Why networks?

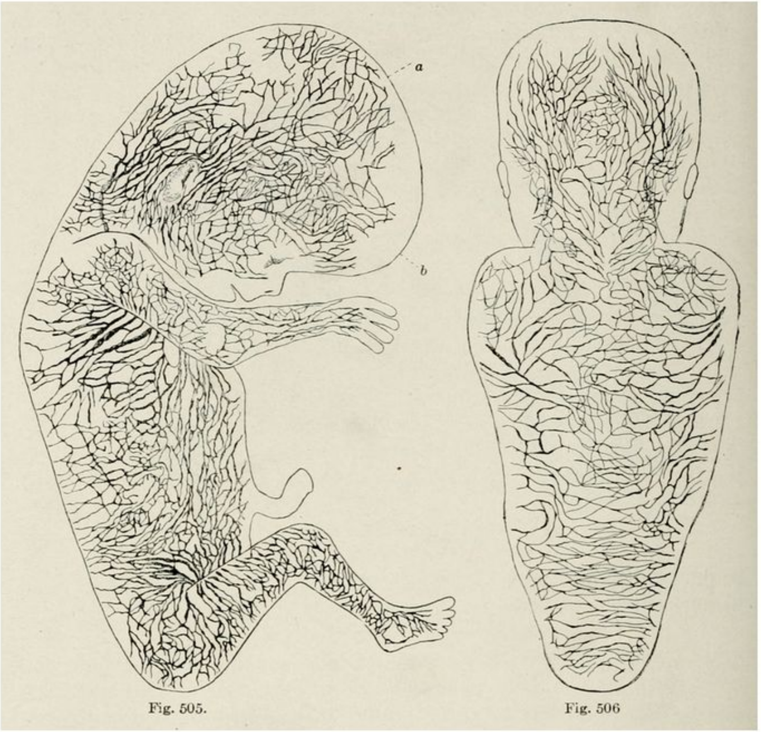
Data course CRI 2020
Marc Santolini
Liubov Tupikina
Class 2

Outline of the day

1. Projects overview
2. Class 2 lecture
3. Notebooks
4. Take home messages round table

Why networks?

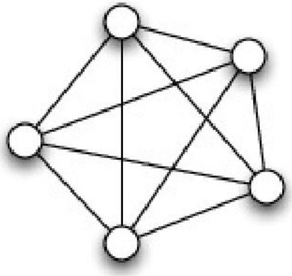
Where you can find them?



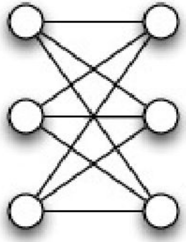
Distension of the lymphatic vessels in the human foetus, from Franz Kreibel, *Manual of human embryology*, 1910

Soft Matter and graphs

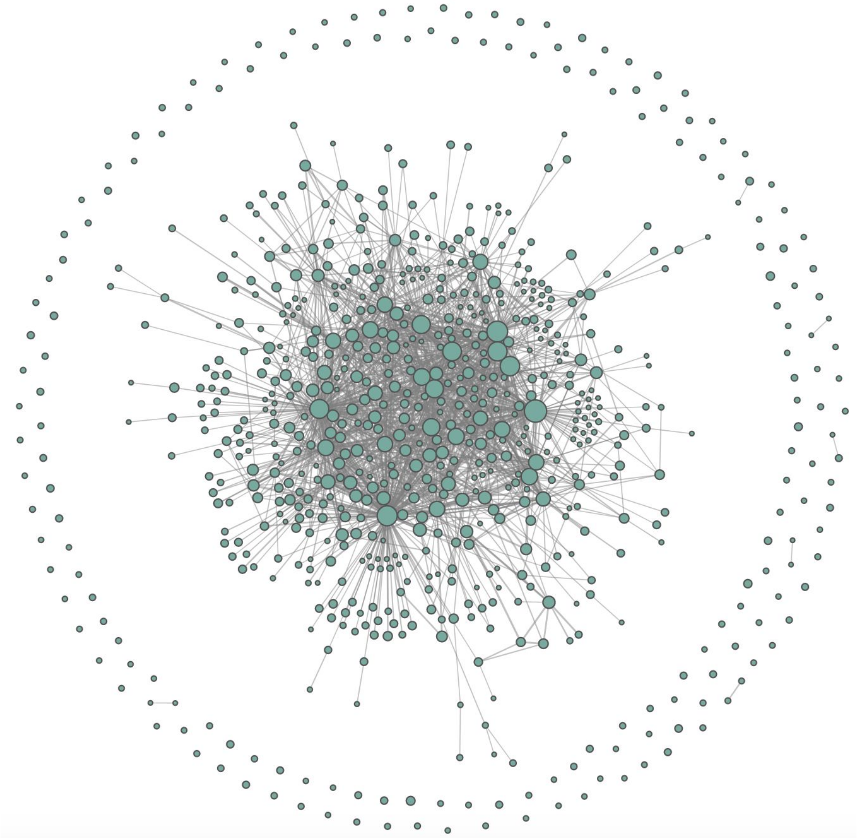
Network history goes back...



K_5

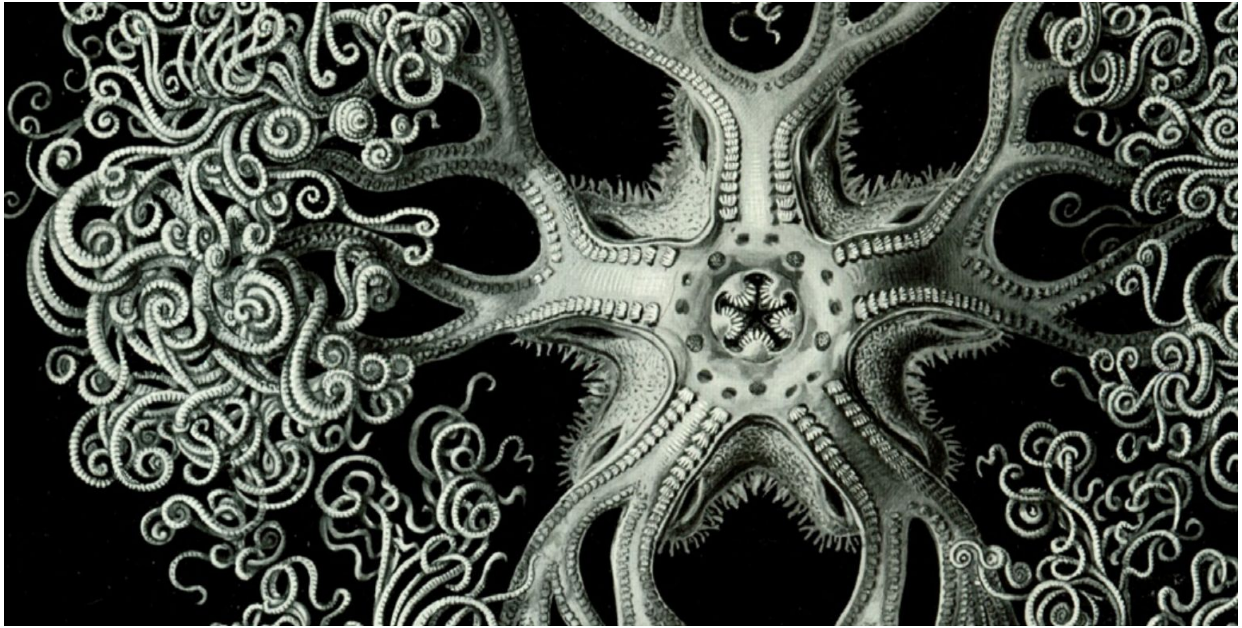


$K_{3,3}$



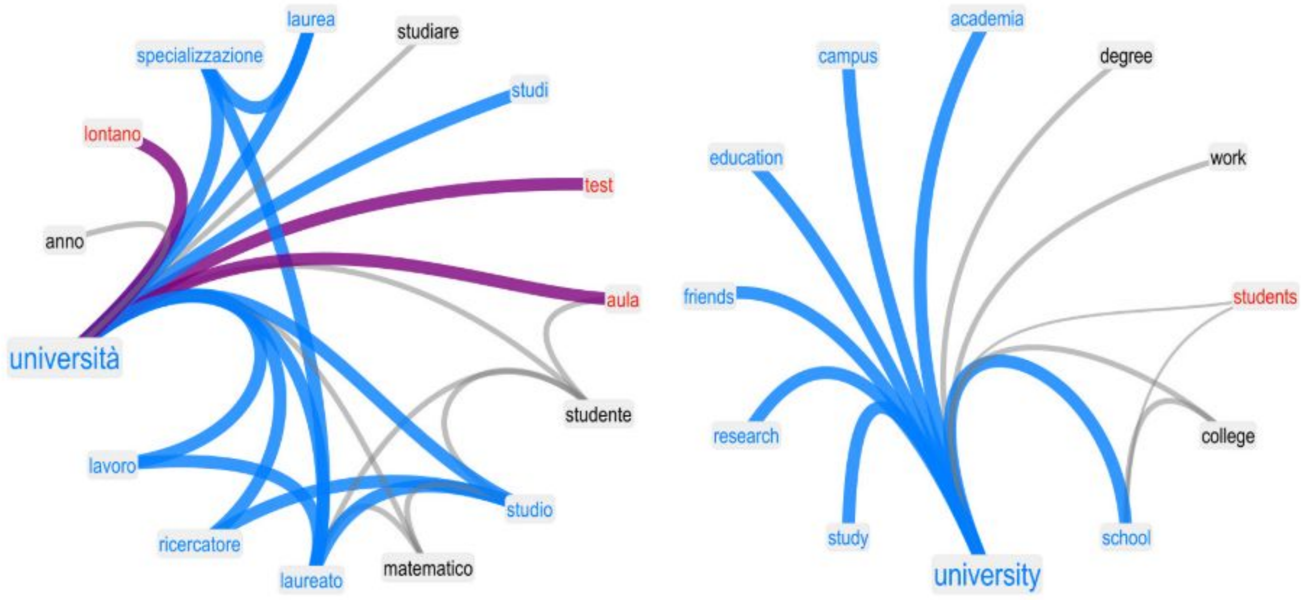
Can anything be presented as a network? What cannot be presented and why?

