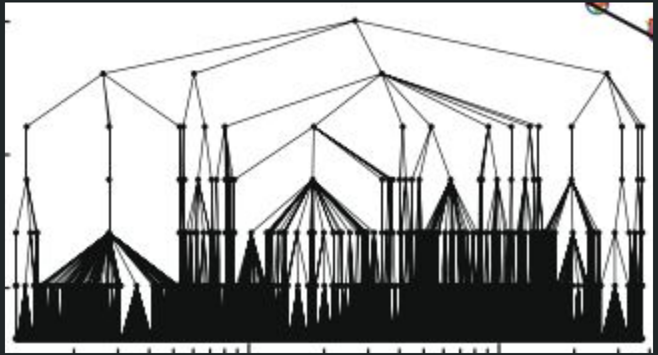
Modeling data with and without networks

Data course CRI 2020
Marc Santolini
Liubov Tupikina
Class 5



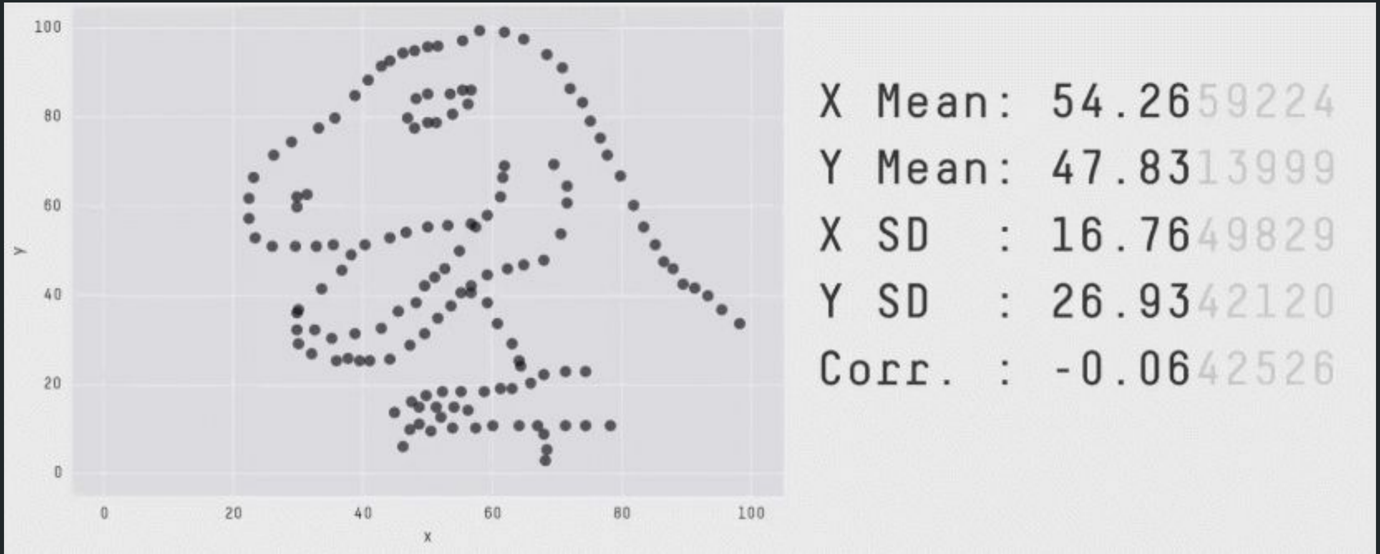
Outline of the day

1. Organisation of the week
2. Projects overview
3. Class 5 lecture
4. Notebooks
5. Take home messages round table

Resources of the day

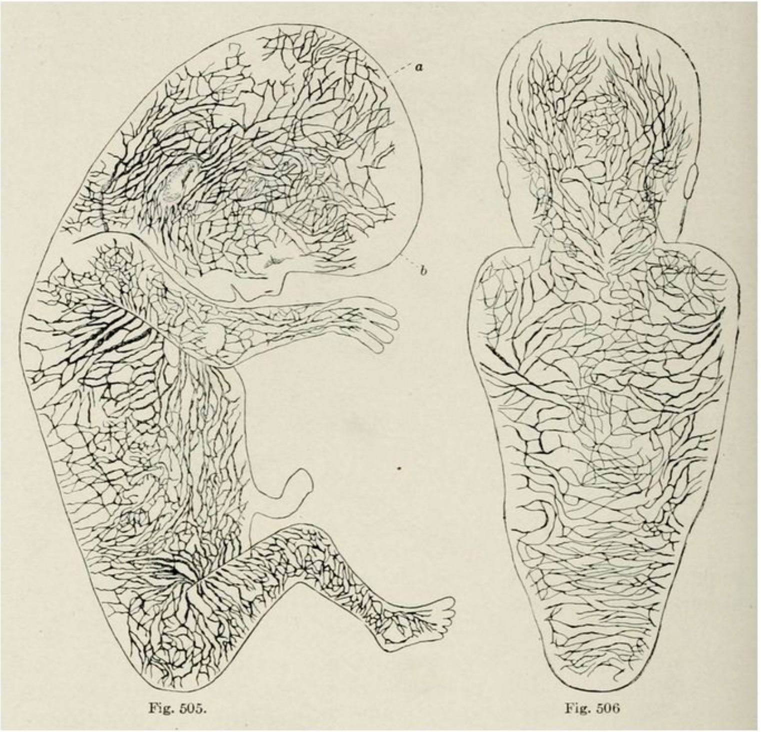
1. Network collection https://snap.stanford.edu/data
2. Networkx library

