Online Social Networks and Media

Introduction

Instructors:

Ευαγγελία Πιτουρά
http://www.cs.uoi.gr/~pitoura
Παναγιώτης Τσαπάρας
http://www.cs.uoi.gr/~tsap

Goal

Understand the importance of networks in life, technology and applications
Study the theory underlying social networks
Learn about algorithms that make use of network structure
Learn about the tools to analyze them

Today:

Some logistics
A taste of the topics to be covered

Introduction to graph theory and notebooks

Logistics

Textbooks:

Easley and Kleinberg free text-book on Networks, Crowds and Markets

M. E. J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003

Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, free text-book on Social Media Mining

Michele Coscia, The Atlas for the Aspiring Network Scientist

Web page:

www.cs.uoi.gr/~tsap/teaching/cs-d6

20% Presentations and class participation

30% Assignments

50% Term Project (in 2 Phases)

No Final Exam

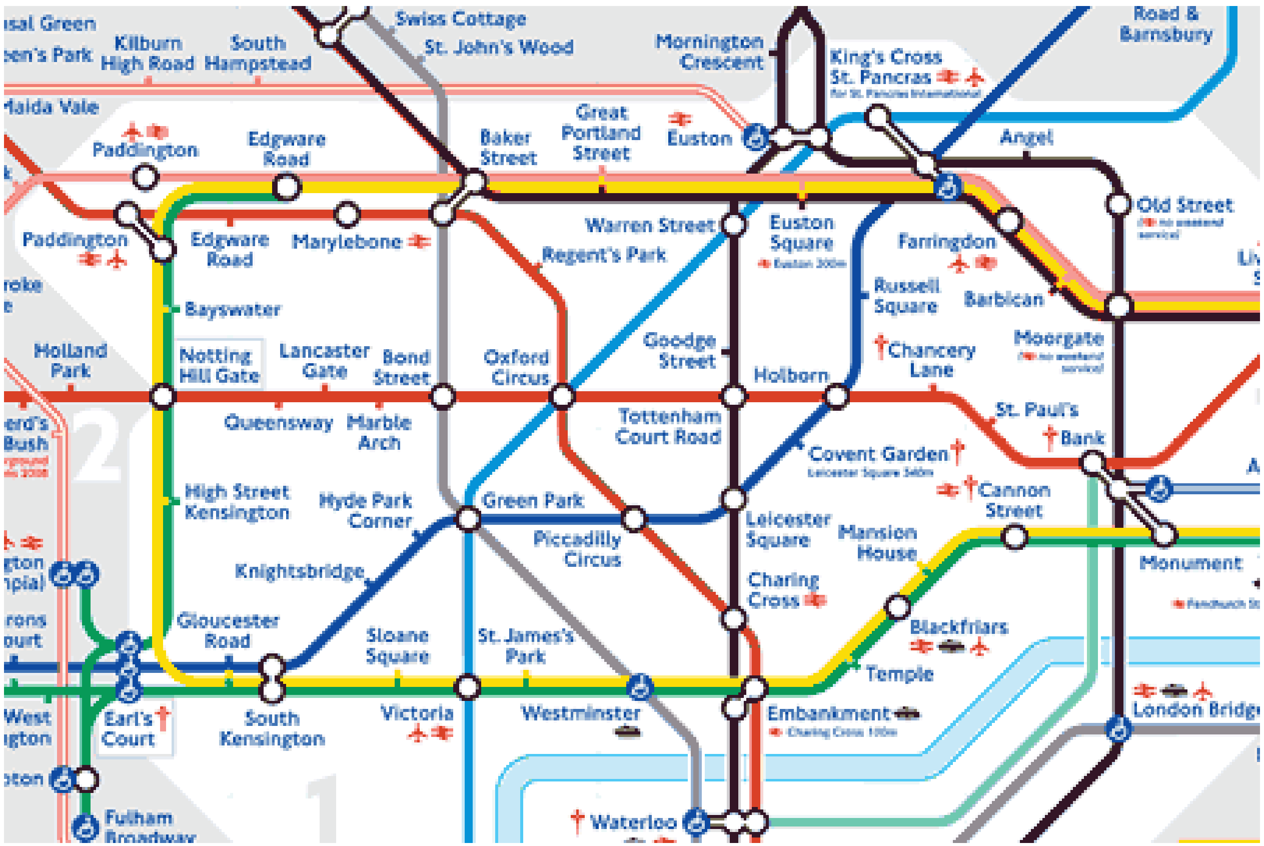
CONSIDER THE FOLLOWING COMPLEX SYSTEMS

The Economy


NASDAQ
KB Home
-2.2%
-1.16
50.96
+1.5%
+0.44
30.89
+2.0%
+0.35
17.84
+1.1%
+0.35
32.08
BUY ORDERS
19.12
+1.10
+6.1%
19.09
+1.0%
+0.17
17.15
+1.2%
+0.27
23.46
NDTV PROFIT
NASDAQ
NASDAQ
NASDAQ
NASDAQ
NASDAQ