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?

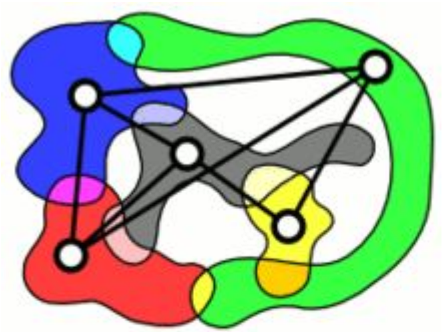
M.Stella et al.

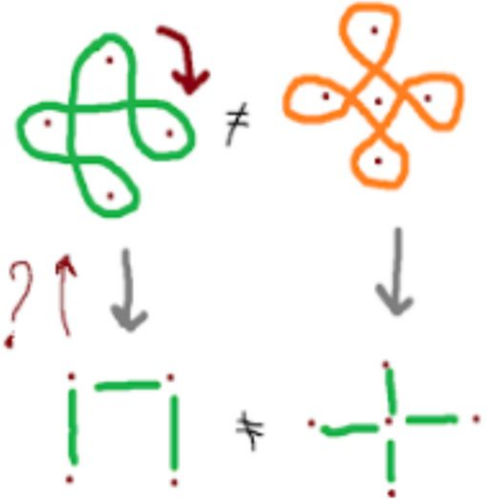


Can anything be presented as a network? What cannot be presented and why?

Shanthi, L.Bauer et al.

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





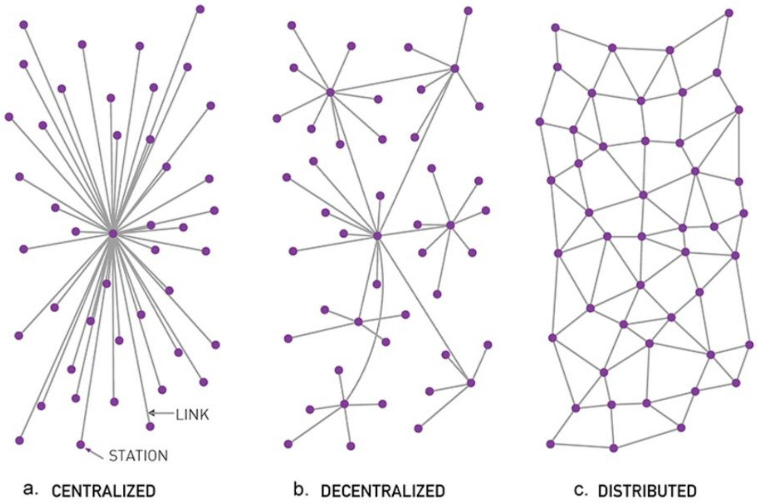
What we will look at in network science?

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

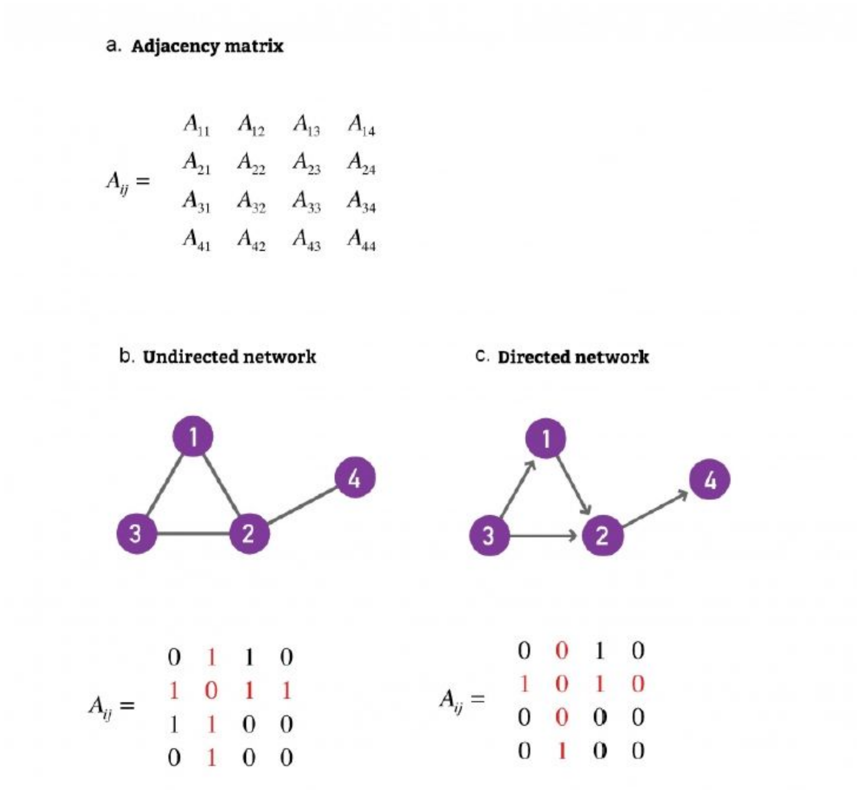
What we will look at in network science?

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

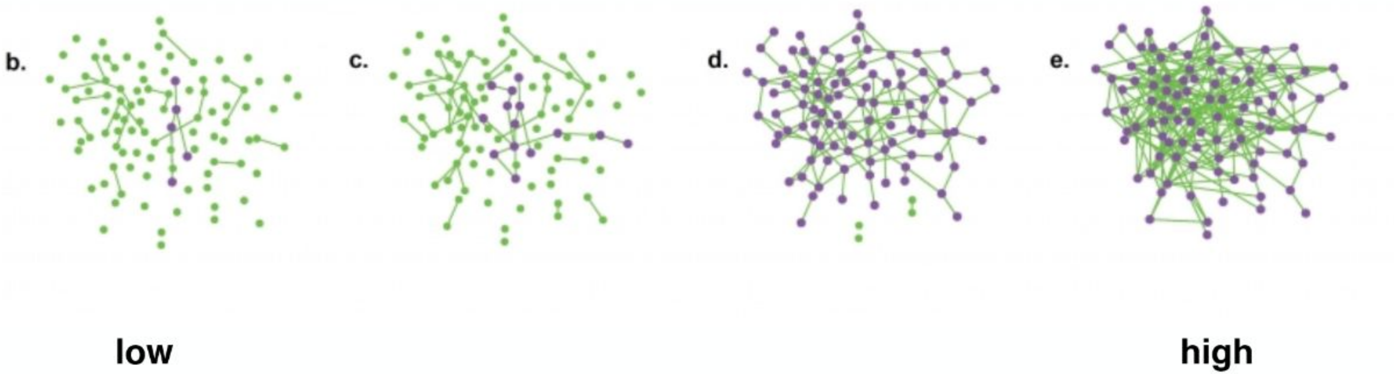
Fig. credits P. Barran.



