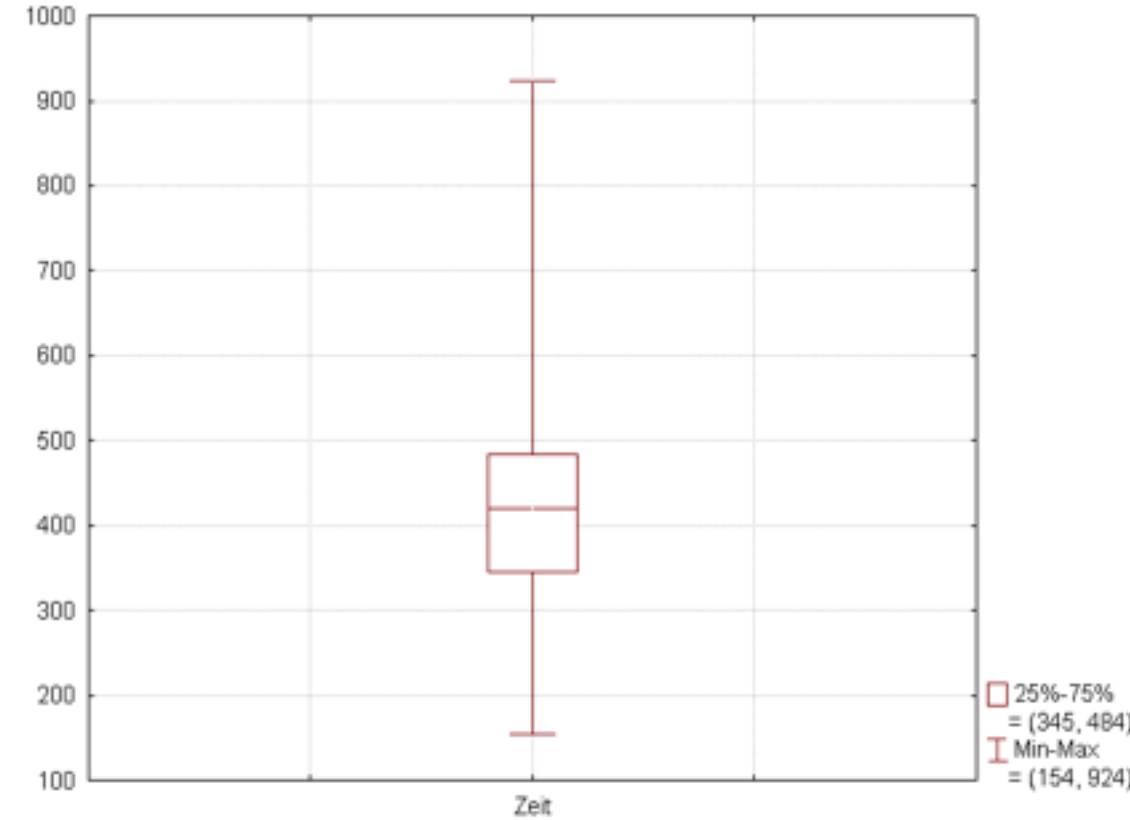


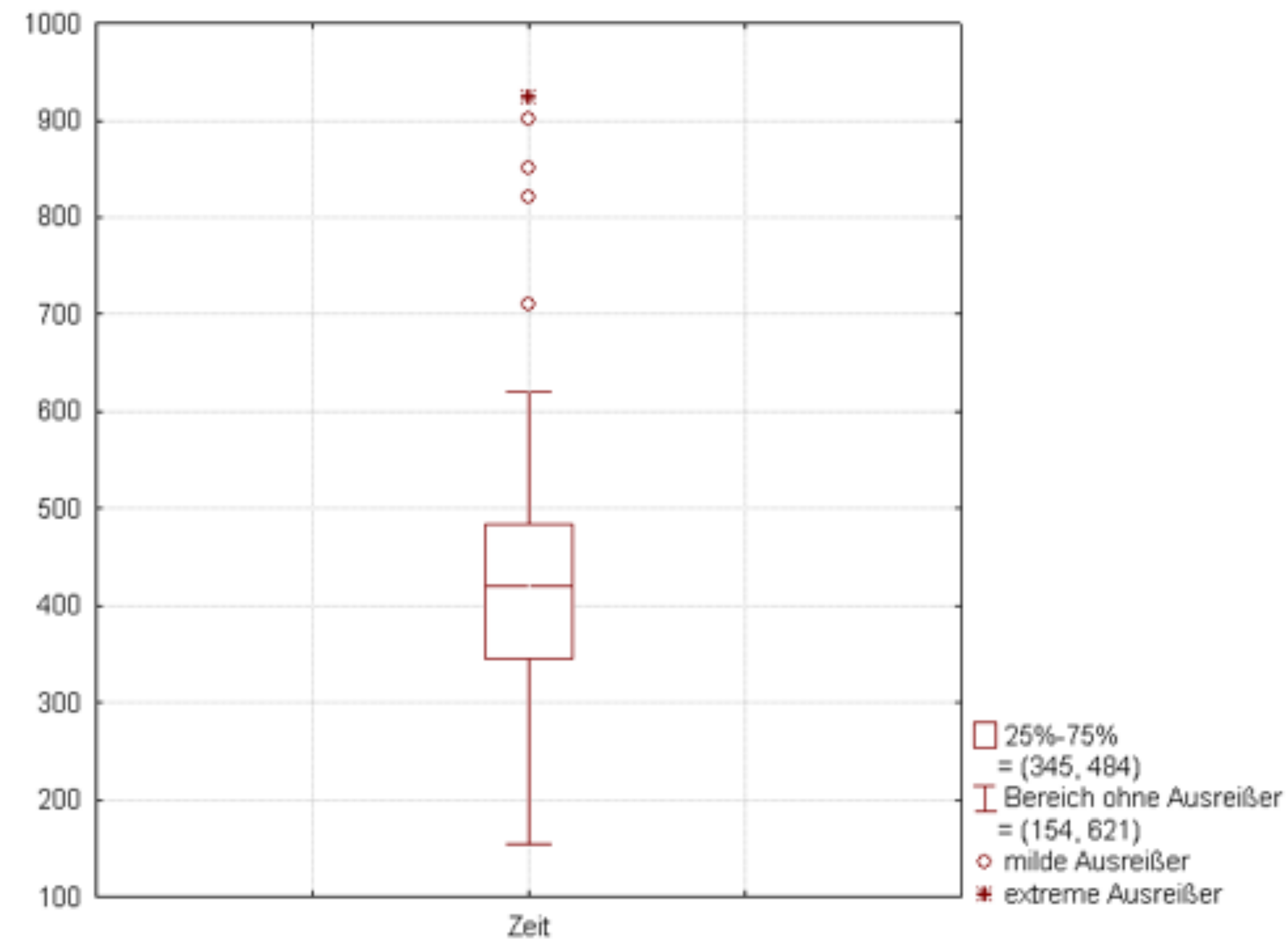
$$R_{int} = R\cos\phi$$
$$R_{ext} = R\sin\phi$$

