Module Detection in Directed Recurrence Networks

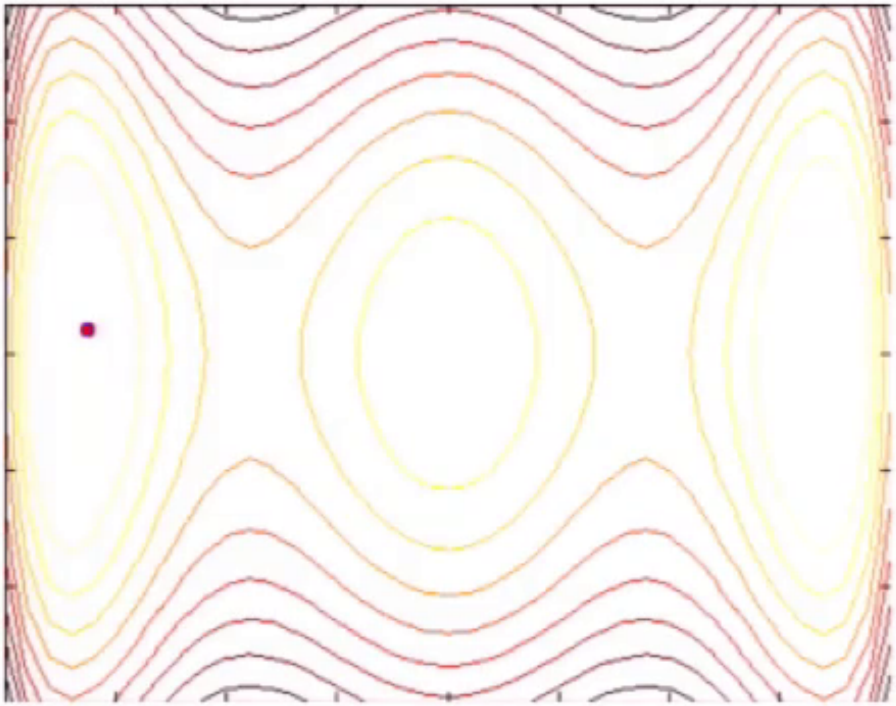
Ralf Banisch

Joint work with Natasa Conrad

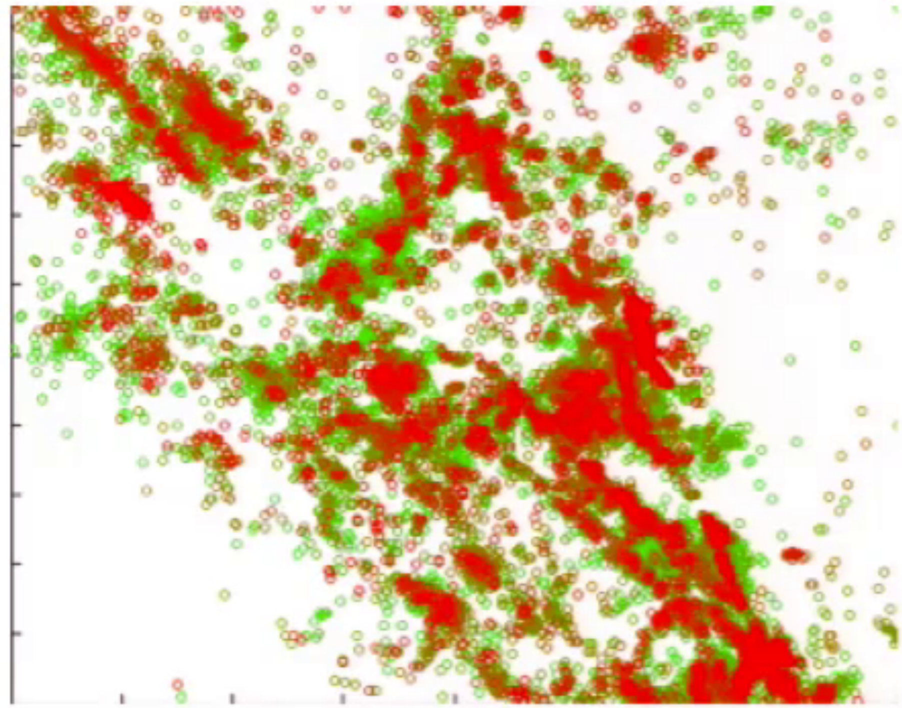
Berlin Mathematical School

Freie Universität Berlin

Motivation: What are time series?



Brownian particle in a potential landscape



earthquake events in Southern California

Motivation: Recurrence Networks

time series $\mathbf{x}_{[0,T]} = \{x_0, \ldots, x_T\}$

$\downarrow$

choose embedding in a metric space (Ω, d)

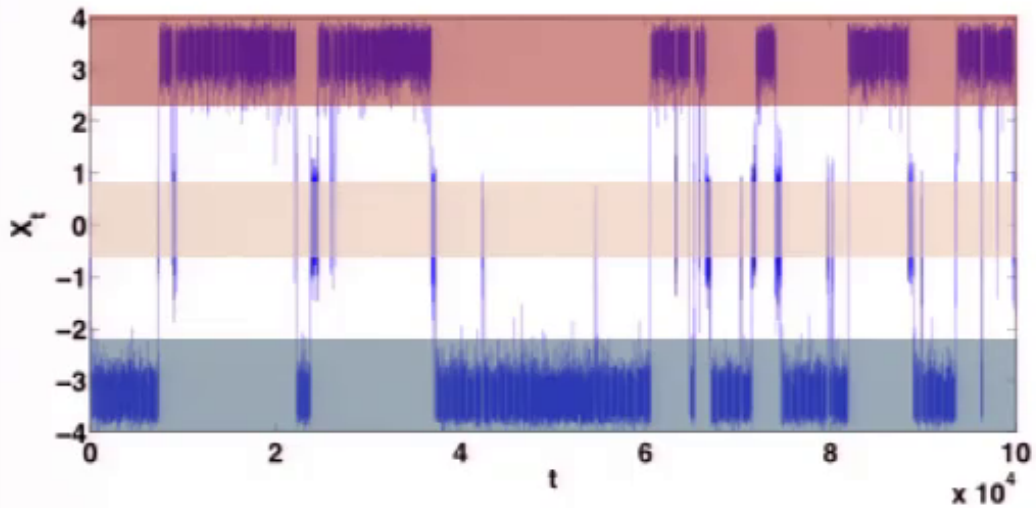
$\downarrow$

construct recurrence network G

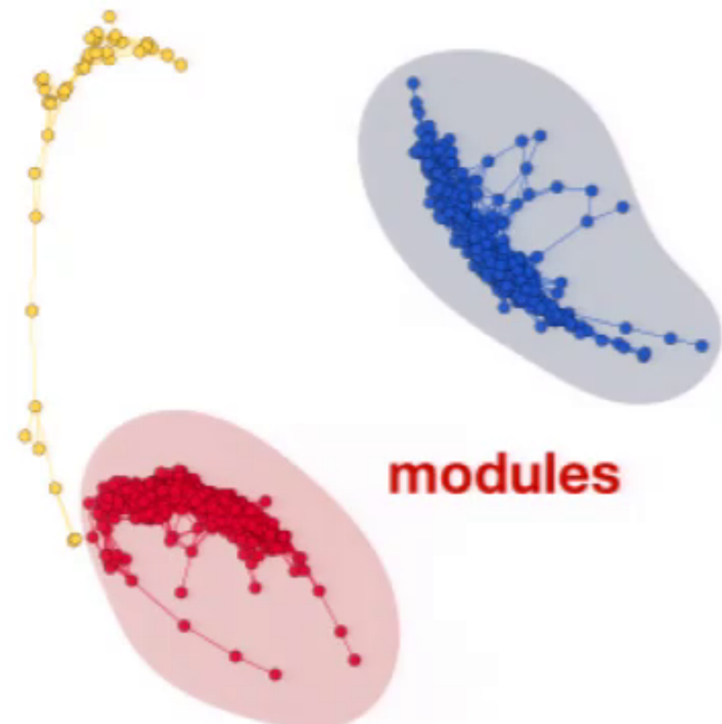
Idea: infer structure in the data from structure in G