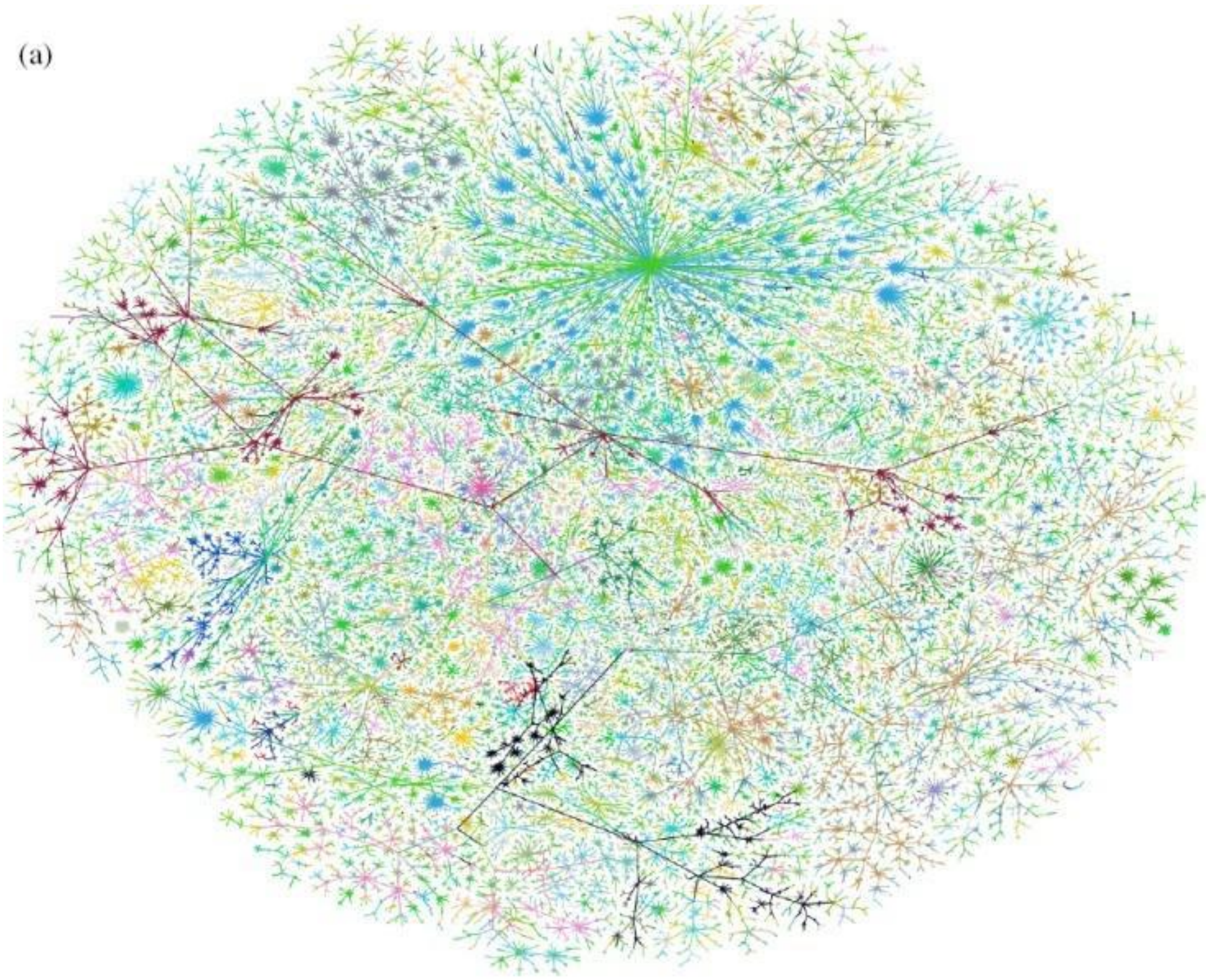
The Human Cell

Traffic and roads

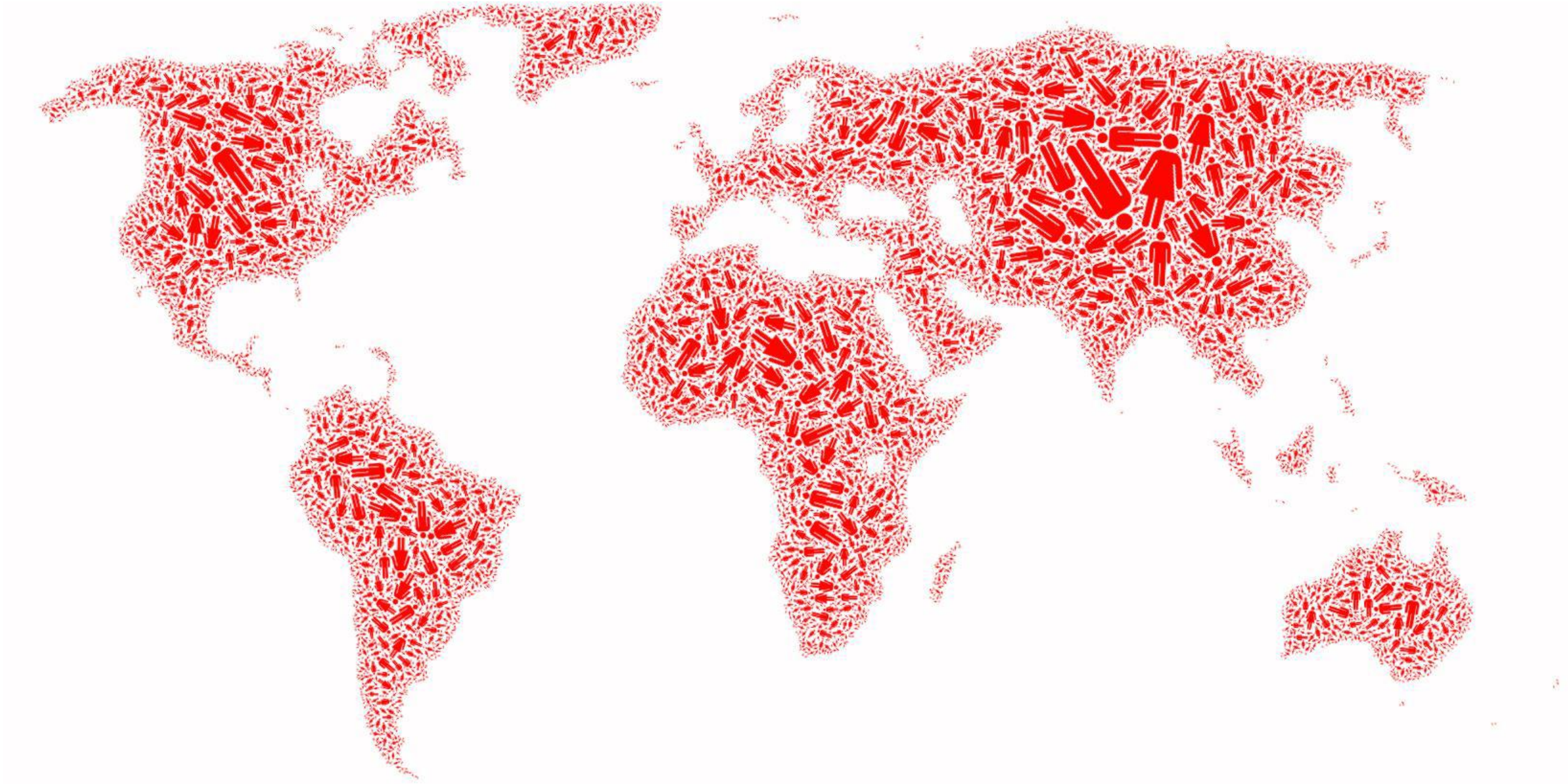


Internet

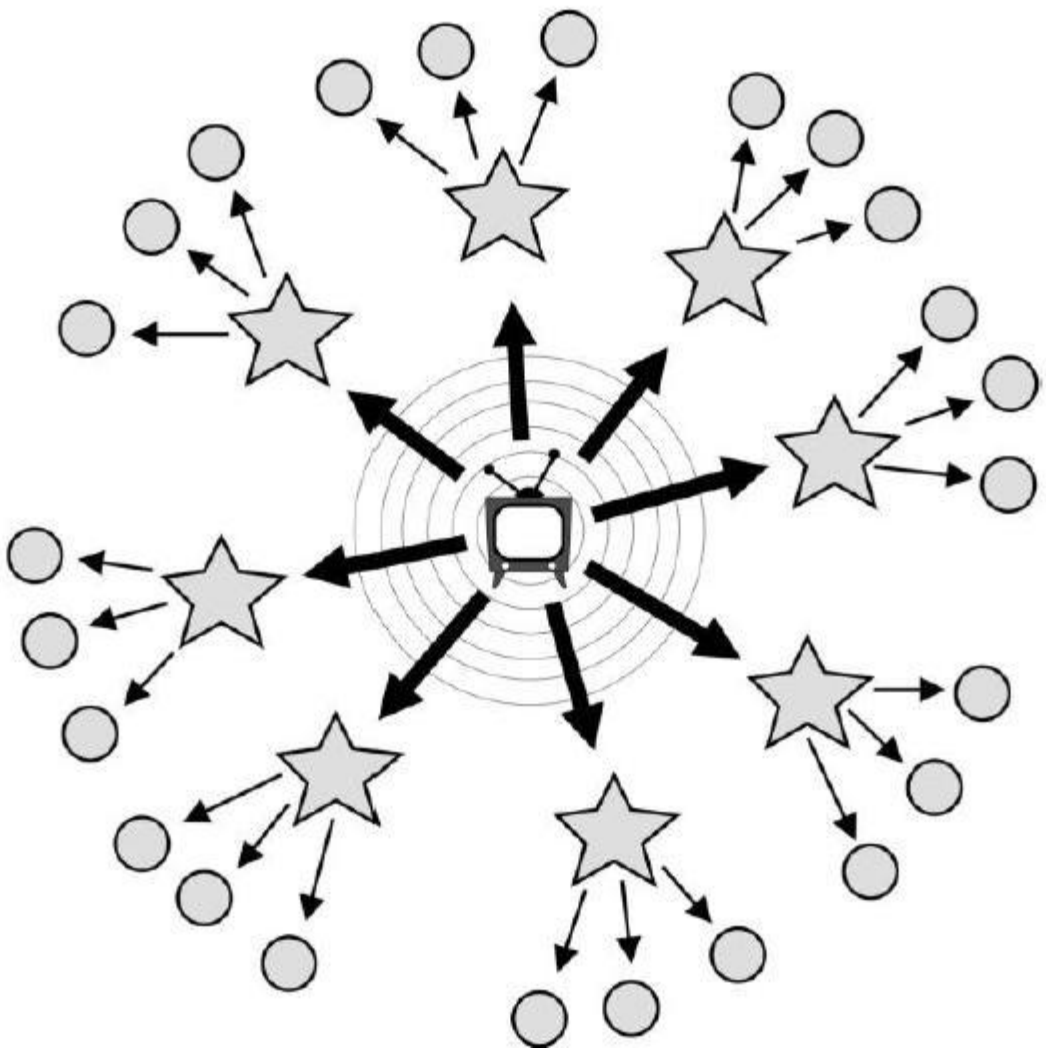
(a)



Society



Media and Information



WHAT DO THEY HAVE IN COMMON?

THE NETWORK!

All of these systems can be modeled as networks

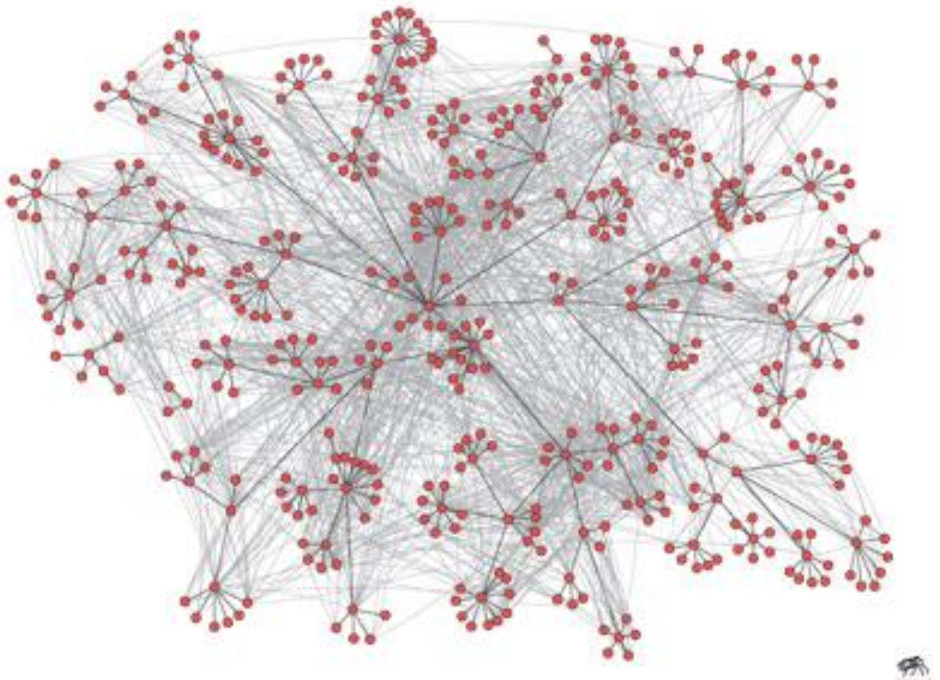
What is a network?

- Network: a collection of entities that are interconnected with links.

Social networks

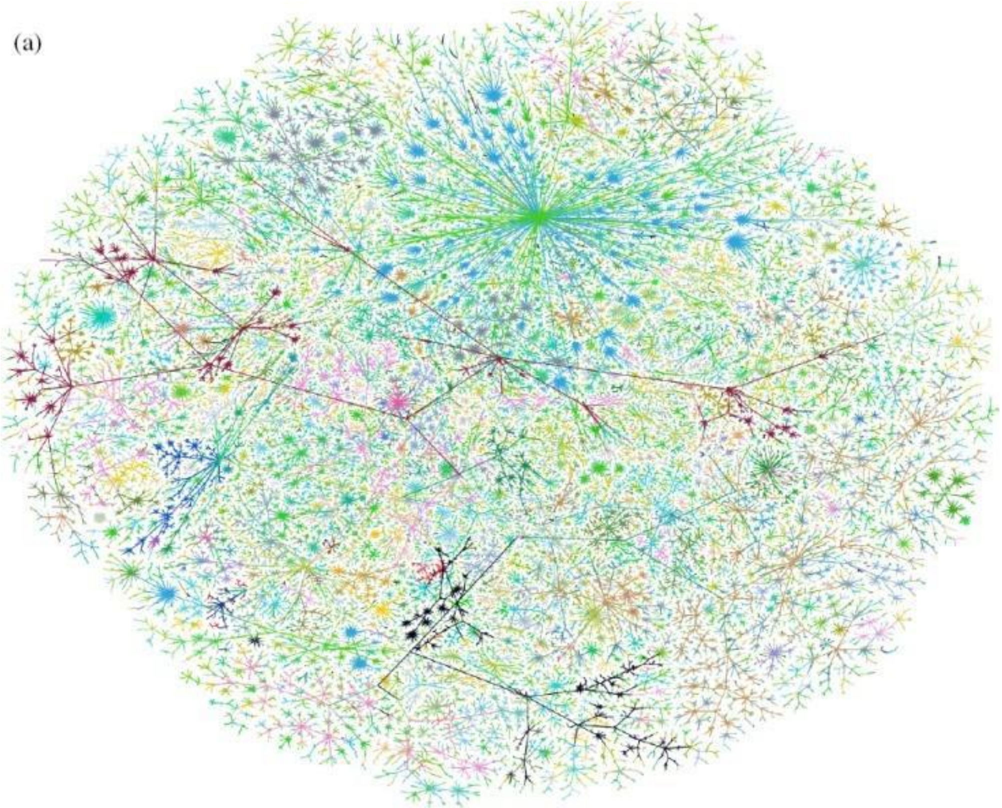
- Entities: People
- Links: Friendships

Communication networks



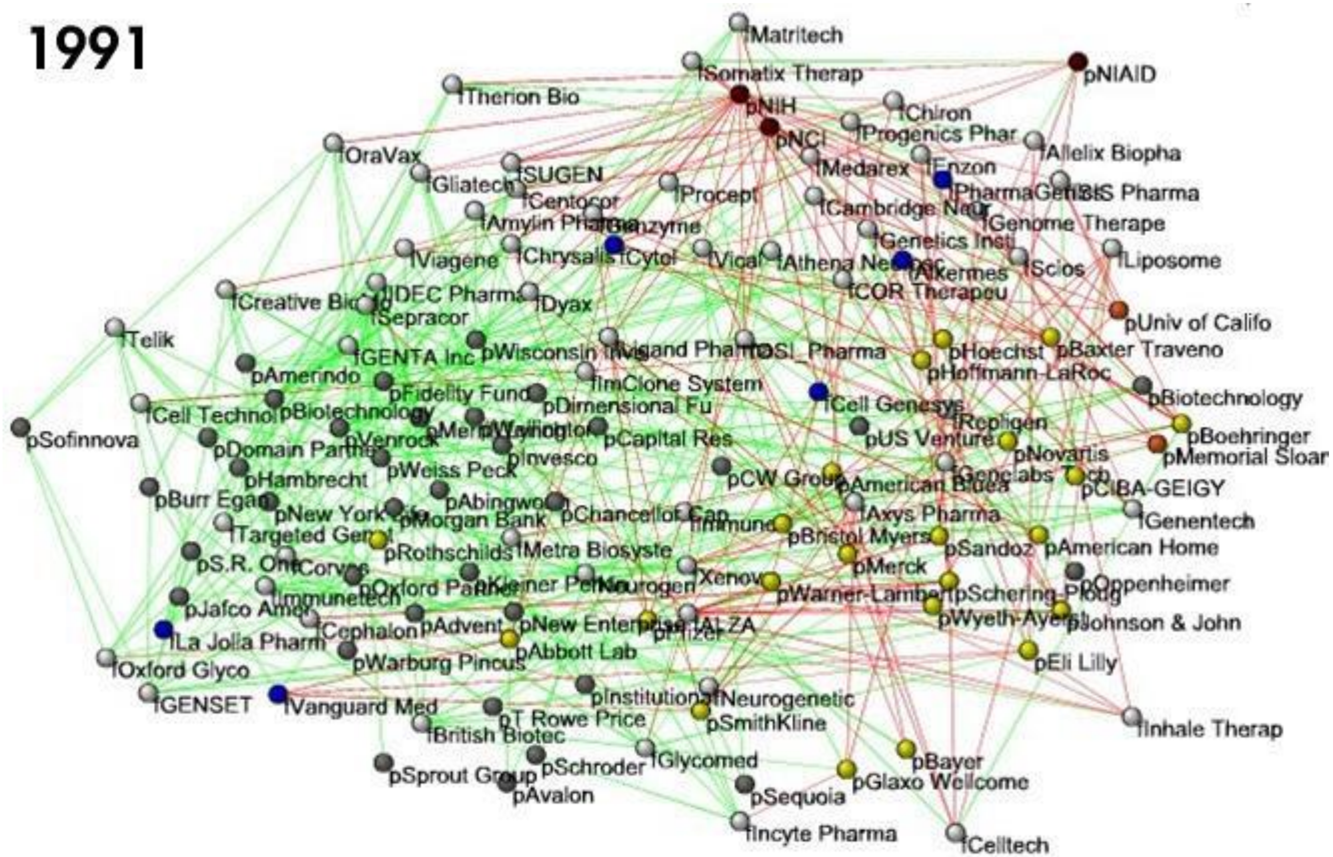
- Entities: People
- Links: email exchange

Communication networks



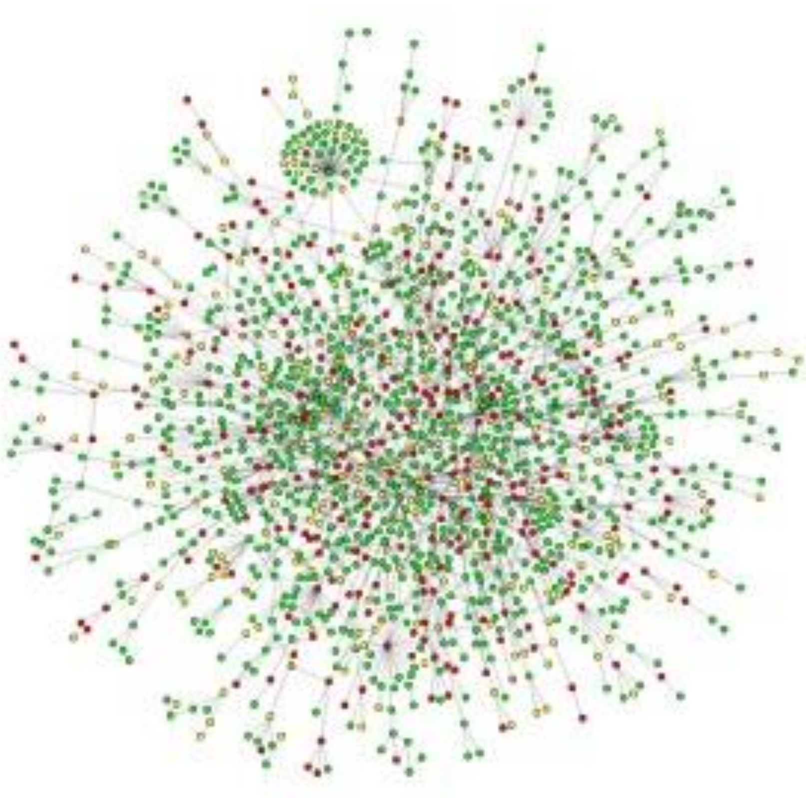
- Entities: Internet nodes
- Links: communication between nodes