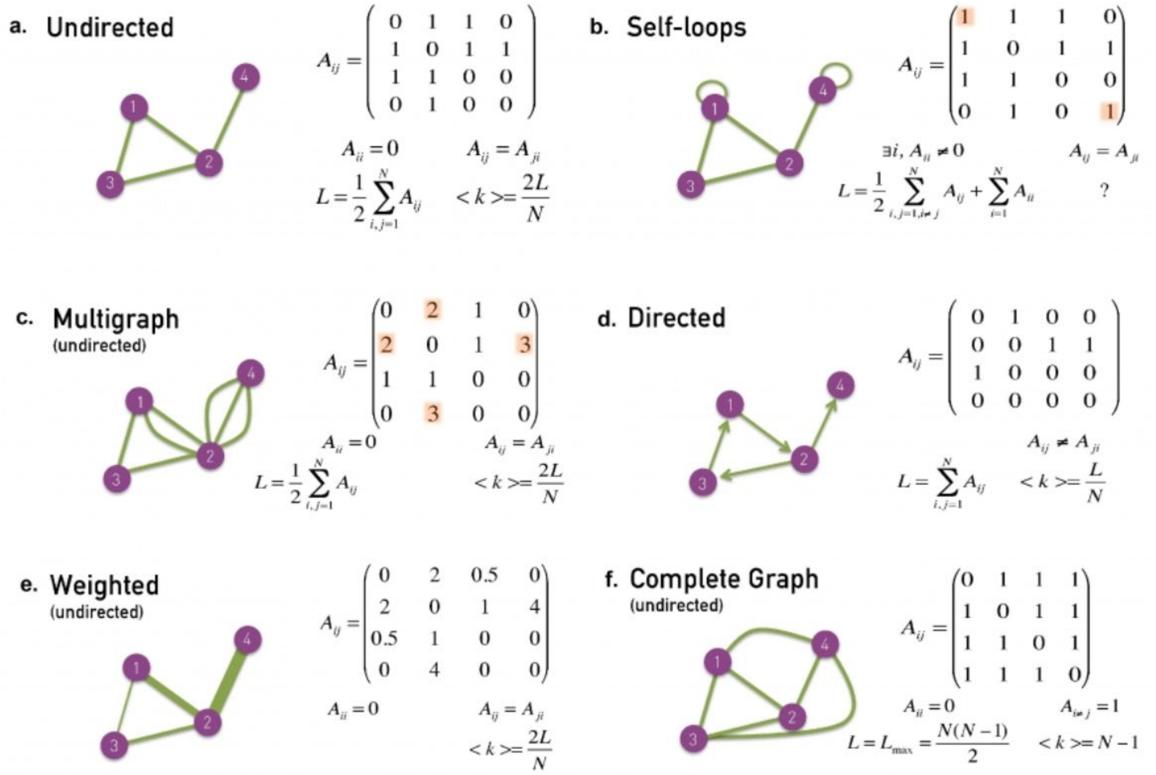
1. Network measures and definitions

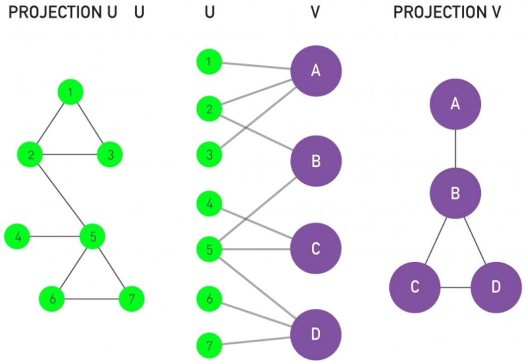


1. Network measures and definitions



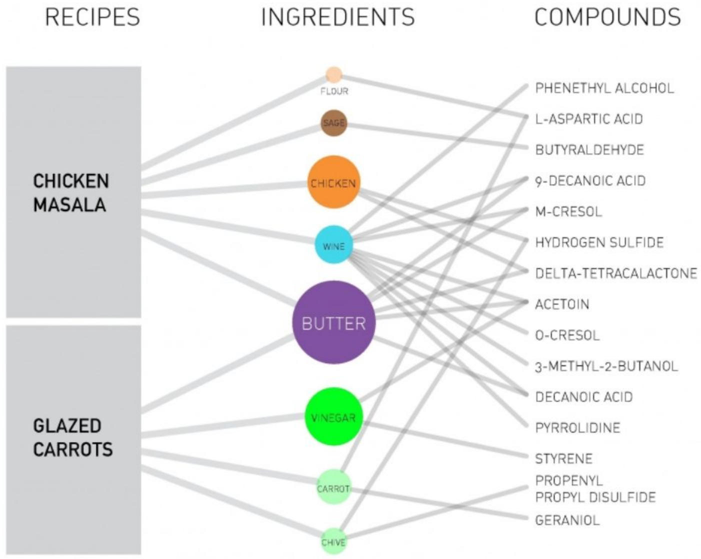
1. Network types



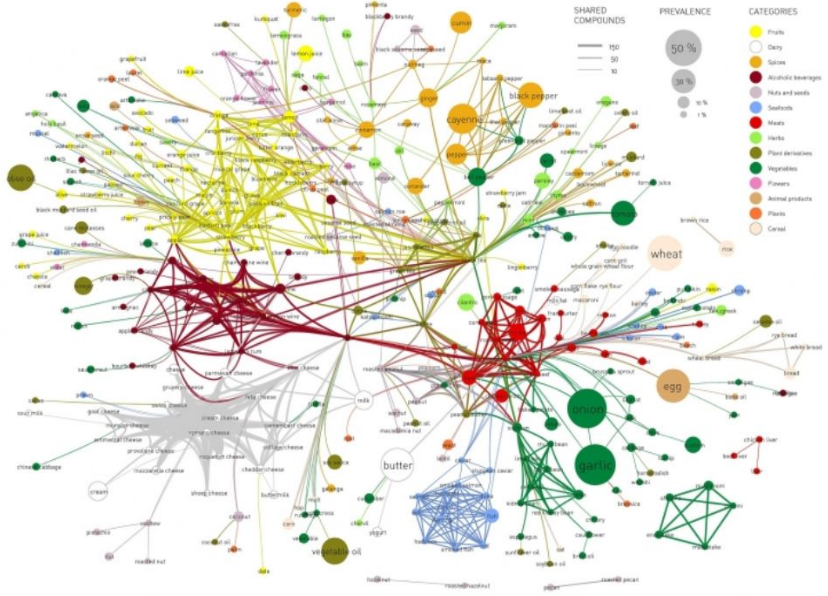


Bipartite networks

a.



b.



1. Network measures:

   Local measures (for each node)
   Global measures (for the whole network)

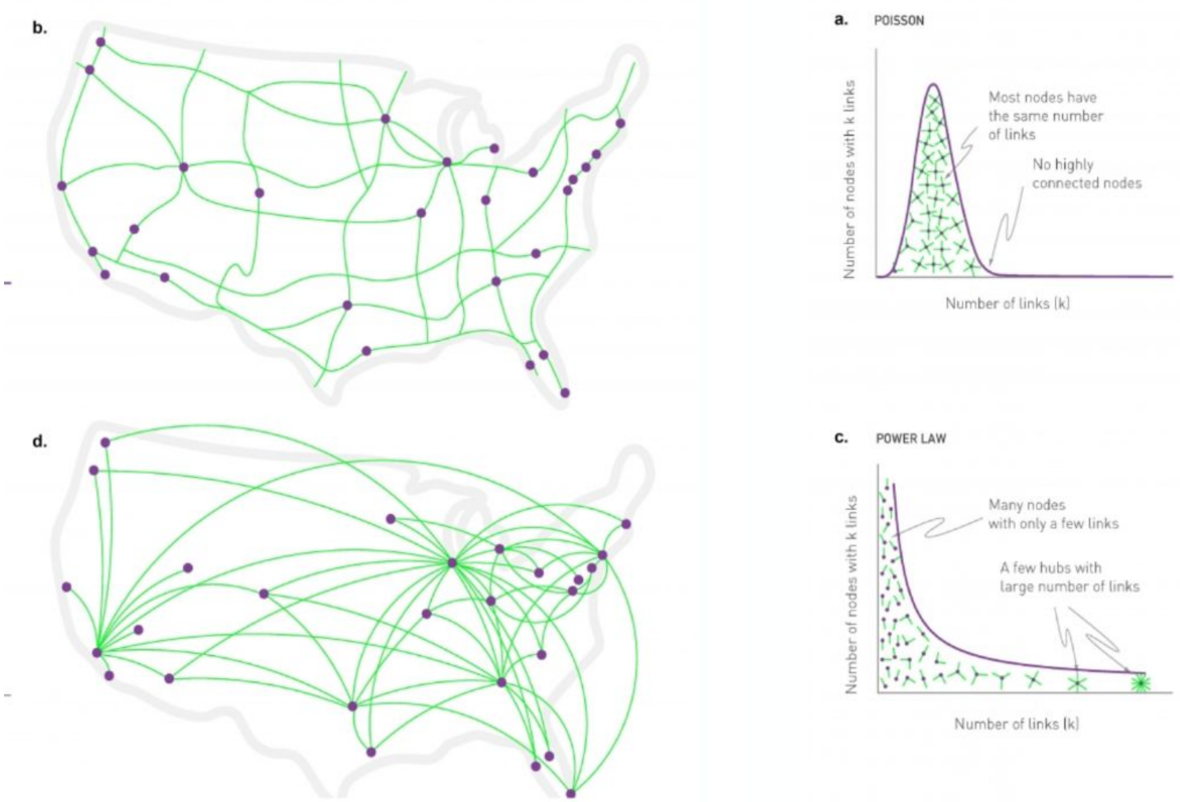
Network measures:
Check calculation on networkx python library

Table 2: Definitions of network science terms and variables.