➟ statistics for each node in each module

Statistics: Box and whiskers

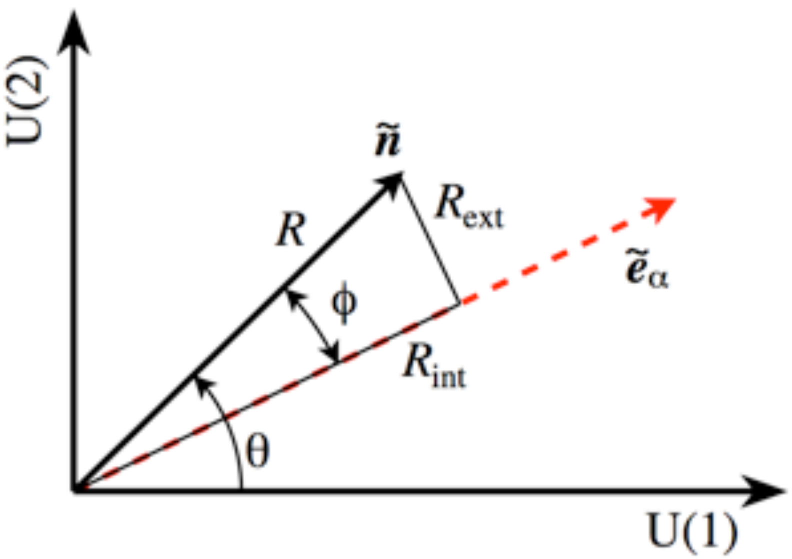
Box plots: Box and whisker plots are uniform in their use of the box. The bottom and top of the box are always the 25th and 75th percentile (the lower and upper quartiles, respectively), and the band near the middle of the box is always the 50th percentile (the median). The lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Any data not included between the whiskers should be plotted as an outlier with a dot.

