COMPLEXIFYING COMPLEX NETWORKS

SPATIAL NETWORKS

Literature

Physics Reports 499 (2011) 1–101

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

ELSEVIER

Spatial networks

Marc Barthélemy *

Institut de Physique Théorique, CEA, IPhT CNRS, URA 2306, F-91191 Gif-sur-Yvette, France
Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 CNRS-EHESS), Ecole des Hautes Etudes en Sciences Sociales, 54 bd. Raspail, F-75270 Paris Cedex 06, France

ARTICLE INFO

Article history:
Accepted 8 November 2010
Available online 23 November 2010
editor: H. Orland

Keywords:

ABSTRACT

Complex systems are very often organized under the form of networks where nodes and edges are embedded in space. Transportation and mobility networks, Internet, mobile phone networks, power grids, social and contact networks, and neural networks, are all examples where space is relevant and where topology alone does not contain all the information. Characterizing and understanding the structure and the evolution of spatial networks is thus crucial for many different fields, ranging from urbanism to epidemiology.

Vol 453 | 5 June 2008 | doi:10.1038/nature06958 nature

LETTERS

Understanding individual human mobility patterns

Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

Science AAAS

Limits of Predictability in Human Mobility
Chaoming Song, *et al.*
Science **327**, 1018 (2010);
DOI: 10.1126/science.1177170

LETTER

doi:10.1038/nature10856

A universal model for mobility and migration patterns

Filippo Simini[1,2,3], Marta C. González[4], Amos Maritan[2] & Albert-László Barabási[1,5,6]

A gravity model for inter-city telephone communication networks

G Krings[1,2], F Calabrese[2], C Ratti[2] and V D Blondel[1]

[1] Université catholique de Louvain (UCL), Department of Applied Mathematics - 4 Avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
[2] SENSEable City Laboratory, Massachusetts Institute of Technology - 77 Massachusetts Avenue, Cambridge, MA 02139, USA

E-mail: `gautier.krings@uclouvain.be`

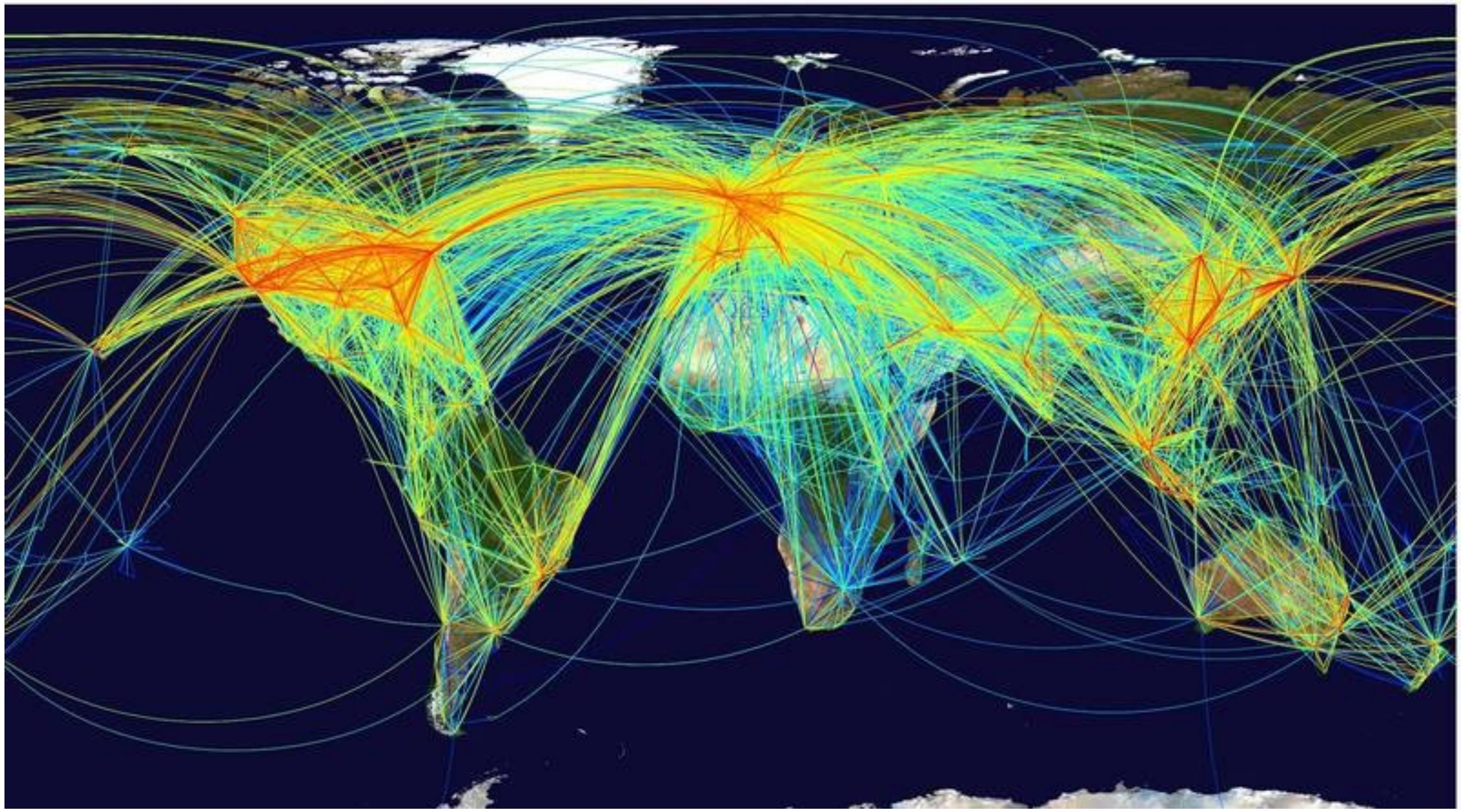
Spatial networks

Networks embedded in space

- Concept
  - Structure alone does not contain all the information about the network
  - Cost: wiring two nodes is not free but has a cost proportional to the distance of the nodes
  - It directly influences the structure of the emerging network

Distance

- Physical distance
- Economical distance
- Social distance
- Difference in professional categories
- …

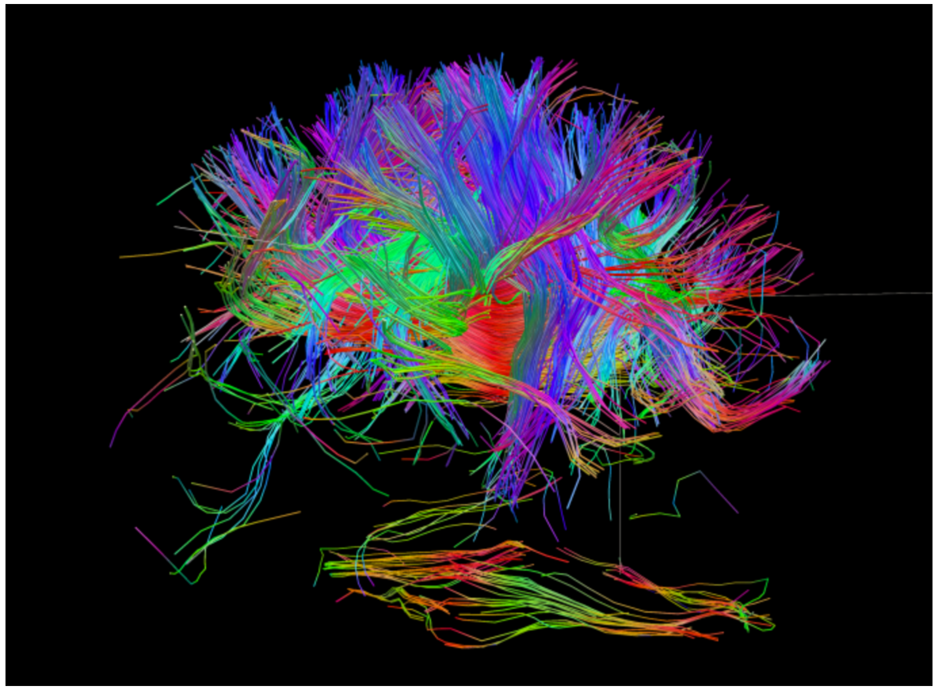


Geisel et.al. (2011)

Spatial networks

Types of spatial networks

- Transportation networks
  - Airline networks
  - Bus, subway, railway, and commuters
  - Cargo ship networks
- Infrastructure networks
  - Road and street networks
  - Power grids and water distribution networks
  - The internet
- Neural networks
- Protein networks
- Mobility networks
- …



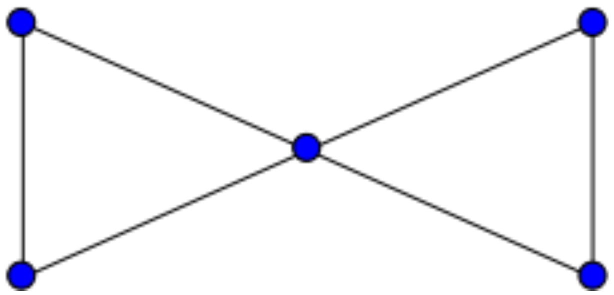
Spatial and Planar networks

Spatial networks

- Nodes are embedded in a metric space
  - $G=(V,E,s(n))$
  - $s: n \rightarrow (x,y)$, where $n \in V$ and (x,y) are coordinates of a metric space (e.g. Euclidean space)
  - probability of finding a link between nodes usually decreases with the distance

Planar networks

- Graph that can be drawn in the plane such that edges do not cross each other but at their endpoints
- Not all spatial networks are planar
- You do not need to know node positions to have a planar graph.



Mixed measures of topology and space

Distance strength

- Cumulative distance from a node i to all neighbours

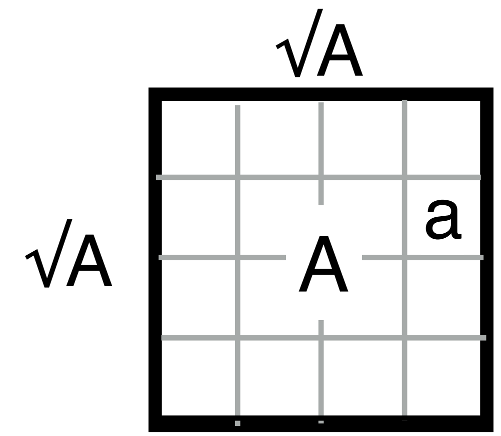
$$s_i^d = \sum_{j \in \Gamma(i)} d_E(i, j)$$

Relation Distance strength - degree

- Nodes with same degrees but different s^d
- Correlation s^d(k) and k ? (Similar to assortativity)
  - Central nodes have higher degrees?
  - Hubs tend to connect to farther nodes?
  - …

Barthélemy, M. (2014). *Spatial networks* (pp. 1967-1976). Springer New York.

Mixed measures of topology and space



Compactness (density measure for cities)

- Measures how much a city is filled with roads

$$\Psi = 1 - \frac{4A}{(\ell_T - 2\sqrt{A})^2}$$

- where A is the area, $L = \sqrt{A}$ is the linear size of a city and ℓ_T is the total length of roads
- $\Psi \in [0,1]$ (for non degenerate cases, i.e., at least a road, see below)
- With only one road circling the area: the total length is $\ell_T = 4\sqrt{A}$ and $\Psi = 0$
- If roads constitute a square grid of spacing a (where $a<\sqrt{A}$): $\ell_T = 2\frac{L^2}{a}$

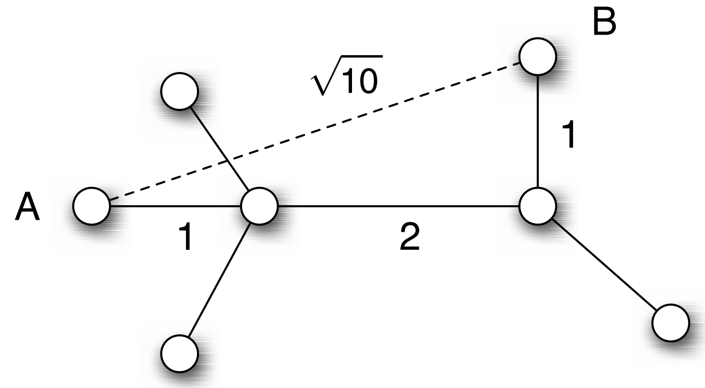
  and $\Psi = 1 - \frac{a^2}{L^2}$ is large

Mixed measures of topology and space

Route distance between two nodes

$$d_R(i, j)$$

- Sum of length of segments which belong to the shortest paths between two nodes in a spatial network



Route factor/Detour index

$$Q(i, j) = \frac{d_R(i, j)}{d_E(i, j)}$$

- d_E=Euclidean distance
- always larger than *1* but the closer to it, the most efficient.

Accessibility

- Route index for a specific node $\langle Q(i) \rangle = \frac{1}{N-1} \sum_j Q(i, j)$
- Average route index for the whole network $\langle Q \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} Q(i, j)$

Simple models

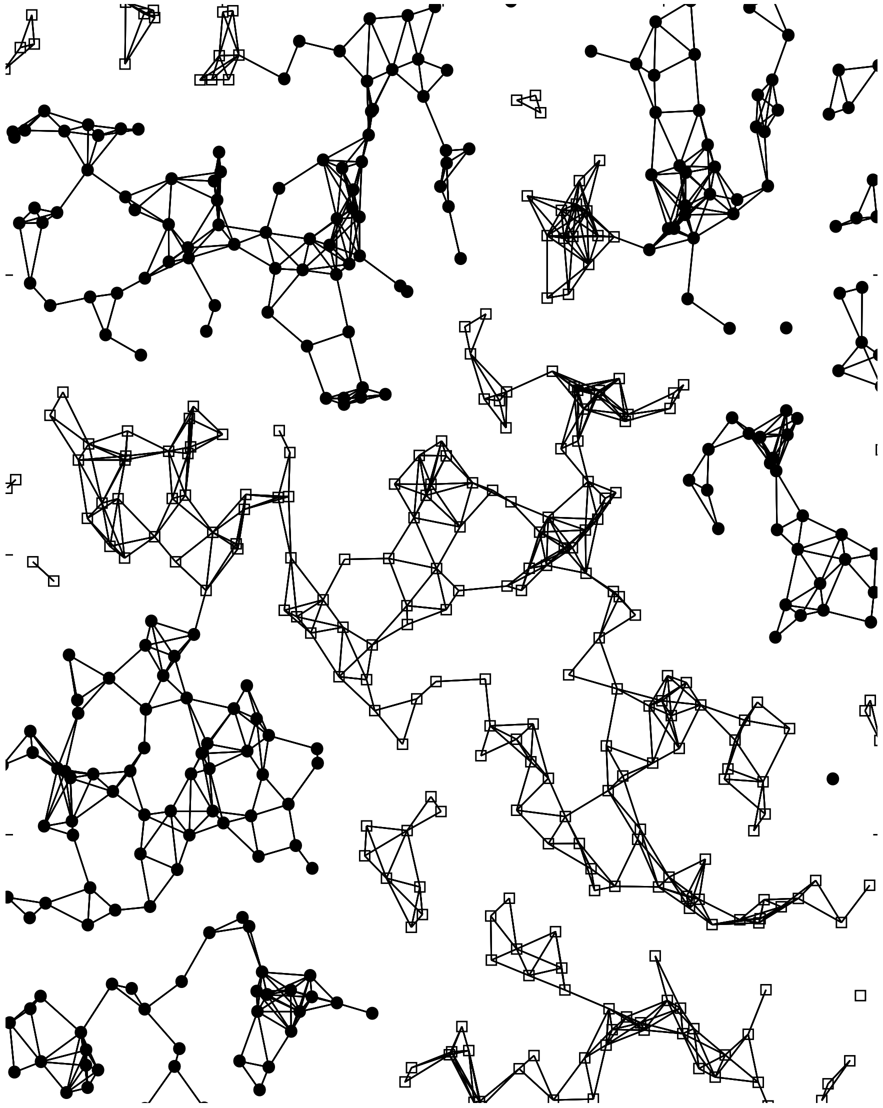
of

spatial networks

Random geometric graphs

General definition:

- Take a space and distribute nodes randomly
- Nodes are small spheres with radius r
- Two nodes are connected if their spheres overlap — separated with distance smaller than 2r
- Also called: disk-percolation
- Application: ad-hoc wireless communications



Degree distribution — Poisson distribution

$$P(k;x,\alpha) = \frac{1}{k!}\alpha^k p(x)^k \mathrm{e}^{-\alpha p(x)}$$

Clustering coefficient

$$\langle C_d \rangle \sim 3\sqrt{\frac{2}{\pi d}}\left(\frac{3}{4}\right)^{\frac{d+1}{2}}$$

Independent of N contrary to random networks

Spatial small world network

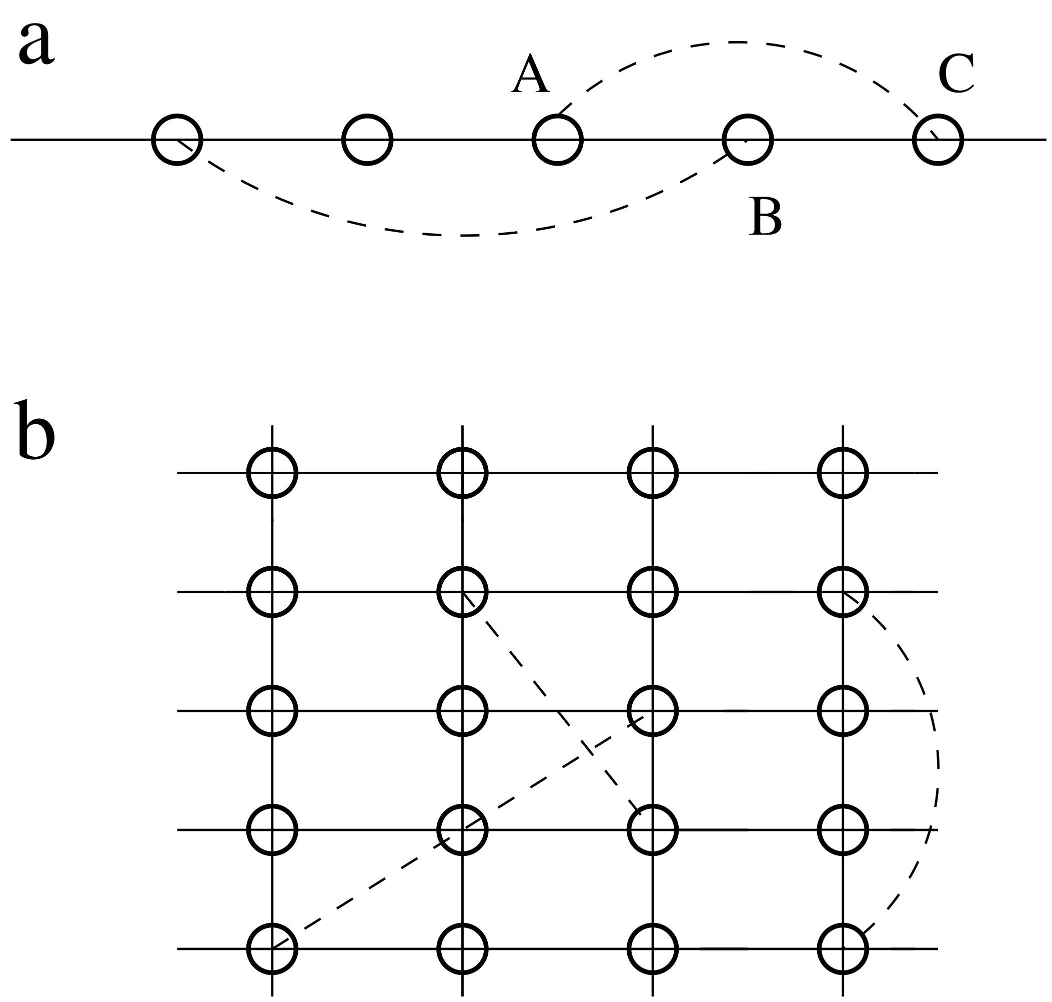
General definition (extension of Watts-Strogatz):

- Take a d dimensional regular lattice
- Assumption: shortcuts are not free but a cost associated with their length
- Links added with a probability q

$$q(\ell) \sim \ell^{-\alpha}$$

Shortest path

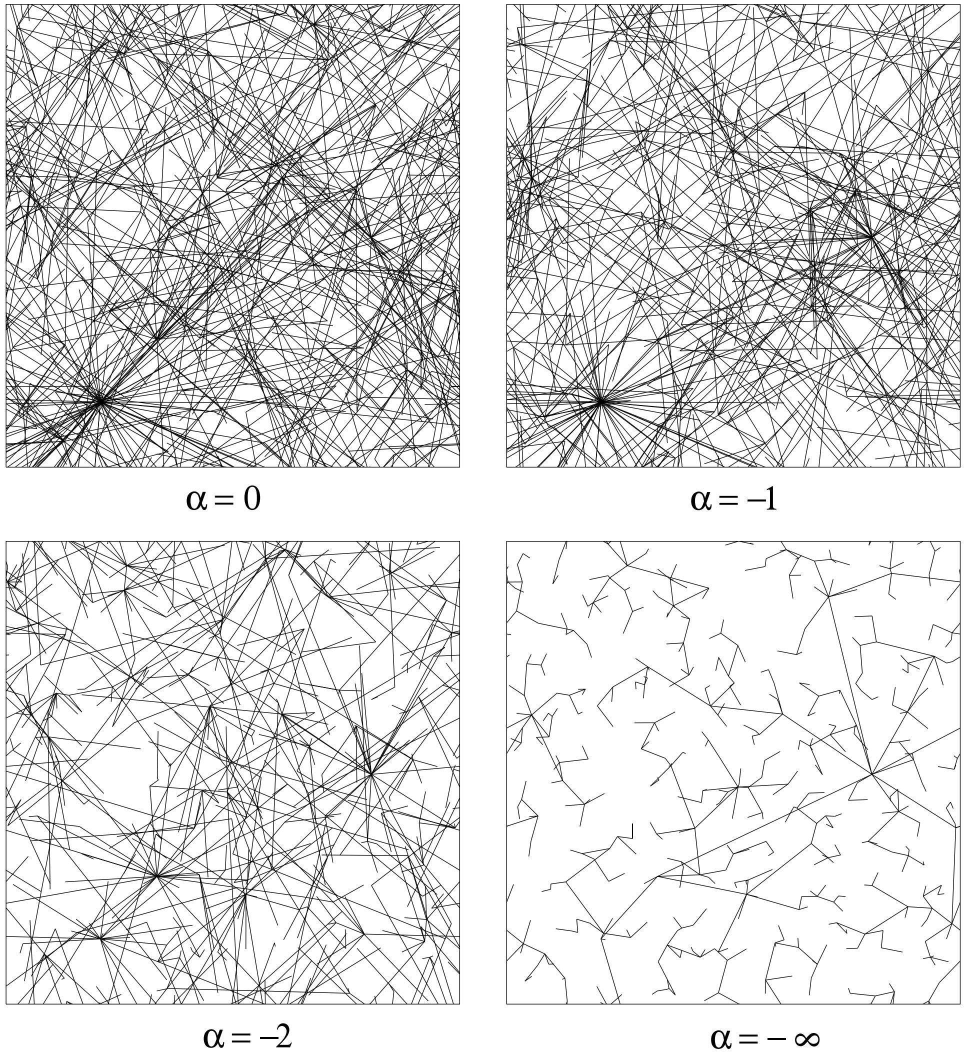
- control parameter is the α exponent
- if α is small: Small world network
- if α is large: Spatial network



Spatial preferential attachement models

General definition:

- Take a space and add successively N nodes at random positions until reaching density ρ
- Link nodes according to preferential attachment, but choosing targets according to both degrees and distance: $p(d, k) = kF(d)$



e.g., power law decaying vision

$$F(d) = d^{\alpha}$$

- control parameter is the α vision exponent
- Degree distribution
  - if $\alpha > -1$ — scale-free
  - if $\alpha < -1$ — non-scale free

The gravity

law

Formal description

Origin-destination matrix

- Describe flow of individuals between locations
- Used since decades by geographers
- Definition:
  - divide the area of interest into zones (cells) labelled by $i=1 \ldots N$
  - count the number of individuals going from location i to location j

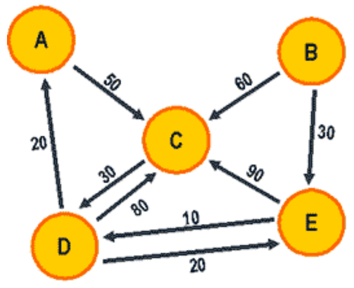
- directed
- weighted
- Beware:
  - time-dependent
  - strongly depends on the zone definition

O/D Matrix

$T(i,j)=$

	A	B	C	D	E	Ti
A	0	0	50	0	0	50
B	0	0	60	0	30	90
C	0	0	0	30	0	30
D	20	0	80	0	20	120
E	0	0	90	10	0	100
Tj	20	0	280	40	50	390

Spatial Interactions



- Difficult to obtain with conventional methods
- Mobile phones, GPS, geosocial apps, …

The gravity law

Number of trips from location i to location j is scaling as

$$T_{ij} = K \frac{P_i P_j}{d_{ij}^{\sigma}}$$

- where $d_{ij} = d_E(i,j)$ is the Euclidean distance between i and j
- $P_{i(j)}$ is the *population size* at location $i(j)$

Inter-city phone communication (Krings et.al.)

- mobile call communication intensity between Belgian cities

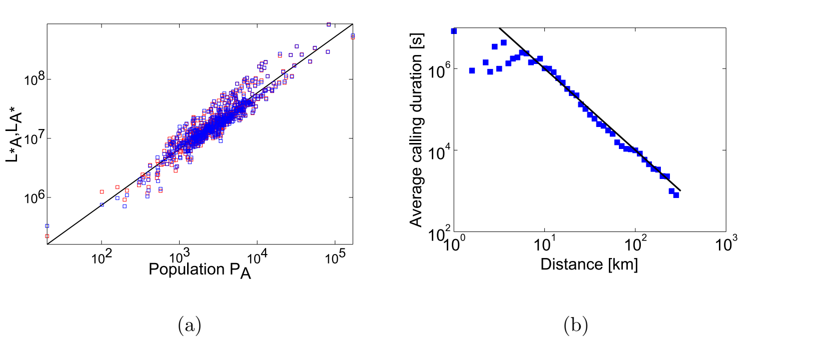
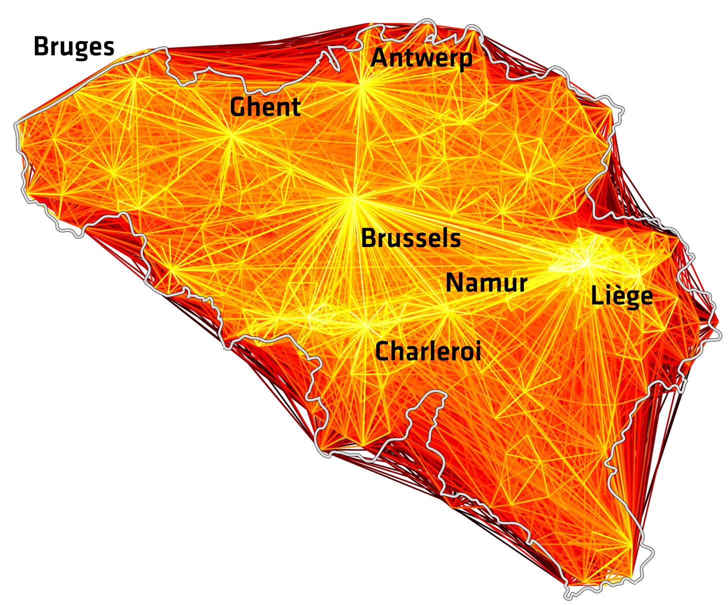
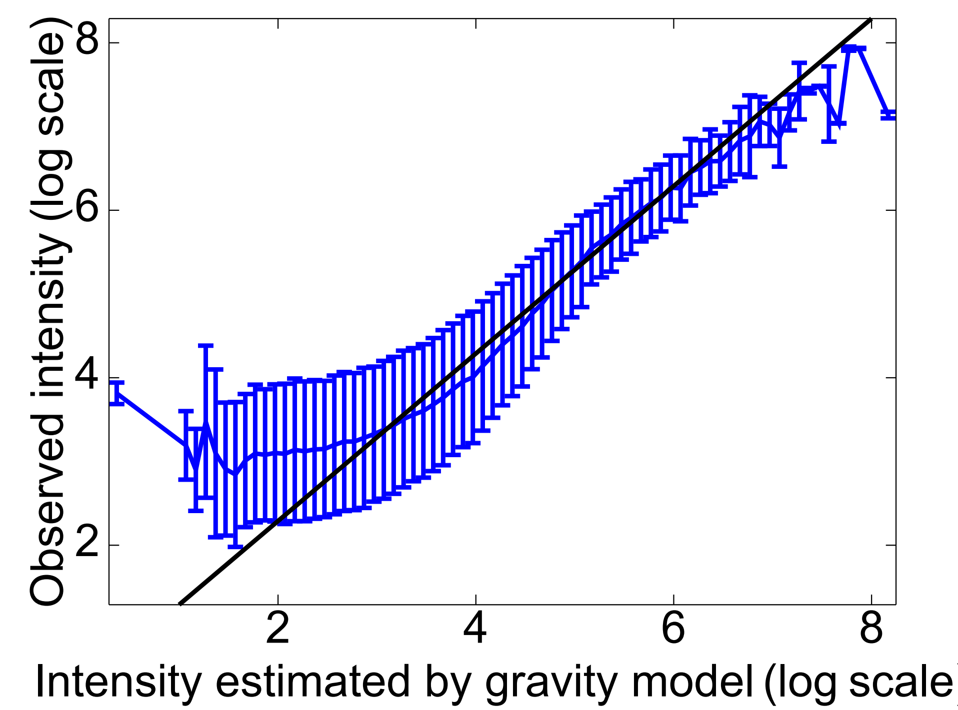


Figure 3. (a) Relation between the outgoing intensity L_{A*} (blue), the incoming intensity L_{*A} (red) and the city population size P_A. (b) Dependence of the average communication intensity between pairs of cities and the average distance separating them. The black line shows a $\frac{1}{d^2}$ decrease.





