Memory and Mesoscopic Structures in Diffusion Processes

Mauro Faccin
Jean-Charles Delvenne

UCL Université catholique de Louvain

ICTEAM, Université Catholique de Louvain, Belgique

CompleNet 2016 - Dijon

Abstract

Here we propose an info-theoretical measure of lumpability. By computing the amount of information that flows from the past of the aggregated Markov chain toward its future one can devise the lumpability of the network on the given block structure, a partition of the original network. The dynamical properties of the aggregated system provide a powerful mechanism to detect underlaying mesoscopic structures that influence the system kinetics. The topology of those structures include communities, core-periphery, bipartite, block stochastic structures.

Definitions:

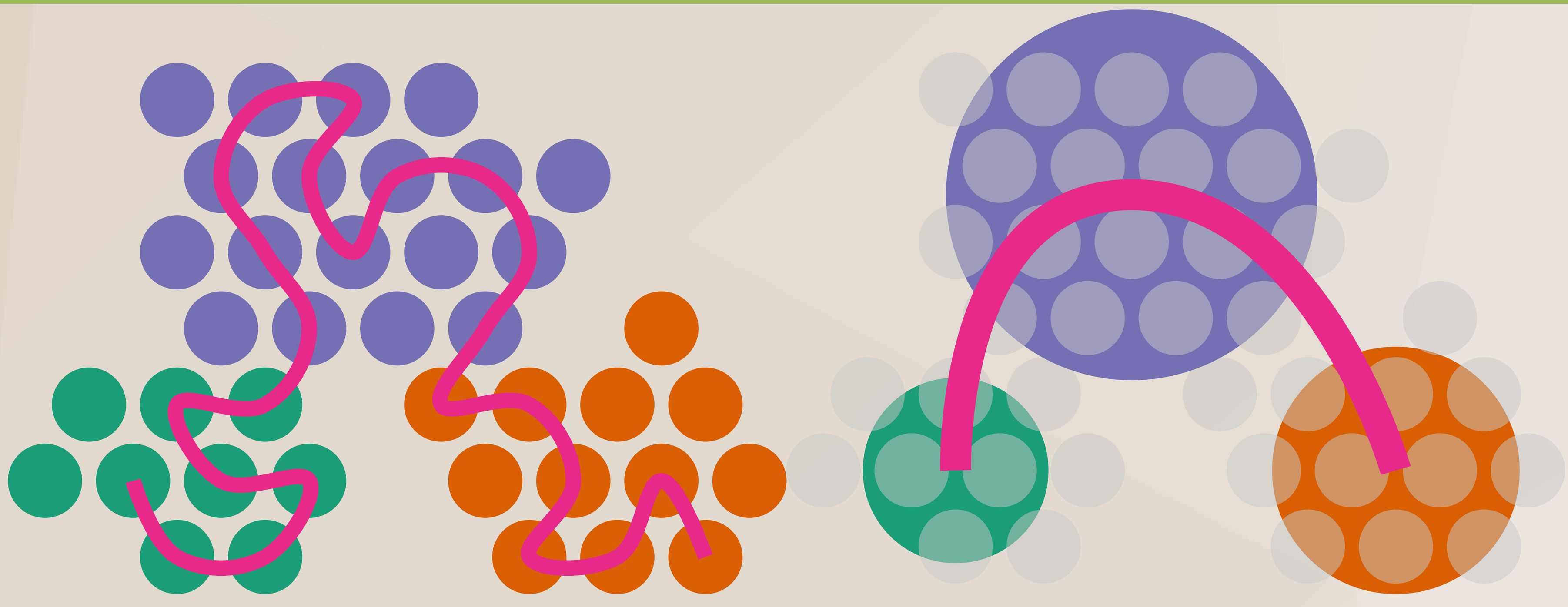
- a graph $G = V, E$
- a Markov Chain on top of G:

$$\mathcal{X} = \{X_0, X_1, \dots\}$$

- a non-overlapping partitioning of the system (such as a community structure)

The Problem:

The projection of the Markov Chain $\mathcal{X}$ on a partition of the system can generate on the aggregated dynamics unwanted effective memory. The latter prevents to map the system dynamics to a memoryless process on the aggregated topology.



