Dynamics and Hierarchical Structure in Networks

R. Lambiotte
Institute for Mathematical Sciences
Imperial College London

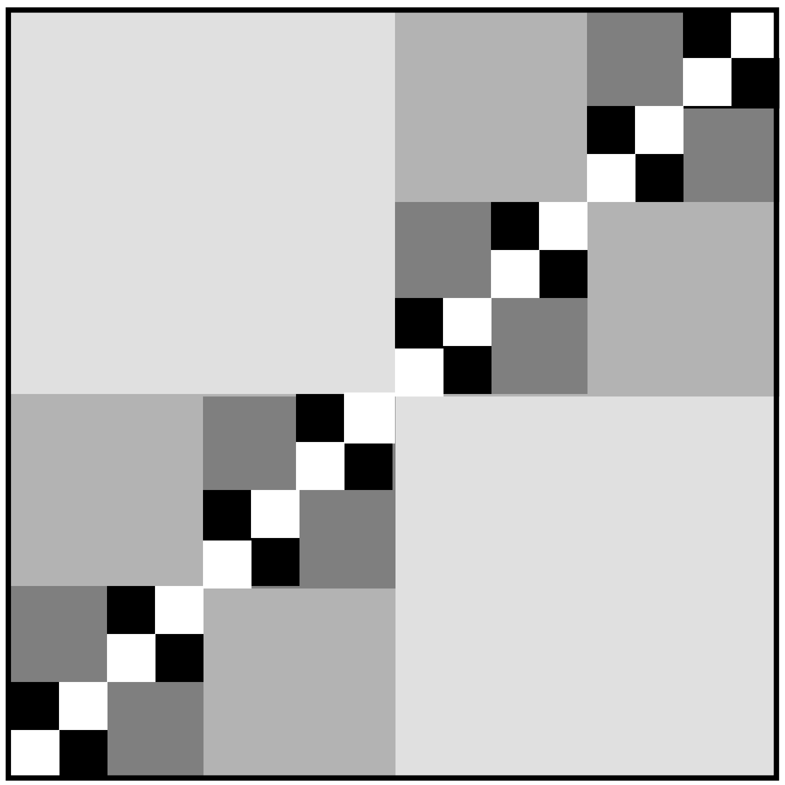
with M. Barahona and J.-C. Delvenne

Imperial College London
100 years of living science
100

R. Lambiotte, J.-C. Delvenne and M. Barahona, *arXiv:0812:1770*

Modular Networks and dynamics

Many networks are "modular" and have a hierarchical structure:
modules within modules



How does such modularity affect dynamics?

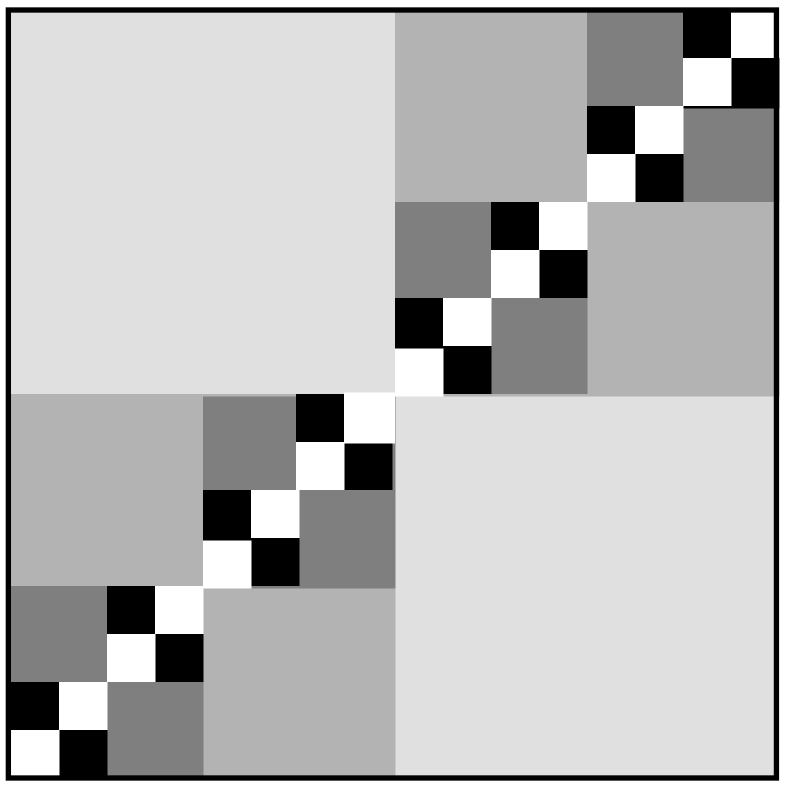
A. Arenas, A. Diaz-Guilera and C.J. Pérez-Vicente, *Phys. Rev. Lett.* **96**, 114102 (2006).
R. Lambiotte, M. Ausloos and J.A. Holyst, *Phys. Rev. E* **75**, 030101(R) (2007).

Is it possible to uncover those modules in large networks?

NG, GN, Walktrap, clique-percolation, Simulated Annealing, etc.

Modular Networks and dynamics

Many networks are “modular” and have a hierarchical structure:
modules within modules



How does such modularity affect dynamics?

A. Arenas, A. Diaz-Guilera and C.J. Pérez-Vicente, *Phys. Rev. Lett.* **96**, 114102 (2006).
R. Lambiotte, M. Ausloos and J.A. Holyst, *Phys. Rev. E* **75**, 030101(R) (2007).

Is it possible to use dynamics to characterize (and uncover?) the modular structure of a network?

e.g. Walktrap (RW exploration), Synchronization of oscillators, Talk of JC Delvenne (Wednesday)

Is it possible to uncover those modules in large networks?

NG, GN, Walktrap, clique-percolation, Simulated Annealing, etc.

