Quantifying diversity in multiplex networks

Laura Carpi,[1] Tiago Schieber,[2] Panos Pardalos,[3]
Gemma Marfany,[4] Cristina Masoller,[5]
Albert Diaz-Guilera,[4] and Martin Ravetti[2]

[1]Centro Federal de Educacao Tecnologica de Minas Gerais
[2]Universidade Federal de Minas Gerais
[3]University of Florida
[4]Universitat de Barcelona
[5]Universitat Politecnica de Catalunya

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

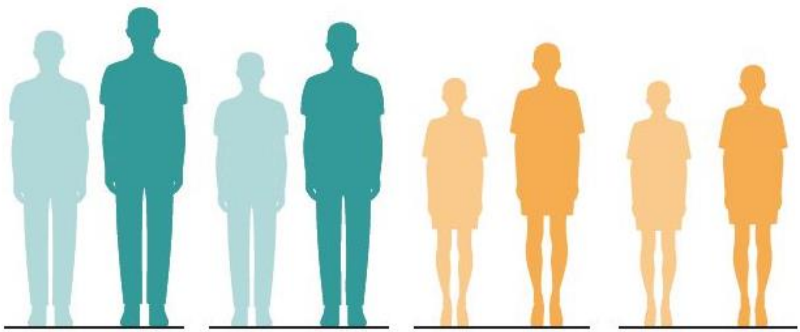
ICREA

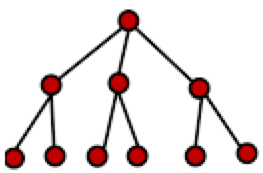
GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

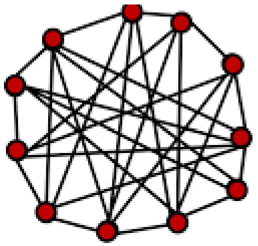
What is diversity?

Takes into account three characteristics of a population:

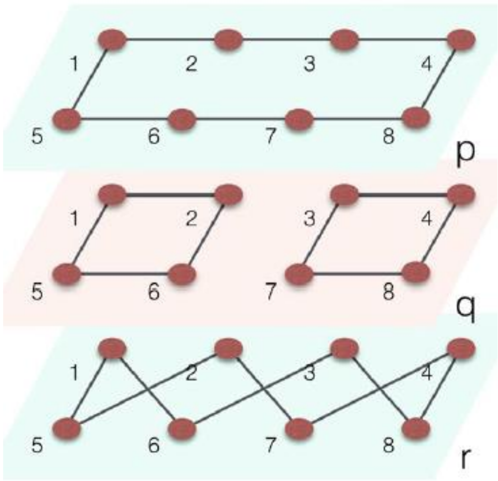
- Diversity in some **attributes** (e.g. atoms with different masses; people with different heights),
- Diversity of **types** (e.g., atoms or molecules; males or females),
- Diversity in **configuration** (e.g., configuration of atoms in a molecule; hierarchical or unstructured relations).







In the context of multiplex networks: what is diversity?



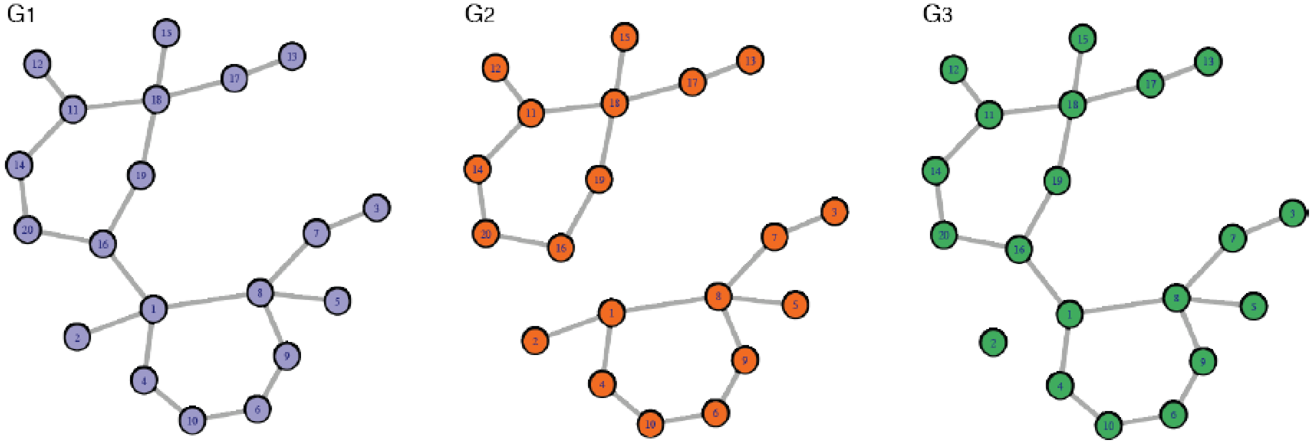
M $(N \times N)$ adjacency matrices, $\mathcal{A} = \{A^{[1]}, A^{[2]}, \ldots, A^{[M]}\}$