Term/variable	Definition
N	number of nodes, N, in graph
E	number of edges, E, in graph
network density	ratio of the number of edges to the maximum number of possible edges $\frac{2E}{N(N-1)}$
distance, $d(n_i, n_j)$	shortest path between node i and node j $d(n_i, n_j)$ where $n_i, n_j \in N$
average shortest path length, L	average length of shortest path between pairs of nodes $L = \frac{1}{N(N-1)} \cdot \sum_{i \neq j} d(n_i, n_j)$
diameter, D	largest shortest path between nodes $D = \max_{n_i \in N, n_j \in N} d(n_i, n_j)$
closeness centrality	inverse of the sum of the length of the shortest paths between node i and all other nodes in the graph $C_i = \frac{1}{\sum_j d(n_i, n_j)}$
degree, k_i	number of edges attached to node i
average degree, $\langle k \rangle$	average number of edges per node in network $\langle k \rangle = \frac{1}{N} \sum_{n=i}^{N} k_i$
local clustering coefficient, c_i	number of edges between the neighbors of node i divided by the maximum number of edges between those neighbors $c_i = \frac{2\lvert e_{jk} \rvert}{k_i(k_i - 1)}$ where $n_j, n_k \in N_i,\ e_{jk} \in E$
average clustering coefficient, $\langle C \rangle$	average clustering coefficient of nodes in the network $\langle C \rangle = \frac{1}{N} \sum_{n=i}^{N} c_i$
modularity, Q	proportion of edges that fall within subgroups of nodes minus the expected proportion if edges were randomly distributed, range $[-1, 1]$
average efficiency, E_G	measure of how efficiently information is exchanged in the network $E_G = \frac{1}{n(n-1)} \sum_{i \neq j \in N} \frac{1}{d(n_i, n_j)}$
largest connected component	largest group of nodes in the network that are connected to each other in a single component
degree distribution, $P(k)$	probability distribution of node degrees in the network
γ	power-law exponent for the degree distribution
Small world structure	network with short average path lengths and relatively high clustering coefficient (relative to a random graph with similar density)
scale-free network	network with a degree distribution that is power-law distributed

Network measures:
Degree or why do we care?

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



Network measures:
Check calculation on networkx python library

TABLE 2: Definitions of network science terms and variables.

Term/variable	Definition
N	number of nodes, N, in graph
E	number of edges, E, in graph
network density	ratio of the number of edges to the maximum number of possible edges $\frac{2E}{N(N-1)}$
distance, $d(n_i, n_j)$	shortest path between node i and node j $d(n_i, n_j)$ where $n_i, n_j \in N$
average shortest path length, L	average length of shortest path between pairs of nodes $L = \frac{1}{N(N-1)} \cdot \sum_{i \neq j} d(n_i, n_j)$
diameter, D	largest shortest path between nodes $D = \max_{n_i \in N, n_j \in N} d(n_i, n_j)$
closeness centrality	inverse of the sum of the length of the shortest paths between node i and all other nodes in the graph $C_i = \frac{1}{\sum_j d(n_i, n_j)}$
degree, k_i	number of edges attached to node i
average degree, $\langle k \rangle$	average number of edges per node in network $\langle k \rangle = \frac{1}{N} \sum_{n=i}^{N} k_i$
local clustering coefficient, c_i	number of edges between the neighbors of node i divided by the maximum number of edges between those neighbors $c_i = \frac{2\lvert e_{jk} \rvert}{k_i(k_i - 1)}$ where $n_j, n_k \in N_i$, $e_{jk} \in E$
average clustering coefficient, $\langle C \rangle$	average clustering coefficient of nodes in the network $\langle C \rangle = \frac{1}{N} \sum_{n=i}^{N} c_i$
modularity, Q	proportion of edges that fall within subgroups of nodes minus the expected proportion if edges were randomly distributed, range $[-1, 1]$
average efficiency, E_G	measure of how efficiently information is exchanged in the network $E_G = \frac{1}{n(n-1)} \sum_{i \neq j \in N} \frac{1}{d(n_i, n_j)}$
largest connected component	largest group of nodes in the network that are connected to each other in a single component
degree distribution, $P(k)$	probability distribution of node degrees in the network
γ	power-law exponent for the degree distribution
Small world structure	network with short average path lengths and relatively high clustering coefficient (relative to a random graph with similar density)
scale-free network	network with a degree distribution that is power-law distributed

Betweenness centrality

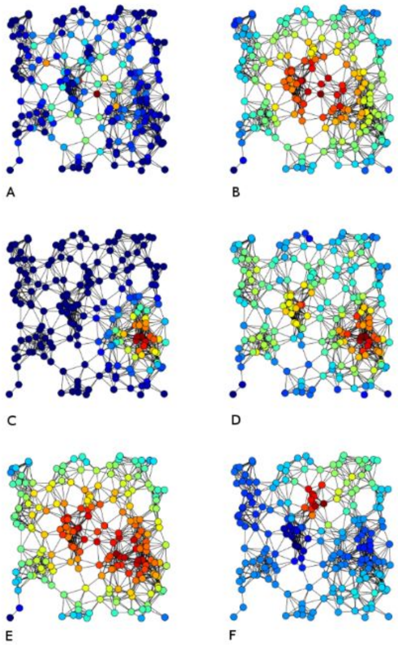
Closeness centrality

Eigenvector centrality

Degree centrality

Harmonic centrality

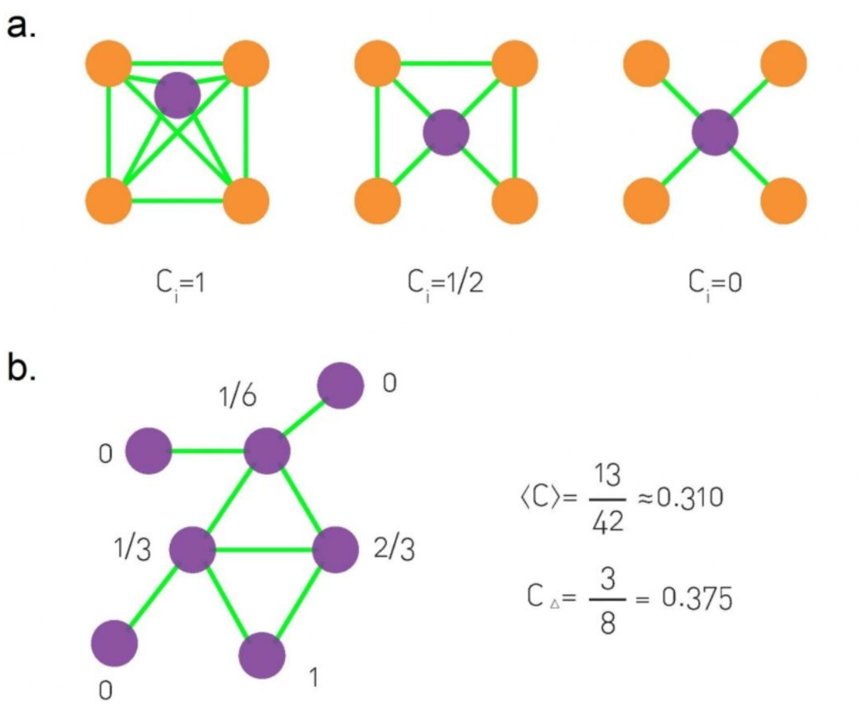
Katz centrality



1. Network measures

Example of clustering

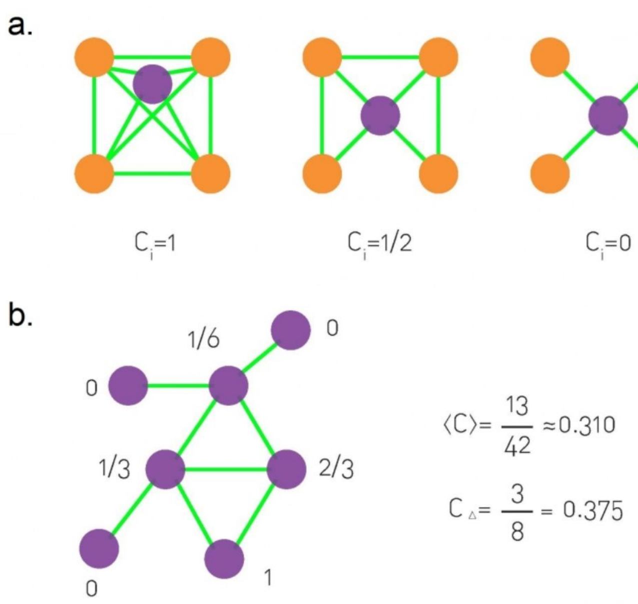
Notebooks at https://github.com/Big-data-course-CRI/materials_big_data_cri_2019