   https://networkx.org/documentation/networkx-1.9/examples



Modeling data

Class 2 reminder



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

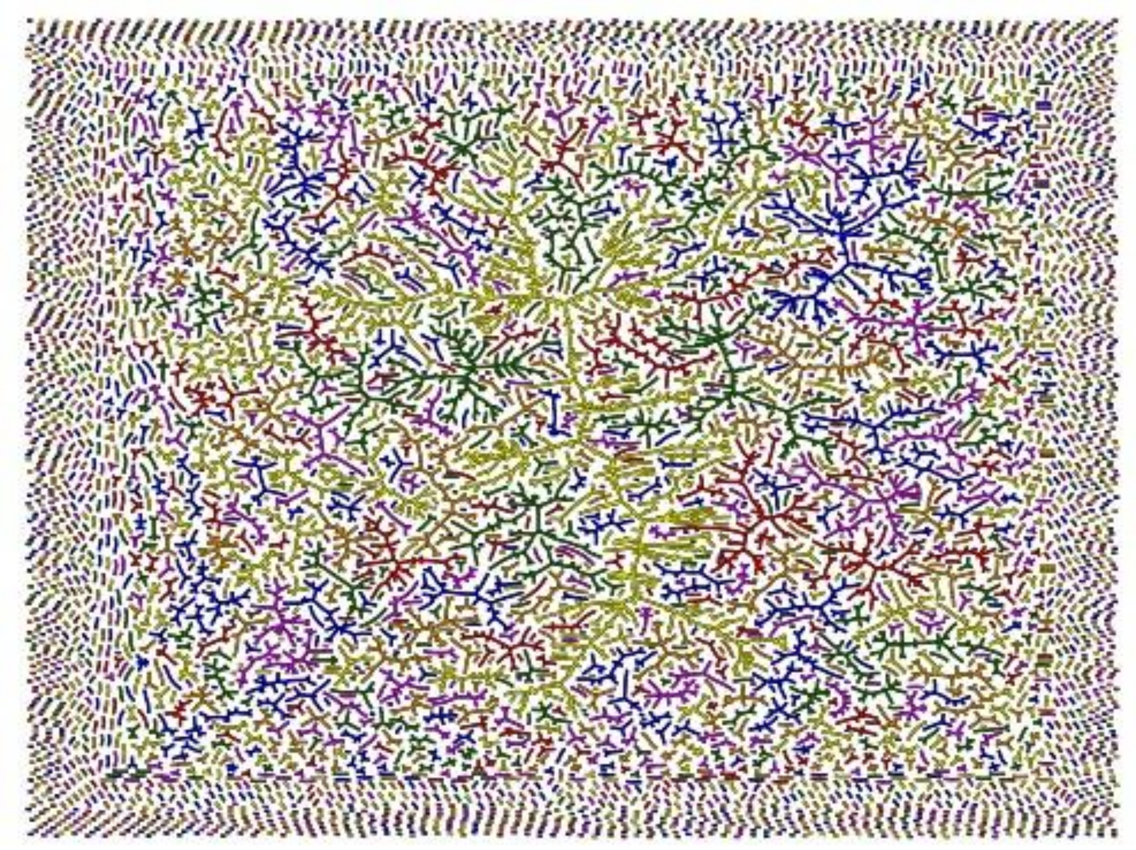
Class 2 reminder

Distribution of degrees
Distribution of other measures

Class 2 reminder

Small exercise
https://snap.stanford.edu/data
Choose a network and show the degree distribution

How random are real networks?



Modeling with networks

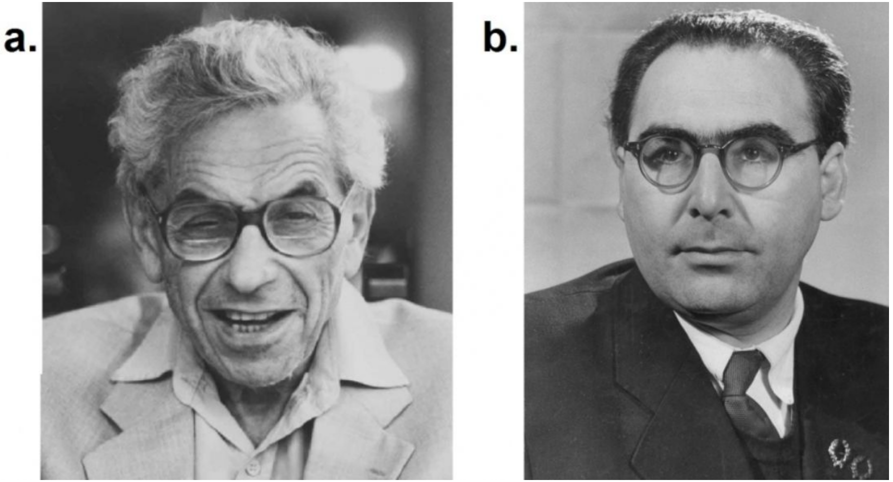
1. Random network models
2. Models of networks in time and space
3. Models of processes on networks

Modeling without networks

1. Random walks models
2. Skewed data models (epidemics spreading example)

Modeling with networks: random network models

1. Erdos Renyi
2. Barabasi Albert
3. Random geometric graph



Main resource

http://networksciencebook.com/chapter/3

Modeling with networks:
Erdos-Renyi random network models

Algorithm:

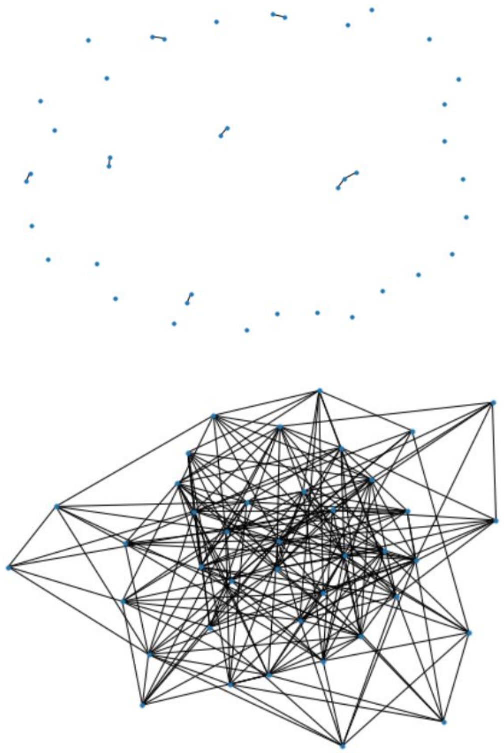
1. Choose any two nodes
2. Toss a coin
3. Draw a link if coin is up

Modeling with networks:
Erdos-Renyi random network models

Algorithm:

