Module Detection in Directed Recurrence Networks

Ralf Banisch

Joint work with Natasa Conrad
Motivation: What are time series?

Brownian particle in a potential landscape

earthquake events in Southern California
Motivation: Recurrence Networks

time series  \( x_{[0,T]} = \{x_0, \ldots, x_T\} \)

choose embedding in a metric space  \((\Omega, d)\)

construct recurrence network  \( G \)

Idea: infer structure in the data from structure in  \( G \)
Motivation: Recurrence networks

structured data

structure in $G$

modules
Reminder: From time series to networks

\[ x_{[0,T]} = \{x_0, \ldots, x_T\} \]

choose embedding in a metric space \((\Omega, d)\) \[\text{[e.g. Takens]}\]

construct recurrence network \(G\)

there are many methods to do this.
Constructing recurrence networks, method 2

Example of a time series:
Constructing recurrence networks, method 2

ii) partition $\Omega$.
Constructing recurrence networks, method 2

iii) identify events in the same block.
Comparison

<table>
<thead>
<tr>
<th>Type</th>
<th>Metric Thresholding</th>
<th>Discretization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrow of Time</td>
<td>$G$ undirected</td>
<td>$G$ directed</td>
</tr>
<tr>
<td>Parameters</td>
<td>not represented</td>
<td>represented by edge directions</td>
</tr>
<tr>
<td>Effort</td>
<td>$O\left(N^2\right)$</td>
<td>$O\left(N\right)$</td>
</tr>
</tbody>
</table>
Outline

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

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

**standard: full partition**

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

every node belongs to exactly one module

assumes perfect structure in the data
Clustering directed networks

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

**full partition**

\[ q_i : V \rightarrow \{0, 1\} \]

every node belongs to exactly one module

assumes perfect structure in the data

**fuzzy partition**

\[ q_i : V \rightarrow [0, 1] \]

some nodes belong to several modules

reflects imperfect structure in the data
Clustering directed networks

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

**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

\[
\begin{align*}
\text{time series} & \quad \text{symbol series} \\
\mathbf{x}_{[0,T]} & = \{x_0, \ldots, x_T\} & \mathbf{s}_{[0,T]} & = \{s_0, \ldots, s_T\}
\end{align*}
\]
Model inference

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

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

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

**assumption:**

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

**ML estimator of** \( P \):

\[ P_{ij} = \frac{N_{ij}^T}{N_i^T} \]

\[ N_{ij}^T = N_{ji}^T \Leftrightarrow P \text{ reversible} \]
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** \( P \):
\[ P_{ij} = \frac{N_{ij}^T}{N_i^T} \]
\[ N_{ij}^T = N_{ji}^T \iff P \text{ reversible} \]

**recurrence network** \( G = \text{graph}(P) \)
- \( V = \text{states of } P \)
- \( E = \{(i,j) \in V \times V : P_{ij} > 0\} \)

**\( P \) reversible \iff \( G \) undirected**
Reversibility

- $P$ reversible $\Rightarrow$ we can cluster based on the metastable sets of $P$.

\[ C \subset V \text{ $\alpha$-metastable } \iff P(s_{t+\alpha} \in C | s_t \in C) \approx 1 \]

- Clustering pipeline:

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

Sarich, Djurdjevac, Schütte 2011
Reversibility

1. select $\alpha$
2. identify $\alpha$-metastable sets $\{C_1, \ldots, C_m\}$
3. compute $m$ committor functions $q_i(v) = P(s_\tau \in C_i | s_0 = v)$

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

\[ \gamma_1 = (1, 2, 3) \]

\[ N_{\gamma_1}^T \rightarrow N_{\gamma_1}^T + 1 \]

[Qian, 2004]
Counting cycles

\[ \gamma_1 = (1, 2, 3) \]
\[ N^T_{\gamma_1} \rightarrow N^T_{\gamma_1} + 1 \]

\[ \gamma_2 = (5, 6) \]
\[ N^T_{\gamma_2} \rightarrow N^T_{\gamma_2} + 1 \]

[Qian, 2004]
Counting cycles

\[ \gamma_1 = (1, 2, 3) \]
\[ N_{\gamma_1}^T \rightarrow N_{\gamma_1}^T + 1 \]

\[ \gamma_2 = (5, 6) \]
\[ N_{\gamma_2}^T \rightarrow N_{\gamma_2}^T + 1 \]

\[ \gamma_3 = (7, 8, 9) \]
\[ N_{\gamma_3}^T \rightarrow N_{\gamma_3}^T + 1 \]

[Qian, 2004]
Main result

\[ \omega(\gamma) := \lim_{T \to \infty} \frac{N_T}{T} \text{ converges a.s.} \]

\[ P_{ij} := \lim_{T \to \infty} \frac{1}{N_i} \sum_{\gamma} \frac{N_T}{|\gamma|} J_\gamma(i) J_\gamma(j) \]

reversible transition matrix

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

membership function

- \( P \) and \( P \) have the same invariant distribution.

- computing \( P \) is \( O(N) \) (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

\[ \alpha \leq 40: \quad 3 \text{ modules} \]
\[ 40 < \alpha < 700: \quad 2 \text{ modules} \]
Example 2: Earthquakes in Southern California

Earthquakes between 1952 and 2012, magnitude ≥ 2.5

\[ \Omega = \text{latitude} - \text{longitude coords} \]

Uniform grid \( h = 0.1^\circ \)