1. Network measures

Example of clustering, paths, betweenness

Notebooks at https://github.com/Big-data-course-CRI/materials_big_data_cri_2019



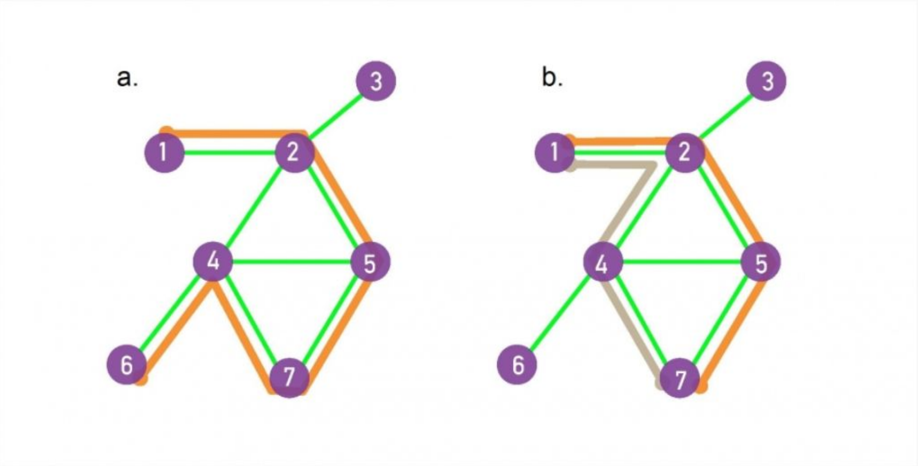


Image 2.12

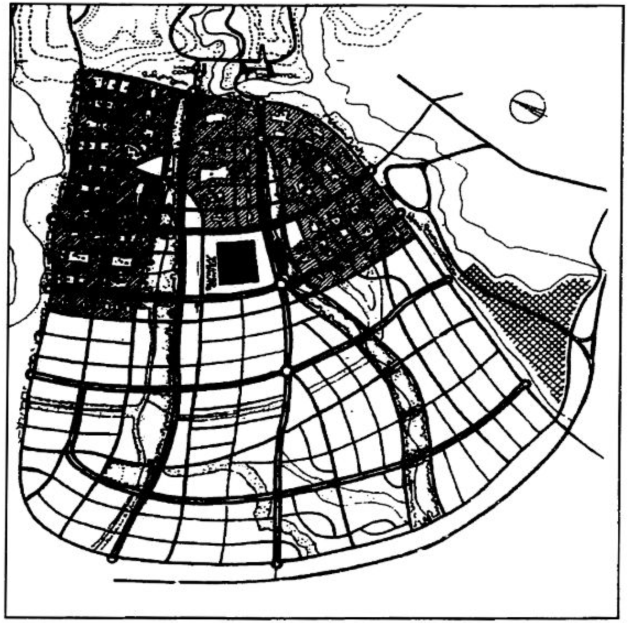
Paths

- A path between nodes i_0 and i_n is an ordered list of n links P = $\{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$. The length of this path is n. The path shown in orange in (a) follows the route 1→2→5→7→4→6, hence its length is n = 5.

What we will look at in network science?

1. Network measures
2. **Networks in time and space**
3. Networks and processes
4. Networks from data

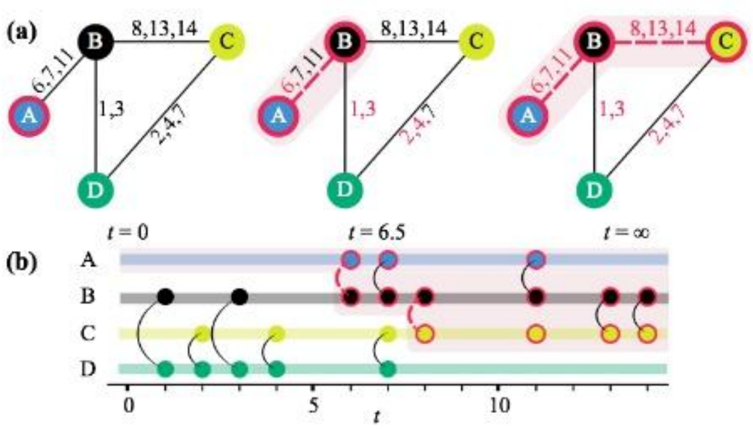
Master Plan for Chandigarh by Albert Mayer RAIC Journal, 1955 (Evenson Norma, Chandigarh, 1966)