1. Choose any two nodes
2. Toss a coin (is it always fair?)
3. Draw a link if coin is up

Modeling with networks:
Erdos-Renyi random network model



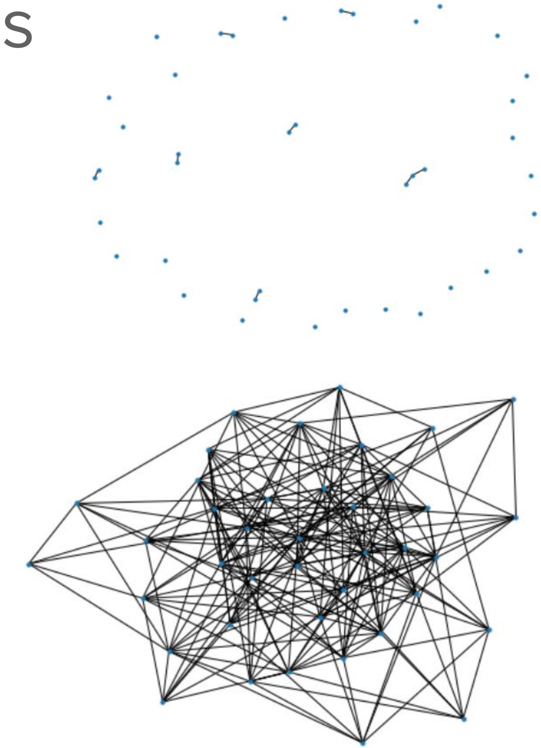
Algorithm:

1. Choose any two nodes from N nodes
2. Toss a coin (with probability p)
3. Draw a link if coin is up

Phase-transition threshold 1/<k>:
below which the network becomes fragmented,
while above which a giant connected component
of order N exists.

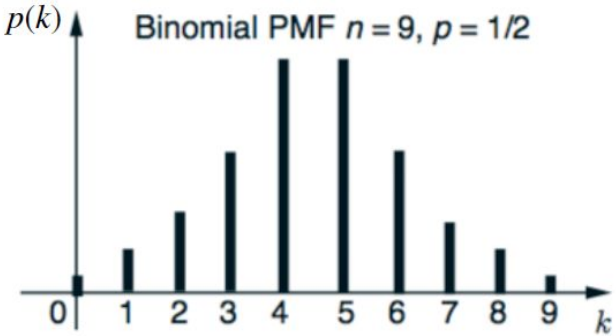
Modeling with networks:
Erdos-Renyi random network models

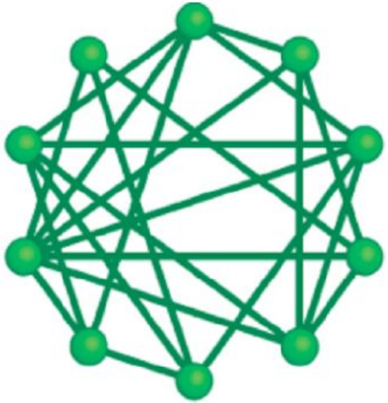
Which degree distribution do you expect?



Modeling with networks:
Erdos-Renyi random network models

Which degree distribution do you expect?





$p = 0.5$

Modeling with networks:
Barabasi-Albert random network models

How could we replicate social interactions?

www.wired.com

Modeling with networks: Barabasi-Albert random network models

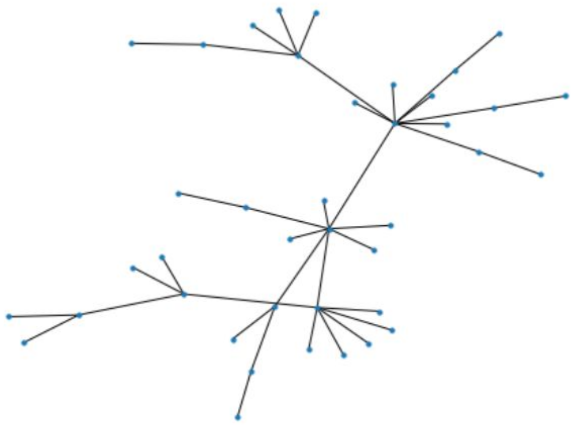
Algorithm:

1. Choose core of m connected nodes
2. Grow with preferential attachment
   to previous core nodes:
   Probability to connect is proportional
   to the number of links that the
   existing nodes
   Reference:
   Albert, Réka; Barabási, Albert-László (2002). "Statistical mechanics of complex networks" (PDF)

Modeling with networks:
Barabasi-Albert random network models

Algorithm:

1. Choose core of m connected nodes
2. Grow with preferential attachment to previous core nodes: Probability to connect is proportional to the number of links that the existing nodes



Reference:

Albert, Réka; Barabási, Albert-László (2002). "Statistical mechanics of complex networks" (PDF)

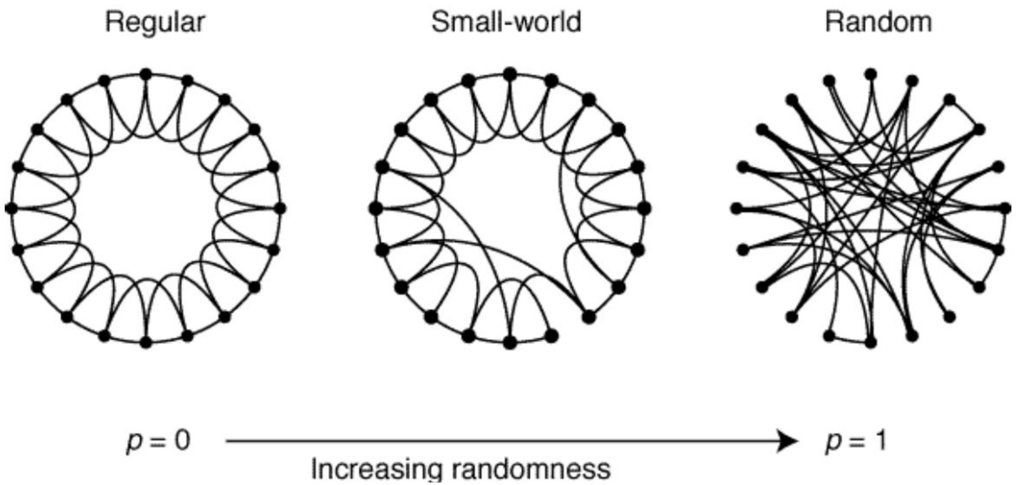
Modeling with networks:
Barabasi-Albert random network

Algorithm:

1. Choose core of connected nodes
2. Grow with preferential attachment
   To previous core nodes

Modeling with networks:
Watts-Strogatz random network

But does it replicate any random network?



