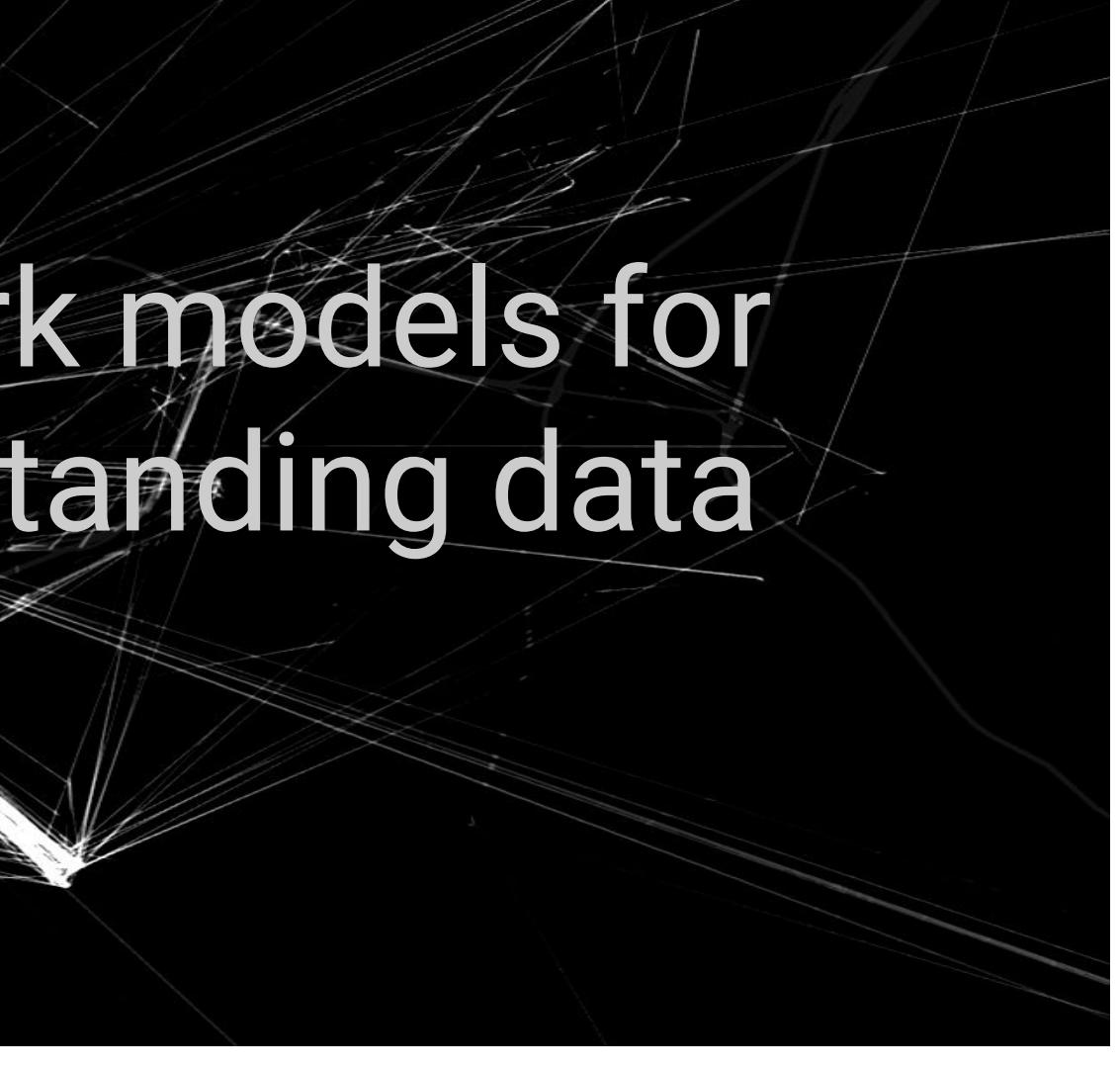


Network models for understanding data



Liu Bauer
CorrelAid Paris, CRI
City Interaction lab

Who am I?

Researcher at CRI Paris France, BLab
#mathematics #physics #computer science
Phd in Theoretical physics

Cofounder City Interaction lab,
Lecturers without borders,
CorrelAid Paris

Twitter @liyubov

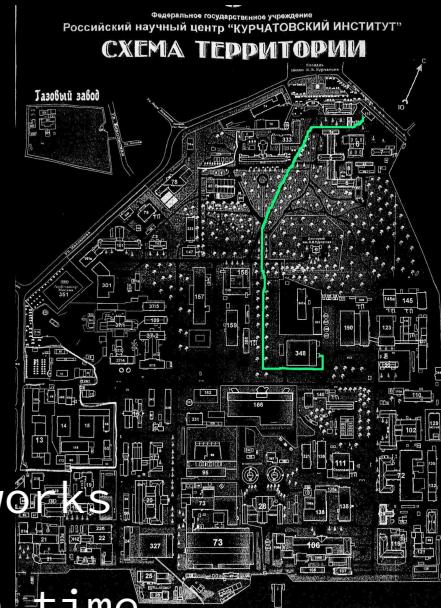
Transport in networks

City evolution in time

Citizen science

Innovations in science

Closed cities ZATOs @strelka



Outline

Network measures and networks in time and space

Network models

Networks from data

Our course details

<https://github.com/Big-data-course-CRI-in-2020>



Outline

1. why networks

-- A network or a graph (G) is a set composed of two subsets: (1) a vertex (V) or node set (e.g. a homeless individual) and (2)

2. network models:

remember that network model is only a model

3. geometric graph model, "very" random networks scale-free model

-- This unit disk graph has been used in computer science to model the topology of ad-hoc wireless communication networks

-- Scale-free models:

Empirical observation: many networks from the World Wide Web to a cell's metabolic system to actors in Hollywood—are dominated by a relatively small number of nodes that are connected to many other sites.

Base outline

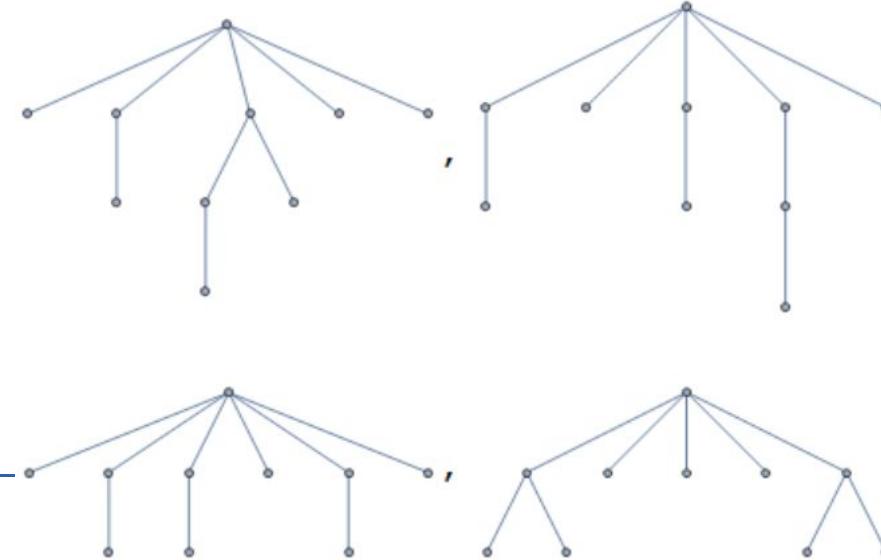
1. Network measures and networks in time and space
2. Network models
3. Networks from data

Our course details

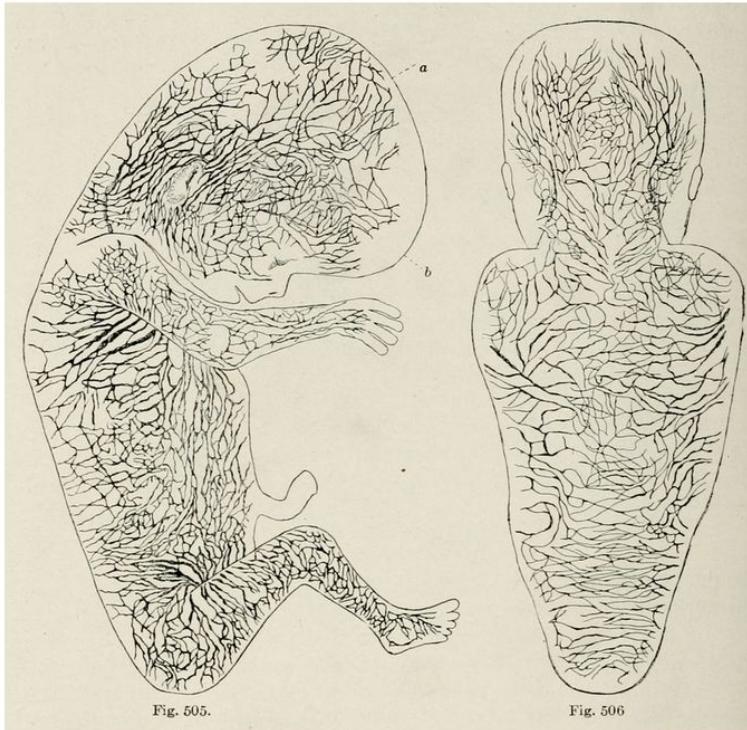
<https://github.com/Big-data-course-CRI-in-2020>



Natural networks



Where you can find networks?

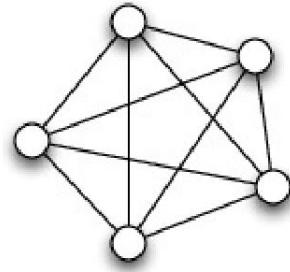


Distension of the lymphatic vessels in the human foetus, from Franz Kreibel, *Manual of human embryology*, 1910

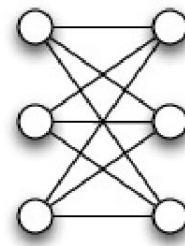


Birth of network science: Soft Matter and graphs

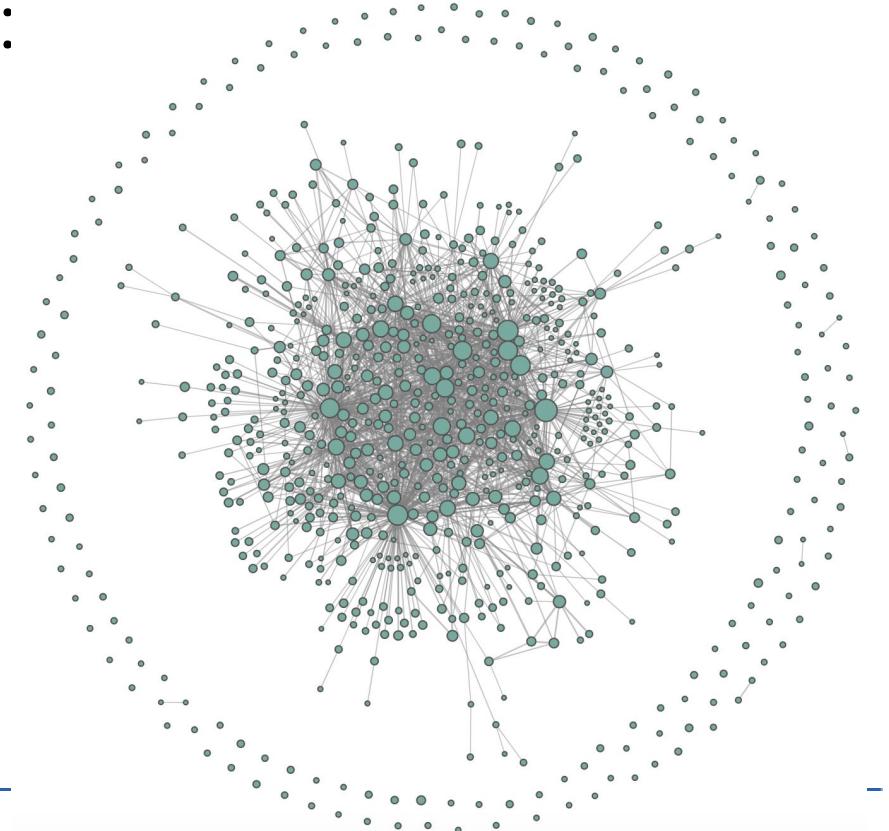
Network history goes back...



K_5



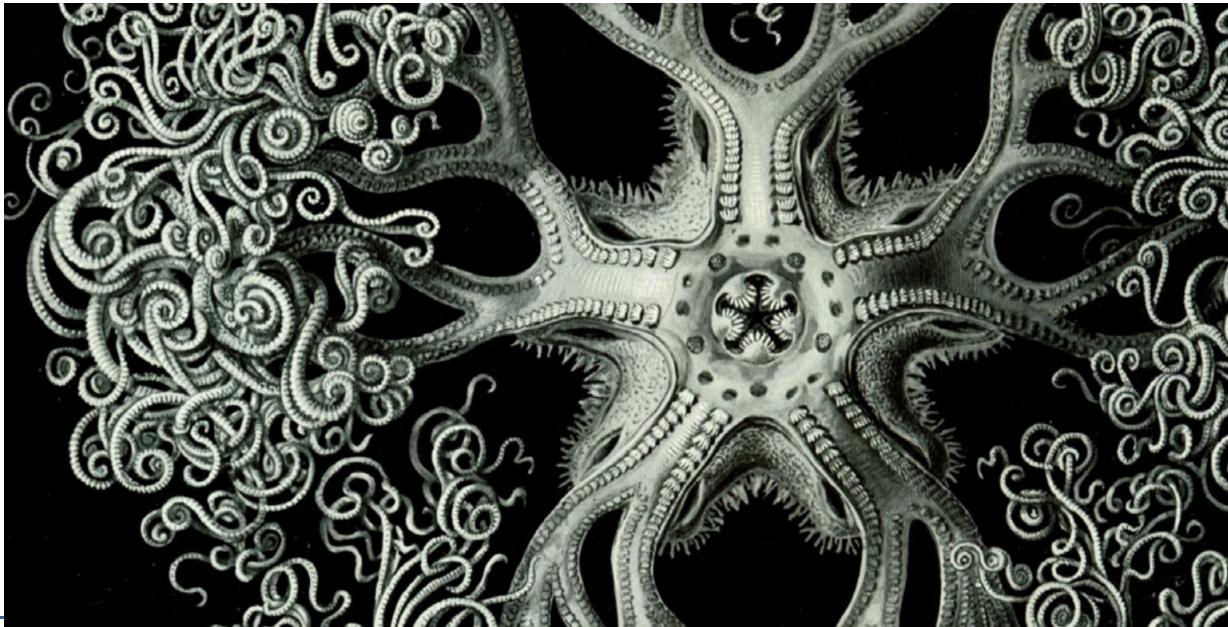
$K_{3,3}$



Can anything be presented as a network? What cannot be presented and why?

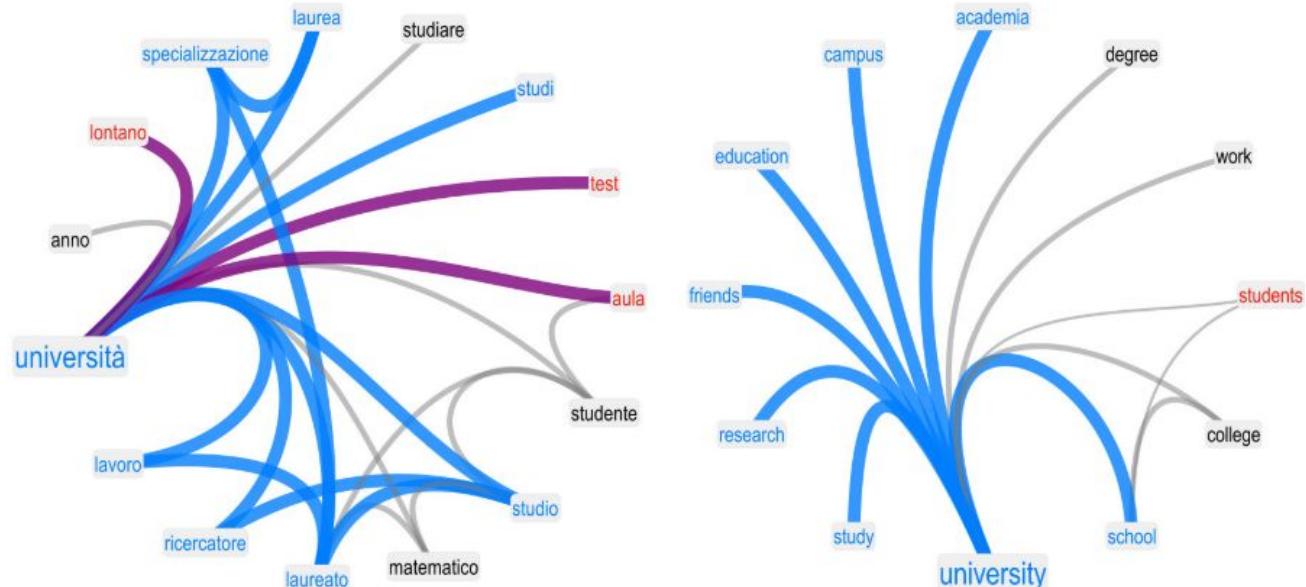


Can anything be presented as a network? What cannot be presented and why?



Can anything be presented as a network? What cannot be presented and why?

