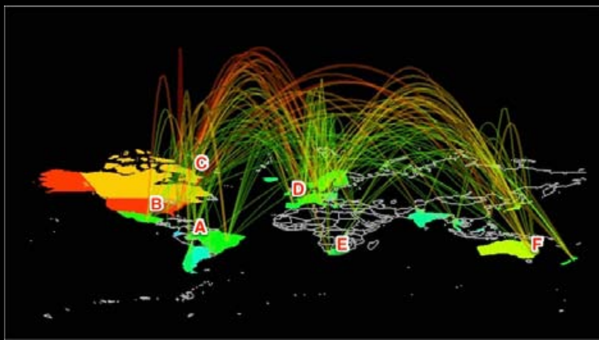
Spectral method for finding dominant graph structures: from undirected to directed graphs

Nataša Djurdjevac Conrad

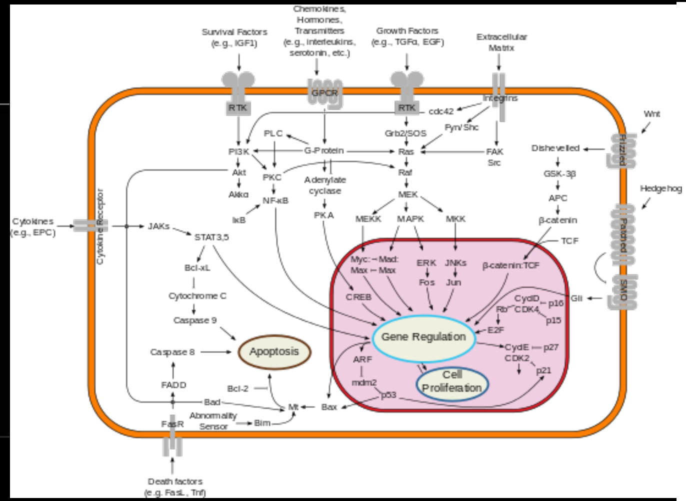
joint work with Marcus Weber and Christof Schütte

Spectra of graphs and applications 2016

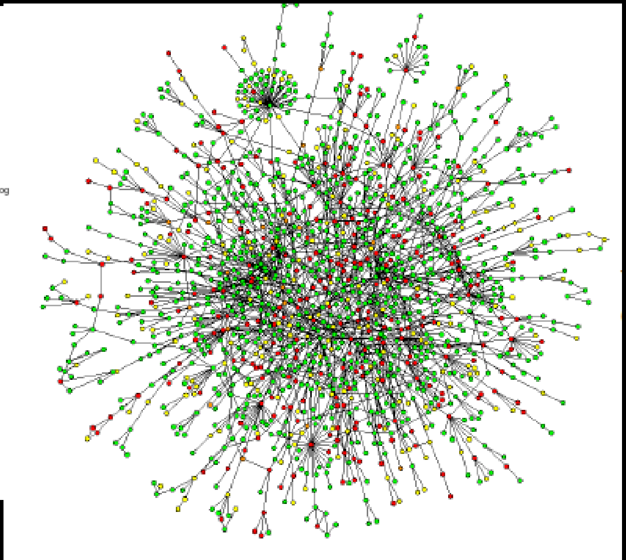
20.05.2016 Belgrade



Internet (www.cellbiol.com)



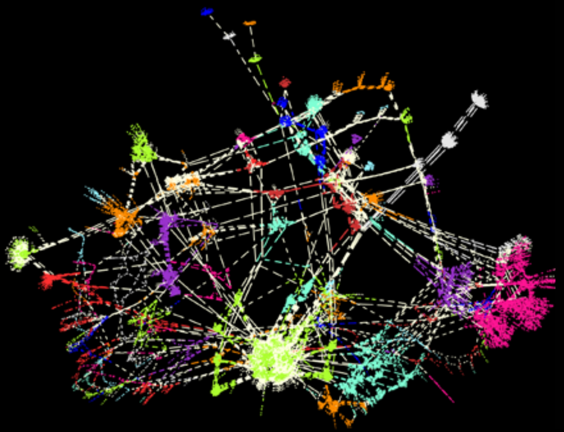
Metabolic network
www.wikipedia.com



Map of yeast protein-protein interactions,
by Hawoong Jeong

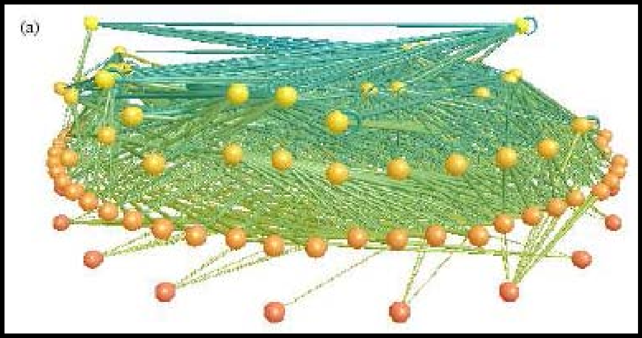
Networks describe a wide range of complex systems in various fields. Understanding the behavior of complex systems starts with understanding the topology of corresponding networks.

