# CE88/CP 88: Lecture 1

# Data Science for Smart Cities

# Marta C. Gonzalez

**Associate Professor of City and Regional Planning and Civil and Environmental Engineering, Faculty Scientist at the LBNL**

My team develops computational models to analyze digital traces to estimate the demand on urban infrastructure in relation to energy and mobility. Examples include: traffic gridlocks and the integration of electric vehicles in the power grid, policy of solar energy adoption, and habits in spending behavior.

http://humnetlab.berkeley.edu/
martag@berkeley.edu


# Martin Liu

Master of Engineering in Civil Engineering (MENG-CE)
https://www.linkedin.com/in/martin-liu-42314417a/
martin97@berkeley.edu



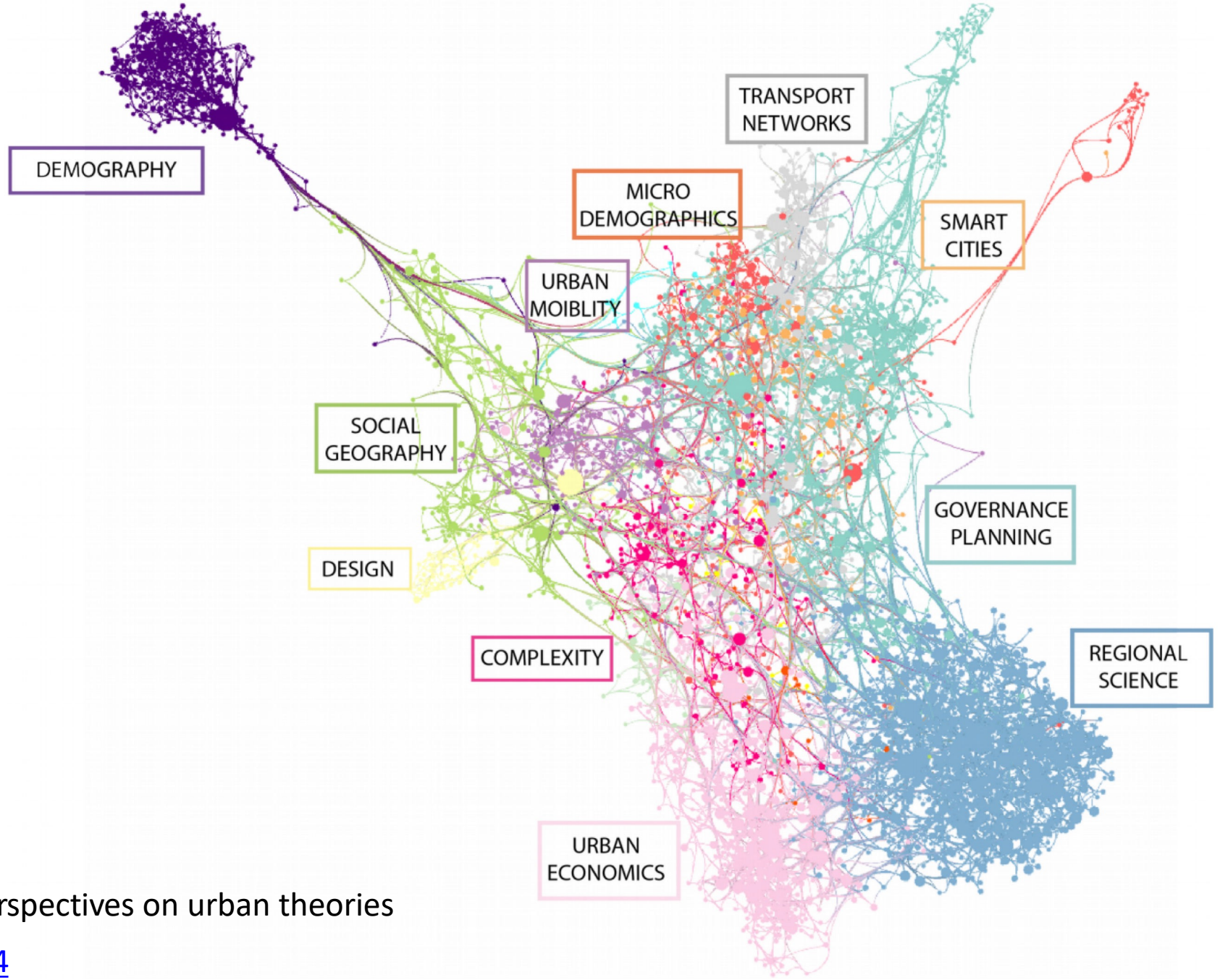
# Giuseppe Perona

Computer Science and City Planning
https://www.linkedin.com/in/giuseppe-perona-709791181
giuseppe.perona@berkeley.edu

Design and operation of smart, efficient, and resilient cities nowadays require data science skills.

This course provides an introduction to working with data generated within transportation systems, power grids, communication networks, as well as data collected via crowd-sensing sensing technologies.

The core Method of this class Network Science in a framework of Complex Systems thinking to study urban systems. Programming language is Python.




Source: Denise Punaim et al., Perspectives on urban theories

https://arxiv.org/abs/1911.02854

**Edward O. Wilson 1929–2021**
**American sociobiologist**

The real problem of humanity is the following: we have Paleolithic emotions, medieval institutions, and god-like technology. Debate at the Harvard Museum of Natural History, Cambridge, Mass., 9 September 2009

# CE88: Data Science for Urban Systems

Urban science seeks to understand the fundamental processes that drive, shape and sustain cities and urbanization.

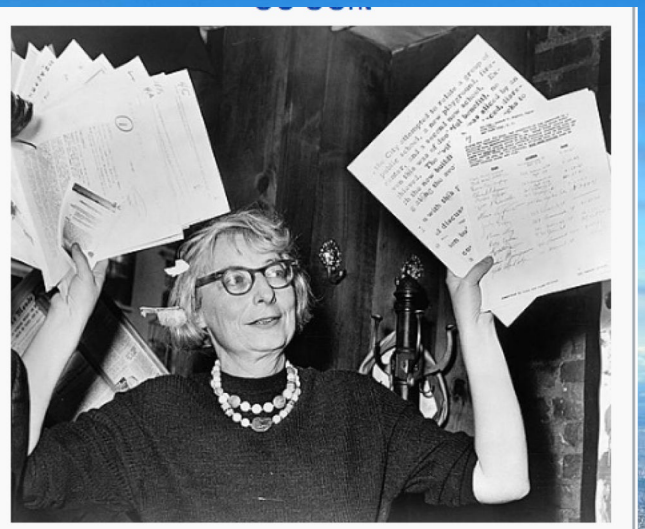
It is a multi/transdisciplinary approach involving concepts, methods and research from the social, natural, engineering and computational sciences, along with the humanities.

source: Report for NSF: Urban Science: Integrated Theory from the First Cities to Sustainable Metropolises

## Note on the report:

Urban analytics is a collection of tools used to analyze and map “urban big data” (generated by social media, crowd sourcing an sensor networks)

# Some History



Jane Jacobs as chairperson of a Greenwich Village civic group at a 1961 press conference

Jacobs fought to prevent Washington Square Park, pictured, from being demolished for a highway

# The Science of Cities

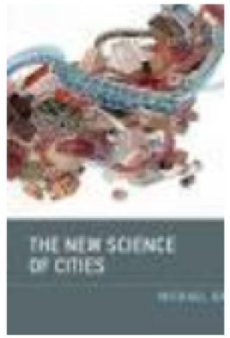
Jane Jacobs (1916 –2006) journalist, author, and activist who influenced urban studies, sociology, and economics.) arguing that urban renewal did not respect the needs of city-dwellers.

Made us think of cities as problems in organized complexity -- organisms that are replete with unexamined, but obviously intricately interconnected, and surely understandable, relationships.

-- **Jane Jacobs *The Life and Death of Great American Cities (1961)***

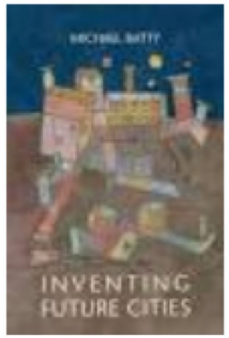
British Urban Planner

He has pioneered the idea of cities as complex systems as well as for his leading role in the conformation of the Science of Cities that combines a wide spectrum of disciplines ranging from Statistical Physics, Mathematics, Architecture and Engineering, to Social Sciences and Economics.)

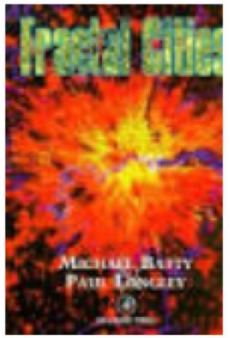



The New Science of Cities
2013

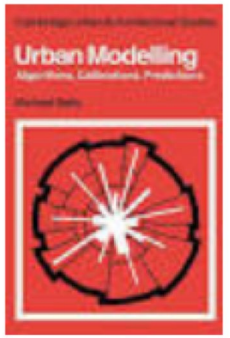
Cities and Complexity
2005




Inventing Future Cities
2018



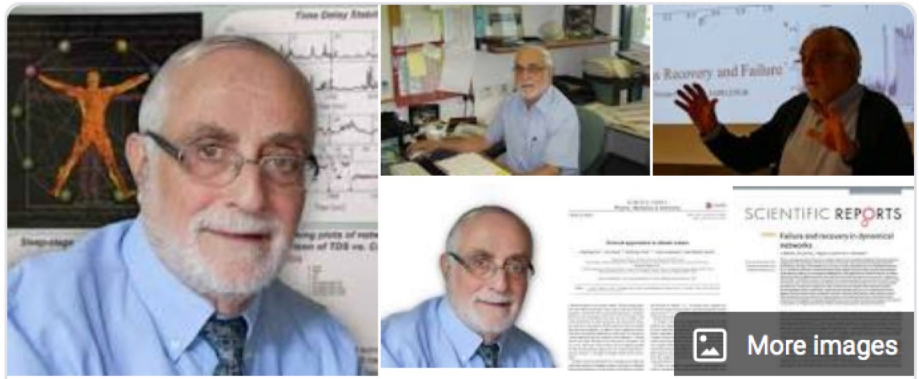

Fractal Cities: A Geometr...
1994




Urban Modelling: Algorith...
1976

Complexity science, also called complex systems science, studies how a large collection of components – locally interacting with each other at small scales – can spontaneously self-organize to exhibit non-trivial global structures and behaviors at larger scales, often without external intervention, central authorities or leaders.

Physics of complex systems connected with Cities…




## Shlomo Havlin

Professor

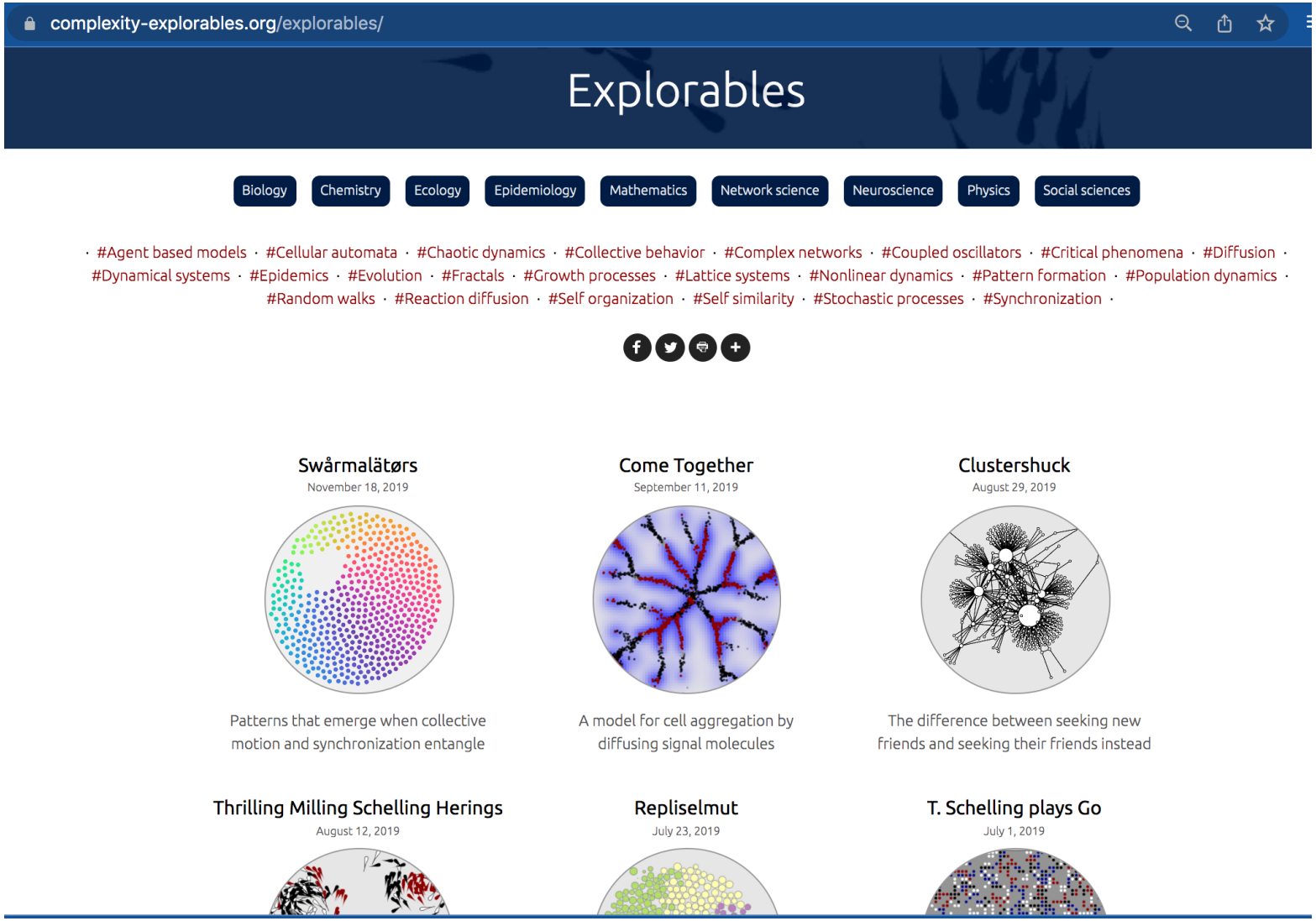
Shlomo Havlin is a Professor in the Department of Physics at Bar-Ilan University, Ramat-Gan, Israel. He served as President of the Israel Physical Society, Dean of Faculty of Exact Sciences, Chairman, Department of Physics. In 2018 he won the Israel Prize for his accomplishments in physics. Wikipedia

## H. Eugene Stanley

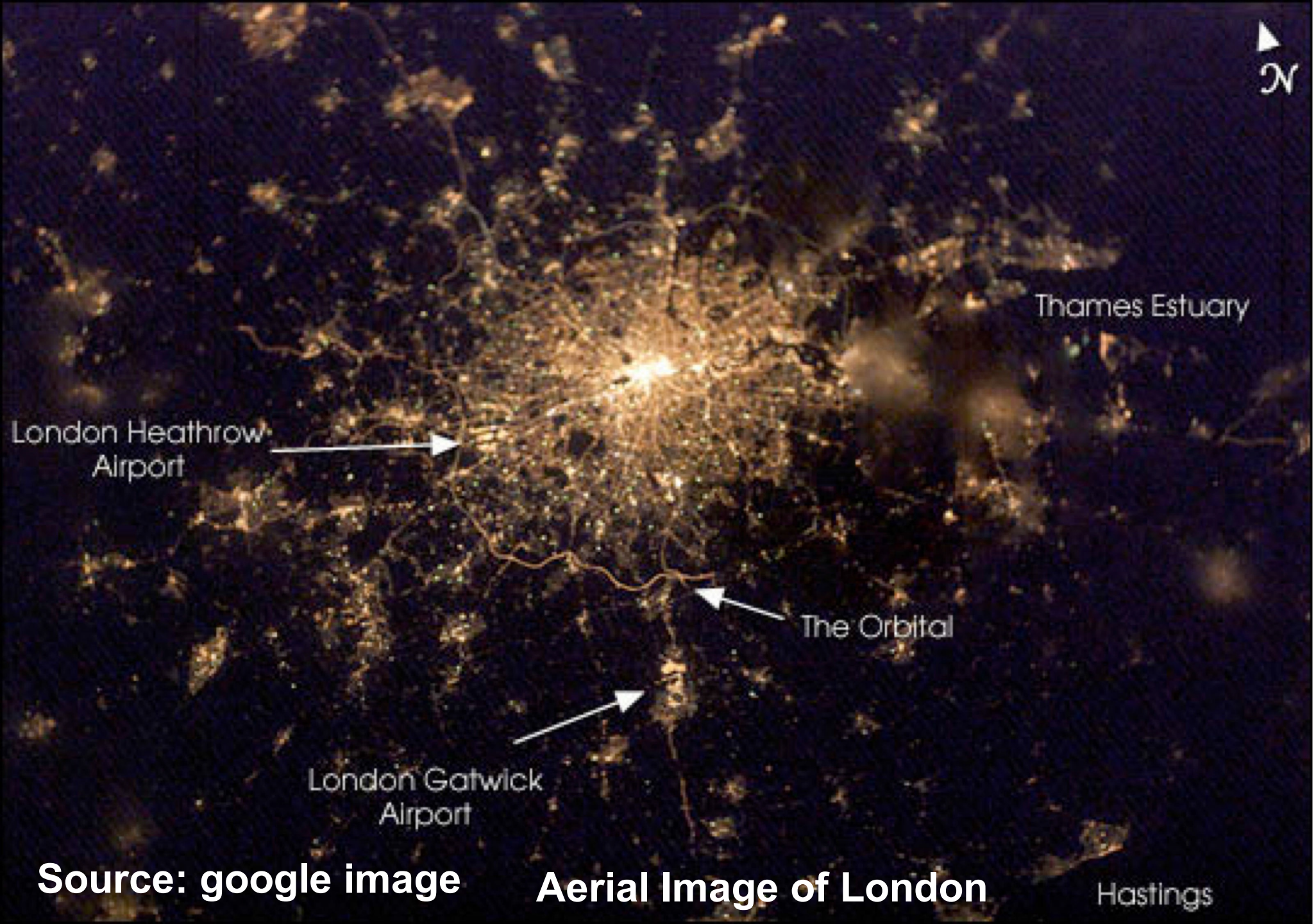
From Wikipedia, the free encyclopedia

**Harry Eugene Stanley** (born March 28, 1941) is an American physicist and University Professor at Boston University. He has made seminal contributions to statistical physics and is one of the pioneers of interdisciplinary science. His current research focuses on understanding the anomalous behavior of liquid water, but he had made fundamental contributions to complex systems, such as quantifying correlations among the constituents of the Alzheimer brain, and quantifying fluctuations in noncoding and coding DNA sequences, interbeat intervals of the healthy and diseased heart. He is one of the founding fathers of econophysics.