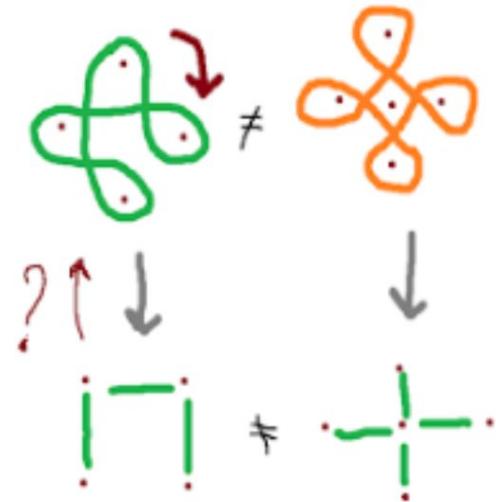
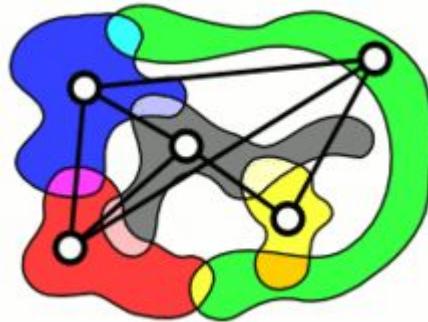
M.Stella et al.



Can anything be presented as a network? What cannot be presented and why?

Shanhi, L.Bauer et al.

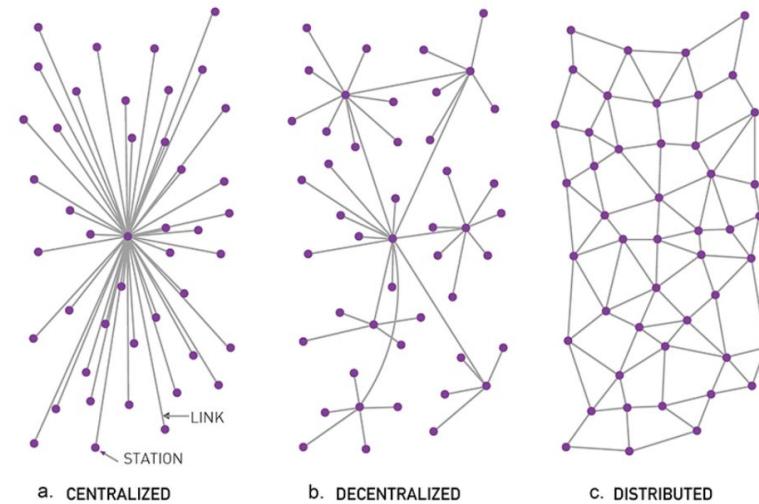
<https://liubauer.medium.com/mathematics-of-kolam-folkloric-graph-theory-4b3acc79d5cb>



What we will look at in network science?

1. Network measures
2. Networks in time and space
3. Networks and processes

Fig. credits P. Barran.



Outline

Network measures and networks in time and space

Network models

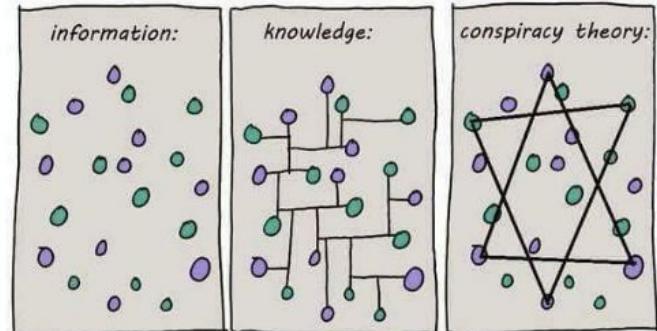
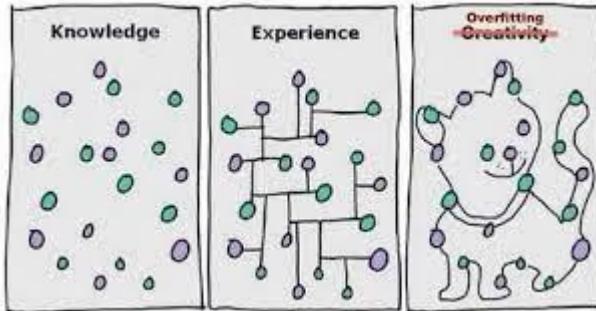
Networks from data

Our course details

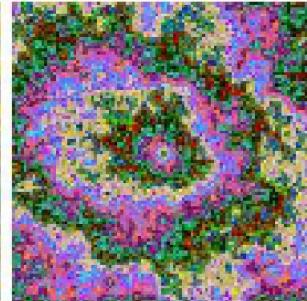
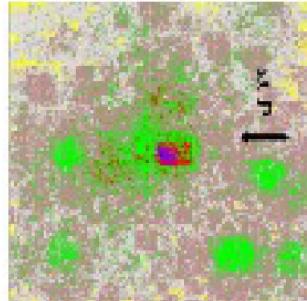
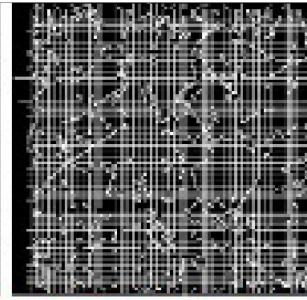
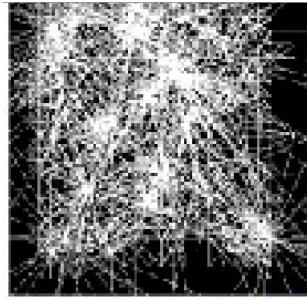
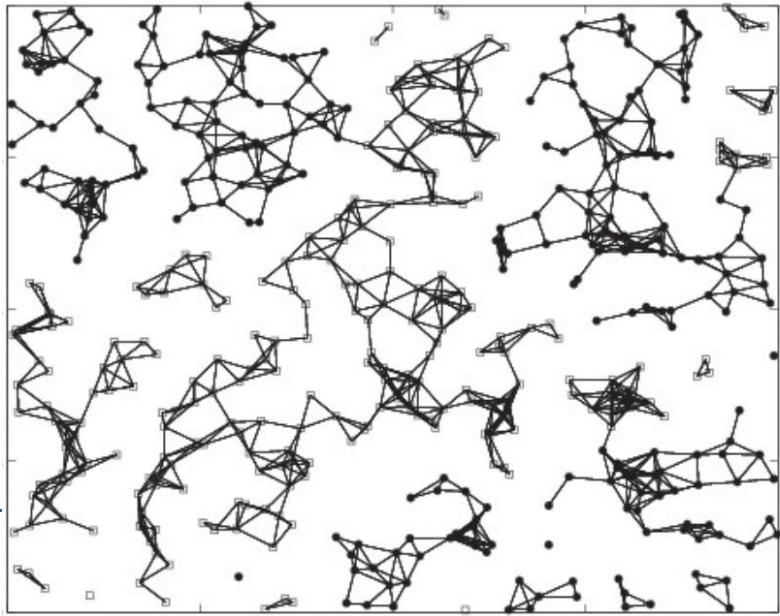
<https://github.com/Big-data-course-CRI-in-2020>



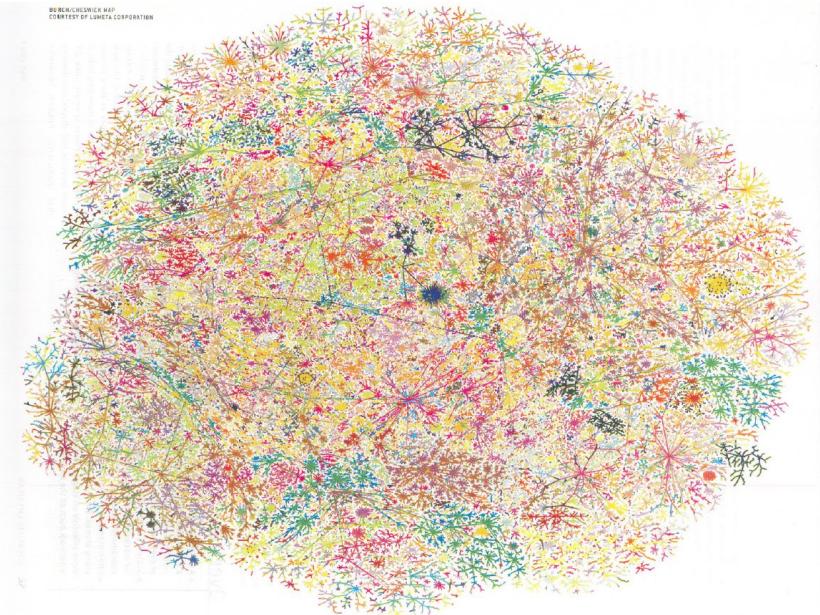
Networks models: pros and cons



Geometric network models



Scale-free network model



THE INTERNET, mapped on the opposite page, is a scale-free network in that some sites [starburst and detail above] have a seemingly unlimited number of connections to other sites. This map, made on February 6, 2003, traces the shortest routes from a test Web site to about 100,000 others, using like colors for similar Web addresses.

Scale-Free Networks

Scientists have recently discovered that various complex systems have an underlying architecture governed by shared organizing principles. This insight has important implications for a host of

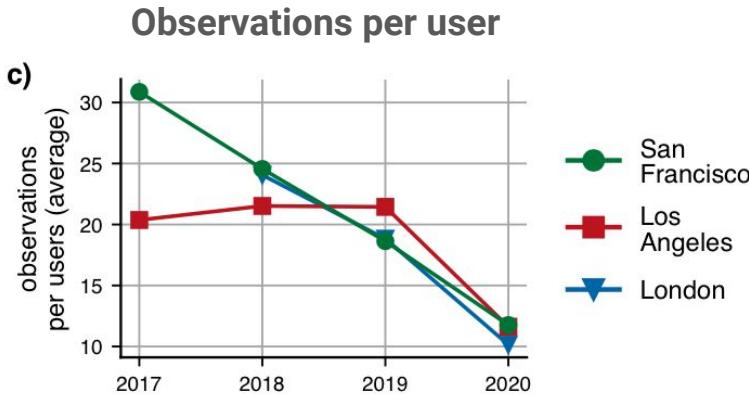
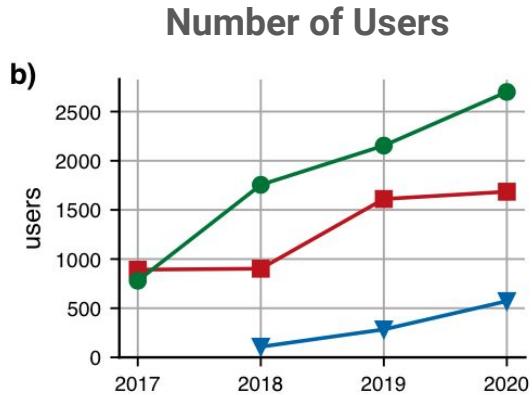
What is it about?

How grassroots organisations can help our cities:

Network analysis of citizen science

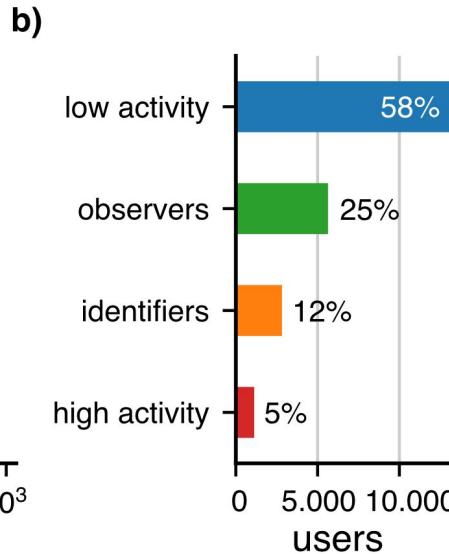
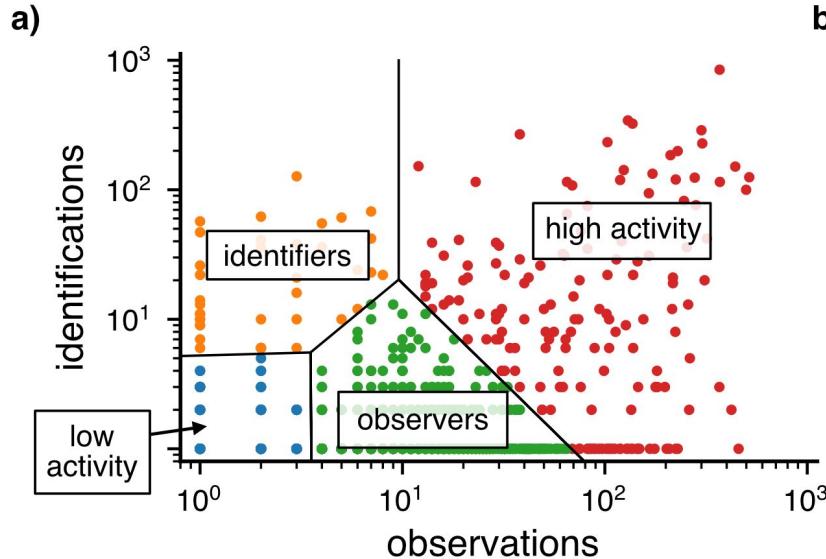


General trends



The number of users increased over the years, but **new users make fewer observations**.

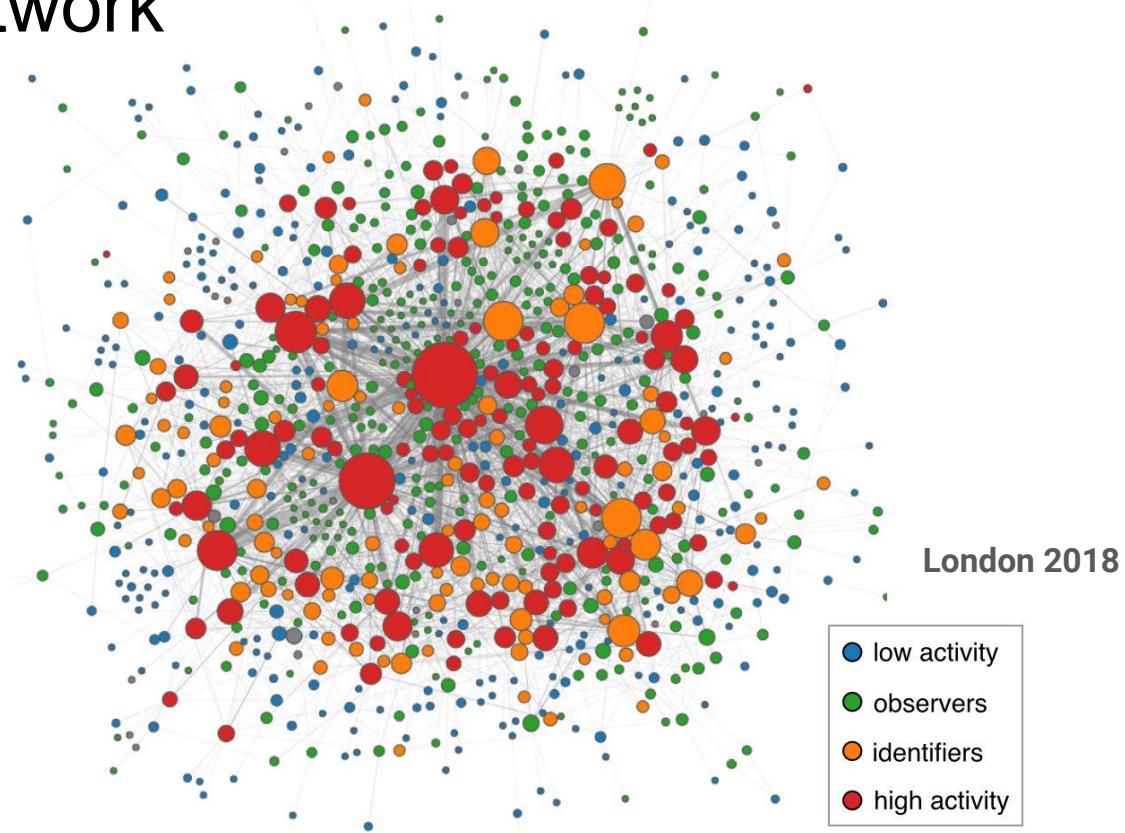
We identified four clusters of users



Most users are **low-activity** users, contributing few observations and identifications.

Only a small fraction are **high-activity**.