Quality of the partition of a network

Let us focus on a weighted network and attribute to each node a module c_i

$$\boxed{Q = \frac{1}{2m} \quad {}_{i,j} \quad A_{ij} - P_{ij} \quad (c_i, c_j)}$$

A_{ij} adjacency matrix

$k_i = \quad {}_{j} \quad A_{ij}$ *degree* of i $\quad m = \frac{1}{2} \quad {}_{ij} \quad A_{ij}$ total weight

P_{ij} adjacency matrix of an equivalent null model

$$Q_b = \frac{1}{2m} \quad {}_{i,j} \quad A_{ij} - \frac{k^{2}}{2m} \quad (c_i, c_j)$$

$$Q_c = \frac{1}{2m} \quad {}_{i,j} \quad A_{ij} - \frac{k_i k_j}{2m} \quad (c_i, c_j)$$

$Q \quad [-1, 1]$

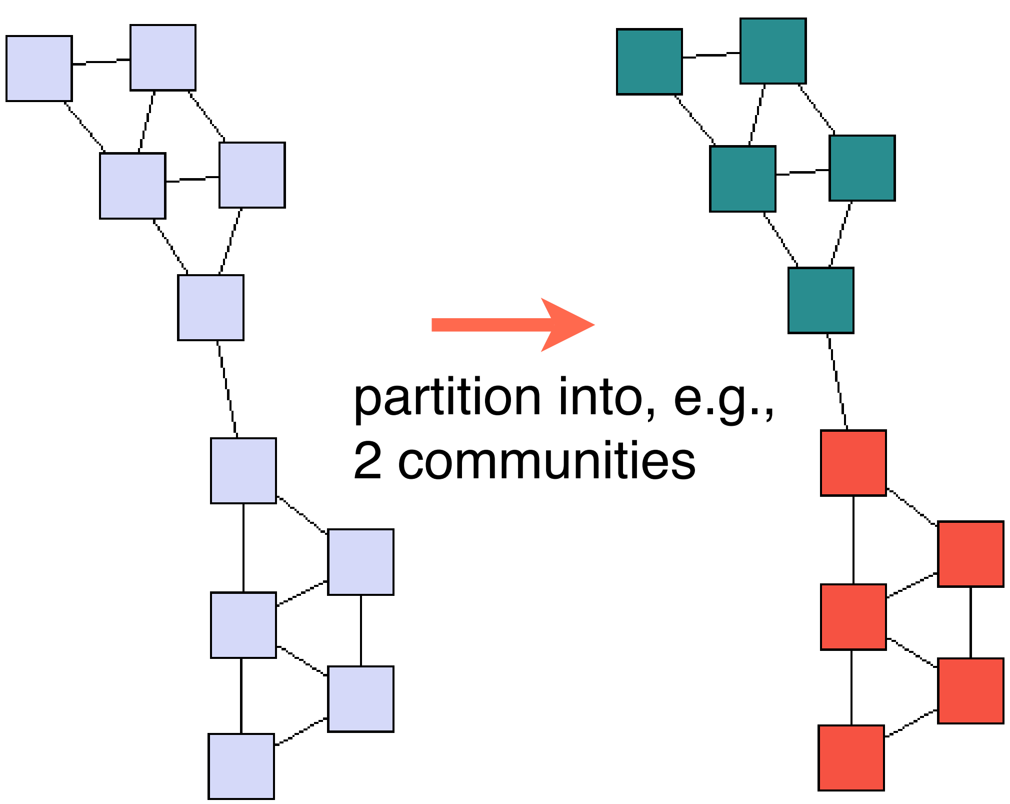
Modularity: dynamical foundations

Let us consider a random walk on a symmetric network: if the partition of a network is “good”, the walker should stay long times in a modules before leaving it

$$p_{i;n+1} = {}_{j} \frac{A_{ij}}{k_j} p_{j;n} \xrightarrow{\text{equilibrium}} p_i = k_i/2m$$

Proba to be in C at *n* $\quad {}_{j \;\; C} \frac{k_j}{2m}$

Proba to be in C at *n* and *n+1* $\quad {}_{i \;\; C \; j \;\; C} \frac{A_{ij}}{k_j} \frac{k_j}{2m}$



$$Q_c = \frac{1}{2m} {}_{C \;\; P \;\; i,j \;\; C} \; A_{ij} - \frac{k_i k_j}{2m}$$

A_{ij}: Probability to be in C at *n* and *n+1*

$\frac{k_i k_j}{2m}$: Same probability for independent walkers

Other processes: CTRW

Modularity is therefore based on the probability to be in the same module over two subsequent time steps, which suggests to generalize the definition over longer time intervals, but also to look at different dynamical processes.

$$\partial_t p_i = \sum_j \frac{A_{ij}}{k_j} p_j - p_i \xrightarrow[\text{equilibrium}]{} p_i = k_i / 2m$$

$$\boxed{R_t = \sum_{i,j} \left(e^{t(B-I)}{}_{ij} \frac{k_j}{2m} - \frac{k_i k_j}{(2m)^2} \right) (c_i, c_j)}$$ with $B_{ij} = A_{ij} / k_j$

What are the optimal partitions of R_t?

t=0 $$Q_0 = 1 - \sum_{i,j} \frac{k_i k_j}{(2m)^2} (c_i, c_j)$$

Communities=single nodes

Other processes: CTRW

Modularity is therefore based on the probability to be in the same module over two subsequent time steps, which suggests to generalize the definition over longer time intervals, but also to look at different dynamical processes.

$$\partial_t p_i = \sum_j \frac{A_{ij}}{k_j} p_j - p_i \xrightarrow{\text{equilibrium}} p_i = k_i / 2m$$

$$\boxed{R_t = \sum_{i,j} \left(e^{t(B-I)} \right)_{ij} \frac{k_j}{2m} - \frac{k_i k_j}{(2m)^2} \quad (c_i, c_j)} \quad \text{with } B_{ij} = A_{ij}/k_j$$

What are the optimal partitions of R_t?

t small $\quad R_t \approx (1 - t) R_0 + t Q_c \quad Q_t$

favours single nodes

modularity

!! Q_t equivalent to the Hamiltonian formulation of Reichardt and Bornholdt (t=1/gamma)