The gravity law

Number of trips from location i to location j is scaling as

$$T_{ij} = K\frac{P_i P_j}{d_{ij}^{\sigma}}$$

- where $d_{ij} = d_E(i,j)$ is the Euclidean distance between i and j
- $P_{i(j)}$ is the *population size* at location $i(j)$

- In a general form: $T_{ij} \sim P_i P_j f(d(i,j))$
  - where $f(d(i,j))$ is the deterrence function describing the effect of space

The gravity law - empirical summary

Both exponential and power-law dependence is observable

Network [Ref.]	N	Gravity law form	Results
Railway express [164]	13	P_iP_j/d_{ij}^{σ}	$\sigma = 1.0$
Korean highways [161]	238	P_iP_j/d_{ij}^{σ}	$\sigma = 2.0$
Global cargo ship [104]	951	$O_iI_jd_{ij}^{-\sigma}\exp(-d_{ij}/\kappa)$	$\sigma = 0.59$
Commuters (worldwide) [162]	n/a	$P_i^{\alpha}P_j^{\gamma}\exp(-d_{ij}/\kappa)$	$(\alpha, \gamma) = (0.46, 0.64)$ for $d < 300$ km
			$(\alpha, \gamma) = (0.35, 0.37)$ for $d > 300$ km
US commuters by county [163]	3109	$P_i^{\alpha}P_j^{\gamma}/d_{ij}^{\sigma}$	$(\alpha, \gamma, \sigma) = (0.30, 0.64, 3.05)$ for $d < 119$ km
			$(\alpha, \gamma, \sigma) = (0.24, 0.14, 0.29)$ for $d > 119$ km
Telecommunication flow [134]	571	$P_iP_jd_{ij}^{-\sigma}$	$\sigma = 2.0$

The gravity law - as a null model

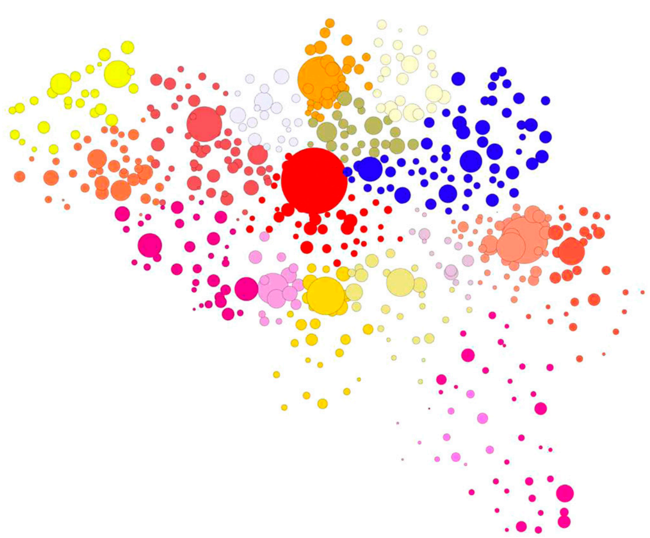
Usage as a network null model

- Consider a spatial network (e.g., phone calls, trips, etc.)
- Fit a gravity model best explaining the observed network. If the population is unknown or not relevant, the degrees of nodes (in/out degrees in directed networks) can be used as a "*population*"
- =>Random model with a given edge probability for each pair of node
- The obtained network is a null model to which the observed network can be compared

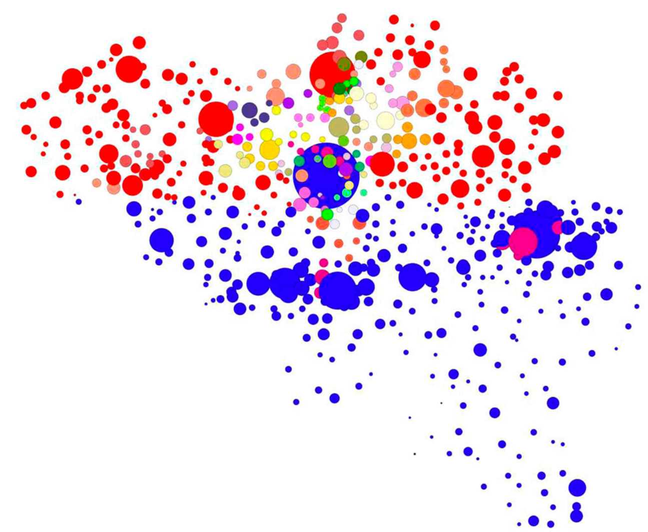
The gravity law - as a null model

Example of application: Space-independent communities

- In the usual modularity, the fraction of internal edges is compared between the observed network and a configuration model.
- One can replace the configuration model by a gravity model



Space-dependent communities



Space-independent communities

Expert, P., Evans, T. S., Blondel, V. D., & Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, *108*(19), 7663-7668.

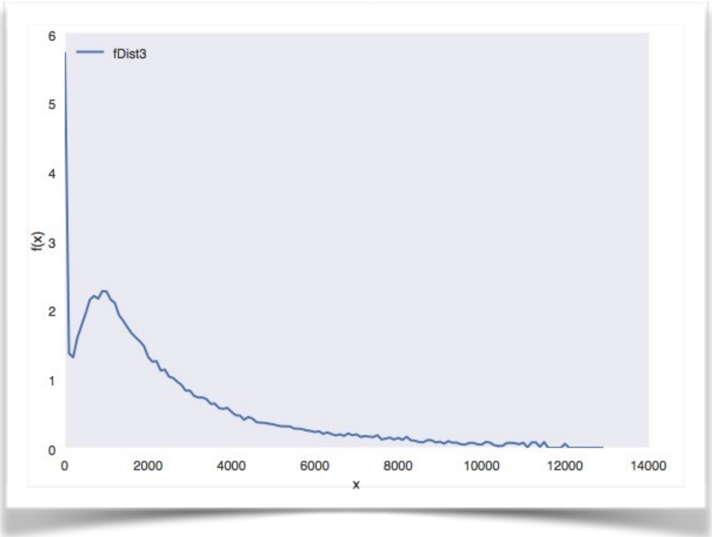
The gravity law - flavors

Agnostic deterrence function

- The influence of distance might be more complex than a power-law or an exponential. In particular, it is often non-monotonic (first increasing, then decreasing. Think of airplanes, bicyles, public transports… unlikely to use for short distances)
- A deterrence function can be learned from data
- Computed by comparing the number of trips observed at a given distance with the number of trip expected if distance has no effect (a configuration model)

$$f(d) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} k_i k_j}$$

$f(d)$

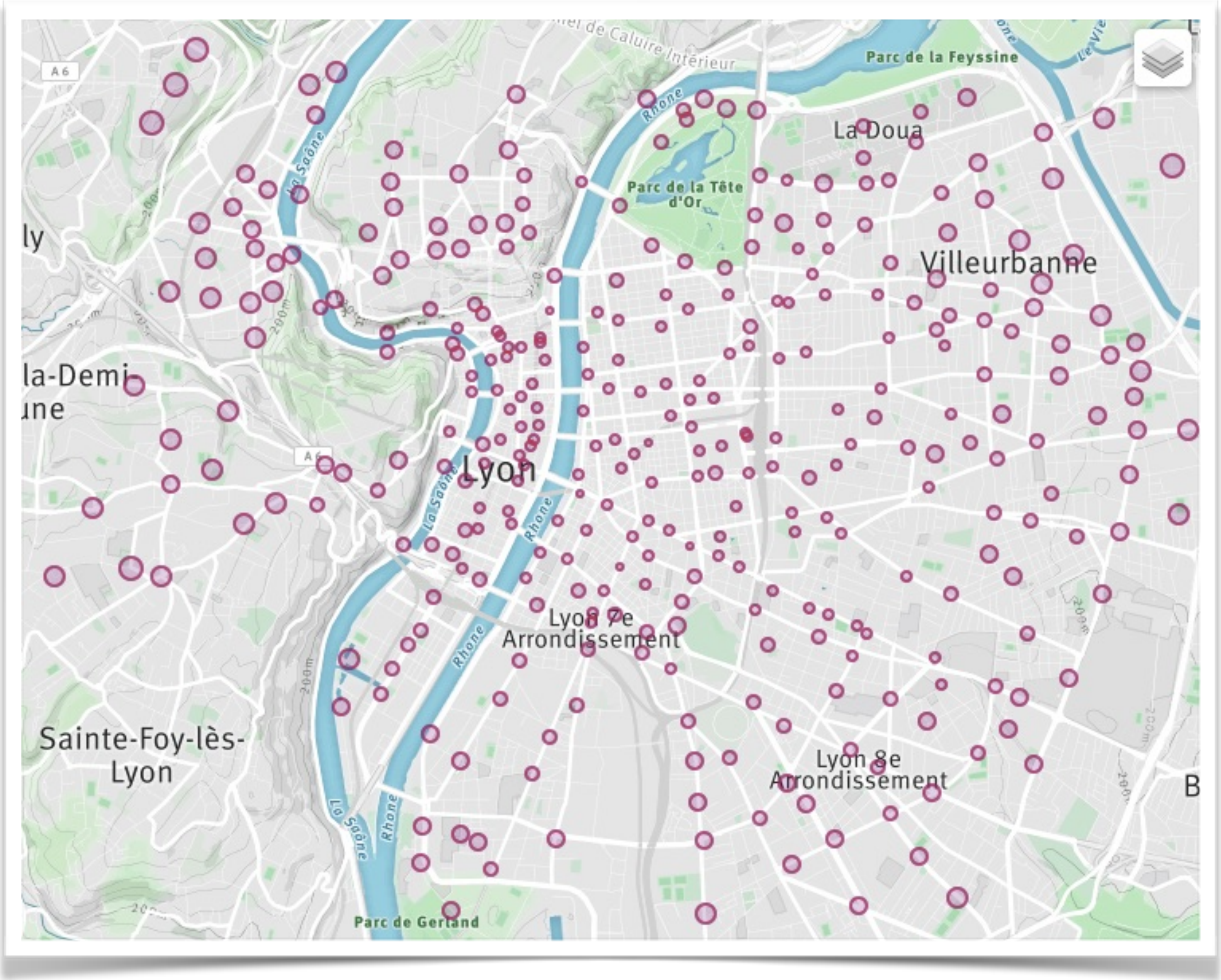


Distance d

Degree corrected gravity model

- The standard gravity model, even fitted to data, do not conserve degrees
- Solution: Methods to uncover the *population* of nodes that would best explain *observed degrees*

The gravity law - example

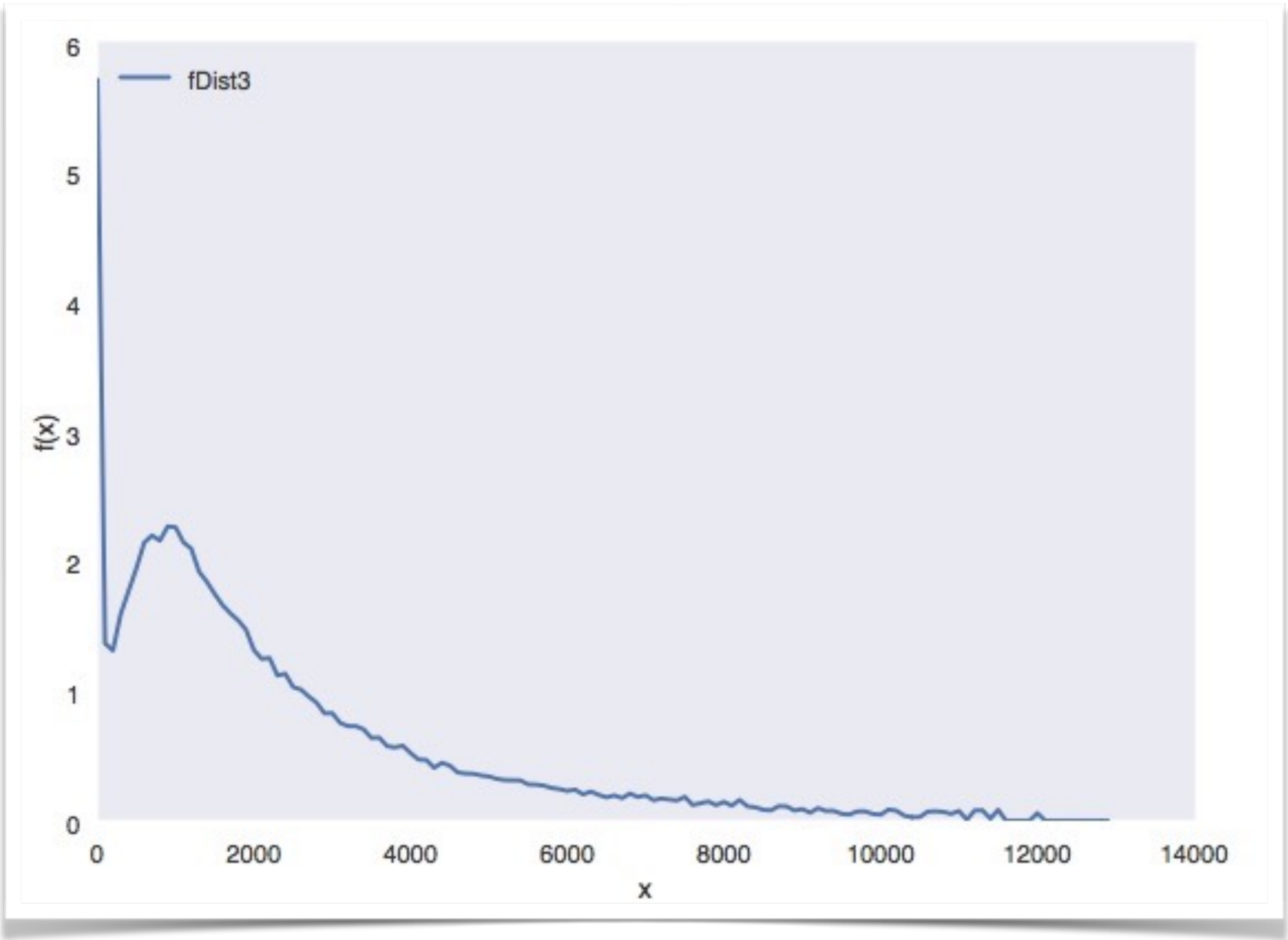


Nodes: station (2D position)