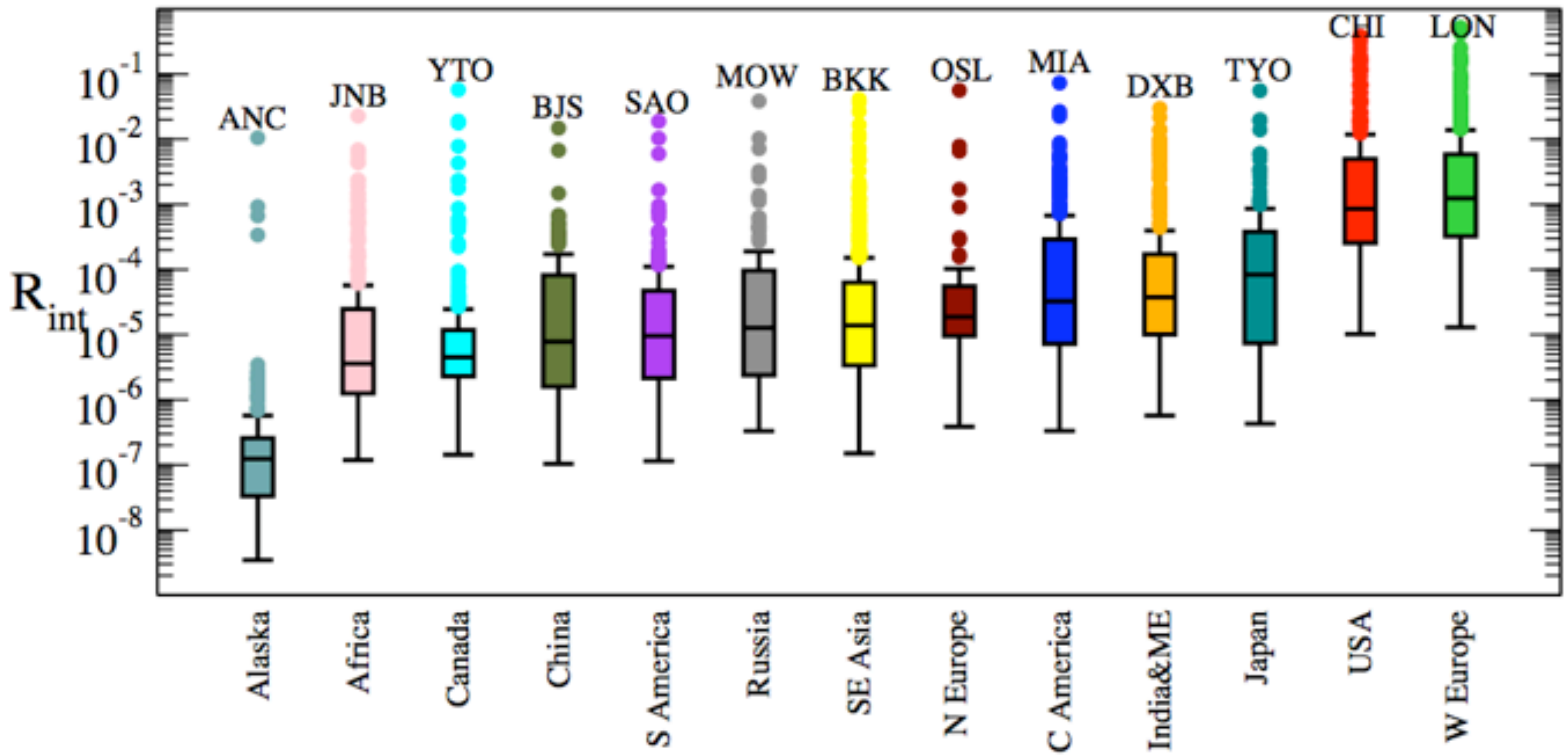




Interpreting TSVD: the structure of individual modules

$$R_{int} = R \cos \phi$$





$N_{SE\text{-}Asia} = 547$

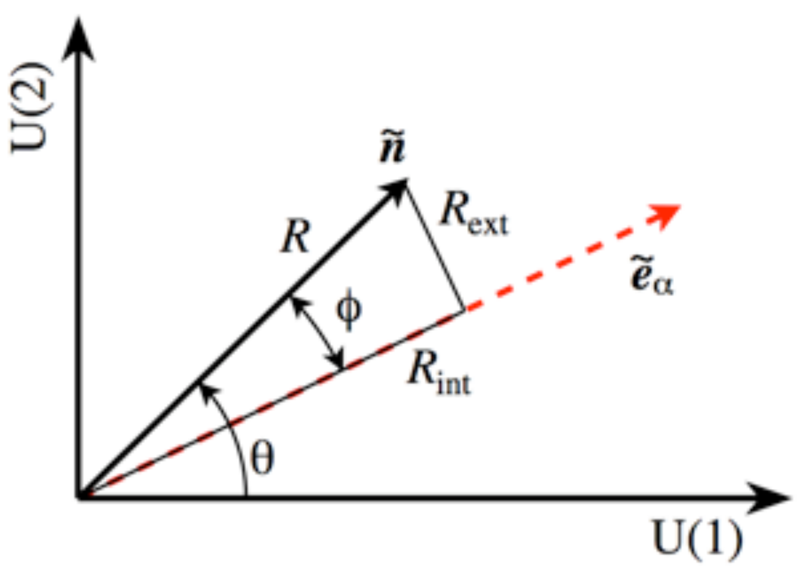
$N_{USA} = 507$

$N_{WE} = 423$

$N_{CA} = 292$

Interpreting TSVD: the structure of individual modules

$$R_{int} = R\cos\phi$$



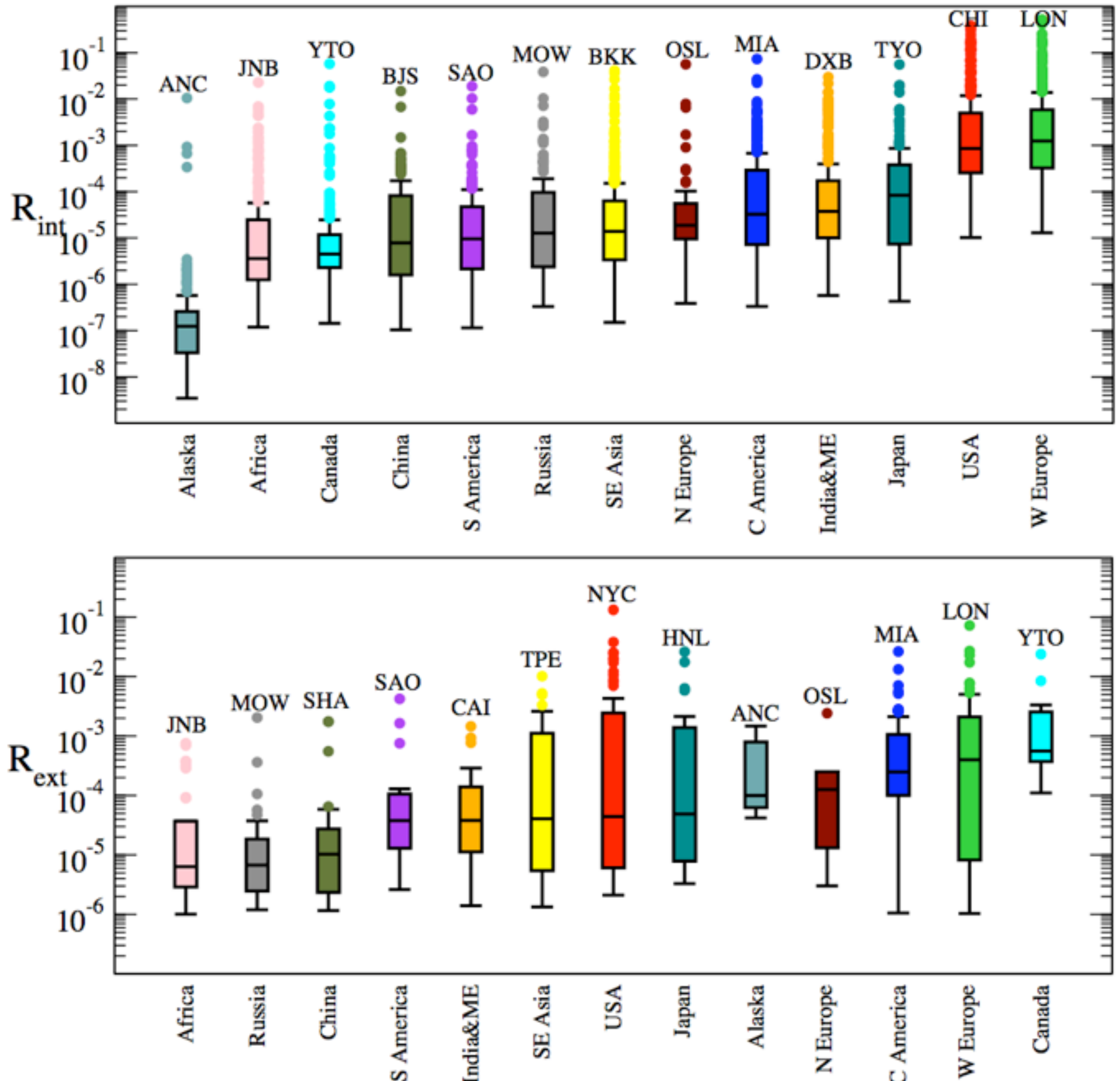
$$R_{ext} = R\sin\phi$$

$N_{SE\text{-}Asia} = 547$