# Definitions: https://complexityexplained.github.io/




**link:** https://www.complexity-explorables.org/explorables/




Source: google image

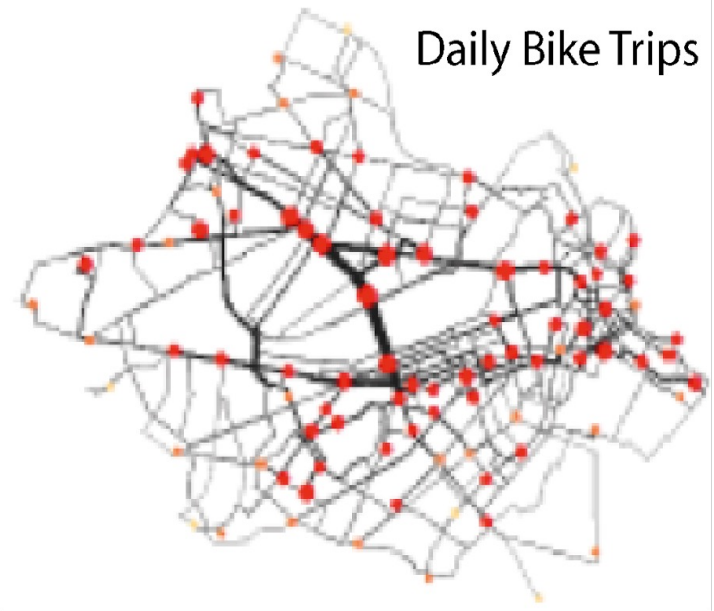
Aerial Image of London

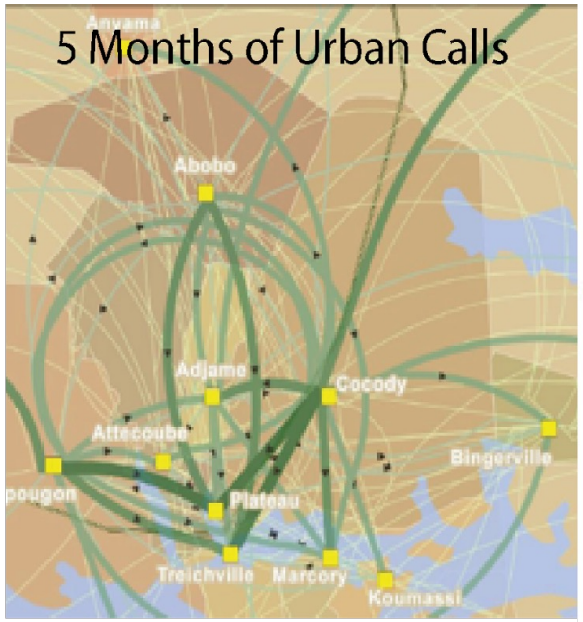
# What we find in Cities:

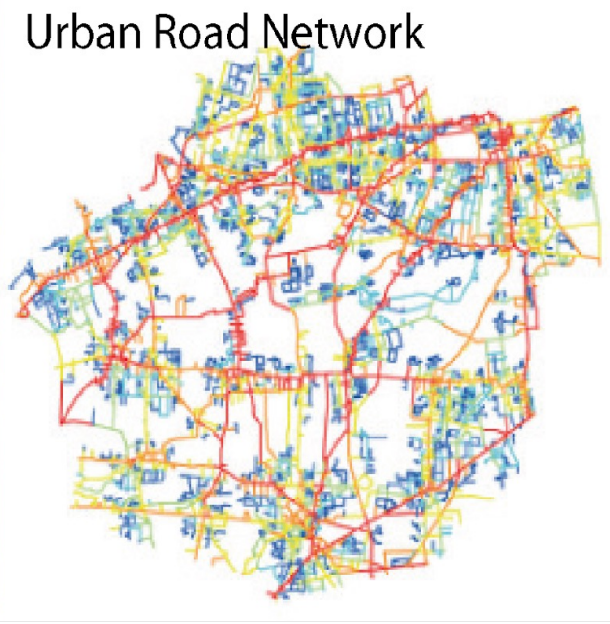
- An Engineered or Natural-physical system
- Social Behavior of people interacting with the system,

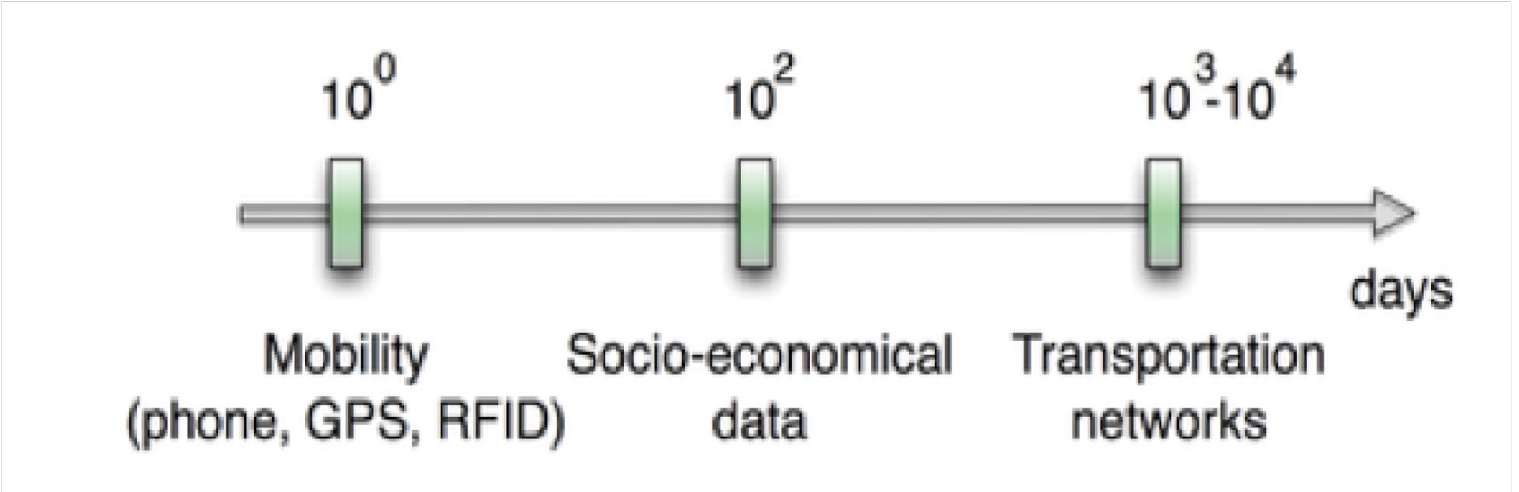
and

- Institutional behavio of organized units such as regulators and Markets governing the System.




Review "Spatial Networks" published in Physics Reports (2011) Free online: https://arxivorg/abs/1010.0302
Book: "Morphogenesis of Spatial Networks" Springer (2018)

# Sample Projects from last years

Sample Project: Spring 2020

# Disease transmission dynamics in a high school: network analysis

**Alex Zhao**[1]

[1]*UC Berkeley, Email: axyzhao@berkeley.edu*

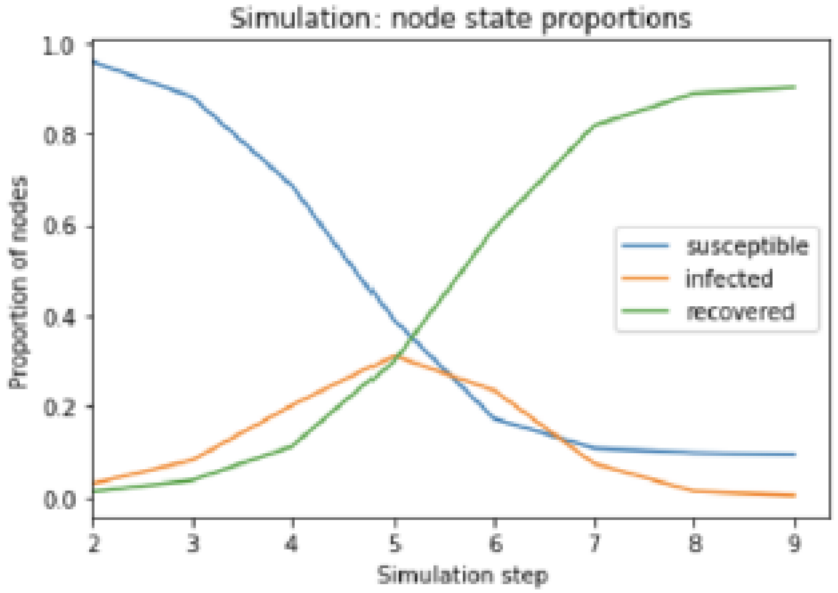
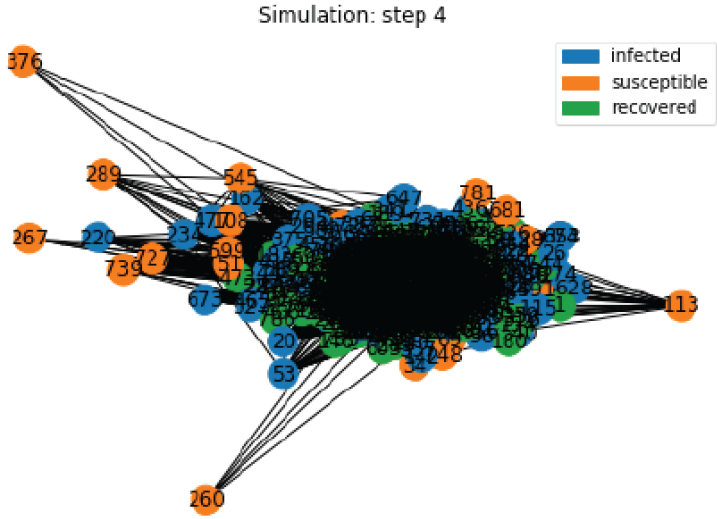
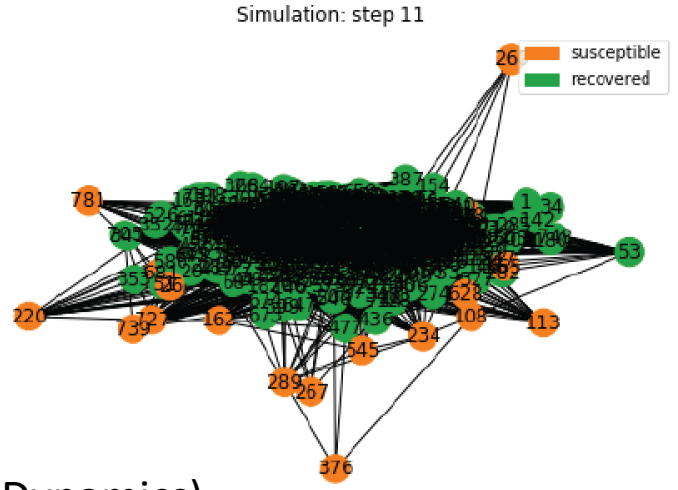



**Figure 7.** Transmission dynamics when first patient has high degree




From: Lecture 2 (Metrics), 3 (Reading Data) and 11 (Network Dynamics)

Applied Network Science

About | **Articles** | Collections | Submission Guidelines

Research | Open Access | Published: 24 November 2021

# A network-based group testing strategy for colleges

Alex Zhao ✉, Kavin Kumaravel, Emanuele Massaro & Marta Gonzalez

*Applied Network Science* **6**, Article number: 93 (2021) | Cite this article



Resulted in a published article Nov 2021

https://appliednetsci.springeropen.com/articles/10.1007/s41109-021-00431-1

**Fig. 2**

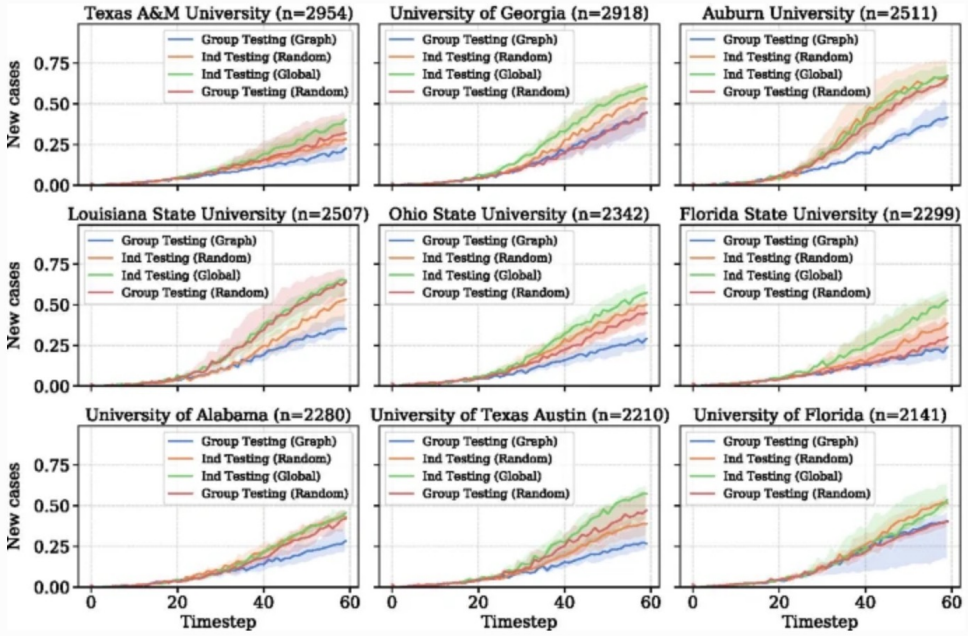
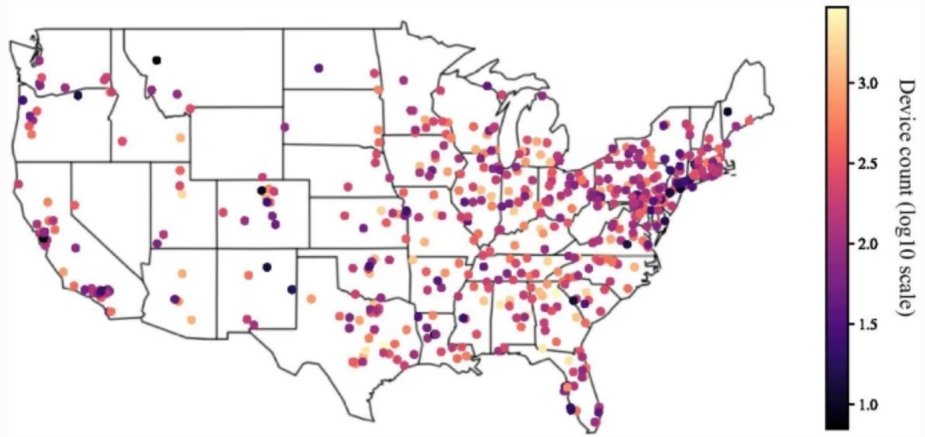