The social network of users



Outline

Network measures and networks in time and space

Network models

Networks from data

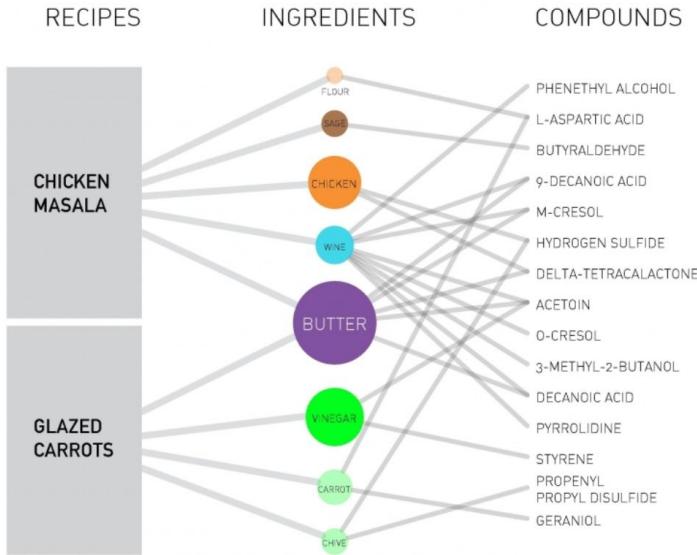
Our course details

<https://github.com/Big-data-course-CRI-in-2020>

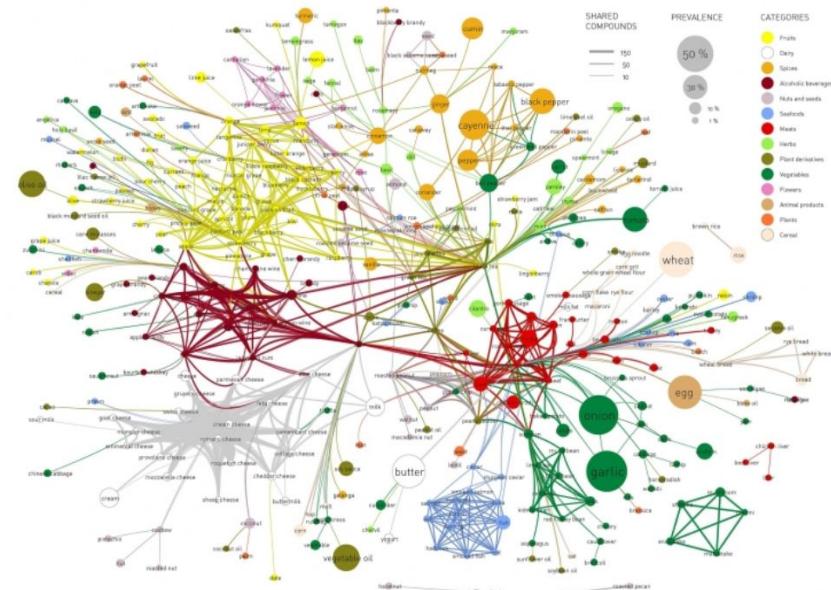


Networks from data

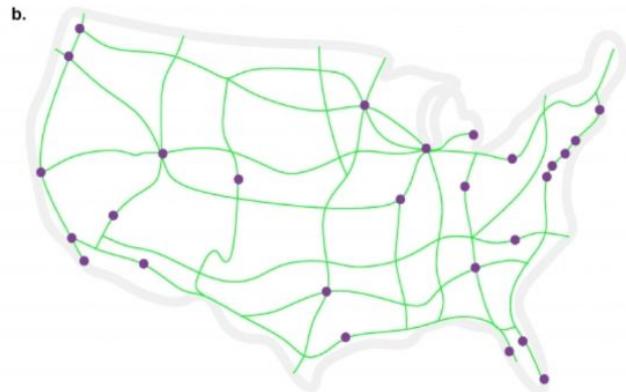
a.



b.

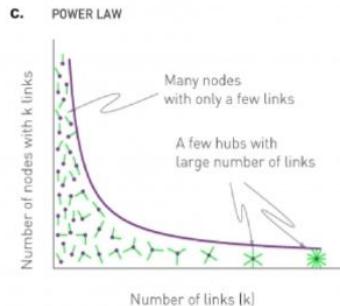
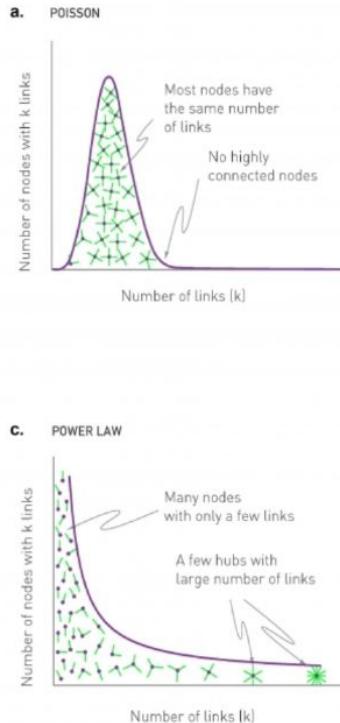
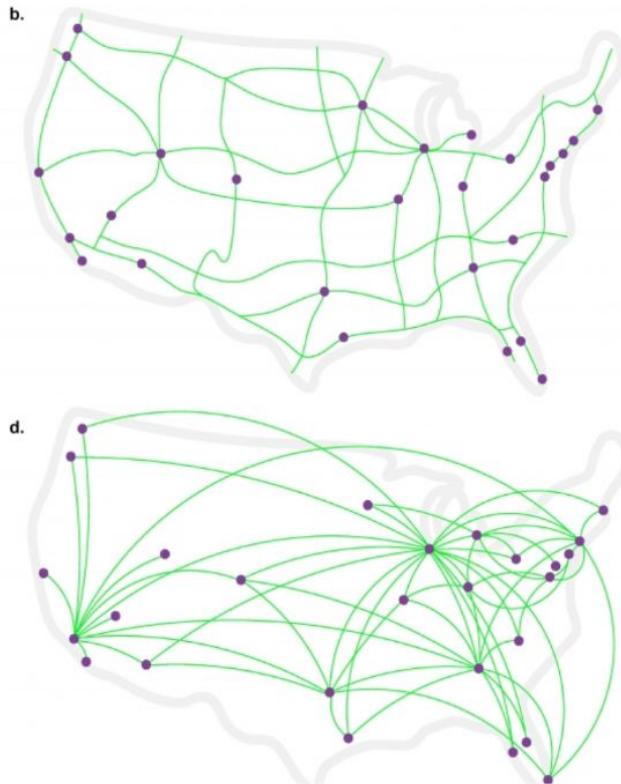


Network measures: Degree or why do we care?



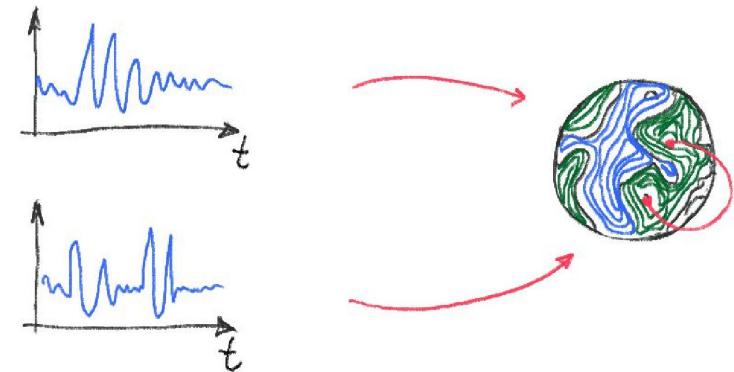
[http://barabasi.com/book
/network-science#netwo
rk-science](http://barabasi.com/book/network-science#network-science)

Network measures: Degree or why do we care? How to look into degree distributions?



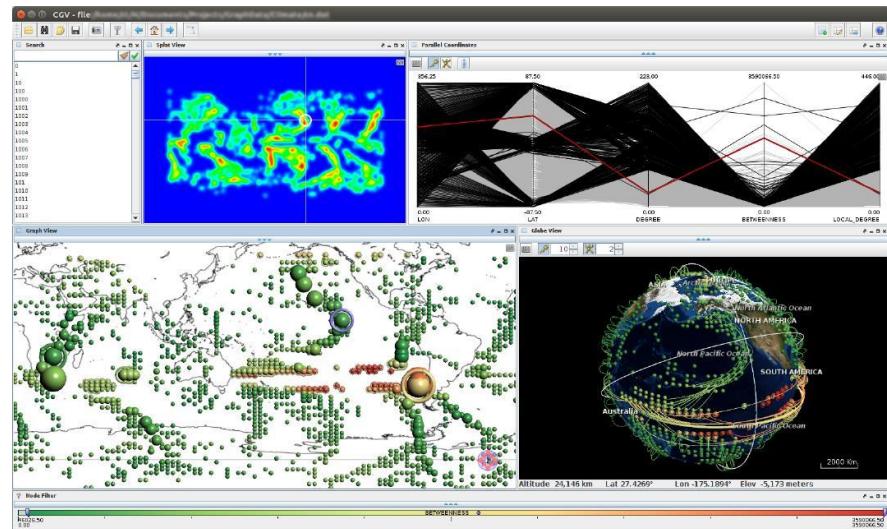
How to construct networks from data?

1. Directly build correspondence between links and edges with data (social networks, flights data)
2. Preprocess data (first build correlation from data)



How to construct networks from data?

1. Directly build correspondence between links and edges with data (social networks, flights data)
2. **Preprocess data (first build correlation from data)**
Working with data from big systems



Want to get involved citizen science analysis?
<https://antiborschewik.info/map>



Interested?

Projects and internships with
City Interaction lab cityinteractionlab/projects
Citizen Science community on CSA connect
Correlaid online community (Paris Chapter)
www.correlaid.org

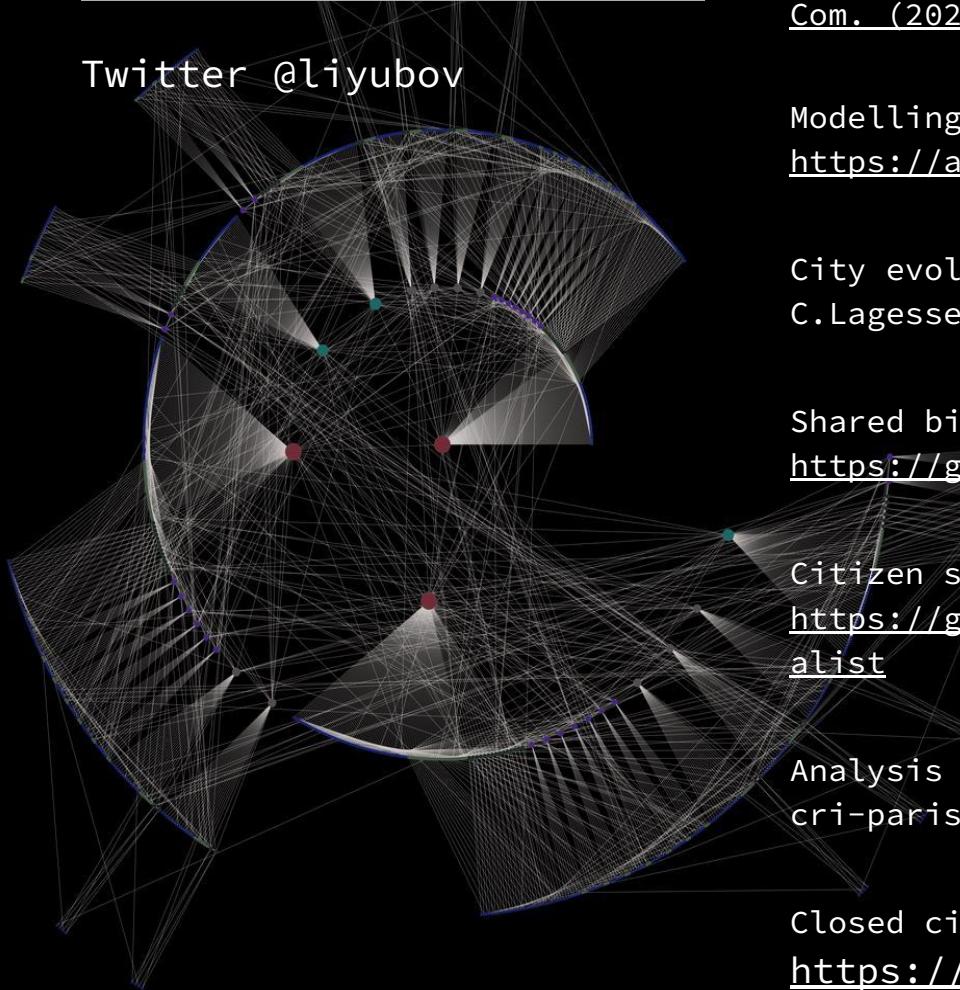
<https://github.com/Big-data-course-CRI-in-2020>

Contact
liubov.tupikina@cri-paris.org
Twitter @Liyubov



References to our projects

Twitter @liyubov



Transport in networks [Berthelot, L. Tupikina et al. Nat. Com. \(2020\)](#)

Modelling railway delay
<https://arxiv.org/abs/2105.06111>

City evolution in time
C.Lagesse, L.Tupikina, P.Bauman

Shared bike systems
<https://github.com/Liyubov/bike-sharing>

Citizen science city analysis
https://github.com/correlaid-paris/citizen_science_inaturalist

Analysis of innovations in science
cri-paris.org

Closed cities ZATOs
<https://theterraforming.strelka.com/>

Network resources

<http://networkrepository.com/networks.php>

http://networksciencebook.com/chapter/3#advanced-b_b

Spatial Networks

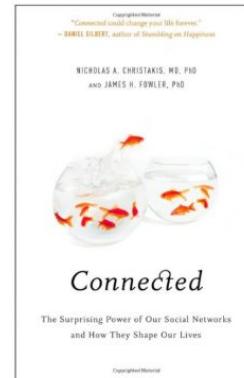
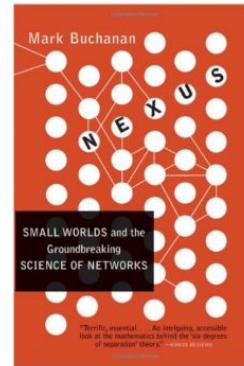
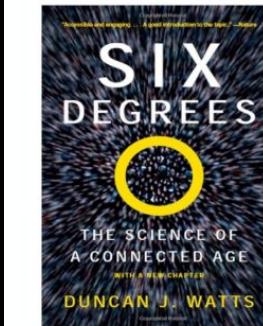
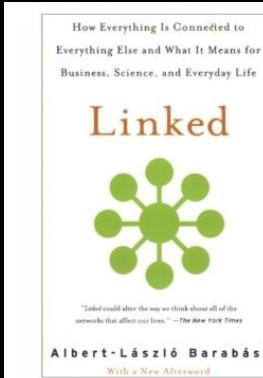
Marc Barthélémy*

Institut de Physique Théorique, CEA, IPHT CNRS, URA 2306 F-91191 Gif-sur-Yvette France and
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.

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, 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. An important consequence of space on networks is that there is a cost associated to the length of edges which in turn has dramatic effects on the topological structure of these networks. We will expose thoroughly the current state of our understanding of how the spatial constraints affect the structure and properties of these networks. We will review the most recent empirical observations and the most important models of spatial networks. We will also discuss various processes which take place on these spatial networks, such as phase transitions, random walks, synchronization, navigation, resilience, and disease spread.

Contents

I. Networks and space	2	1. Erdos-Renyi graph	39
A. Introduction	2	2. Planar Erdos-Renyi graph	39
B. Quantitative geography and networks	2	3. The hidden variable model for spatial networks	40
C. What this review is (not) about	2	4. The Waxman model	41
II. Characterizing spatial networks	3	C. Spatial small worlds	41
A. Generalities on planar networks	3	1. The Watts-Strogatz model	41
1. Spatial and planar networks	3	2. Spatial generalizations	42
2. Classical results for planar networks	3	D. Spatial growth models	43
3. Voronoi tessellation	4	1. Generalities	43
		2. Preferential attachment and distance selection	43
		3. Growth and local optimization	46
		E. Optimal networks	49



Resources page: papers

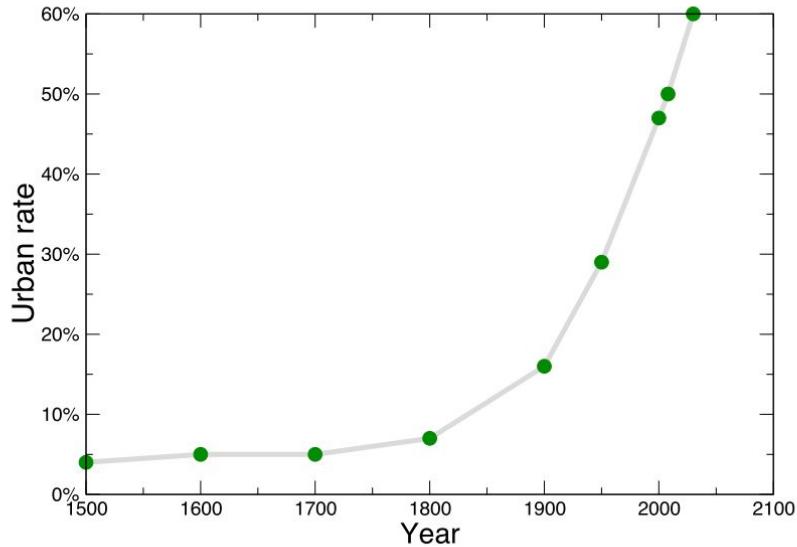
Paper on human mobility review <https://arxiv.org/abs/1710.00004>

Paper on spatial planning

Paper on spatial networks <https://arxiv.org/abs/1010.0302>

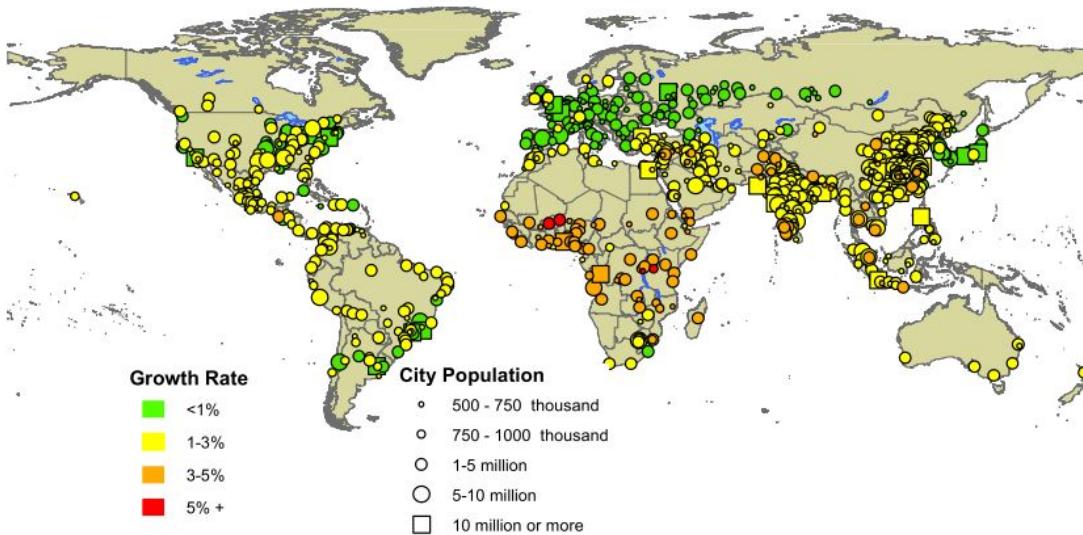
Why city science?