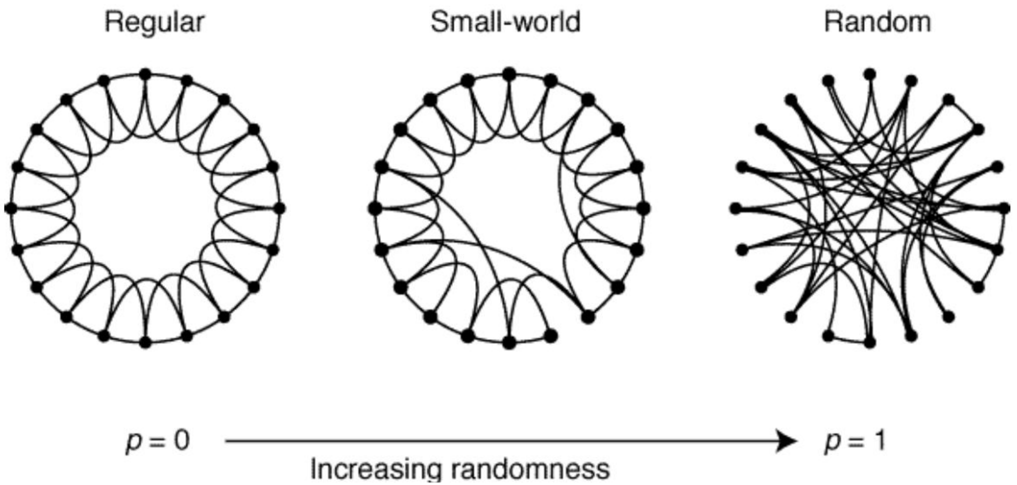
Modeling with networks: Watts-Strogatz random network

Algorithm:

1. Start from a regular k-circle-network
2. Rewire links from a regular circle

References:

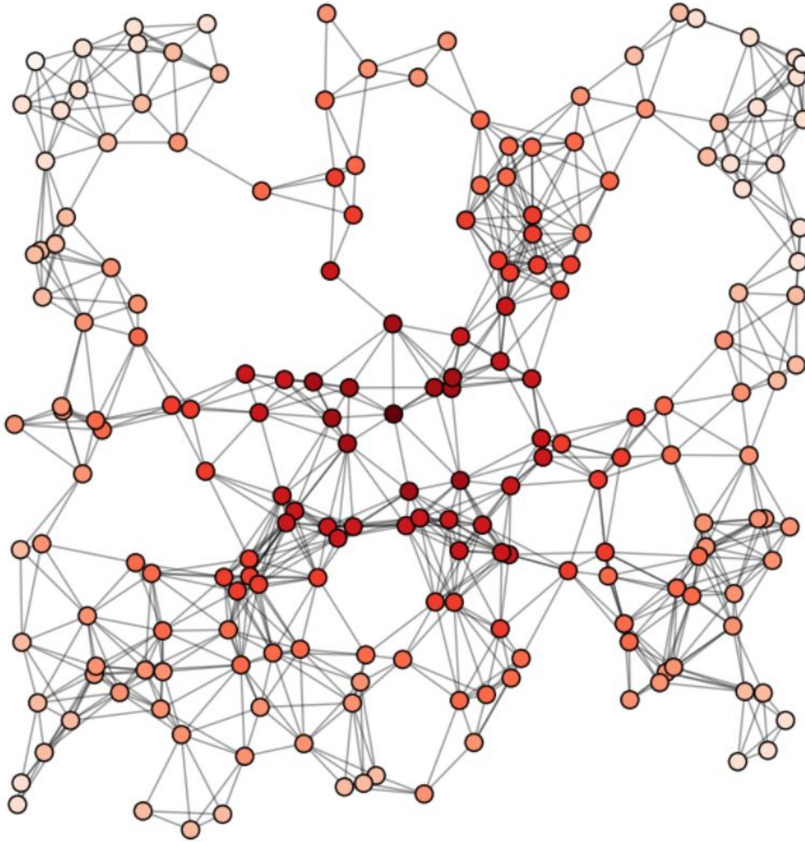
Watts, D. J.; Strogatz, S. H. (1998).
"Collective dynamics of 'small-world' networks" (PDF).



Modeling with networks: Geometric random network

Algorithm:

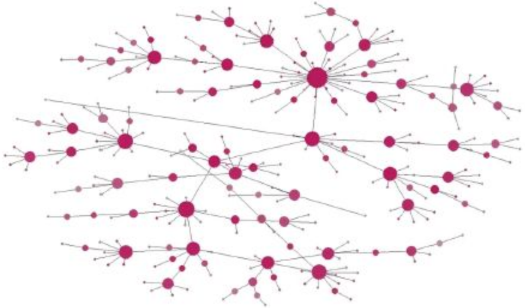
1. Choose any two nodes within R (radius)
2. Toss a coin
3. Draw a link if coin is up and nodes are close enough