**Fig. 1**

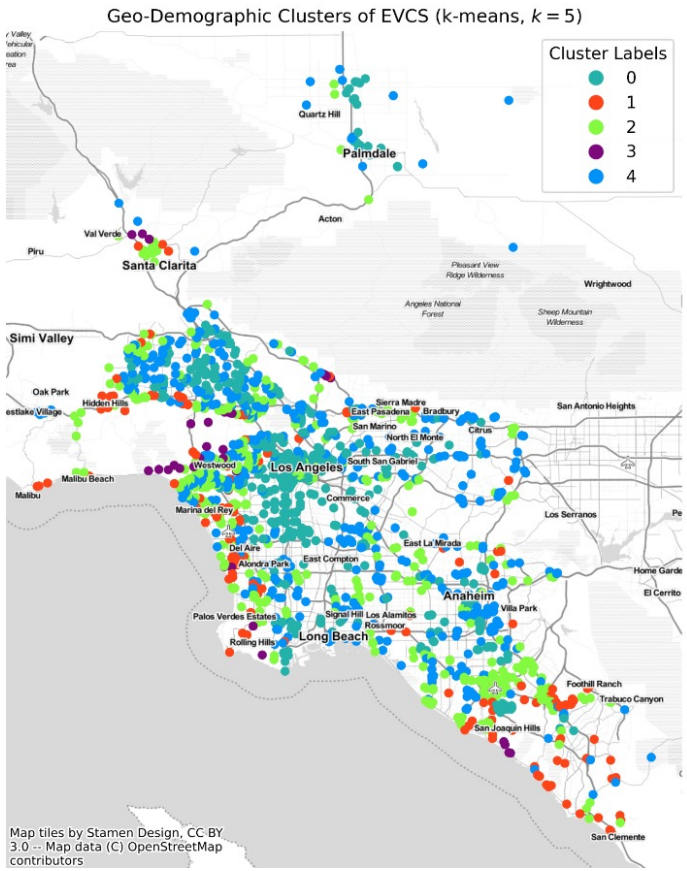



Device count and location of college clusters. Though the simulation includes a small number of agents located in Puerto Rico and Hawaii, they were excluded from this figure

Sample Project: Spring 2021

# Southern California Electric Vehicle Charging Station Network Analysis

Group: Jarvis Yuan, Justin Wong, Danny Ha, Jing Xu



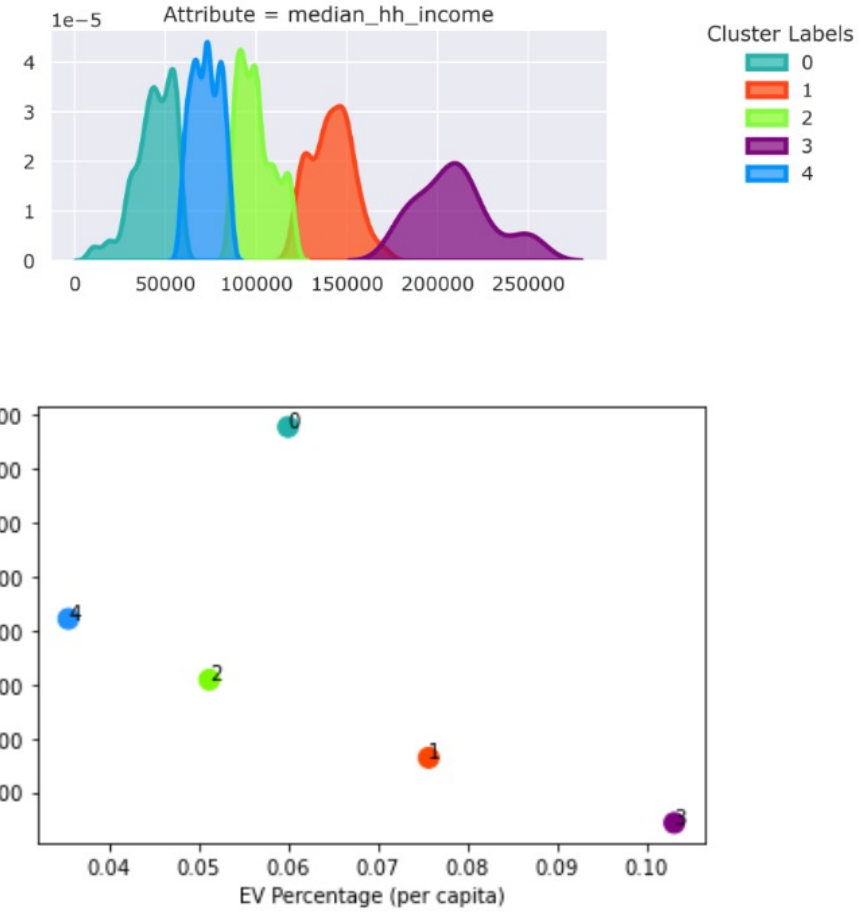




**Figure 31.** EV Ownership Percentage and Population Density Correlation (by Cluster)

From: Lecture 5 and 6 (Clustering Census Data)

# Half Marathon Running Routes in Berkeley

by Conner McGraw, Juan Hernandez, Arman Ramezani

**Objective:** we set out to design routes with less change in elevation, reached every neighborhood, and have roundess.

**Paths Between Communities:**

We use each of the designated community nodes and develops all the shortest paths between one another. Ideally, this would create a rounder route more typical of a running route.

## All Paths Shortest Path Matrix

```
for node in range(0, len(all_betweenness_nodes)):
    for n in range(0, len(all_betweenness_dest)):
        # First find route from origin node to destination and add to list `routes2`.
        rt2 = ox.shortest_path(G_new, all_betweenness_nodes[node], all_betweenness_dest[n], weight="grade_abs")
        # This turns all NoneType into 0.
        if rt2 is None:
            rt2 = [0]
        betweenness_routes.append(rt2)
        # Next find the lengths of these routes.
        l2 = ox.utils_graph.get_route_edge_attributes(G_new, betweenness_routes[node], "length")
        betweenness_routes_lengths.append(round(sum(l2)))
```

From Lecture: 2(Metrics),3(Reading Data),10 (OSMNx), and 12 (Centralities)

# What former students say?

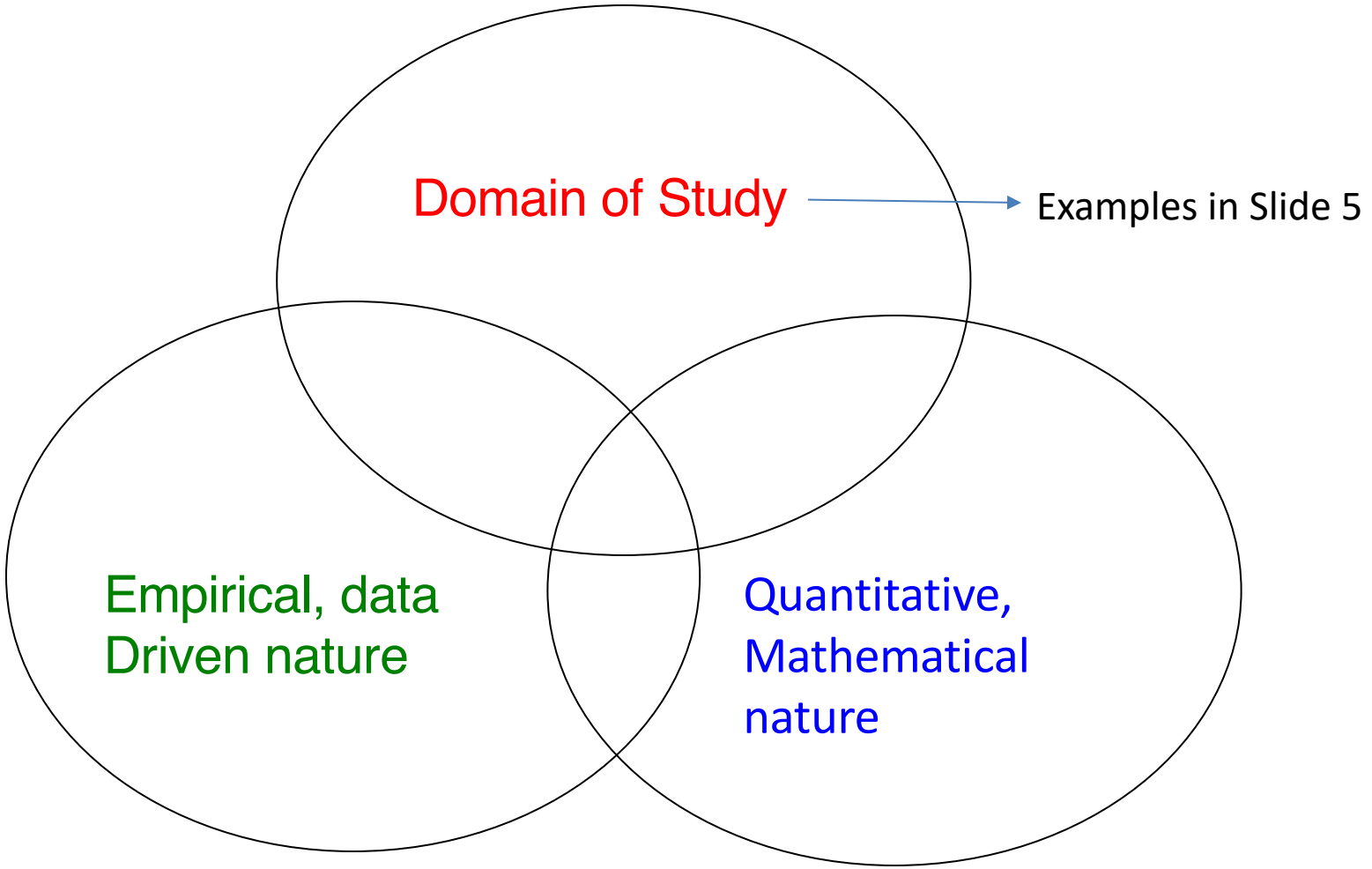
**Please provide additional comments or suggestions about the course**

| Comments |
| --- |
| It would be nice to utilize the GSI's more. There were 3 for such a small class! Maybe have them host labs, tighten up the assignments, give them a chance to guest lecture on an interesting topic, etc. Prof Eric Van Dusen's 88 class on economic models did a really great job of this so you could always exchange ideas with him. |
| I really enjoyed this class A LOT. The only data science class I have taken before was Data8 and this class does an excellent job bridging data science into other fields, especially urban and city planning. As an urban studies major, through this class, I have grown an appreciation for data science and inspired me to take on more projects in the future. This class we had a final project where we applied everything we learned from the semester into one capstone–like experience. I really enjoyed applying everything I learned from class into one final project. From this experience, I will want to continue working with what I learned from class to a possible future project or senior thesis. |
| not having coding prereqs makes the classs really inefficient since theres so much emphasis on coding in the assignments and class but so much range in coding levels, doesnt make sense to me. |

This class has a co–req of data8 but the way it is currently taught, I think it would be appropriate to have a coding class prereq like cs61b and data science co–req like data100. For example, we were working with clustering around the same time data100 started teaching the material.

Despite being on Zoom, Professor Gonzalez made everything manageable. As mentioned earlier, I am a total beginner to the class material and I only took Data8 as a prerequisite for this class. Coding was difficult for me but everything was well planned and the GSIs and professor was accessible, able to help explain any concerns with class material.

# Characteristics of Network Science



# The core Method of CE88: Network Science in a framework of Complex Systems thinking to study urban systems.

| | | |
|---|---|---|
| Lec 1: January 17th | Introduction to Network Science | |
| Lec 2: January 24th | Network Metrics | |
| Lec 3:January 31st | Network Models | Assignment 1 |
| Lec 4: February 7th | Network Models | |
| Lec 5: February 14th | Clustering Census Data | Assignment 2 |
| Lec 6: February 21st | Clustering Census Data | |
| Lec 7: February 28th | Discussion of Research Ideas | Assignment 3_I |
| Lec 8: March 7th | Weighted Networks | |
| Lec 9: March 14th | Scaling Transportation Networks | Assignment 3_II |
| Lec 10: March 21st | OSMNx | |
| March 28th | Spring Recess | |
| Lec 11: April 4th | Dynamics on Networks | Assignment 4 |
| Lec 12: April 11th | Centralities and Network Resilience | |
| Lec 13: April 18th | Discussion of Project Ideas | Paper Draft |
| Lec 14: April 25th | Class Summary | |
| Final Presentations | Project Presentations (May 2nd, 3rd and 4th) | |

Assignments 1-4: 60%
Participation: 10%
Project Preparation: 15%
Final Paper: 10% and Presentation 5%

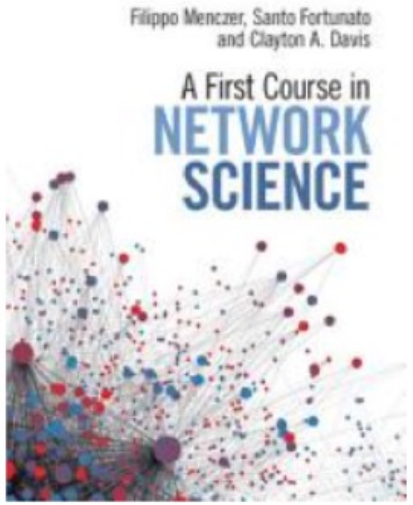
Problem sets

| | Covers | Assigned | Due |
|---|---|---|---|
| Assignment 1 | L1-L2 | 01/20 | 01/31 |
| Assignment 2 | L3-L4 | 02/03 | 02/14 |
| Assignment 3 part 1 | L5-L6 | 02/17 | 02/28 |
| Assignment 3 part 2 | L5-L6 | 03/03 | 03/14 |
| Assignment 4 | L7-L10 | 03/17 | 04/04 |
| Checkup (paper draft) | L1-L12 | 04/07 | 04/18 |
| Final Project | L1-14 | | 05/02, 05/03 and 05/04 |

For example:

Textbook:

*FIRST COURSE IN NETWORK SCIENCE,* ***Author:*** *MENCZER,* ***ISBN:*** *9781108471138*




**FIRST COURSE IN NETWORK SCIENCE**

**Author:** MENCZER

**ISBN:** 9781108471138

Purchase

Rent

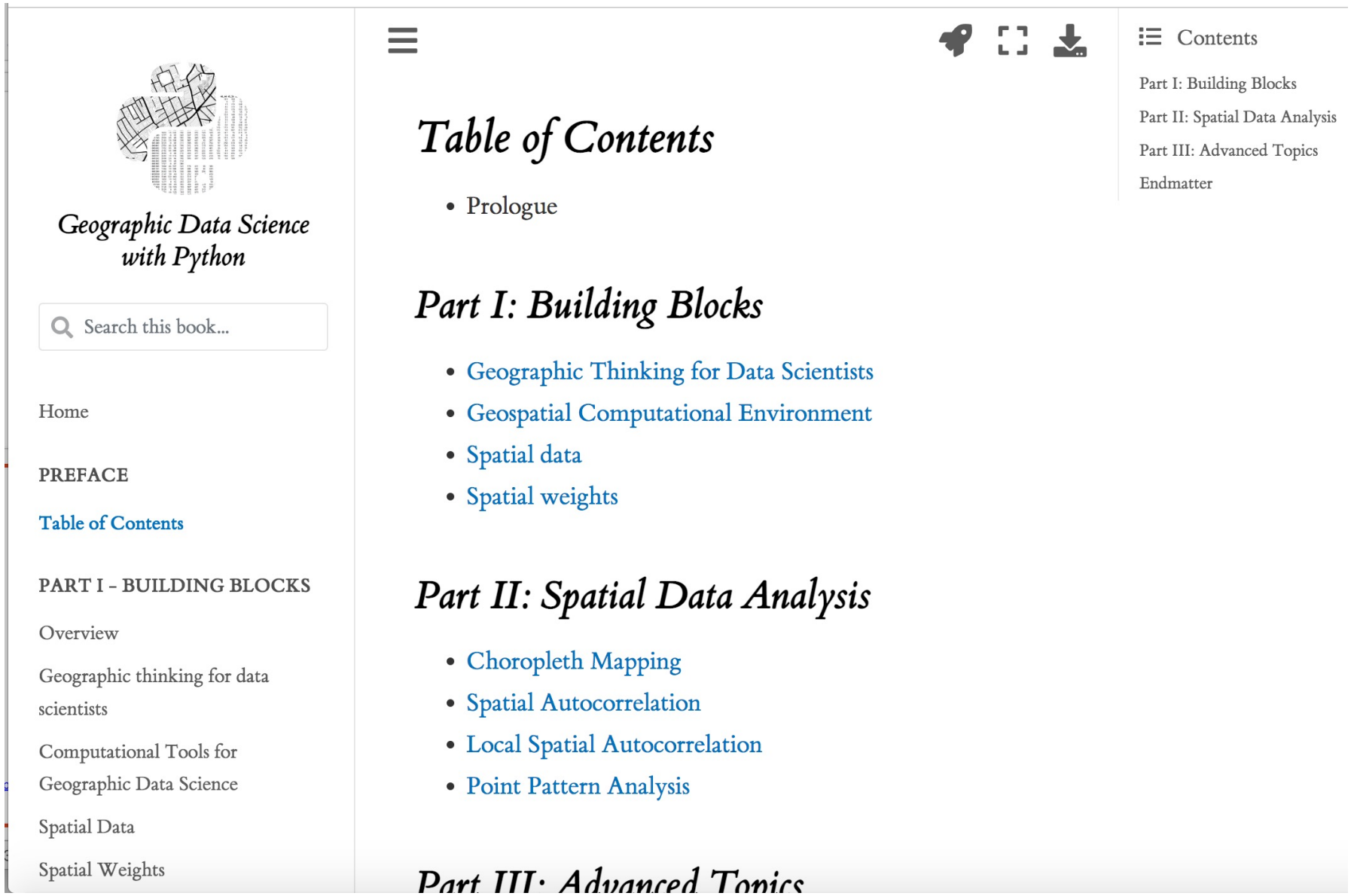
Borrow

Purchase

Purchase

$40.49

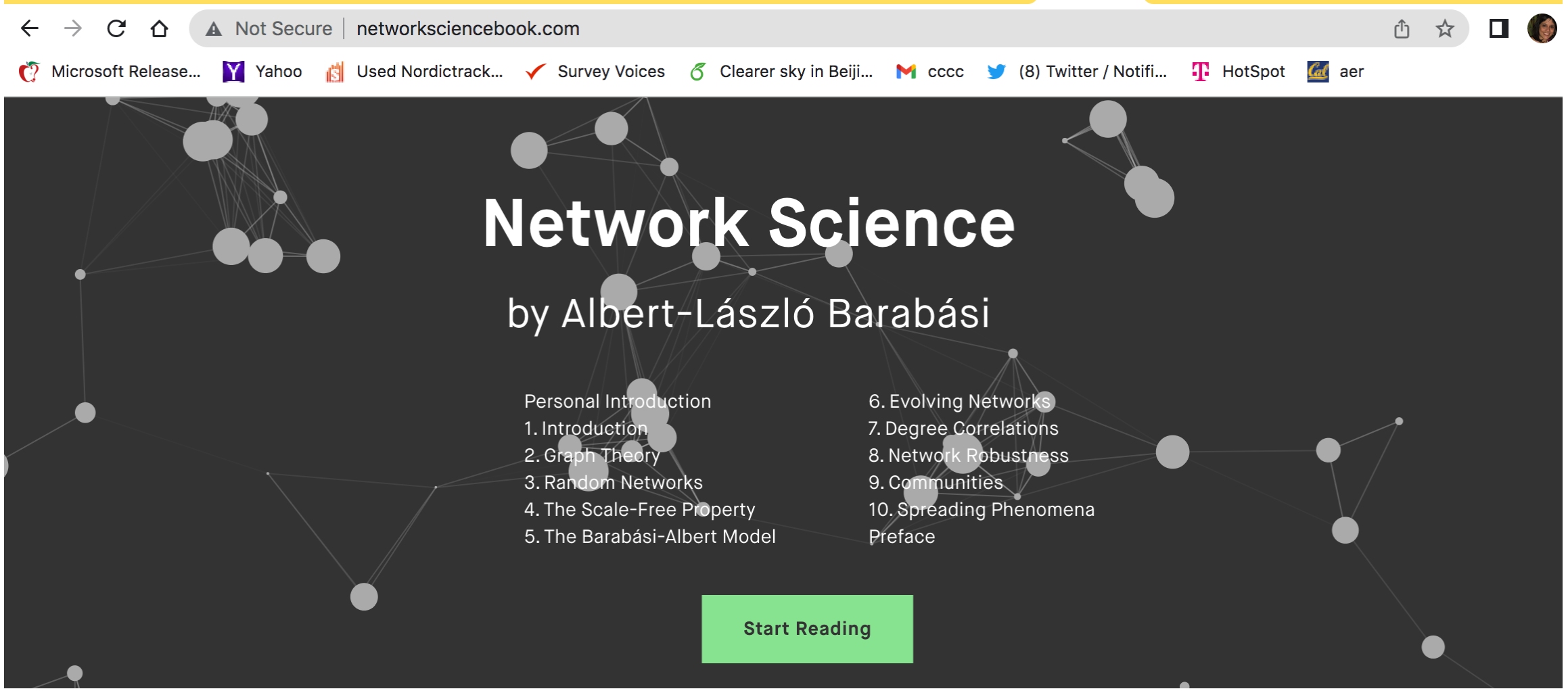
For most lectures




https://geographicdata.science/book/notebooks/10_clustering_and_regionalization.html

For Lectures 5 and 6

http://networksciencebook.com/




Free on-line book

**Course Policy:**

- If assignments are uploaded after the grace period 20% will be deducted and the new date needs to be agreed with the instructor (exceptions to the deduction apply)
- A laptop is needed for class exercises. If you plan to take the class and do not have one, please see the instructor in advance.
- Please sign up at bcourses. I will confirm your enrollment for the course, then you will be able to see the course page.
- Regular attendance and participation are essential and expected.
- It is expected that the students are familiar with academic honesty, lack of knowledge of the policy is not a reasonable explanation for a violation.

Unless stated otherwise, problem sets should be solved individually. Electronic versions must be uploaded to the bcourses site.

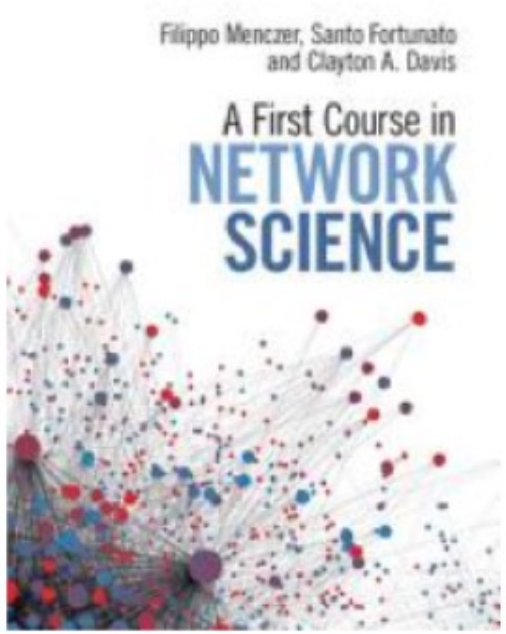
20% for late assignments reduction for late assignments more than 48 hours (it means there is a 48hrs grace period)

Students that need extra time for deadlines (w/wo disability request) are granted it

Prof. Marta C. Gonzalez Office Hours:
Tuesdays 4:00-5:30 PM @Room 406c and @ Zoom
Thursdays 9:00-10:00 AM by appointment via email (20 min slots) @ Zoom
Room 406c; martag@berkeley.edu,
https://berkeley.zoom.us/j/4124372482

Textbook:

*FIRST COURSE IN NETWORK SCIENCE,* ***Author:*** *MENCZER,* ***ISBN:*** *9781108471138*




**FIRST COURSE IN NETWORK SCIENCE**

**Author:** MENCZER

**ISBN:** 9781108471138

Cal Purchase

Rent

Borrow

Purchase

Purchase

https://github.com/CambridgeUniversityPress/FirstCourseNetworkScience/tree/master/tutorials
https://www.amazon.com/First-Course-Network-Science/dp/1108471137

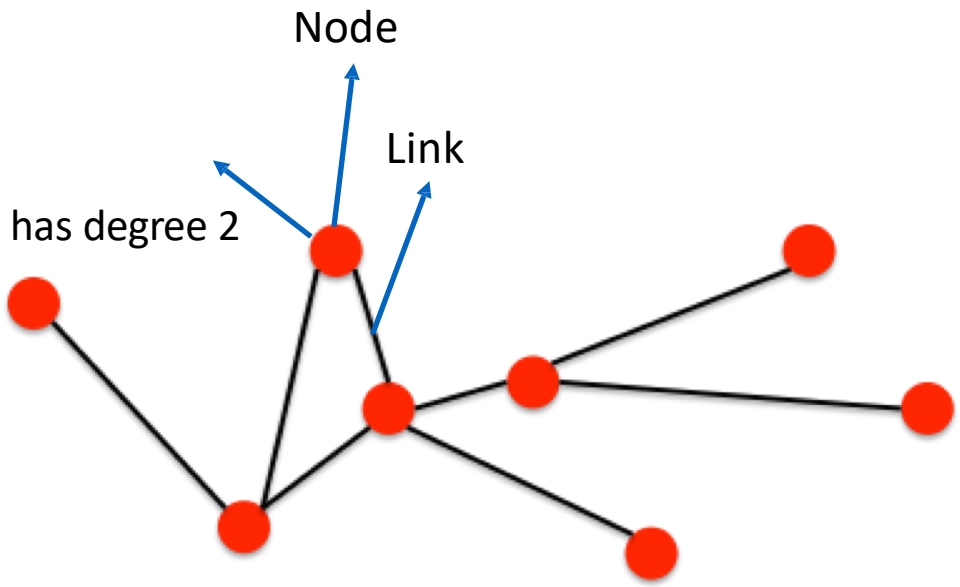
# Lecture 1: Introduction to Network

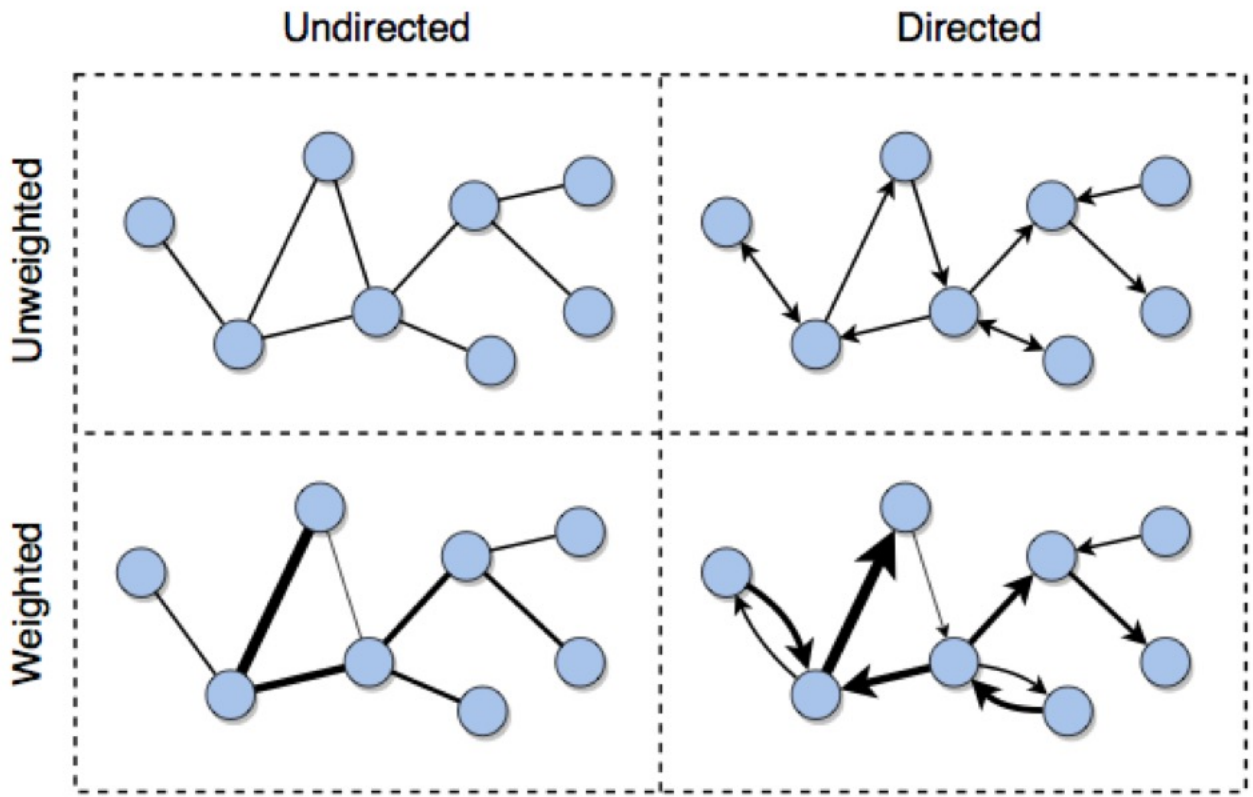
(Lecture 2: Network Metrics)

Assignment 1: Lectures 1 and 2
Due: January 31st

# Definition

- A **network** (graph) is a structure used to model pairwise relations between objects. A network is made up of entities (aka **nodes**, vertices) and the relationships between them (aka **links**, edges)
- Number of edges of incident to the vertex is **the node degree**






**Can you think of a few examples in each of these categories?**

# Formal definitions

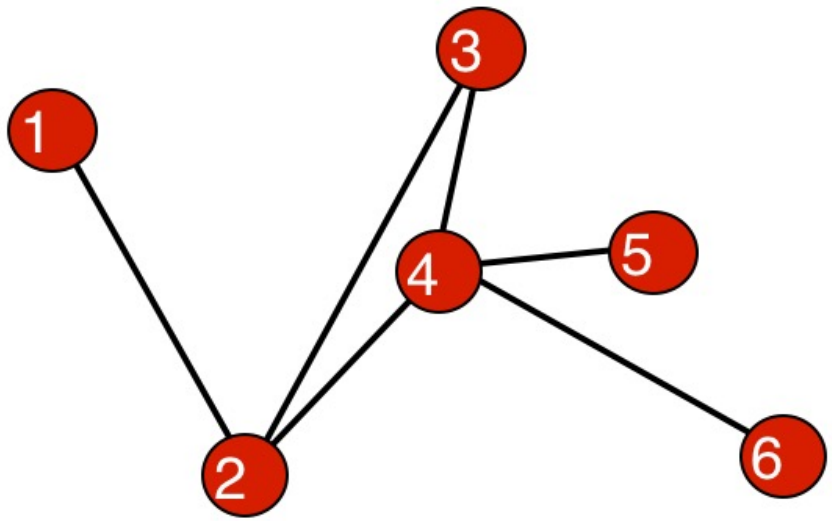
- A **network** or **graph** *G* has two parts, a set of *N* elements, called **nodes** or **vertices**, and a set of *L* pairs of nodes, called **links** or **edges**. The link *(i, j)* joins the nodes *i* and *j*. Two nodes are **adjacent** or **connected** or **neighbors** if there is a link between them.

- A network can be **undirected** or **directed**. A directed network is also called a **digraph**. In directed networks, links are called **directed links** and the order of the nodes in a link reflects the direction: the link *(i, j)* goes from the **source** node *i* to the **target** node *j*. In undirected networks, all links are bi-directional and the order of the two nodes in a link does not matter.

- A network can be **unweighted** or **weighted**. In a weighted network, links have associated **weights**: the **weighted link** *(i,j,w)* between nodes *i* and *j* has weight *w*. A network can be both directed and weighted, in which case it has directed weighted links.

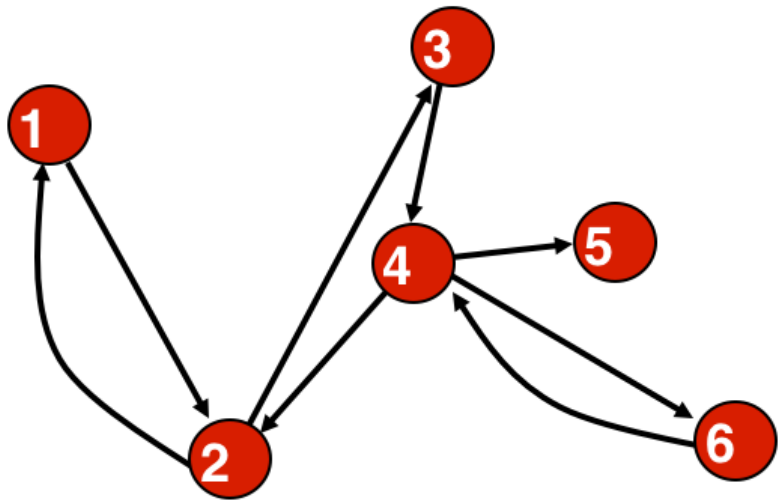
Source: A first course in Network Science Chapter 1

# Python and NetworkX




```
import networkx as nx # always!

G = nx.Graph()
G.add_node(1)
G.add_nodes_from([2,3,…])
…
G.add_edge(1,2)
G.add_edges_from([(2,3),(2,4),…])
…
G.nodes()
G.edges()
G.neighbors(4)

for n in G.nodes:
    print(n, G.neighbors(n))
for u,v in G.edges:
    print(u, v)
```