*J. Reichardt and S. Bornholdt, Phys. Rev. E **74**, 016110 (2006)*
Statistical mechanics of community detection

Other processes: CTRW

Modularity is therefore based on the probability to be in the same module over two subsequent time steps, which suggests to generalize the definition over longer time intervals, but also to look at different dynamical processes.

$$\partial_t p_i = \sum_j \frac{A_{ij}}{k_j} p_j - p_i \xrightarrow[\text{equilibrium}]{} p_i = k_i/2m$$

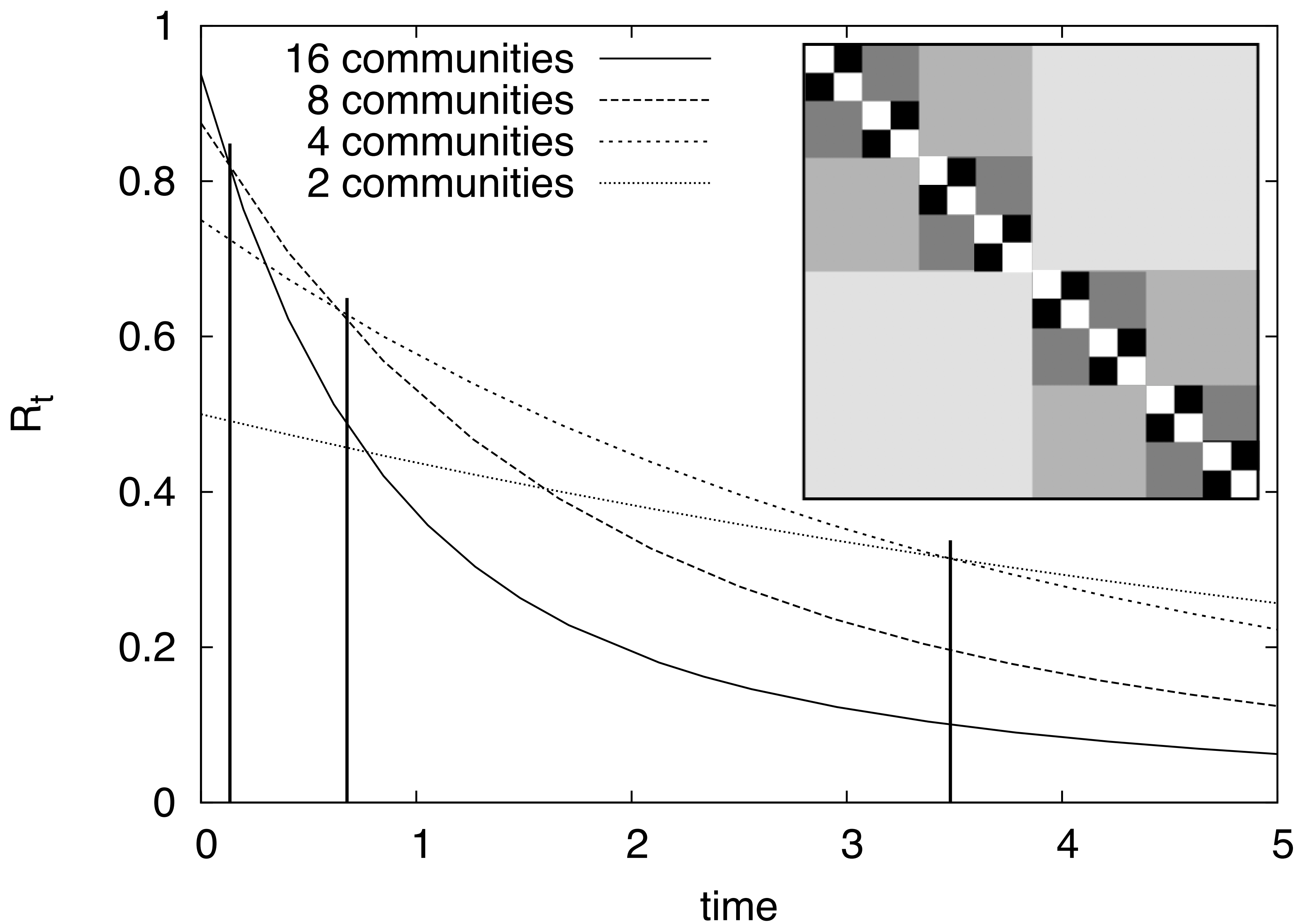
$$\boxed{R_t = \sum_{i,j} \left[e^{t(B-I)}_{ij} \frac{k_j}{2m} - \frac{k_i k_j}{(2m)^2} \right] (c_i, c_j)} \quad \text{with } B_{ij} = A_{ij}/k_j$$

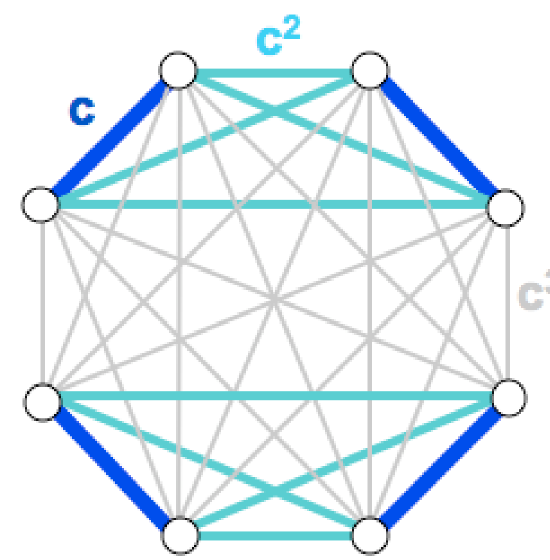
What are the optimal partitions of R_t?