Importance of cities: urbanization rate

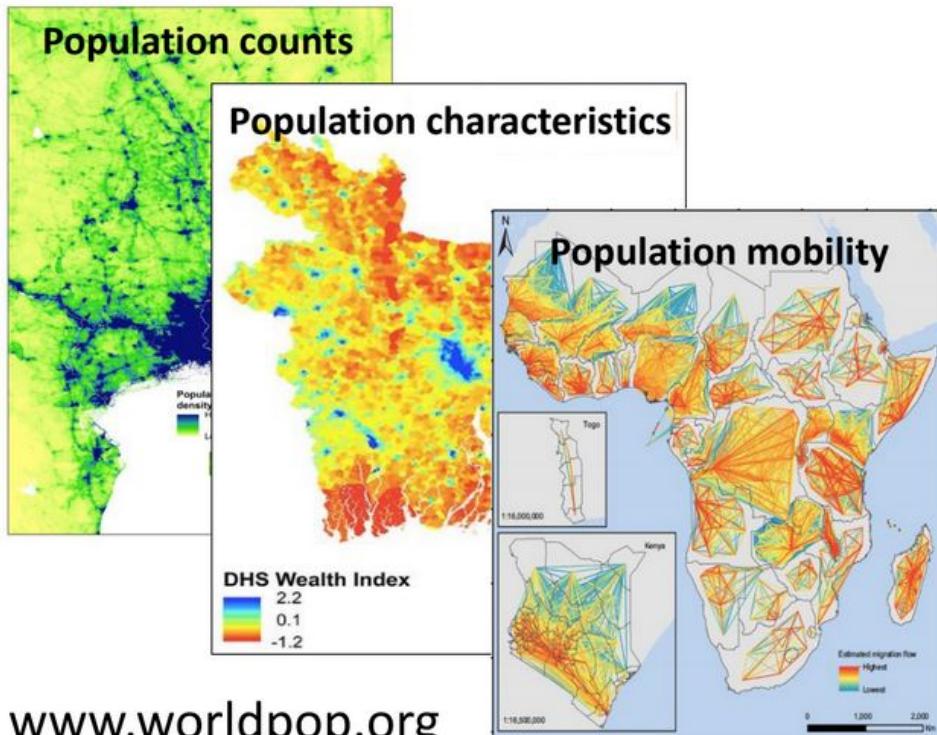


Projection: in 2050: 70% of the world population lives in cities

Why city science?



Why spatial analysis?



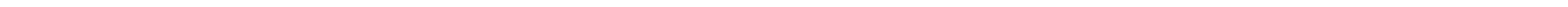
Applied research and implementation group

Mapping small area population distributions, demographics and dynamics

Open data, open peer-reviewed statistical methods, user engagement

Application in epidemiology, maternal/newborn health, childhood vaccination

Human mobility



Human mobility

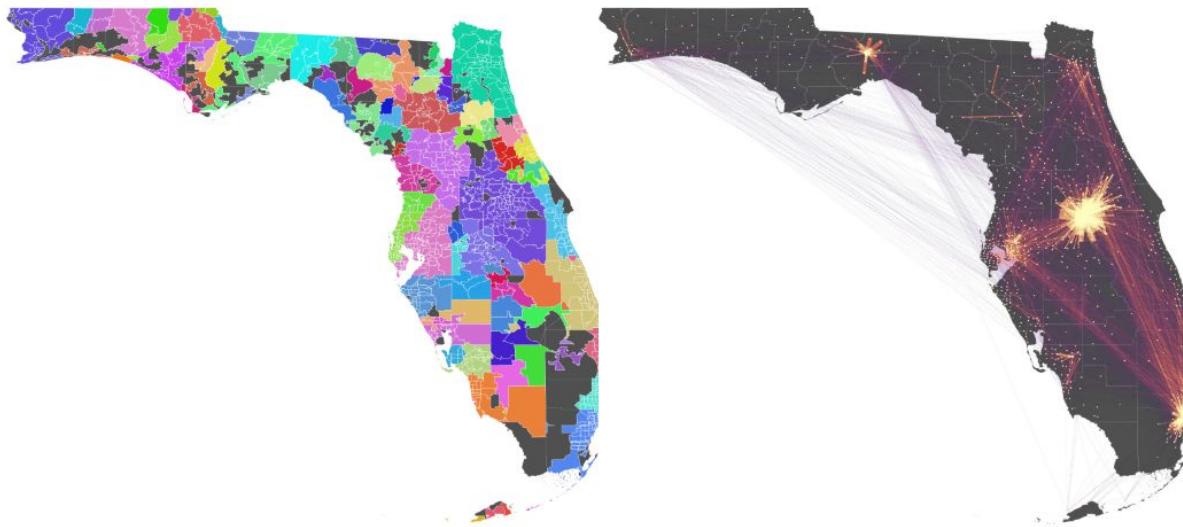


Fig. 3: Commuting flows compiled from census data. Left panel: The state of Florida partitioned according to its counties. Right panel: Commuting flows between counties, where thickness of lines correspond to volume of flow. Data compiled from the United States Census Bureau.

Human mobility

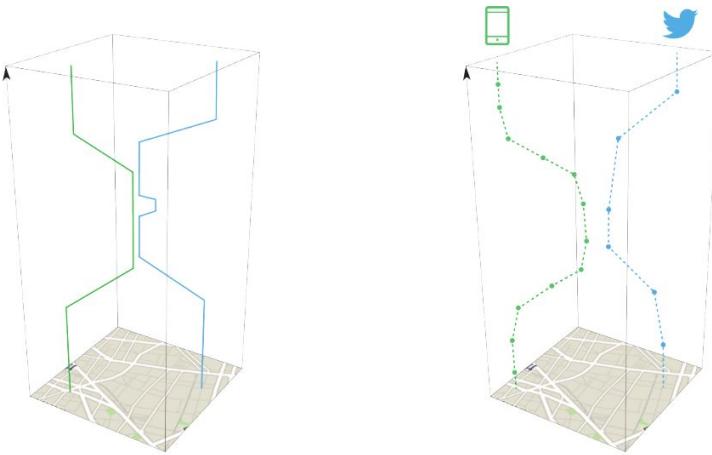
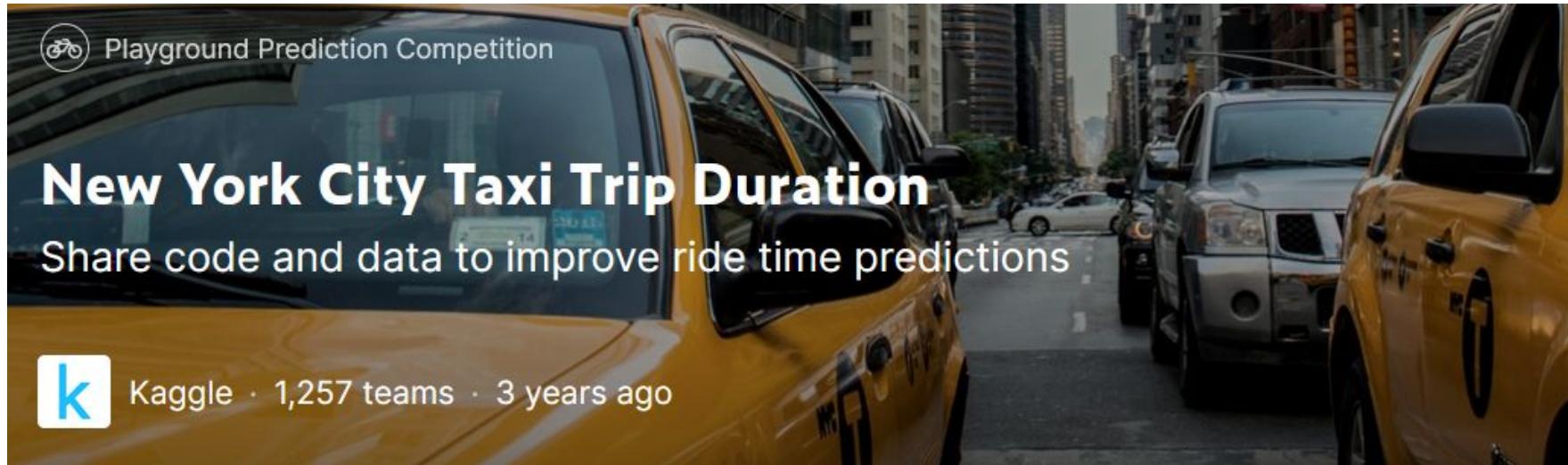


Fig. 1: The cubes of time geography, as first proposed by Torsten Hägerstrand in [36]. The geographical space is represented by the 2D plan, while time is figured by the vertical axis. (Left) The two curves represent the daily space-time trajectories of two individuals living in the same neighborhood and working in the same place. (Right) The geographical footprints continuously and passively produced by individuals through the use of their ICT devices allow to approximate their trajectories. While these re-constructed trajectories are partial and contain errors that might mislead the understanding of underlying trajectories, they are nonetheless more precise nowadays than they were 10 years ago, and produced by a constantly growing number of individuals worldwide.

Human mobility: age of data



Playground Prediction Competition

New York City Taxi Trip Duration

Share code and data to improve ride time predictions

Kaggle · 1,257 teams · 3 years ago

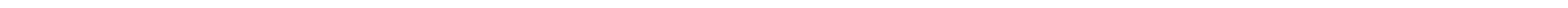
Overview Data Notebooks Discussion Leaderboard Rules

Some available data sources:

<http://www.census.gov/hhes/commuting/data/commutingflows.html>

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/>

Human mobility: What are possible types of users?



Human mobility: which measure to use?

Messages:

- Just plotting trajectory does not really work
- Measure should be splitting users into groups
- Measures should be independent from space

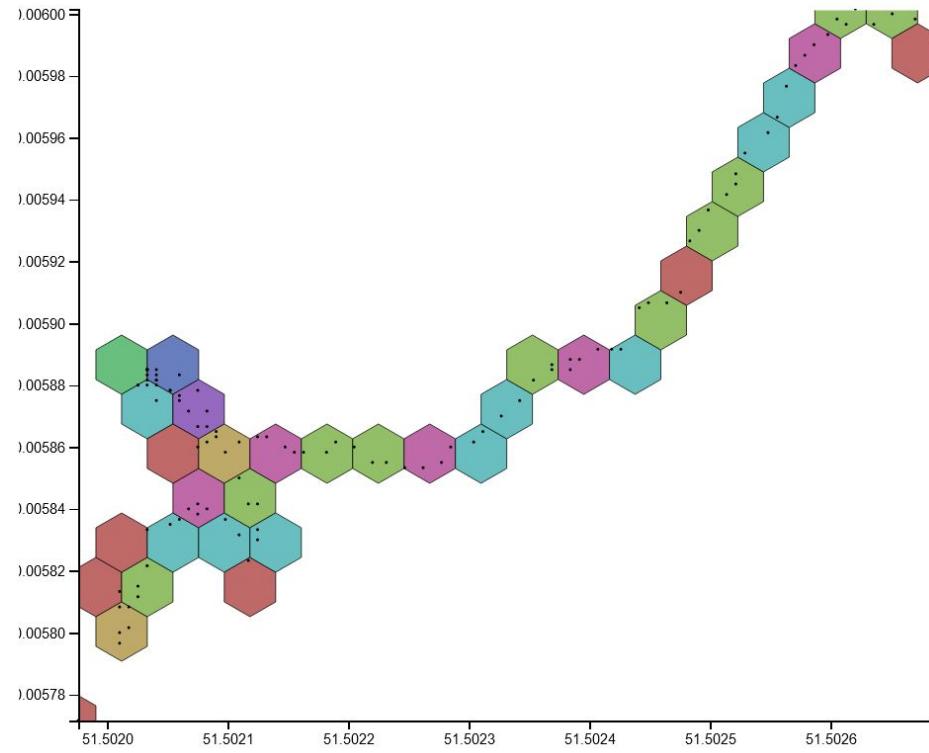
Returners and explorers dichotomy in human mobility

<https://www.nature.com/articles/ncomms9166/>

Data walking project resource

Reminder from class 5

How to analyse trajectories?



Human mobility: which measure to use?

Messages:

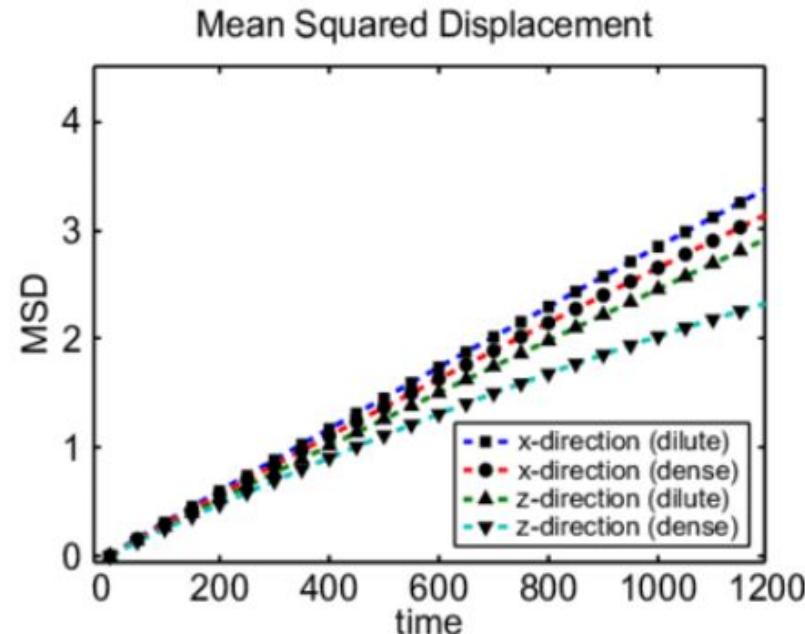
- Just plotting trajectory does not really work
- Measure should be splitting users into groups
- Measures should be independent from space:
E.g mean squared displacement (MSD)
is a measure of the deviation of the position

Returns and explorers dichotomy in human mobility

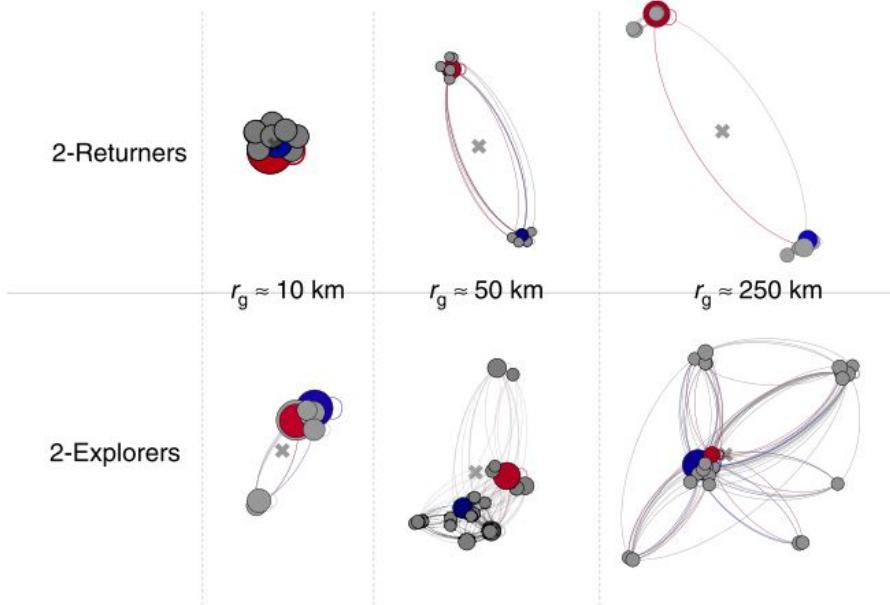
<https://www.nature.com/articles/ncomms9166/>

Reminder from class 5

How to analyse trajectories?



Human mobility: which measure to use?