# Directed networks



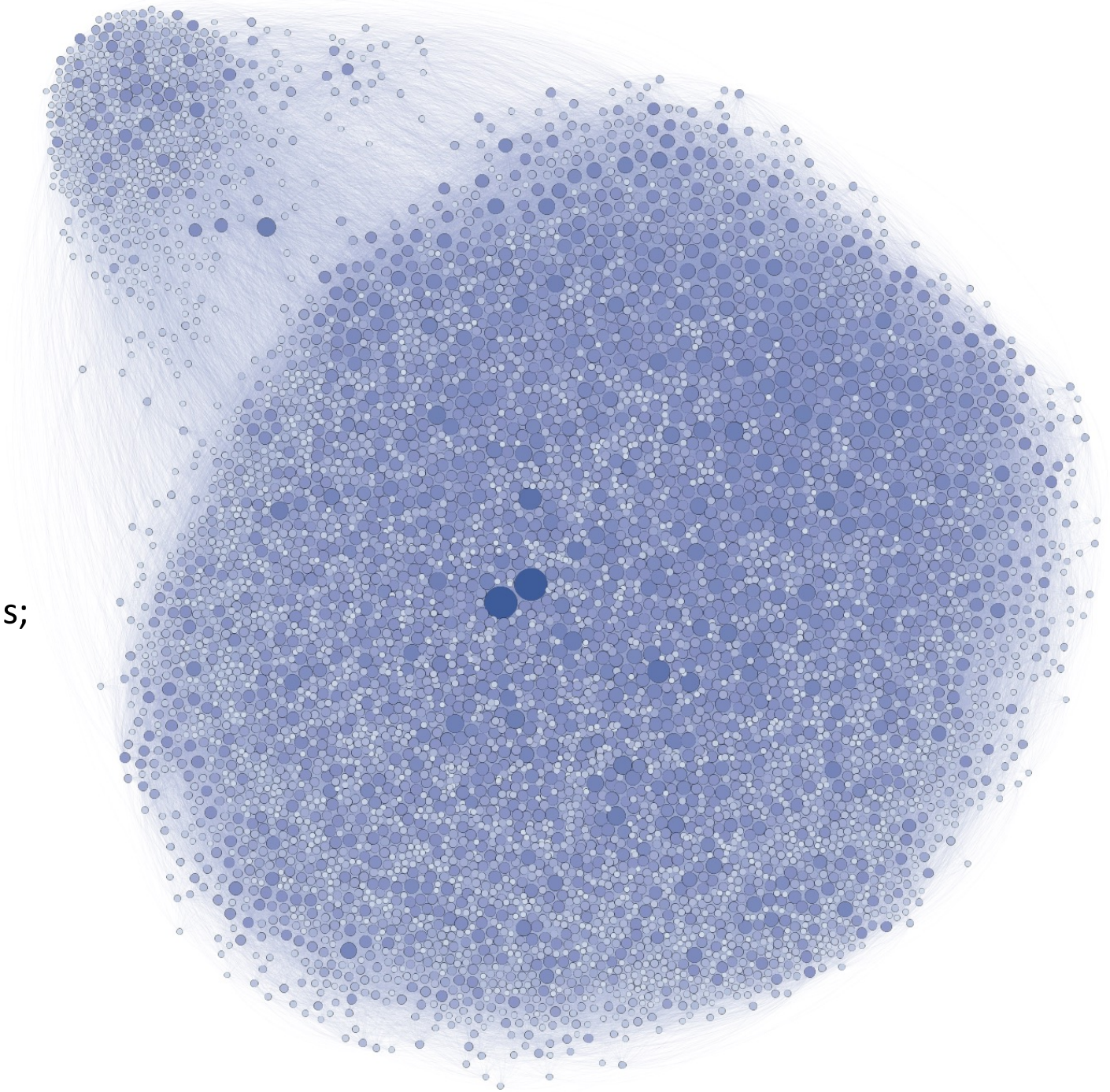

```
import networkx as nx # don't forget!

D = nx.DiGraph()
D.add_edge(1,2)
D.add_edge(2,1)
D.add_edges_from([(2,3),(3,4),…])
…
D.number_of_nodes()
D.number_of_edges()
D.edges()
D.successors(2)
D.predecessors(2)
D.neighbors(2)
```

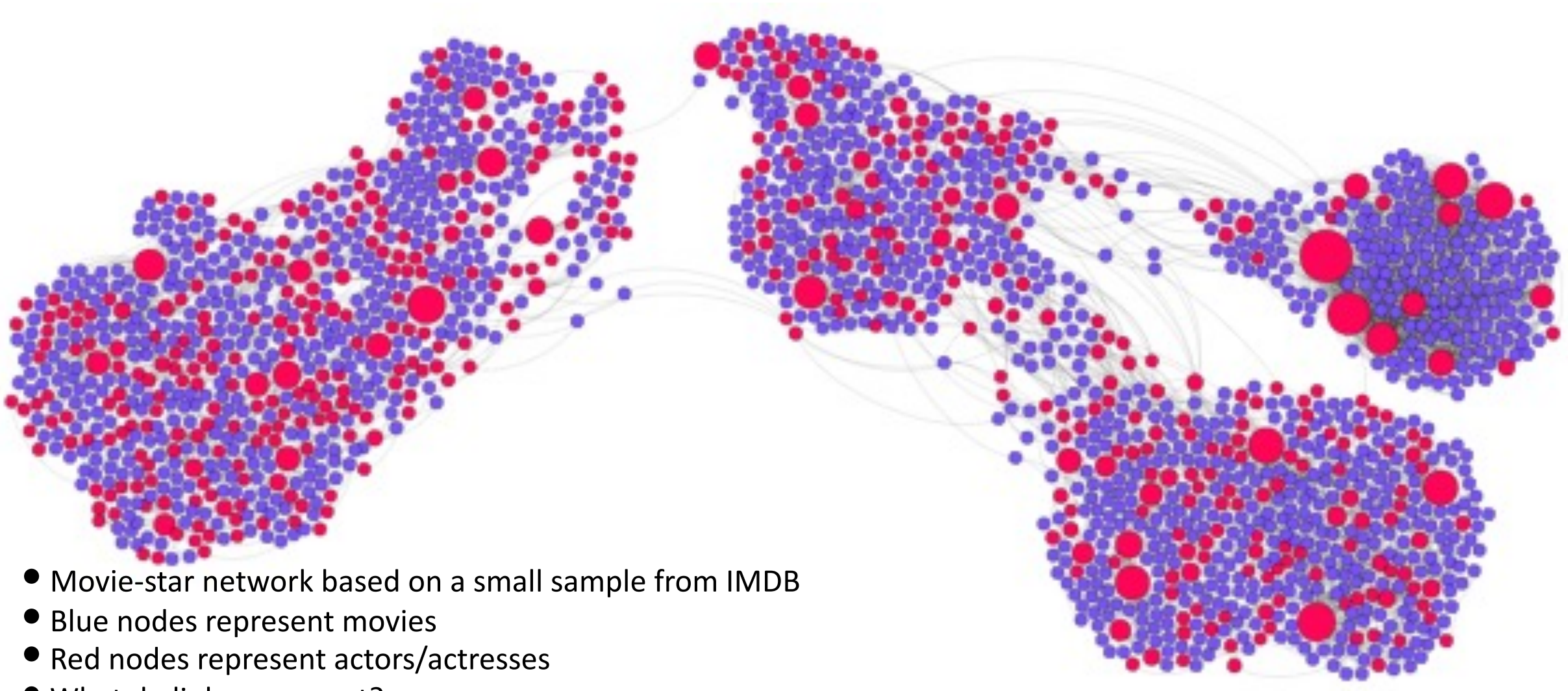
Activity 1:

Separate in 12 Groups and discuss the Networks in slides  22 to 33 reply the posted questions.

- Facebook users at Northwestern University
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger, darker nodes have more connections; what does that represent?
- What do the two clusters tell us?

Bonus: Approx. number of nodes and Links?

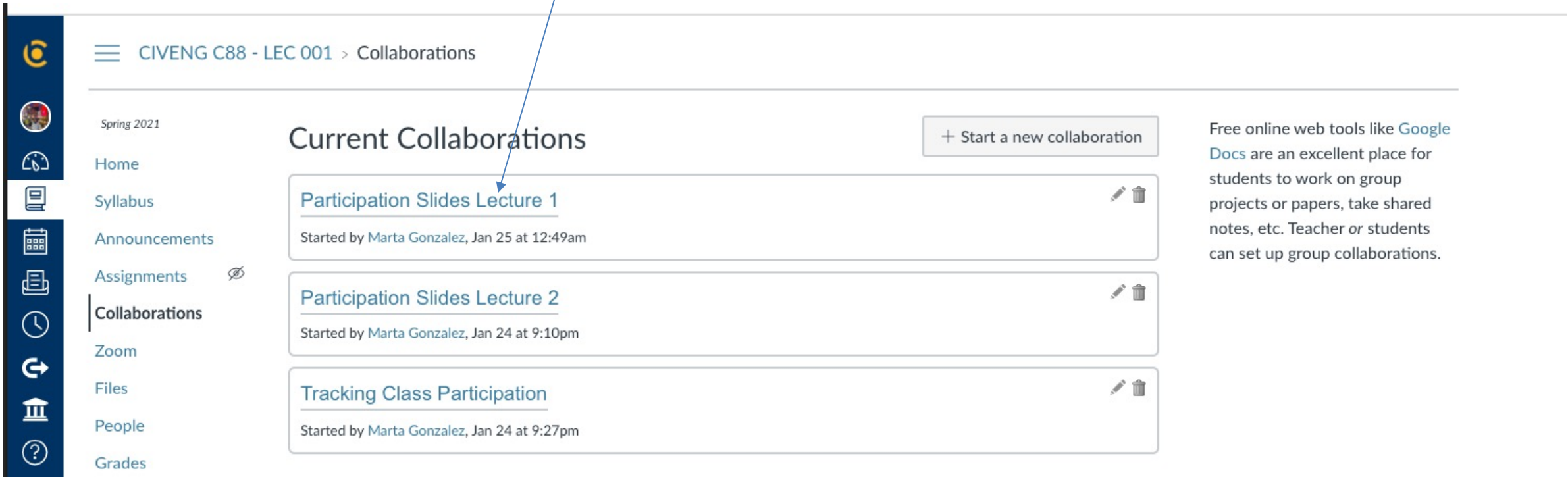




- Movie-star network based on a small sample from IMDB
- Blue nodes represent movies
- Red nodes represent actors/actresses
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what does that mean?
- What do the clusters represent?

Bonus: Approx. number of nodes and Links?

Click there to open the slides and add the answers



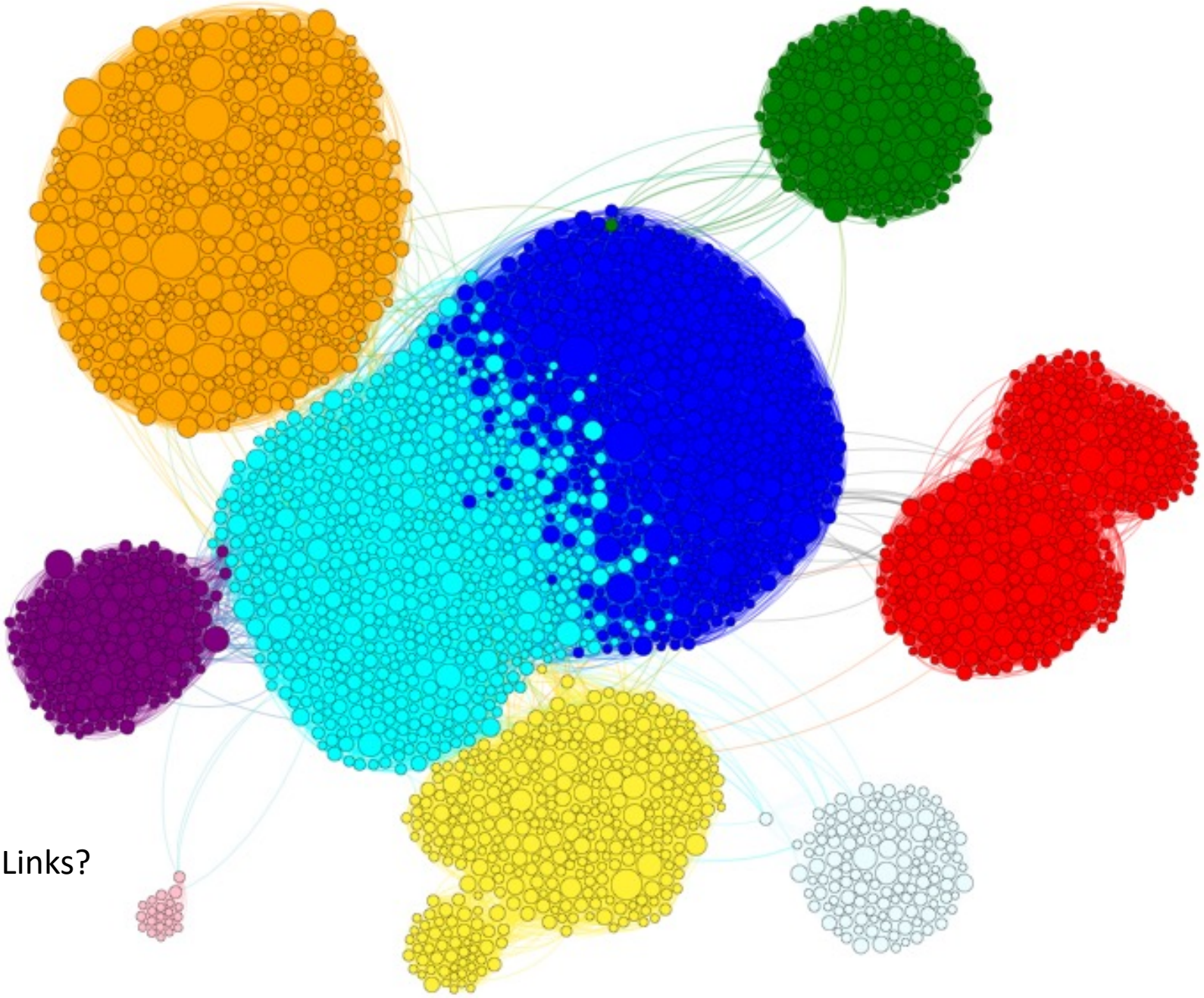

https://docs.google.com/presentation/d/1GDjVUUvgFOBS9QfG28E6Y_DVJuuETFeWtB1tBsSlHsI/edit?usp=sharing

# Activity 1:

https://docs.google.com/presentation/d/1l3d4jR_qMng37gtCR_b4US6S6GhGSEihB1Y9sMRAVZk/edit?usp=sharing

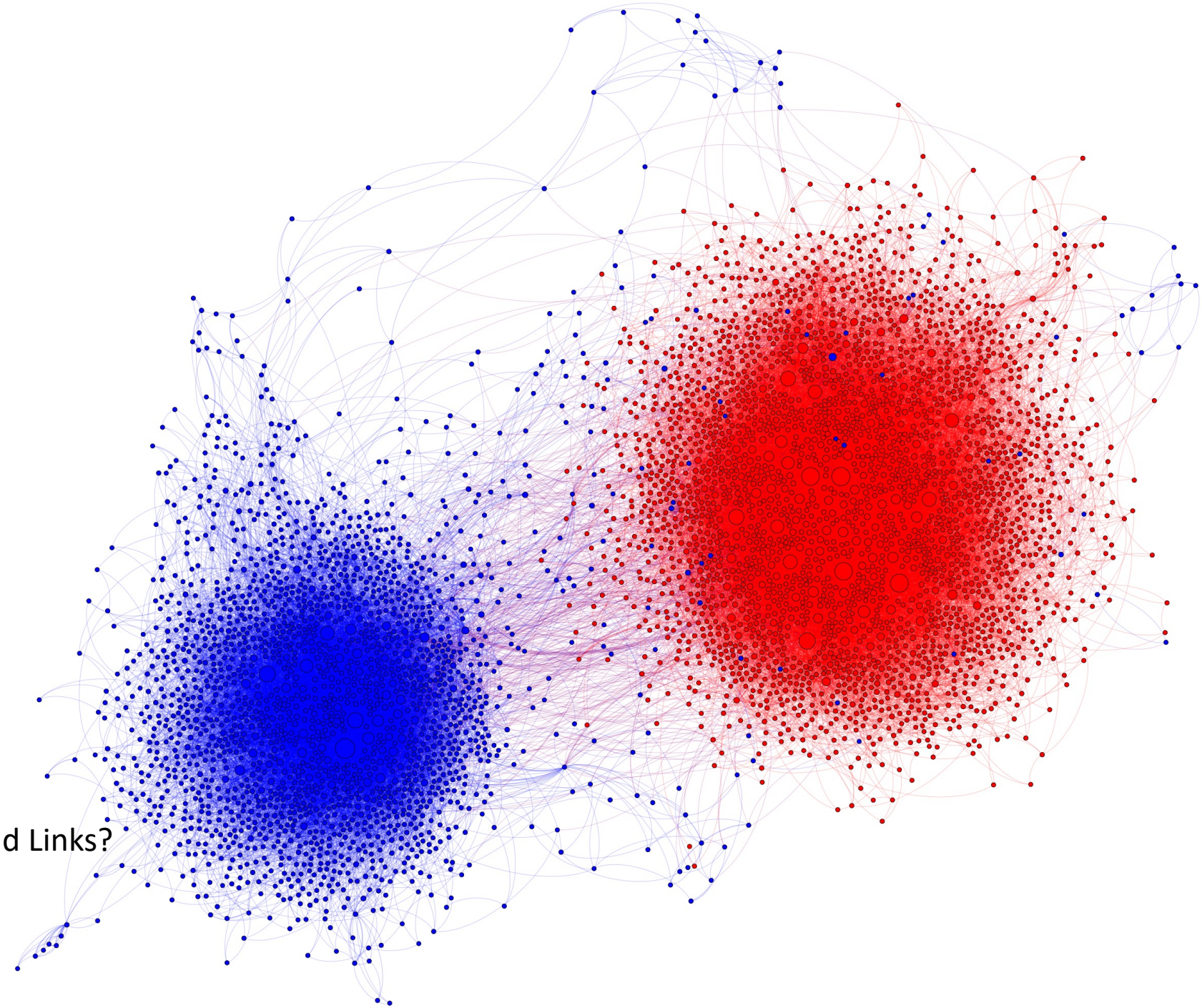
- Movie co-star network based on a small sample from IMDB
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what does that mean?
- What do the clusters represent?