Networks in time and space

https://arxiv.org/pdf/1108.1780.pdf

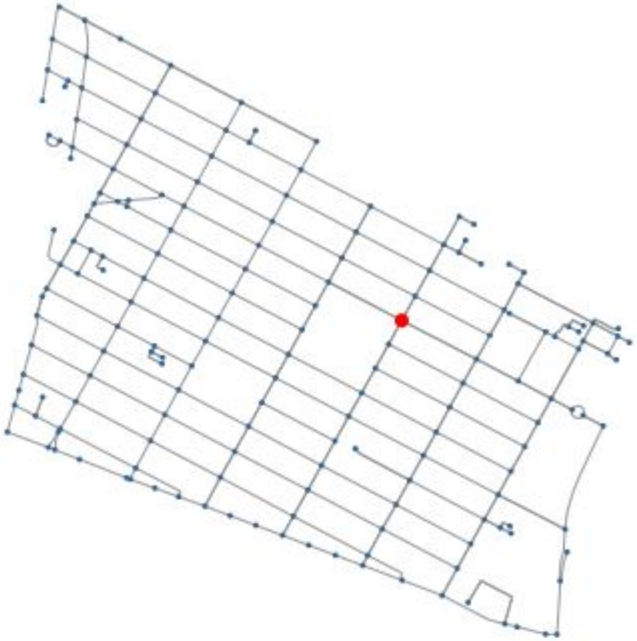
Temporality matters: reachability issue

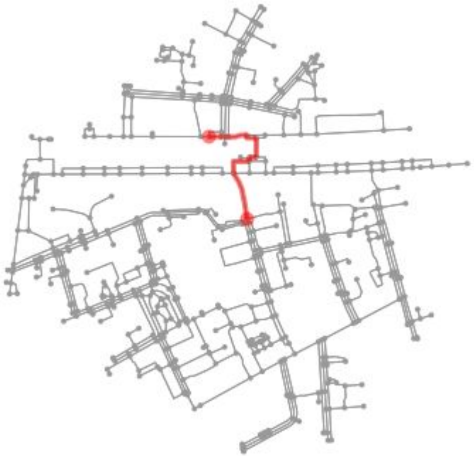


Review Holme-Saramaki, Phys. Rep. (2012), arXiv:1108.1780

Networks in time and space

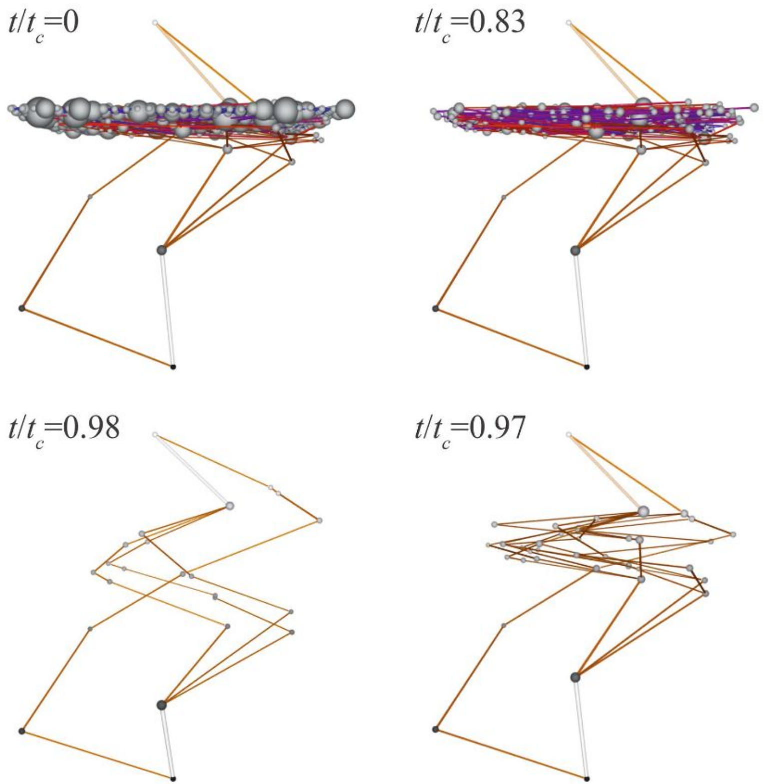
https://arxiv.org/abs/1010.0302



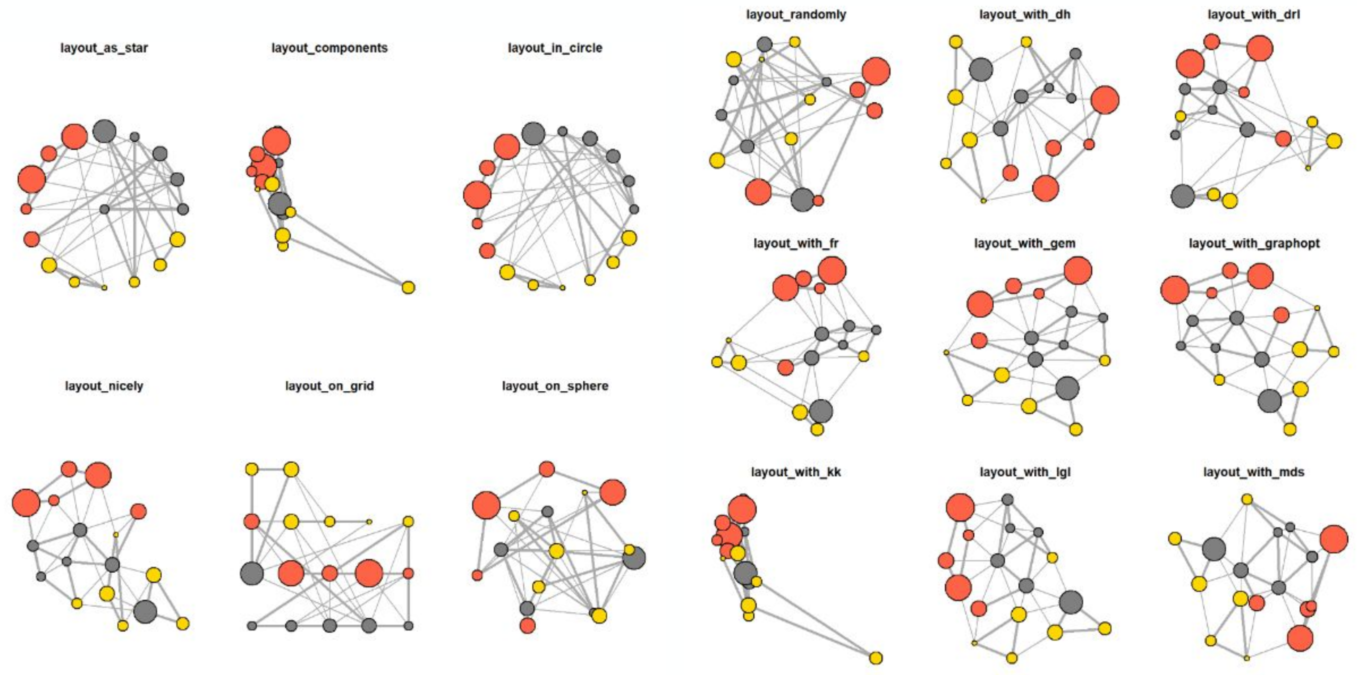


Networks in time and space

https://www.nature.com/articles/s41598-019-44701-6



Networks layout (also for Day 3)



Networks in time and space

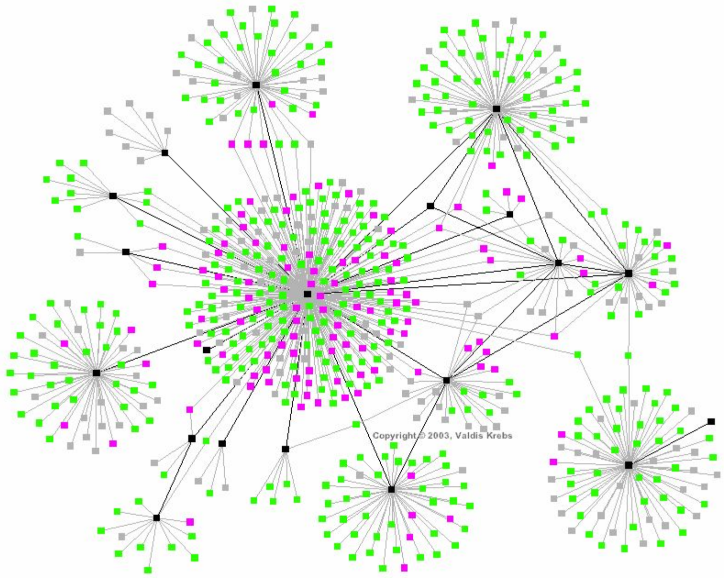
Good resource on spatial networks

M.Barthelemy “Spatial networks”

Good resource on temporal networks

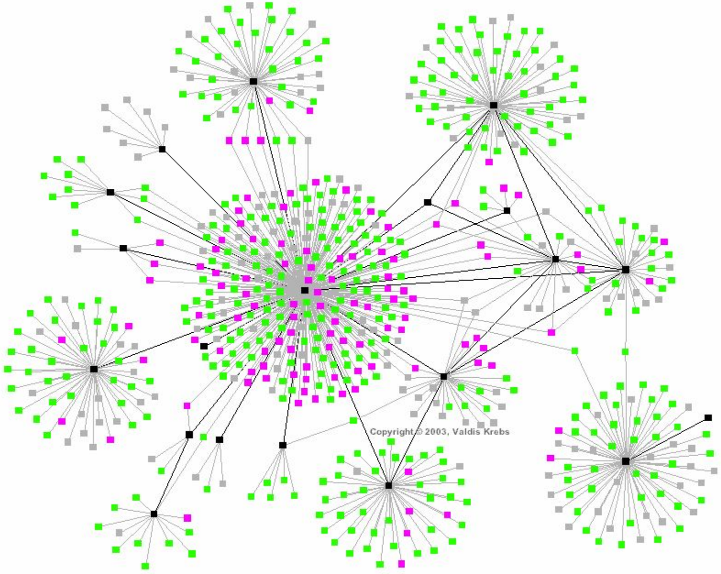
What we will look at in network science?

1. Network measures
2. Networks in time and space
3. **Networks and processes**
4. Networks from data



What are spreading processes on networks?

Any examples of spreading processes?



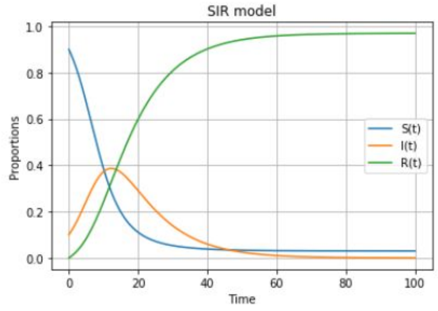
SIR model

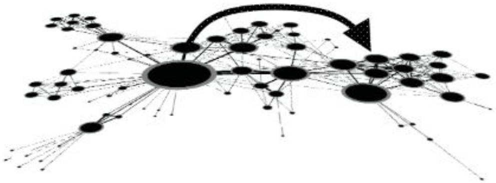
of epidemics spreading.

Make your own simulations of susceptible individuals count and on a network or in population

https://github.com/Big-data-course-CRI/materials_big_data_cri_2019/blob/master/resources%20Python/spreading_on_networks.ipynb

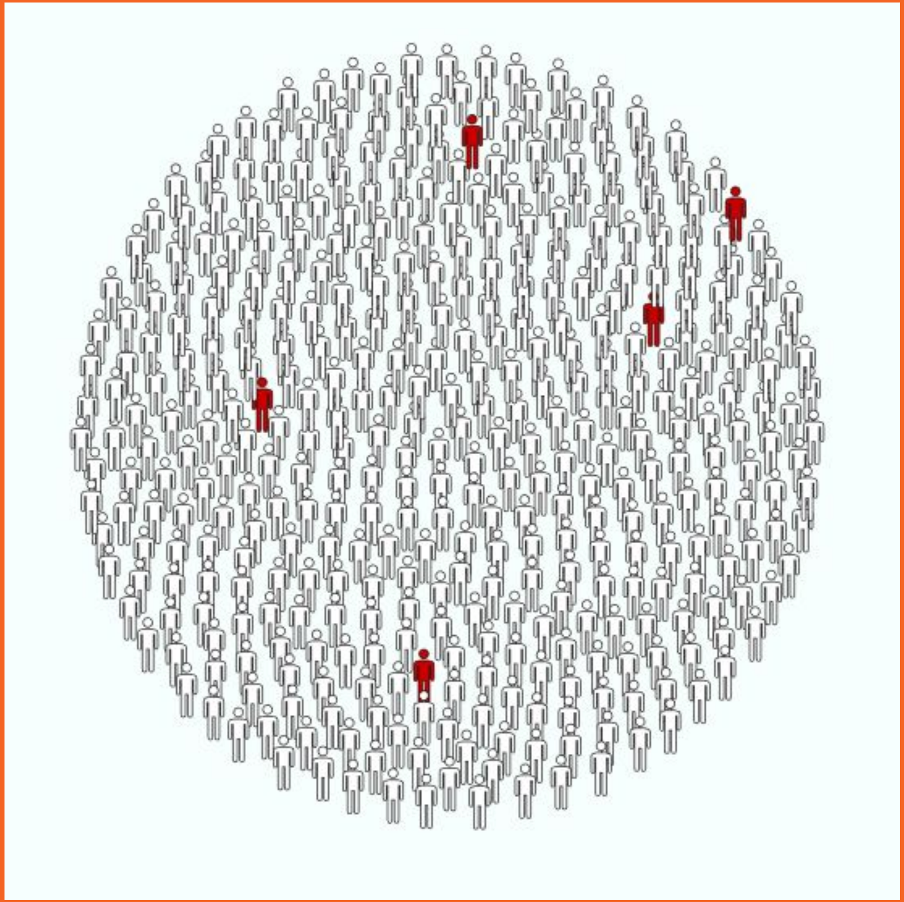
```
solution=scipy.integrate.odeint(SIR_model,[S0,I0,R0],t,args=(beta,gamma))
solution=numpy.array(solution)

plt.figure(figsize=[6,4])
plt.plot(t,solution[:,0],label="S(t)")
plt.plot(t,solution[:,1],label="I(t)")
plt.plot(t,solution[:,2],label="R(t)")
plt.grid()
plt.legend()
plt.xlabel("Time")
plt.ylabel("Proportions")
plt.title("SIR model")
plt.show()
```