- Different **configurations** (connection paths) of the set of elements
  - the nodes
  - the layers
- Two types:
  - local: paths that a node has in the different layers;
  - global: how diverse the layers are.
- To quantify diversity we first need to define a “distance” to compare
  - the paths of a node in the different layers,
  - the different layers.

Some network distances

$$d_{\text{Hamming}}(\boldsymbol{y}_1, \boldsymbol{y}_2) = \sum_{i \neq j}^{N} \left[A_{ij}^{(1)} \neq A_{ij}^{(2)} \right]$$

- Graph Edit distance: operations required to make graph G isomorphic to graph H.
- Entropy distances: entropy-like measures associated with distributions extracted from the graphs.
- Spearman’s correlation coefficient: rank correlation between sorted lists of vertices of the two graphs.
- Etc.
- Main problem: not all the links have the same importance.



Node distance distribution

- NDD of node **i** in layer **p**: $N^p_i(d)$ is the fraction of nodes that are at distance d (shortest path) from node **i** in layer **p**.
- The set of N distributions, $N^p_1 \ldots. N^p_N$, contains information about the global topology of layer **p**, in a compact way, and can be used to define a distance between unlabeled graphs.
- If two graphs have the same set of NDDs $\Rightarrow$ they have the same diameter, average path length, etc.

T. A. Schieber et al, Nat. Comm. 8, 13928 (2017)

Transition matrix of a layer

- T^p is the adjacency matrix of layer **p**, rescaled by the degree of each node.
- T^p_{ij} is the probability that, in layer **p**, node **j** is reached, in one step, by a random walker located at node **i** in **p**.
- It contains information about the local topology of layer **p**.

1) To quantify the differences of the connectivity paths of a node in two layers,

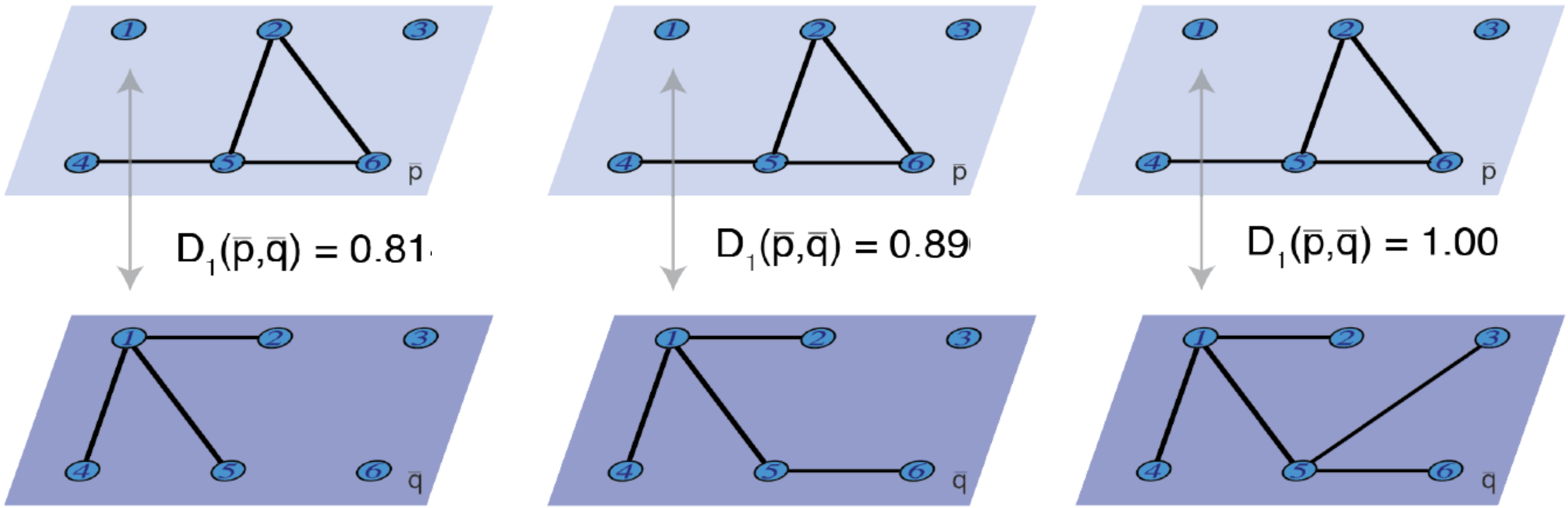
$$\mathcal{D}_i(\overline{p},\overline{q}) = \frac{\sqrt{\mathcal{J}(\mathcal{N}_i^{\overline{p}},\mathcal{N}_i^{\overline{q}})} + \sqrt{\mathcal{J}(T_i^{\overline{p}},T_i^{\overline{q}})}}{2\sqrt{\log(2)}}$$

