SCHOOL FOR ADVANCED STUDIES LUCCA

Networks Days
Bridging micro with macro
Padua, 24-25 October 2024

Diffusion processes allow edges classification and clustering in complex networks

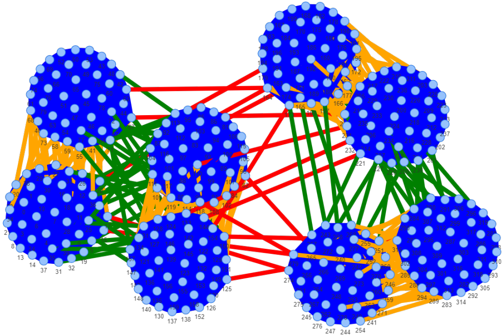
Francesca Santucci, Pablo Villegas, Andrea Gabrielli, Guido Caldarelli and Tommaso Gili

25/10/2024,
Padova, Italia

Information diffusion and edge classes

We propose an **edge-centric** description of a **diffusivity process** to **identify classes of edges** characterized by particular **diffusion timescales**.

→ Decompose diffusion into **Laplacian eigenmodes**
→ Extract the typical **diffusion timescales**.
→ Identify the **classes of edges characterized by each timescale**.
→ Explain the tight **link between edge classes**, diffusion timescales, and graph **hierarchical modular structure**.



From diffusion to the density operator

Information diffusion

$$\boldsymbol{x}(\tau) = \underbrace{e^{-\tau\hat{L}}}_{\text{Network propagator}} \boldsymbol{x}(0)$$

Where:
- $\hat{A}$: **Adjacency matrix** (undirected);
- $\hat{L} = \hat{D} - \hat{A}$, with $D_{ij} = \delta_{ij} \sum_k A_{ik}$: **(Symmetric) graph Laplacian**;
- $x(t)$: state of the network (x_i = info at node i);

Density operator $\hat{\rho}(\tau)$

Effective integrated information at time τ [a]

$$\hat{\rho}(\tau) = \frac{e^{-\tau\hat{L}}}{Tr(e^{-\tau\hat{L}})}$$

[a] De Domenico and Biamonte, "Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison".

Information integration matrix $\hat{\zeta}(\tau)$

$$\zeta_{ij}(\tau) = H\left(\frac{\rho_{ij}(\tau)}{min(\rho_{ii}, \rho_{jj})} - 1\right)$$

$H(\cdot)$ is the Heaviside function

Two nodes effectively share information when they have integrated an amount greater or equal than the information contained in one of the two nodes.

Information diffusion as a sequence of phase transitions

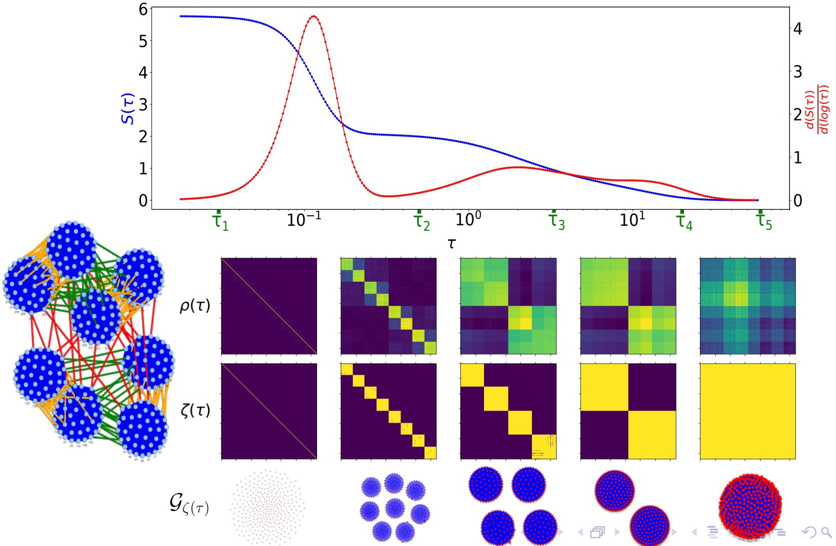
Spectral entropy

Order parameter for 2nd order transitions (peaks in derivative)[a]

$$S[\hat{\rho}(\tau)] = -Tr(\hat{\rho})\ln(\hat{\rho})$$

[a] Villegas et al., “Laplacian paths in complex networks: Information core emerges from entropic transitions”.