Radius of gyration - root mean square distance of the object's parts from either its center of mass.

$$r_g^{(k)} = \sqrt{\frac{1}{N_k} \sum_{j=1}^k n_j (\mathbf{r}_j - \mathbf{r}_{cm}^{(k)})^2}$$

Returns and explorers dichotomy in human mobility
<https://www.nature.com/articles/ncomms9166/>

Human mobility: types of users

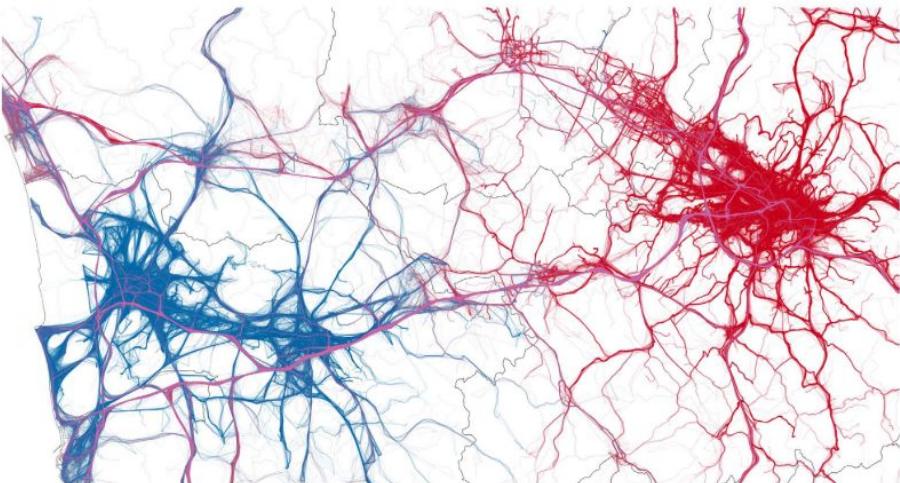
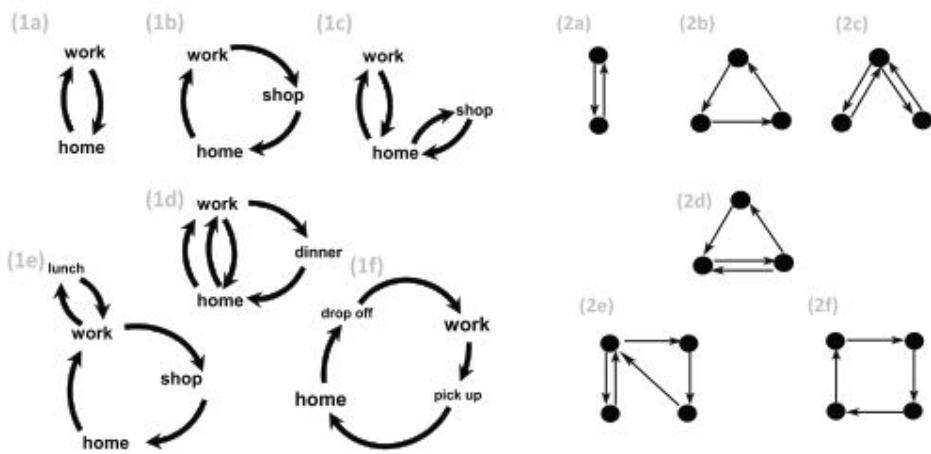


Figure. 1 | A visualization of the complexity of the explored mobility patterns. A fragment of the GPS trajectories used in our study, displaying trips originating in the metropolitan areas of Pisa (in blue) and Florence (red). This plain geo-referenced visualization of experimental data reveals the confrontation of two 'competing' metropolitan areas. It also demonstrates the ability of Big Data to portray social complexity. This map has been generated through the QGIS software, available at <http://www.qgis.org/en/site/>.

Returns and explorers dichotomy in human mobility

<https://www.nature.com/articles/ncomms9166/>

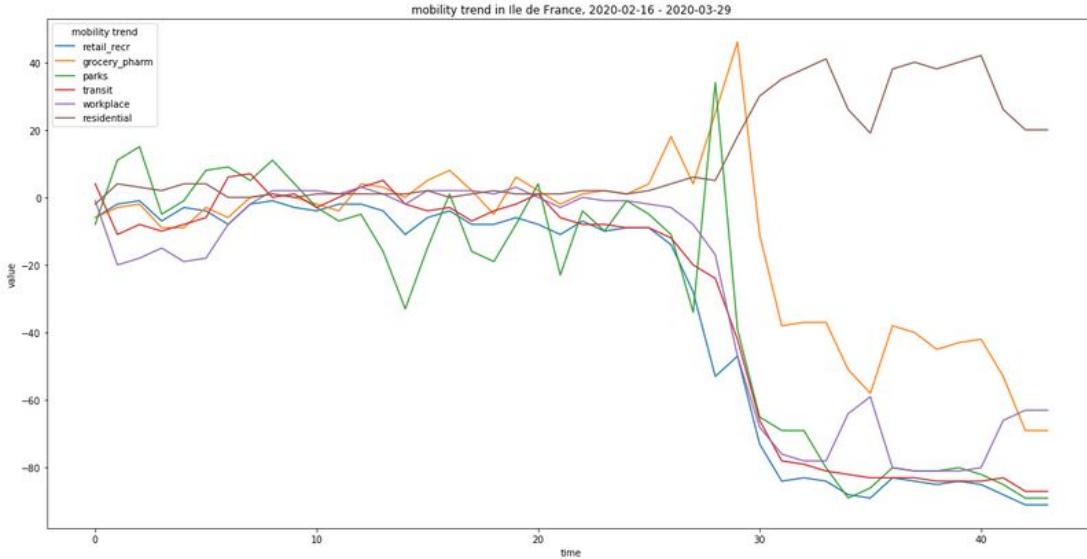
Human mobility: which measure to use?



Most Frequented Locations and Motifs - mobility privacy

Are all human very different in the way they are walking and traveling every day?

Human mobility: what did COVID change?



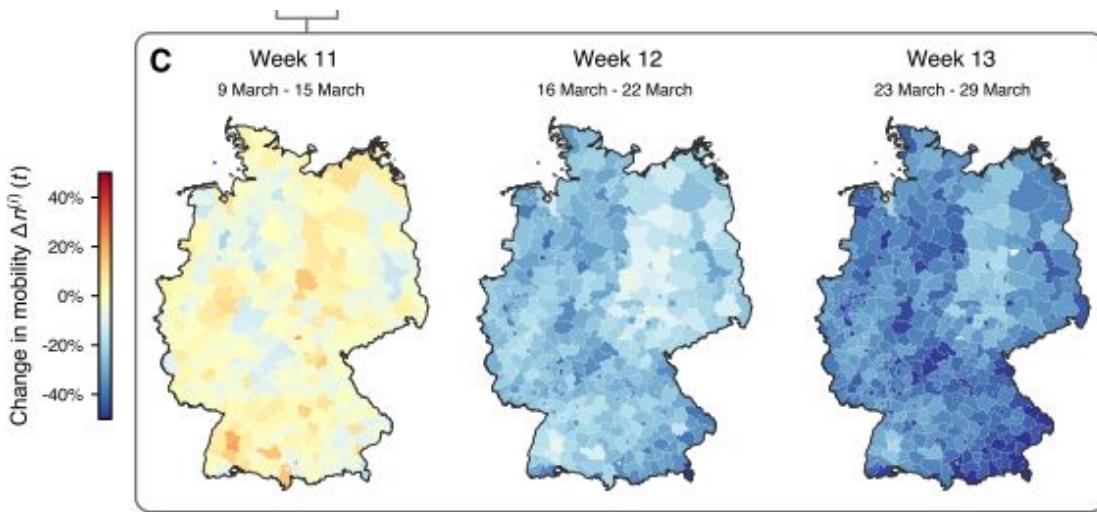
Most Frequented Locations and Motifs - mobility privacy

Are all human very different in the way they are walking and traveling every day?

Code link data

<https://github.com/Liyubov/google-mobility-reports-data?organization=Liyubov&organization=Liyubov>

Human mobility: what did COVID change?



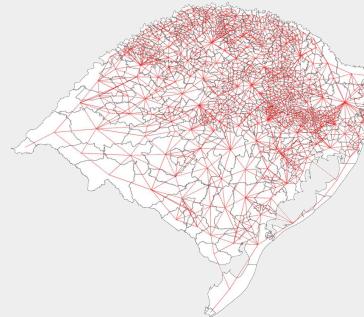
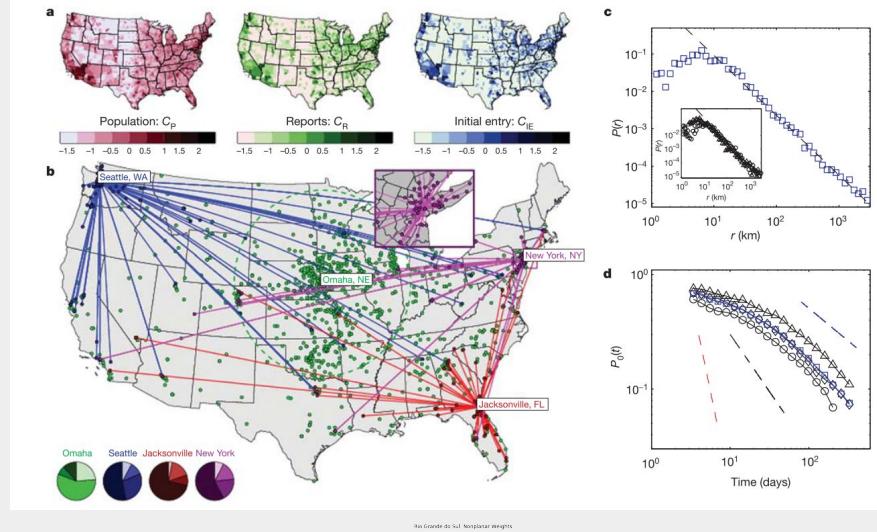
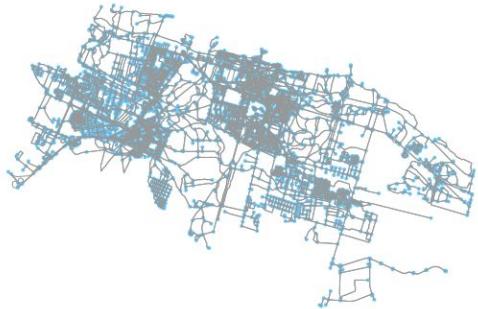
Global changes of COVID19

Networks of mobility, changes in human behaviour, city vs. countryside, people from different countries

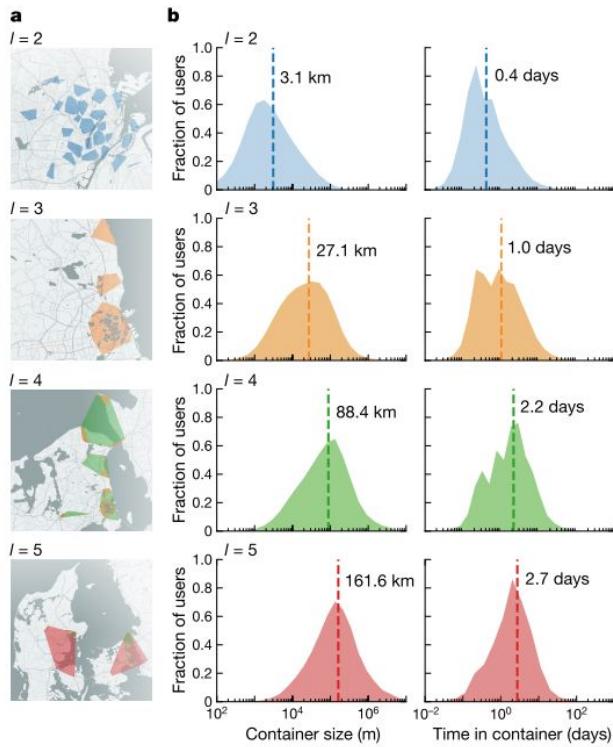
Human mobility scales

Spatial and temporal analysis:

How distributions of trips length durations look like?

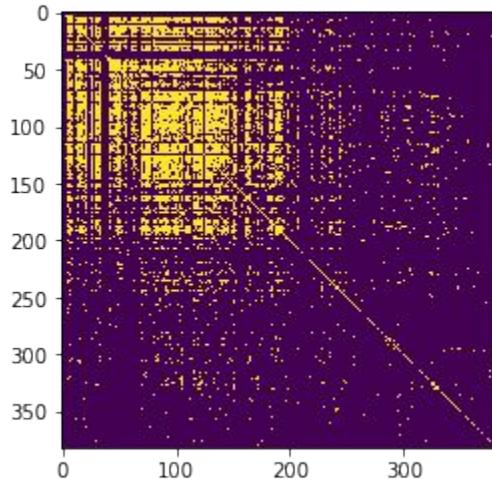
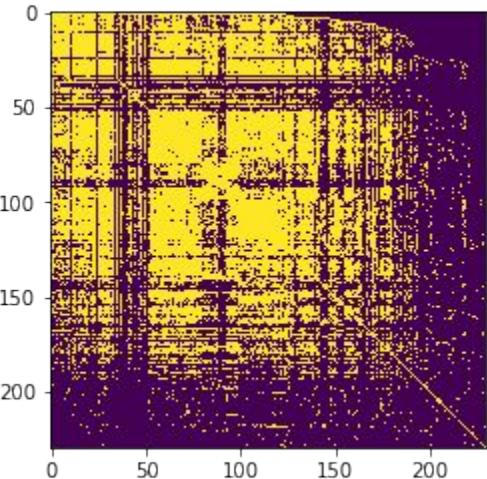


Human mobility: which measure to use?

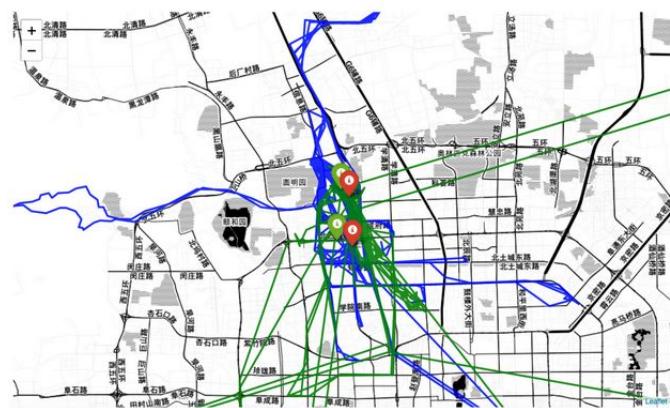


Scales of mobility:
How to show universality of human
travels at different scales?

Human mobility: which measure to use?



Origin-Destination Matrix - the number of individuals traveling between locations in a given area, over a given period of time.



Human mobility: virtual



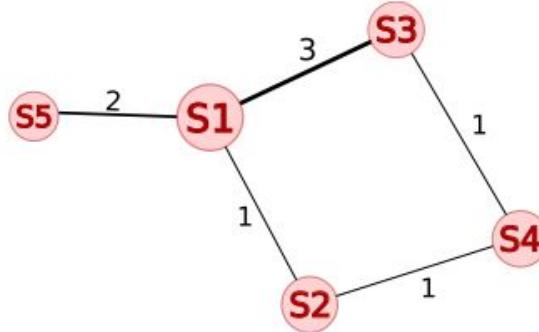
Is it only valid for spatial locations or also for internet?

Returners and Explorers Dichotomy in Web Browsing Behavior – A Human Mobility Approach

https://link.springer.com/chapter/10.1007%2F978-3-319-30569-1_13



mobility: virtual



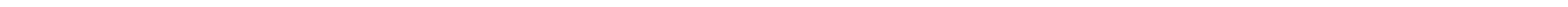
Is it only valid for spatial locations or also for internet?

Returns and Explorers Dichotomy in Web Browsing Behavior – A Human Mobility Approach

https://link.springer.com/chapter/10.1007%2F978-3-319-30569-1_13

Data on web browsing history: The dataset consists of more than 5M anonymized Web browsing history entries corresponding to visits to 187,680 hosts by 524 users between September 21, 2010 to May 24, 2014.

Spatial networks



What is special about spatial networks?

For the course on mobility and spatial networks:

1. Network measures specific for spatial networks:
(1) Closeness, (2) Betweenness centrality, (3) Straightness,
(4) Information (from blue to red the centrality increases).

Which else network measures could you imagine?

2. OSMNX package for openstreetmaps API

Question:

how can we integrate the information from network theory into maps
in order to make it more usable?

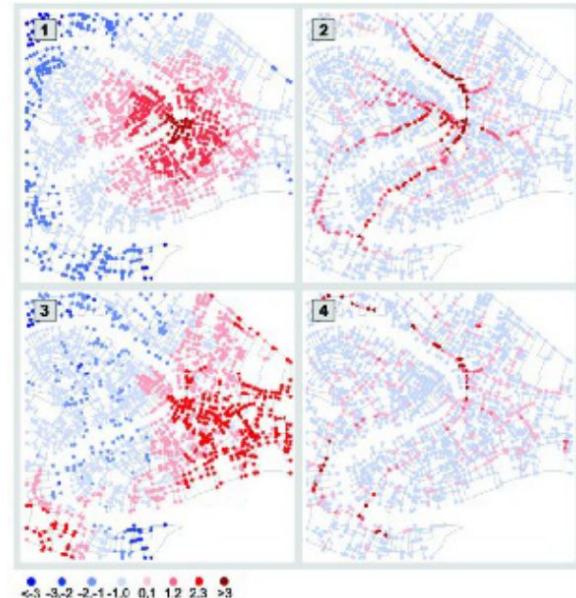
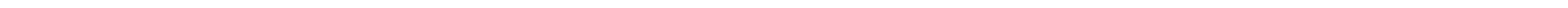
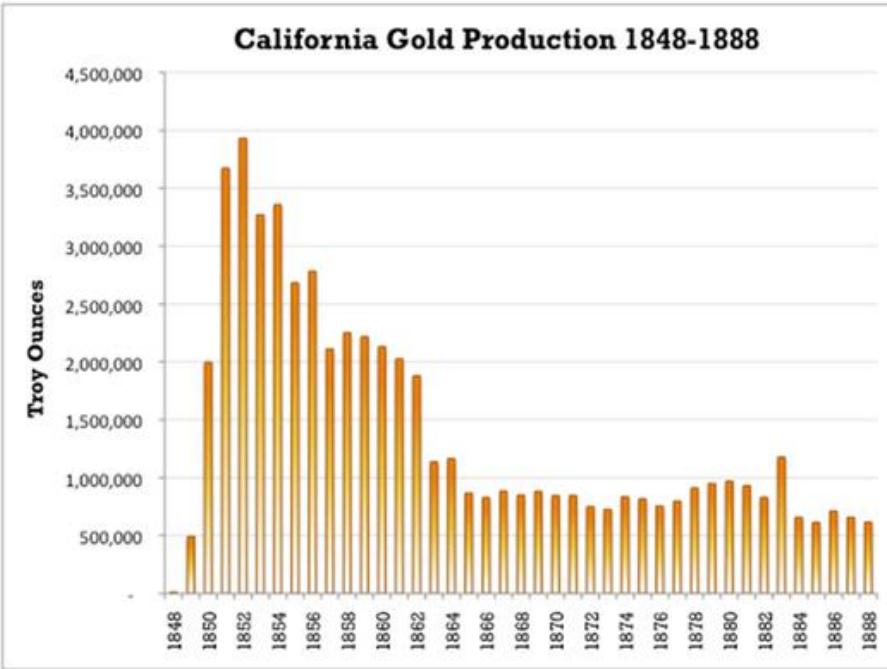


FIG. 24 Colour-coded maps for different centralities in Venice, Italy (see [102]). (1) Closeness, (2) Betweenness centrality, (3) Straightness, (4) Information (from blue to red the centrality increases). From [102].