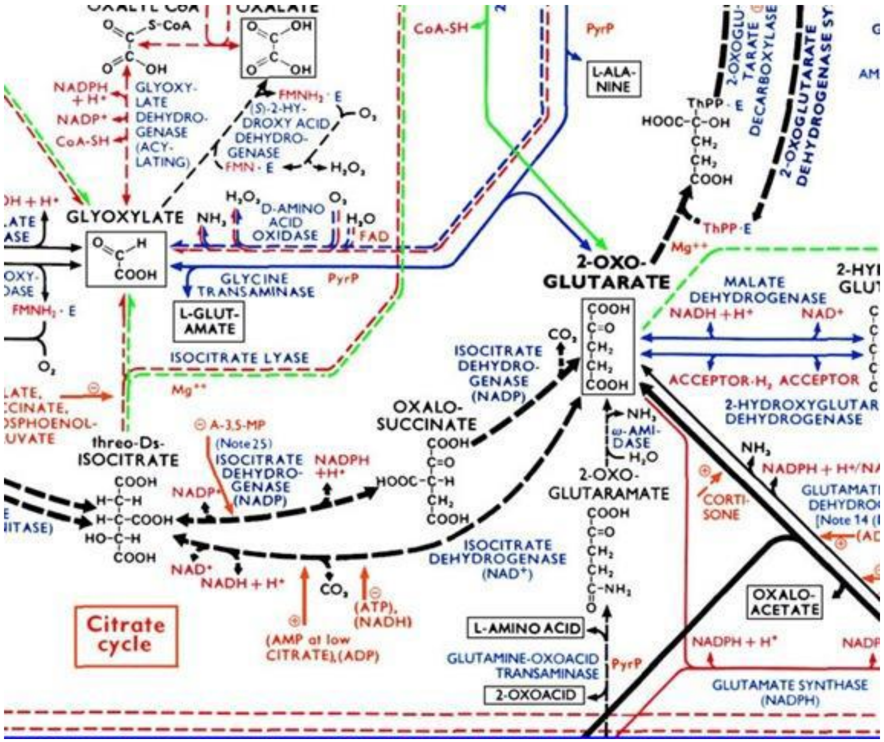
Financial Networks



- Entities: Companies
- Links: relationships (financial, collaboration)

Biological networks

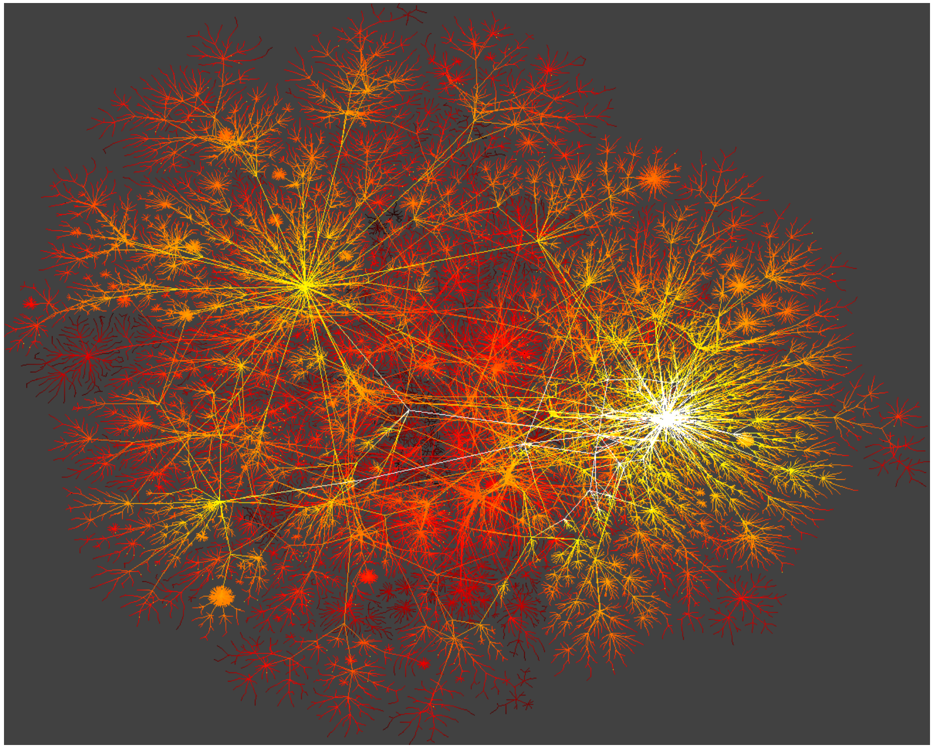




- Entities: Proteins
- Links: interactions

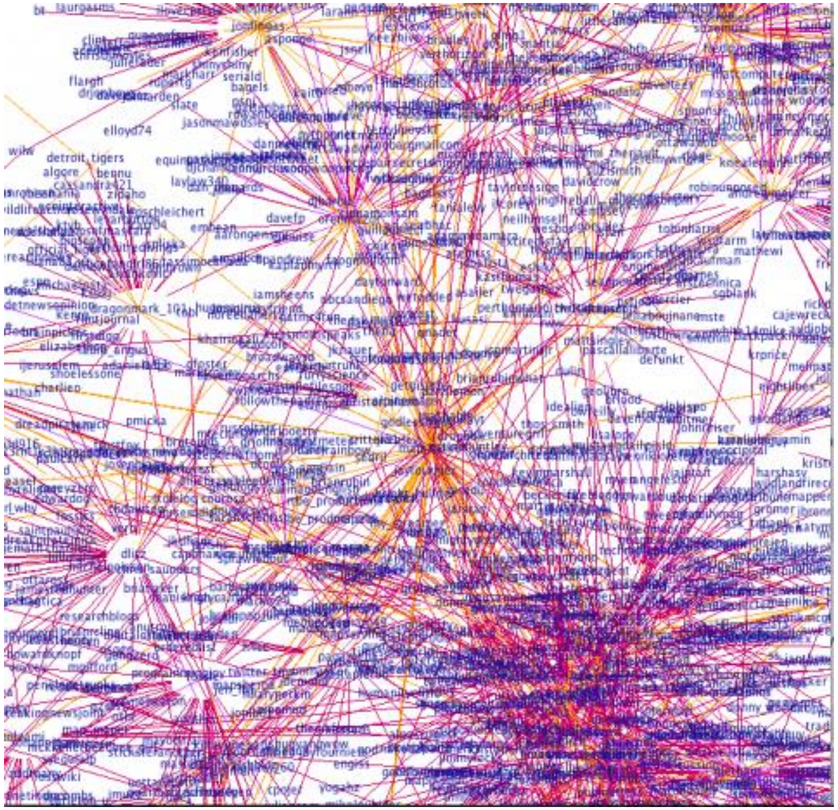
- Entities: metabolites, enzymes
- Links: chemical reactions

Information networks



- Entities: Web Pages
- Links: Links

Information/Media networks



- Entities: Twitter users
- Links: Follows/conversations

Many more

- Wikipedia
- Brain
- Highways
- Software
- Etc…

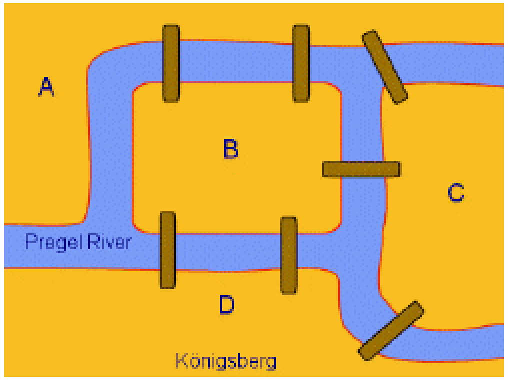
Why networks are important?

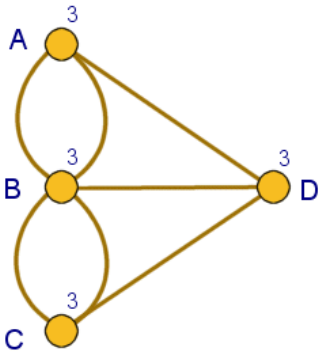
- We cannot truly understand a complex system unless we understand the underlying network.
  - Everything is connected, studying individual entities gives only a partial view of a system

- Two main themes:
  - What are the structural properties of the network?
  - How do processes happen in the network?
  - How can we use networks to extract knowledge

Graphs

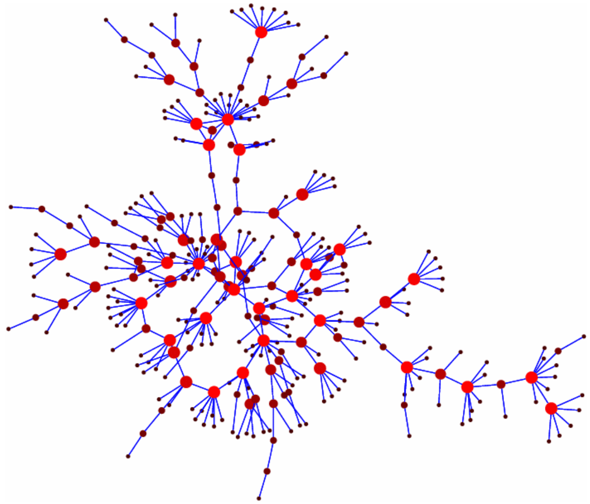
- In mathematics, networks are called graphs, the entities are nodes, and the links are edges
- Graph theory starts in the 18th century, with Leonhard Euler
  - The problem of Königsberg bridges
  - Since then graphs have been studied extensively.





Networks in the past

- Graphs have been used in the past to model existing networks (e.g., networks of highways, social networks)
  - usually these networks were small
  - network can be studied visual inspection can reveal a lot of information



Networks now

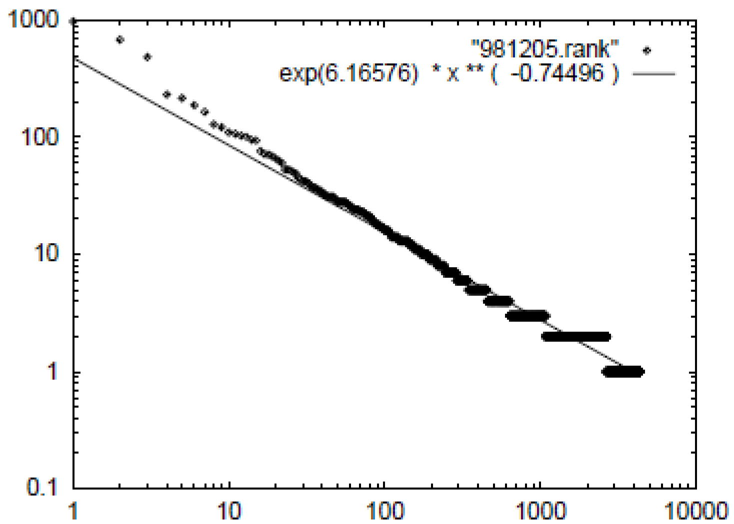
- More and larger networks appear
  - Products of technology
    - e.g., Internet, Web, Facebook, Twitter
  - Result of our ability to collect more, better, and more complex data
    - e.g., gene regulatory networks
  - Result of the willingness of users to contribute data
    - e.g., users making their relationships public online
- Networks of thousands, millions, or billions of nodes
  - Impossible to process visually
  - Problems become harder
  - Processes are more complex
- Networks as a source of data rather than combinatorial objects

Topics

- Measuring Real Networks
- Modeling the evolution and creation of networks
- Identifying important nodes in the network
- Understanding information cascades and virus contagions
- Finding communities in graphs
- Link Prediction
- Storing and processing huge networks
- Network embeddings
- Machine learning in networks
- Network Fairness
- Other special topics

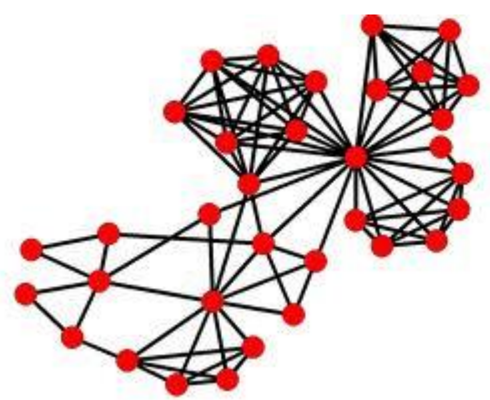
Understanding large graphs

- What does a network look like?
  - Measure different properties to understand the structure



(a) Int-12-98

degree of nodes



Triangles in the graph

Real network properties

- Most nodes have only a small number of neighbors (degree), but there are some nodes with very high degree (power-law degree distribution)
  - scale-free networks
- If a node x is connected to y and z, then y and z are likely to be connected
  - high clustering coefficient
- Most nodes are just a few edges away on average.
  - small world networks
- Networks from very diverse areas (from internet to biological networks) have similar properties
  - Is it possible that there is a unifying underlying generative process?