Edges: number of trips over a period

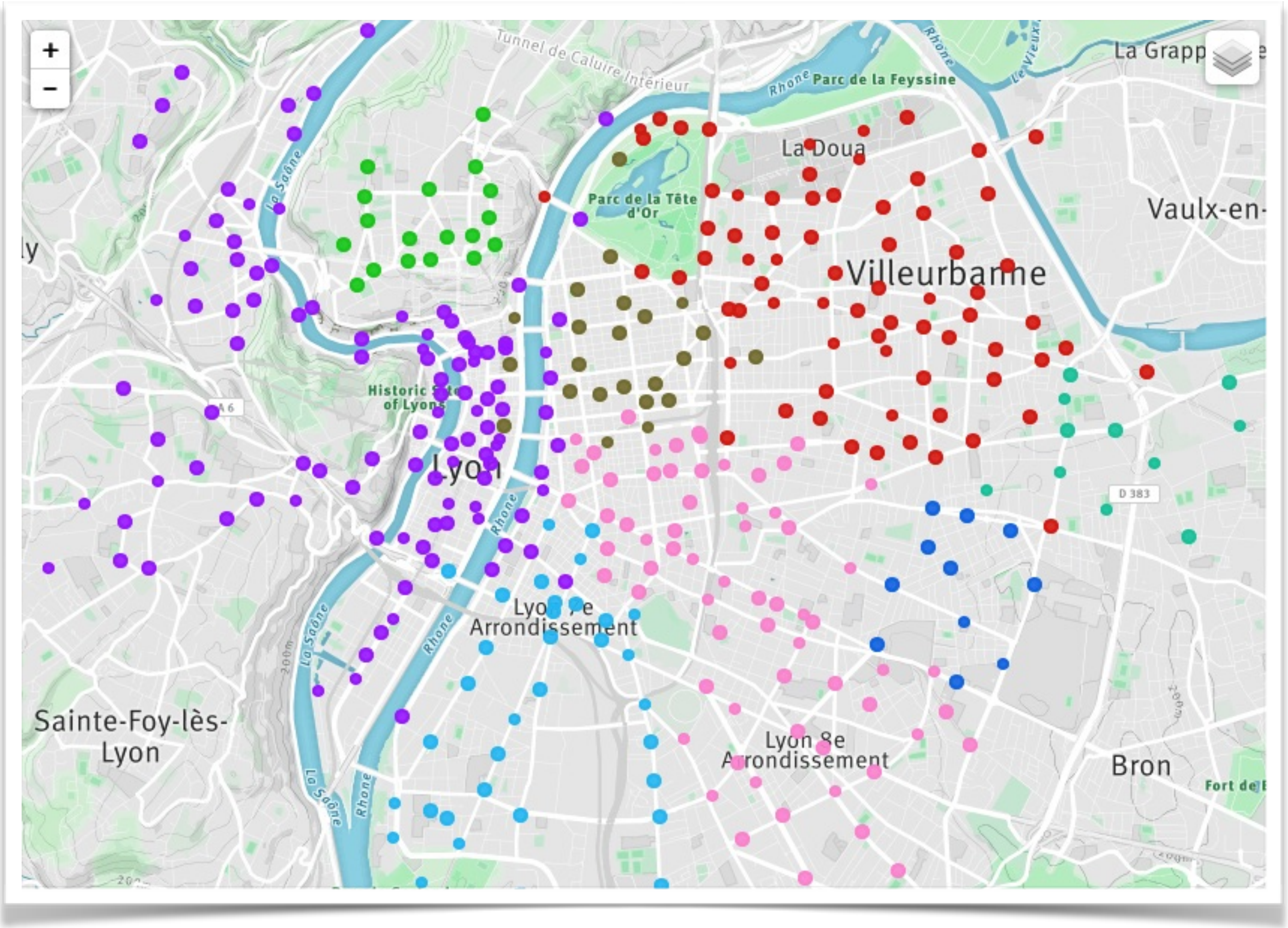
The gravity law - example

$f(d)$



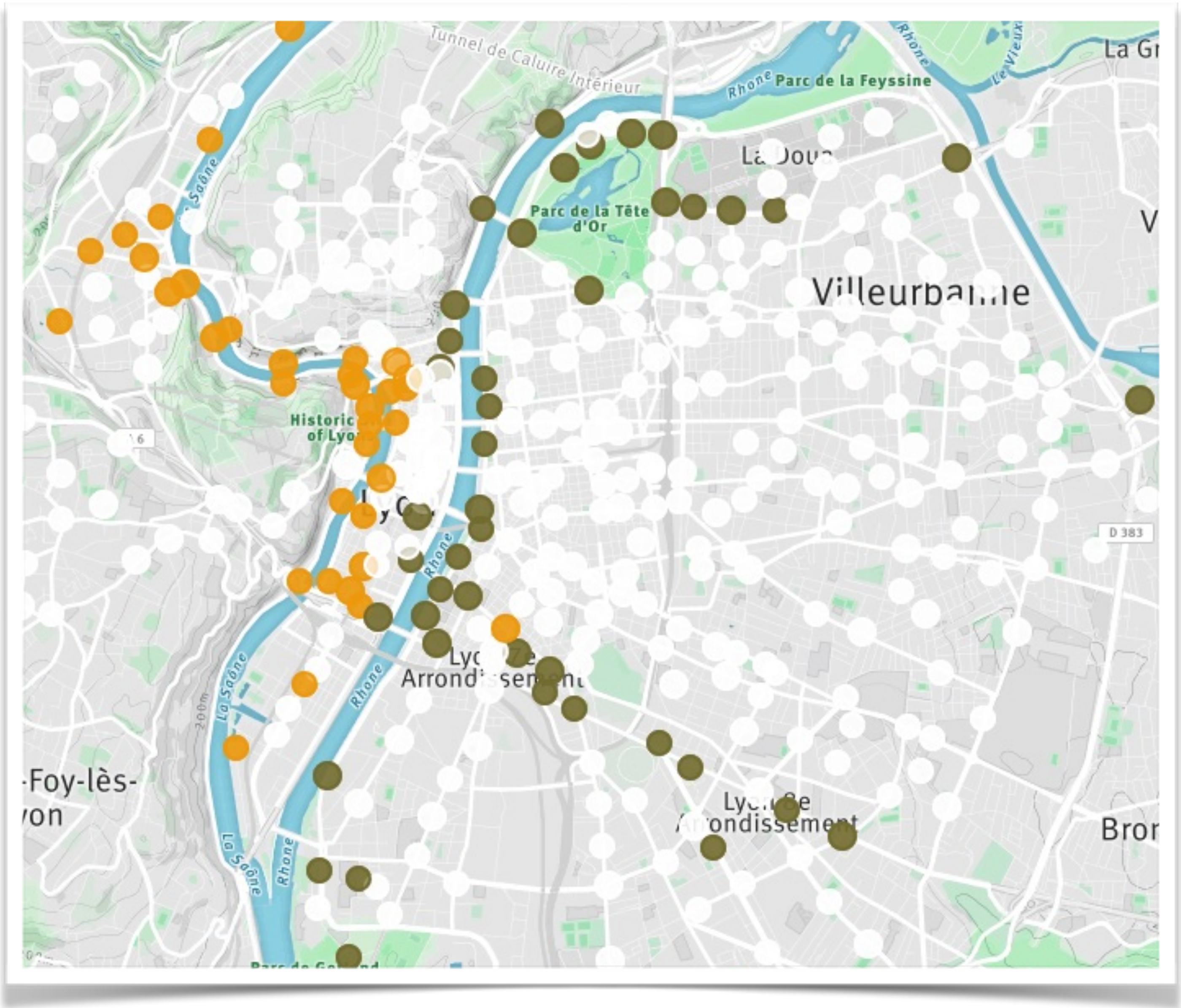
Distance *d* *(meters)*

The gravity law - example



Space-dependent communities

The gravity law - example



Some (social) space-independent communities that were previously *hidden* by spatial constraints

The radiation law

The radiation law

Limitations of the gravity law

1. Requires previous data to fit

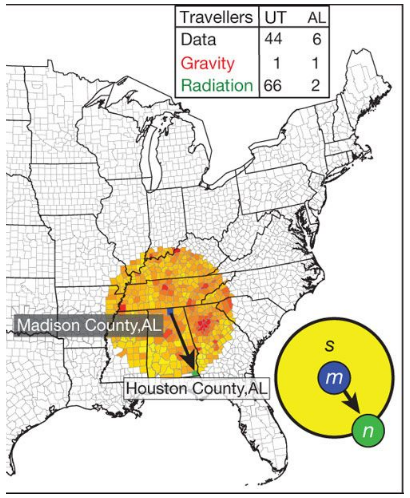
2. The number of travelers between destinations depends only on their populations and distances. In reality, this value depends probably of other *opportunities*

The radiation law

Intuition: Model how people move for jobs

1. Individuals look for job in all cities
2. Each city has a number of job opportunities
    - Each job has a value of *interest*, considered random
3. What is the probability for a job-seeker to choose a job in city *c* located at distance *d?*
    - Depends only on how many jobs offered in cities at a distance equal or lower than *d* (probability to find a better job closer)

The model is parameter-free!

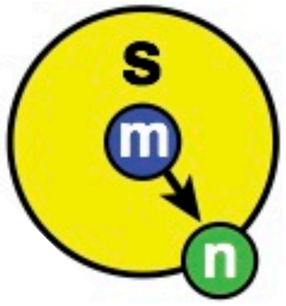


The radiation law

The model can be formulated in terms of radiation and absorption

- take locations i and j with populations (in-degree) m_i and n_j and at distance r_{ij}
- denote s_{ij} the total population in the circle with radius r_{ij} centered at i (excluding the source and destination population)
- T_i is the number of commuters (out-degree)

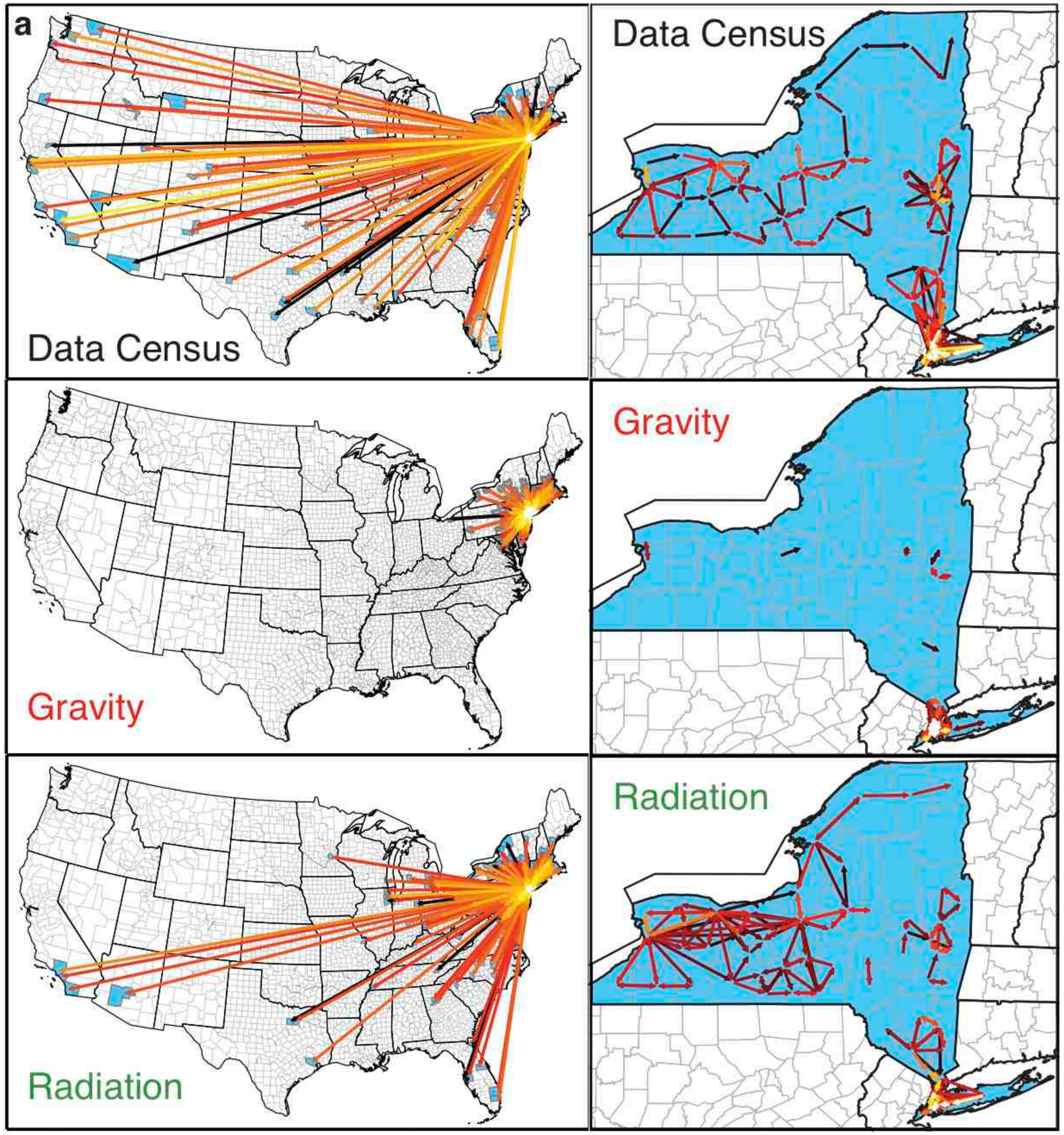
$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}.$$

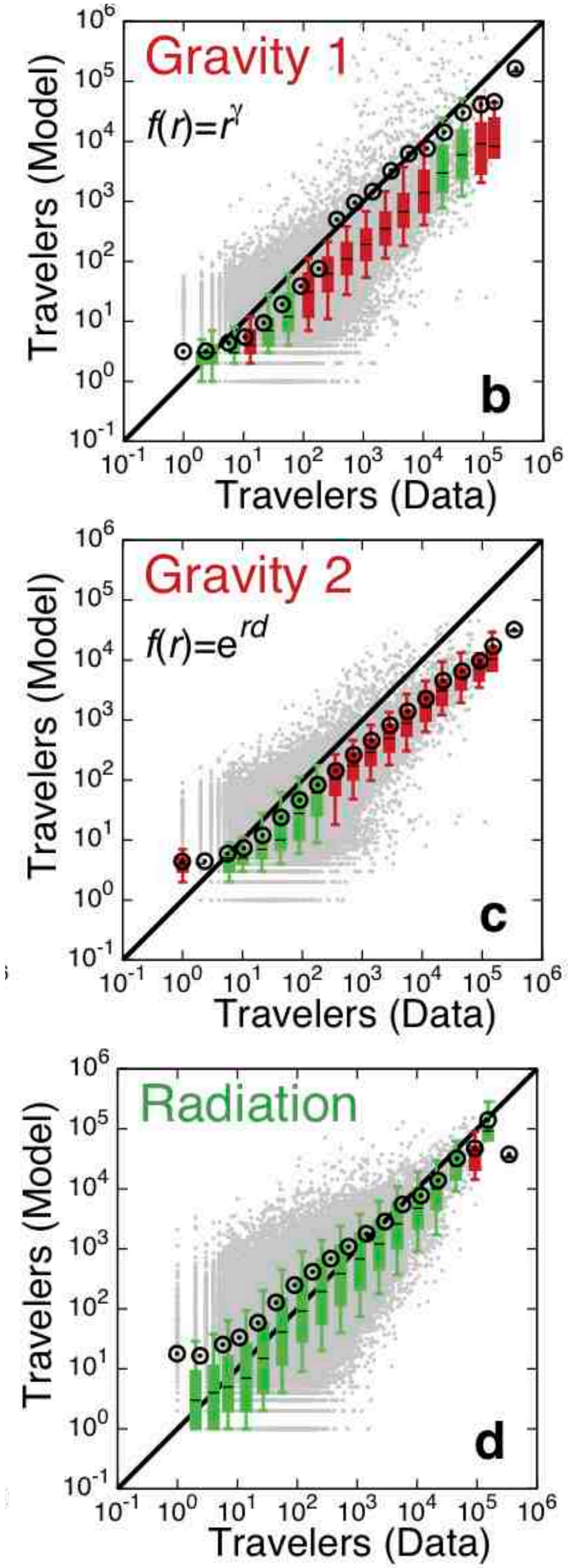


The radiation law

Comparison with census data and the gravity law predictions

Simini. et.al, Nature 2010





The radiation law

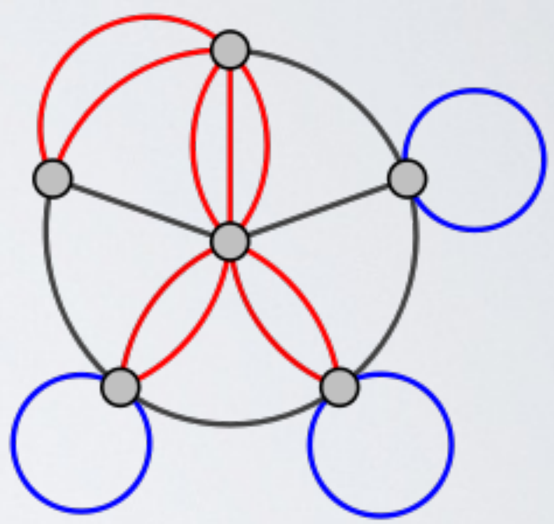
Beware

- The real relevance of the radiation law in general cases is disputed. Parameter free is an advantage, but also a strong drawback in many situations… The real world is complex !

MULTI-GRAPHS

MULTIGRAPH

- Multi-graph: several edges allowed between same nodes
- Often used in conjunction with labels:
  - One edge for friendship,
  - one edge for family,
  - one edge for co-worker…



- Without labels, can be simplified as a weighted graph
- If several labels, can be considered as separate graphs for analysis

MULTI-PARTITE GRAPHS

MULTI-PARTITE GRAPHS

- Bi-partite: there exists 2 kinds of nodes, and links can be only between nodes of different types
  - Multi-partite: similar but with more than 2 types. Much less common
- Not strictly different from normal graphs: if you don't know the two categories of nodes, it looks like any network
- The problem is that some definitions of normal graphs become meaningless
  - =>Clustering coefficient

MULTI-PARTITE GRAPHS

- Bi-partite networks are quite commonly use
  - Actors - Films
  - Clients - Products
  - Resercher**s** - conferences/institutions
  - …
- Normal methods work but sometimes give unintuitive results:
  - Specific variants have been proposed

MULTI-PARTITE GRAPHS

Modularity: do not count pairs of nodes of same types

$$Q_B = \frac{1}{m}\sum_{u=1}^{r}\sum_{v=1}^{c}(\tilde{A}_{uv} - P_{uv})\delta(g_u, h_v) = \frac{1}{m}\sum_{u=1}^{r}\sum_{v=1}^{c}\left(\tilde{A}_{uv} - \frac{k_u d_v}{m}\right)\delta(g_u, h_v),$$

MULTI-PARTITE GRAPHS

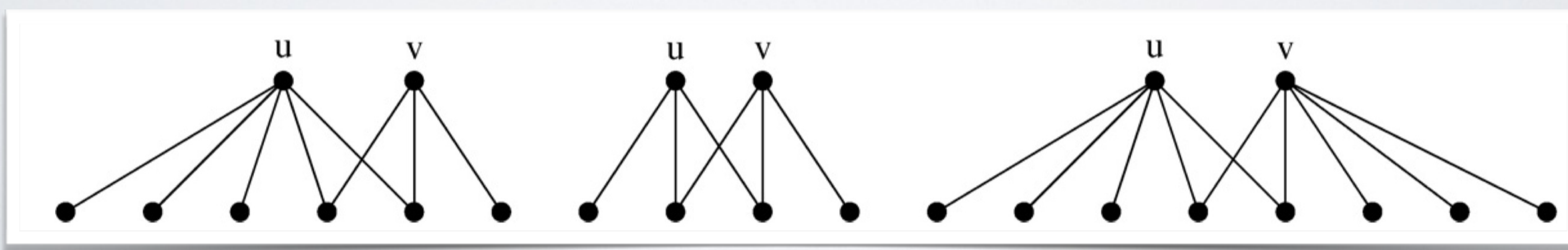
Clustering Coefficient:

Of a pair

$$\mathrm{cc}_{\bullet}(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

Of a Node: Average among nodes N at distance 2

$$\mathrm{cc}_{\bullet}(u) = \frac{\sum_{v \in N(N(u))} \mathrm{cc}_{\bullet}(u, v)}{|N(N(u))|}$$



2/6 2/4 2/8

MULTI-PARTITE GRAPHS

- Large literature on the topic, in particular applications to **recommendation**
  - Users - products => propose the right products to the right user

Kunegis, J., De Luca, E. W., & Albayrak, S. (2010, June). The link prediction problem in bipartite networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems* (pp. 380-389). Springer, Berlin, Heidelberg.