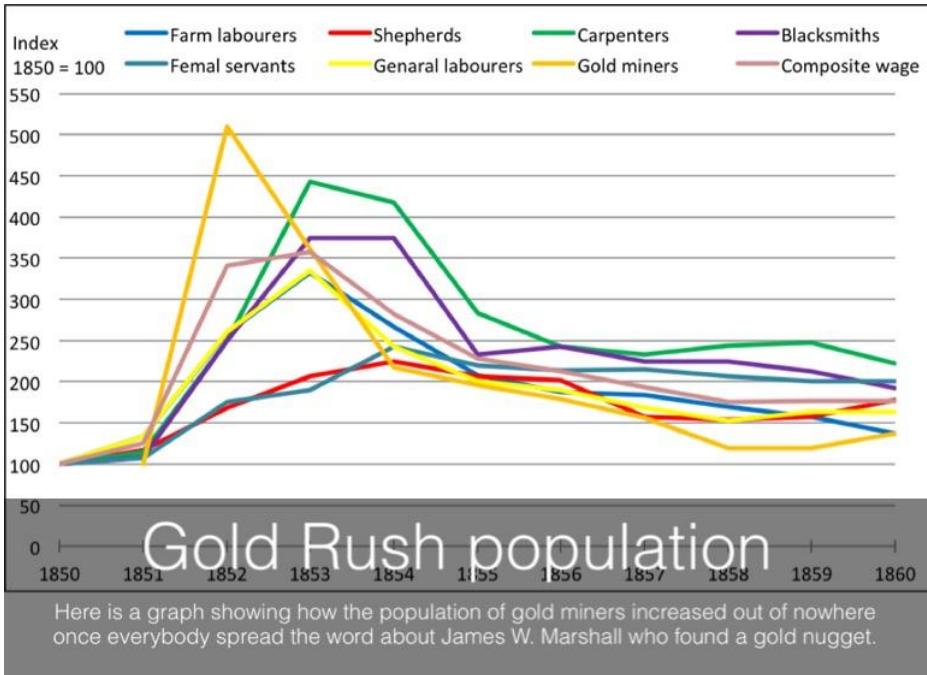
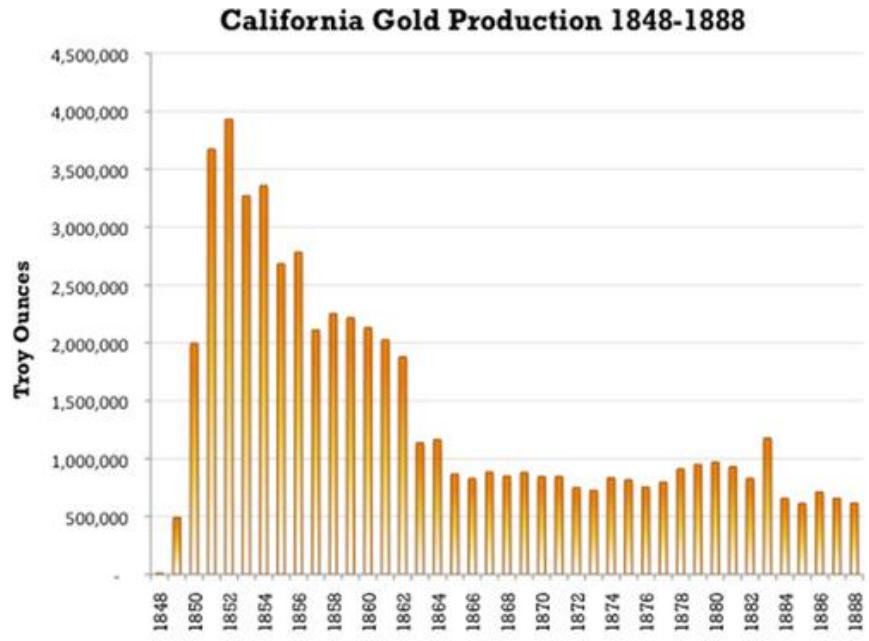
Can we model a city?



Can we model a city?



Can we model a city?



Can we model city growth?

Can we brain-storm here? What can we use?

1. Zipf's law:

Zipf's law implies that in any country the city with the largest population is generally twice as large as the next- biggest, and so on...

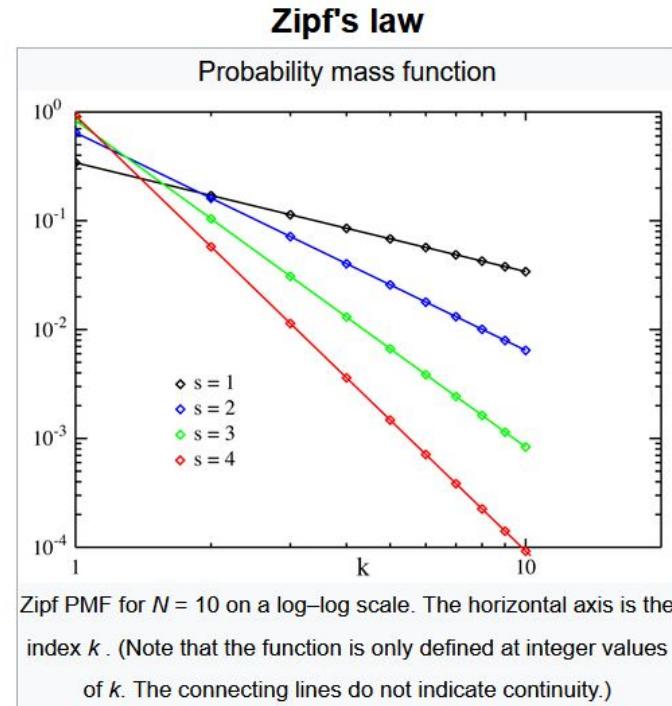
Discovered by q.linguist Zipf:

"the frequency of any word is inversely proportional to its rank in the frequency table."

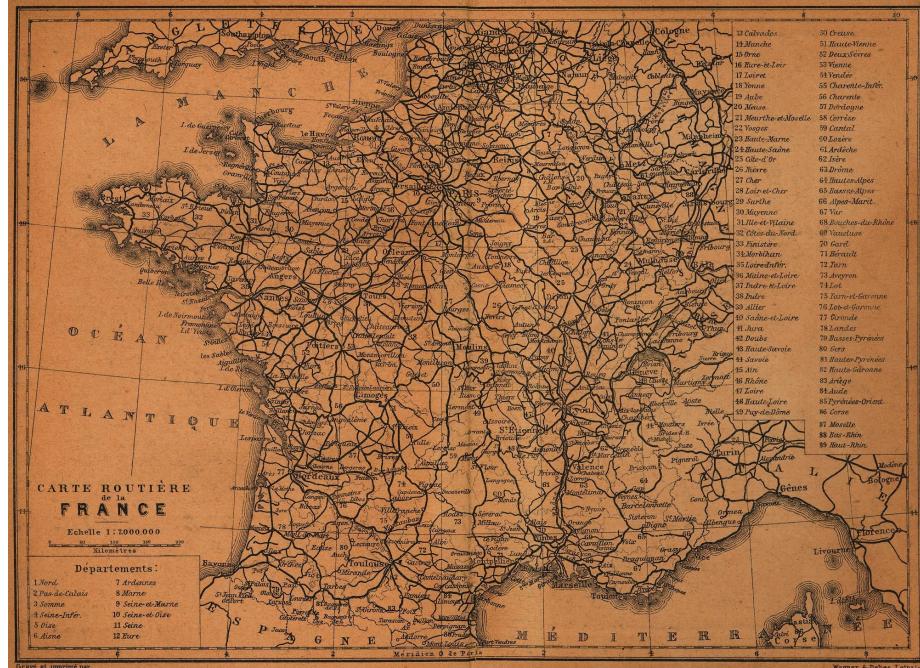
vs.

Gibrat's law...

2. Network growth models

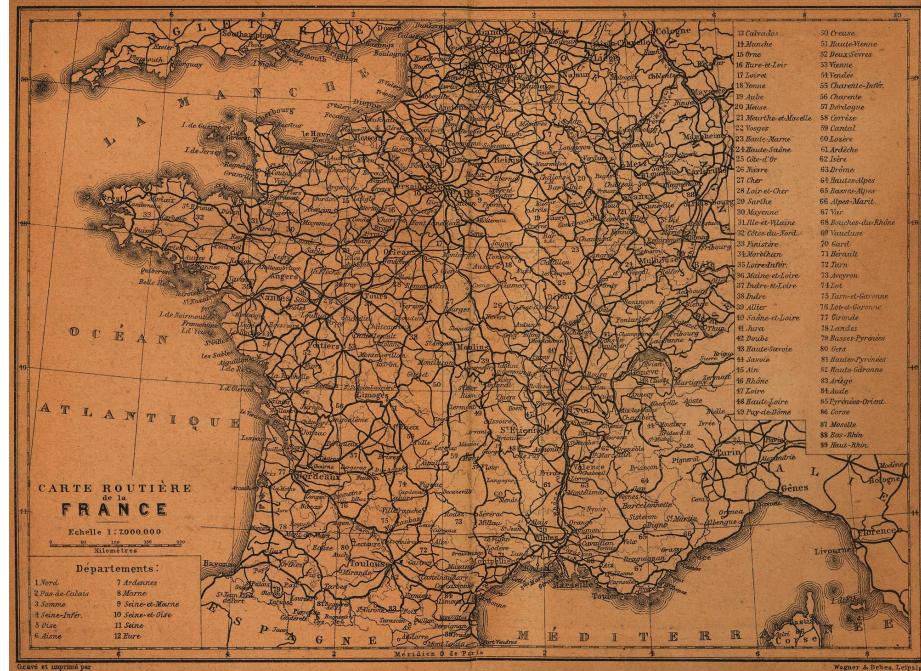


Can we model a city? Or “what is common between France and USA”?



“The growth equation of cities” Vincent Verbavatz, Marc Barthelemy
<https://arxiv.org/pdf/2011.09403>

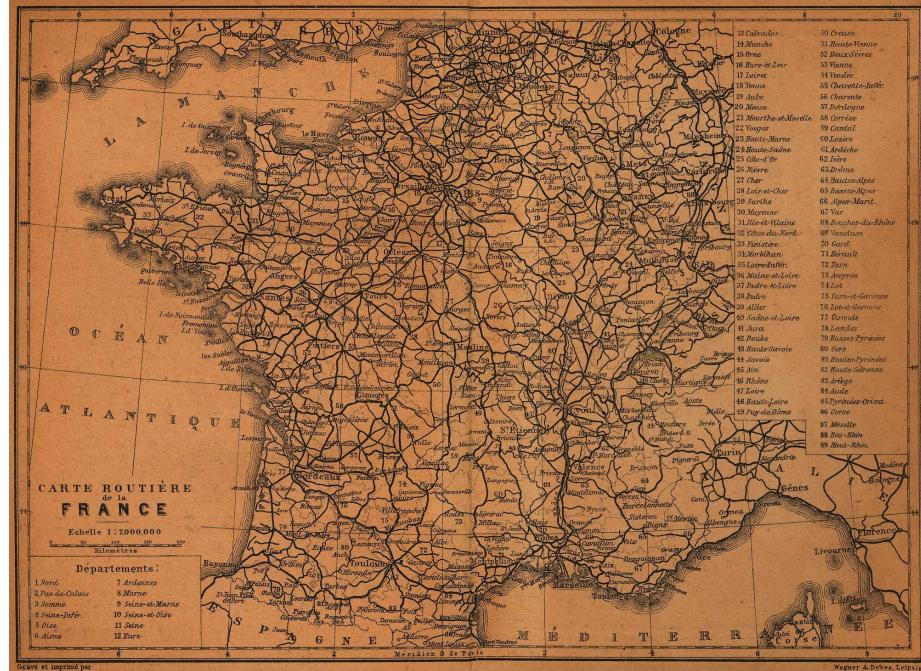
Can we model a city? Or “what is common between France and USA”?



Gibrat's law or the law of proportionate effect is a rule defined by [Robert Gibrat](#) in 1931 stating that the proportional rate of growth of a firm is independent of its absolute size.

Main take-home message
the city size distribution is log-normal.

Can we model a city? Or “what is common between France and USA”?



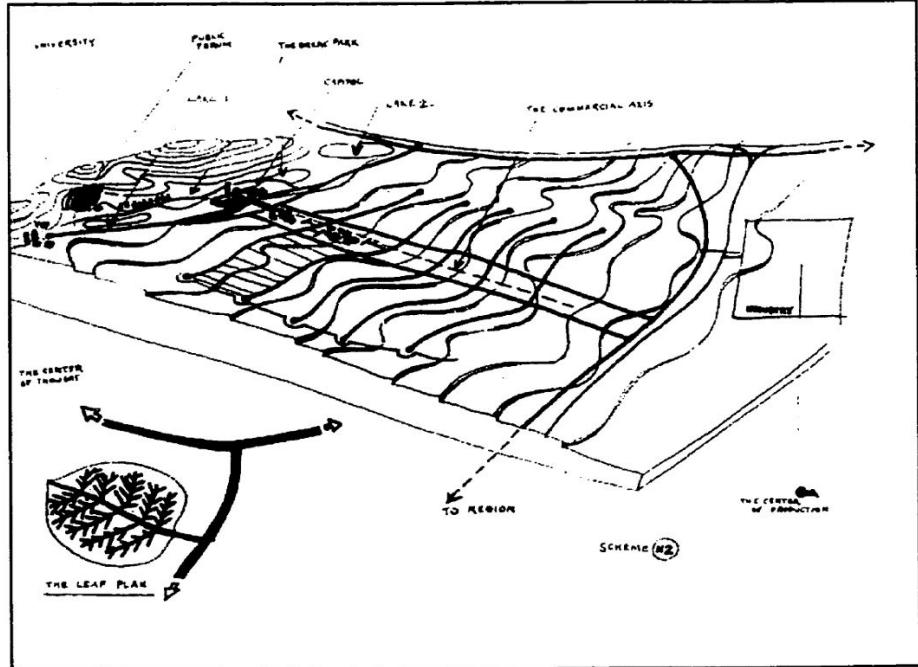
Main idea to incorporate time and space:

$$\partial_t S_i = \eta_i S_i + \sum_{j \in N(i)} J_{j \rightarrow i} - J_{i \rightarrow j}$$

“The growth equation of cities” Vincent Verbavatz, Marc Barthélémy
<https://arxiv.org/pdf/2011.09403>

City planning models

What can architects use to design more sustainable cities?



Why Le Courbousier may be wrong?

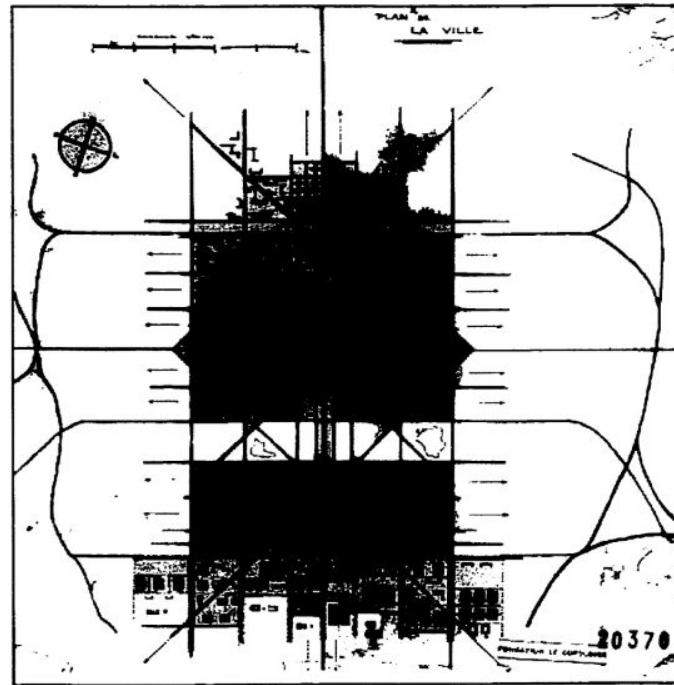
Can we brainstorm here? What can we use?

1. **Zipf's law:**

Zipf's law implies that in any country the city with the largest population is generally twice as large as the next-biggest, and so on...

2. **Network growth models** - does it bring new ideas?