Motivation: Recurrence networks

structured data



structure in G



Reminder: From time series to networks

time series $\mathbf{x}_{[0,T]} = \{x_0, \ldots, x_T\}$

$\downarrow$

choose embedding in a metric space (Ω, d) [e.g. Takens]

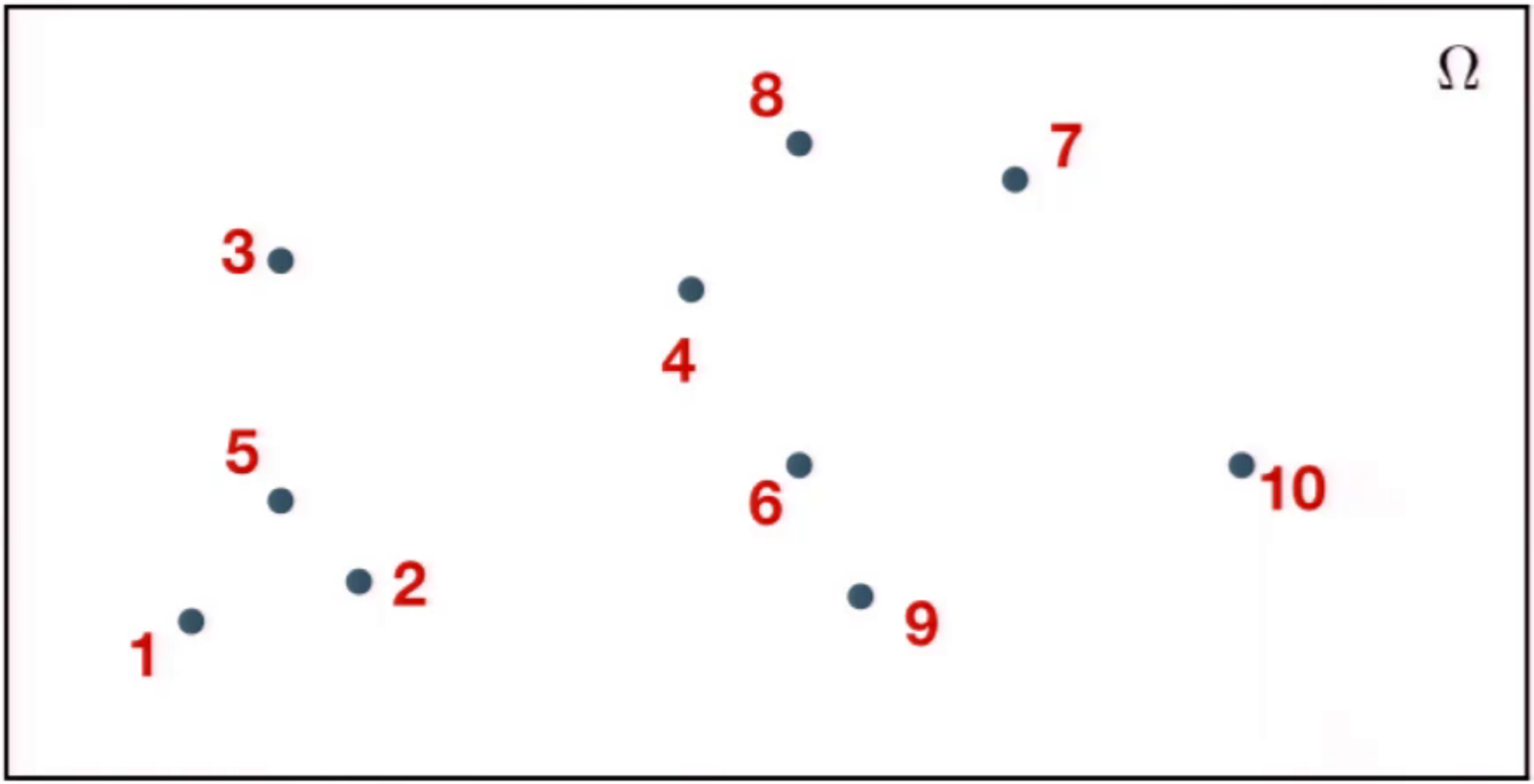
$\downarrow$

construct recurrence network G

there are many methods to do this.

Constructing recurrence networks, method 2

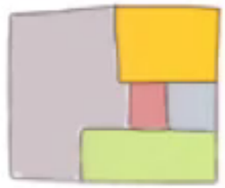
Example of a time series:

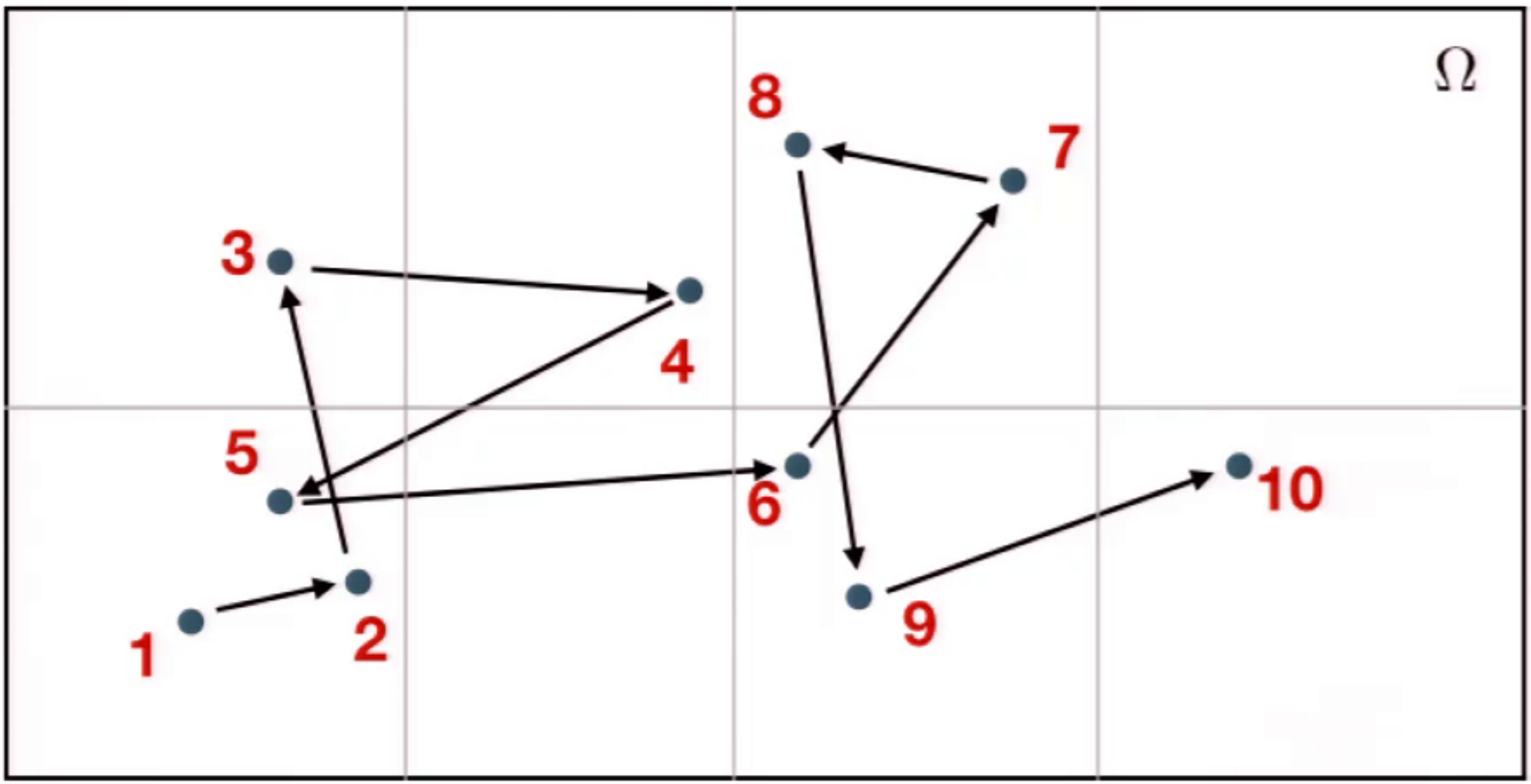


[S. Abe, N. Suzuki]

Constructing recurrence networks, method 2

ii) partition Ω .