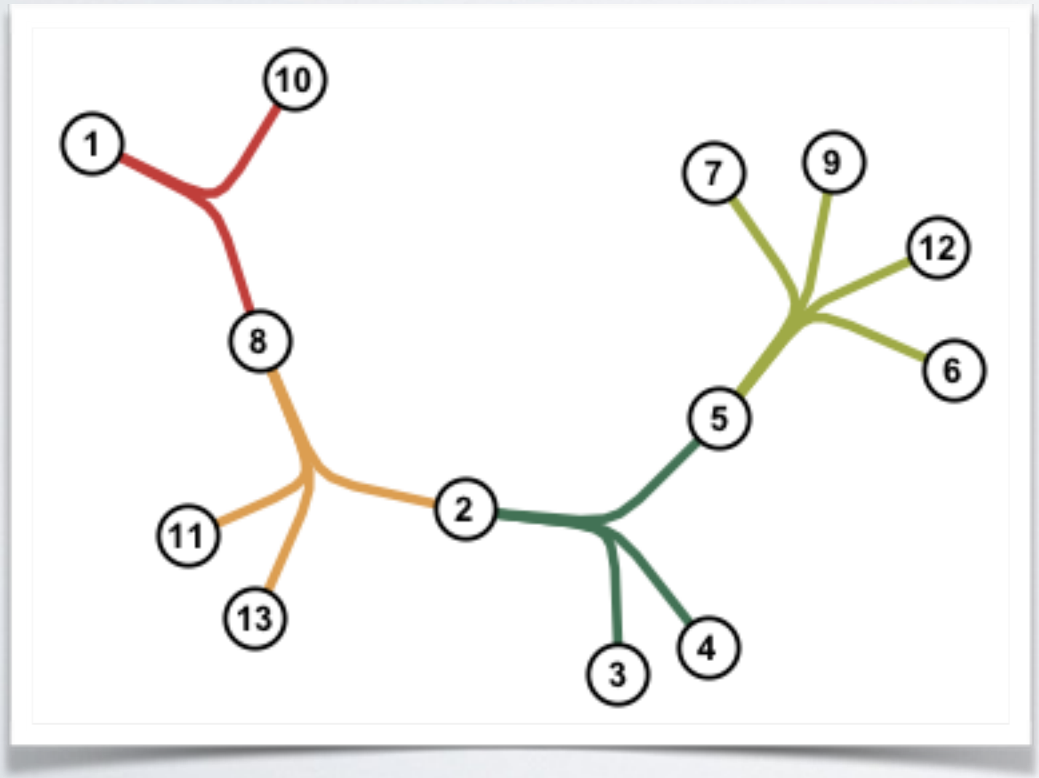
Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, *76*(6), 066102.

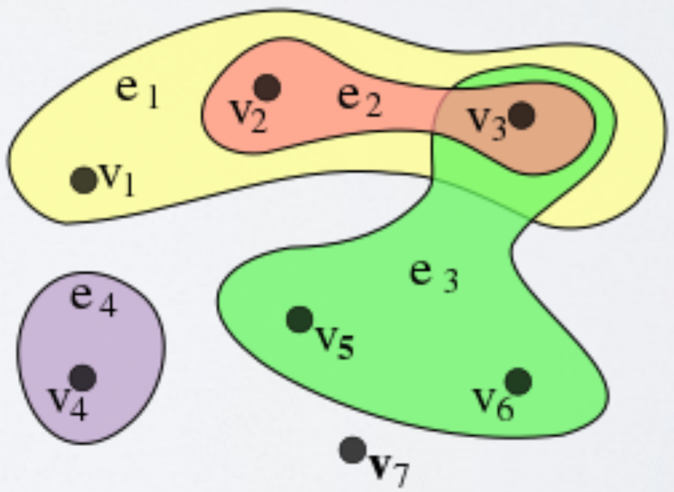
Zhang, P., Wang, J., Li, X., Li, M., Di, Z., & Fan, Y. (2008). Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications*, *387*(27), 6869-6875.

HYPERGRAPHS

HYPERGRAPHS

- "Generalization" of graph
- An edge is not limited to 2 extremities





HYPERGRAPHS

- Most common usage: represent a single event involving several nodes

- In social networks: 10 students attending a same course A
  - Normal network: 45 undirected edges. Giant clique. Very dense
    - Problem: if 5 attend another course B and the others another course C => no way to see who worked with whom (a single clique, with double links or weights=2)
  - Hypergraph: A single link with ten endpoints
    - And we can add 2 single links with 5 endpoints and still differentiate attendances

- Another example: in Bitcoin, transactions are multi-input, multi-output. Some transactions have 1000 input, 1000 output
  - 500 000 links for a single transaction in a normal network!

HYPERGRAPHS

- In practice, very few direct usages
  - Too difficult to handle ? To different from normal networks?
- Hypergraphs can be transformed in bi-partite graphs
  - Social Network: student nodes and class nodes
  - Bitcoin: transaction nodes and address nodes

MULTILAYER NETWORKS

MULTILAYER NETWORKS

- Multiplex network
- Multislice network
- Multitype network
- Heterogenous information network

[Kivela 2014]

MULTILAYER NETWORKS

- Can be used to represent:
  - Several types of relations between the same nodes
    - Bus transportation network
    - Bicycle transportation network
    - Car transportation network
    - …

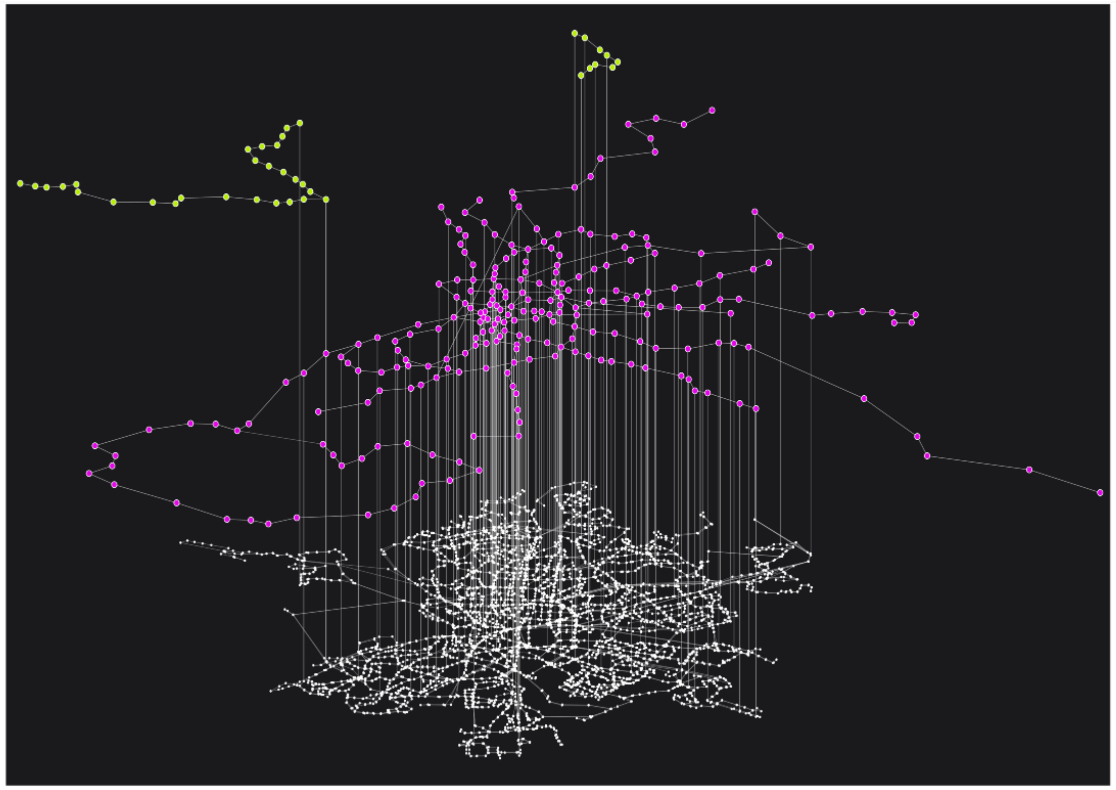
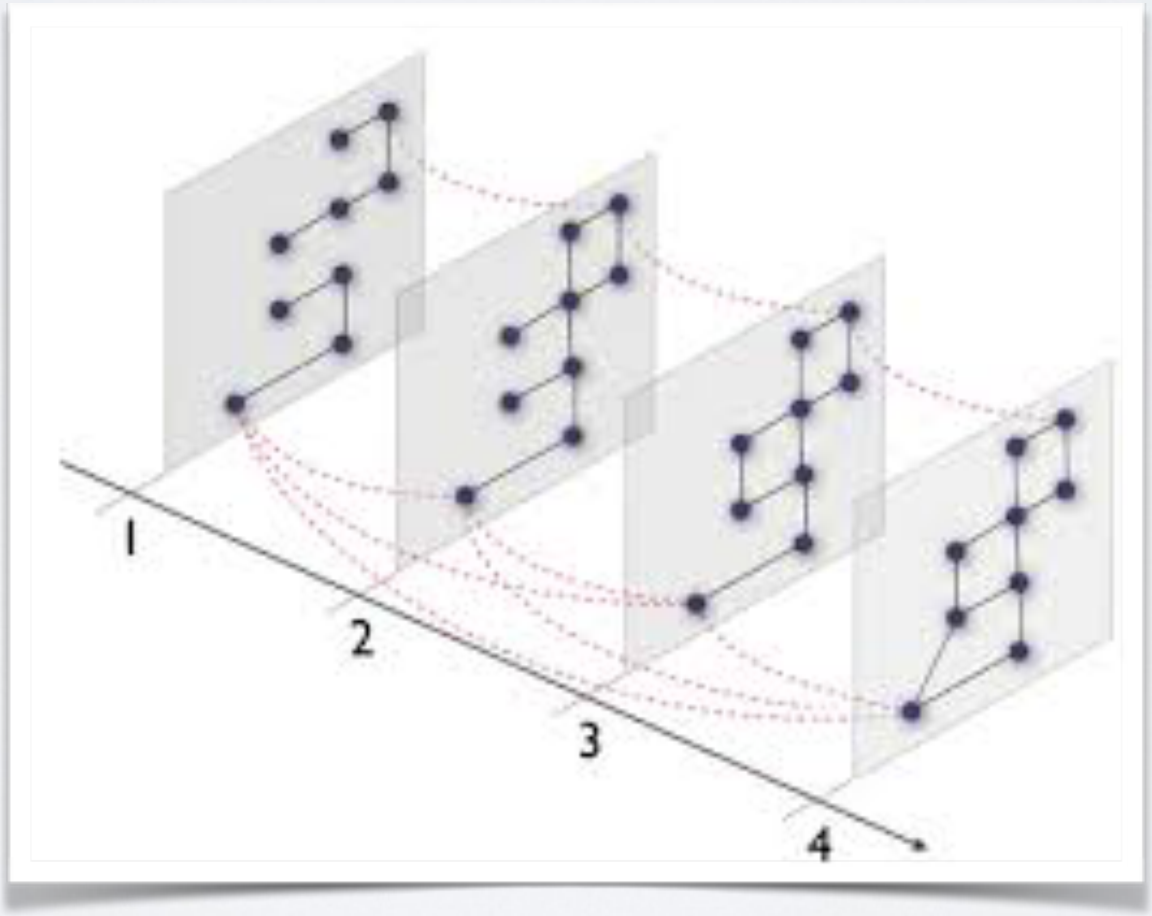


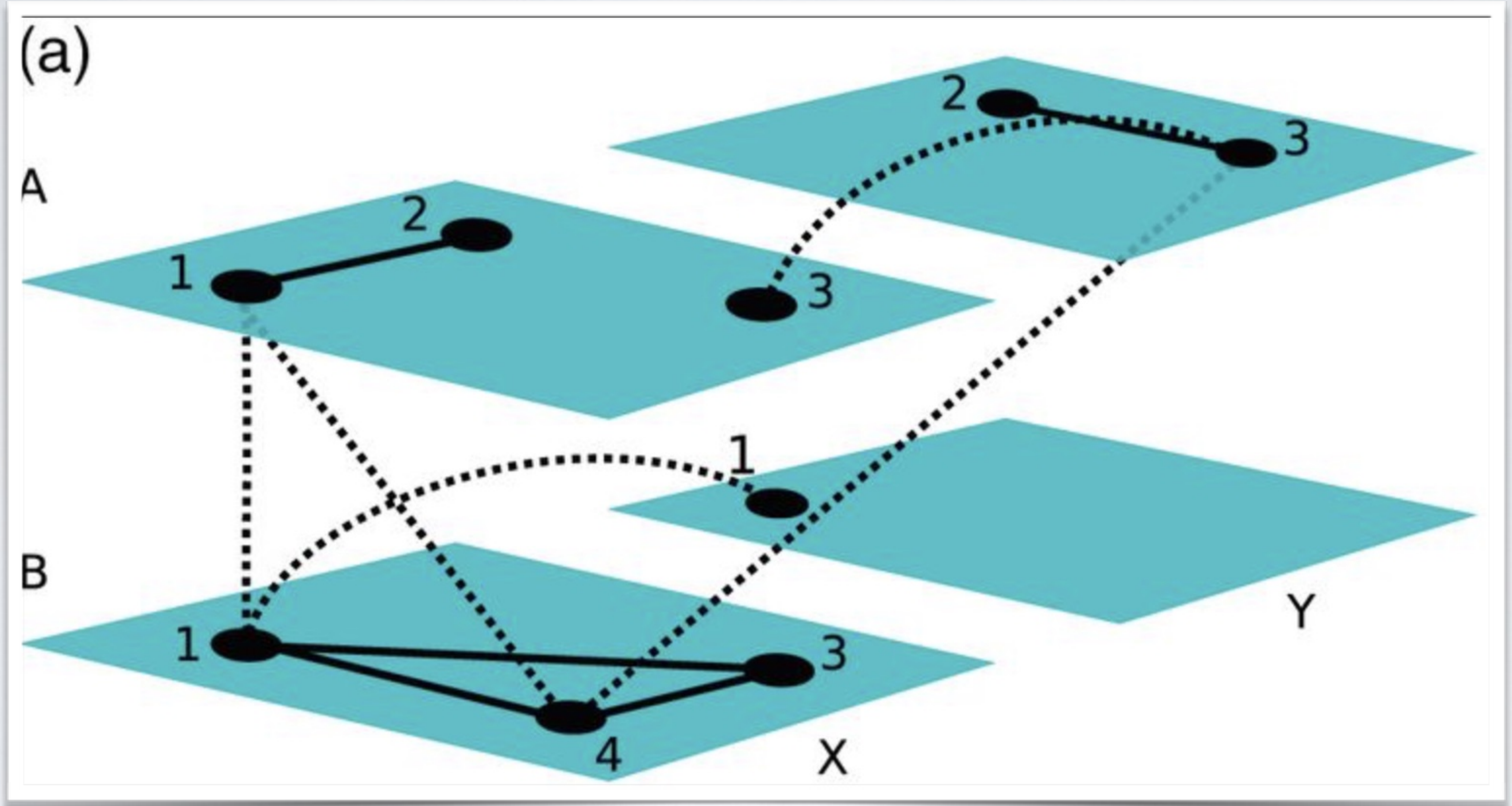
Figure 2. Superlayer representation of the Madrid transportation system. The figure represents the three transportation modes considered: tram (yellow nodes, upper layer), metro (purple nodes, mid layer) and buses (white nodes, bottom layer). See Table1 for statistics of these layers.