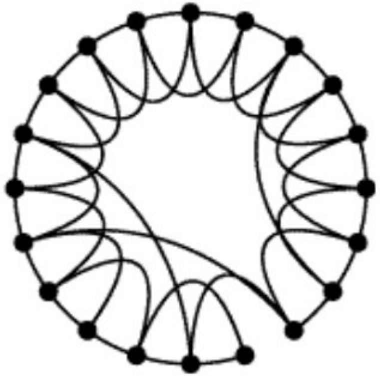
Take-home messages

Erdős-Rényi networks

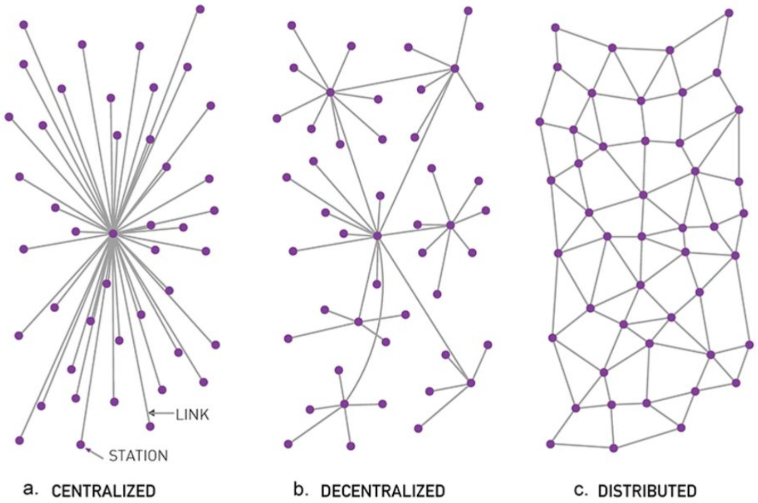
Barabási-Albert networks



Watts-Strogatz networks



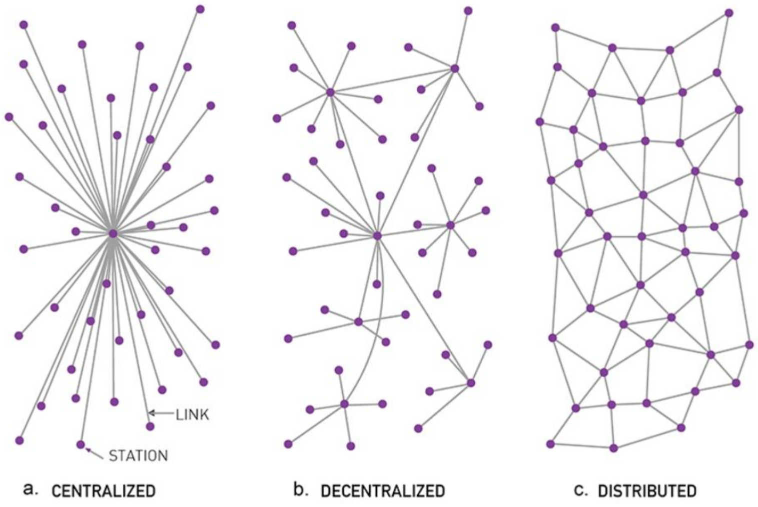
How to generate robust random network?



How to generate robust random network?

Ideas:

- Remove links from network
- Detect changes globally and locally
- Define measures for characterisation of network robustness

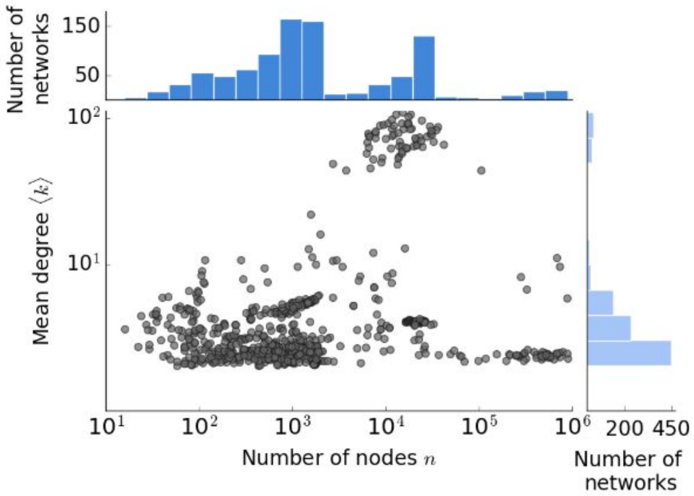


How do random networks help?

Main ideas:

- Not to model the exact network, but the global properties
- To grasp average characteristics

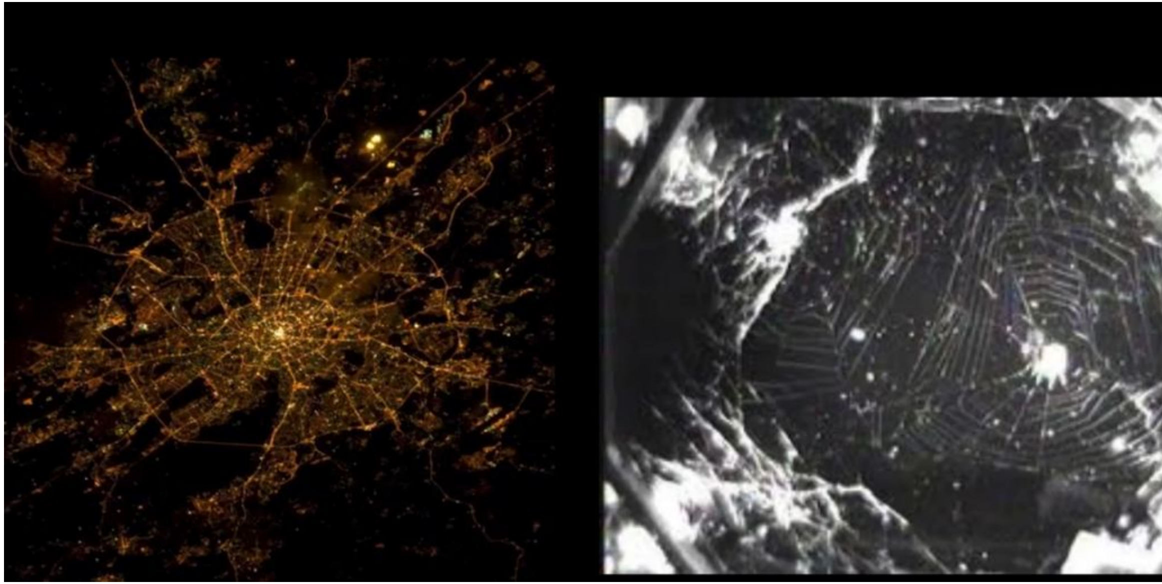
A.Broido et al. on scale-free networks



How do random networks help?