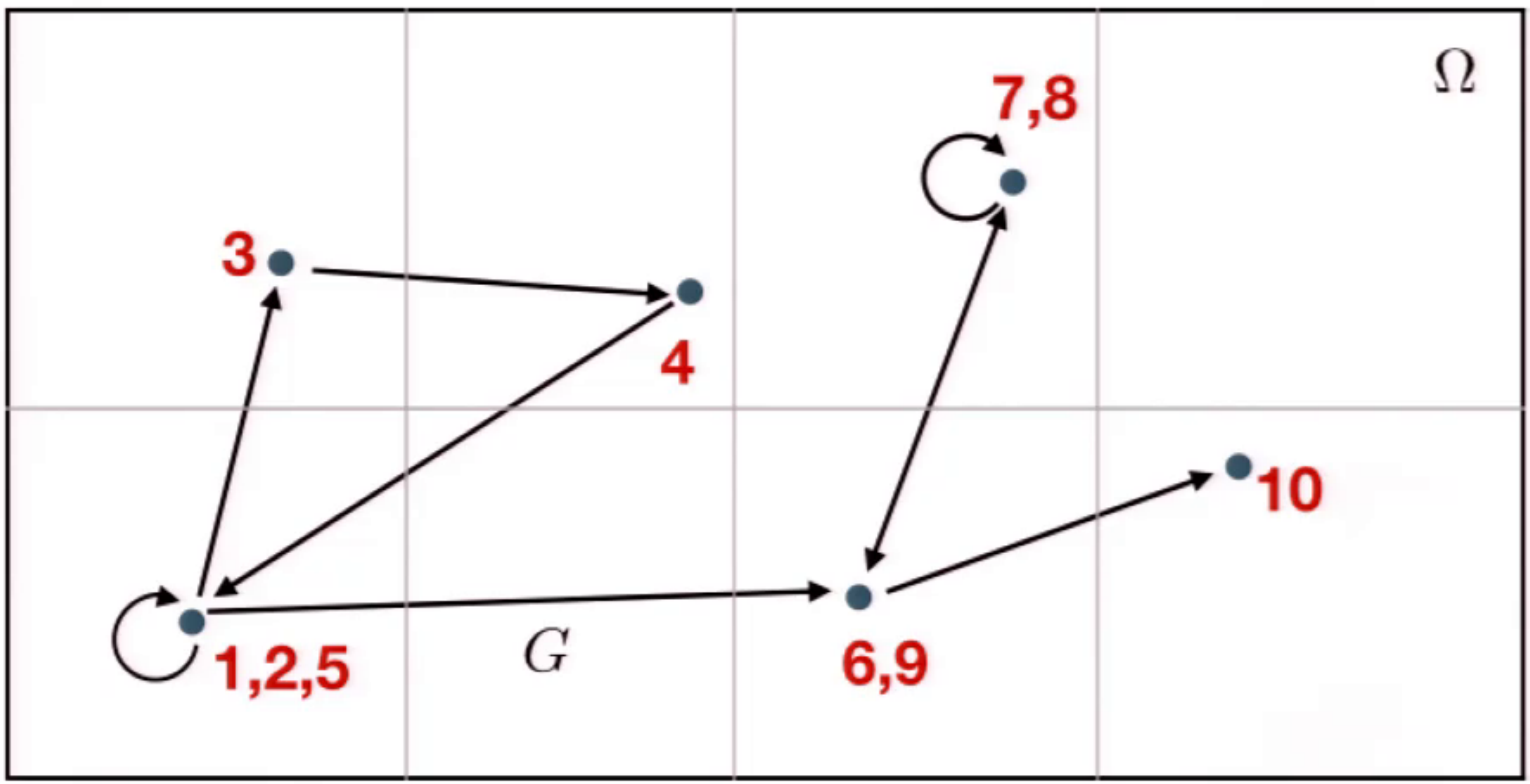




[S. Abe, N. Suzuki]

Constructing recurrence networks, method 2

iii) identify events in the same block.



[S. Abe, N. Suzuki]

Comparison

ε

	metric tresholding	**discretization**
type	G undirected	G directed
arrow of time	not represented	represented by edge directions
parameters	treshold ε	grid spacing h
effort	$\mathcal{O}\left(N^2\right)$	$\mathcal{O}\left(N\right)$

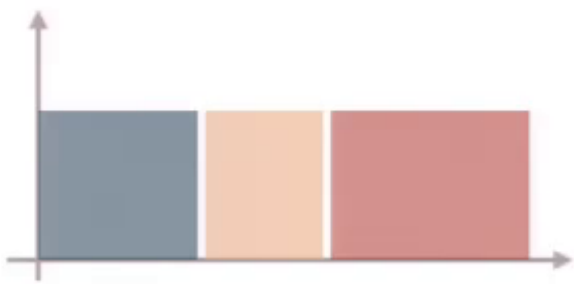
Outline

- Recurrence networks
- **Module detection**
- Method: Counting cycles
- Examples