MULTILAYER NETWORKS

- Can be used to represent:
  - Several snapshots of the same network



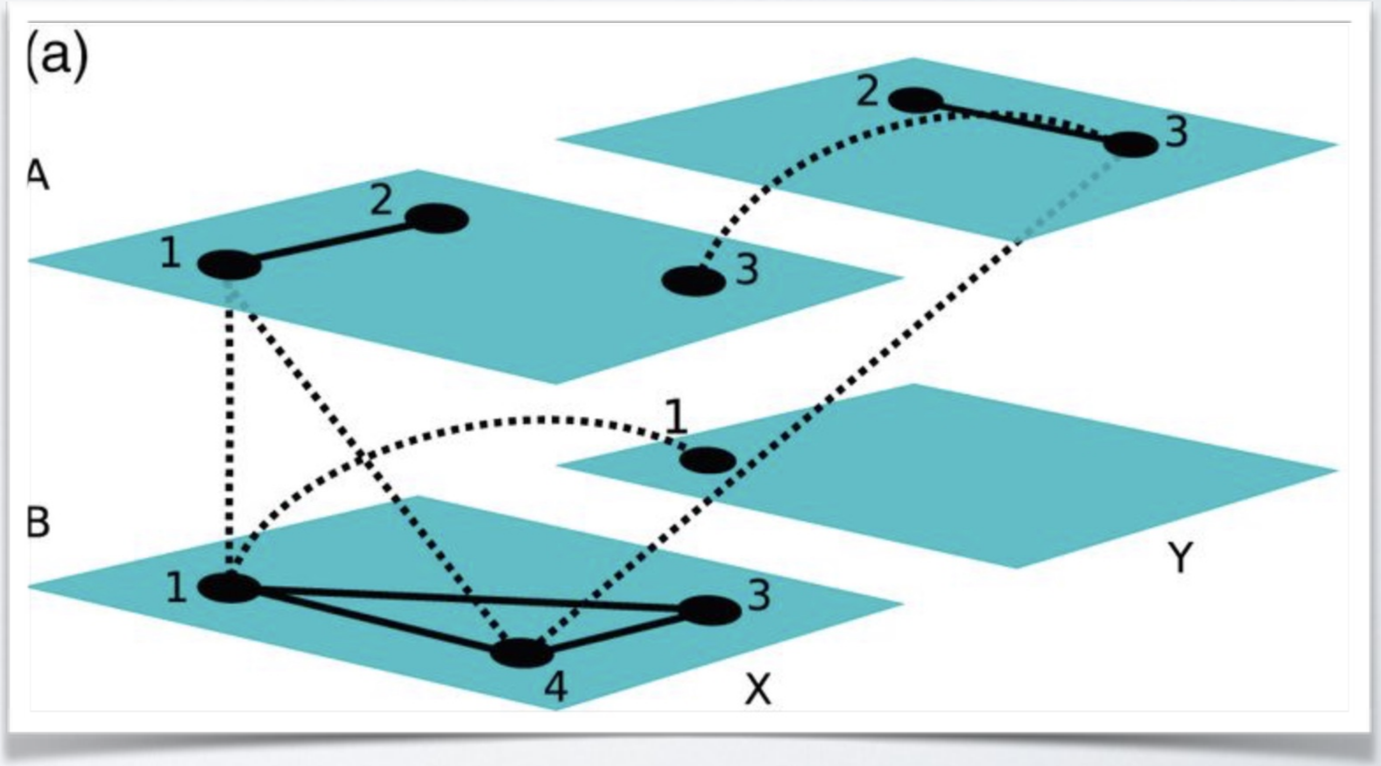
MULTILAYER NETWORKS



Both/Other

MULTILAYER NETWORKS

- Relations can be:
  - Only between **same** nodes in different layers
    - Public transport interconnection
  - Between different nodes in different layers
    - Information transfert form person A on Facebook to person B on Instagram.

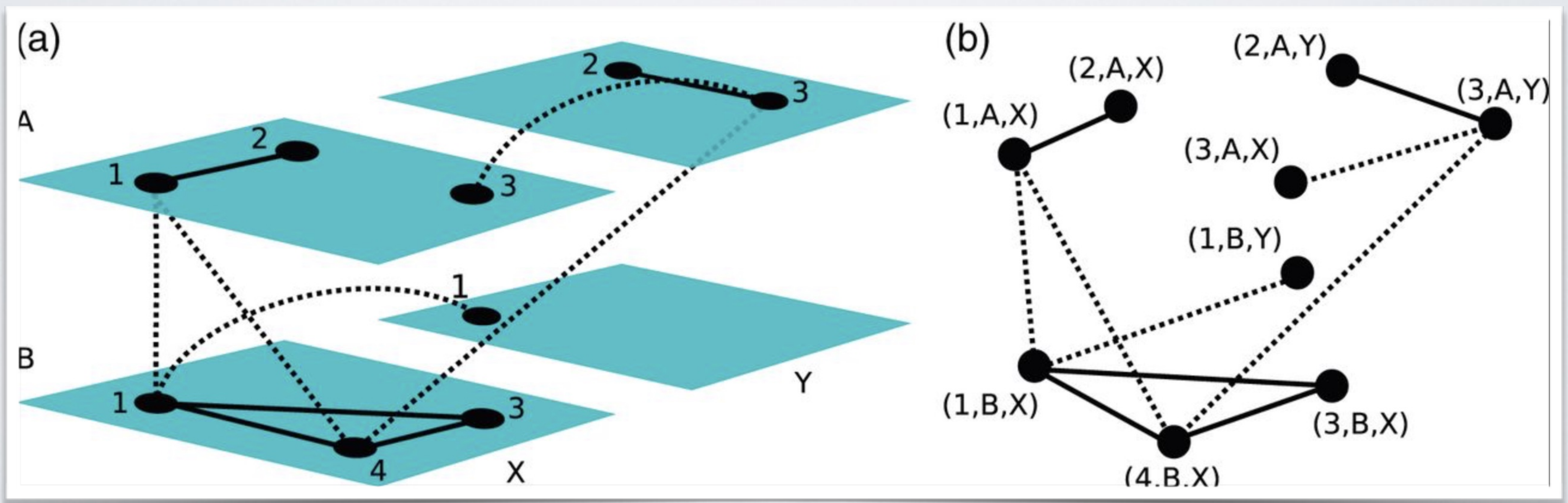


MULTILAYER NETWORKS

- All usual definitions on static networks can be extended to multilayer networks
  - Degree, clustering coefficient, community detection…
- The problem is that there are many ways to do it, and it depends on what your layers represent
  - Degree of a person on a multilayer network of facebook, Twitter, Linked-in?
- If you used a multilayer network, it is because it was not well summarized by a single network…
  - Same definition for multilayer dynamic and multilayer different types?

MULTILAYER NETWORKS

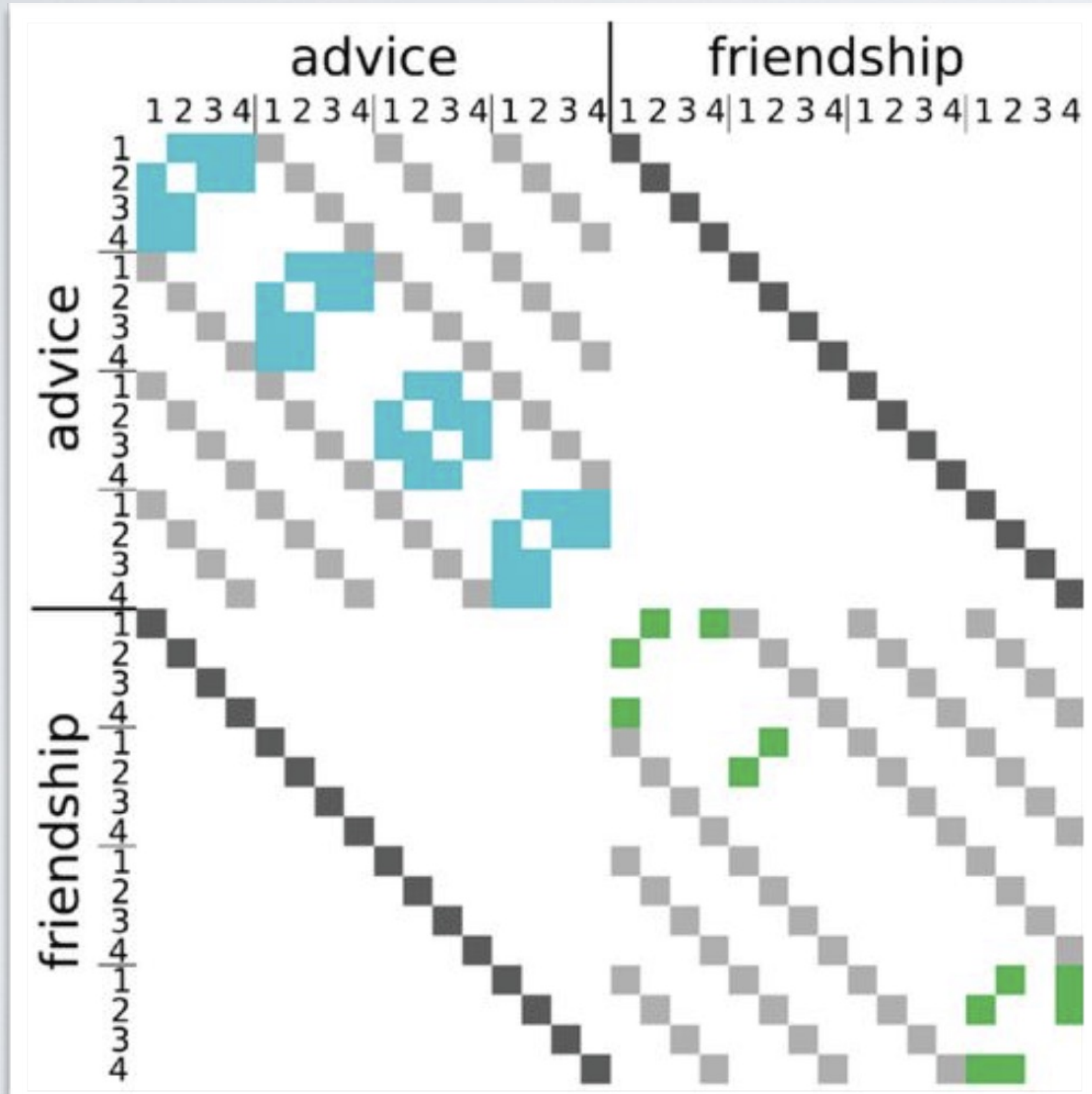
A simple idea: multilayers networks can be transformed into simple networks



MULTILAYER NETWORKS

- Matrix representation:
  - Many algorithms on networks work on adjacency matrices
- Solution 1: Tensors
  - Be careful if not all nodes in all layers!
  - Interesting when only links between same nodes in <> layers
- Solution 2: Supra-adjacency matrix
  - Or flattened tensors

MULTILAYER NETWORKS



Blue, green: intra-layer

gray: inter-layer 1

black: inter-layer2

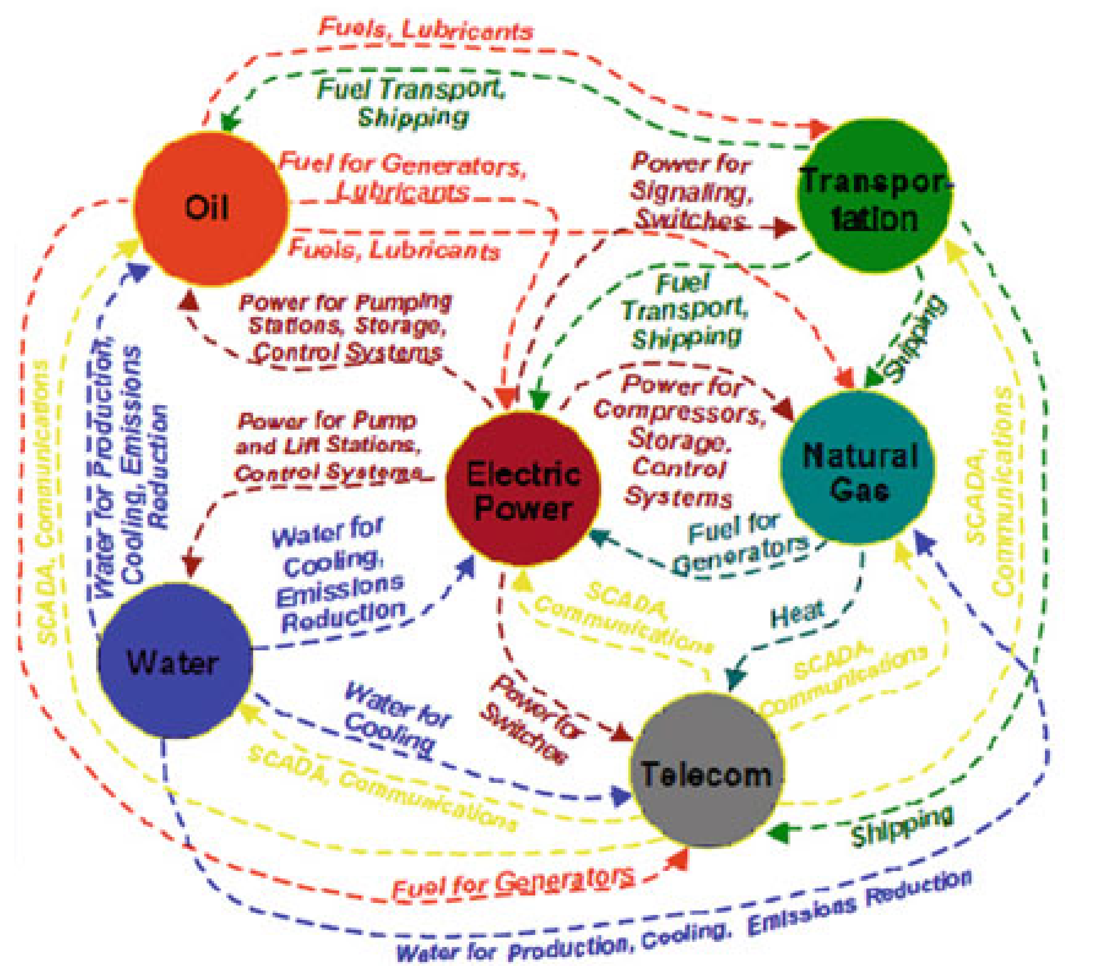
Cognitive map: relations between
4 people seen by each of these 4 people