Main ideas:

- Not to model the exact network, but the global properties
- To grasp average characteristics



Look for analogies:
Web - Web ....

How do random networks help?

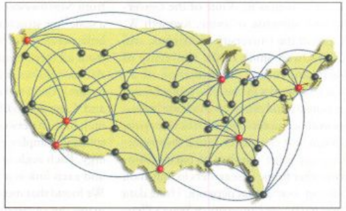
RANDOM VERSUS SCALE-FREE NETWORKS

RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

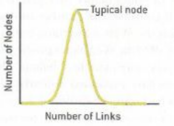
In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)—nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

Random Network

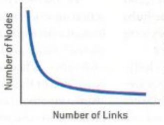
Scale-Free Network

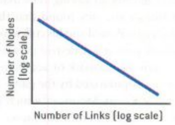


Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages





Specifically, a power law does not have a peak, as a bell curve does, but is instead described by a continuously decreasing function. When plotted on a double-logarithmic scale, a power law is a straight line [*see illustration above*]. In contrast to the democratic distribution of links seen in random networks, power laws describe

Scale-Free Networks Abound

OVER THE PAST several years, researchers have uncovered scale-free structures in a stunning range of systems. When we studied the World Wide Web, we looked at the *virtual* network of Web pages connected to one another by hyperlinks. In contrast, Michalis Faloutsos

some social networks are scale-free. A collaboration between scientists from Boston University and Stockholm University, for instance, has shown that a network of sexual relationships among people in Sweden followed a power law: although most individuals had only a few sexual partners during their lifetime, a few (the

PERSPECTIVE

Revisiting the Foundations of Network Analysis

Carter T. Butts

Network analysis has emerged as a powerful way of studying phenomena as diverse as interpersonal interaction, connections among neurons, and the structure of the Internet. Appropriate use of network analysis depends, however, on choosing the right network representation for the problem at hand.

The past decade has seen a dramatic surge of interest in the study of networks, with much of it in fields outside the "traditional" areas of mathematics, computer science, and the social sciences (*1*, *2*). By providing a formal mechanism for representation, measurement, and modeling of relational structure, the use of network analytic methods in these new domains (including the reductive nature of graphical structure that has facilitated its rich mathematical development (*3*) and associated scientific applications (*4*, *5*).

Extensions and relaxations of this basic framework designed to accommodate more complex situations are many and varied. We may avoid the assumption of dichotomous relationships by allowing edges to carry different weights [such as

to make. The biologist's method of defining potential feeding sites will greatly influence the structure of the interaction network.

The basic problem is the definition of the class of distinct entities on which one's relation of interest will be defined. The mere act of positing such a class, of course, smuggles in the tacit assumption that such a class can be defined (and moreover, that it is scientifically useful to do so). The choice of individual humans as nodes in studies of friendship (*11*) or kinship (*12*) networks and the use of individual publications in citation studies (*13*) are examples in which this assumption is well-justified. On the other hand, studies of interactions between aggregates such as groups (*14*), households (*15*), or organizations may encounter problems due to the fluidity of the interacting units and the fact that subunits of a larger unit may themselves interact with others both within and without the "parent."

As in the biological example, collapsing all potentially interacting elements into a single unit

Pushing Networks to the Limit

PERSPECTIVE

to make. The biologist's method of defining po-
tential feeding sites will greatly influence the

Revisiting the Fou
of Network Analy

Carter T. Butts

Network analysis has emerged as a powerful
as interpersonal interaction, connections am
Appropriate use of network analysis depends
representation for the problem at hand.

The past decade has seen a dramatic sur
of interest in the study of networks, w
much of it in fields outside the "tradition
areas of mathematics, computer science, and t
social sciences (*1*, *2*). By providing a formal mec
anism for representation, measurement, and mc
eling of relational structure, the use of netwo
analytic methods in these new domains (includi