When t goes to infinity, the optimal partition is made of 2 communities (by spectral decomposition)

Modularity: dynamical foundations

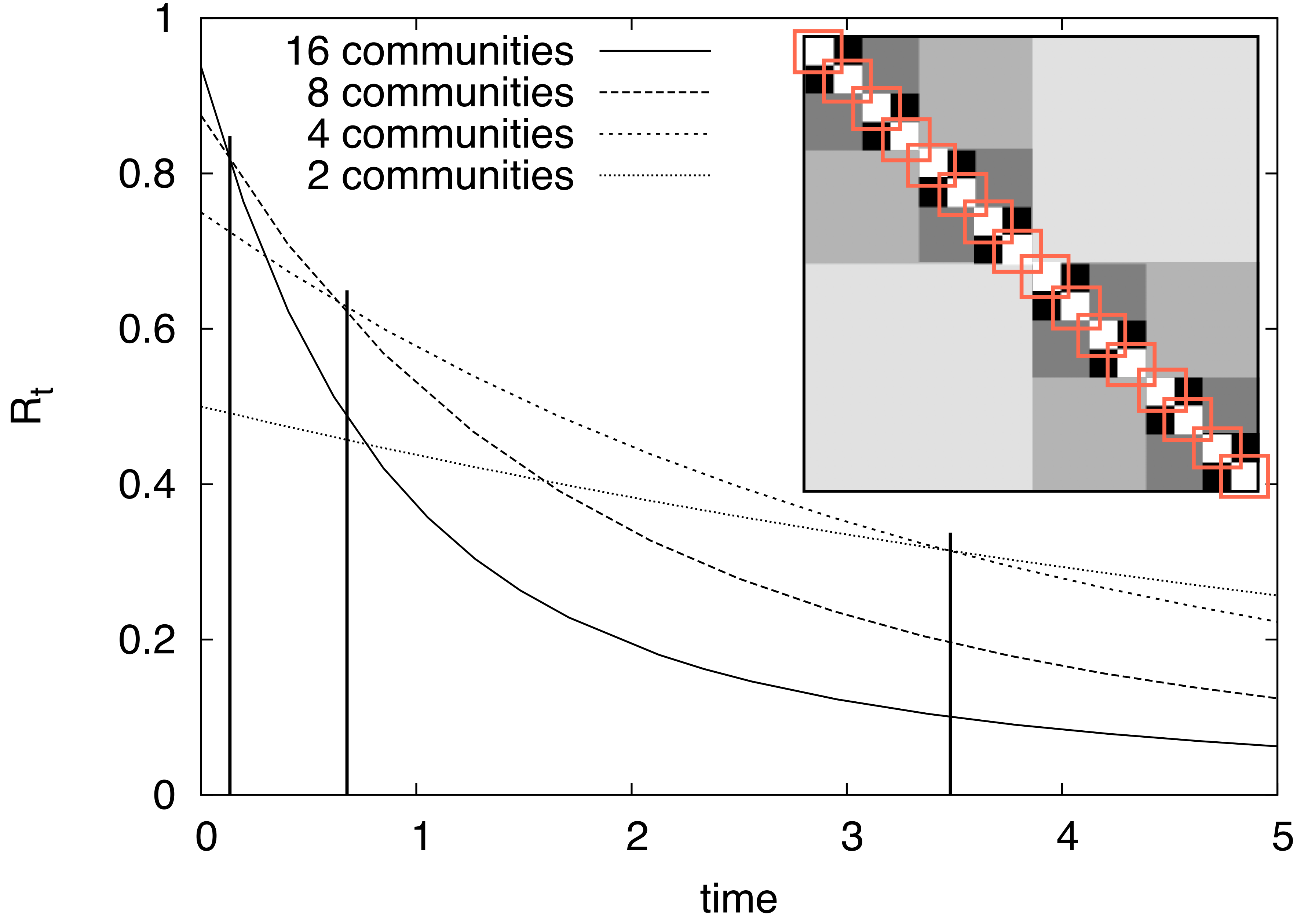
Time is a "resolution parameter": larger and larger communities when time is increased