Generating random graphs

- Classic graph theory model (Erdös-Renyi)
  - each edge is generated independently with probability p

- Very well studied model but not realistic:
  - most vertices have about the same degree
  - the probability of two nodes being linked is independent of whether they share a neighbor
  - the average paths are short

Modeling real networks

- Real life networks are not "random"
- Can we define a model that generates graphs with statistical properties similar to those in real life?

- The rich-get-richer model

We need to accurately model the mechanisms that govern the evolution of networks (for prediction, simulations, understanding)

Ranking of nodes on the Web

- Is my home page as important as the facebook page?
- We need algorithms to compute the importance of nodes in a graph
- The PageRank Algorithm
  - A success story of network use

It is impossible to create a web search engine without understanding the web graph

Information/Virus Cascade

- How do viruses spread between individuals? How can we stop them?

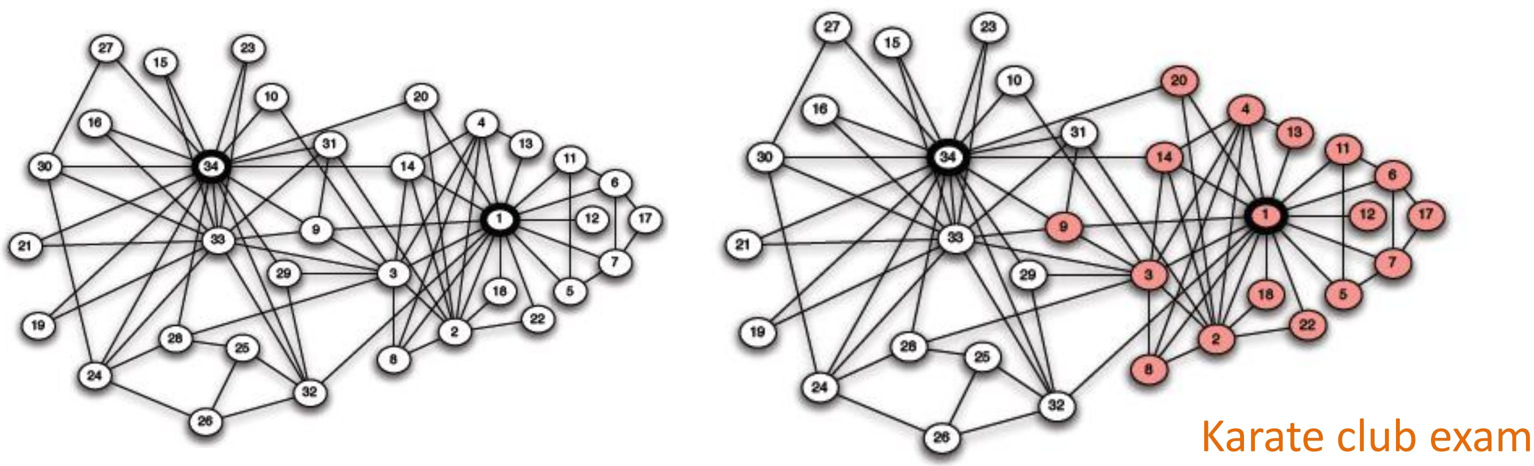
- How does information propagate in social and information networks? What items become viral? Who are the influencers and trend-setters?

- We need models and algorithms to answer these questions

Online advertising relies heavily on online social networks and word-of-mouth marketing.
COVID-19 demonstrated the importance of understanding virus spread.

Clustering and Finding Communities

- What is community?
  - "Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties." [Wasserman & Faust '97]



Karate club example
[W. Zachary, 1970]

How do we define and discover communities in large graphs?

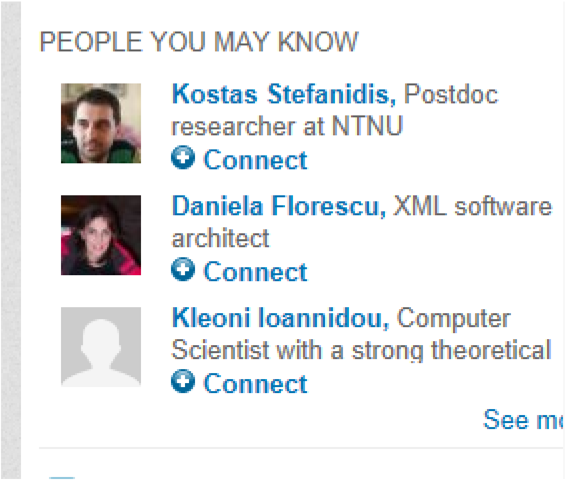
Homophily

- **Homophily:**"Birds of a feather flock together". People with similar interests tend to connect to each other
  - Caused by two related social forces [Friedkin98, Lazarsfeld54]
    - *Social influence:* People become similar to those they interact with
    - *Selection:* People seek out similar people to interact with
- Homophily is widely observed in social networks and has important implications on the network
- Applications in online marketing
  - *viral marketing* relies upon social influence affecting behavior
  - *recommender systems* predict behavior based on similarity

How does homophily affect the properties and the processes in the network?

Link Prediction

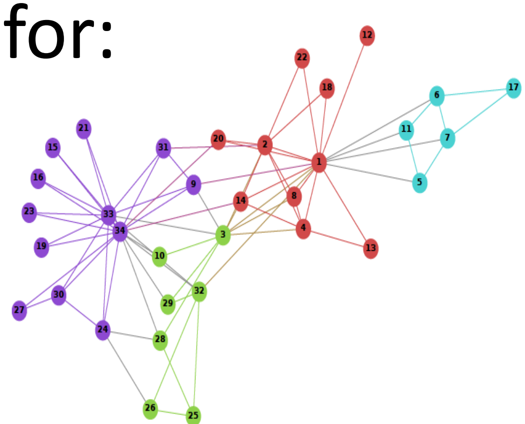
- Given a snapshot of a social network at time *t*, we seek to accurately predict the edges that will be added to the network during the interval from time *t* to a given future time *t'*.
- Applications:
  - Accelerate the growth of a social network (e.g., Facebook, LinkedIn, Twitter) that would otherwise take longer to form.
  - Identify suspect relationships



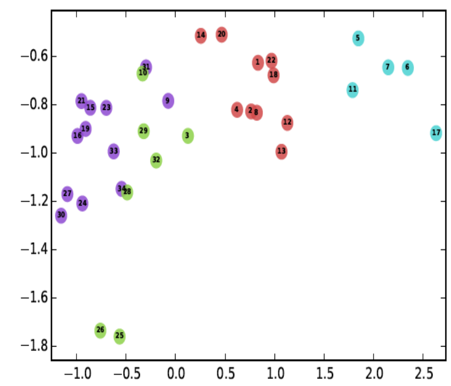
How do we predict future links?
How do we make friendship recommendations?

Network Embeddings

- Project the nodes of the network into a multidimensional real space, such that the nodes that are close to each other in the network (or, semantically similar) are mapped to near-by points in this space
- An application of machine learning on graphs.
- Multiple applications for:
  - Link prediction
  - Node classification
  - More



Zachary's Karate Club Network:



2-dimensional node embedding

Image from: Perozzi et al.. DeepWalk: Online Learning of Social Representations. KDD 2014.

Machine Learning on Graphs

- The network information can be used for Machine Learning tasks such as classification:
  - "You will be known by the company you keep": Use the network information to classify the nodes of the graph.
  - Simple label propagation techniques have been used very successfully in the past
- Graph Machine Learning (GML): Application of sophisticated ML techniques on graphs. The AI revolution has also reached networks
  - GNN: Graph Neural Networks. Incorporate the network structure into the Neural Network computation.

Network content

- Users on online social networks generate content.
- Mining the content in conjunction with the network can be useful
  - Do friends post similar content on Facebook?
  - Can we understand a user's interests by looking at those of their friends?
    - The importance of homophily
  - Social recommendations: Can we predict a movie rating using the social network?

Social Media

- Today Social Media (Twitter, Facebook, Instagram) have supplanted the traditional media sources
  - Information is generated and disseminated mostly online by users
  - Twitter has become a global "sensor" detecting and reporting everything
  - Social media have become the main platform for disseminating ideas
- Interesting problems:
  - Automatically detect events using Twitter
  - Sentiment mining
  - Misinformation and fake news detection

The dark side of Social Networks

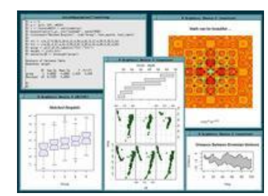
- Do algorithms make fair and correct decisions?
- Do algorithms create filter bubbles and echo chambers, increase polarization, and promote misinformation?
- Are online social networks and media a threat to democracy and our mental health?

Tools

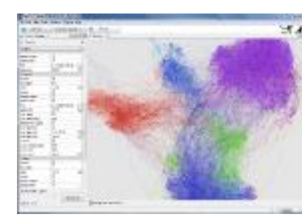
NetworkX: a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. http://networkx.lanl.gov/

Stanford Network Analysis Platform (SNAP): general purpose, high performance system for analysis and manipulation of large networks written in C++ http://snap.stanford.edu/snap/index.html

R: free software environment for statistical computing and graphics. http://www.r-project.org/



Gephi: interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs http://gephi.org/



Frameworks for Processing Large Graphs

Large scale (in some cases billions of vertices, trillions of edges)

How to process graphs in parallel?

- Write *your own code*
- Use *MapReduce (general parallel processing) **
- *Pregel* (bulk synchronous parallel model) introduced by Google in 2010*
- *Giraph* http://incubator.apache.org/giraph/ (part of Hadoop software)

Storage?

*J. Dean, S. Ghemawat. *MapReduce: Simplified Data Processing on Large Clusters.* OSDI 2004: 137-150
** G. Malewicz, M. H. Austern, A. J. C. Bik, J. C. Dehnert, I. Horn, N. Leiser: *Pregel: a system for large-scale graph processing*. SIGMOD Conference 2010: 135-146

Data

Collected using available APIs (Twitter, Facebook, etc)

Using existing collections, e.g., from SNAP (more in the webpage), permission may be required

Stanford Large Network Dataset Collection
60 large social and information network datasets
Coauthorship and Citation Networks
DBLP: Collaboration network of computer scientists
KDD Cup Dataset
Internet Topology
AS Graphs: AS-level connectivities inferred from Oregon route-views, Looking glass data and Routing registry data
Yelp Data
Yelp Review Data: reviews of the 250 closest businesses for 30 universities for students and academics to explore and research
Youtube dataset
Youtube data: YouTube videos as nodes. Edge a->b means video b is in the related video list (first 20 only) of a video a.
Amazon product copurchasing networks and metadata
Amazon Data: The data was collected by crawling Amazon website and contains product metadata and review information about 548,552 different products (Books, music CDs, DVDs and VHS video tapes).
Wikipedia
Wikipedia page to page link data: A list of all page-to-page links in Wikipedia
DBPedia: The DBpedia data set uses a large multi-domain ontology which has been derived from Wikipedia.
Edits and talks: Complete edit history (all revisions, all pages) of Wikipedia since its inception till January 2008.
Movie Ratings
IMDB database: Movie ratings from IMDB
User rating data: Movie ratings from MovieLens

Acknowledgements

- Thanks to Jure Leskovec for some of the material from his course notes.

- M. E. J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003

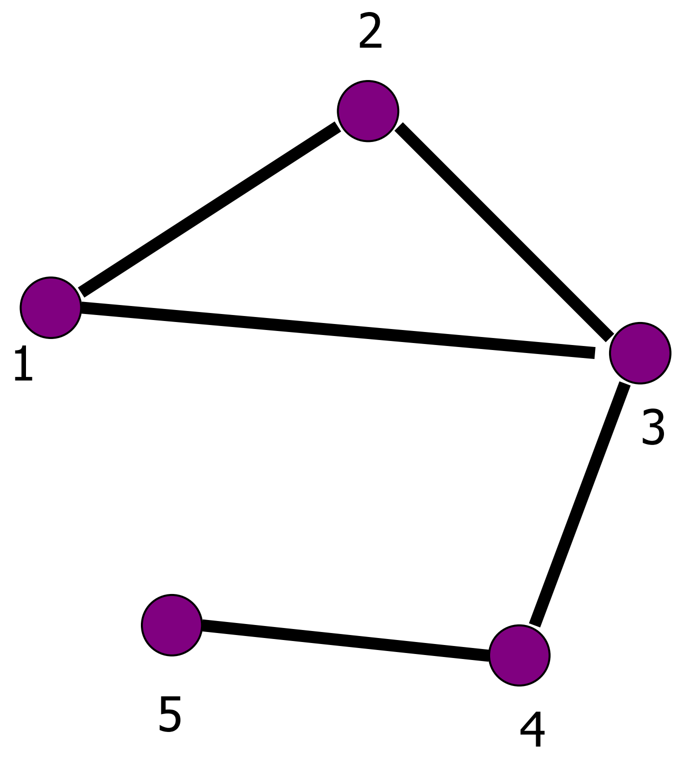
Graph Theory Reminder

For a more extensive reminder

(pptx, pdf)