Whole system　　　　Aggregated system

The Solution:

We must enforce the partition or community structure to reflect the dynamical properties of the original process. In this case the aggregated network represents a ***good model*** for the original system.

Given the Markov Chain $\mathcal{X}$, its projection to the aggregated topology is

$$\mathcal{Y} = \{Y_0, Y_1, \dots\}$$

The flow of information as means of mutual information

$$I_k = I(Y_{t+1}; Y_{t-k}, \dots | Y_t, \dots, Y_{t-k+1})$$

provides a proxy of the lumpability of the stochastic process on the aggregated network. Here I_k denotes the amount of information that flows from the process past toward its future and gives a measure of the memory order k of the aggregated dynamics.
The highest value of k with non-negligible I_k represents the Markovianity order of the aggregated dynamics.

Recipe:

Maximize: $I_0 = I(Y_{t+1}; Y_t, \dots)$ the information flow from the past to the immediate future (provides predictability of the process)

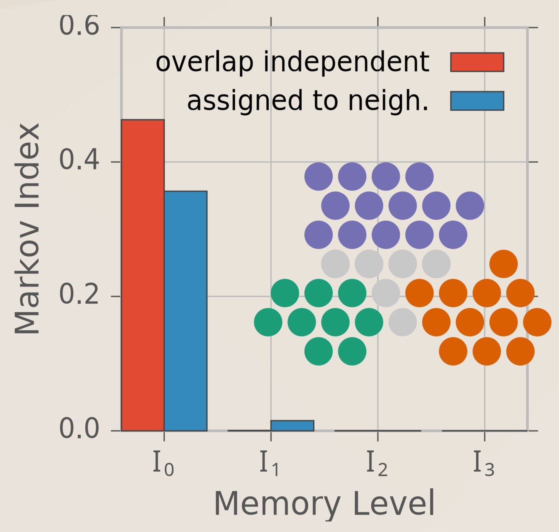
Minimize: $I_k, \forall k > 0$ the information flow from past toward future given the knowledge of the present (ensures Markovianity of the aggregated process)

Aknowledgement:

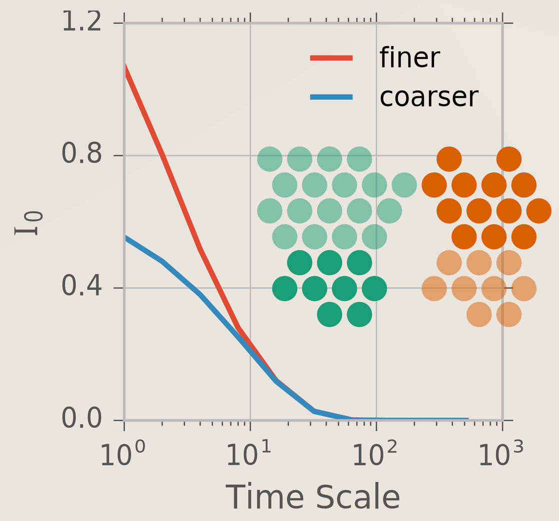
This work was supported by the Belgian Programme of Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office and an Action de Recherche Concertée (ARC) of the French Community of Belgium.

Examples:

Community structure in a scale free network. The community structure represents a good model with high I_0 and negligible higher order effects. Other partitioning attempts lead to lower information on flow or higher order memory effects.



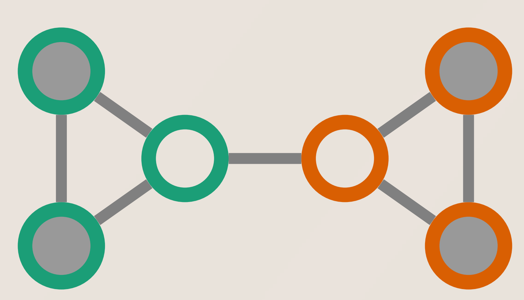
Overlapping Communities. The uncertainty on community attribution of the overlapping nodes is solved when those nodes are assigned to a new independent block acting as a bridge. Assigning those nodes to the closer block gives an aggregated dynamics with higher order memory, incompatible with the original process.