Edge-centric perspective: Diffusion as progressive integration of groups of edges in ζ

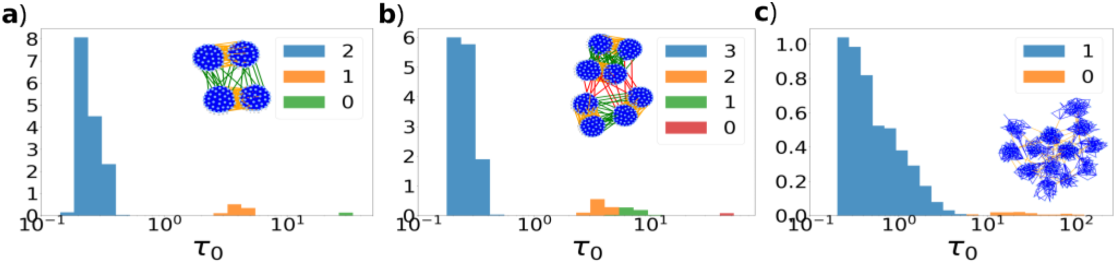


Integration time edge classes

Integration time

Time ij effectively aggregates the information at "its sides"

$$\tau_0^{ij} = \tau \text{ s.t } f(\tau) = \max_{\iota \in \{i,j\}} \{\rho_{ij}(\tau) - \rho_{\iota\iota}(\tau)\} = 0 \quad (1)$$



Eigenmode decomposition of the information integration condition

Nodes i and j share information if $\zeta_{ij} = 1$, i.e if $\max_{\iota \in \{i,j\}}(\rho_{ij} - \rho_{\iota\iota}) \geq 0$

Remember:

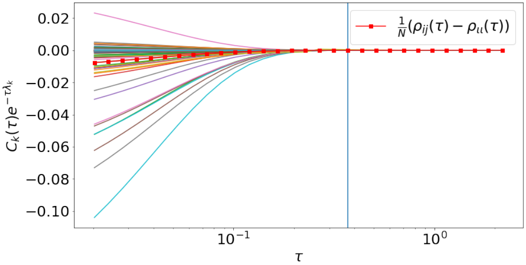
$\hat{\rho} = \frac{e^{-\tau \hat{L}}}{Z}$

Since $\hat{L}$ is symmetric, $\in \Re$ and semi-positive defined:

$\hat{L} = U \Lambda U^T \Rightarrow \rho_{ij} \propto \sum_{k=0}^{N-1} u_{ik} u_{jk} e^{-\tau \lambda_k}$

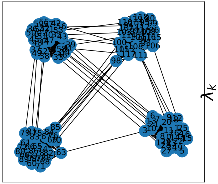
$$\zeta_{ij} = 1 \Leftrightarrow \max_{\iota \in \{i,j\}} \Big(\sum_{k=0}^{N-1} e^{-\tau \lambda_k} \underbrace{(u_{ik} u_{jk} - u_{\iota k}^2)}_{c_{k\iota}^{(ij)}} \Big) \geq 0$$

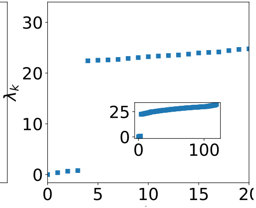
$$\Rightarrow \boxed{\max_{\iota \in \{i,j\}} \sum_{k=0}^{N-1} e^{-\tau \lambda_k} c_{k\iota}^{(ij)} \geq 0}$$



- c_k^{ij}: mode's amplitude, **edge specific**.

Information integration eigenmodes and modular structure





In this case, we can approximate the diffusion condition $\max_{\iota\in\{i,j\}} \sum_k^{N-1} e^{-\tau\lambda_k} c_{k\iota}^{ij}(\tau) \geq 0$ as:

$$\max_{\iota\in\{i,j\}} \Big(e^{-\tau\lambda_1} \underbrace{\gamma_{1\iota}^{ij}}_{\sum_{k=1}^{M-1} c_{k\iota}^{ij}} + \sum_{k\geq M}^{N-1} e^{-\tau\lambda_k} c_{k\iota}^{ij}\Big) \geq 0$$

necessary condition (less stringent) $\Leftarrow$

Verified for $\tau_0' < \tau_0$

$$\boxed{e^{-\tau\lambda_1} \underbrace{\max_{\iota\in\{i,j\}} (\gamma_{1\iota}^{ij})}_{\equiv \gamma_1^{ij}} + \sum_{k=1}^{N-1} e^{-\tau\lambda_k} \underbrace{\max_{\iota\in\{i,j\}} (c_{k\iota}^{ij})}_{\equiv c_k^{ij}} \geq 0}$$

- **Advantage: removing the time** in the evaluation of the maximum.