Clustering directed networks

Clustering = m functions $q_i : V \to [0,1], \quad \sum_{i=1}^{m} q_i(x) = 1$

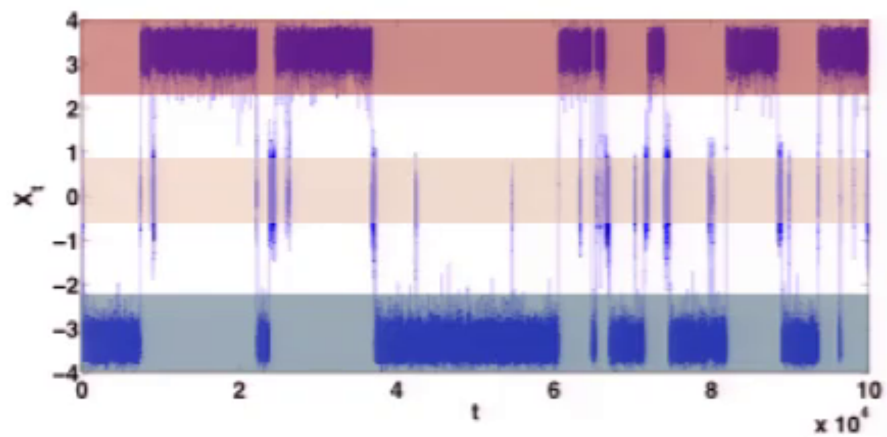
standard: full partition



$q_i : V \to \{0,1\}$

every node belongs to exactly one module

assumes perfect structure in the data



Clustering directed networks

Clustering = m functions $q_i : V \to [0,1], \quad \sum_{i=1}^{m} q_i(x) = 1$

full partition

$q_i : V \to \{0,1\}$

every node belongs to
exactly one module

assumes perfect structure
in the data

fuzzy partition

$q_i : V \to [0,1]$

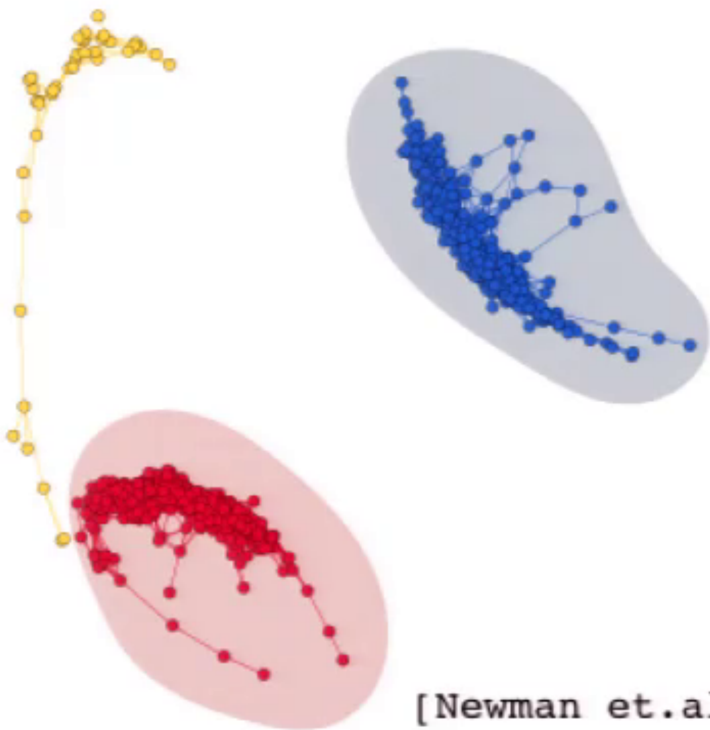
some nodes belong
to several modules

reflects imperfect structure
in the data

Clustering directed networks

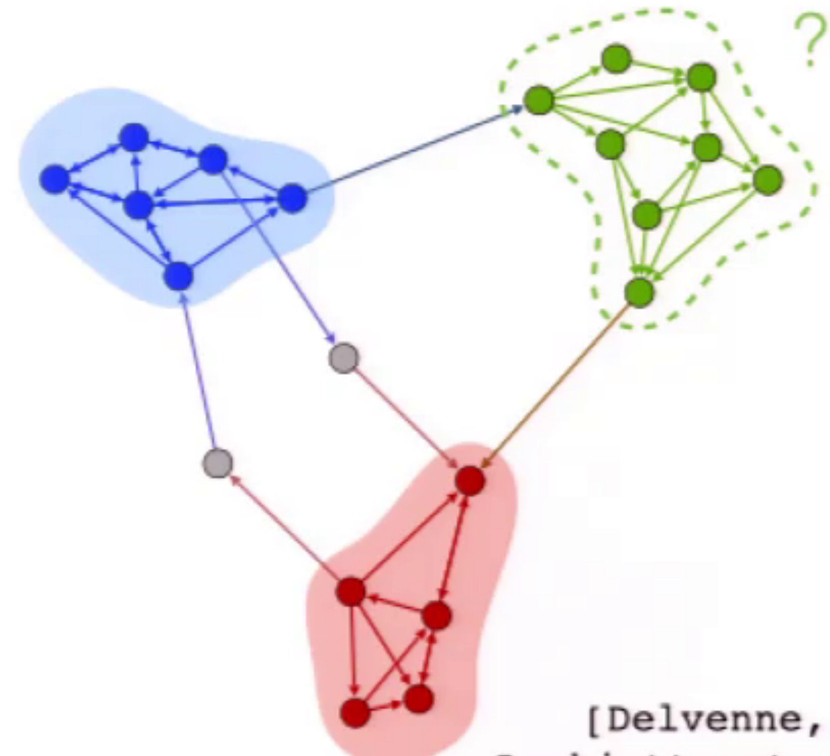
Clustering = m functions $q_i : V \to [0,1], \quad \sum_{i=1}^{m} q_i(x) = 1$

density based



[Newman et.al.]

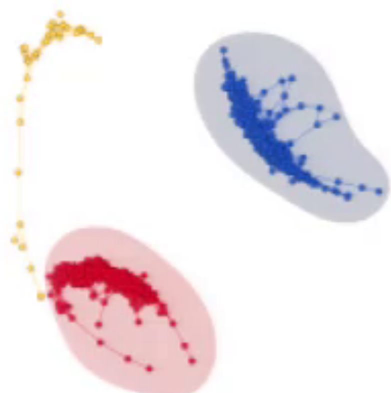
pattern based



[Delvenne, Lambiotte et.al.]

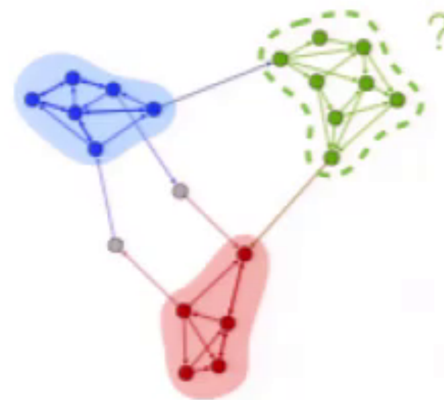
Clustering directed networks

density based

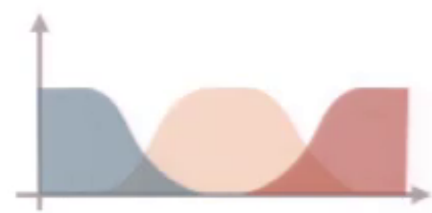


full partition

pattern based



fuzzy partition



Outline

- Recurrence networks
- Module detection
- **Method: Counting cycles**
- Examples

Model inference

fuzzy partition should reflect uncertainty -> to quantify uncertainty, we need to estimate a model of how the data was generated

Model inference

fuzzy partition should reflect uncertainty -> to quantify uncertainty, we need to estimate a model of how the data was generated

time series

$$\mathbf{x}_{[0,T]} = \{x_0, \ldots, x_T\}$$

symbol series

$$\mathbf{s}_{[0,T]} = \{s_0, \ldots, s_T\}$$

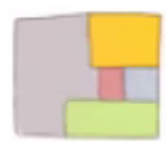
[Robinson et.al.]

Model inference

fuzzy partition should reflect uncertainty -> to quantify uncertainty, we need to estimate a model of how the data was generated.

time series

$$\mathbf{x}_{[0,T]} = \{x_0, \ldots, x_T\}$$



symbol series

$$\mathbf{s}_{[0,T]} = \{s_0, \ldots, s_T\}$$

assumption:

$\{s_0, \ldots, s_T\}$ was generated by a Markov chain

ML estimator of P:

$$P_{ij} = \frac{N_{ij}^T}{N_i^T}$$

$N_{ij}^T = N_{ji}^T \Leftrightarrow P$ reversible

Model inference

fuzzy partition should reflect uncertainty -> to quantify uncertainty, we need to estimate a model of how the data was generated.

assumption:

$\{s_0, \ldots, s_T\}$ was generated by a Markov chain

ML estimator of P:

$$P_{ij} = \frac{N_{ij}^T}{N_i^T}$$

$N_{ij}^T = N_{ji}^T \Leftrightarrow$ P reversible

recurrence network $G = \mathrm{graph}(P)$

$$V = \text{states of } P$$

$$E = \{(i,j) \in V \times V : P_{ij} > 0\}$$

P reversible $\Leftrightarrow$ G undirected

Reversibility

- P reversible -> we can cluster based on the metastable sets of P.

$$C \subset V \ \text{α-metastable} \Leftrightarrow \mathbf{P}\left(s_{t+\alpha} \in C | s_t \in C\right) \approx 1$$