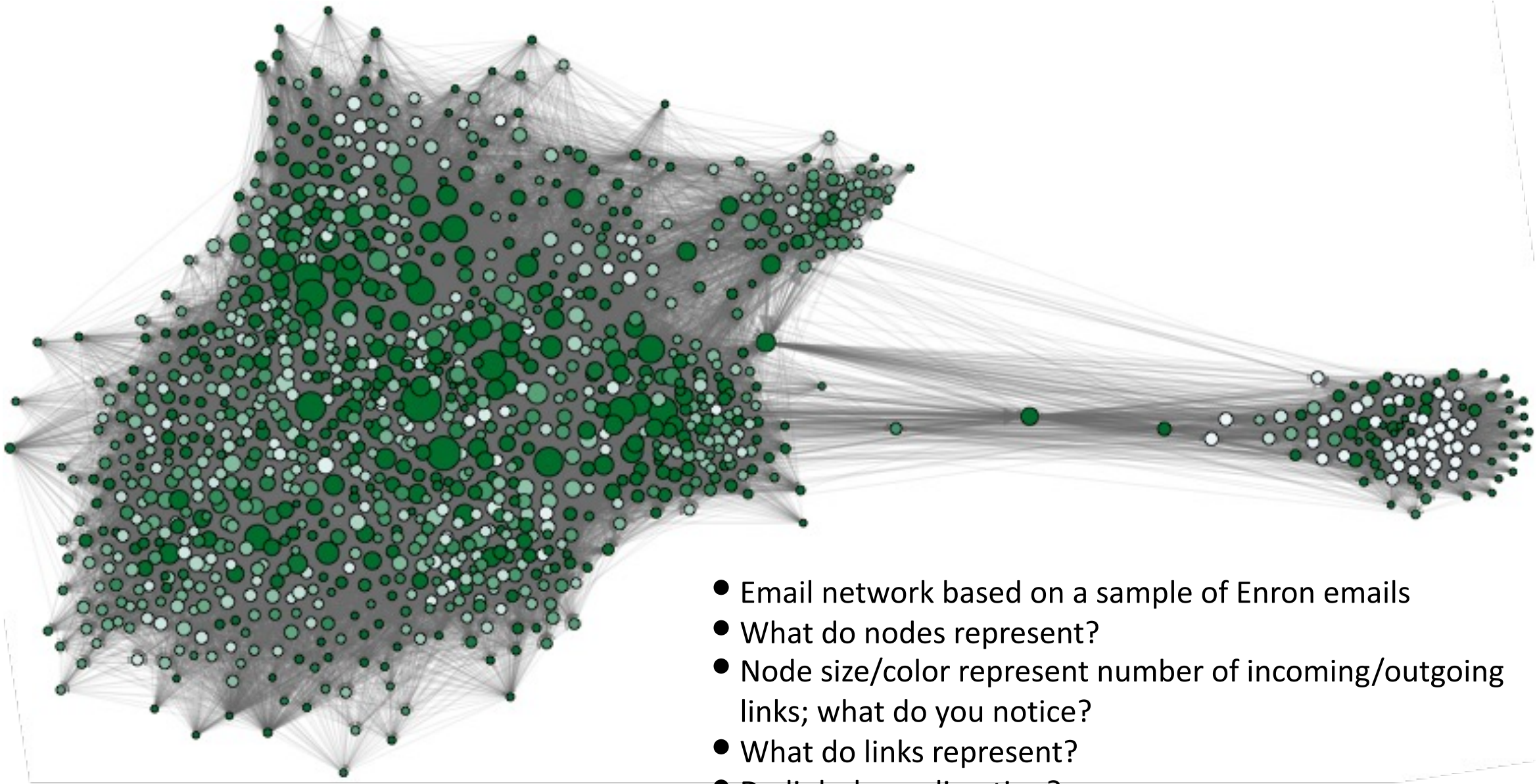
Bonus: Approx. number of nodes and Links?



- Retweet network on Twitter, based on political posts during 2010 US election
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what does that mean?
- What do the clusters and colors represent?

Bonus: Approx. number of nodes and Links?

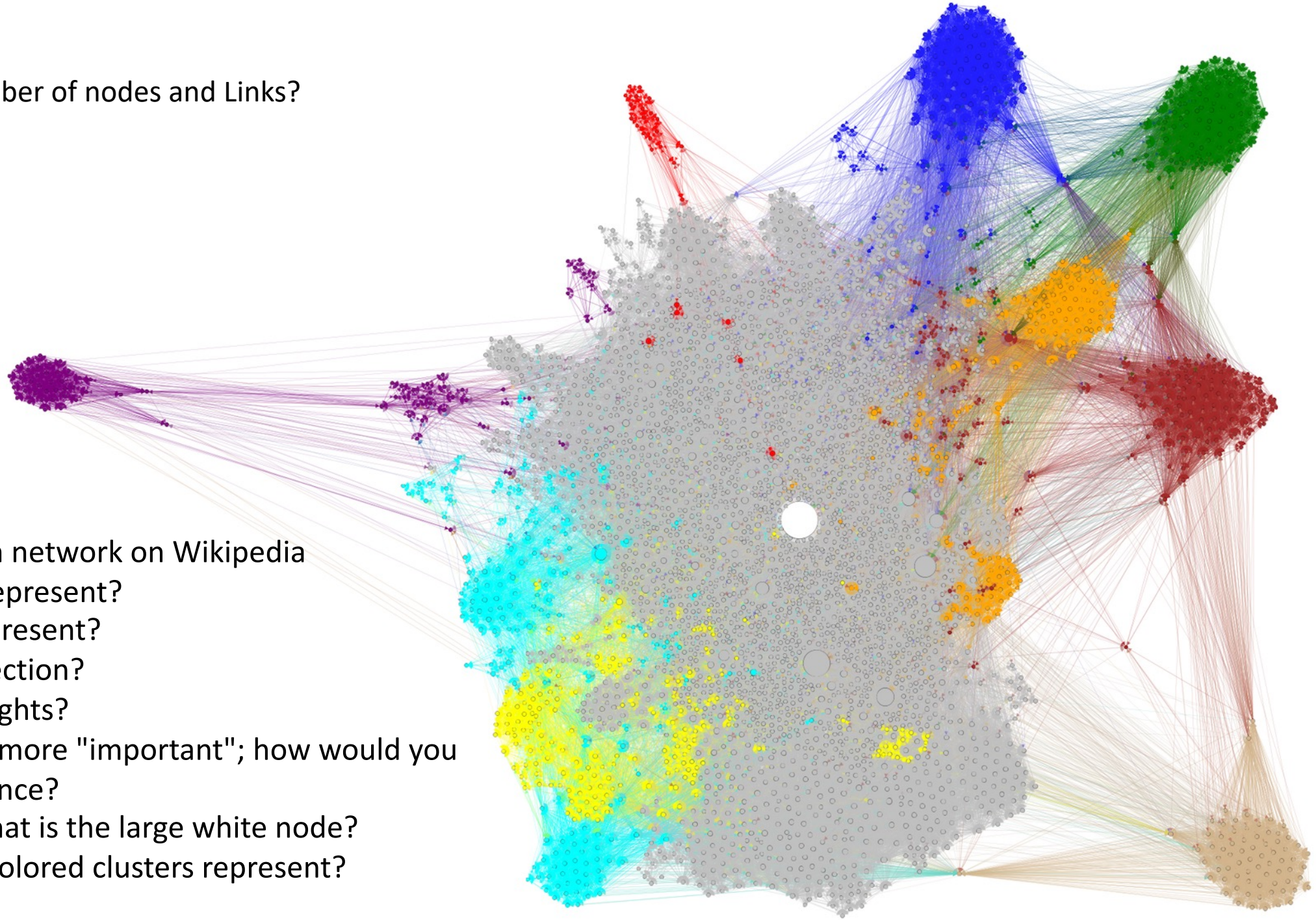




- Email network based on a sample of Enron emails
- What do nodes represent?
- Node size/color represent number of incoming/outgoing links; what do you notice?
- What do links represent?
- Do links have direction?
- Do links have weights?
- What do the clusters represent?

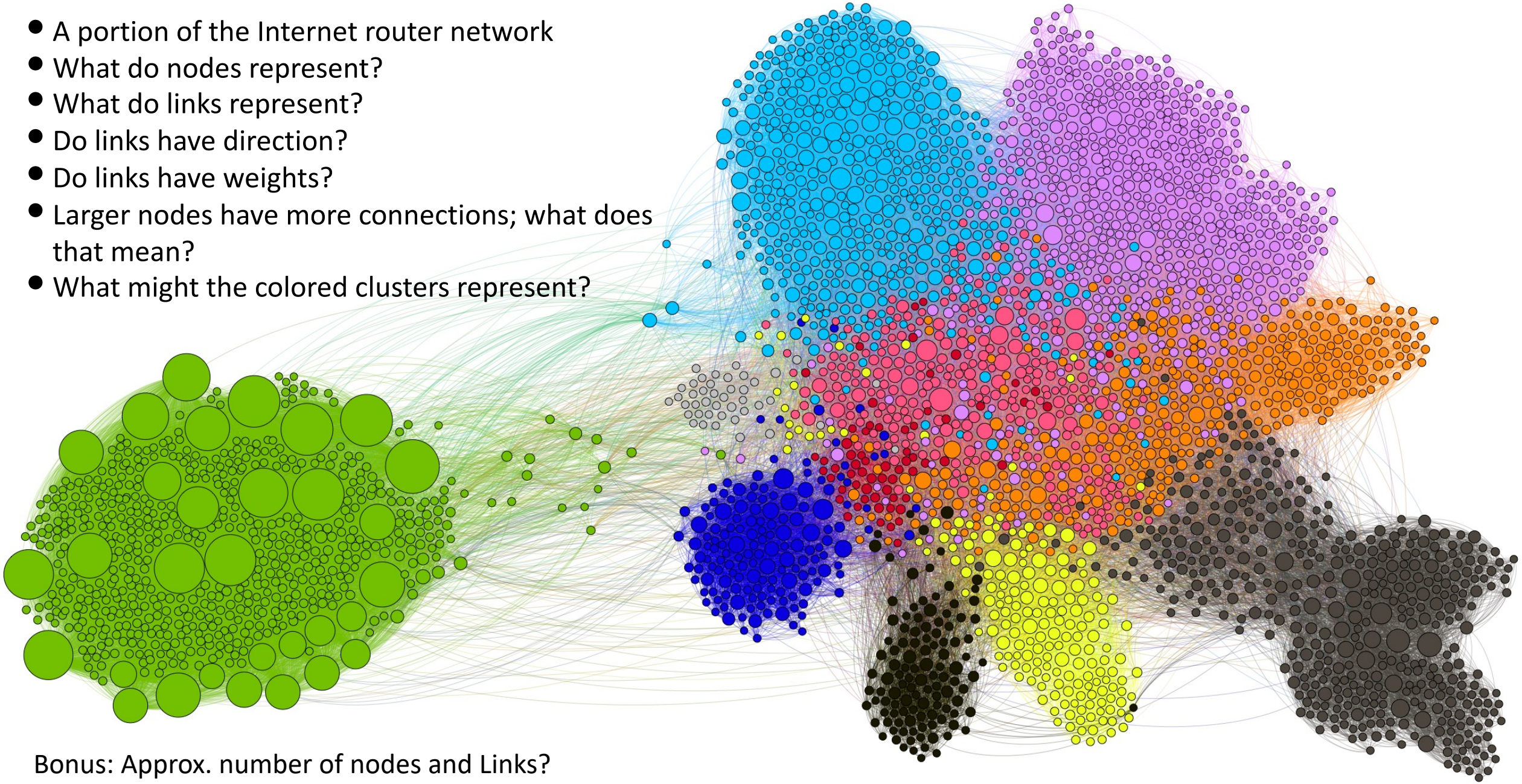
Bonus: Approx. number of nodes and Links?

Bonus: Approx. number of nodes and Links?

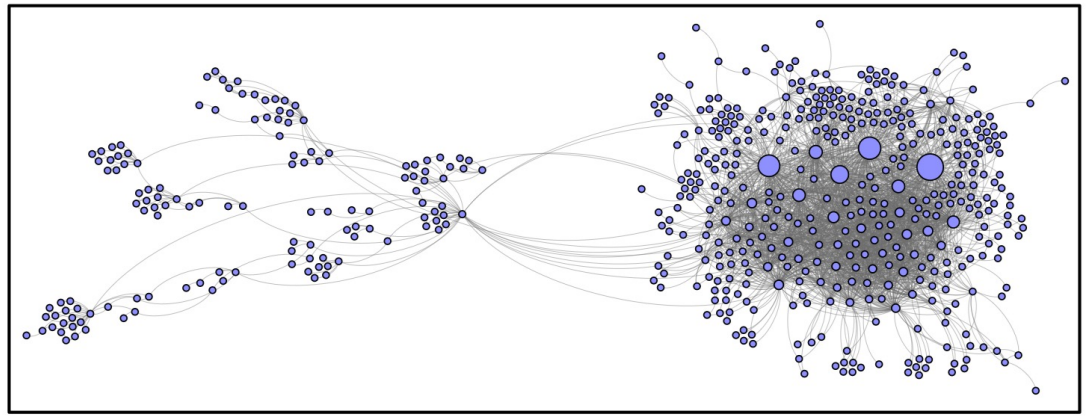


- Math information network on Wikipedia
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes are more "important"; how would you measure importance?
- Can you guess what is the large white node?
- What might the colored clusters represent?

- A portion of the Internet router network
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what does that mean?
- What might the colored clusters represent?



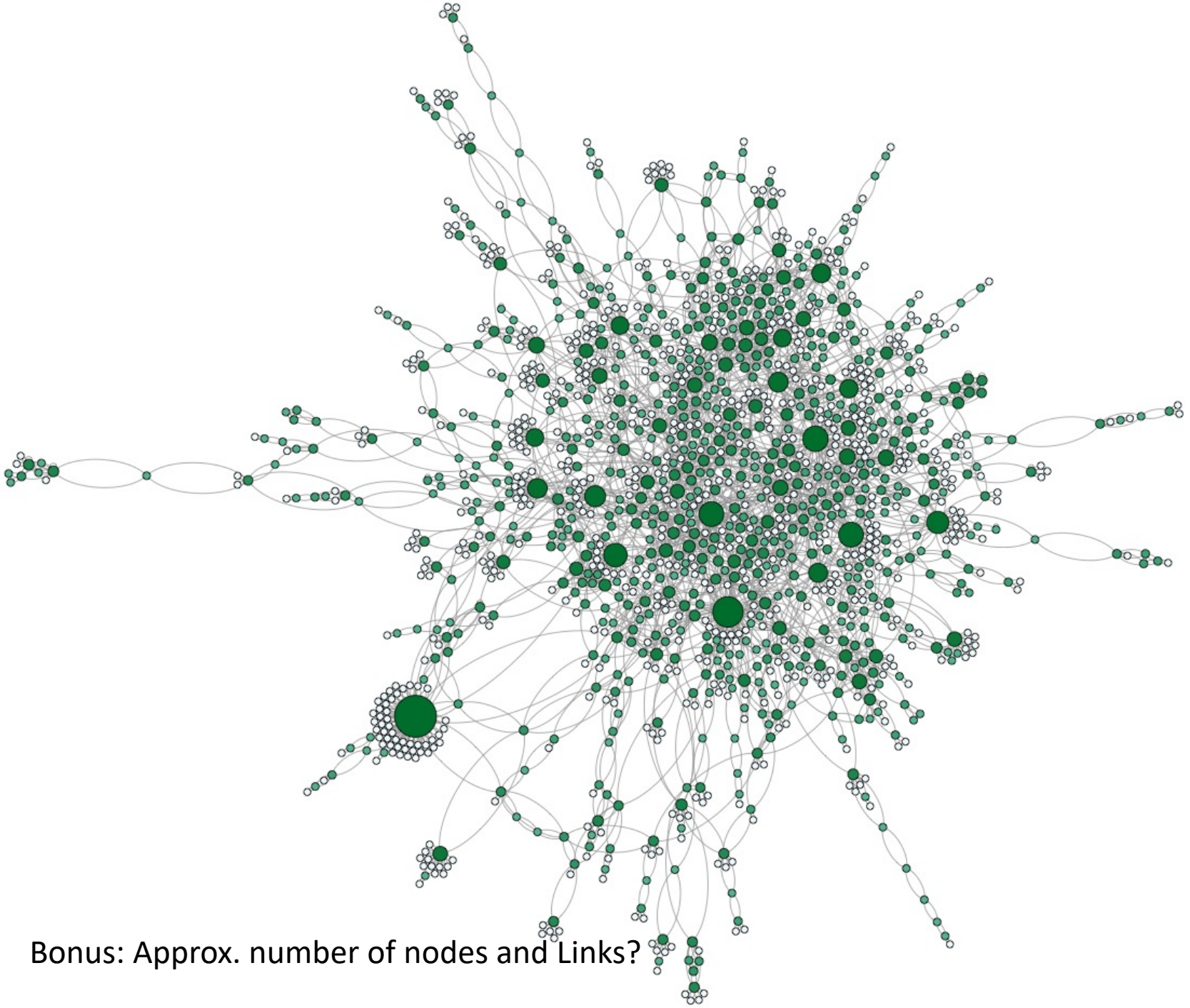
Bonus: Approx. number of nodes and Links?