Social network
(http://www.facebook.com)



Transportation network

Human brain network
(www.sciencephoto.com)

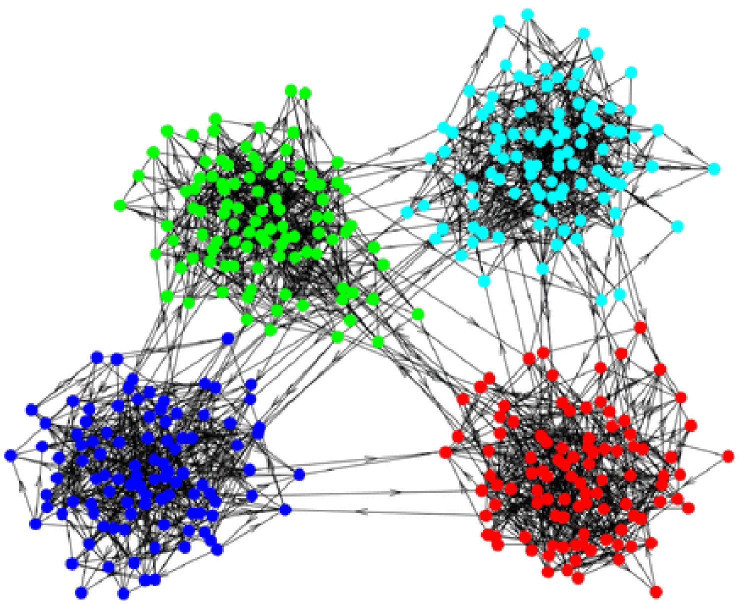


Food Web (http://www.foodwebs.org)

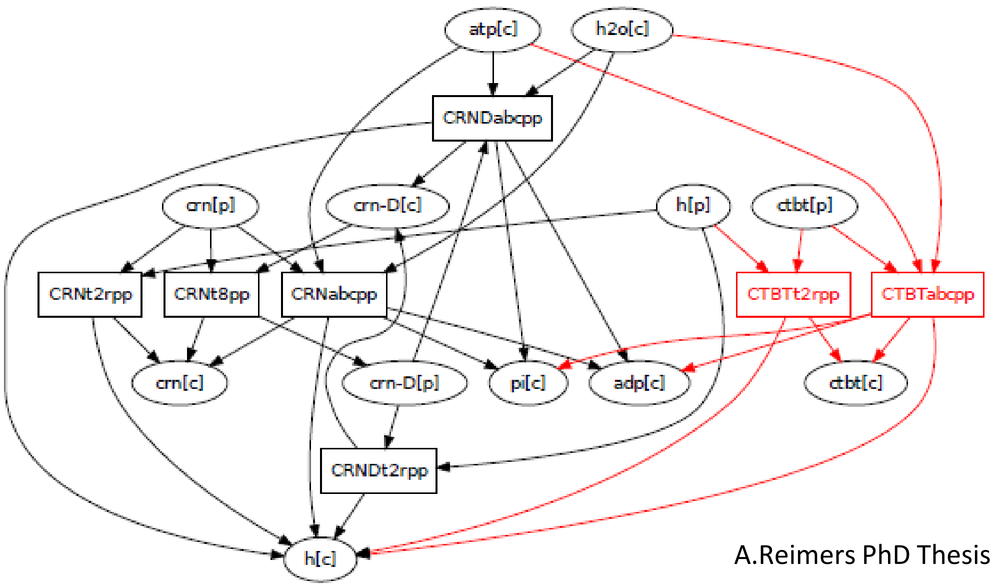
Finding dominant structures

<u>Graph theory:</u> **Modules (communities, clusters)** are densely connected subgraphs that are loosely interconnected with the rest of the network.

<u>Applications:</u> Groups of nodes which share common properties in the system.
→**Biological systems: functional units (protein complexes, biochemical pathways)**