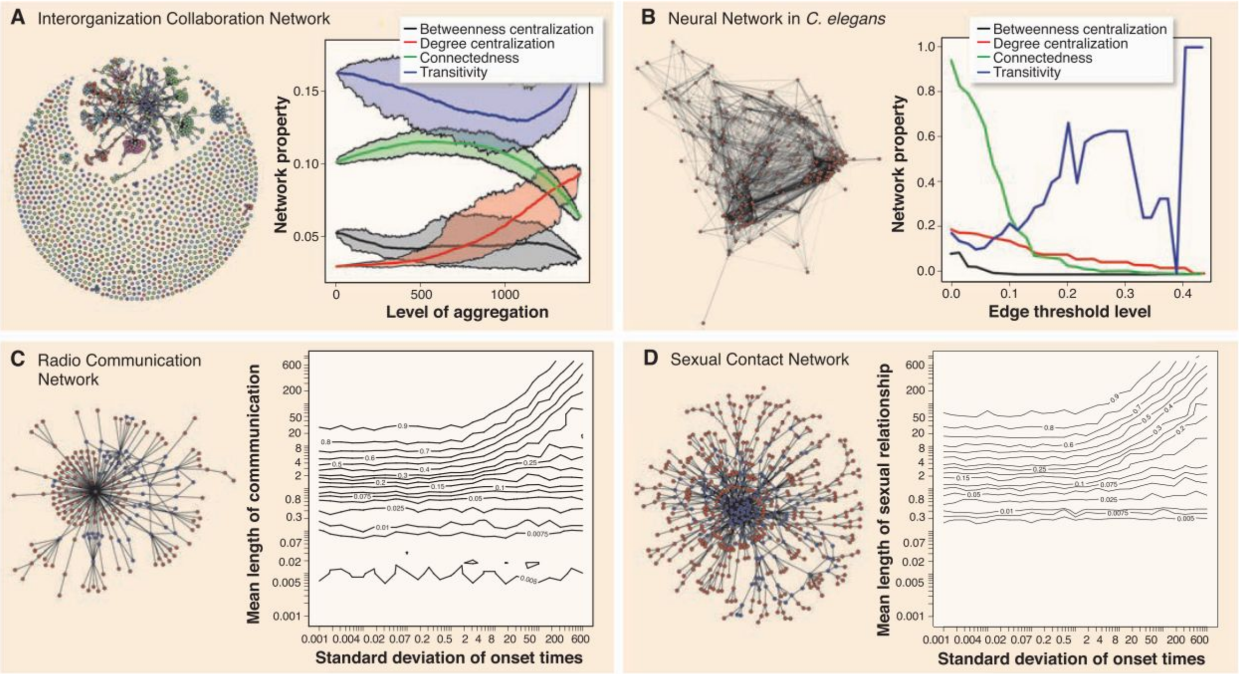


Fig. 1. Effects of changing definitions of "node" and "edge." **(A)** Network of interorganizational collaboration in the first 13 days of the Hurricane Katrina response (*39*) illustrates potential consequences of node aggregation; edges different structural properties (right). (**C**) and (**D**) show the effects of edge timing, depicted as contour plots (right), in systems in which the edges are not static; each line represents the fraction of the population reached by the diffusion process.

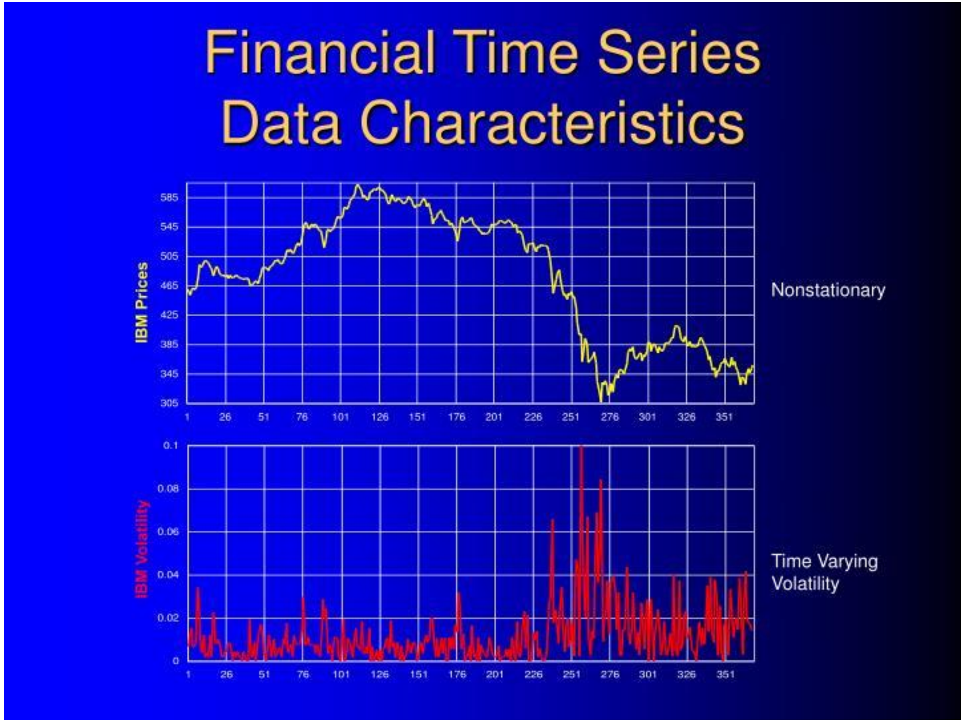
How to generate random network?

1. Random network models
2. Models of networks in time and space
3. **Models of processes on networks and beyond**

Random walks models

Main ideas:

- Not to model the exact time-series, but the global properties
- To grasp average characteristics



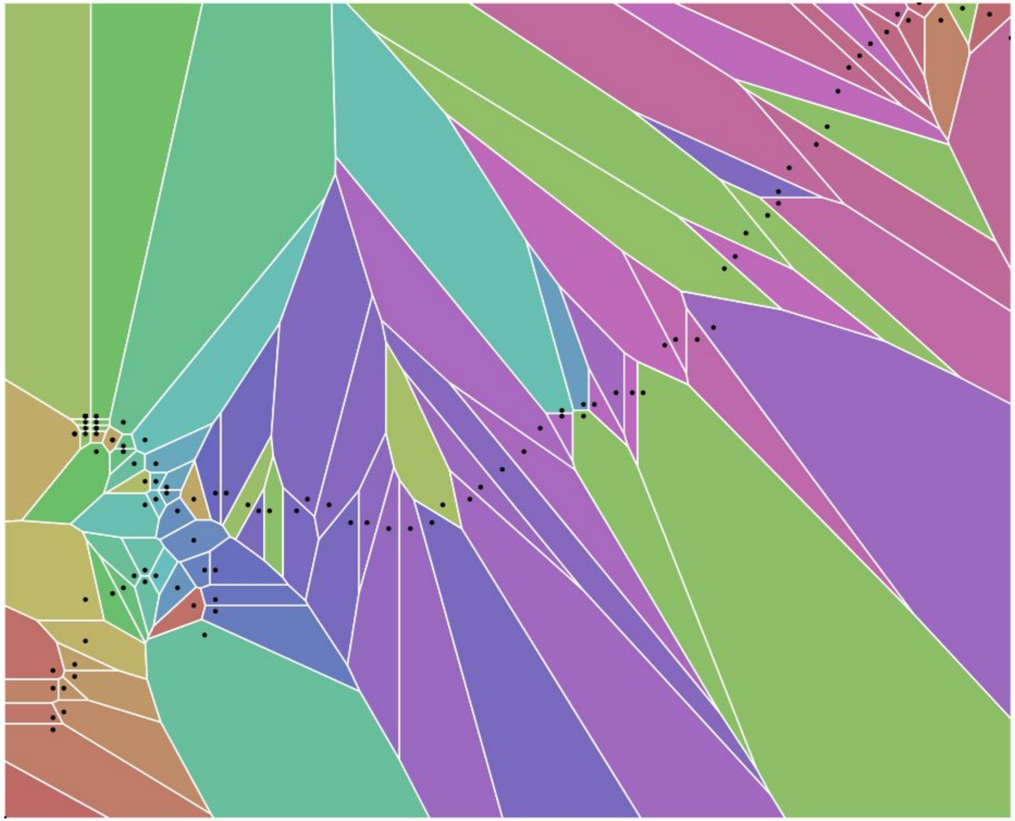
Klafter et al. on “First steps in random walks”