Le Corbusier 'Response to Moscow'; the proposed city on the site of Moscow in relationship to the 'Green City' 1930, Application of the principles of Ville Radieuse, Foundation Le Corbusier, 20471, Jean-Louis Cohen; Le Corbusier and the Mystique of USSR, p. 142



Why Openstreetmaps?

...and presentation of osmnx python package

Why Citizen Science and openstreetmaps?

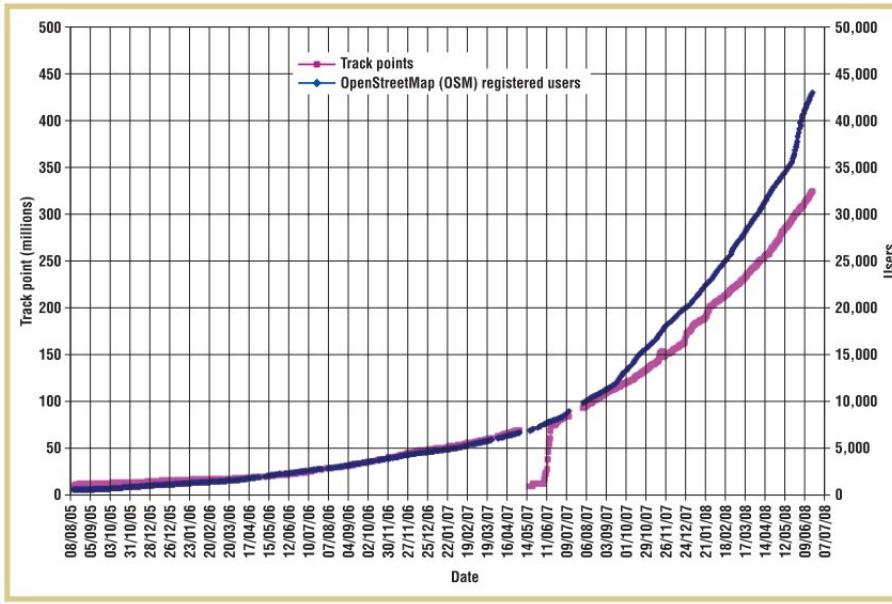


Figure 1. Graph of user and contributions growth to OSM on a monthly basis. The graph shows the accelerating growth in number of users and the rapid increase in data entry measured in track points (source: <http://wiki.openstreetmap.org>).

M.Haklay et al. "OpenStreet map: User-generated street maps" (2008)

Openstreetmaps analysis

Links to github

https://github.com/cityinteractionlab/openstreetmaps_osmnx_workshop

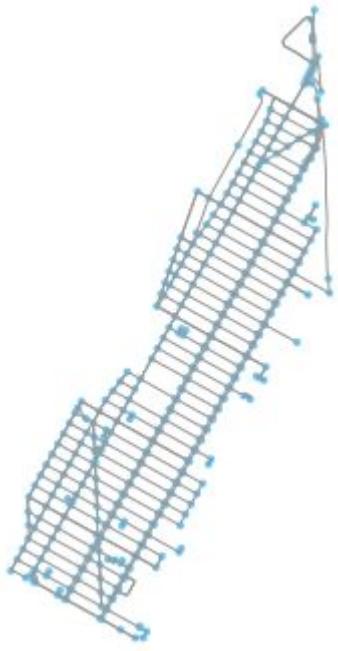
<https://github.com/gboeing/osmnx-examples>

https://github.com/cityinteractionlab/openstreetmaps_osmnx_workshop

Google colab

(careful, many packages required)

https://colab.research.google.com/drive/18aLg-YgrxMPUpR4iagZ5YnUnryi0_b30#scrollTo=KwoM_vordZ



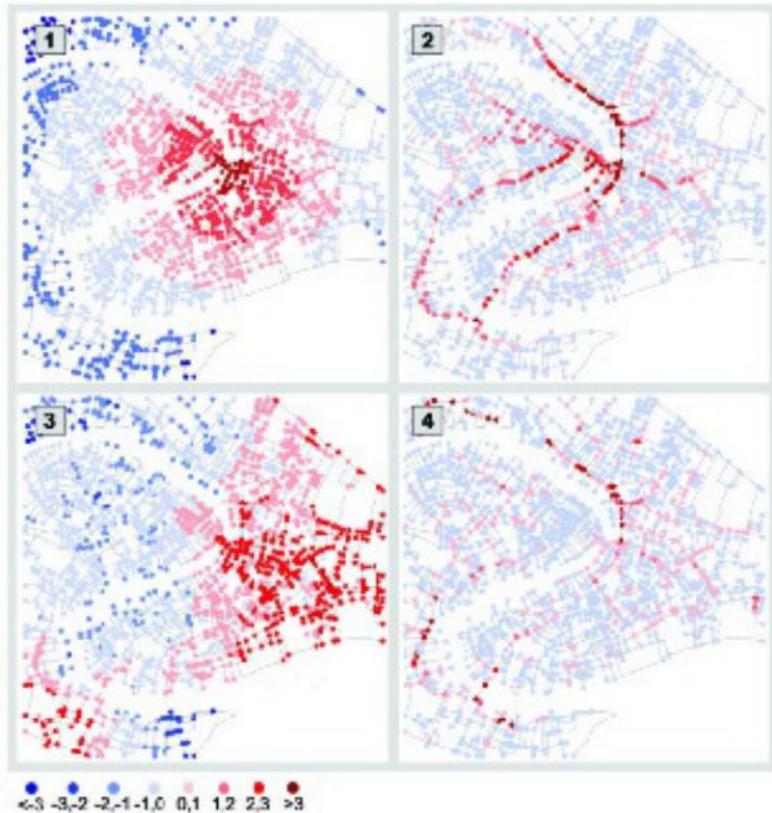
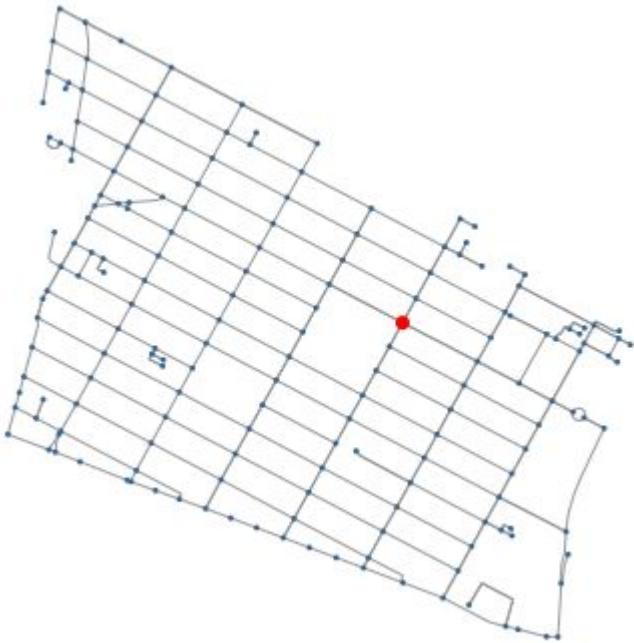


FIG. 24 Colour-coded maps for different centralities in Venice, Italy (see [102]). (1) Closeness, (2) Betweenness centrality, (3) Straightness, (4) Information (from blue to red the centrality increases). From [102].





City labs projects



Some information for city volunteers

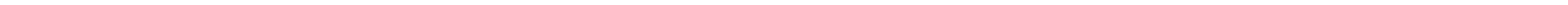
For Berlin <https://www.citylab-berlin.org/> and Paris

Mobility analysis

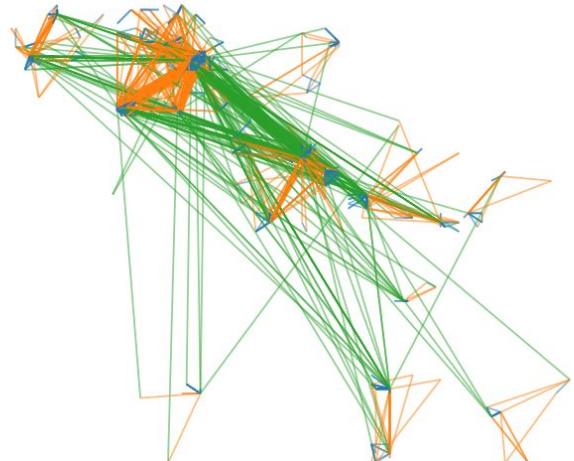
How can we do better than Uber?

Notebooks time:

1. Kepler.gl
2. Python spatial networks
3. Openstreetmaps



Mobility analysis



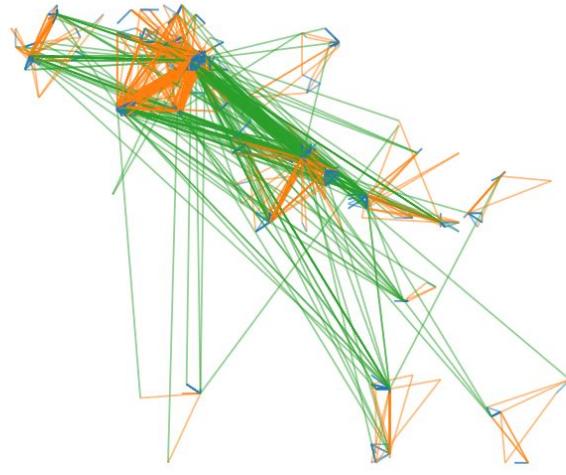
Animated Show grid Random seed

L

Number of levels

<https://observablehq.com/@ulfaslak/a-model-for-generating-multiscale-mobility-traces>

Mobility analysis



Animated Show grid Random seed

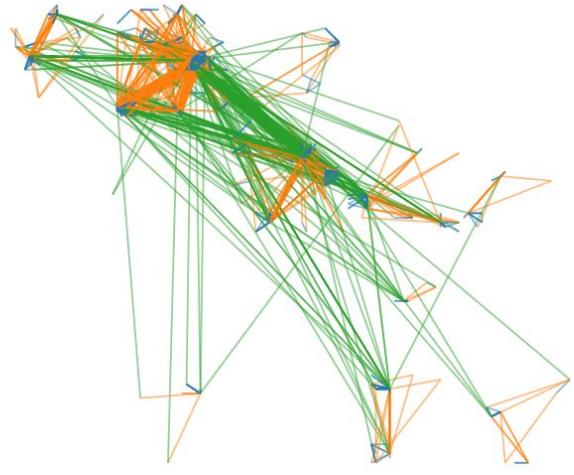
L
Number of levels

Mapbox and uber based design

<https://github.com/mapbox/>
Kepler.gl



Mobility analysis open data



Animated Show grid Random seed

L
Number of levels

3

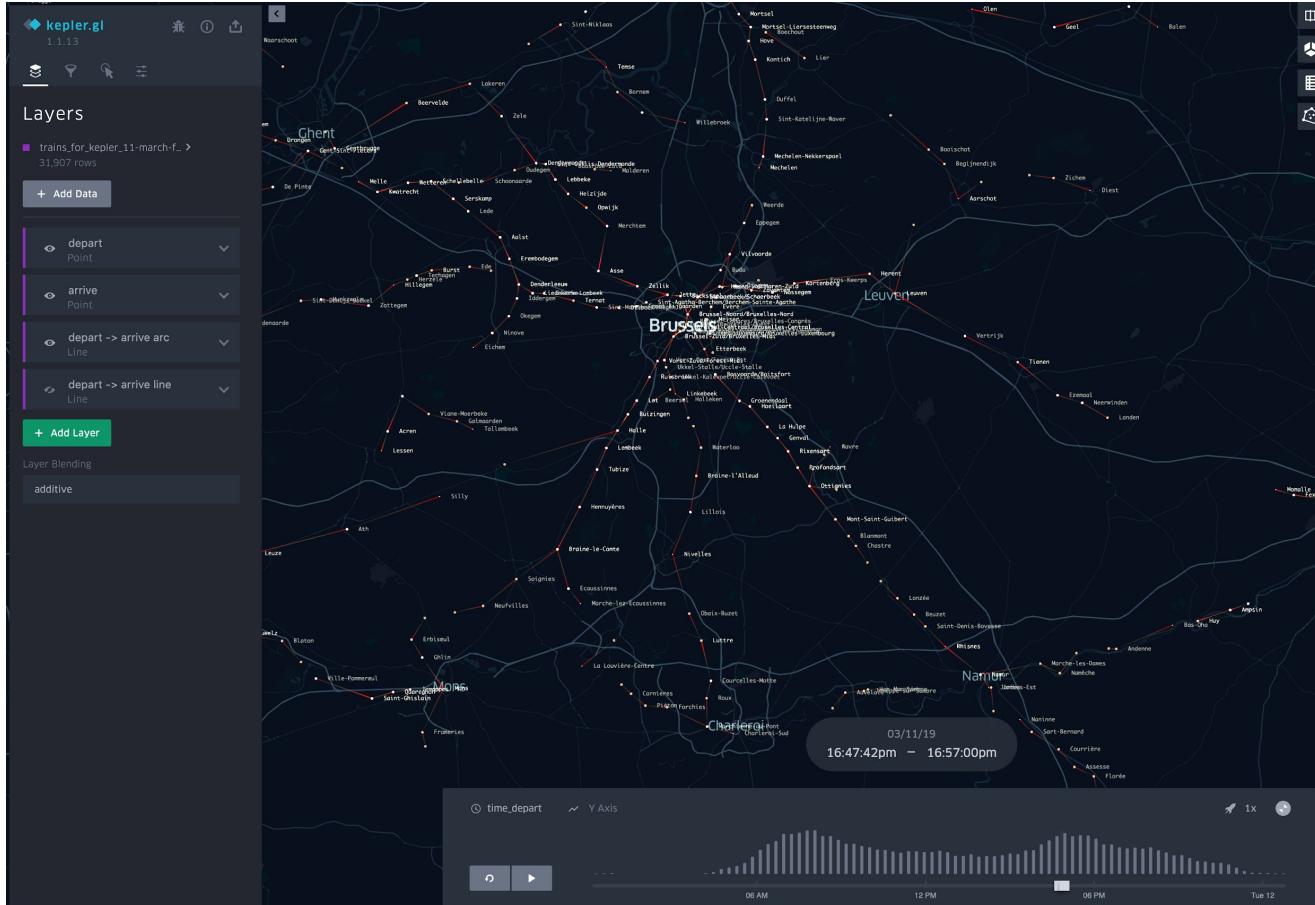


Open walking data <http://datawalking.com/>

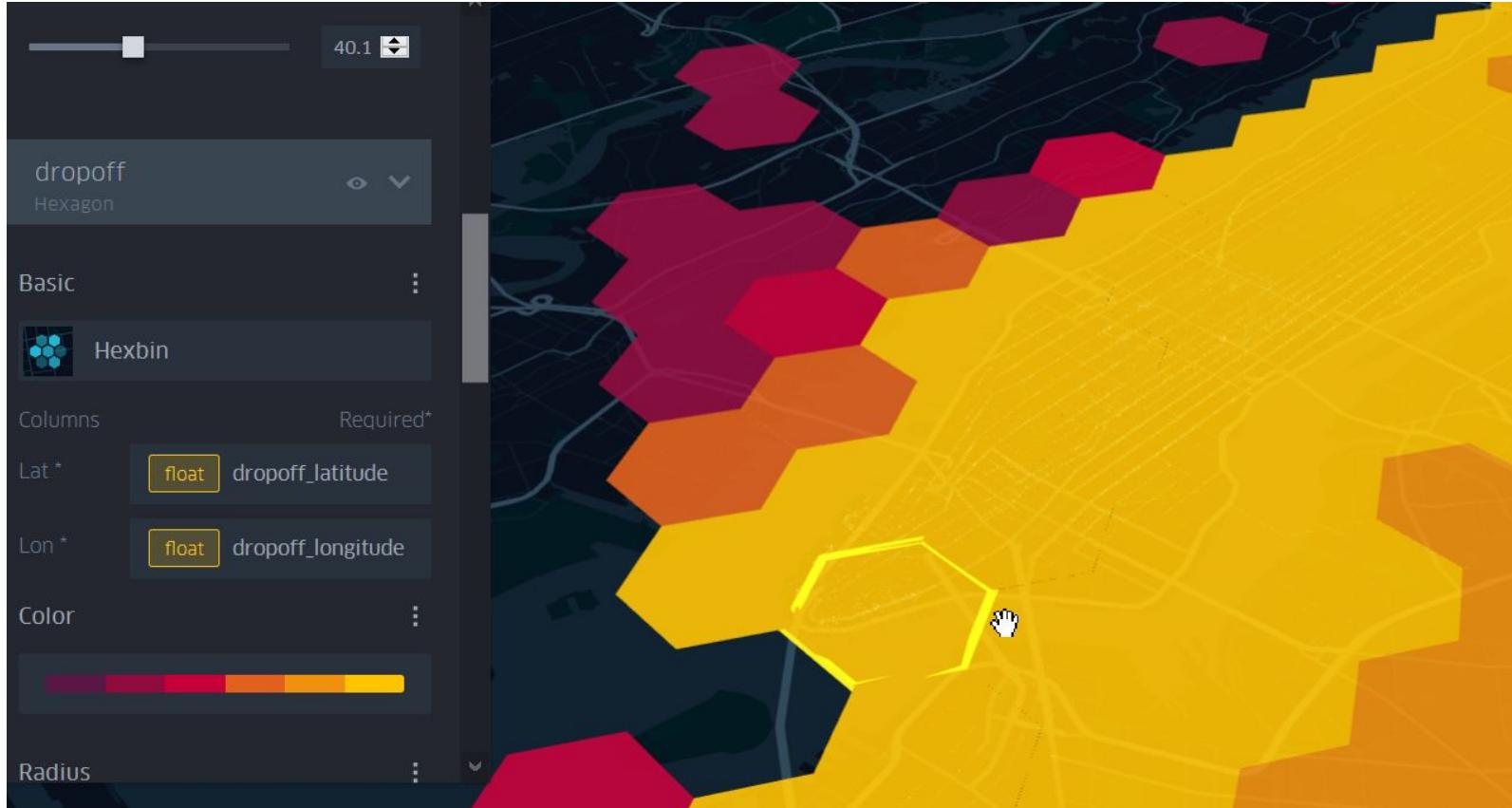
Open taxi data <https://github.com/uber-web/kepler.gl-data/blob/master/nyctrips/data.csv>