Interdependent networks

Example 1: Infrastructure networks

Interdependent infrastructure networks

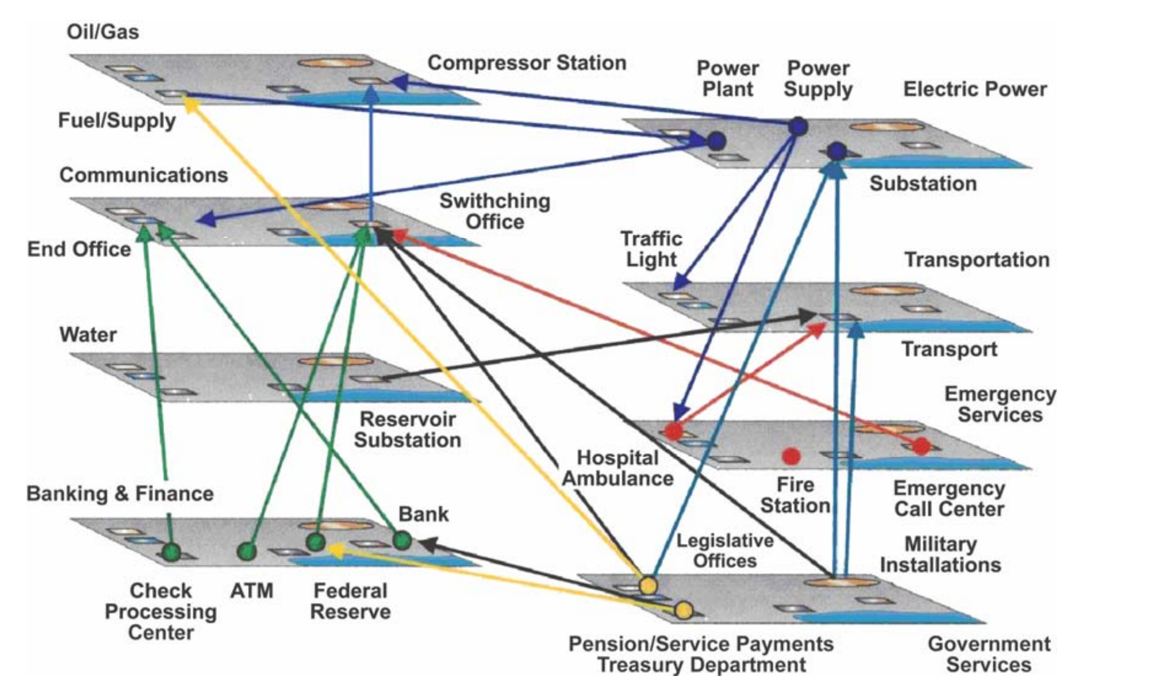
- Power-grid networks
- Communication networks
- railway networks
- Water supply
- Gas supply
- Transportation and fuel



Peerenboom, Fisher, and Whitfield, 2001

Motivation

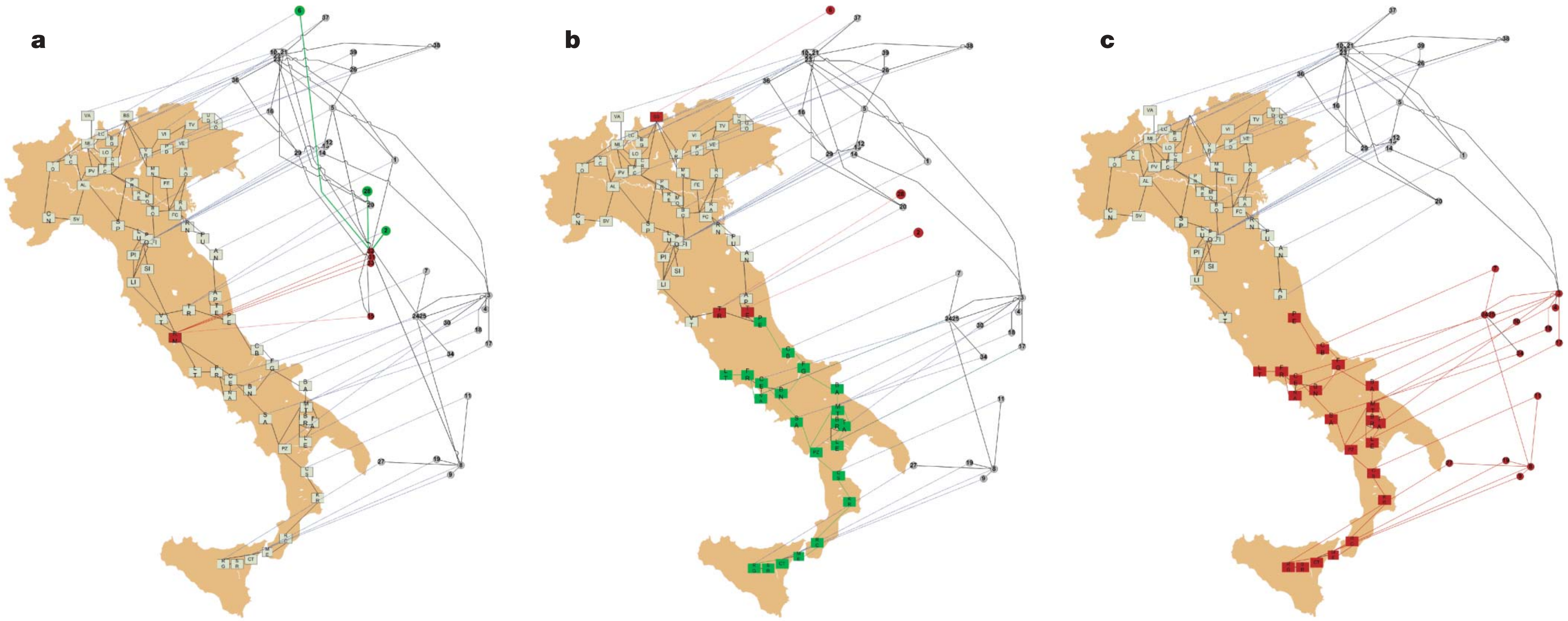
- To understand correlated failure
- To assess risk of interdependency
- To design robust interdependent networks against attack and random failure



Example 1: Infrastructure networks

Example: 2003 Italy blackout

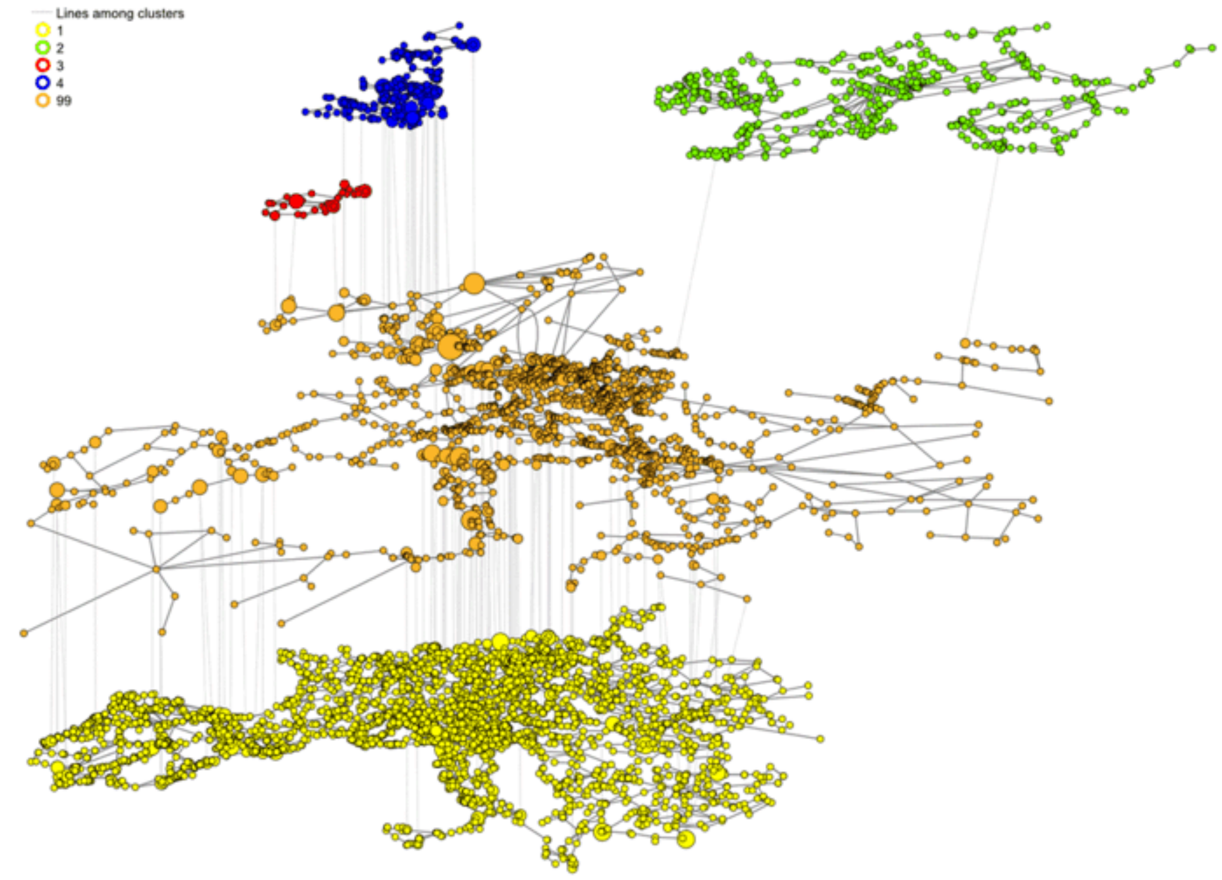
- A power line between I and CH was damaged by storm
- Power outage for 12 hours in Italy and spread to Switzerland for 3 hours
- 56 millions of people without electricity
- 110 trains cancelled
- All flights were cancelled
- People stuck and sleeping in the metro



Interdependent networks

Network structure

- co-existing networks with
  - links exist between nodes of the same network
  - links between networks assign the dependency between nodes in different layers
- The identities of nodes are not necessarily the same in different layers



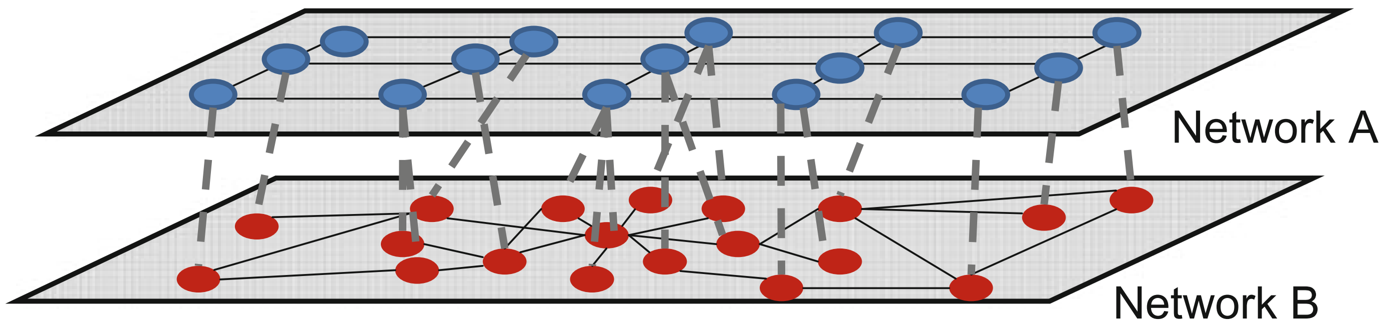
Interdependent networks - definition

Take two networks A and B

- Nodes in A depend on one or more nodes in B and vice versa

Types of links

- **Connectivity links**
  - Connect nodes from the same layer (A or B)
  - They allow the information to spread between nodes of the same layer
- **Dependency links**
  - Connect nodes of different layers (between A and B)
  - Assumption: for a node to function in one layer needs support from another node from another layer
- **Direction of dependency links**
  - $A_i \rightarrow B_j$
  - *i* node in *A* provides a critical resource to *j* node in *B*
  - If node *i* fails, the supported node *j* fails as well



HIGHER ORDER NETWORKS (HON)

HIGHER ORDER NETWORKS

- Recent and very active field of research
- Many networks are built using logs of *sequence of items* encountered by *actors*
  - People travelling in public transport go through stations
  - Consumer buy products on amazon one after the other
- Normal network: split sequences in pairs
  - Higher order: conserve the memory of previous items
  - From Markovian to non-Markovian

HIGHER ORDER NETWORKS

- Typical example: air traffic.
- Many cities does not have direct trips
  - e.g.: Lyon->Tokyo
  - Flight goes through stopover: Lyon->Paris->Tokyo
- If we want to create a weighted network of trips:
  - We count the number of trips Lyon->Paris, and Paris->Tokyo
  - We forget the information of previous/Next step

HIGHER ORDER NETWORKS

- This transformation to a Markovian process induce an important information loss

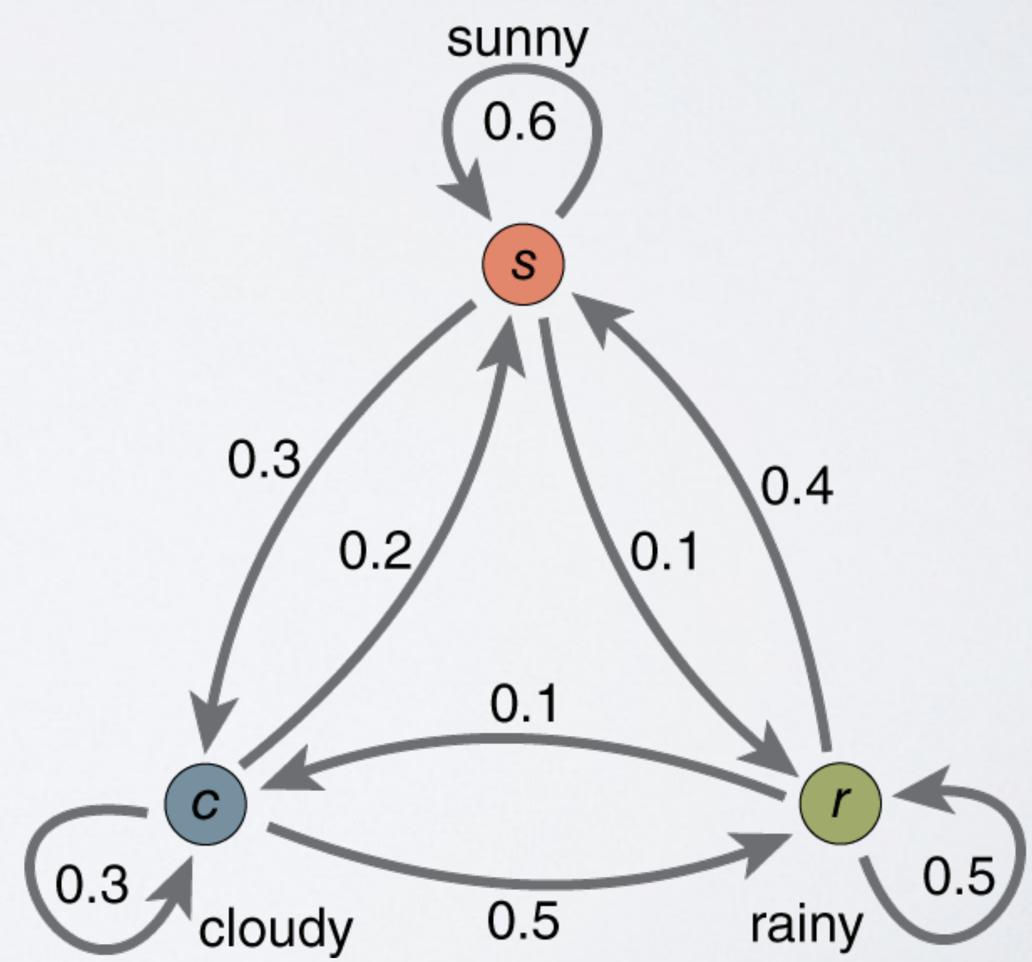
- From Paris,
  - 1% passengers go to Tokyo
  - 1% passengers go to Geneva

- You know that a passenger leaving Paris comes from Lyon
  - 1% probability to go to Tokyo
  - 1% probability to go to Geneva
  - =>Wrong!
  - If the passenger comes from lyon, much more likely to go to Tokyo than to Geneva!