Random walks models

Main ideas:

- Not to model the exact time-series, but the global properties
- To grasp average characteristics

Klafter et al. on "First steps in random walks"
Data walking visualisations
https://github.com/Liyubov/DATA_WALKING

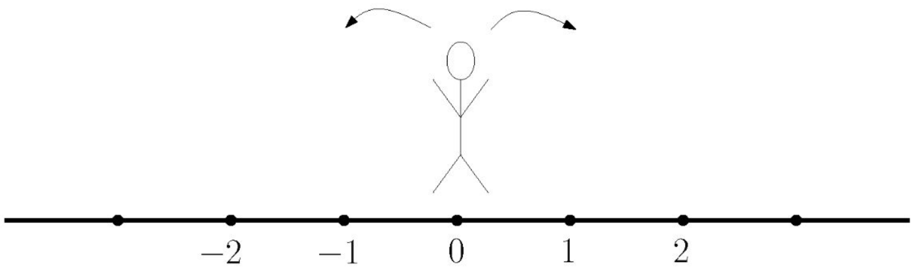


Can simple random walk model help?

Main ideas:

- Not to model the exact process but the global properties
- To grasp average characteristics

For more processes modeling go to https://github.com/Liyubov/heterogeneous-dynamics-on-networks



Network resources

http://networkrepository.com/networks.php
http://networksciencebook.com/chapter/3#advanced-b

Network resources

http://networkrepository.com/networks.php
http://networksciencebook.com/chapter/3#advanced-b

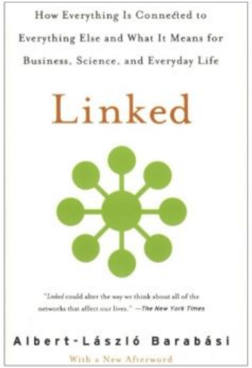
Spatial Networks

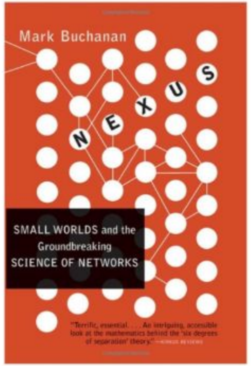
Marc Barthélemy*

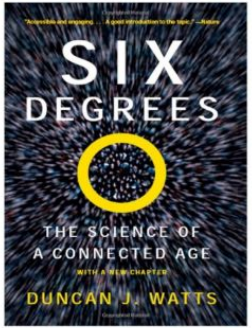
Institut de Physique Théorique, CEA, IPhT CNRS, URA 2306 F-91191 Gif-sur-Yvette France and Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 CNRS-EHESS) Ecole des Hautes Etudes en Sciences Sociales, 54 bd. Raspail, F-75270 Paris Cedex 06, France.

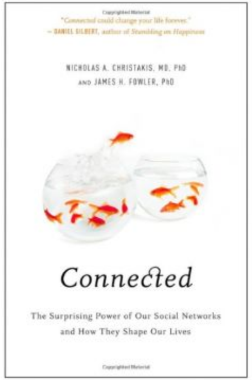
Complex systems are very often organized under the form of networks where nodes and edges are embedded in space. Transportation and mobility networks, Internet, mobile phone networks, power grids, social and contact networks, neural networks, are all examples where space is relevant and where topology 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





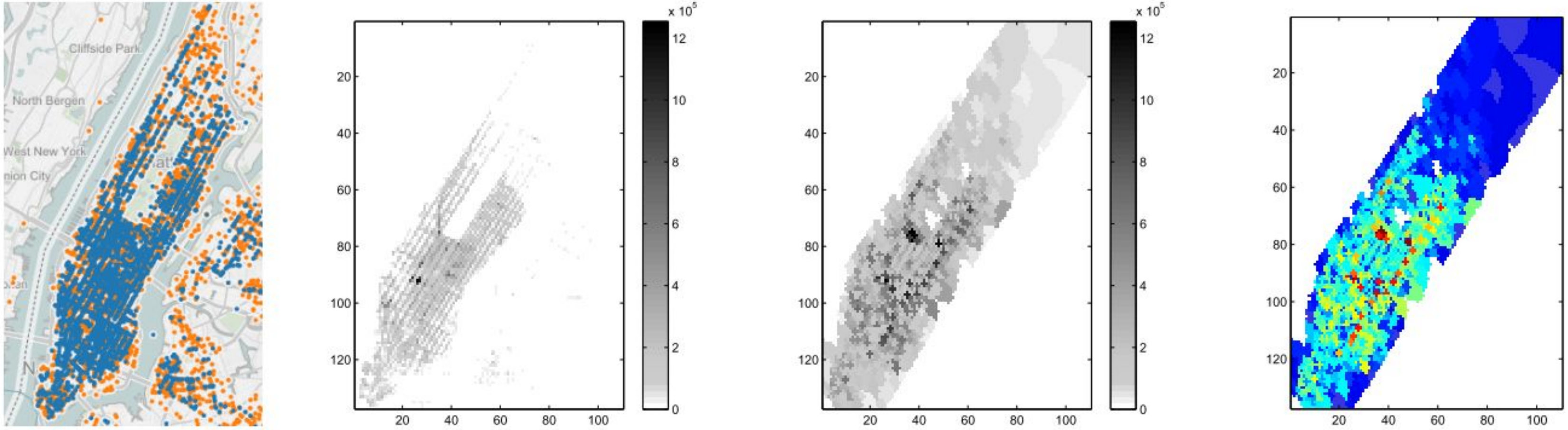




Jupyter notebooks:

1. Random graphs
2. Spatial network modeling

Models of transport (jupyter notebook with Chakresh on taxi moves)



(a) (b) (c) (d)

Jupyter notebooks and resources

https://classroom.google.com/u/0/w/MjA3ODcyNzc4NDYy/t/all