Open COVID data <https://github.com/google-research/open-covid-19-data>

Kepler.gl - geodata visualisation

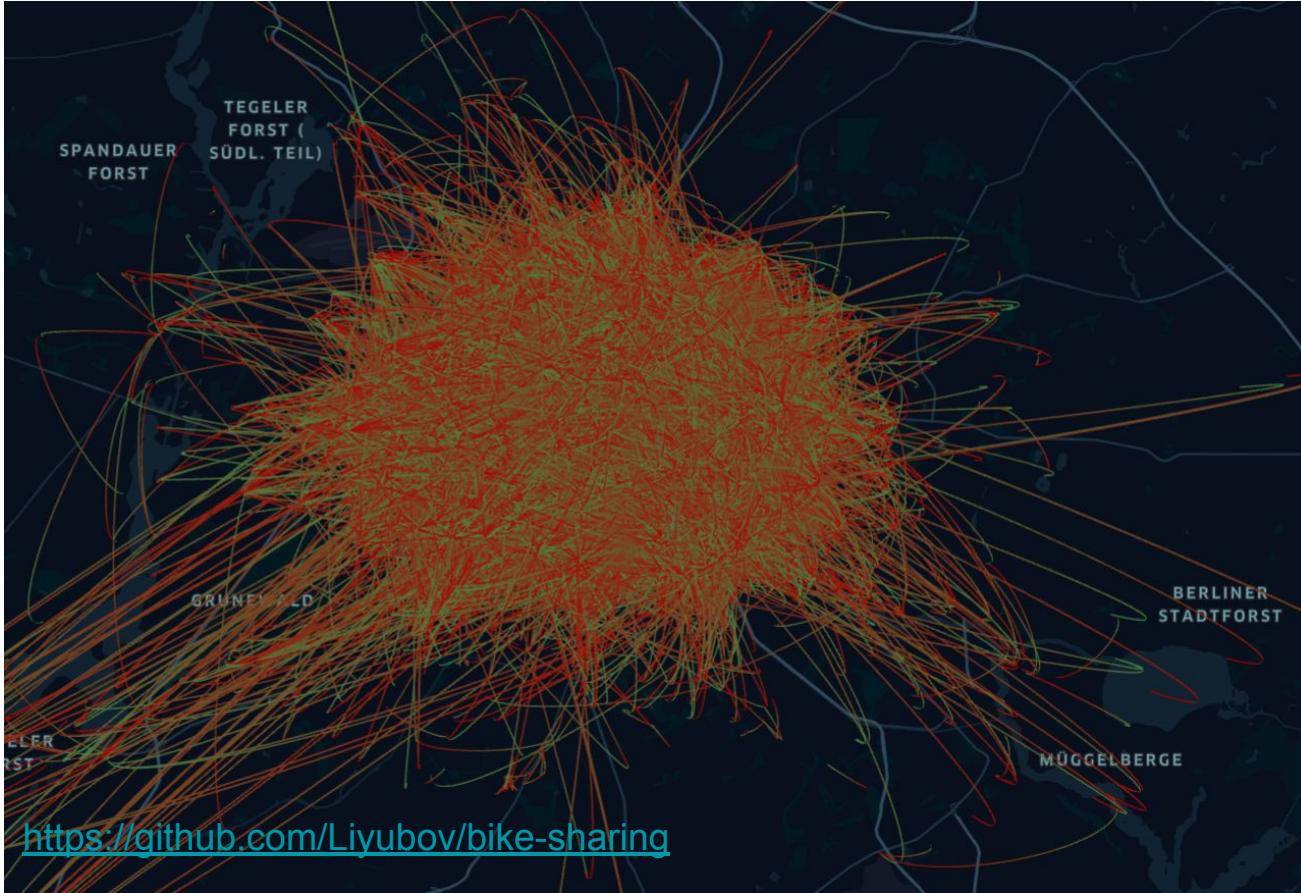


Kepler.gl - geodata visualisation



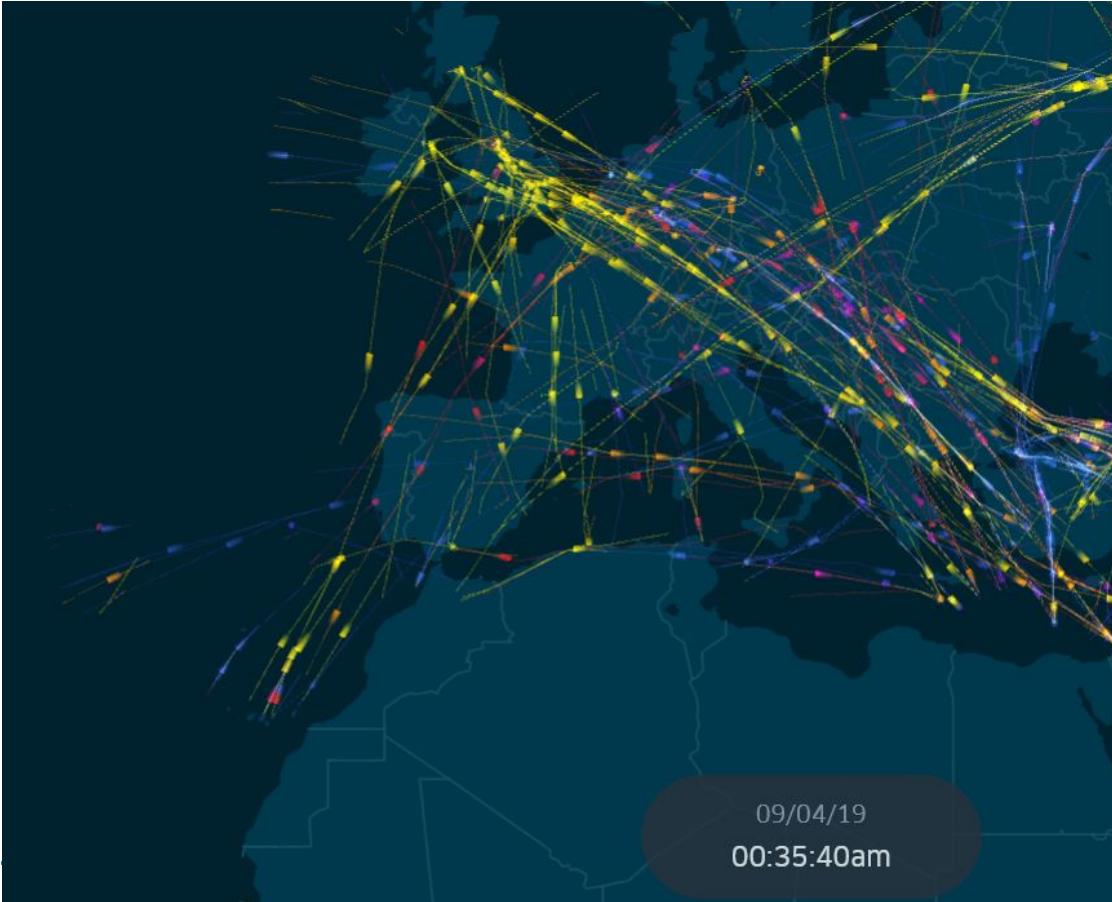
Open taxi data <https://github.com/uber-web/kepler.gl-data/blob/master/nyctrips/data.csv>

Kepler.gl - example visualisation



<https://github.com/Liyubov/bike-sharing>

Kepler.gl - example visualisation



Data visualisation of mobility data

Part I: statistics from mobility data,
application on open taxi data

Part II: data visualisation with kepler.gl,
python etc.



Transportational structures you see
from the airplane

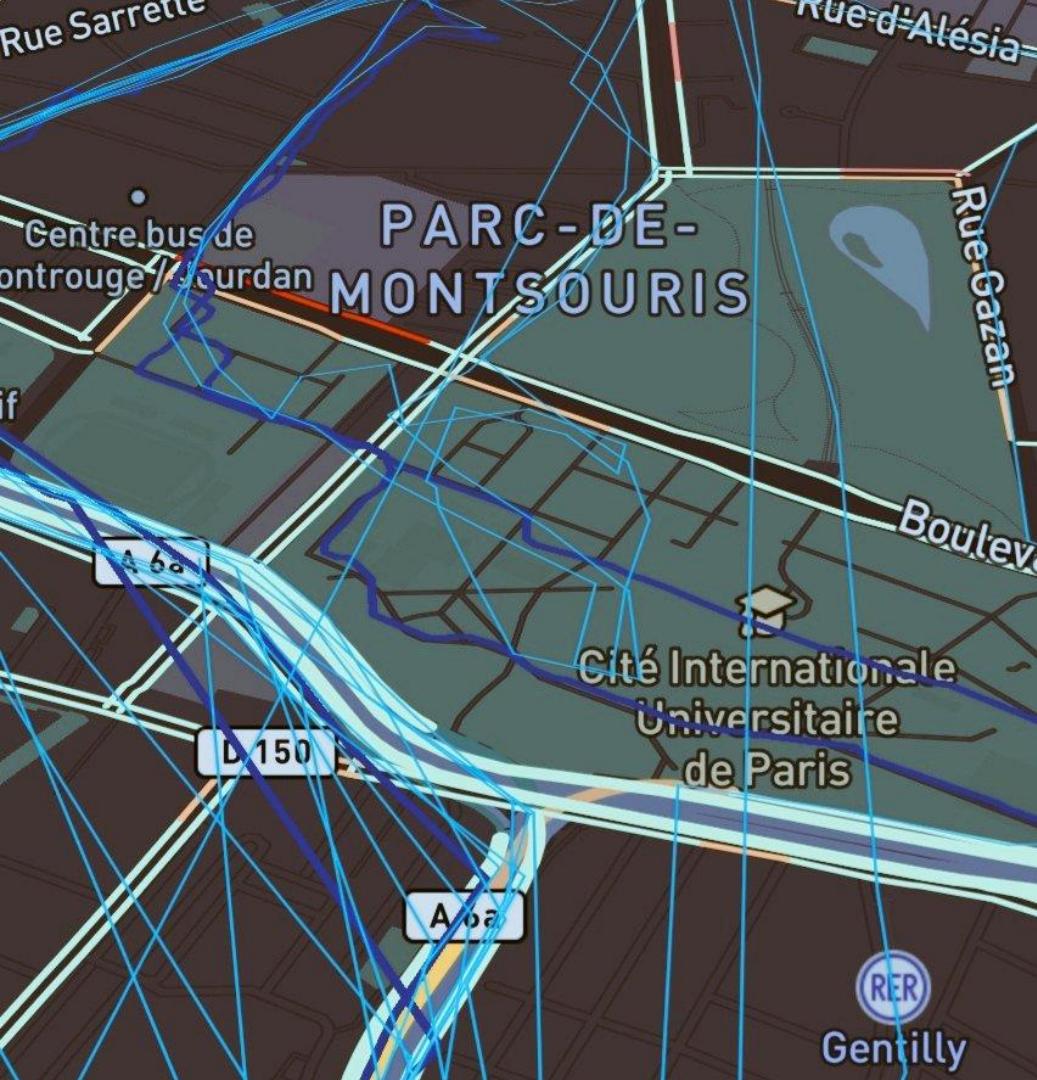


Notebooks

Github link

https://github.com/Liyubov/visualisation_transport_flow

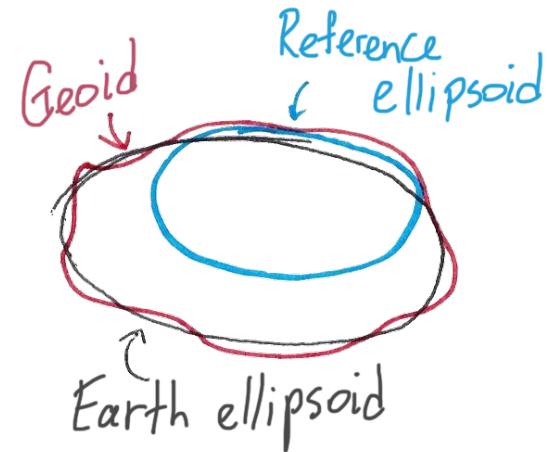
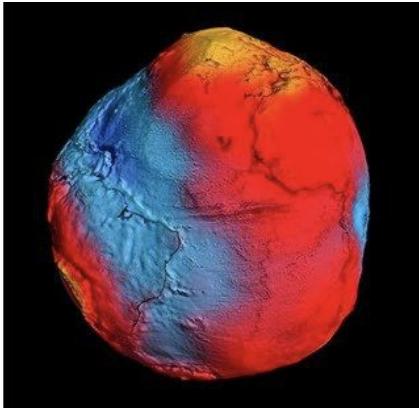
Figure from Move in Saclay app



Some important notes on mobility data



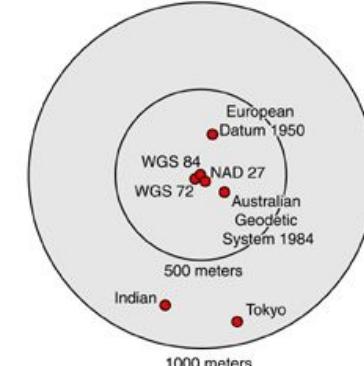
Earth is not (exactly) round!



Different coordinates systems:

WGS 84, NAD83, ED50, GRS80...

(used in GPS)



Practical aspects

Coordinates representation:

- (7.1757822, 46.6177318)
longitude latitude
- ex: GPS, Google
- $46^{\circ} 37' 06.9''$ N, $7^{\circ} 10' 36.0''$ E
deg min sec
- ex: GoPro



Conversion formula:

👎👎 46.6177318 ≠ 46.37069

💪💪 46.6177318 = 46 + 37/60 + 6.9/3600



GeoJson - universal format for geodata

Collection of featured geometrical objects.

Python: geojson (<https://python-geojson.readthedocs.io/>)
geopandas (<http://geopandas.org/>)
shapely (<https://shapely.readthedocs.io/>) -- libpysal

```
{  
    "type": "Feature",  
    "geometry": {  
        "type": "Point",  
        "coordinates": [125.6, 10.1]  
    },  
    "properties": {  
        "name": "You are here!"  
    }  
}
```

Point,
LineString,
Polygon,
MultiPoint,
MultiLineString,
MultiPolygon

Activity for 3rd hour: Try yourself session - choose activity

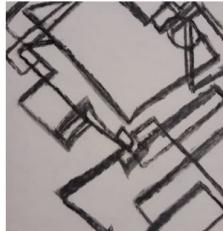
1. Download and consider mobility data
2. Look into resources on open mobility www.openhumans.org
3. Visualise your data with www.Kepler.gl based on deck.gl
4. Look into hackathons on open data and look into <https://liyubov.github.io/healthycityhack.github.io/>
https://github.com/useLiyubov/open_data https://github.com/correlaid-paris/citizen_science_inaturalist
5. Work with the humanitarian data
www.worldpop.org spatial population analysis and hdx platform
6. Add information to open tutorials
<https://github.com/an-medvedev/open-mobility-tutorial>



City Interaction lab

<https://sites.google.com/view/liubovkmatematike/city-interaction-lab?authuser=0>

City Interaction lab projects



Project 1

We are currently working together with our partners on analysis of citizen science data from around the globe. If you are happy, join us, fork repository

<https://github.com/cityinteractionlab/>

And explore the open data from iNaturalist and citizen science platforms like SciStarter Project is organised thanks to CorrelAid Germany and is lead by CorrelAid France (L.Bauer) together with UCL



Project 2

Project on Researchers mobility with Openhumans and CRI:

Have a look at related publications

<https://euraxess.ec.europa.eu/worldwide/australia-nz/survey-covid-19-will-have-huge-impact-researcher-mobility-and>

The webpage of participative project

<https://www.openhumans.org/activity/mobility-data-of-researchers/>

Open positions: coordinator and data visualisation manager
Join us

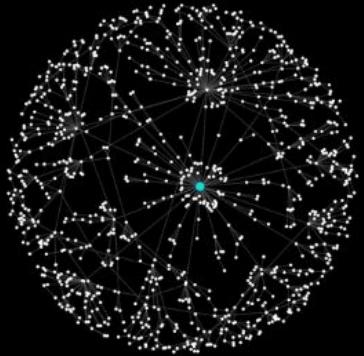
verainthecity@yahoo.com liubov.tupikina@cri-paris.org

PARIS, BERLIN, ONLINE
**NEO URBAN NETWORKS
WORKSHOP
2020**

NEO URBAN NETWORK WORKSHOP

JOIN

SCIENCE URBAN COMMUNITIES
OF NEW CITIES



**NOVEMBER
27, 2020**

VIRTUAL PAVILION
REGISTER FOR ZOOM LINK ATTENDENCE
[HTTPS://FORMS.GLE/DUV6SS2SN3TDIDAN7](https://forms.gle/DUV6SS2SN3TDIDAN7)

City Interaction lab events

27th November online

<https://sites.google.com/view/liubovkmatematike/city-interaction-lab/neo-urban-workshop?authuser=0>