- clustering pipeline:

select α → identify α-metastable sets $\{C_1, \ldots, C_m\}$ → compute m committor functions $q_i(v) = \mathbf{P}\left(s_\tau \in C_i | s_0 = v\right)$

module cores

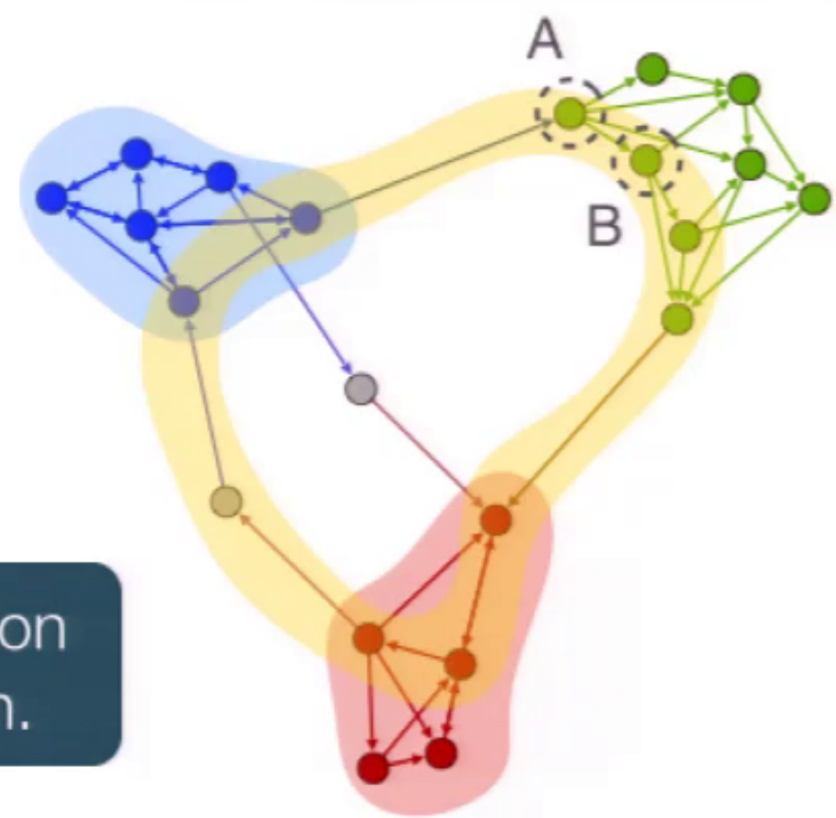
fuzzy affiliation functions

[Sarich, Djurdjevac, Schütte 2011]

Reversibility

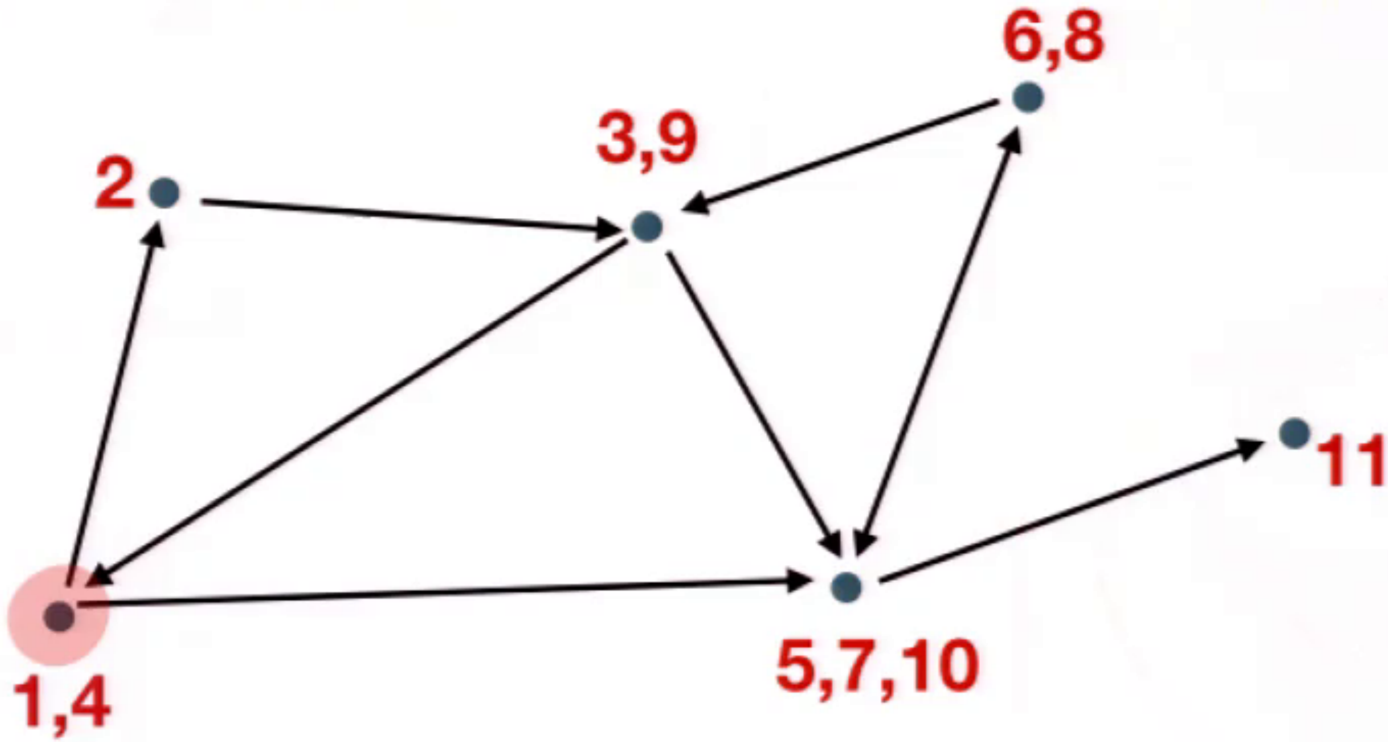
select α → identify α-metastable sets $\{C_1, \ldots, C_m\}$ → compute m committor functions $q_i(v) = \mathbf{P}\left(s_\tau \in C_i | s_0 = v\right)$

this **only** works if P is reversible.



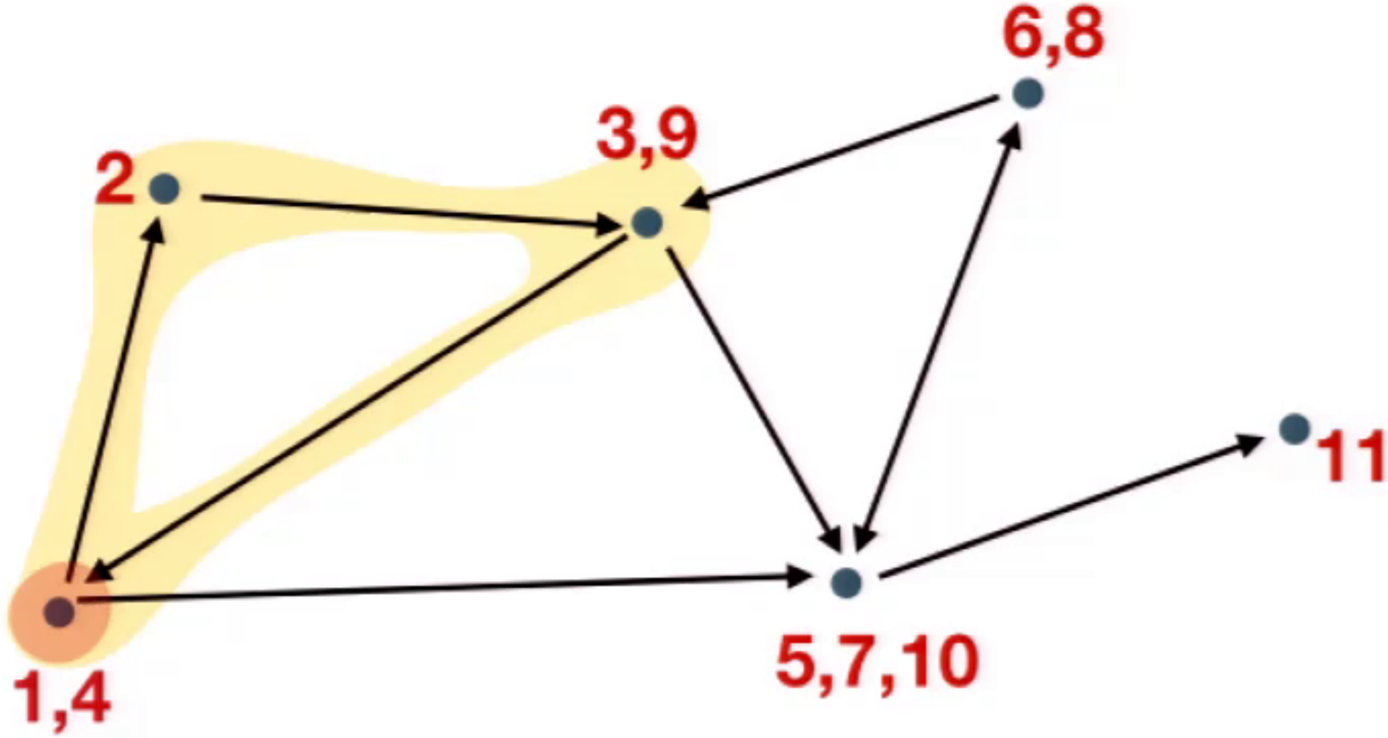
Goal: construct reversible approximation of P that keeps directional information.