Hierarchical Communities. Finer block decomposition captures more information at finer time resolution (*finer* data). When temporal sampling is applied, on the other hand, both the *finer* and the *coarser* models provide comparable insight into the system dynamics. Higher order memory is negligible in both cases.

Equitable Partitions:

The following small graph represents a challenge: community-wise the green/orange partition is preferred for its high modularity; on the other hand the white/gray division is an equitable partition with perfect Markovian dynamics.

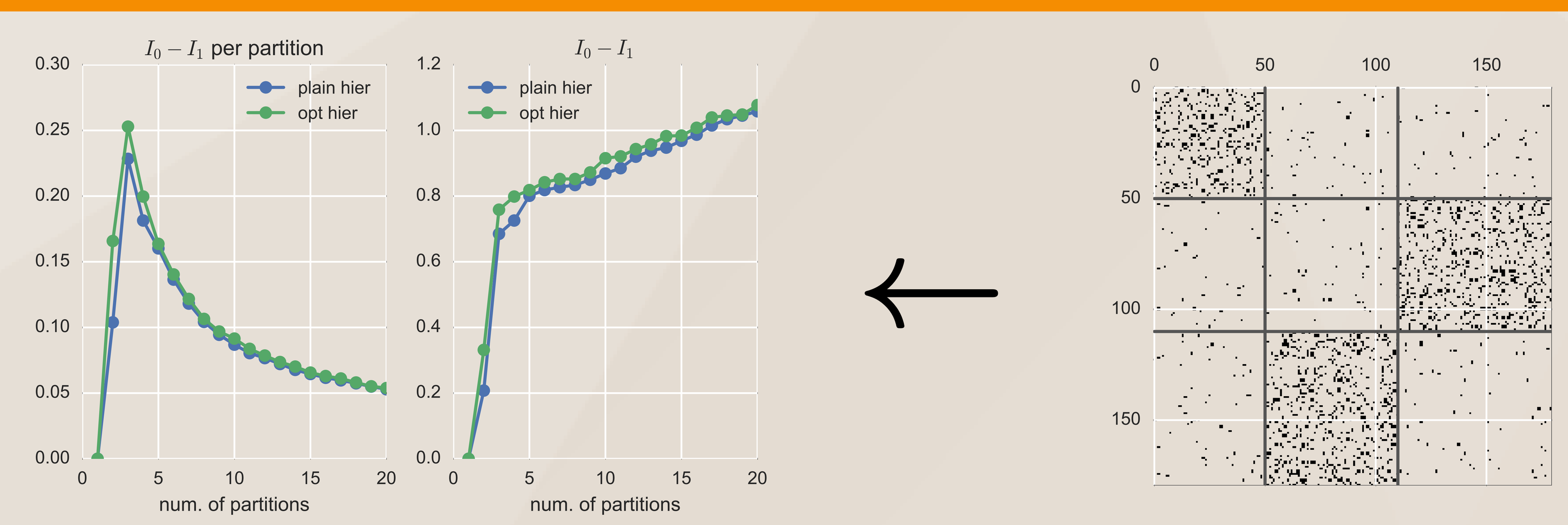


white/gray perfect equitable partition, perfect Markovian aggregated dynamics, recovered with cost function $f = I_1$

green/orange good partition with negligible memory effects, higher predictability of the dynamics, recovered with cost function $f = I_1 - I_0$

Partition Detection:

Given an undirected network with block stochastic structure (adj. mat. below), the underlaying blocks are recovered with a simple hierarchical algorithm and cost function $f = I_1 - I_0$.



The best partition corresponds to the given structure (with negligible number of mis-assigned nodes).