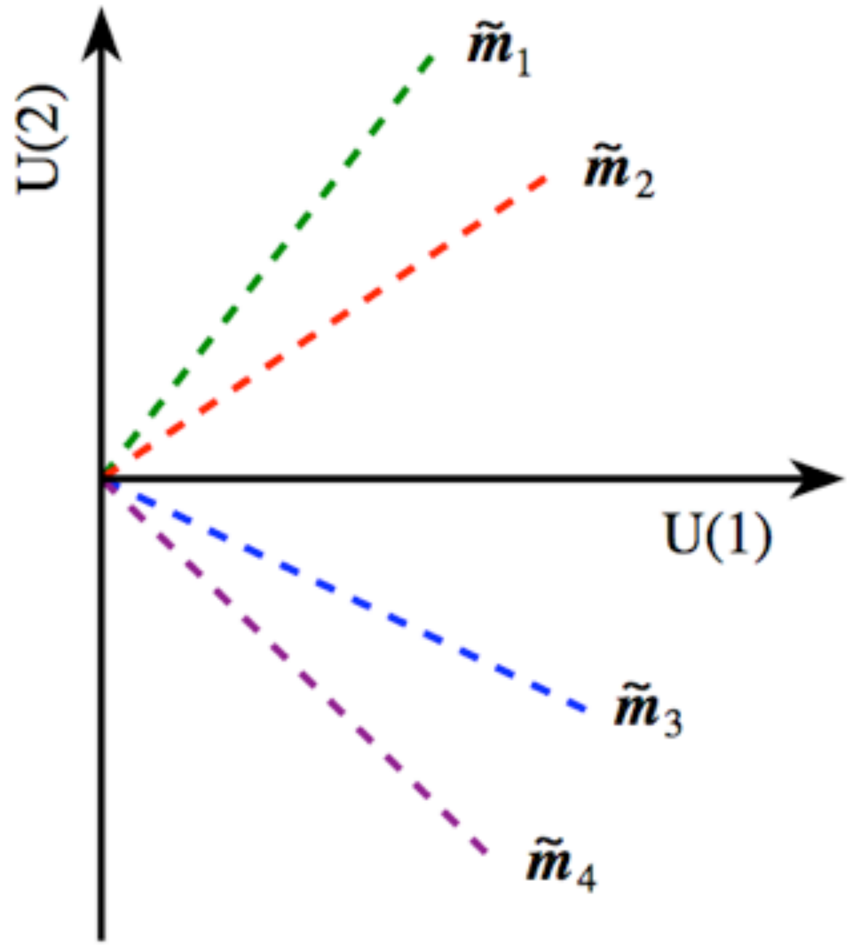
$N_{USA} = 507$

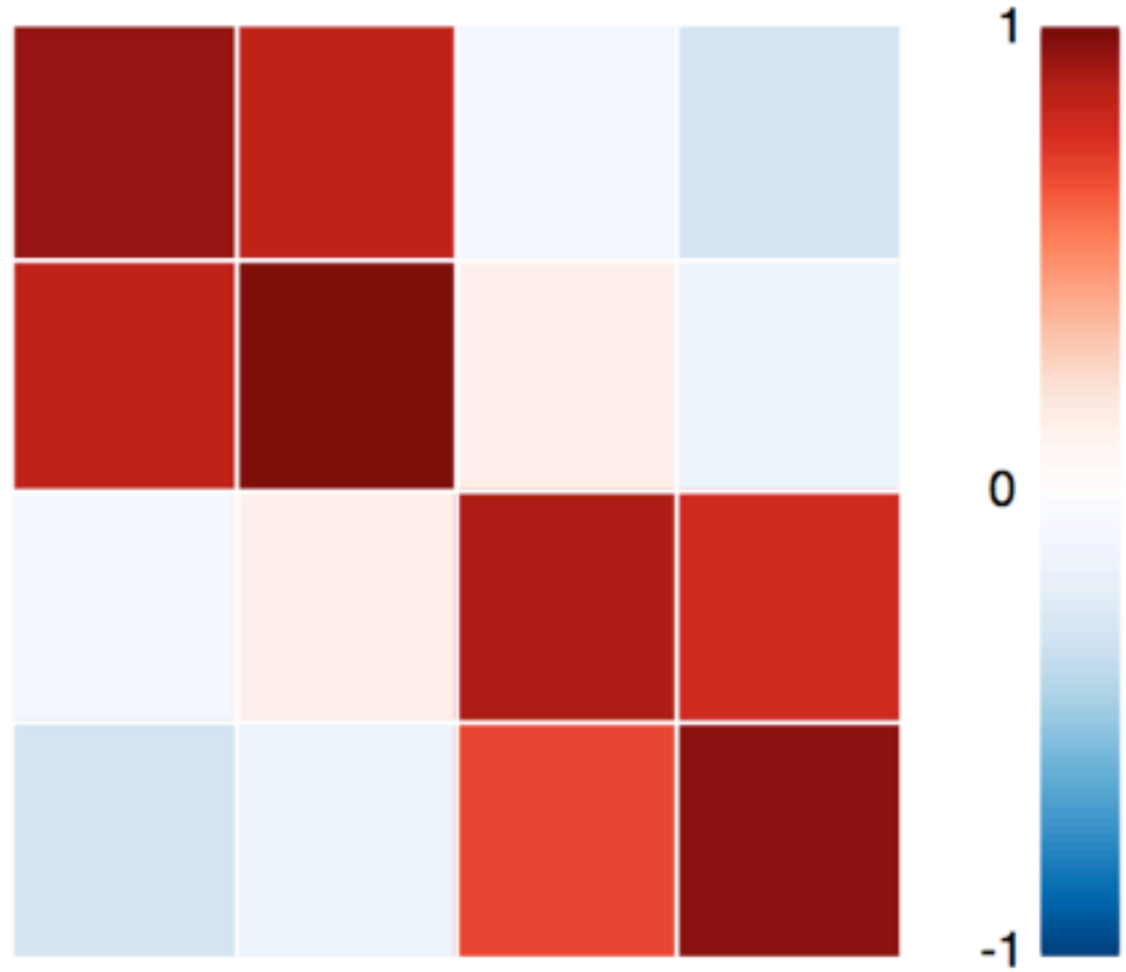
$N_{WE} = 423$

$N_{CA} = 292$



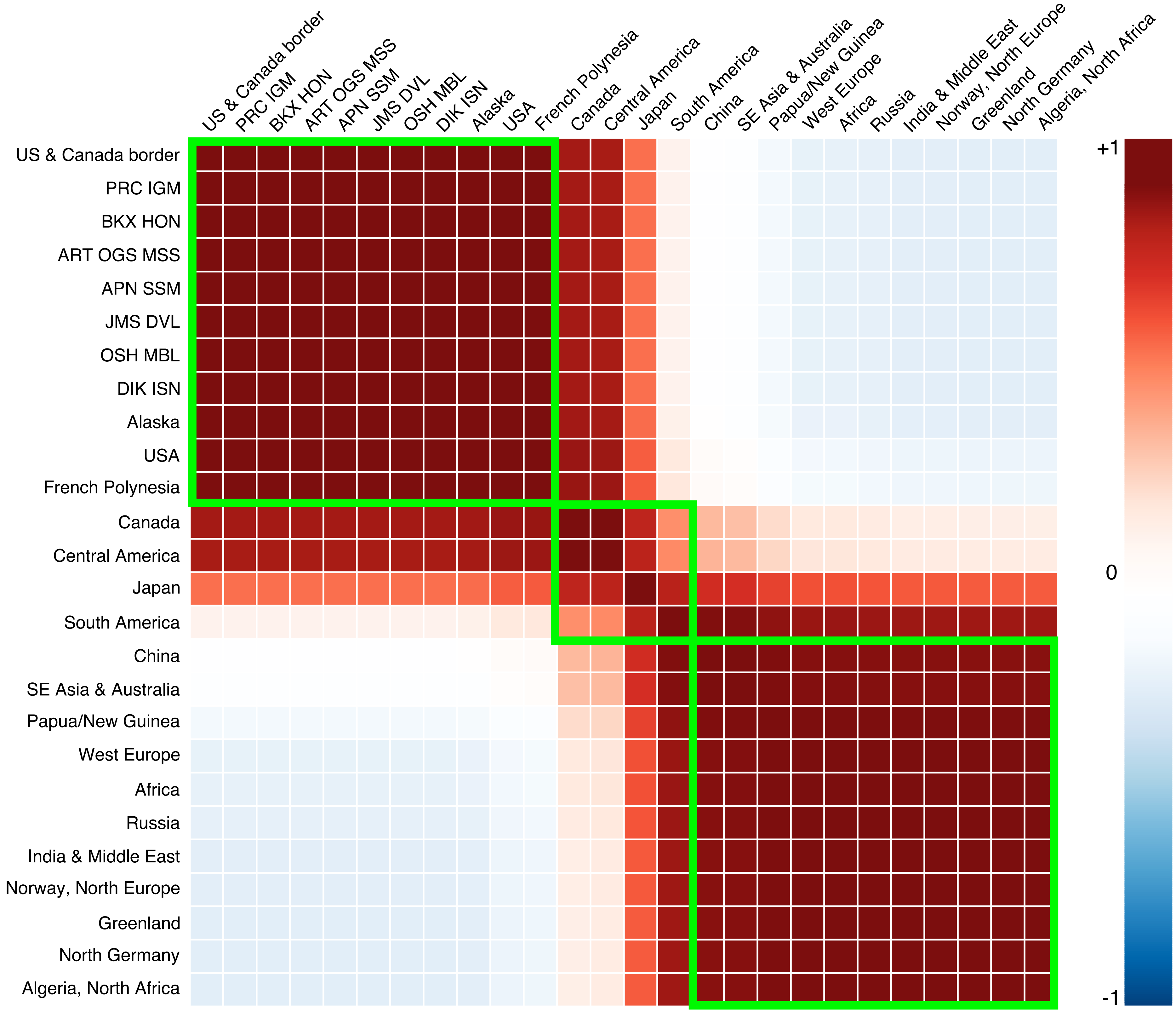
Interpreting TSVD: interrelations between modules





$$\tilde{m}_\alpha = \sum_{i \in \alpha} \tilde{n}_i$$

Interpreting TSVD: interrelations between modules



Summary

M

N

2	0	4	2
2	2	0	0
3	5	0	1
6	6	0	1
4	3	2	3
1	0	0	2
0	0	1	7

C

(contribution matrix)