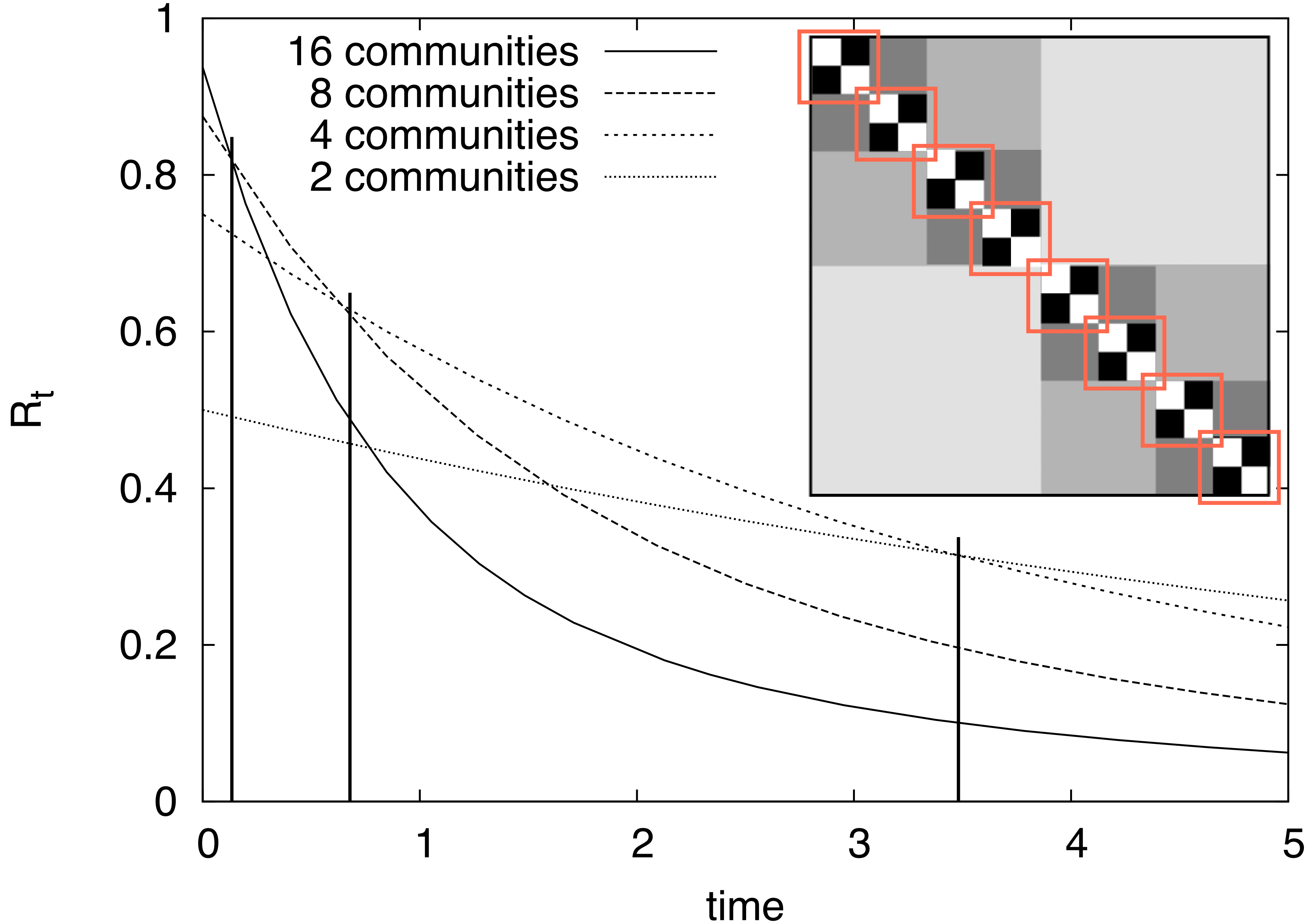
Modularity: dynamical foundations

Time is a “resolution parameter”: larger and larger communities when time is increased



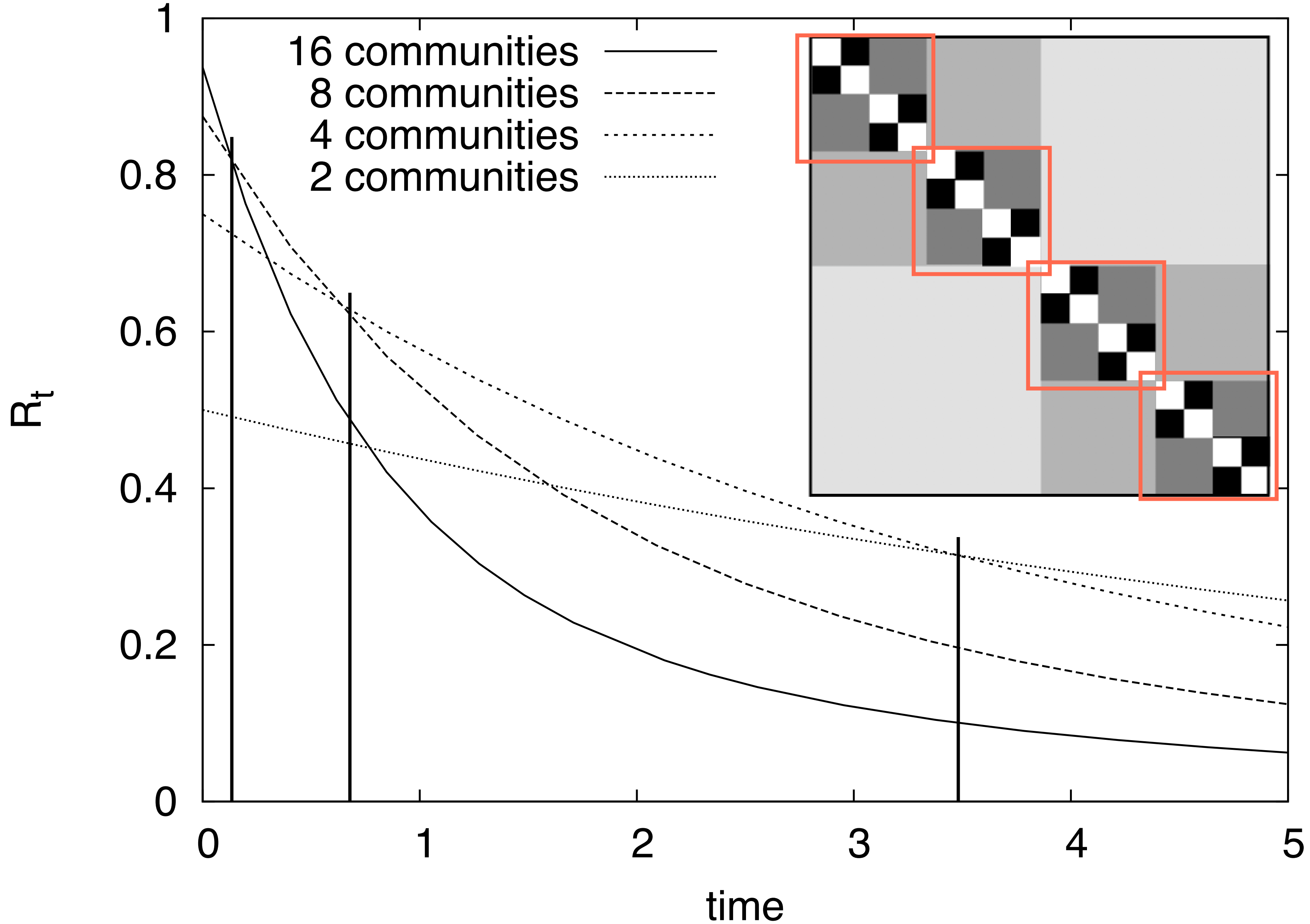
Modularity: dynamical foundations

Time is a "resolution parameter": larger and larger communities when time is increased