Undirected Graph

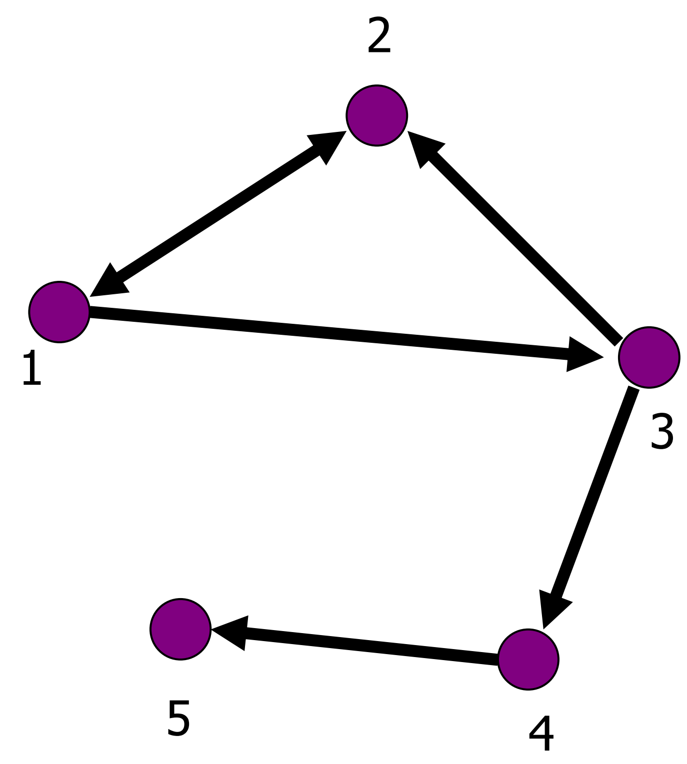
- Graph G=(V,E)
  - V = set of vertices (nodes)
  - E = set of edges



undirected graph
V = {1, 2, 3, 4, 5}
E={(1,2),(1,3),(2,3),(3,4),(4,5)}

Directed Graph

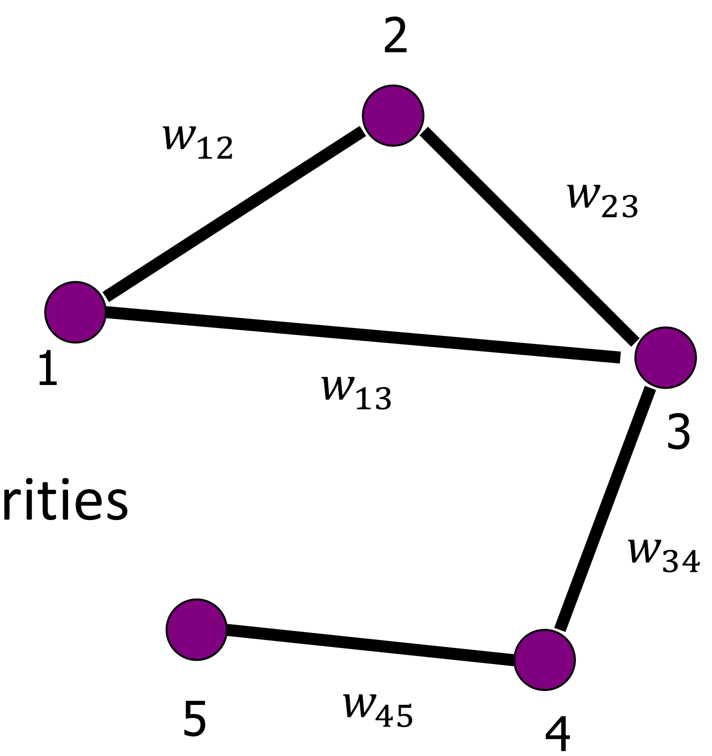
- Graph G=(V,E)
  - V = set of vertices (nodes)
  - E = set of edges



directed graph
V = {1, 2, 3, 4, 5}
E={‹1,2›, ‹2,1› ‹1,3›, ‹3,2›, ‹3,4›, ‹4,5›}

Weighted Graph

- Graph G=(V,E)
  - V = set of vertices (nodes)
  - E = set of edges and their weights



Weights can be either distances or similarities

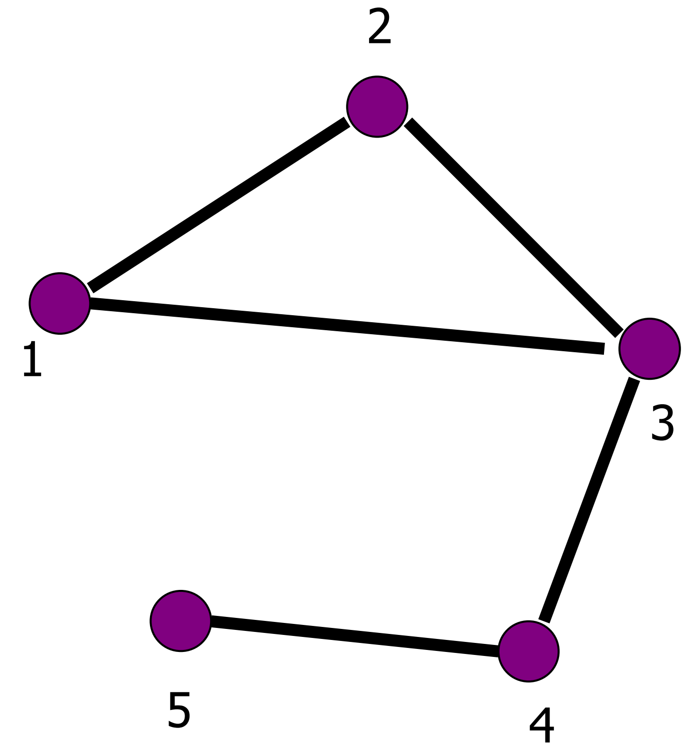
Weighted graph

V = {1, 2, 3, 4, 5}

E={(1,2, w_{12}),(1,3, w_{12}),(2,3, w_{12}),(3,4, w_{12}),(4,5, w_{12})}

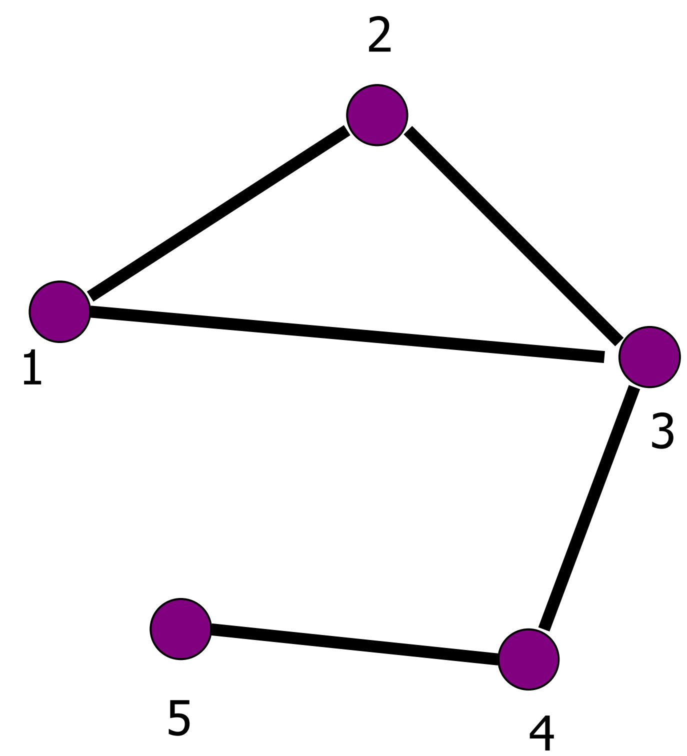
Undirected graph

- Neighborhood N(i) of node i
  - Set of nodes adjacent to i
- degree d(i) of node i
  - Size of N(i)
  - number of edges incident on i

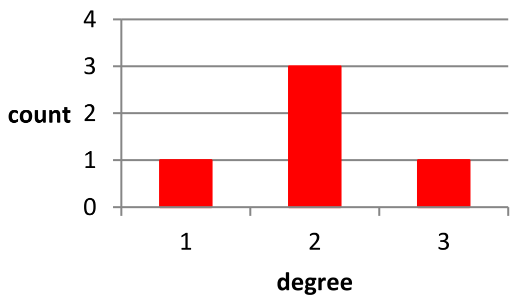


Undirected graph

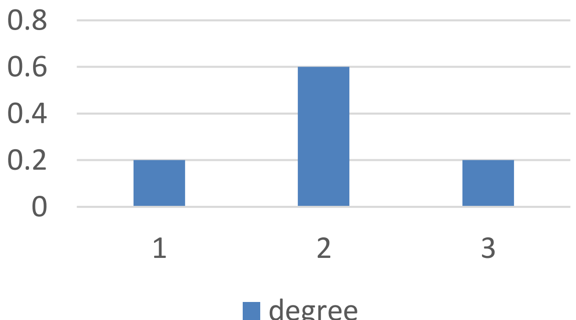
- degree sequence
  - [d(1),d(2),d(3),d(4),d(5)]
  - [2,2,3,2,1]



- degree histogram
  - [(1:1),(2:3),(3,1)]

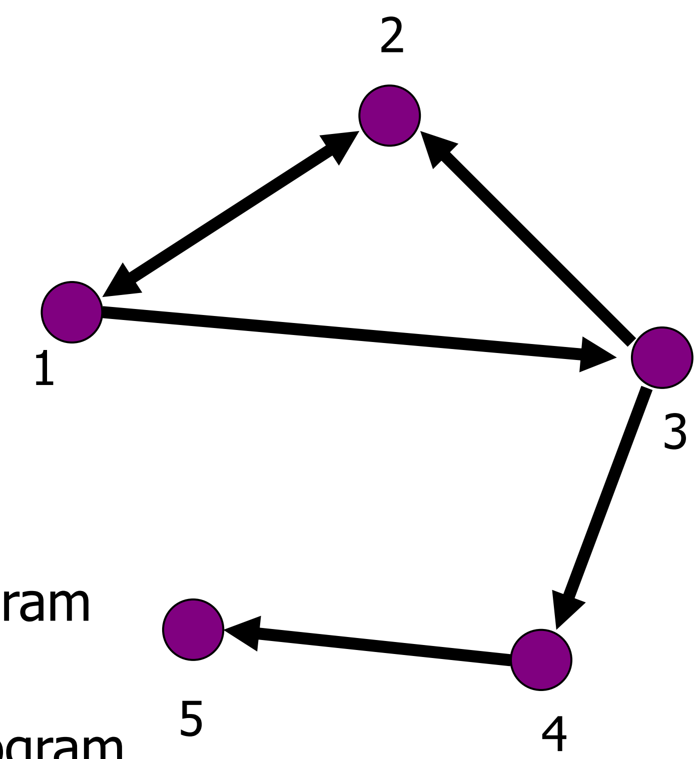


- degree distribution
  - [(1:0.2),(2:0.6),(3,0.2)]



Directed Graph

- in-degree $d_{in}(i)$ of node i
  - number of edges incoming to node i
- out-degree $d_{out}(i)$ of node i
  - number of edges leaving node i



- in-degree sequence
  - [1,2,1,1,1]
- out-degree sequence
  - [2,1,2,1,0]
- in-degree histogram
  - [(1:4),(2:1)]
- out-degree histogram
  - [(0:1),(1:2),(2:2)]

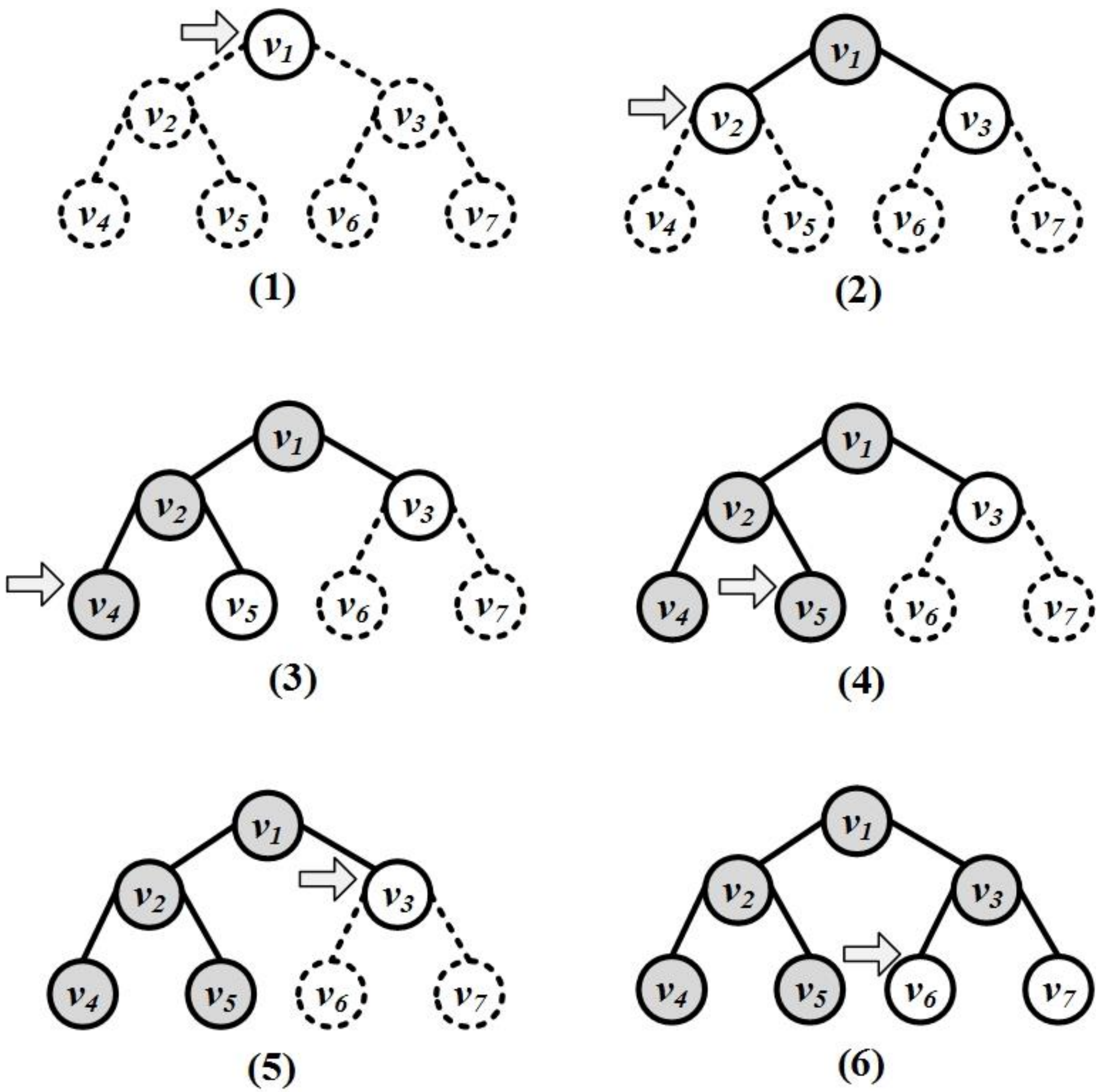
Graph Traversals

- A traversal is a procedure for visiting (going through) all the nodes in a graph