Counting cycles



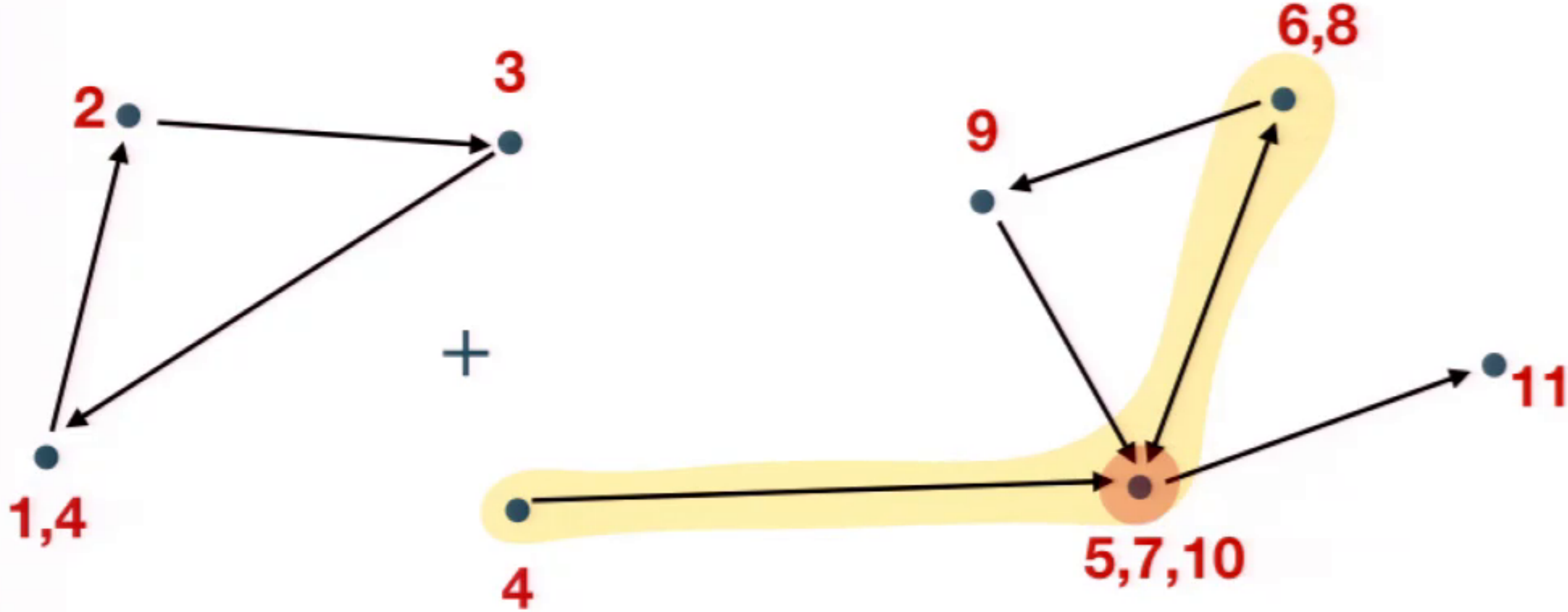
[Qian, 2004]

Counting cycles



[Qian, 2004]

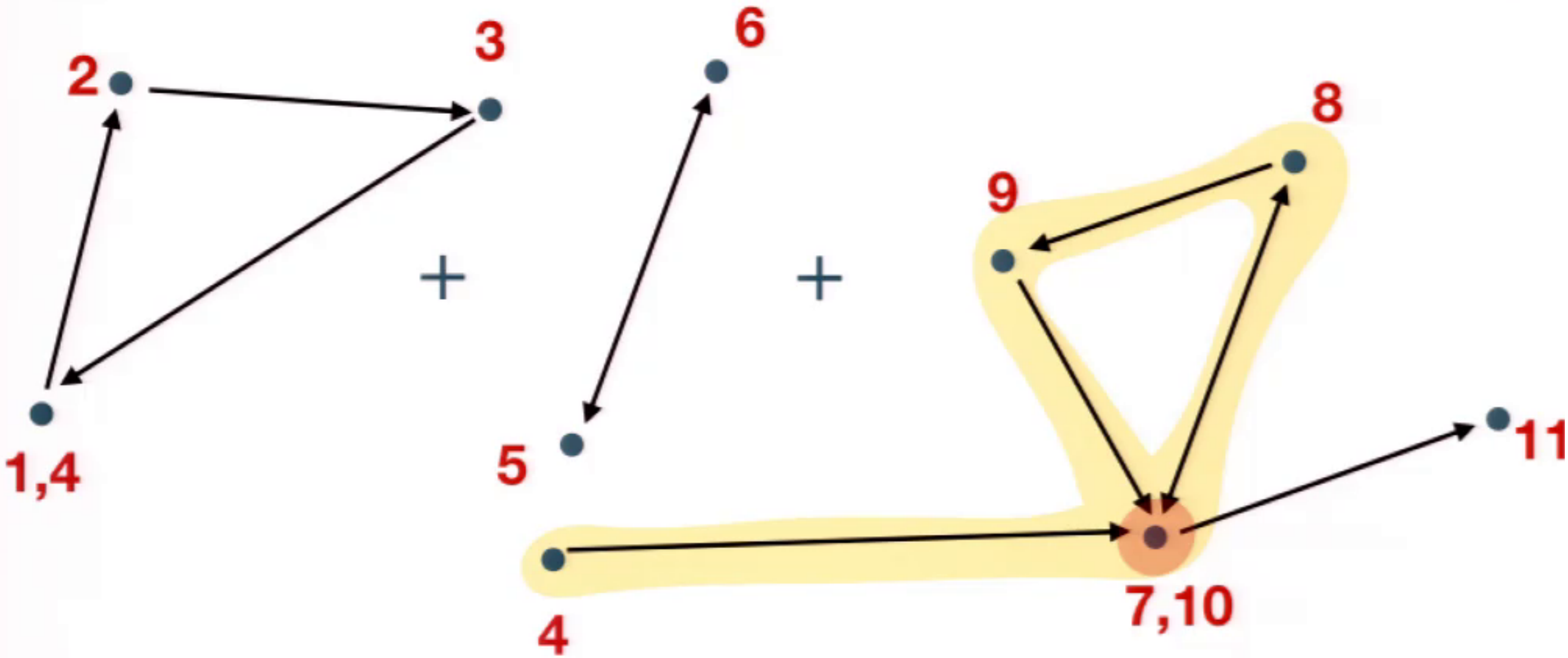
Counting cycles



$$\gamma_1 = (1,2,3)$$
$$N^T_{\gamma_1} \to N^T_{\gamma_1} + 1$$

[Qian, 2004]