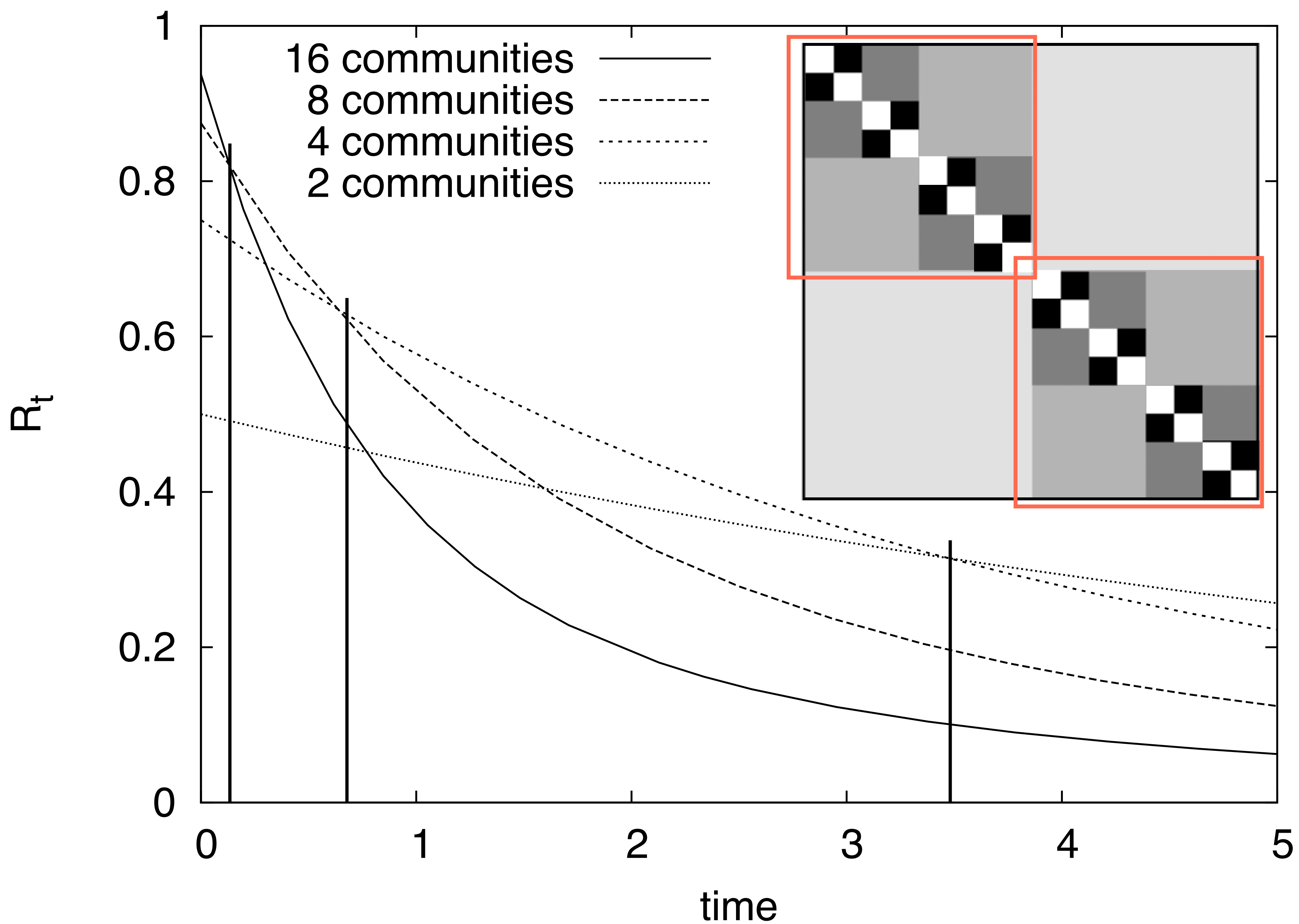
Modularity: dynamical foundations

Time is a “resolution parameter”: larger and larger communities when time is increased



Modularity: dynamical foundations

Time is a "resolution parameter": larger and larger communities when time is increased



Other processes: Kuramoto model

Modularity is therefore based on the probability to be in the same module over two subsequent time steps, which suggests to generalize the definition over longer time intervals, but also to look at different dynamical processes.

$$\partial_t \; {}_i = \; {}_i + \underset{j}{\;\;} A_{ij} \sin(\; {}_j - \; {}_i)$$

Approach toward synchronization when $\; {}_i =$

$$\; {}_i = \; {}_\infty + p_i$$

$$\partial_t p_i = \underset{j}{\;\;} \frac{A_{ij}}{k} p_j - \frac{k_i}{k} p_i$$

the probability to leave a node is proportional to the degree of the node

$$p_i = 1/N \qquad k \;/2m$$

which is different from:

$$\partial_t p_i = \underset{j}{\;\;} \frac{A_{ij}}{k_j} p_j - p_i$$

the probability to leave a node does not depend on the degree of the node

$$p_i = k_i/2m$$

Alex Arenas, Albert Diaz-Guilera, C Pérez-Vicente, Phys. Rev. Lett. 96, 114102 (2006).
Synchronization Reveals Topological Scales in Complex Networks

Other processes: Kuramoto model

By reapplying the previous steps, one finds the quality function:

$$\boxed{R_t = \quad {}_{i,j} \quad e^{t/\ k\ (A-K)} \quad {}_{ij} \ \frac{1}{N} - \frac{1}{N^2} \quad (c_i, c_j)}$$

When t is small:

$$R_t \approx (1 - t) R_0 + t Q_b \quad Q_t$$

This quantity is equivalent to the Reichardt & Bornholdt approach, but also to the Arenas approach consisting in optimising modularity but modifying the network with self-loops: $Q_b(A_{ij} + r I_{ij})$

$$t = \ k \ /(\ k \ + r)$$

!! Such an equivalence between the 3 approaches is only valid for R'_t and not R_t