Depth First Search Traversal

- Depth-First Search (DFS) starts from a node i, selects one of its neighbors j from N(i) and performs Depth-First Search on j before visiting other neighbors in N(i).
  - The algorithm can be implemented using a *stack structure*

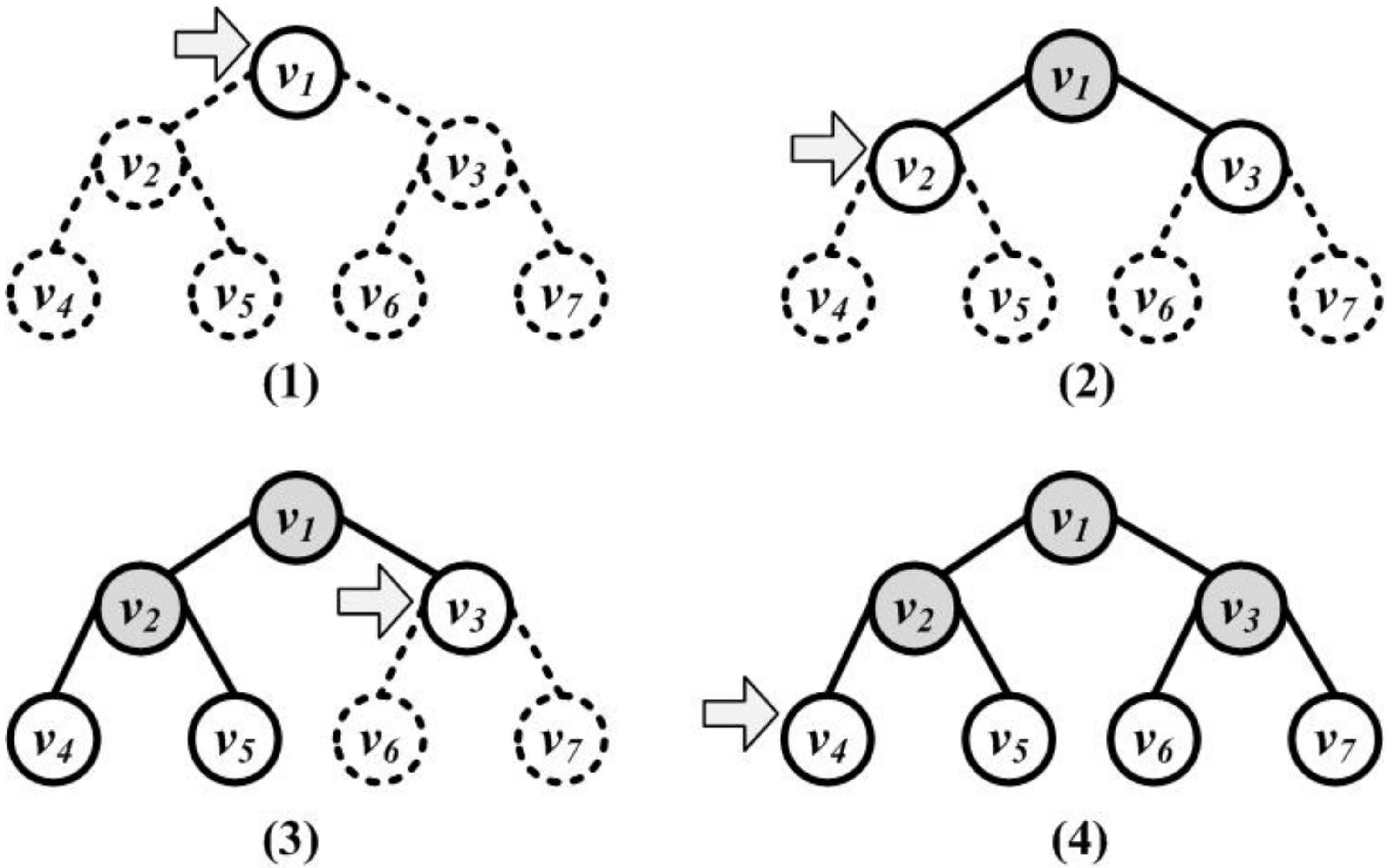
Example for a tree graph



Breadth First Search Traversal

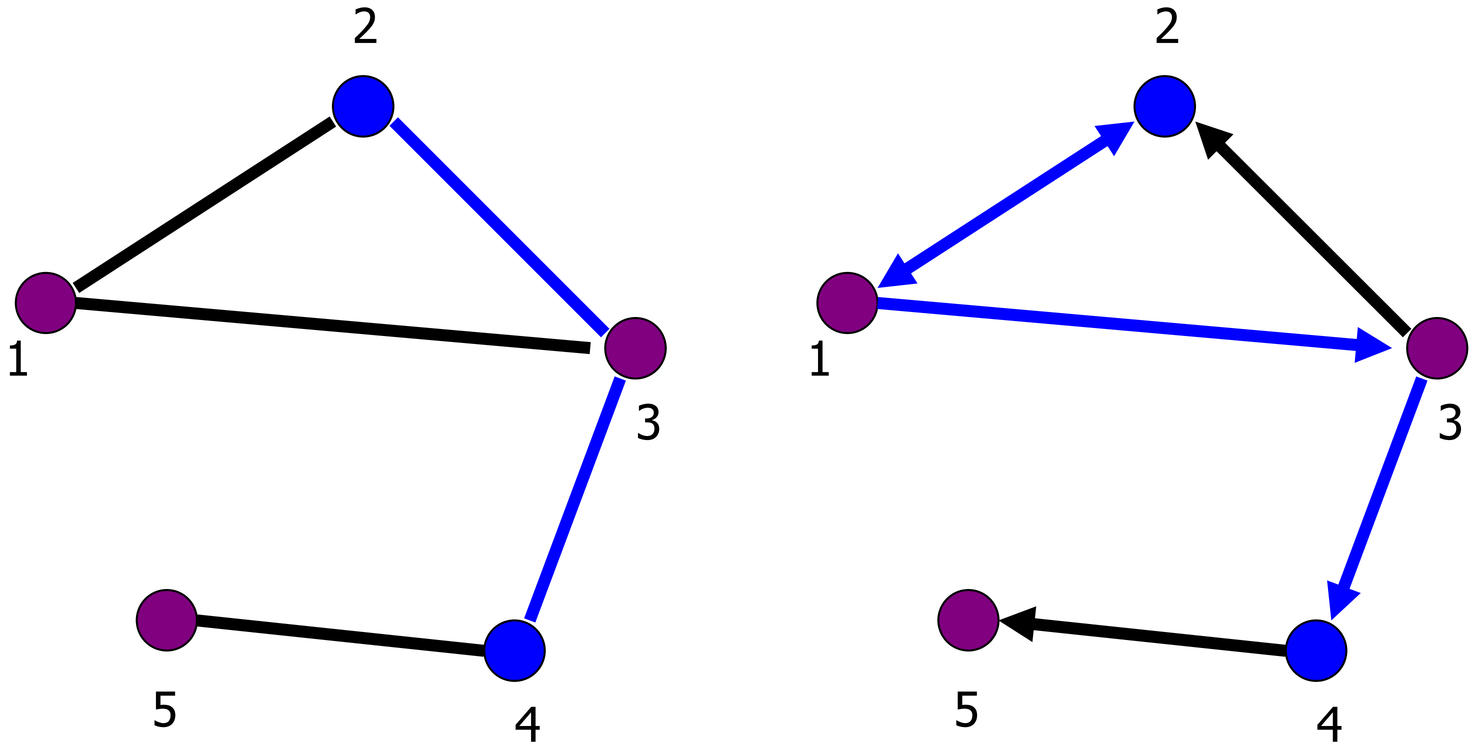
- Breadth-First-Search (BFS) starts from a node, visits all its immediate neighbors first, and then moves to the second level by traversing their neighbors.
  - The algorithm can be implemented using a *queue structure*

Example of BFS on a tree



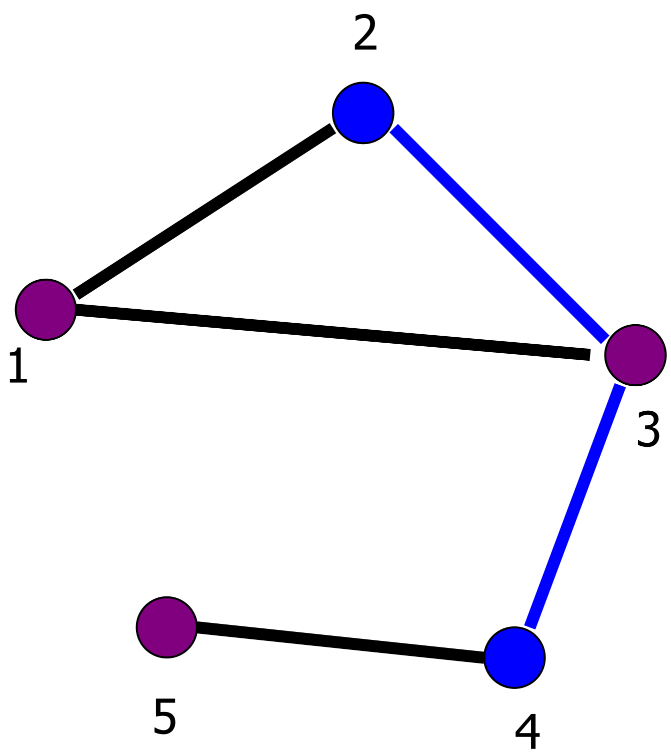
Paths

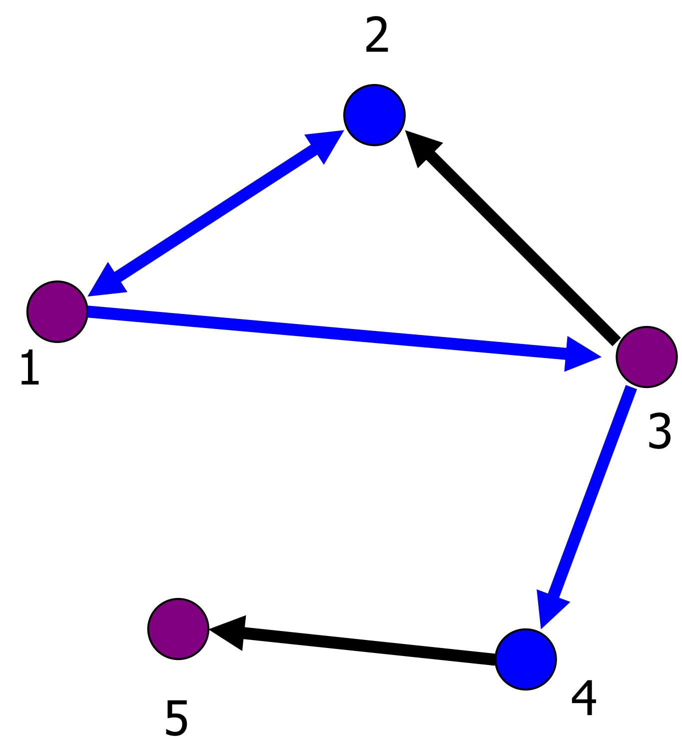
- Path from node i to node j: a sequence of edges (directed or undirected) from node i to node j
  - path length: number of edges on the path
  - nodes i and j are connected
  - cycle: a path that starts and ends at the same node



Shortest Paths

- Shortest Path from node i to node j
  - also known as BFS path, or geodesic path
  - We can find all shortest paths from a node using BFS