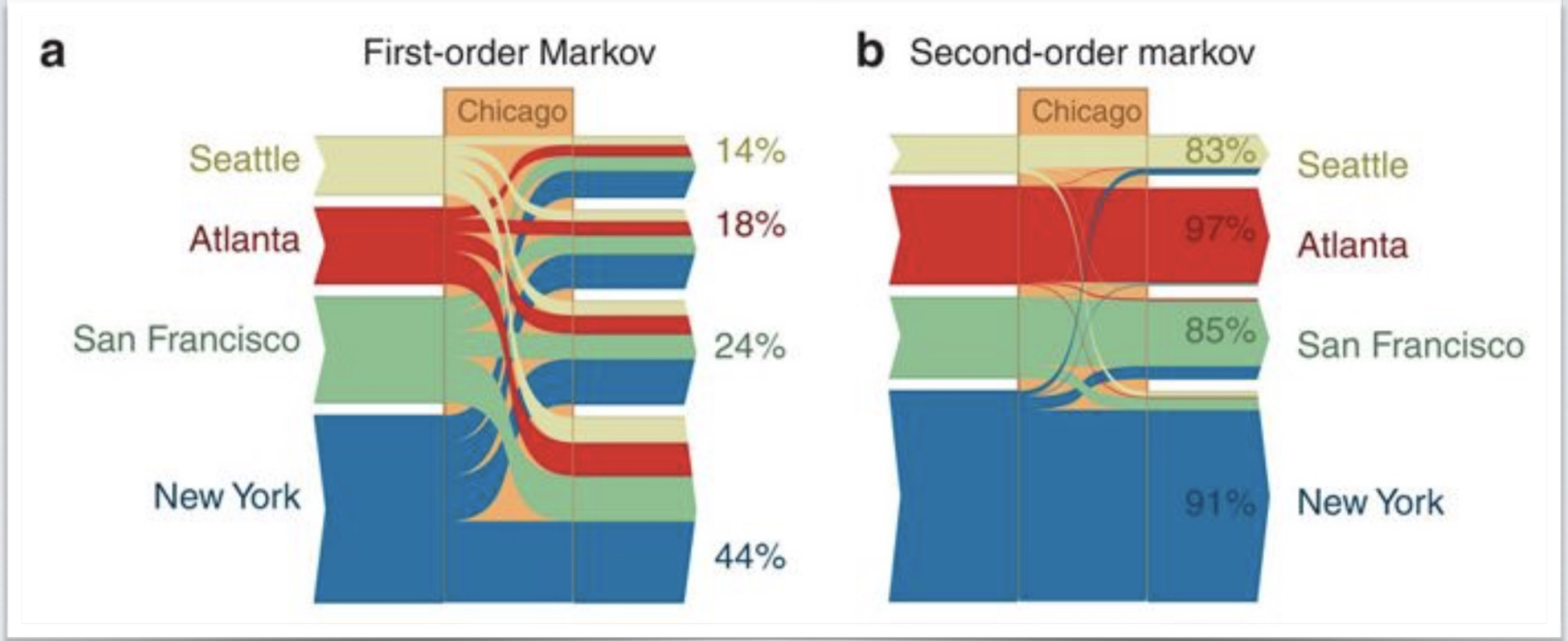
HIGHER ORDER NETWORKS

- **Non-markovian process**

- Markovian process:
  - A random process in which the future is independent from the past
  - Describe a process: Markov chains



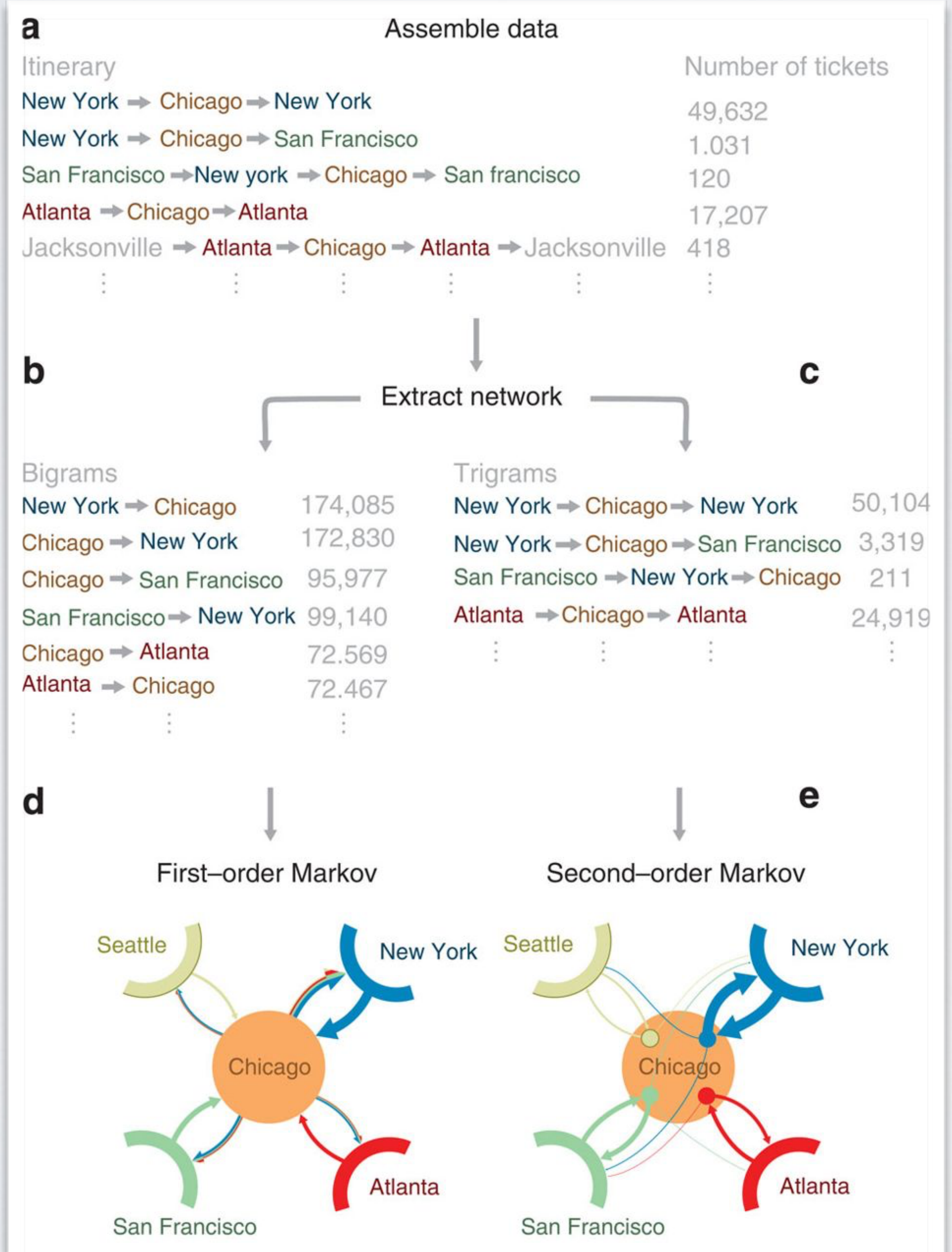
HIGHER ORDER NETWORKS



Round trips

Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., & Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications*, *5*, 4630.

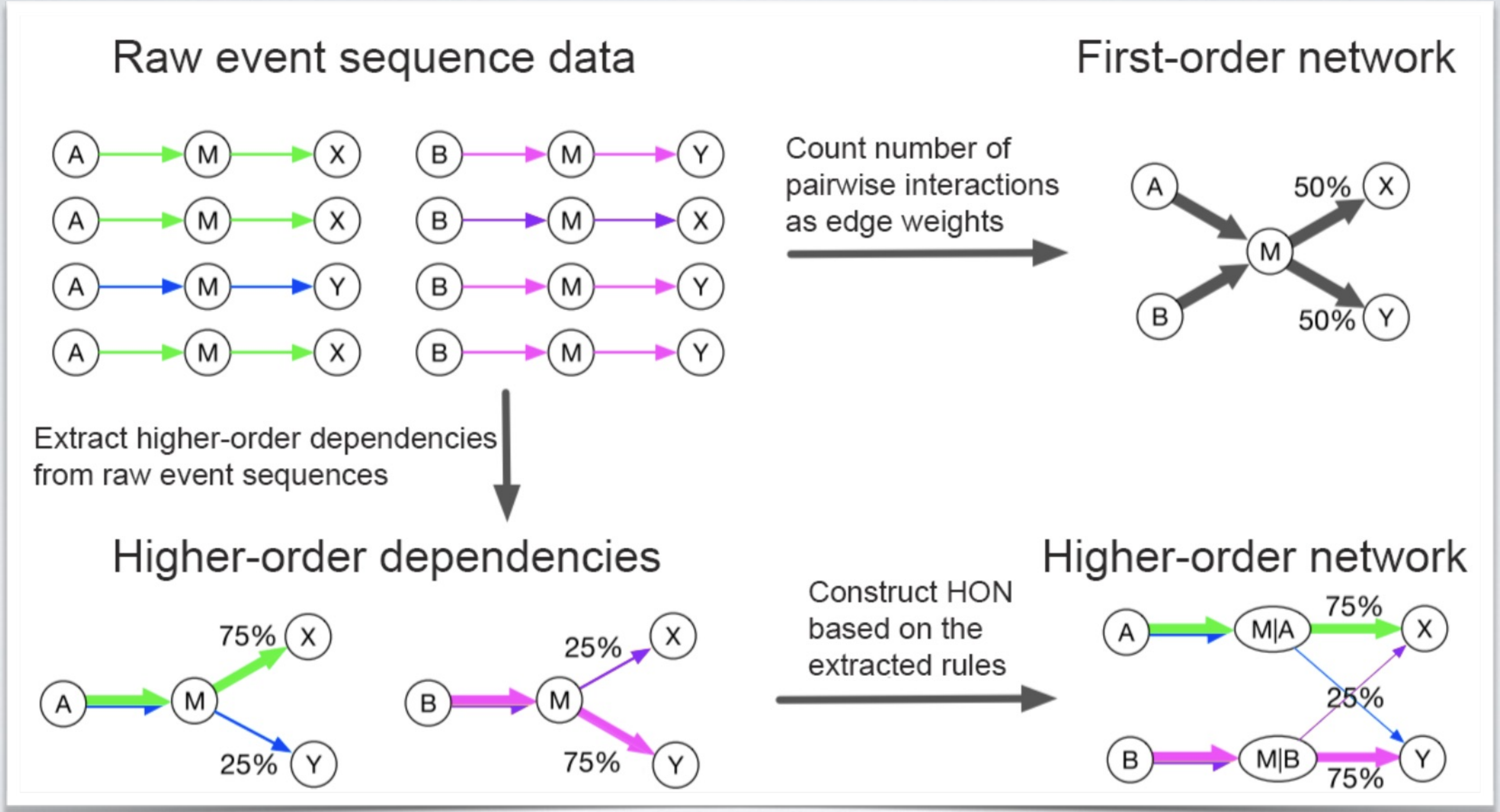
HIGHER ORDER NETWORKS



HIGHER ORDER NETWORKS

- New nodes are created
  - Do not correspond to an original item (a city)
  - Correspond to an item AND an origin
- A single element of memory: second-order network
- Two elements of memory: third-order network
- etc.

HIGHER ORDER NETWORKS



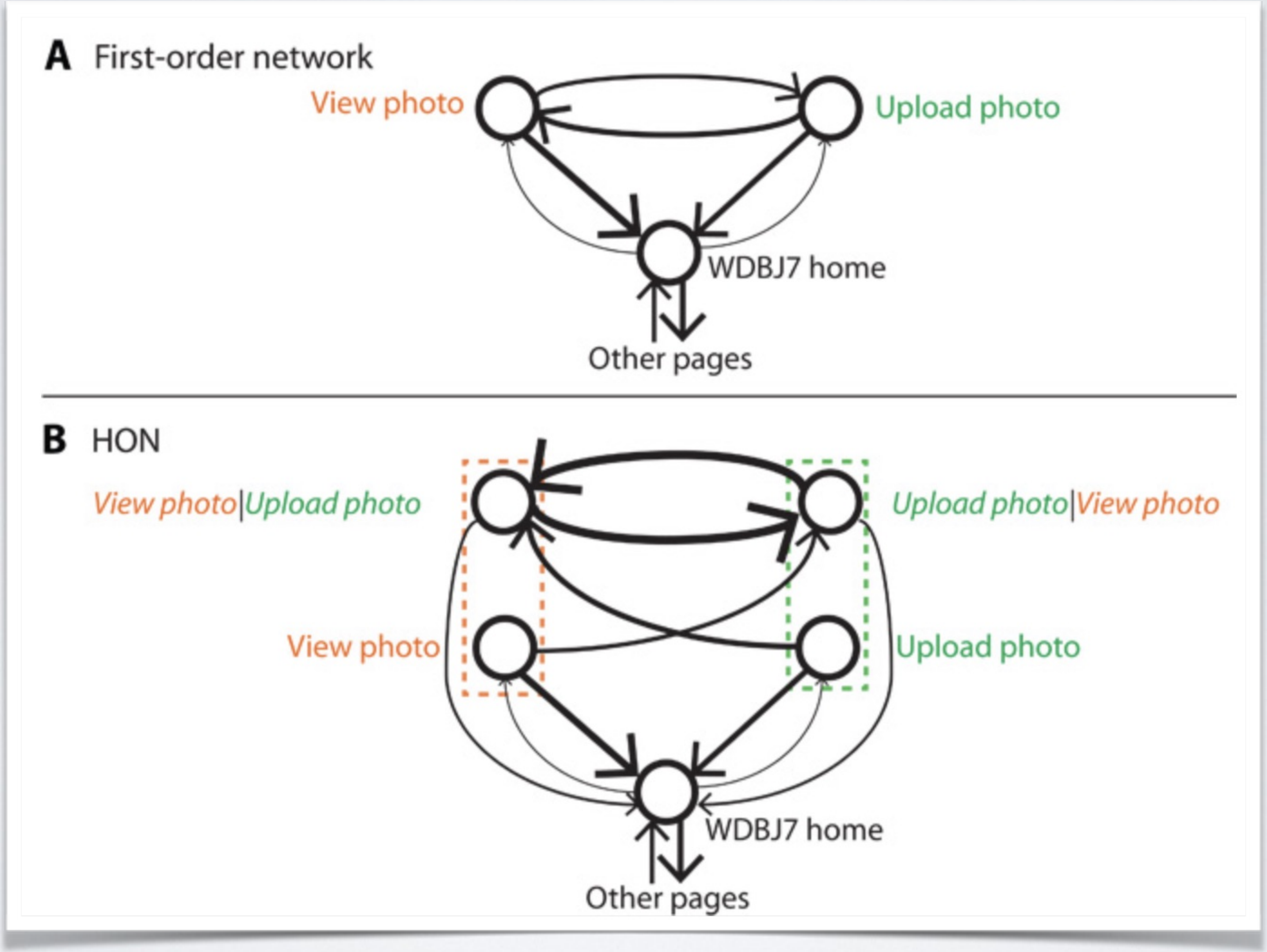
HIGHER ORDER NETWORKS

- Random walk approaches generalize naturally to higher order networks
  - Centrality: PageRank
  - Communities: Infomap
- They are based on the principle of a markovian processes
  - At each step, the *random walker* decides to follow an out-going link
  - This probability can depend on the walker origin

HIGHER ORDER NETWORKS

- Applying a community detection algorithm to a HON
- A node is now a tuple (node, history)
- If we apply a community detection algorithm:
  - Communities are composed of (node,history) vertices
  - We can go back to a traditional community partition:
    - We forget the memory part of nodes
    - Several instances of same nodes in same community
    - Same node in different communities
    - =>Overlapping communities

HIGHER ORDER NETWORKS



HIGHER ORDER NETWORKS

- Weakness: complexity
- Number of nodes multiplied by number of possible arrival sources
- => Rare cases could be ignored, current issue

Lambiotte, R., Rosvall, M., & Scholtes, I. (2019). From networks to optimal higher-order models of complex systems. *Nature physics*, *15*(4), 313-320.