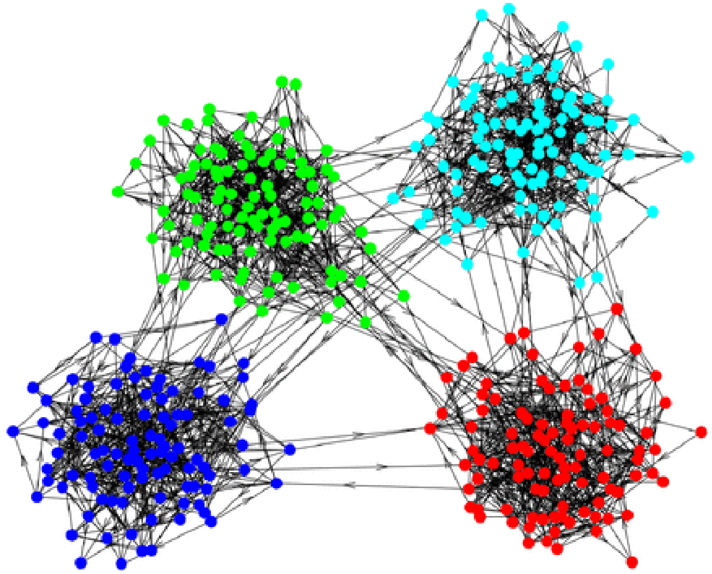
Beber et al. 2012 Interface



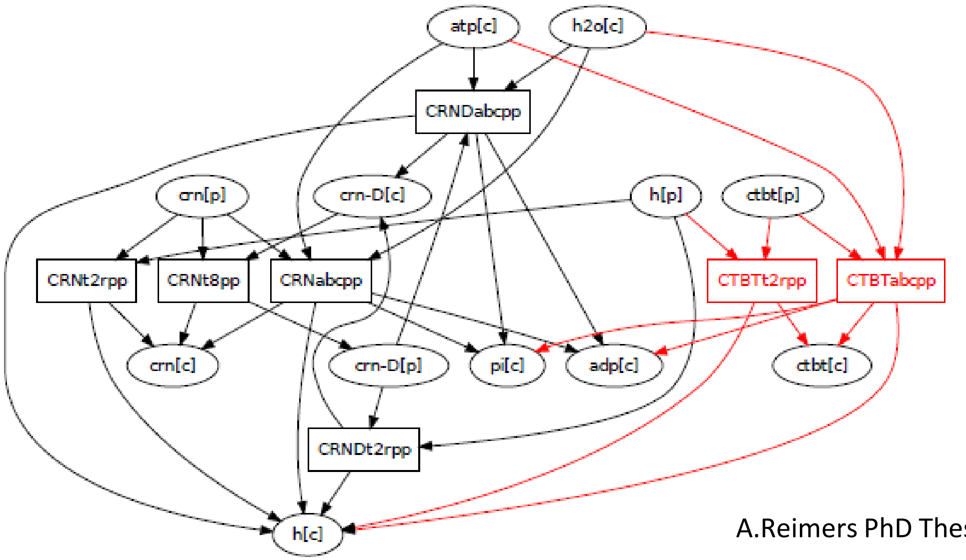
A.Reimers PhD Thesis

What are dominant graph structures?

- Modules
- Dominant (important) cycles/pathways



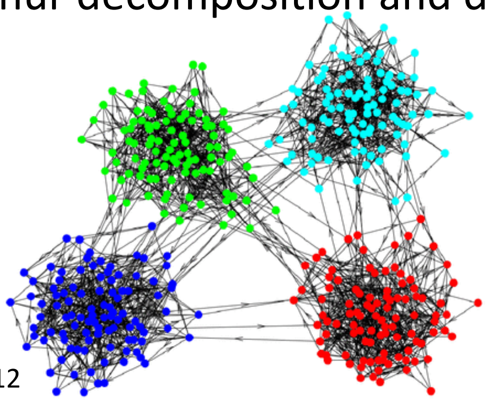
Beber et al. 2012 Interface



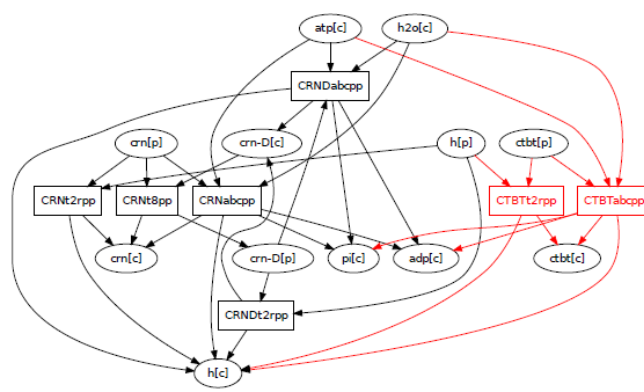
A.Reimers PhD Thesis

Outline:

- **What are dominant network structures?**
  - definition in terms of a random walk process
  - relations to spectral properties
- **How do dominant structures appear?**
  - Eigenvalue perturbation:
    - real to complex eigenvalues
    - 1- and 2-cycle perturbations
- **How can we find dominant network structures?**
  - Schur decomposition and dominant structures



Beber et al. 2012
Interface



A.Reimers PhD Thesis

Dominant network structures: random walk process

Graph $G = (V, E)$ is defined with:

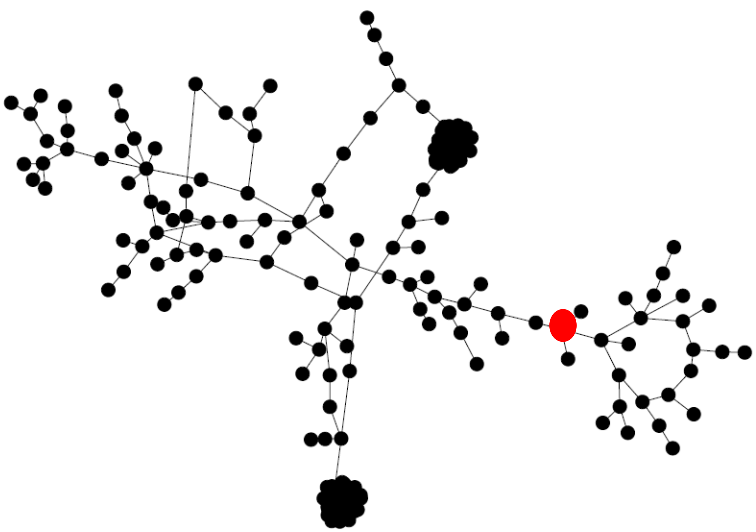
- a set of nodes (states) $V = \{1, \dots, n\}$,
- a set of edges $E \subset V \times V$.