- 0 indicates that node **i** has identical connectivity paths in layers **p** and **q**,
- 1 indicates that node **i** is not connected (or not active) in one layer, while there are paths connecting **i** to all nodes in the other layer

2) To quantify differences between layers p and q

$$\mathcal{D}(\overline{p},\overline{q}) = \langle \mathcal{D}_i(\overline{p},\overline{q}) \rangle_i$$

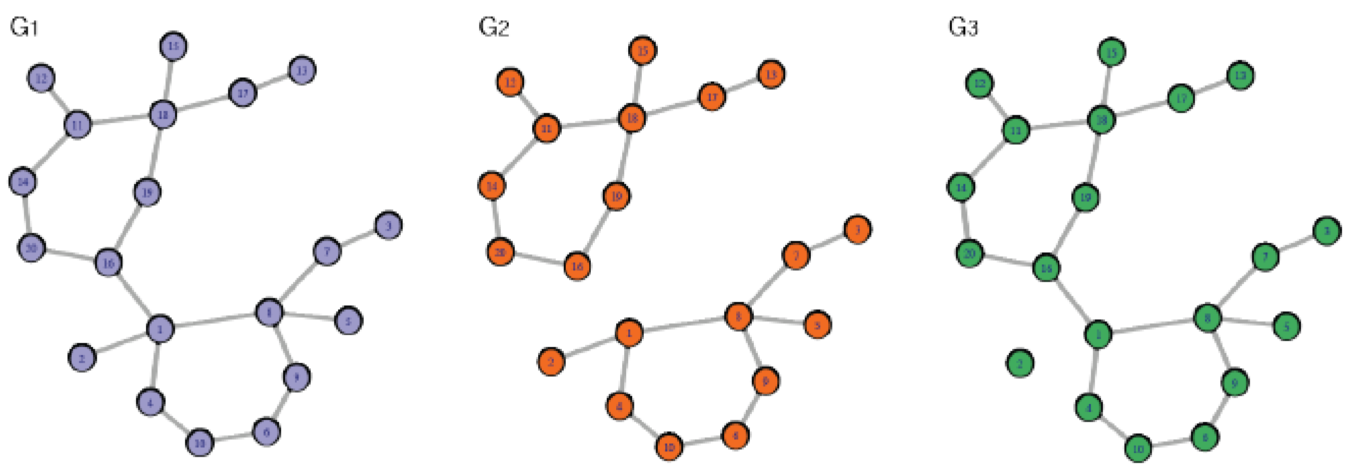
Examples



$$\mathcal{D}(\overline{G1},\overline{G3}) = 0.14$$

$$\mathcal{D}(\overline{G2},\overline{G3}) = 0.34$$

$$\mathcal{D}(\overline{G1},\overline{G2}) = 0.36$$



Diversity measure

- Main idea: the diversity of a system is defined by the distances between its elements: the larger the distances, the more different the elements are, and the more diverse the system is.