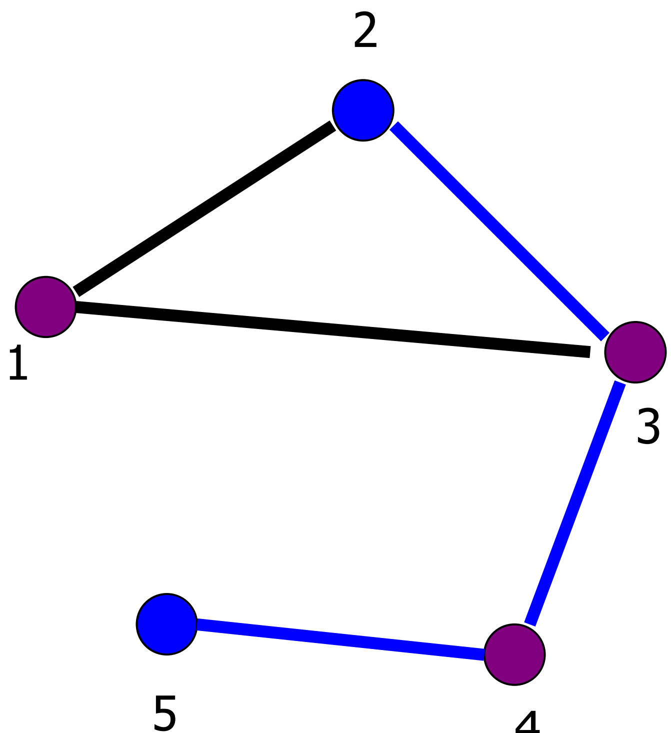


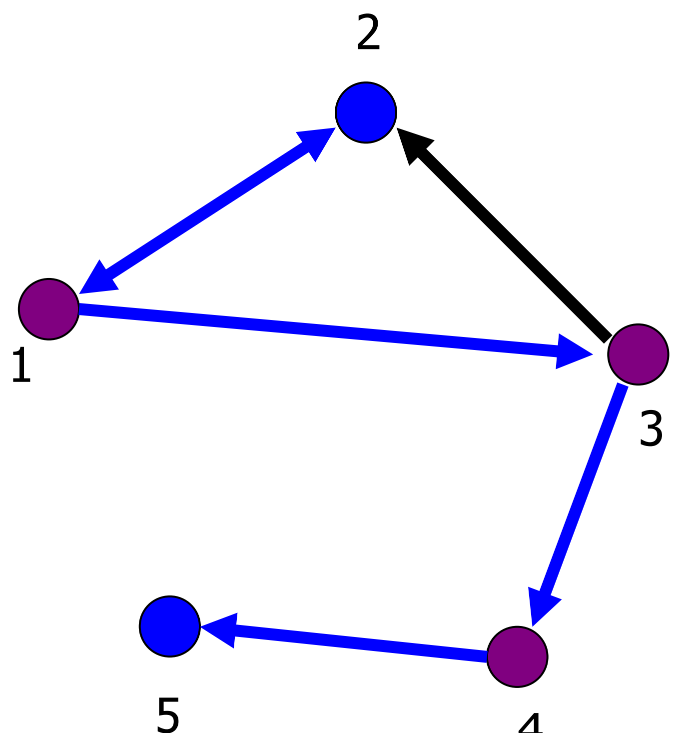
Shortest paths on weighted graphs

- Shortest paths on weighted graphs are harder to construct
  - There are several well known algorithms for finding single-source, or all-pairs shortest paths
  - For example: Dijkstra's Algorithm

Diameter

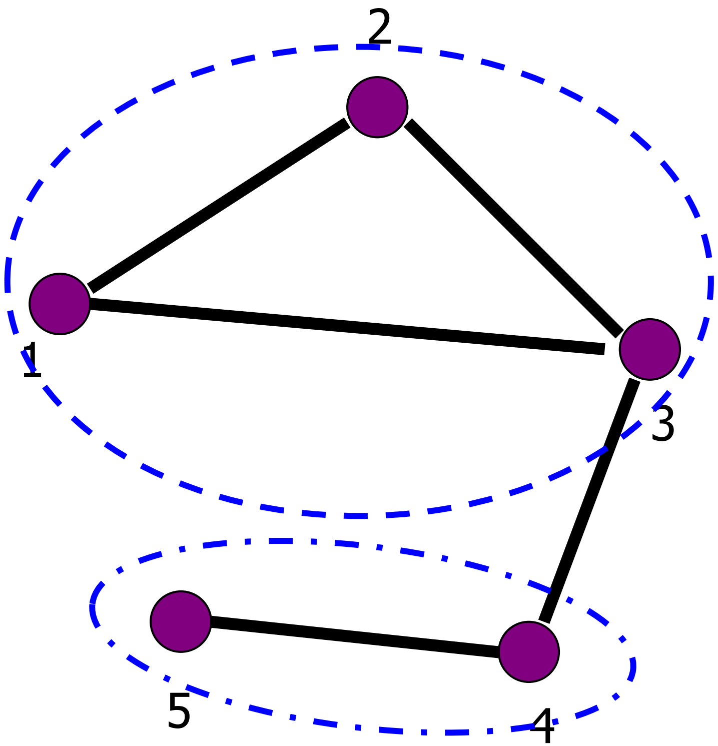
- The longest shortest path in the graph





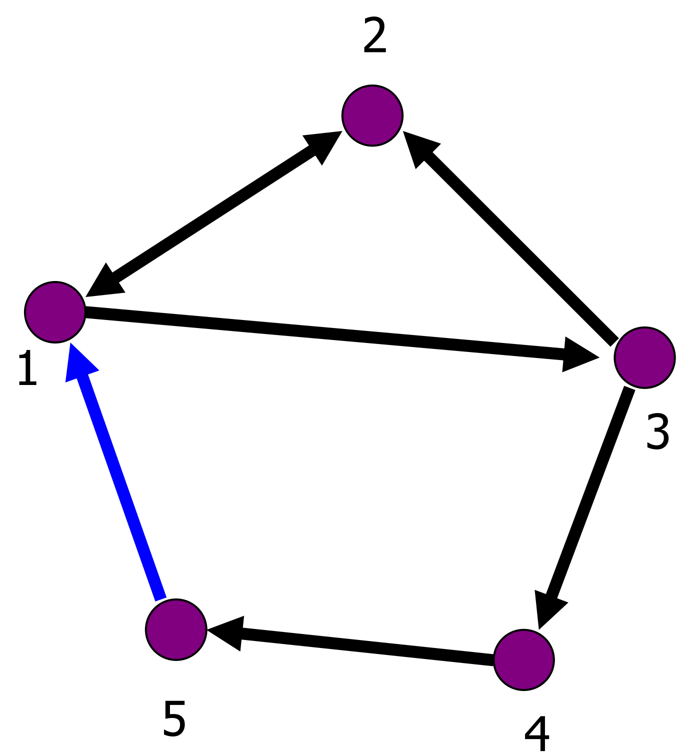
Undirected graph

- Connected graph: a graph where there every pair of nodes is connected
- Disconnected graph: a graph that is not connected
- Connected Components: subsets of vertices that are connected



Directed Graph

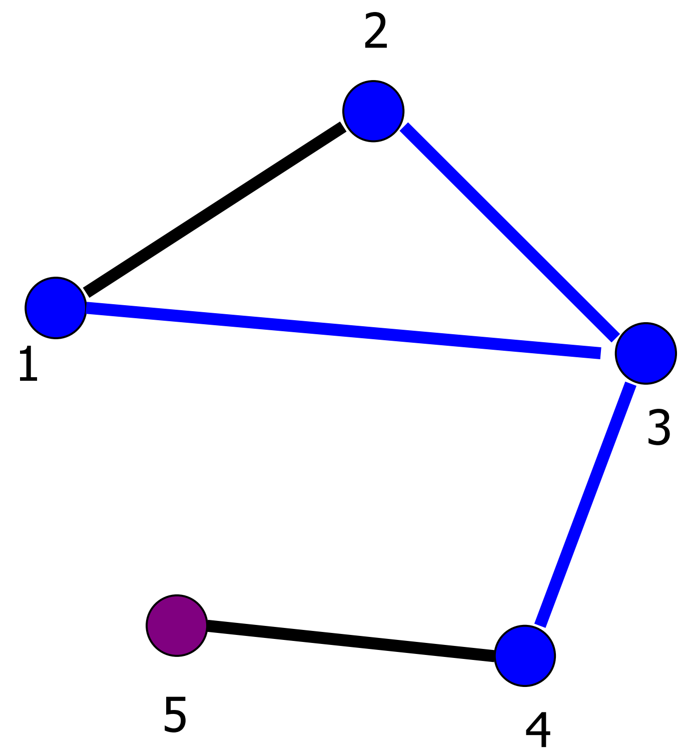
- Strongly connected graph: there exists a path from every i to every j
- Weakly connected graph: If edges are made to be undirected the graph is connected



Subgraphs

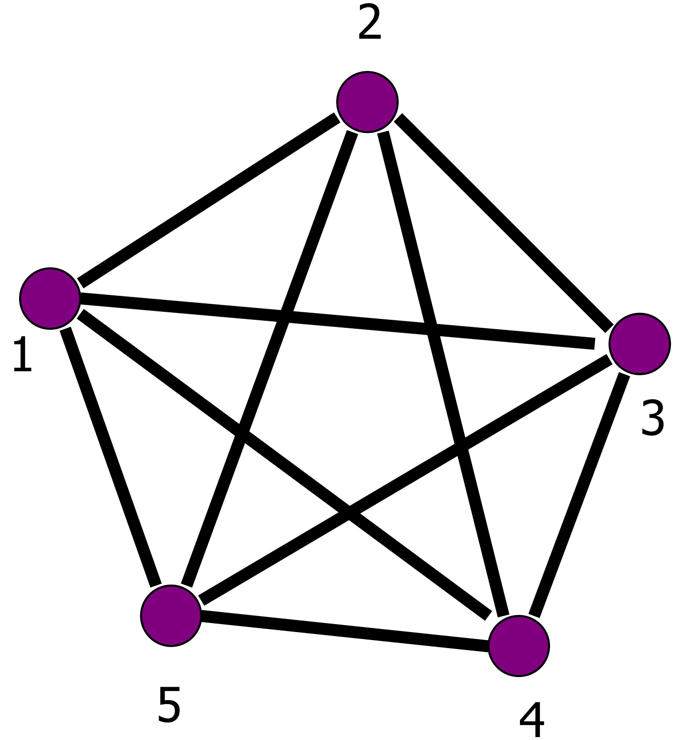
- Subgraph: Given V′ $\subseteq$ V, and E′ $\subseteq$ E, the graph G′=(V′,E′) is a subgraph of G.

- Induced subgraph: Given V′ $\subseteq$ V, let E′ $\subseteq$ E is the set of all edges between the nodes in V′. The graph G′=(V′,E′), is an induced subgraph of G



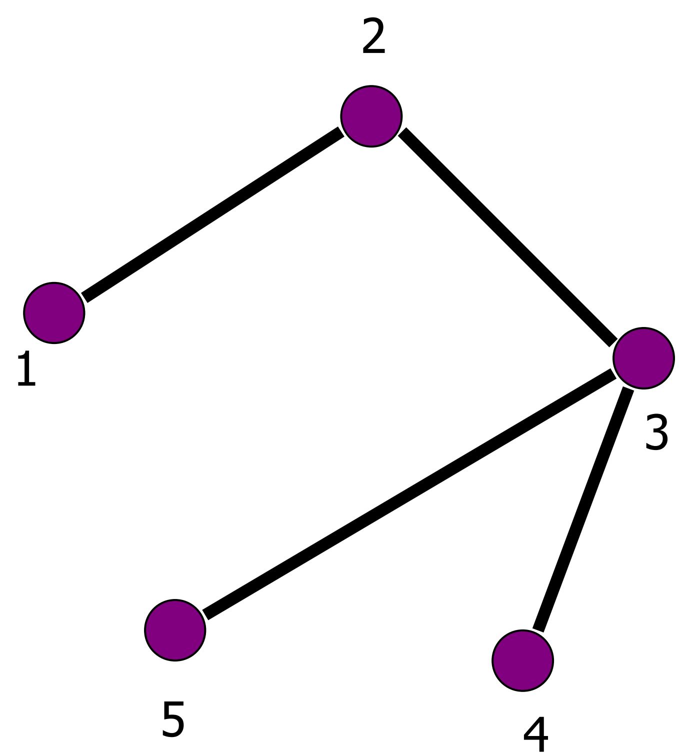
Fully Connected Graph

- Clique K_n
- A graph that has all possible $n(n-1)/2$ edges



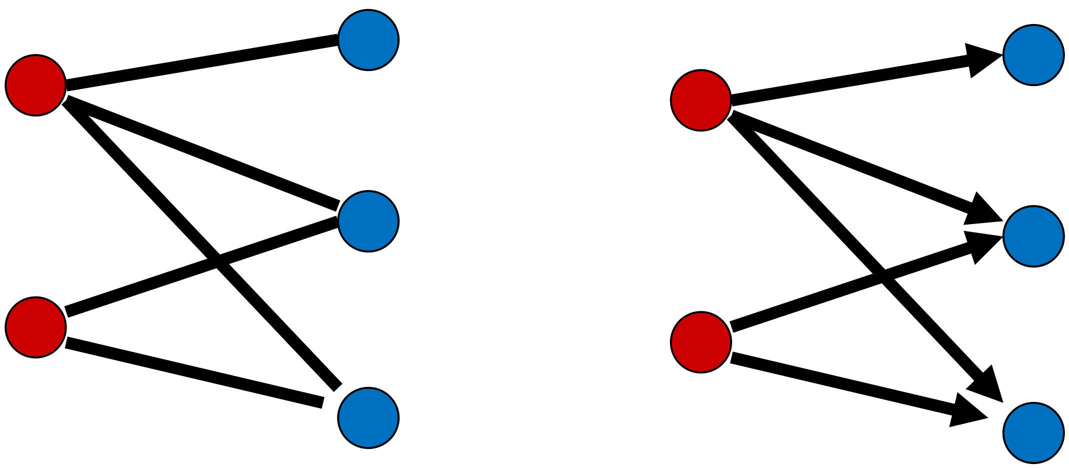
Trees

- Connected Undirected graphs without cycles



Bipartite graphs

- Graphs where the set of nodes V can be partitioned into two sets L and R, such that there are edges only between nodes in L and R, and there is no edge within L or R