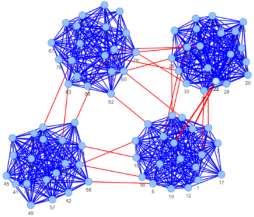
Sign of γ_1, communication time scale and edge role

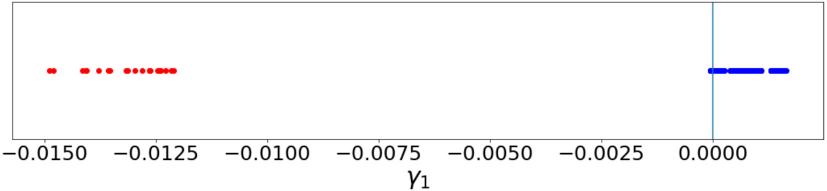
$\gamma_1^{ij} < 0 \;\Rightarrow\;$ Slower communication ($\tau_0' > \frac{1}{\lambda_1}$)

$\gamma_1^{ij} \geq 0 \;\Rightarrow\;$ Faster communication

Remember the activation condition:

$$e^{-\tau\lambda_1}\boxed{\gamma_1^{ij}} + \sum_{k=1}^{N} e^{-\tau\lambda_k} c_k^{ij} \geq 0$$

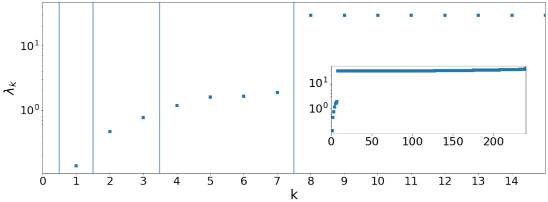




Linked to the **modular structure**: Due to Laplacian eigenvectors features (see spectral clustering), bridges tend to be associated with $\gamma_1 < 0$, and vanish otherwise.
$\rightarrow$ We **use** γ_1^{ij} **to classify the edges based on a "faster" (internal) and "slower" (external) time scale.**

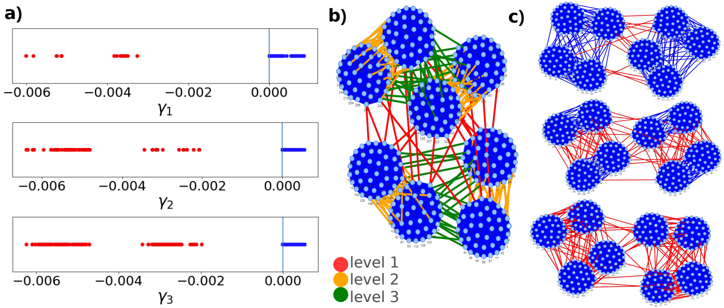
Hierarchical modularity based edge classification algorithm

i) Identify the indexes K_l corresponding with the gaps in the Laplacian spectrum;

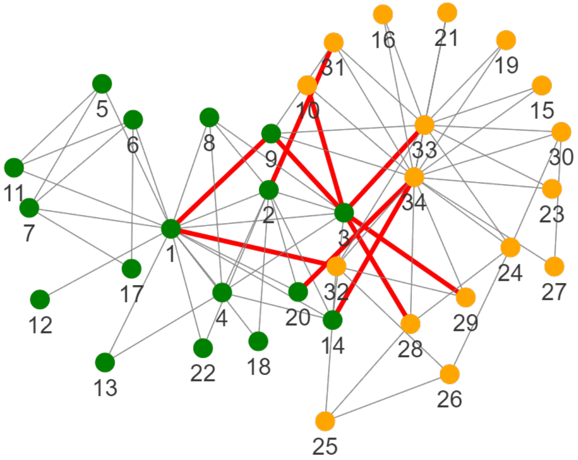


ii) Compute for each edge ij
$\gamma_l^{ij} = \max_{\iota \in \{i,j\}} (\sum_{K_l \le k < K_{l+1}} c_{k\iota}^{ij});$

iii) Assign each edge to the class characterized by a time-scale longer than $1/\lambda_l$, where λ_l is the smallest eigenvalue associated with a $\gamma_l < 0$.