- The distance between the element **g** $\not\subset$ S and the set S, **D(g; S)**, is the smallest distance between **g** and any of the elements of S, $\mathcal{D}(\overline{g}, S) = \min_{\overline{s_i} \in S} \mathcal{D}(\overline{g}, \overline{s_i})$

- Recurrent def.: $U(S) = max_{\overline{s_i} \in S}\{U(S \setminus \overline{s_i}) + \mathcal{D}(\overline{s_i}, S \setminus \overline{s_i})\}$

  $U(S)=0$ if $|S|=1$

- Diversity increases when a different element is included

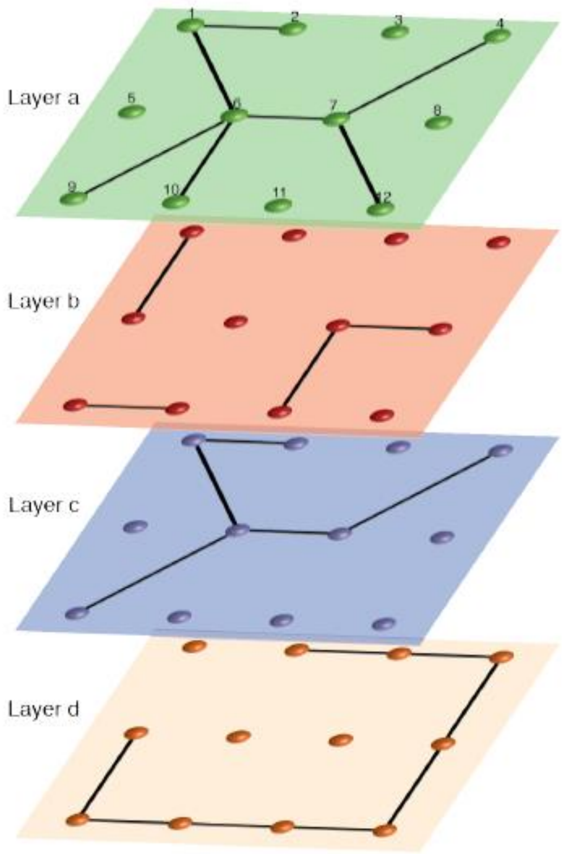
$$U(S \cup \overline{g}) \geq U(S) + D(\overline{g}, S)$$

Example

- Step 0: D(a, b)=0.70, D(a, c)=0.21, D(a, d)=0.79, D(b, c)=0.58, D(b, d) =0.69, D(c, d)=0.80
- Step 1: S = {a, b, c, d}, layers a and c have the smallest distance value, and c is the layer that less contributes to the diversity of S, as it is closer to the remaining layers. Then, the first step gives U(S) = U(S1) + 0.21 where S1 = S-c.
- Step 2: In S1 layers b and d present the smallest distance value, and b is the layer that less contributes to the diversity of S1. Therefore,