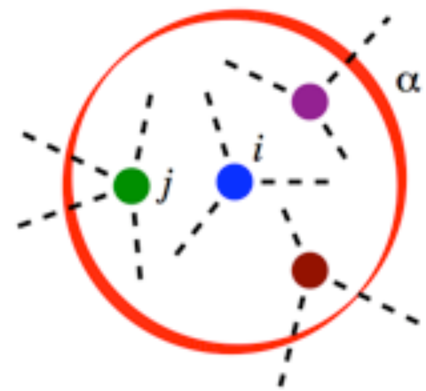
N

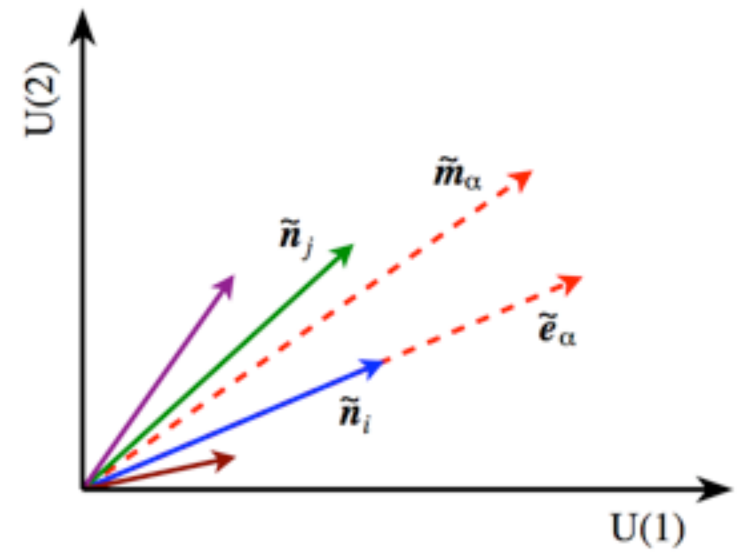
N

.2	-.3	.8	-.2	-.3	.1	0
.2	.1	0	.1	0	.2	.9
.4	.2	-.2	-.7	0	.4	-.1
			⋮			

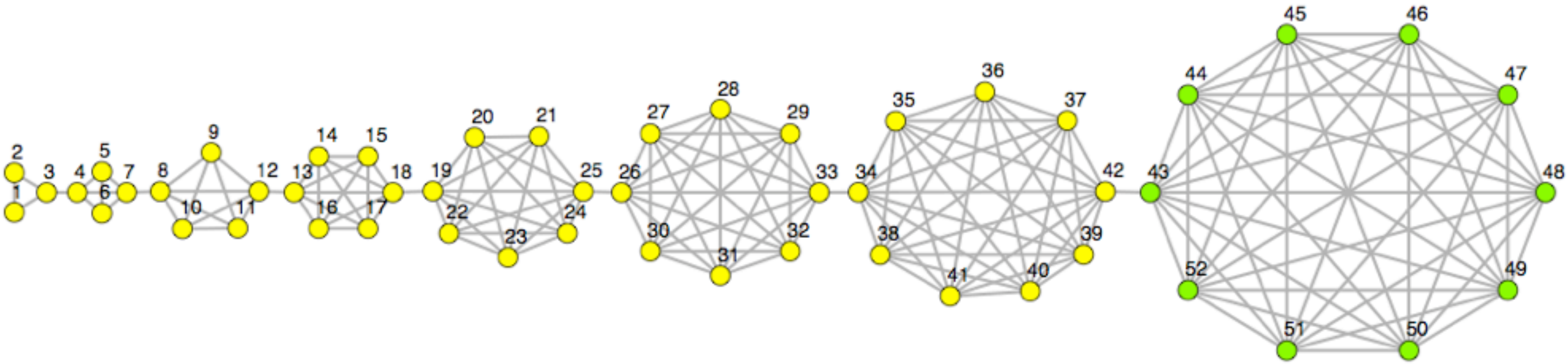
U

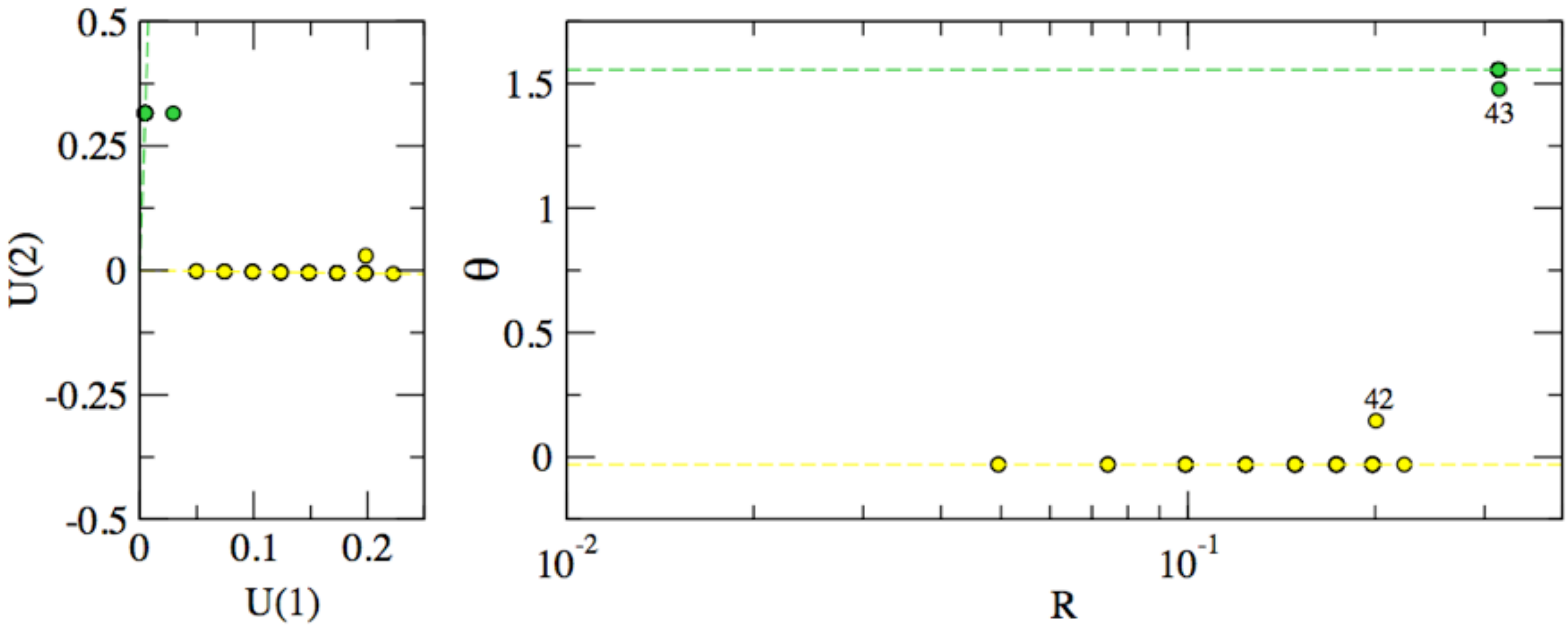
least squares optimal



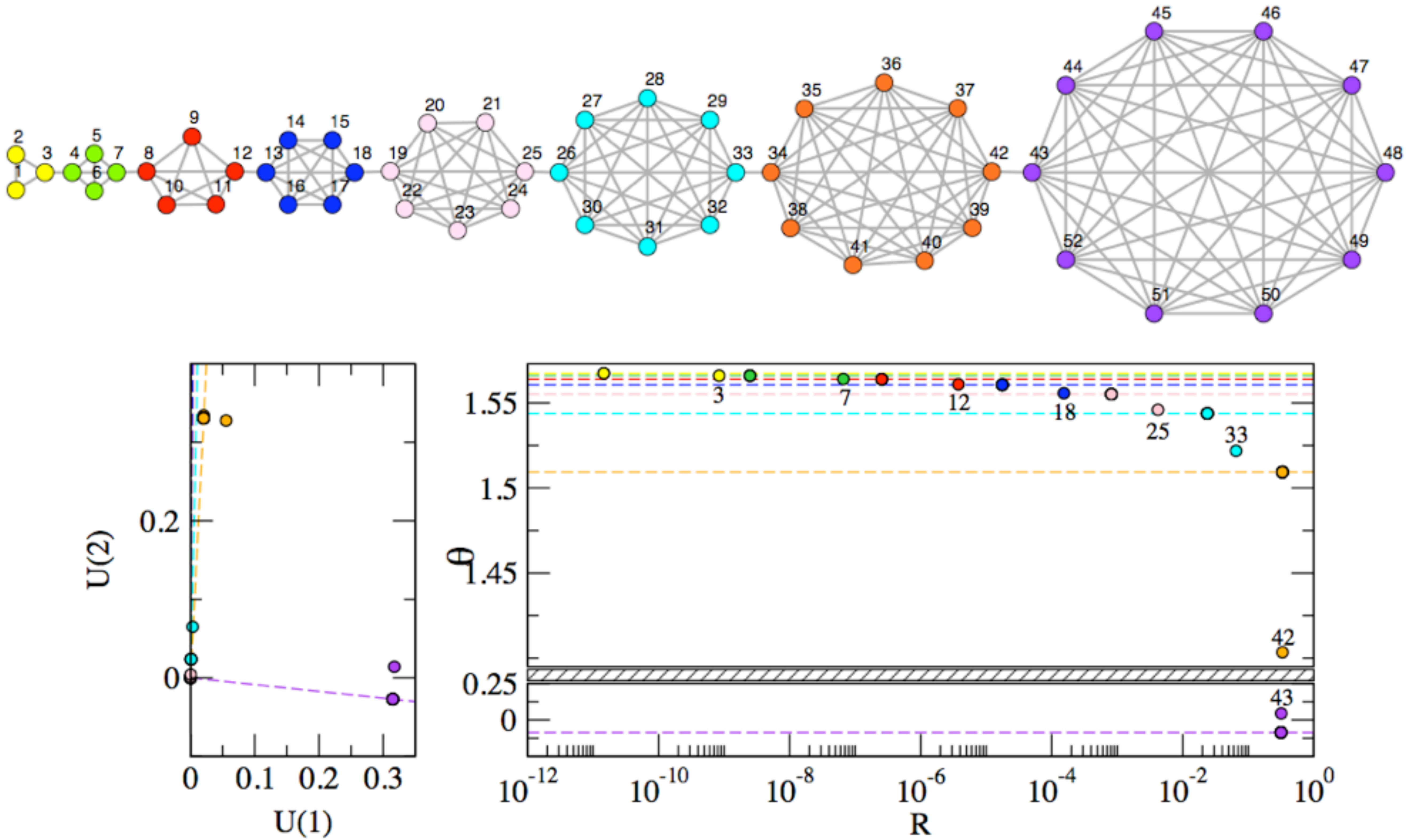


The output of TSVD

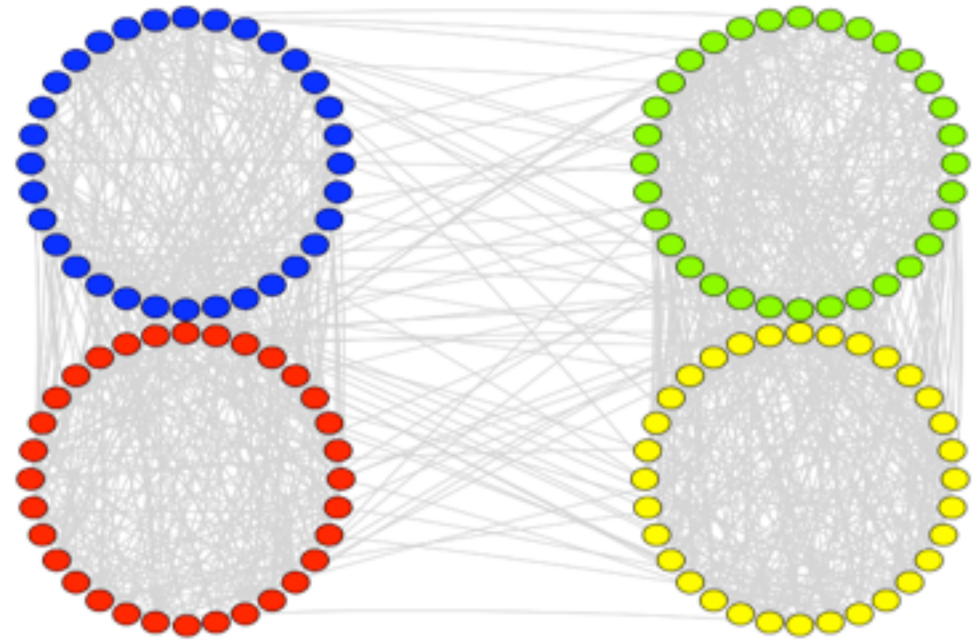


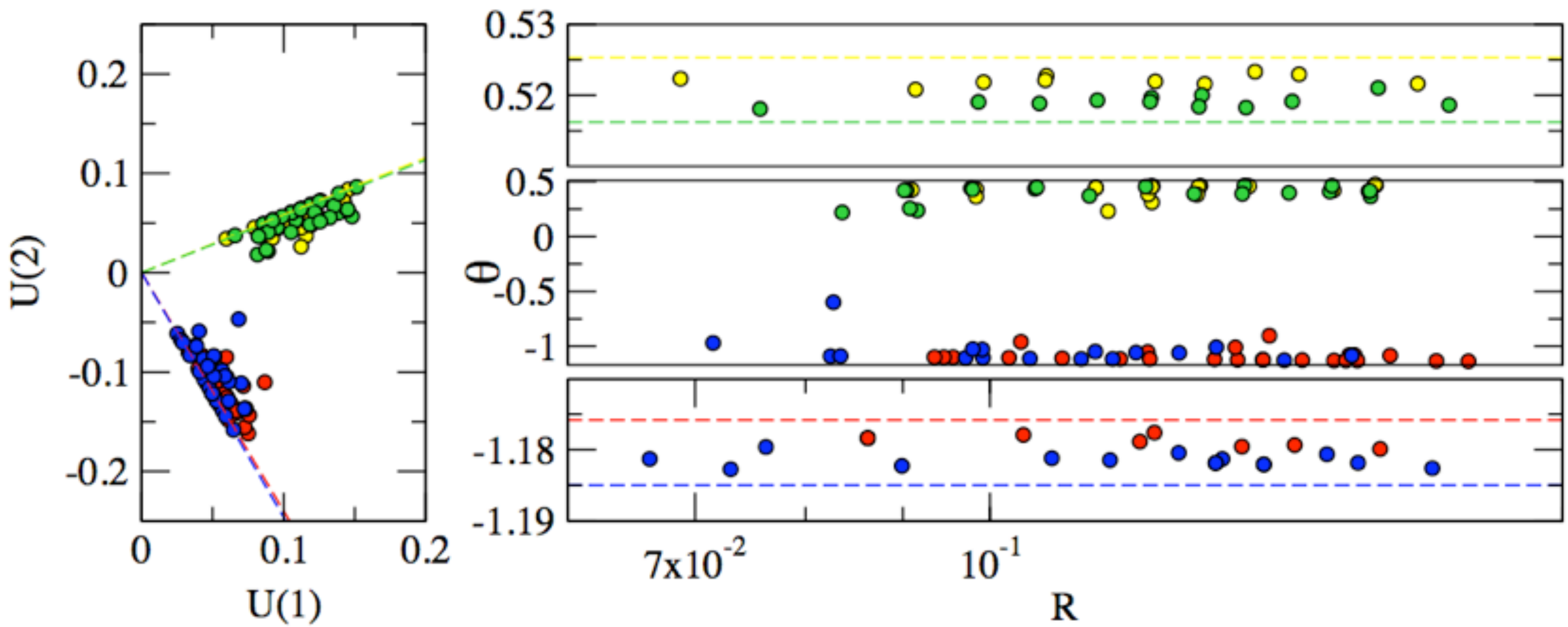


The output of TSVD



The output of TSVD