- US air transportation network
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what do they represent?
- What do the layouts represent?

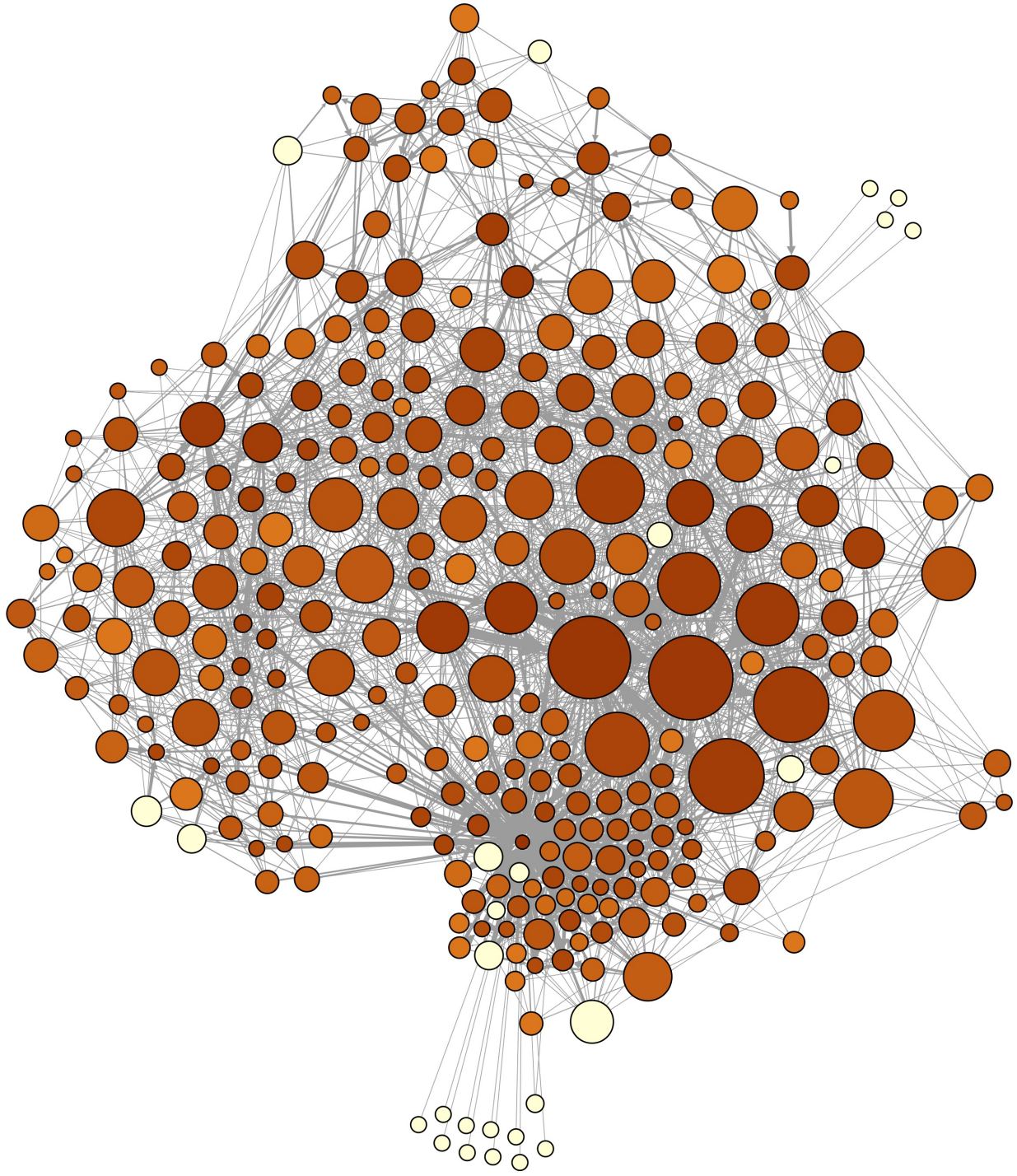
Bonus: Approx. number of nodes and Links?



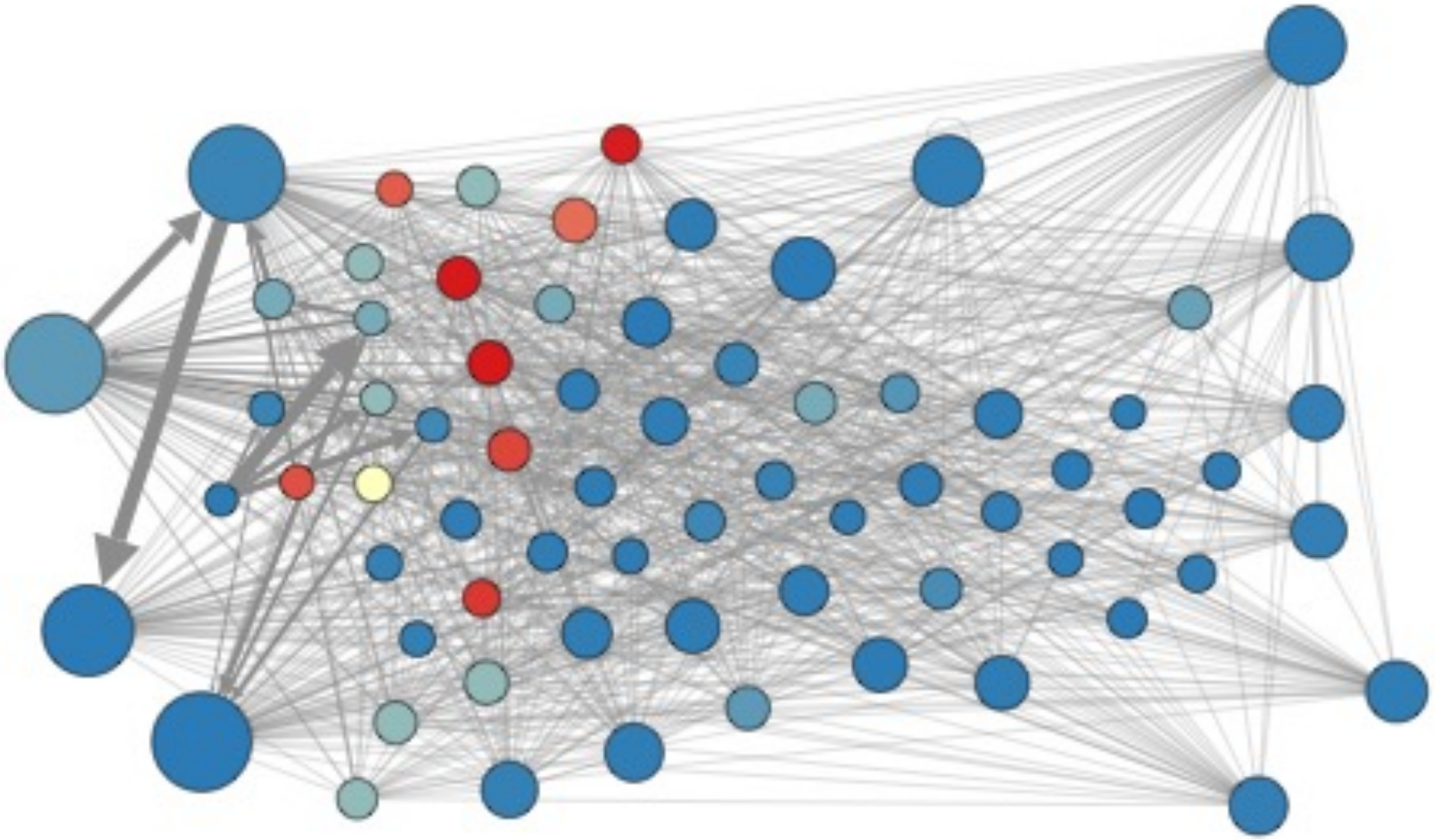
- Protein interaction network of yeast
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger nodes have more connections; what does that mean?
- What do the clusters represent?

Bonus: Approx. number of nodes and Links?

- Neural network of the roundworm c. elegans
- What do nodes represent?
- What do links represent?
- Do links have direction?
- Do links have weights?
- Larger/darker nodes have more outgoing/incoming connections; what does that mean?



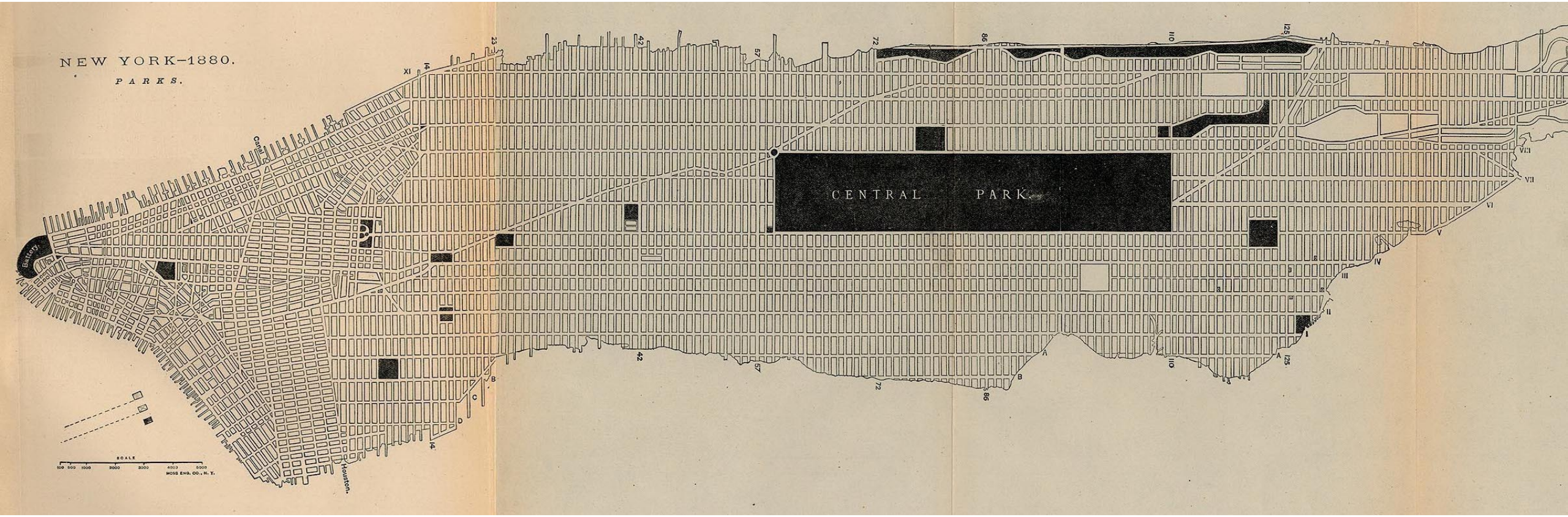
Bonus: Approx. number of nodes and Links?



- Food web of species in the Florida Everglades
- What do nodes represent?
- What do links represent?
- Do links have direction? What does it represent?
- Do links have weights? What do they represent?
- Larger nodes have more incoming links; what are they?
- Red nodes have more outgoing links; what are they?

Bonus: Approx. number of nodes and Links?

# What are the nodes and links of the network represented in this street map?




Map of New York in 1880. From Report on the Social Statistics of Cities, Compiled by George E. Waring, Jr., United States Census Office, 1886. Image courtesy of University of Texas Libraries

Bonus: Approx. number of nodes and Links?

Option 1: Work remotely


https://datahub.berkeley.edu/hub/login?next=%2Fhub%2F

jupyterhub

Logout | Control Panel

Files | Running | Clusters | Nbextensions

Select items to perform actions on them.

Download Directory | Upload | New | ↻

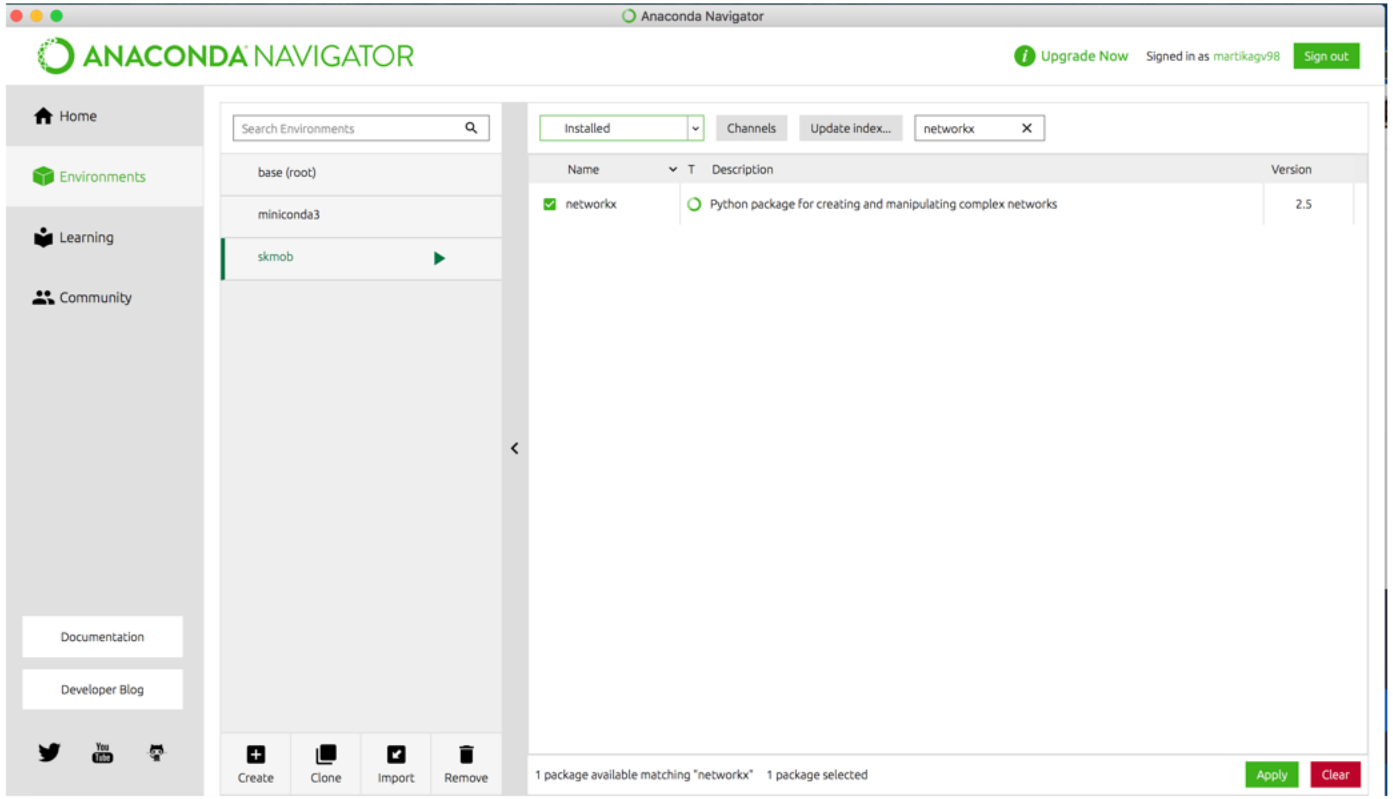
0 / CE88-CYPLAN88_Spring23 / Lecture 1

| | Name ↓ | Last Modified | File size |
|---|---|---|---|
| | .. | seconds ago | |
| ☐ | MyFirstNetwork_Exercise-Soln.ipynb | seconds ago | 60.1 kB |
| ☐ | MyFirstNetwork_Exercise-woSoln.ipynb | seconds ago | 62 kB |