SIR model:

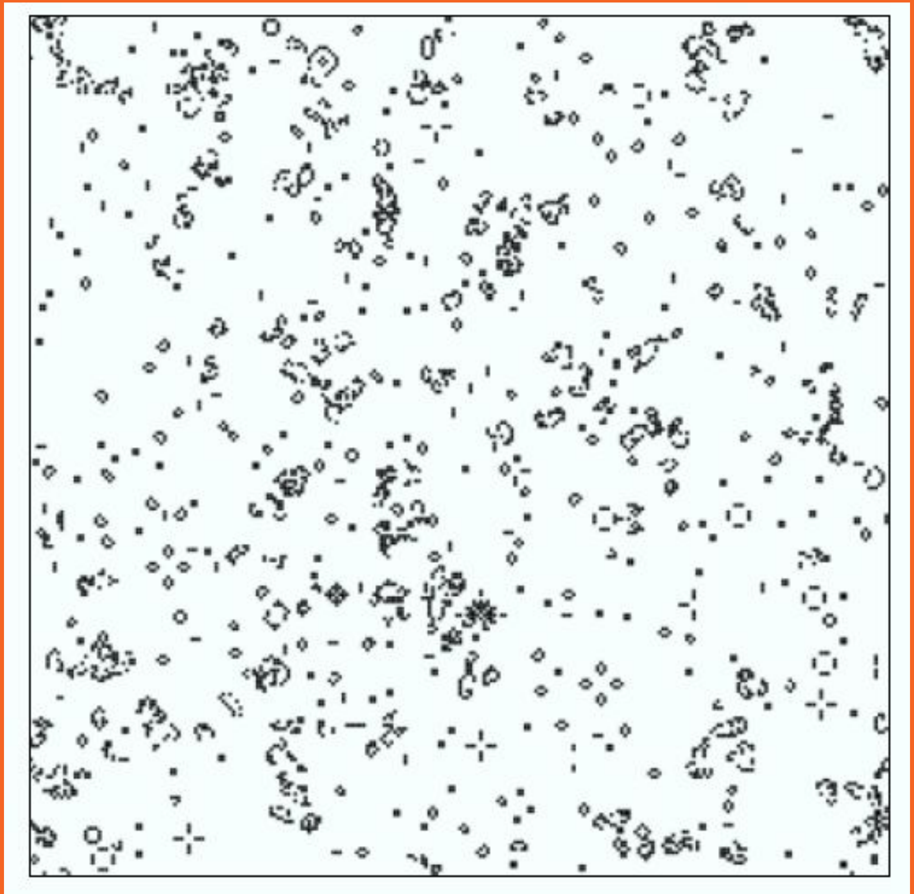
Susceptible individuals (S) can be infected by coming in contact with other infected (I) individuals.
Once infected they can transmit the disease until they recover (R) and become immune. After some time immunity wanes and individuals become susceptible again:



Conway game of life model

An attempt to describe life from deterministic point of view of cellular automata

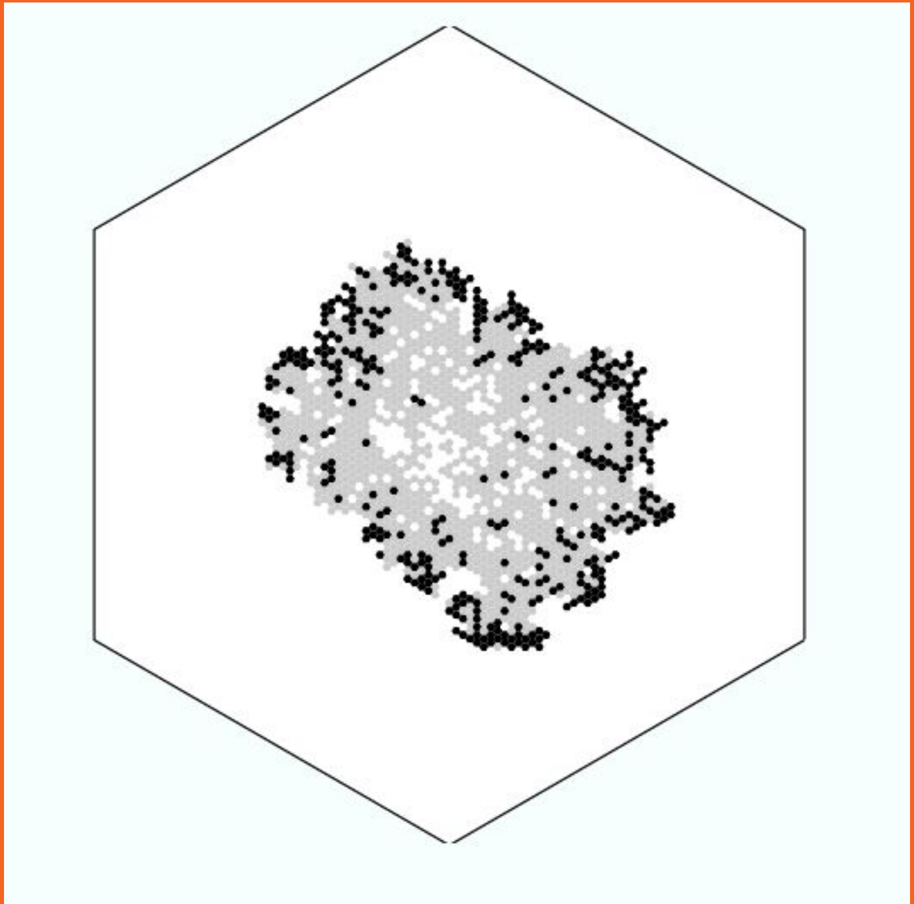
https://mathworld.wolfram.com/GameofLife.html



SIR model

of epidemics spreading with spatial parameter

https://www.complexity-explorables.org/slides/critical-hexsirsize/



What we will look at in network science?

1. Network measures
2. Networks in time and space
3. Networks and processes
4. **Networks from data**

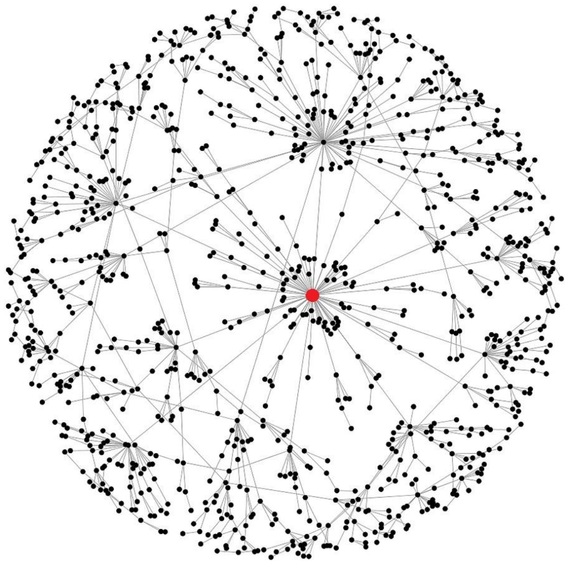
Figure 7.11
Aaron Koblin's *Flight Patterns* (2005): visualization of the flight paths of aircraft crossing North America

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)

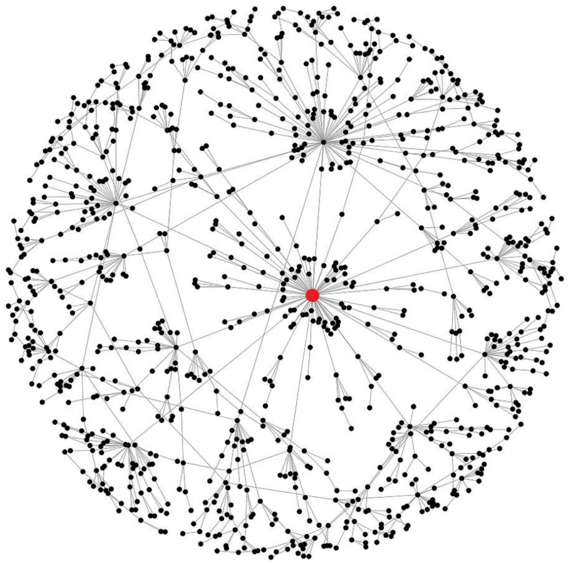
How to construct networks from data?