Adjacency matrix A of a graph G is

$$A(x,y) = \begin{cases} > 0, & (x,y) \in E; \\ = 0, & (x,y) \notin E. \end{cases}$$

Introduce a random walk process (RW) on a graph G as a Markov chain $(X_n)_{n \in \mathbb{N}}$, with stochastic matrix P

$$P(x,y) = \frac{A(x,y)}{d(x)}, \quad d(x) = \sum_{y \in V} A(x,y).$$



Transition rule:

In every time-step, the process jumps to one of his neighbors with the same probability.

Dominant structures are metastable sets of a RW process.

Dominant network structures:

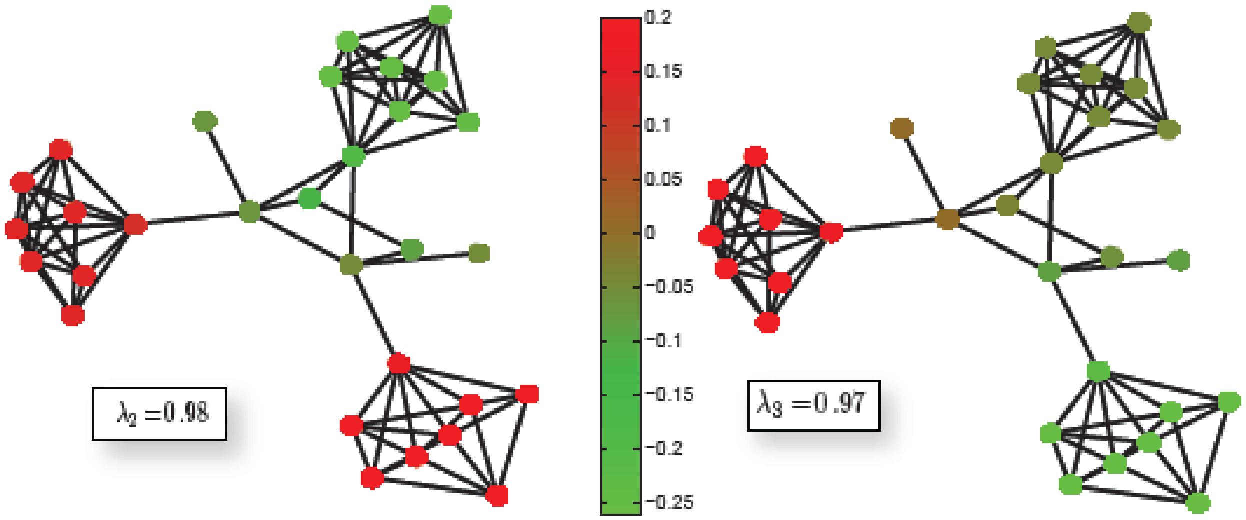
relations to spectral properties

- In terms of the random walk process on a graph, modules are metastable sets.

Undirected graphs:

- Use spectral clustering methods to identify modules: number of dominant eigenvalues of P represent number of modules; corresponding eigenvectors indicate the elements of modules.

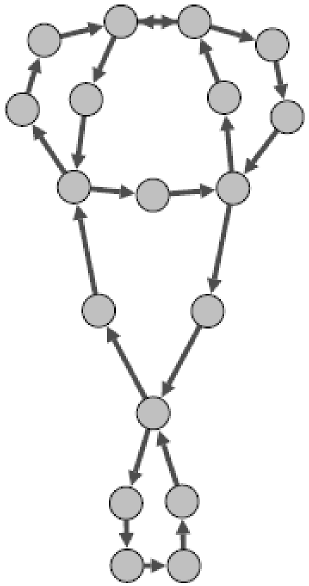
$$\lambda_1 = 1 \quad \lambda_2 = 0.98 \quad \lambda_3 = 0.97 \quad \lambda_4 = 0.76 \quad \lambda_5 = 0.41 \dots$$



Dominant network structures:
relations to spectral properties

Directed graphs:

However, for directed graphs P can have complex eigenvalues due to cycles that appear in the graph.
Metastability of non-reversible Markov processes is **NOT** very well understood.
Dominant spectrum is connected to modules and dominant cycles.



i	λ_i
1	1
2	0.7156 + 0.0053i
3	0.7156 - 0.0053i
4	0.3593 + 0.8579i
5	0.3593 - 0.8579i
6	0.3260 + 0.6374i
7	0.3260 - 0.6374i

Outline:

- **What are dominant network structures?**
  - definition in terms of a random walk process
  - relations to spectral properties
- **How do dominant structures appear?**
  - Eigenvalue perturbation:
    - real to complex eigenvalues
    - 1- ,2- and 3-cycle perturbations
- **How can we find dominant network structures?**
  - Schur decomposition and dominant structures

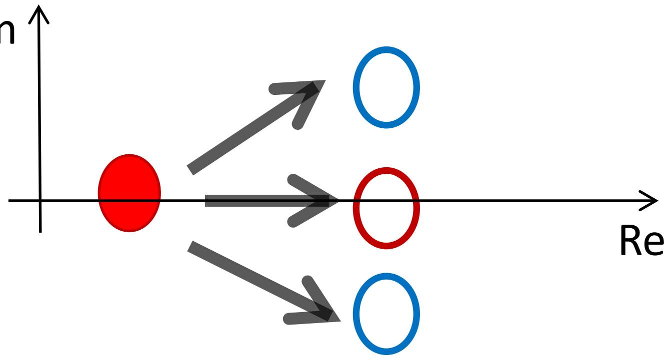
Eigenvalue perturbation:
real to complex eigenvalues

Consider the perturbation

$$P_\epsilon = P + \epsilon L$$

with a generator matrix L, $L_{i,j} \geq 0, i \neq j$ and $\sum_j L_{ij} = 0, \forall i \in V$.
We assume the following asymptotic expansions

$$\begin{aligned} \lambda_j^\epsilon &= \lambda_j + \epsilon \eta_j + \mathcal{O}(\epsilon^2) \\ u_j^\epsilon &= \tilde{u}_j + \epsilon v_j + \mathcal{O}(\epsilon^2). \end{aligned}$$



Then, one can show:

- simple eigenvalue λ_j: $\eta_j = \langle u_j, L u_j \rangle$ - a shift of λ_j along the real line.
- degenerate eigenvalue $\lambda_j = \lambda_{j+1}$: $\begin{cases} \text{remain two-fold;} \\ \text{split along real line;} \\ \textbf{split into complex plane.} \end{cases}$

Eigenvalue perturbation:
real to complex eigenvalues

For a two-fold eigenvalue $\lambda_j = \lambda_{j+1}$ and $\lambda_j^\epsilon = \lambda_j + \epsilon\eta_j + \mathcal{O}(\epsilon^2)$, η_j are given by the eigenvalues of the matrix

$$\hat{L} = \begin{pmatrix} \langle u_j, Lu_j \rangle & \langle u_j, Lu_{j+1} \rangle \\ \langle u_{j+1}, Lu_j \rangle & \langle u_{j+1}, Lu_{j+1} \rangle \end{pmatrix}.$$

Using the anti-symmetric part $\hat{L}^A = (\hat{L} - \hat{L}^T)/2$ of $\hat{L}$,

$$\hat{L}^A = \frac{1}{2}\begin{pmatrix} 0 & \delta \\ -\delta & 0 \end{pmatrix}, \delta = \hat{L}_{12} - \hat{L}_{21}.$$

we can show that the perturbation drives the two-fold real eigenvalue into the complex plane iff the deviation from symmetry of $\hat{L}$ is strong enough

$$\delta^2 > \frac{1}{2}(\hat{L}_{11} - \hat{L}_{22})^2 + \hat{L}_{12}^2 + \hat{L}_{21}^2. \qquad (1)$$

Eigenvalue perturbation: real to complex eigenvalues

Consequence: perturbations of a single entry of P or of a 2-cycle can never move a double eigenvalue into the complex plane.

Example: Let us consider the reversible transition matrix

$$P = \begin{pmatrix} 0.8566 & 0.1195 & 0.0239 \\ 0.0566 & 0.9195 & 0.0239 \\ 0.0566 & 0.1195 & 0.8239 \end{pmatrix},$$

with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = 0.8$.
Perturbation of P using the 2-cycle $\gamma = (1,2)$:

$$L_1 = \sigma_1 \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \sigma_1 = 0.1242.$$

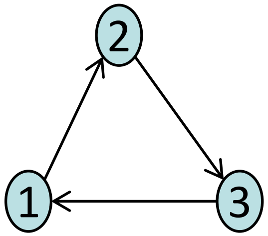
Then, eigenvalues of $P_\epsilon = P + \epsilon L_1$ are real-valued:

$$\lambda_1 = 1, \lambda_2 = 0.8, \lambda_3 \begin{cases} < 0.8 & \text{if } \epsilon > 0 \\ = 0.8 & \text{if } \epsilon = 0 \\ > 0.8 & \text{if } \epsilon < 0 \end{cases}$$