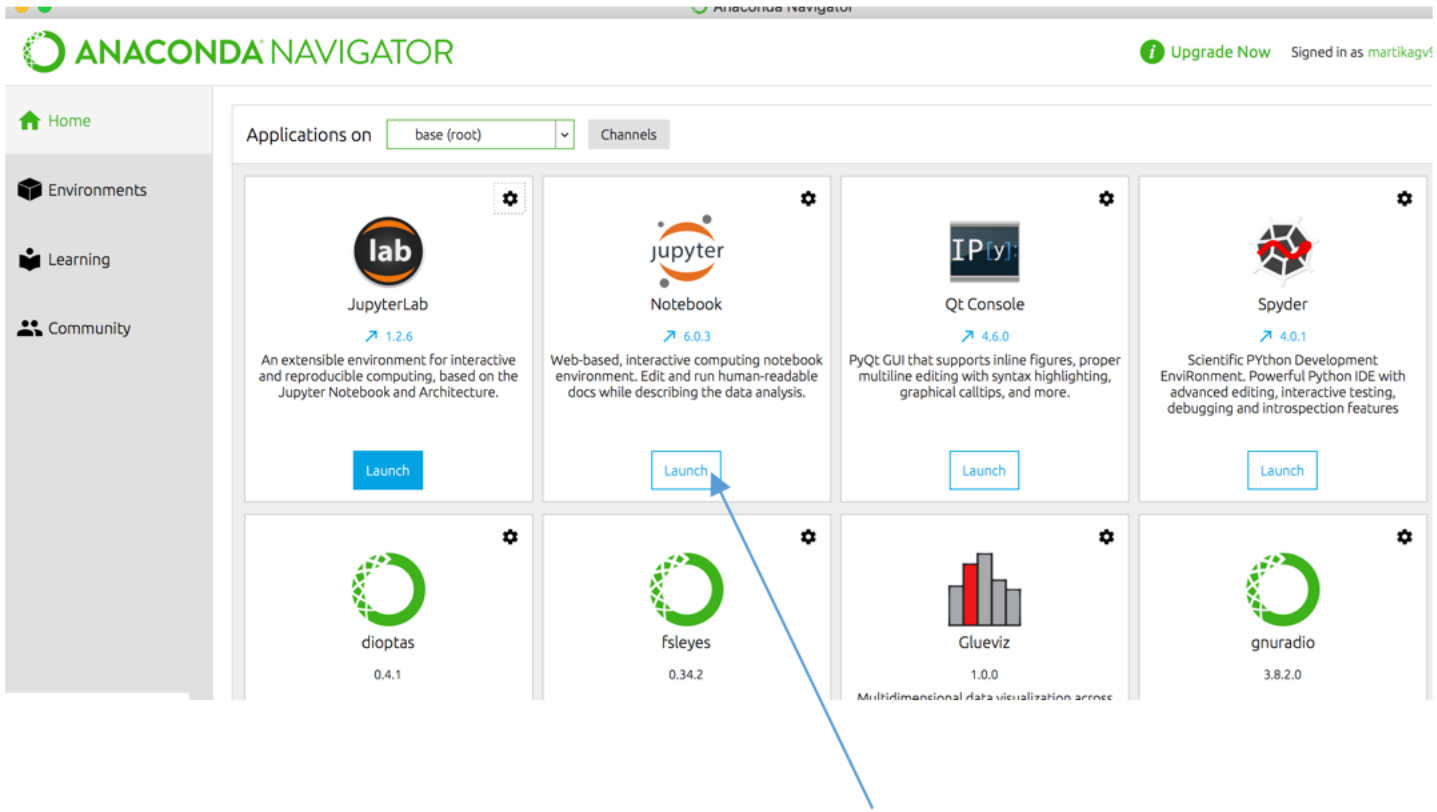
Option 2:

Follow [InstallationInstructions_v2.pdf]

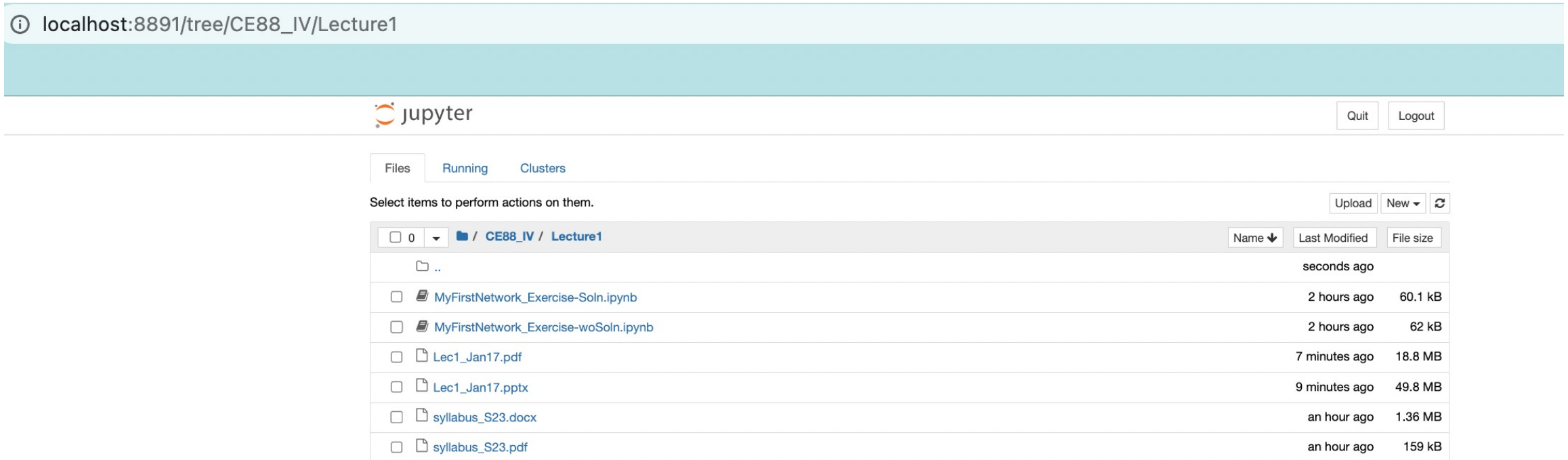
Install and Open the Anaconda Navigator to Install Networkx



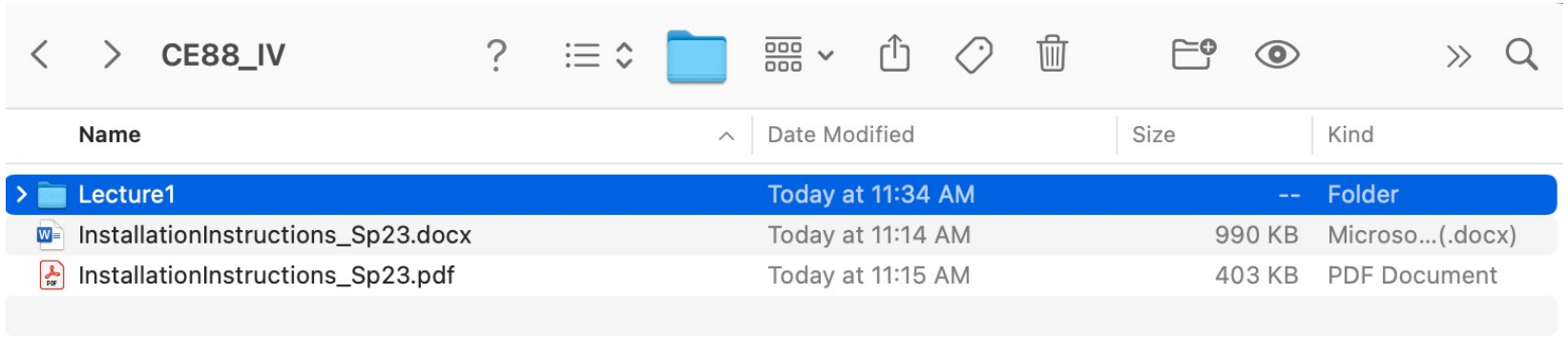
Launch from there the Jupyter Notebook

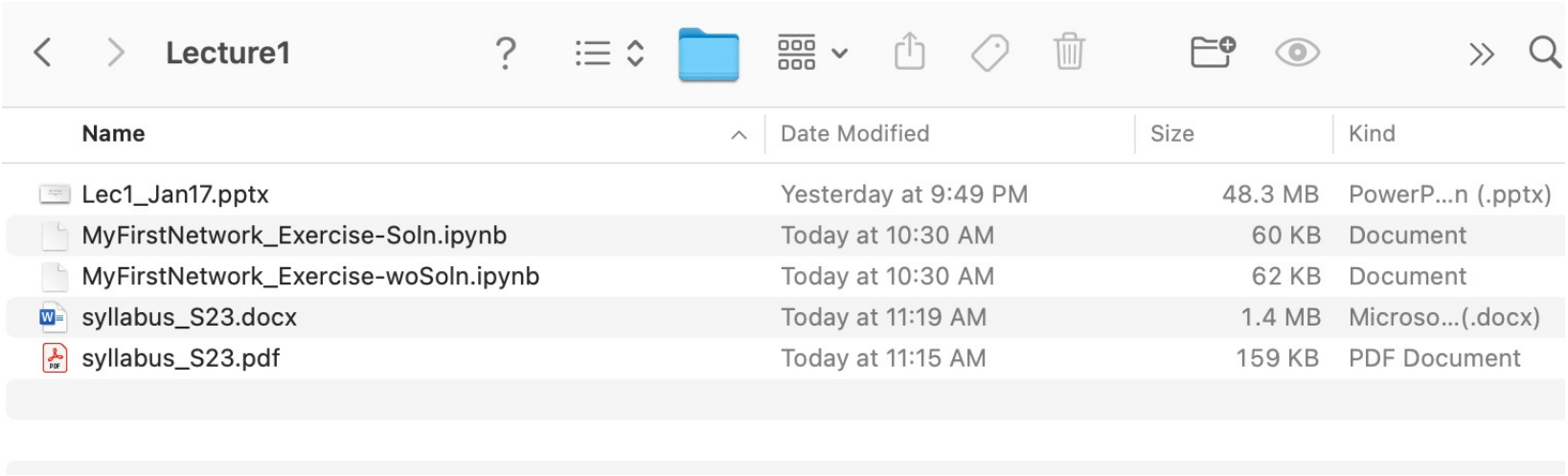


Navigate to the folder where you have the script



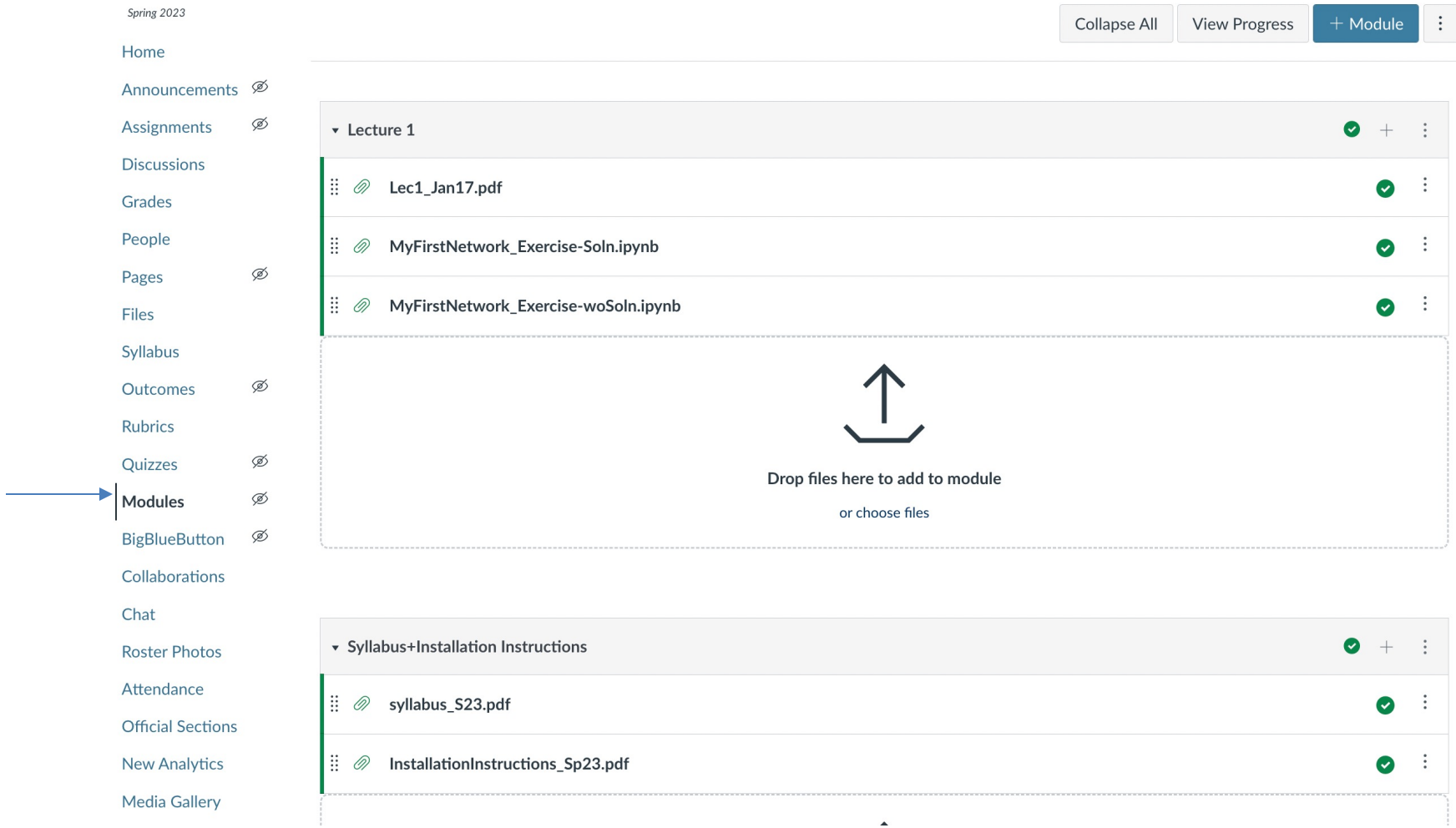
**You must keep  1 folder per Lecture**



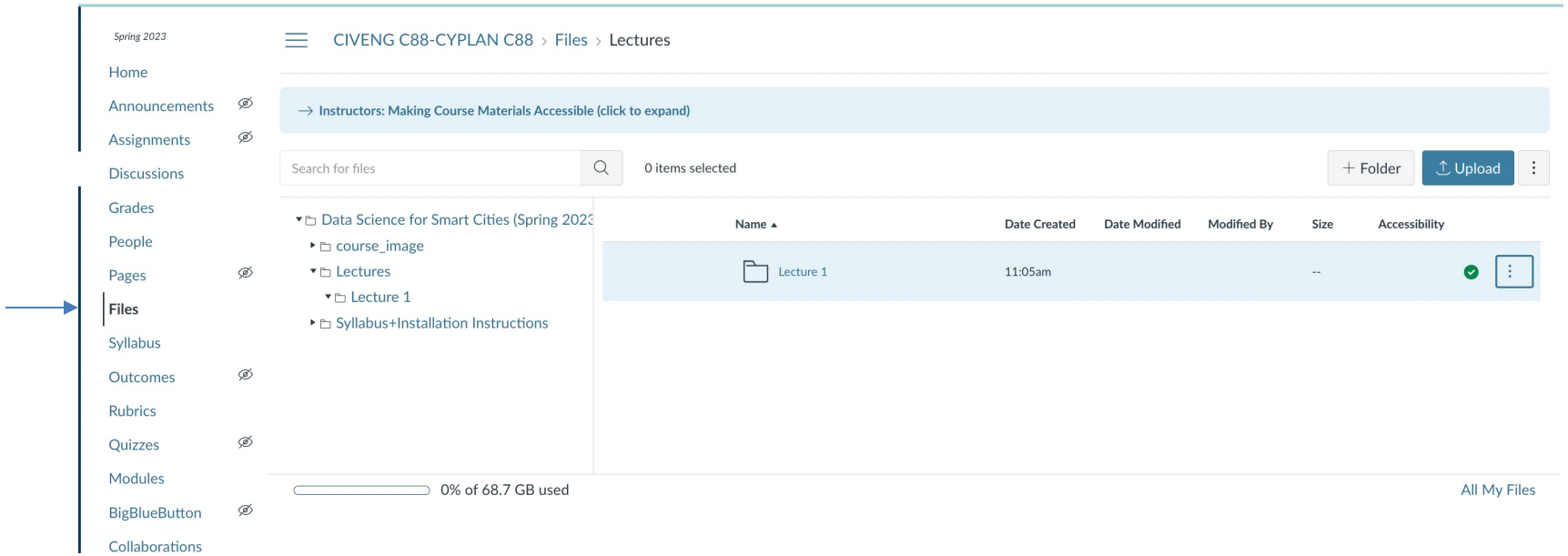



# Navigating b-courses to find Lecture notes

# See content in Modules:



# Download Lecture Folder as a Zip File



## People also ask

How do I make Chrome download to a specific folder?

**Change download locations**

1. On your computer, open Chrome.
2. At the top right, click More. Settings.
3. At the bottom, click Advanced. Downloads.
4. Adjust your download settings: To change the default download location, click Change and select where you'd like your files to be saved.

# For next Tuesday:

1) Decide if you will install jupyter notebook and work locally (recommended) or remotely via Berkeley DataHub

2) Finalize the solution of MyFirstNetwork_Exercise-woSoln.ipynb (or read the script MyFirstNetwork_Exercise-woSoln.ipynb)

3)Learn to create folders and download data in the target folder, open the scripts in jupyter notebok

4) See this short Python Tutorial: https://github.com/CambridgeUniversityPress/FirstCourseNetworkScience/blob/master/tutorials/Appendix%20-%20Python%20Tutorial.ipynb