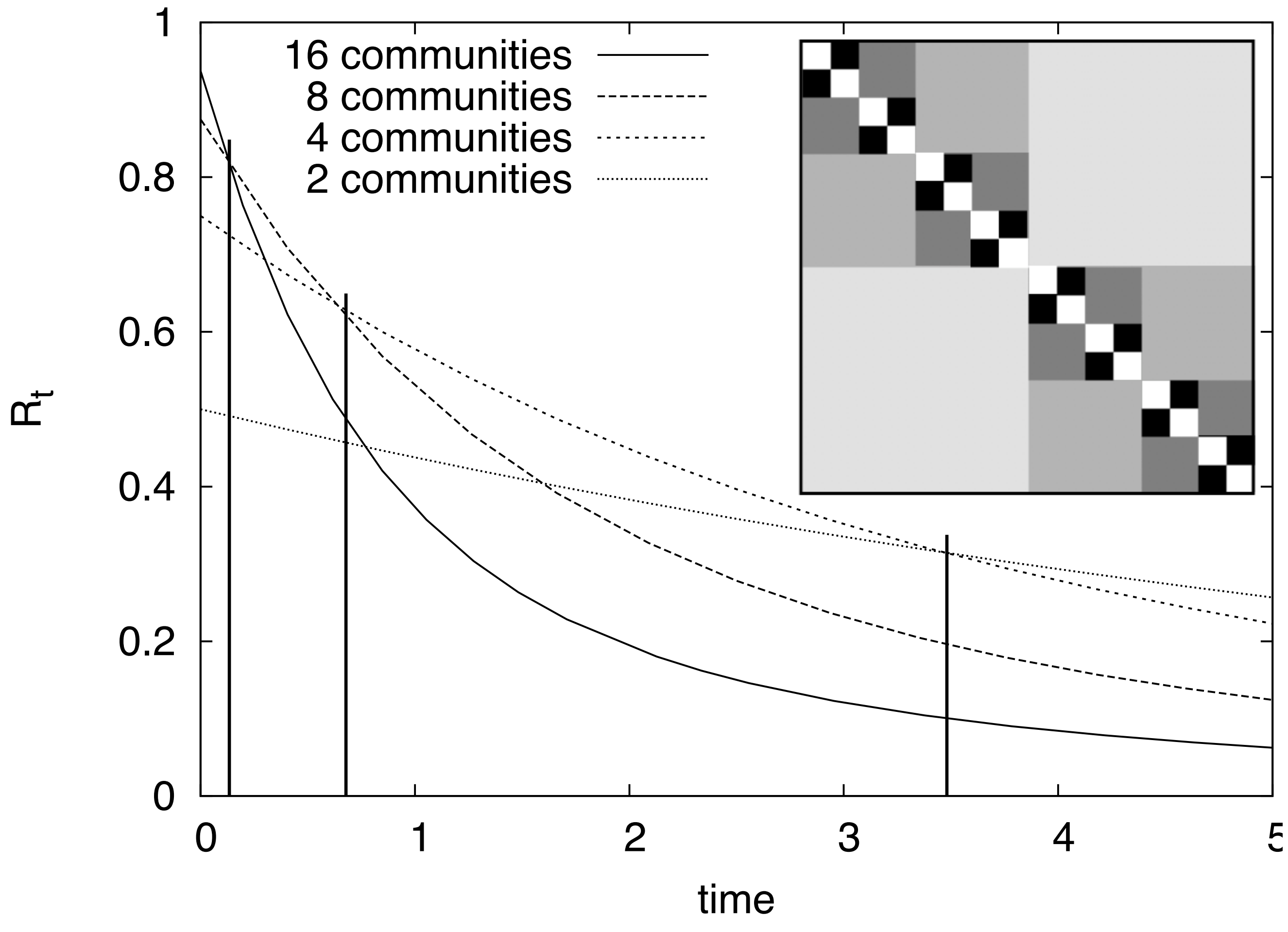
$$R_t \approx (1 - t) R_0 + t Q_c \quad Q_t$$