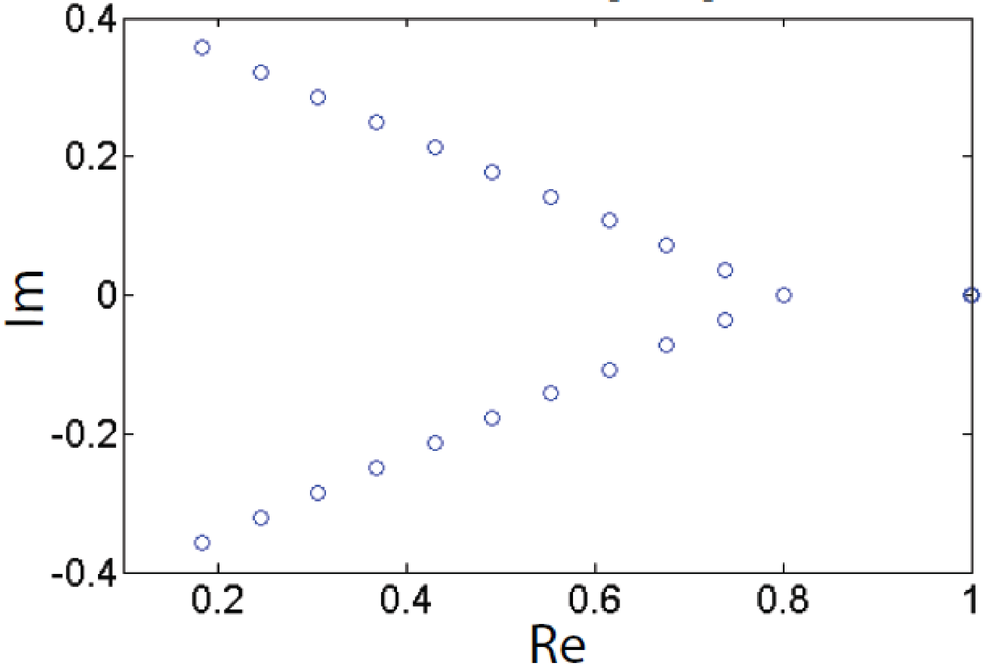
Eigenvalue perturbation:
real to complex eigenvalues

Perturbation of P using the 3-cycle $\gamma = (1,2,3)$:

$$L_2 = \sigma_2 \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}, \quad \sigma_1 = 0.4119.$$



Then, eigenvalues of $P_\epsilon = P + \epsilon L_2, \epsilon \in [0,1]$ are



Eigenvalue perturbation:
real to complex eigenvalues

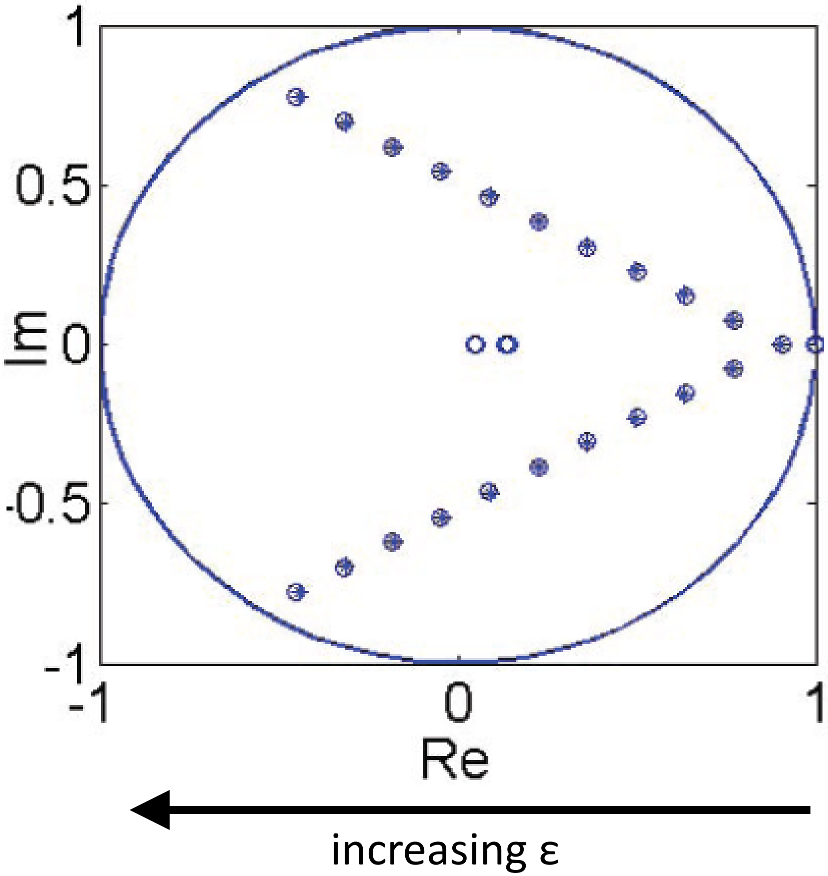
From metastable sets to dominant cycles: Let us consider the reversible transition matrix

$$P = \begin{pmatrix} 0.1437 & 0.0401 & 0.3494 & 0.4344 & 0.0323 \\ 0.0189 & 0.0740 & 0.2585 & 0.4861 & 0.1625 \\ 0.0074 & 0.0115 & 0.9385 & 0.0184 & 0.0242 \\ 0.0130 & 0.0307 & 0.0261 & 0.9094 & 0.0209 \\ 0.0010 & 0.0109 & 0.0365 & 0.0223 & 0.9292 \end{pmatrix},$$

with eigenvalues $\lambda = 1,\ 0.9048,\ 0.9048,\ 0.1353,\ 0.0498.$

Eigenvalue perturbation:
real to complex eigenvalues

After the perturbation of a 3-cycle $\gamma = (3,4,5)$: $L = \sigma(C_\gamma - I_\gamma)$, $\sigma = 0.9$. the eigenvalues of $P_\epsilon = P + \epsilon L, \epsilon \in [0,1]$ are