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



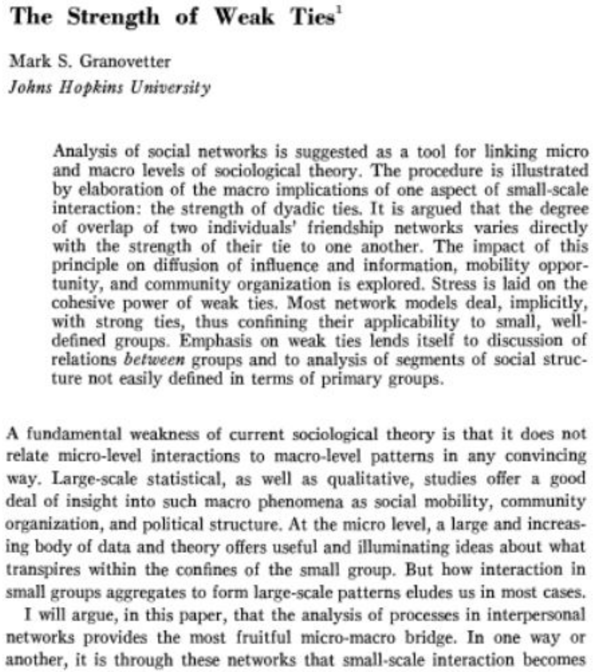
How to construct networks from data?

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



Remember that networks do not give one-to-one
Correspondence of your data.
Hence do not generalize

Social networks analysis



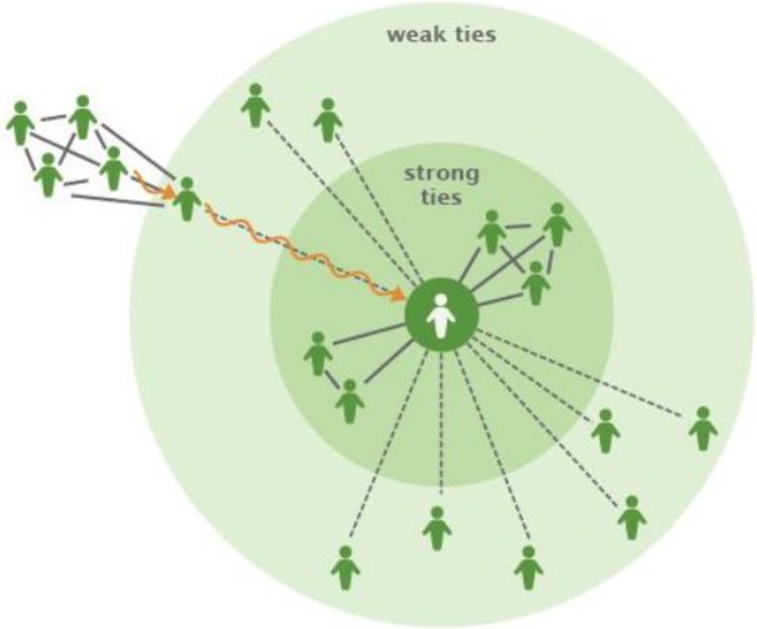
The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

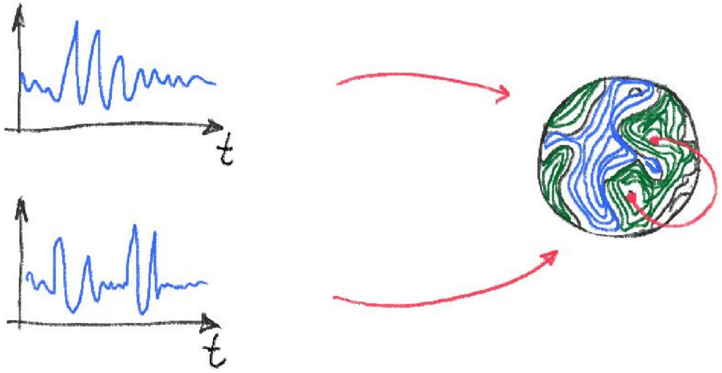
A fundamental weakness of current sociological theory is that it does not relate micro-level interactions to macro-level patterns in any convincing way. Large-scale statistical, as well as qualitative, studies offer a good deal of insight into such macro phenomena as social mobility, community organization, and political structure. At the micro level, a large and increasing body of data and theory offers useful and illuminating ideas about what transpires within the confines of the small group. But how interaction in small groups aggregates to form large-scale patterns eludes us in most cases.

I will argue, in this paper, that the analysis of processes in interpersonal networks provides the most fruitful micro-macro bridge. In one way or another, it is through these networks that small-scale interaction becomes



How to construct networks from data?

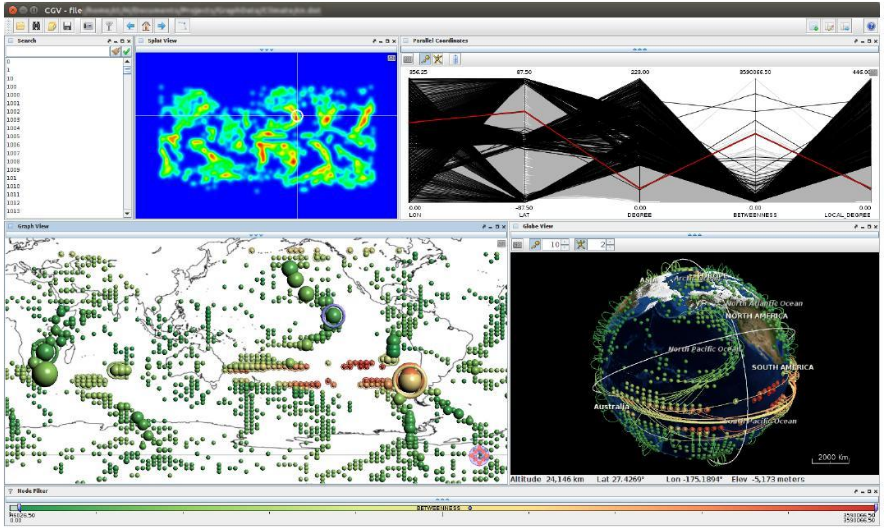
1. Directly build correspondence between links and edges with data (social networks, flights data)
2. Preprocess data (first build correlation from data)



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

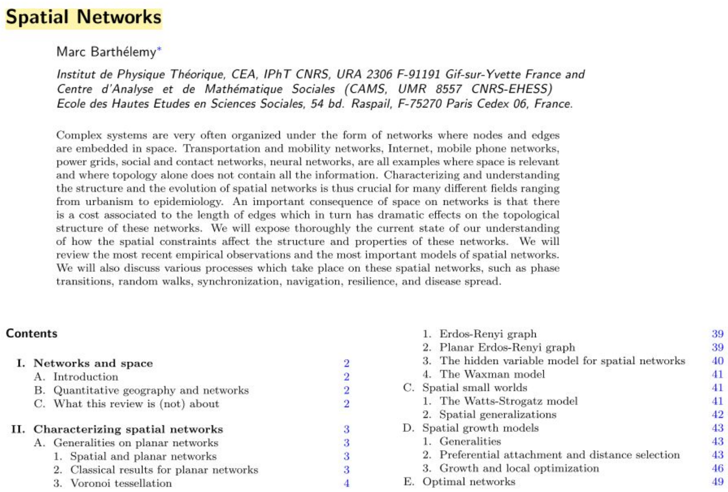


Network resources

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



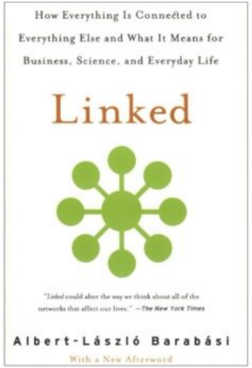
Spatial Networks

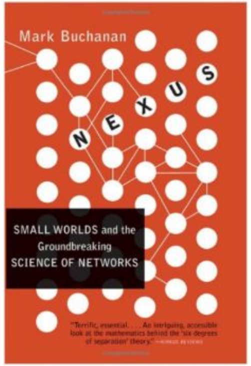
Marc Barthélemy*

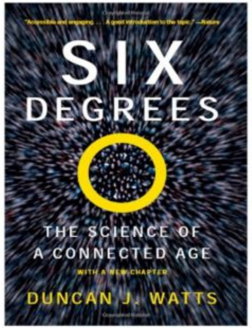
Institut de Physique Théorique, CEA, IPhT CNRS, URA 2306 F-91191 Gif-sur-Yvette France and Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 CNRS-EHESS) Ecole des Hautes Etudes en Sciences Sociales, 54 bd. Raspail, F-75270 Paris Cedex 06, France.

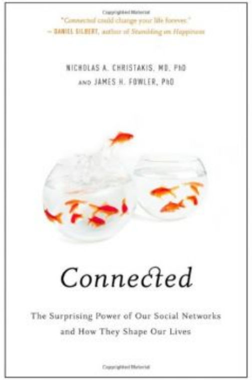
Complex systems are very often organized under the form of networks where nodes and edges are embedded in space. Transportation and mobility networks, Internet, mobile phone networks, power grids, social and contact networks, neural networks, are all examples where space is relevant and where topology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

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