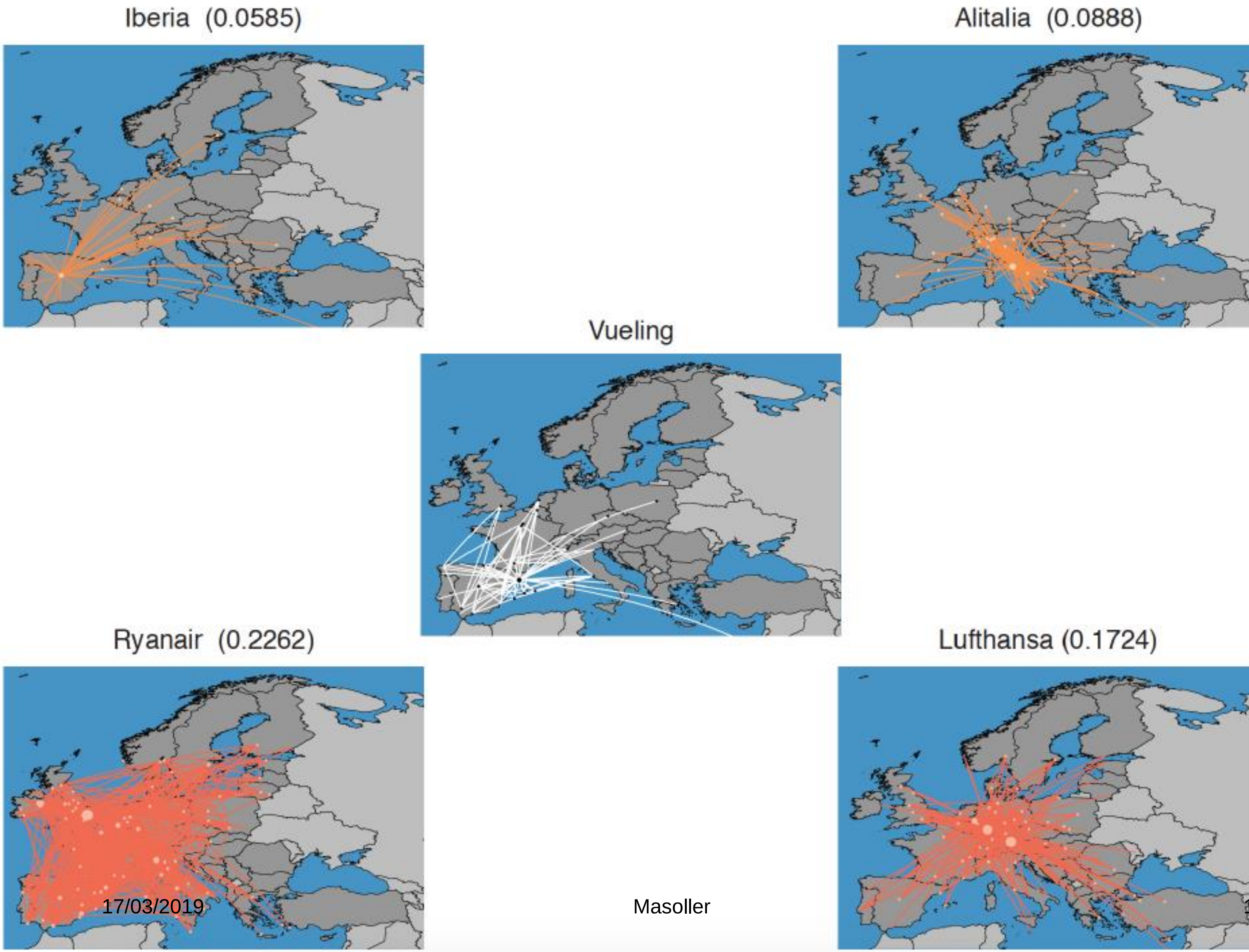
  U(S1) = U(S2) + 0.69 where S2 = S1 - b = {a, d}.
- Step 3: U(S2) = D(a, d).

$$\Rightarrow U(S) = 0.21 + 0.69 + 0.79 = 1.69$$



Carpi et al, Sci. Reports 9, 4511 (2019)

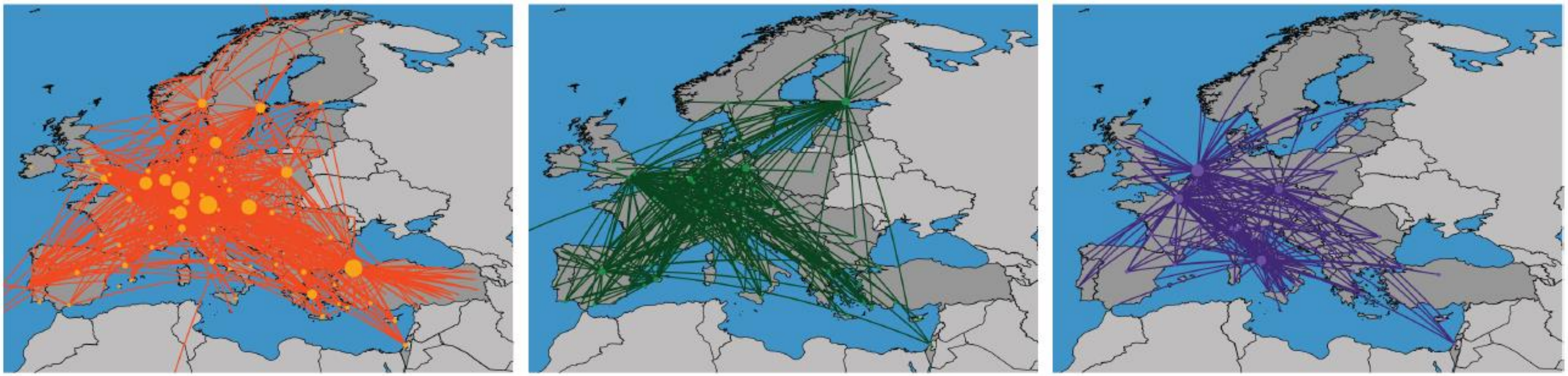
Analysis of Europe air traffic network



More similar routes to Vueling

More different routes to Vueling

Diversity ordering: Elements can be ordered in terms of their contribution to the diversity of the set.