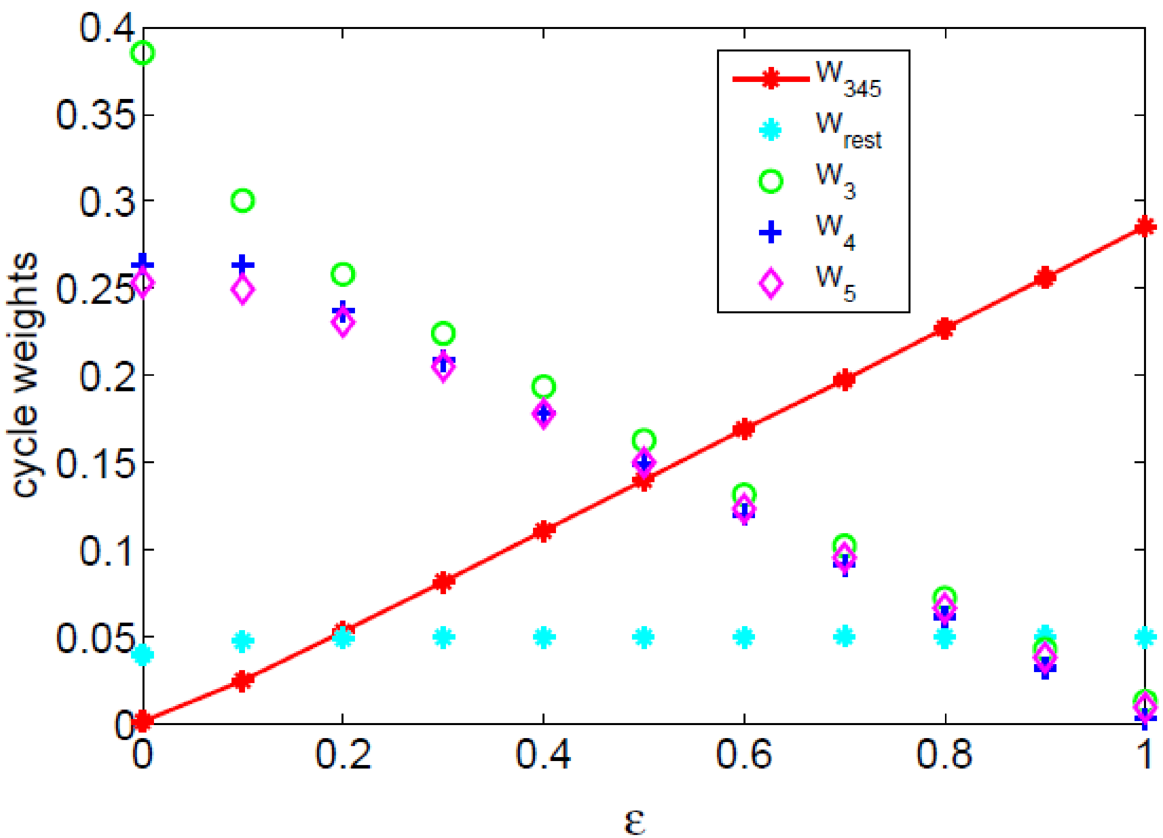


$$P = \begin{pmatrix} 0.1437 & 0.0401 & 0.3494 & 0.4344 & 0.0323 \\ 0.0189 & 0.0740 & 0.2585 & 0.4861 & 0.1625 \\ 0.0074 & 0.0115 & 0.9385 & \boxed{0.0184} & 0.0242 \\ 0.0130 & 0.0307 & 0.0261 & 0.9094 & \boxed{0.0209} \\ 0.0010 & 0.0109 & \boxed{0.0365} & 0.0223 & 0.9292 \end{pmatrix}$$

Eigenvalue perturbation:

real to complex eigenvalues

The change of dominant structures from **three metastable sets (for $\epsilon = 0$) to a dominant cycle $(3, 4, 5)$ (for $\epsilon = 1$)** can be observed by following changes of cycle weights (induced by visits of the RW).