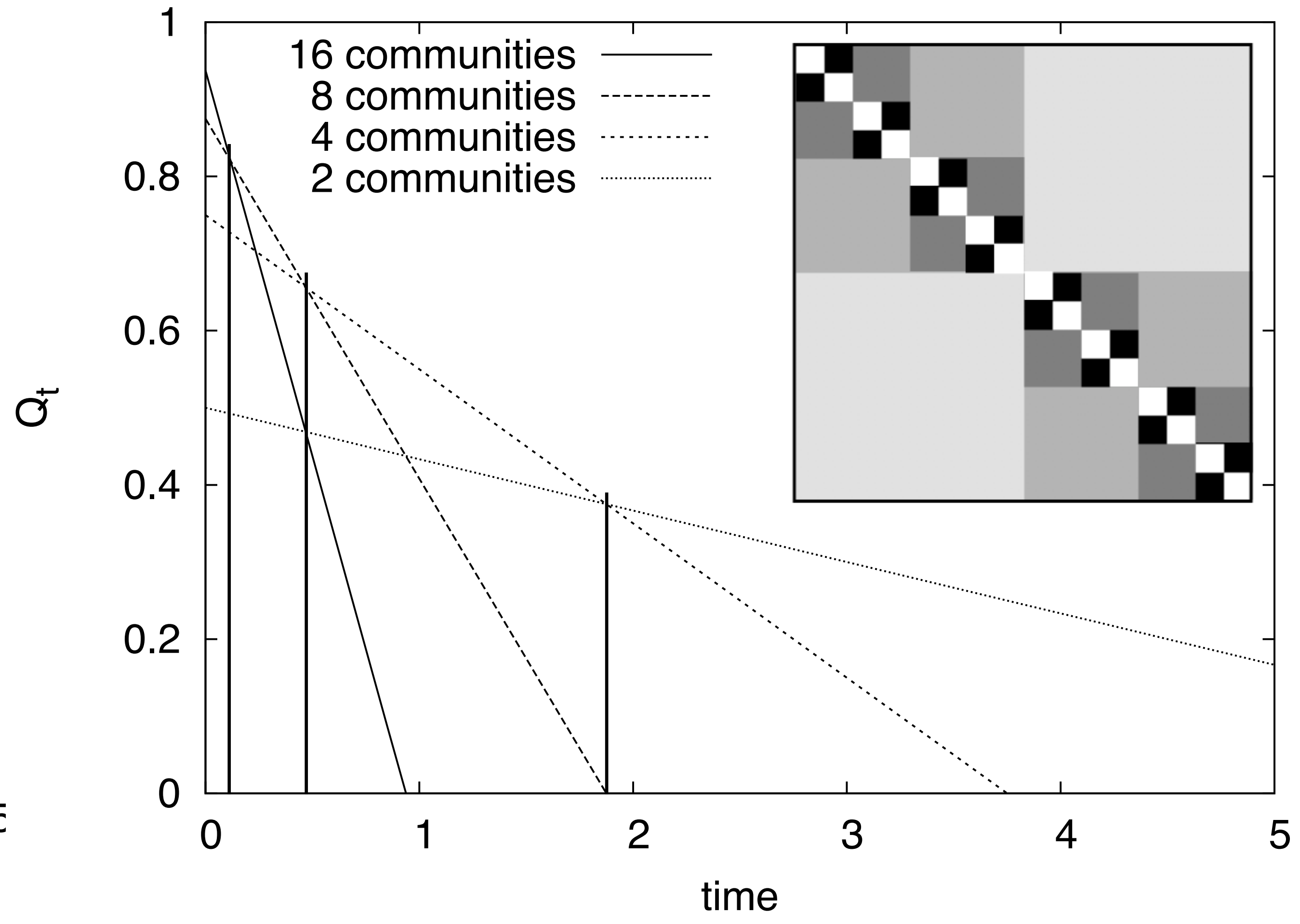
A Arenas, A Fernandez, S Gomez, New J. Phys. ***10****, 053039 (2008).*
Analysis of the structure of complex networks at different resolution levels

In practice: how to optimize these quantities?

As a first step, we have developed two methods to optimise

$$R_t \approx (1-t)R_0 + tQ_c \qquad Q_t$$





In practice: how to optimize these quantities?

As a first step, we have developed two methods to optimise

$$R_t \approx (1 - t)R_0 + tQ_c \qquad Q_t$$

Simulated Annealing method

~ a few thousand nodes

“Louvain” method: greedy algorithm

virtually no limit in the size of the network (10^9 links)

R. Guimera, M. Sales and L.A.N. Amaral, *Phys. Rev. E* **70,** 025101 (2004).
V.D. Blondel, J.-L. Guillaume, RL and E. Lefebvre, *J. Stat. Mech.*, P10008 (2008).

Advantages of the Louvain method

Known to perform very well for optimising modularity:

	Karate	Arxiv	Internet	Web nd.edu	Phone	Web uk-2005	Web WebBase 2001
Nodes/links	34/77	9k/24k	70k/351k	325k/1M	2.04M/5.4M	39M/783M	118M/1B
CNM	.38/0s	.772/3.6s	.692/799s	.927/5034s	-/-	-/-	-/-
PL	.42/0s	.757/3.3s	.729/575s	.895/6666s	-/-	-/-	-/-
WT	.42/0s	.761/0.7s	.667/62s	.898/248s	.553/367s	-/-	-/-
Our algorithm	.42/0s	.813/0s	.781/1s	.935/3s	.76/44s	.979/738s	.984/152mn

Also works quite well for Q_t. E.g., football network:

SA, *iteration factor=1.0*; *cooling factor=0.995*

0.0 115 0.991245 115 23
0.1 115 0.894656 33 63
0.2 115 0.850513 14 58
0.3 115 0.817482 12 40
0.4 115 0.786487 12 36
0.5 115 0.755492 12 36
0.6 115 0.724497 12 51
0.7 115 0.693602 11 36
0.8 115 0.663422 11 33
0.9 115 0.633770 10 58
1.0 115 0.601357 9 79

Louvain method:

0.0 115 0.991245 115 0
0.1 115 0.894616 33 0
0.2 115 0.850513 14 0
0.3 115 0.817482 12 0
0.4 115 0.786487 12 0
0.5 115 0.755492 12 0
0.6 115 0.724497 12 0
0.7 115 0.693602 11 0
0.8 115 0.663422 11 0
0.9 115 0.633792 10 0
1.0 115 0.604570 10 0

V.D. Blondel, J.-L. Guillaume, RL and E. Lefebvre, *J. Stat. Mech.*, P10008 (2008).

Conclusion

- Relation between dynamics and the hierarchical structure of networks
- Different dynamics lead to different quality functions for the partition of a graph
- Changing time allows to zoom in and out
- Algorithms developed in order to study very large networks

Original Louvain method available on *http://findcommunities.googlepages.com*

Generalized codes to optimise Q_t available on http://*www.lambiotte.be*

Thanks to R. Guimera and J.-L. Guillaume (for providing their codes)

R. Lambiotte, J.-C. Delvenne and M. Barahona, *arXiv:0812:1770*

Conclusion

N.B. Different dynamics lead to different quality functions for the partition of a graph (but most of the interesting cases should be related to a Laplacian of the adjacency matrix)

$$\partial_t p_i = \underset{j}{} \frac{A_{ij}}{k} p_j - \frac{k_i}{k} p_i$$

$$\partial_t p_i = \underset{j}{} \frac{A_{ij}}{k_j} p_j - p_i$$

Conclusion

- Relation between dynamics and the hierarchical structure of networks
- Different dynamics lead to different quality functions for the partition of a graph
- Changing time allows to zoom in and out
- Algorithms developed in order to study very large networks

Original Louvain method available on *http://findcommunities.googlepages.com*

Generalized codes to optimise R_t available on http://*www.lambiotte.be*

Thanks to R. Guimera and J.-L. Guillaume (for providing their codes)