a Star Alliance b Oneworld c Skyteam

U=0.97 U=0.34 U=0.33

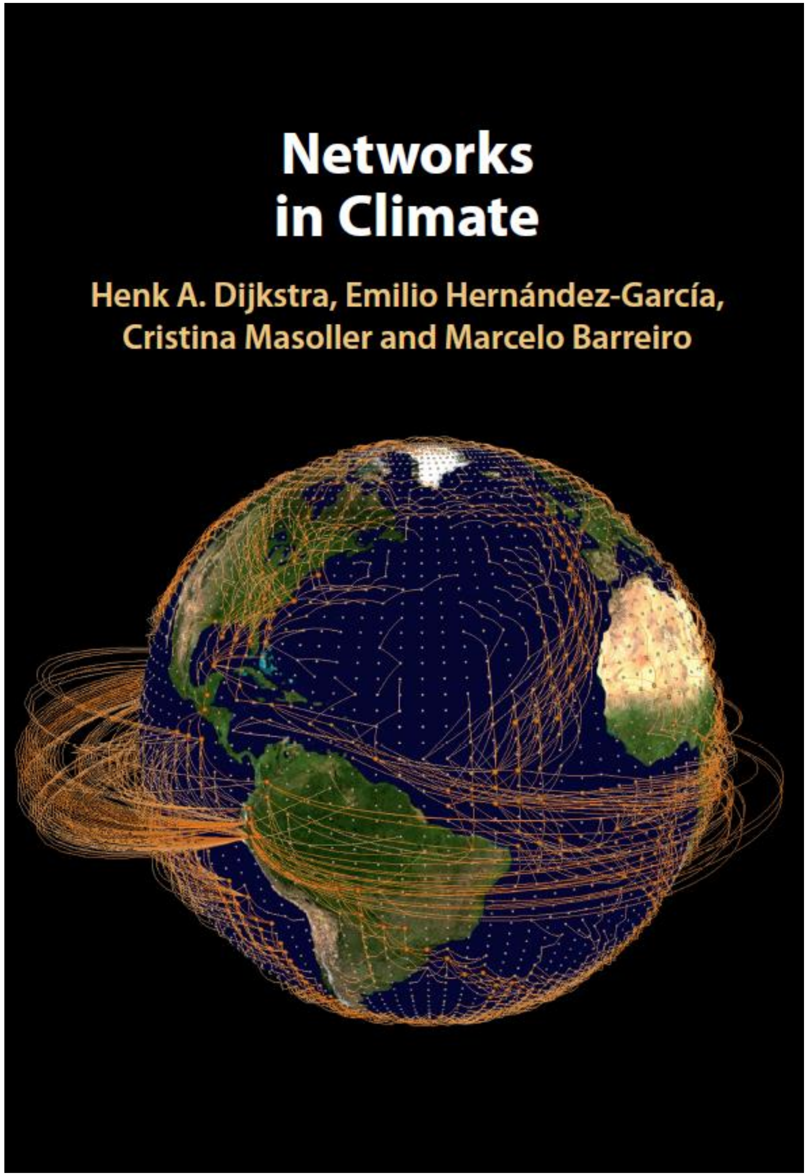
Diversity ordering of Star Alliance (from less to more contribution):

Brussels Airlines (BEL), Swiss Air (SWR), Polish Airlines (LOT), Air Portugal (TAP), Aegean Airlines (AEE), Austrian Airlines (AUA), Scandinavian Airlines (SAS), Turkish Airlines (THY), Lufthansa (DLH)

To summarize

- New distance between unweighted labelled graphs.
- Used to define the diversity of the connectivity paths of a node in the different layers, and the diversity of the connectivity paths of the whole set of layers.
- Analysis of air alliances reveals which airlines, when joining an alliance, optimally increase the diversity, bringing new routes while minimizing overlapping ones; and which ones, when leaving the alliance, less compromise the diversity of the routes offered by the alliance.

- Other applications: Carpi et al, Sci. Reports 9, 4511 (2019)



Cambridge University Press 2019

Advertisement

EU FP7 LINC (Learning about interacting networks in climate)

⇓

H2020 CAFÉ (Climate advanced forecasting of extremes)

Marie Curie PhD positions available

<cristina.masoller@upc.edu>
http://www.fisica.edu.uy/~cris/

THANK YOU FOR YOUR ATTENTION !