Karate-club

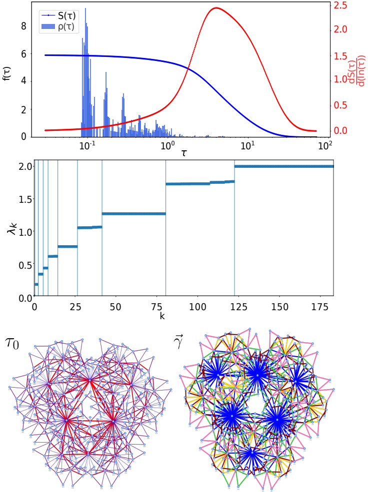


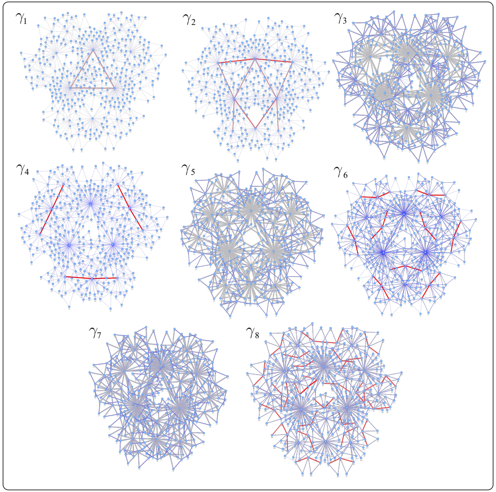
16
21
31
19
5
10
15
6
8
33
9
30
11
34
2
23
7
3
1
24
27
17
32
12
20
29
4
14
28
22
18
13
26
25

Dorogostev-Golstev-Mendes graph[1]

Diffusion modes patterns

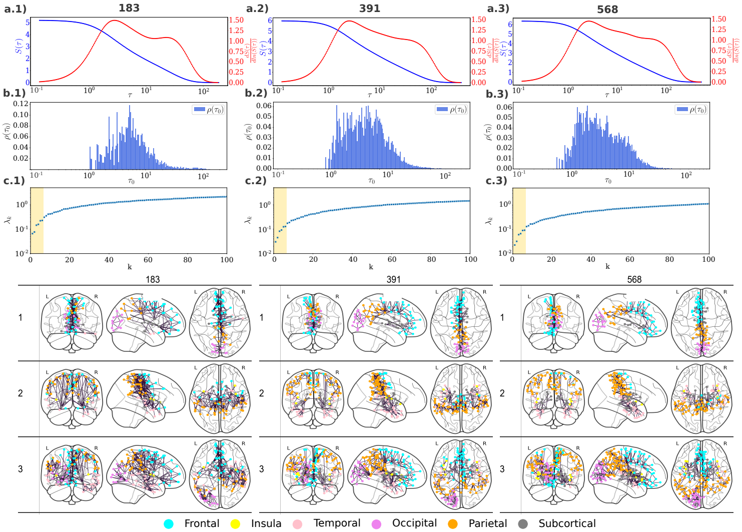
$$\sum_{k=l}^{L} e^{-\tau\lambda_l} \boxed{\gamma_l^{ij}} \geq 0$$





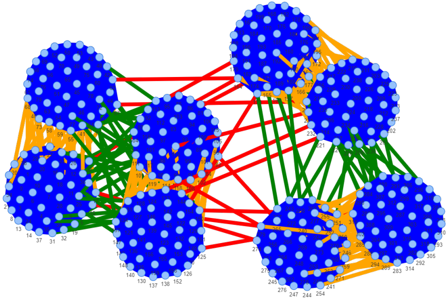
1 Dorogovtsev, Goltsev, and Mendes, “Pseudofractal Scale-free Web”.

Structural connectivity brain network



Conclusions

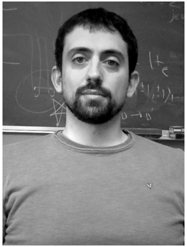
- We developed tools for the analysis of **diffusion pathways**, based on the identification of **edge classes**, which uncover the **aggregation phases** of the diffusion process.
- Edge classes can be identified by:
  - The **integration time**, provided the diffusion time scales are separated enough.
  - The **Laplacian eigenmodes**. In the case of a **hierarchical modular graph**, it is possible to associate Laplacaian eigenmodes, **diffusion time scales**, and nested structures.
- Laplacian eigenmodes allow the identification of edge patterns in complex networks (see **Dorogostev-Golstev-Mendes** graph and the **Human brain** network)



Thank you for your attention!