We can observe module/cycle transitions using our cycle weight function

Outline:

- **What are dominant network structures?**
  - definition in terms of a random walk process
  - relations to spectral properties
- **How do dominant structures appear?**
  - Eigenvalue perturbation:
    - real to complex eigenvalues
    - 1- and 2-cycle perturbations
- **How can we find dominant network structures?**
  - Schur decomposition and dominant structures

Perturbation of the Schur decomposition

The Schur decomposition (X_ϵ, R_ϵ) of P_ϵ has the form

$$P_\epsilon X_\epsilon = X_\epsilon R_\epsilon, \quad \text{s.t.} \quad X_\epsilon^T D_\epsilon X_\epsilon = \mathrm{Id},$$

so that the columns of X_ϵ form an $\langle\cdot,\cdot\rangle$-orthonormal basis (as the eigenvectors did for reversible P_ϵ) and where R_ϵ has the following form

$$R_\epsilon = \begin{pmatrix} B_\epsilon & U_\epsilon \\ 0 & \tilde{\Lambda}_\epsilon \end{pmatrix}.$$

Theorem

Up to first order in ϵ the upper 2×2 block of the Schur form R_ϵ of P_ϵ is given by $\lambda \mathrm{Id}_{2\times 2} + \epsilon \hat{L}_s$ where $\hat{L}_s$ denotes the Schur form of $\hat{L}$ (and thus has the same eigenvalues). Thus, the Schur form $\hat{L}_s$ of the projection of L to the eigenspace of a double eigenvalue of the reversible matrix P is sufficient to see whether the perturbation leads to a complex conjugated pair of eigenvalues.

Perturbation of the Schur decomposition

The Schur decomposition (X_ϵ, R_ϵ) of P_ϵ has the form

$$P_\epsilon X_\epsilon = X_\epsilon R_\epsilon, \quad \text{s.t.} \quad X_\epsilon^T D_\epsilon X_\epsilon = \mathrm{Id},$$

so that the columns of X_ϵ form an $\langle \cdot, \cdot \rangle$-orthonormal basis (as the eigenvectors did for reversible P_ϵ) and where R_ϵ has the following form

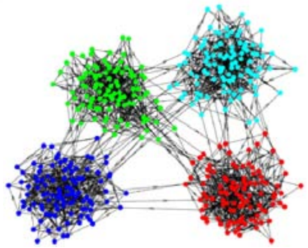
$$R_\epsilon = \begin{pmatrix} B_\epsilon & U_\epsilon \\ 0 & \tilde{\Lambda}_\epsilon \end{pmatrix}.$$

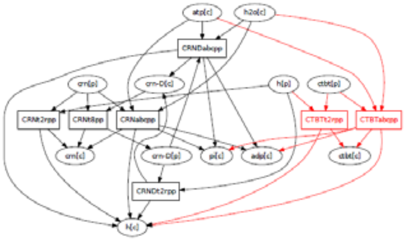
Our approach:

- Analysis: we can use Schur form R_ϵ to conclude whether the perturbation L forces a pair of eigenvalues into the complex plane.
- Algorithm: we use leading Schur vectors X_ϵ as membership functions in spectral clustering (PCCA+).

Summary:

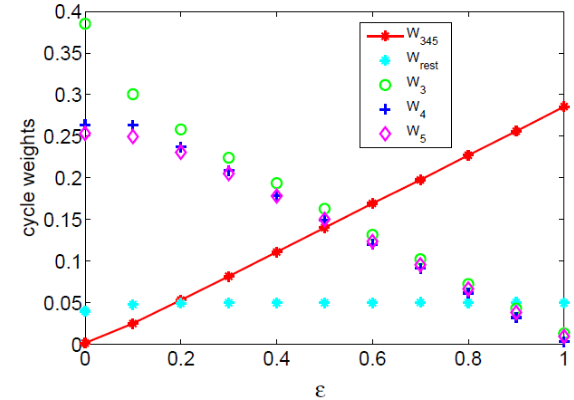
Applications:



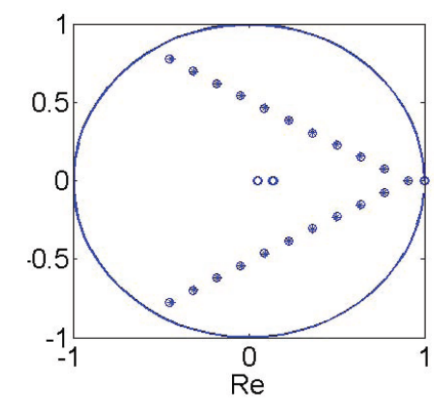


Dominant network structures can help us find important parts of complex systems (protein complexes, biochemical pathways)

Theory:



- Dominant network structures:
  - often visited by random walk process
  - relations to spectral properties: modules vs. dominant cycles
- Eigenvalue perturbation:
  - real to complex eigenvalues using cycle petrurbations
  - measure the importance of a structure by cycle weights
- Schur decomposition and dominant structures
  - alternative to the spectral decomposition
  - encodes important perturbation information



References:

- N. Djurdjevac Conrad, Marcus Weber and Ch. Schütte, ***Finding dominant structures of nonreversible Markov processes***, submitted 2015.
- R. Banisch and N. Djurdjevac Conrad, ***Cycle flow based module detection in directed recurrence networks****, EPL 2014.*
- Marcus Weber, Konstantin Fackeldey, ***G-PCCA: Spectral Clustering for Non-reversible Markov Chains***, submitted 2015.

Thank you

for your attention

Questions?

Email: natasa.conrad@zib.de