Counting cycles

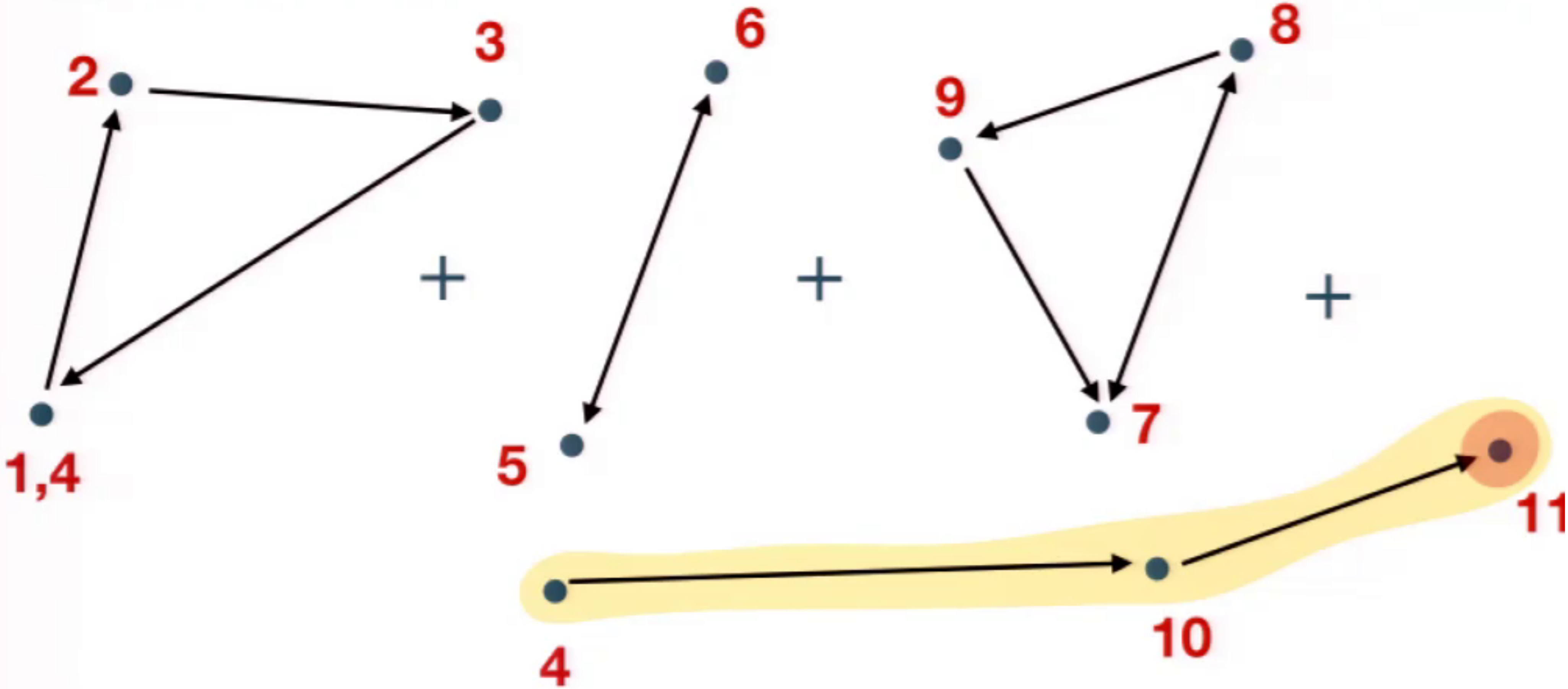


$\gamma_1 = (1, 2, 3)$
$N^T_{\gamma_1} \to N^T_{\gamma_1} + 1$

$\gamma_2 = (5, 6)$
$N^T_{\gamma_2} \to N^T_{\gamma_2} + 1$

[Qian, 2004]

Counting cycles



$\gamma_1 = (1,2,3)$
$N^T_{\gamma_1} \to N^T_{\gamma_1} + 1$

$\gamma_2 = (5,6)$
$N^T_{\gamma_2} \to N^T_{\gamma_2} + 1$

$\gamma_3 = (7,8,9)$
$N^T_{\gamma_3} \to N^T_{\gamma_3} + 1$

[Qian, 2004]

Main result

$$\omega(\gamma) := \lim_{T\to\infty} \frac{N_\gamma^T}{T}$$ converges a.s.

$$\mathcal{P}_{ij} := \lim_{T\to\infty} \frac{1}{N_i^T} \sum_\gamma \frac{N_\gamma^T}{|\gamma|} J_\gamma(i) J_\gamma(j)$$ reversible transition matrix

$$J_\gamma(i) = \begin{cases} 1 & i \in \gamma, \\ 0 & \text{else.} \end{cases}$$ membership function

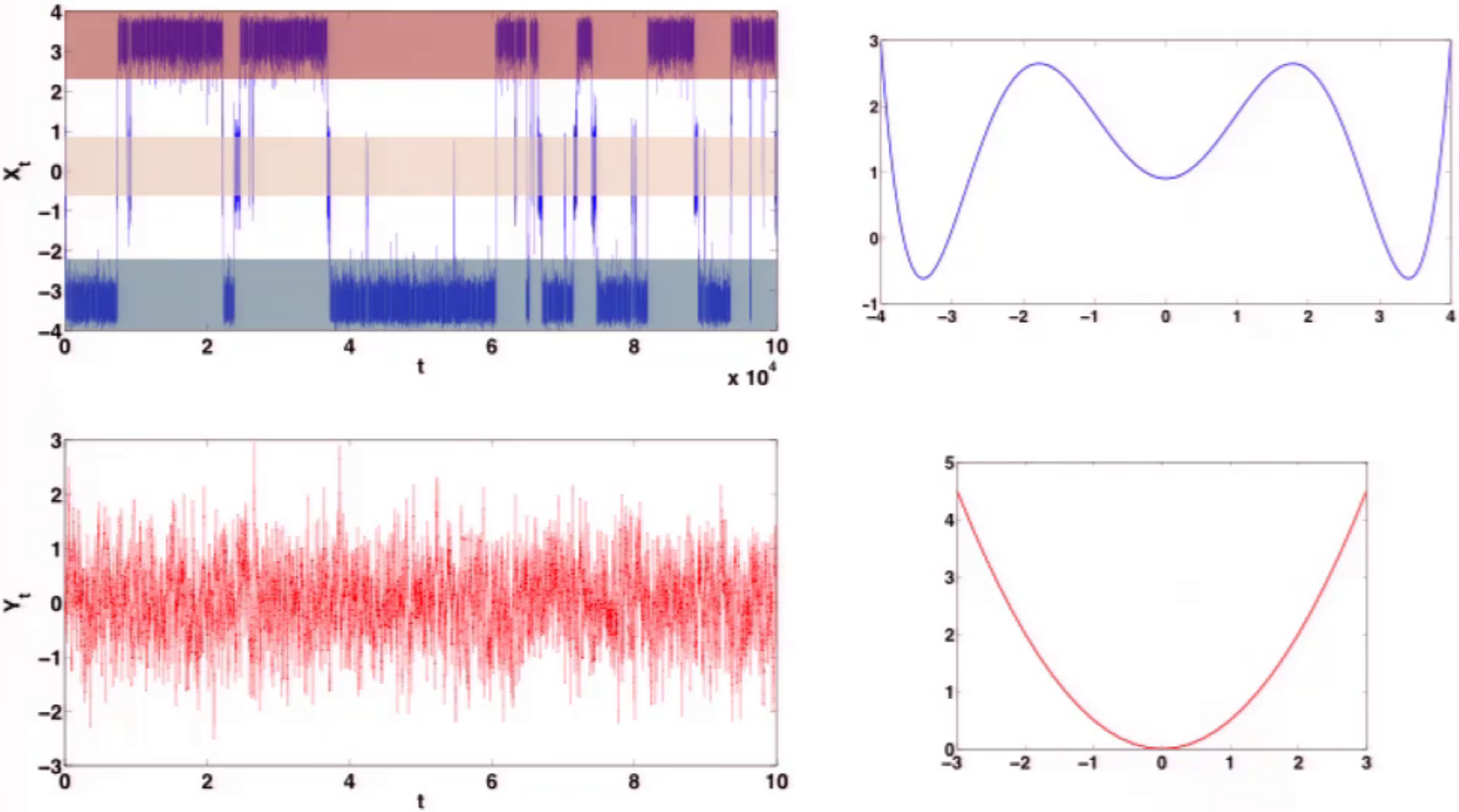
- $\mathcal{P}$ and P have the same invariant distribution.
- computing $\mathcal{P}$ is $\mathcal{O}(N)$ (and thus as expensive as computing P).