Francesca Santucci, PhD student at [1],
francesca.santucci@imtlucca.it

The rest of the team:



Tommaso Gili[1] Pablo Villegas[2] Guido Caldarelli[3] Andrea Gabrielli [45]

[a]Networks Unit, IMT School for Advanced Studies, Lucca, Italy

[b]'Enrico Fermi' Research Center (CREF), Rome, Italy

[c]4DSMN Ca'Foscari, University of Venice, Venezia Mestre, Italy

[d]Dipartimento di ingegneria Civile, Informatica e delle Tecnologie Aeronautiche, Università degli Studi Roma Tre, Rome, Italy

[e]Istituto dei Sistemi Complessi - CNR, Rome, Italy.

Diffusion on the graph

$$
\begin{aligned}
\frac{dx_i}{dt} = & -k\sum_j A_{ij}(x_i - x_j) = -k[x_i\sum_j A_{ij} - \sum_j A_{ij}x_j] = \\
& -k[x_i k_j - \sum_j A_{ij}x_j] = -k[\sum_j (k_j\delta_{ij} - A_{ij})x_j] = \\
& -k[\sum_j L_{ij}x_j] \\
& \rightarrow \frac{d\boldsymbol{x}}{dt} = -k\hat{L}\boldsymbol{x}
\end{aligned} \tag{2}
$$

Diffusion and eigenmodes

Diffusion equation: $\dot{\boldsymbol{x}} = -\hat{L}\boldsymbol{x} \implies \boldsymbol{x(t)} = e^{-\tau\hat{L}}\boldsymbol{x(0)}$

Diagonalizing: $U^T\boldsymbol{x(t)} = U^T e^{-\tau\hat{L}} U U^T \boldsymbol{x(0)}$

$$y_i(t) = e^{-\tau\lambda_i} y_i(0) \tag{3}$$

$\rightarrow$ Each y_i evolves independently, decaying with λ_i: as τ grows, the modes corresponding to bigger λ_i are 'turned off', and in the end, only the stationary mode ($\lambda_N = 0$, $u_N = \frac{1}{\sqrt{N}}$) survives.

The contribution of different eigenmodes to x_i is given by the i components of the different modes eigenvectors.

$$x_i(\tau) = \sum_k u_{ik} y_k(\tau) \tag{4}$$

Network propagator

$$\begin{aligned} \boldsymbol{x}(\boldsymbol{\tau}) &= \underbrace{e^{-\tau \hat{L}}}_{K} \boldsymbol{x}(\boldsymbol{0}) \\ x_i(\tau) &= \sum_{j}^{N} K_{ij}(\tau) x_j(0) \end{aligned} \tag{5}$$

K_{ij} counts the number of diffusion paths between j and i at time τ

Spectral entropy

$$S[\rho(\hat{\tau})] = \frac{1}{\ln(N)} \sum_{i}^{N} \mu_i(\tau) \ln \mu_i(\tau) \tag{6}$$

$\mu_i(\tau)$ are $\hat{\rho}(\tau)$'s eigenvalues, related to the Laplacian eigenvalues λ_i by: $\mu_i(\tau) = \frac{e^{-\tau\lambda_i}}{\sum_k e^{-\lambda_k \tau}}$

Specific heat:

$$C = \frac{dS}{d(\ln \tau)} \tag{7}$$

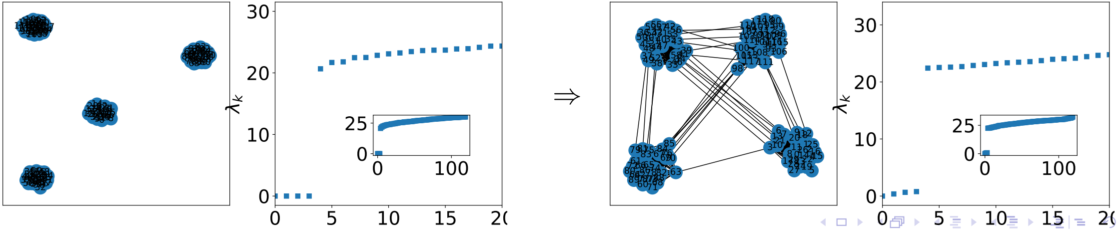
Its peak signals a deceleration of information diffusion, separating different information flow rate regions of the graphs.

Spectral properties of $\hat{L}$

- $\lambda_0 = 0$
- **multiplicity** of $\lambda = 0 \rightarrow$ **number of connected components** of G, and its eigenspace is **spanned by** multiples of **the indicator vectors** of those components $\rightarrow$ stationary mode(s).

If we slightly modify a graph with M connected components to obtain a connected graph, we expect the spectrum of the new graph to be still close to the original one (CFR spectral clustering)[a].

[a] Ulrike Von Luxburg. "A tutorial on spectral clustering". In: *Stat Comput* (2007).



Star-clique