Information loss

In the case of a rank $r = 2$ approximation, the unicity of the two-ranked decomposition is ensured if singular values satisfy $\sigma_1 > \sigma_2 > \sigma_3$

Loss of information of this projection compared to the initial data by computing the relative difference between the Frobenius norms:

$$E_r = \frac{\|\boldsymbol{C}\|_F - \|\boldsymbol{C_r}\|_F}{\|\boldsymbol{C}\|_F} = \frac{\sum_{\alpha=1}^{M} \sigma_\alpha^2 - \sum_{\alpha=1}^{r} \sigma_\alpha^2}{\sum_{\alpha=1}^{M} \sigma_\alpha^2}$$

airports: 18.2%

AS-P2P: 15.8%

Estructural navigability

R_ext: How externally connected a node is

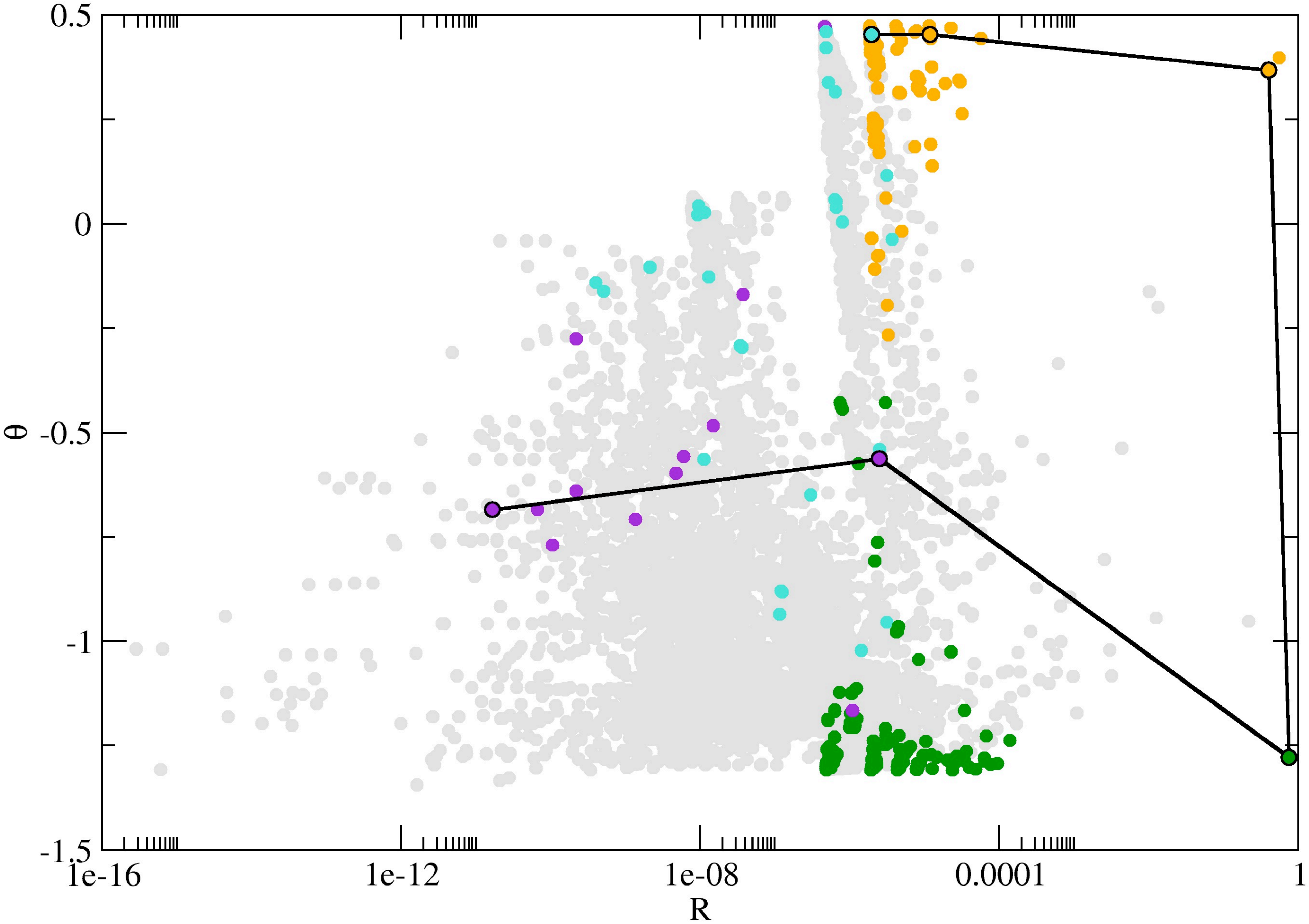
θ to what neighborhood a node belongs

Greedy routing: select a neighbor that minimizes

$$cost_k = \begin{cases} \beta\left(\dfrac{\lambda + |\Delta\theta_{k\to j}|}{R_{int_k}}\right) & \text{if } k \in \alpha_j, \\ \dfrac{|\Delta\theta_{k\to j}|}{R_{ext_k}} & \text{otherwise.} \end{cases}$$