[RB and N. Conrad, 2014]

Outline

- Recurrence networks
- Module detection
- Method: Counting cycles
- **Examples**

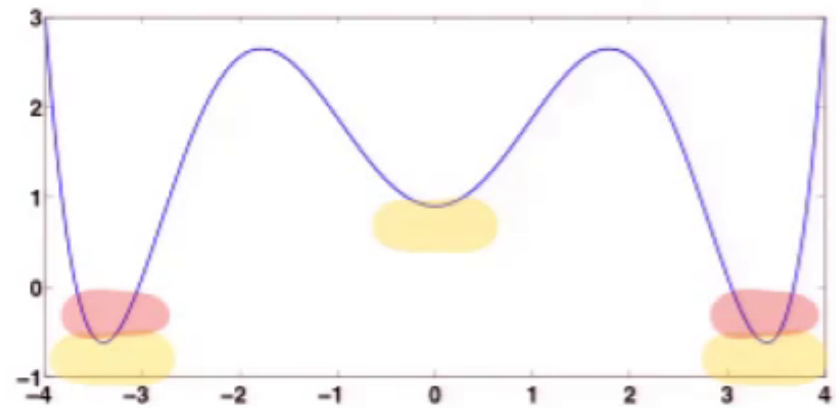
Example: Brownian motion in potential

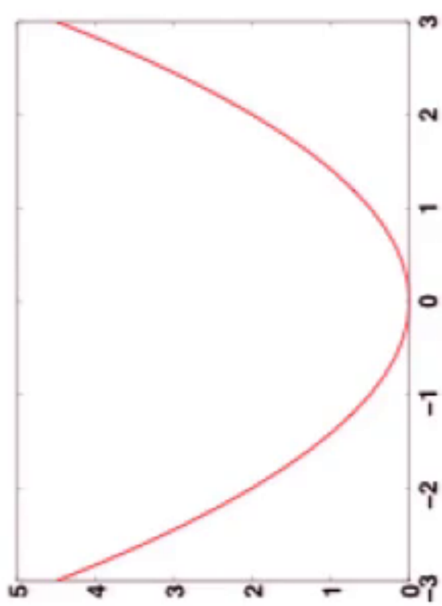


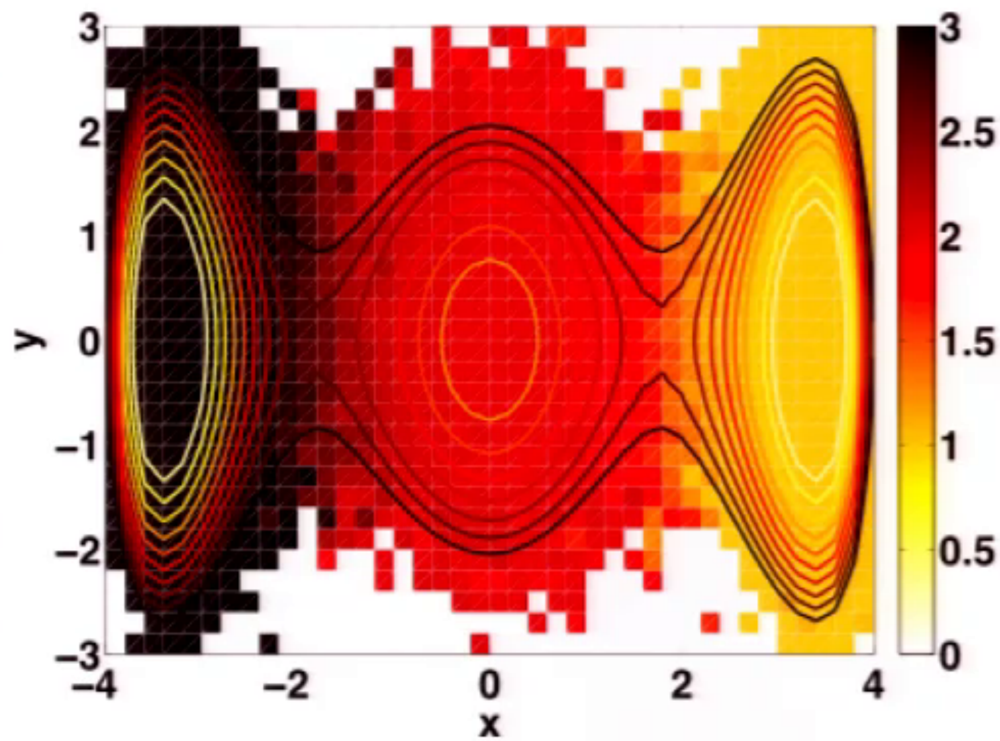
Example: Brownian motion in potential

$\alpha \leq 40$: 3 modules

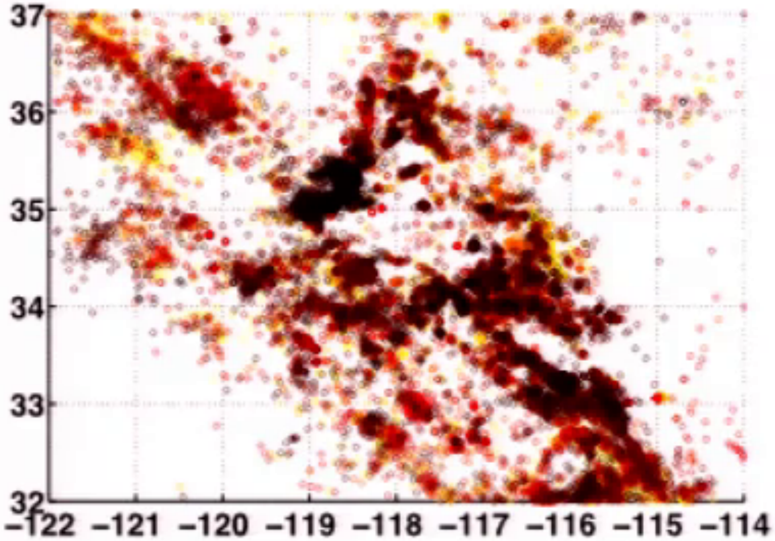
$40 < \alpha < 700$: 2 modules

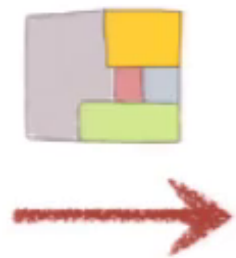


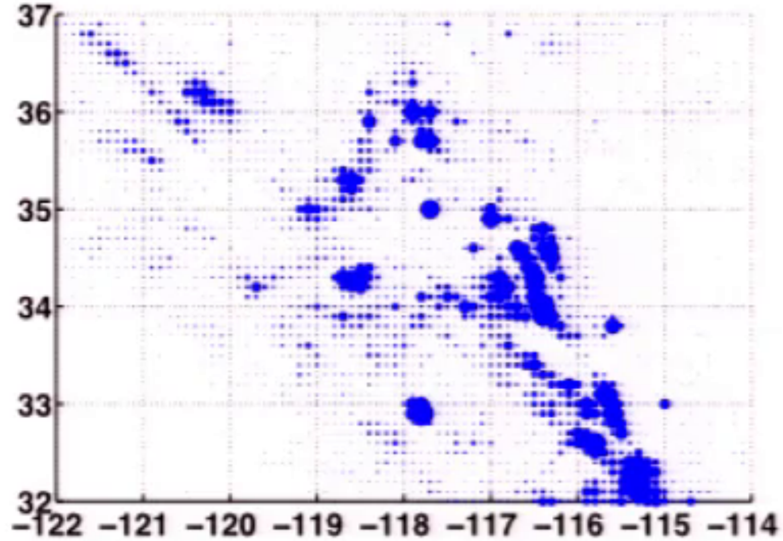




Example 2: Earthquakes in Southern California







earthquakes between 1952 and 2012 magnitude ≥ 2.5

Ω = latitude - longitude coords

uniform grid $h = 0.1^\circ$