Preliminary results of local routing on the AS network

average path length 5.1. Success ratio 97.2%



Summary

M

N

2	0	4	2
2	2	0	0
3	5	0	1
6	6	0	1
4	3	2	3
1	0	0	2
0	0	1	7

C

(contribution matrix)

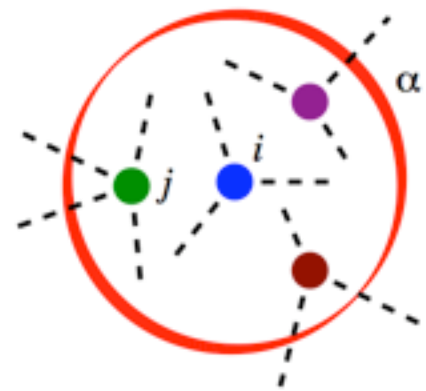
N

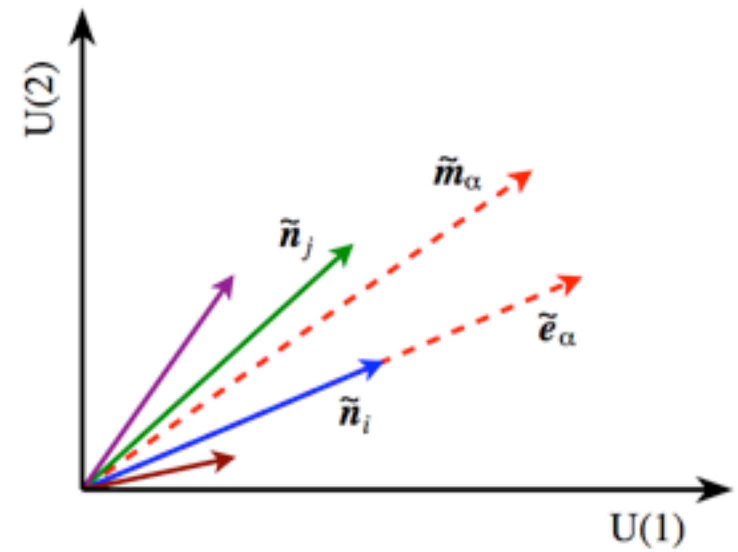
N

.2	-.3	.8	-.2	-.3	.1	0
.2	.1	0	.1	0	.2	.9
.4	.2	-.2	-.7	0	.4	-.1
			⋮			

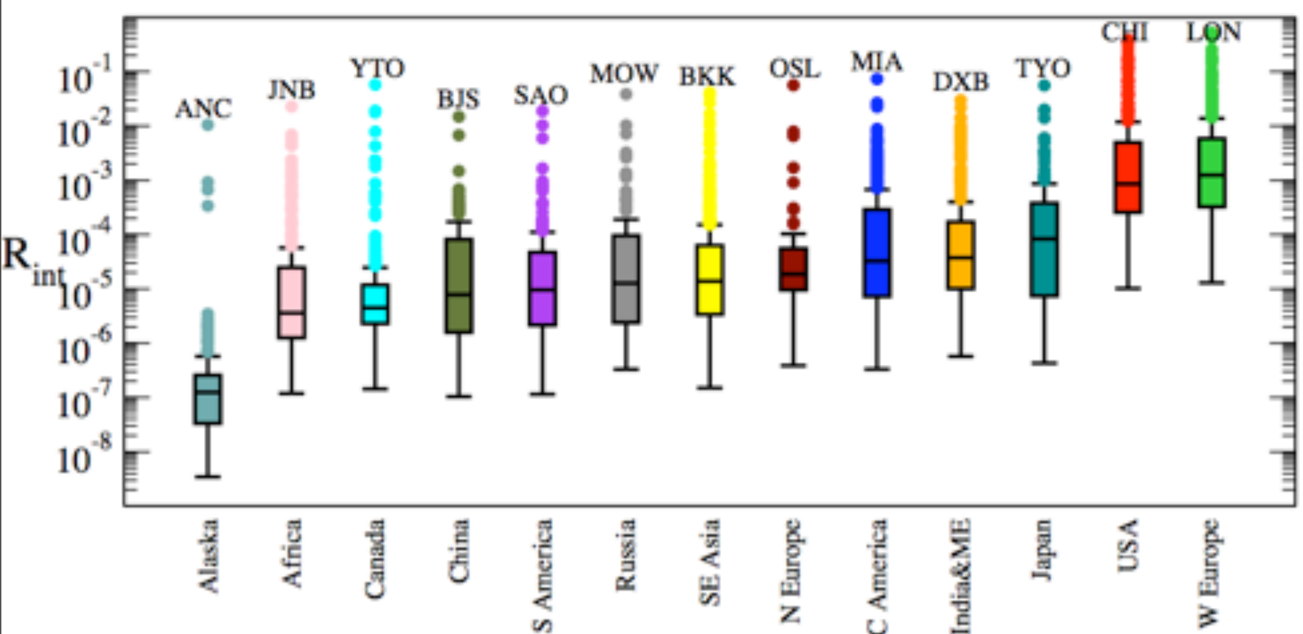
U

least squares optimal

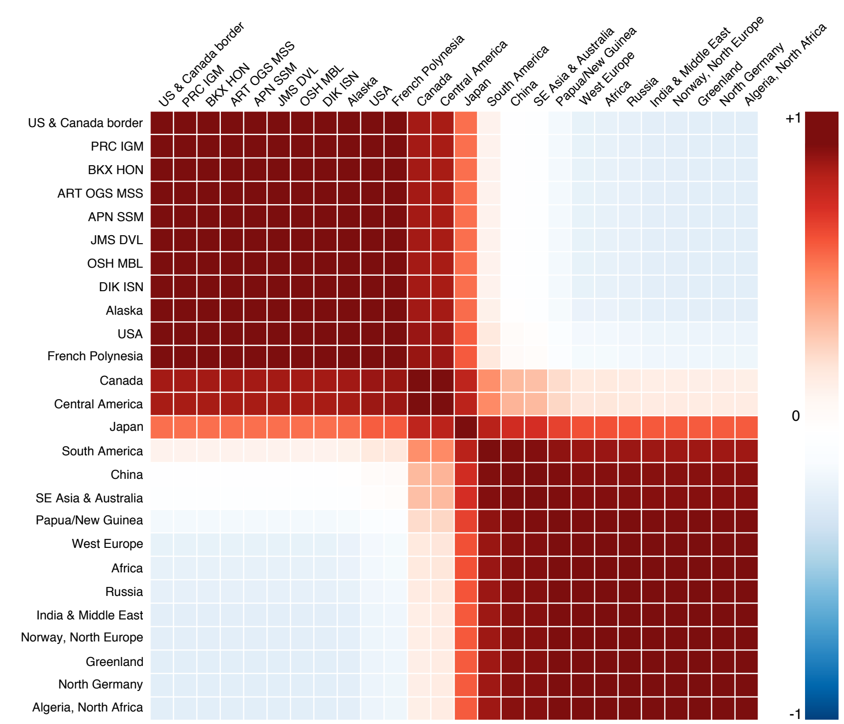




structure of individual modules



interrelations between modules



Thanks for your attention

A. Arenas, J. Borge-Holthoefer, S. Gómez and G. Zamora

New Journal of Physics, 12, 053009 (2010)