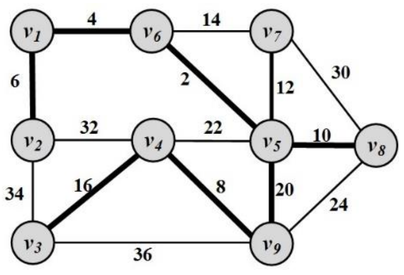


Spanning Tree

- For any connected graph, the spanning tree is a subgraph and a tree that includes all the nodes of the graph
- There may exist multiple spanning trees for a graph.
- For a weighted graph and one of its spanning tree, the weight of that spanning tree is the summation of the edge weights in the tree.
- Among the many spanning trees found for a weighted graph, the one with the minimum weight

  is called the

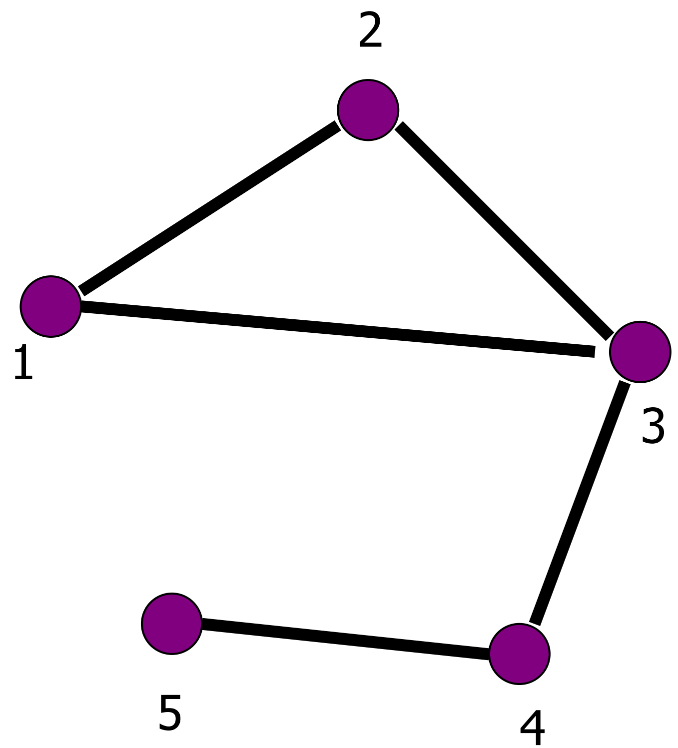
minimum spanning tree (MST)



Graph Representation

- Adjacency Matrix
  - symmetric matrix for undirected graphs

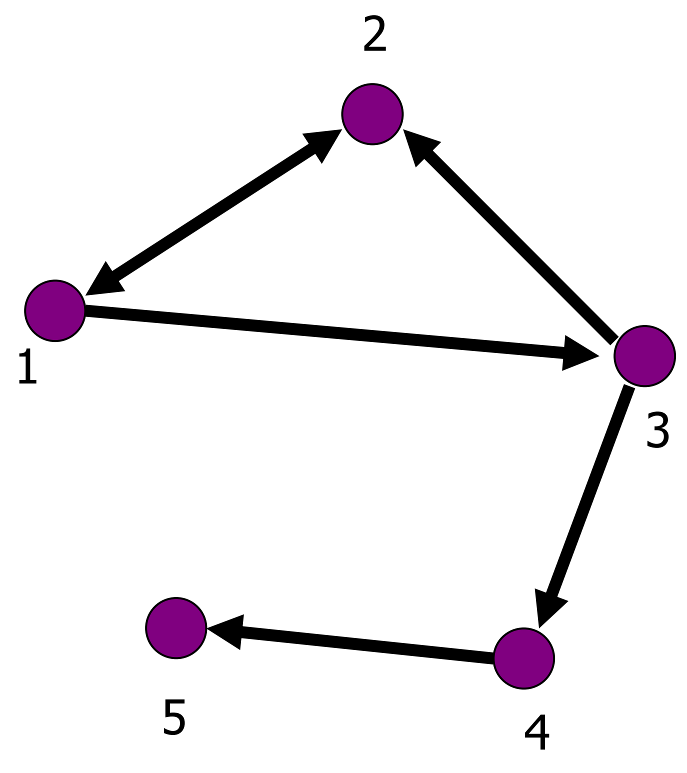
$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



Graph Representation

- Adjacency Matrix
  - unsymmetric matrix for undirected graphs

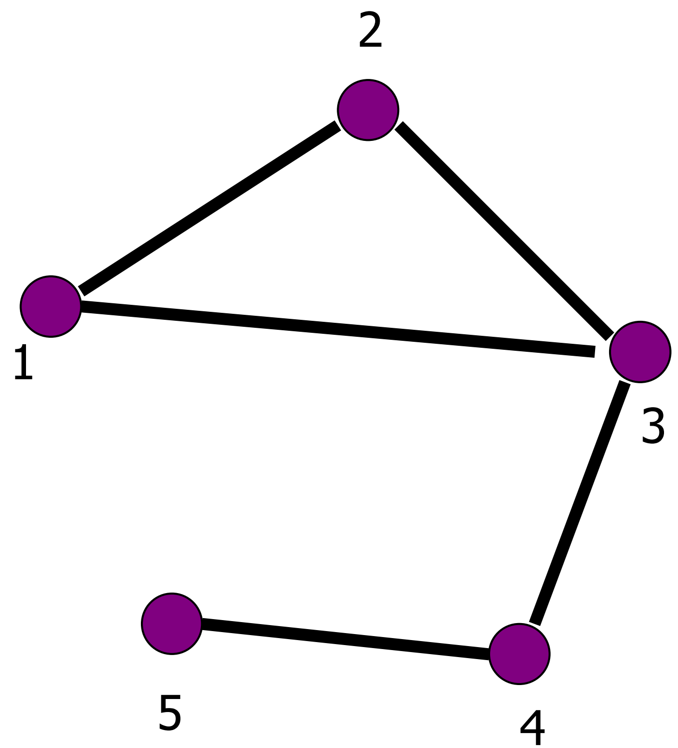
$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Graph Representation

- Adjacency List
  - For each node keep a list with neighboring nodes

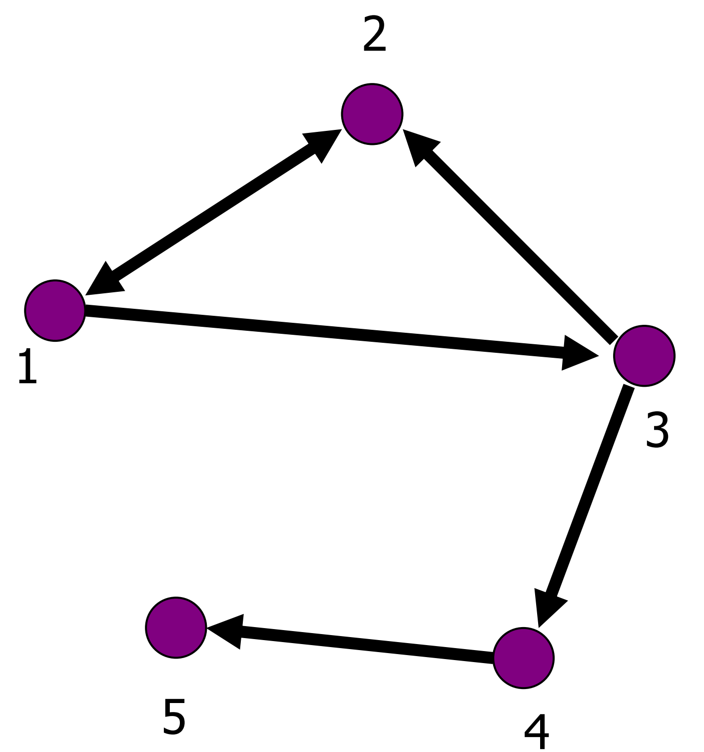
1: [2, 3]
2: [1, 3]
3: [1, 2, 4]
4: [3, 5]
5: [4]



Graph Representation

- Adjacency List
  - For each node keep a list of the nodes it points to

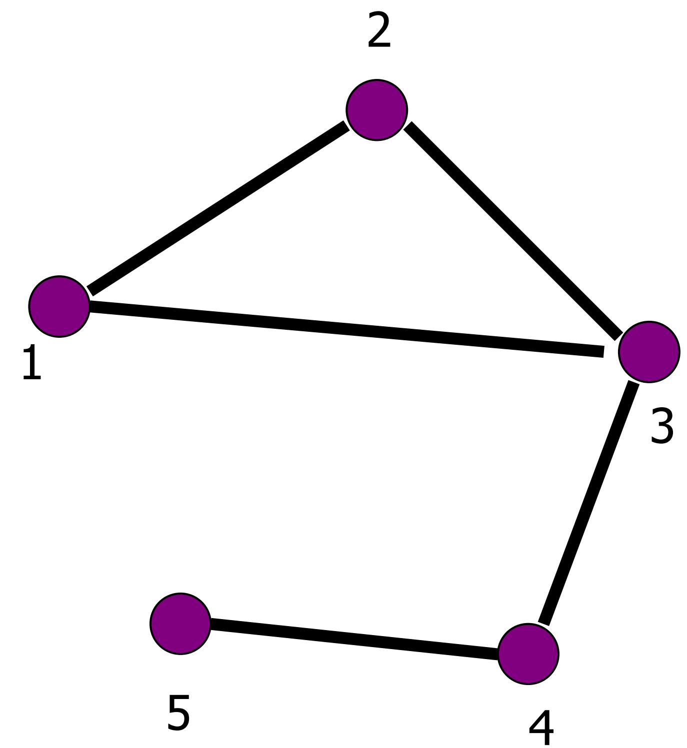
1: [2, 3]
2: [1]
3: [2, 4]
4: [5]
5: [null]



Graph Representation

- List of edges
  - Keep a list of all the edges in the graph

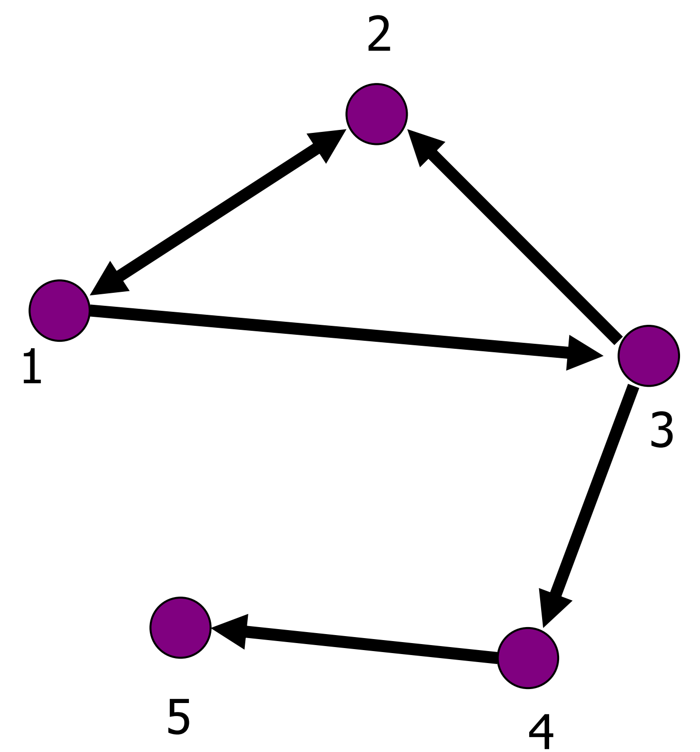
(1,2)
(2,3)
(1,3)
(3,4)
(4,5)



Graph Representation

- List of Edges
  - Keep a list of all the directed edges in the graph

(1,2)
(2,1)
(1,3)
(3,2)
(3,4)
(4,5)



P and **NP**

- **P**: the class of problems that can be solved in polynomial time
- **NP**: the class of problems that can be verified in polynomial time, but there is no known solution in polynomial time
- **NP**-hard: problems that are at least as hard as any problem in **NP**

Introduction to Notebooks

Notebooks

- Notebooks offer an interactive Python programming environment for coding, that combines code, text, and output.
- It is often used in Data Science for presenting the analysis output
- We will use Notebooks in class for testing in practice the different algorithms and models we present.
- For the course assignment you will maintain one or more notebooks that you will update with assignments on the latest material we cover in class.

Python

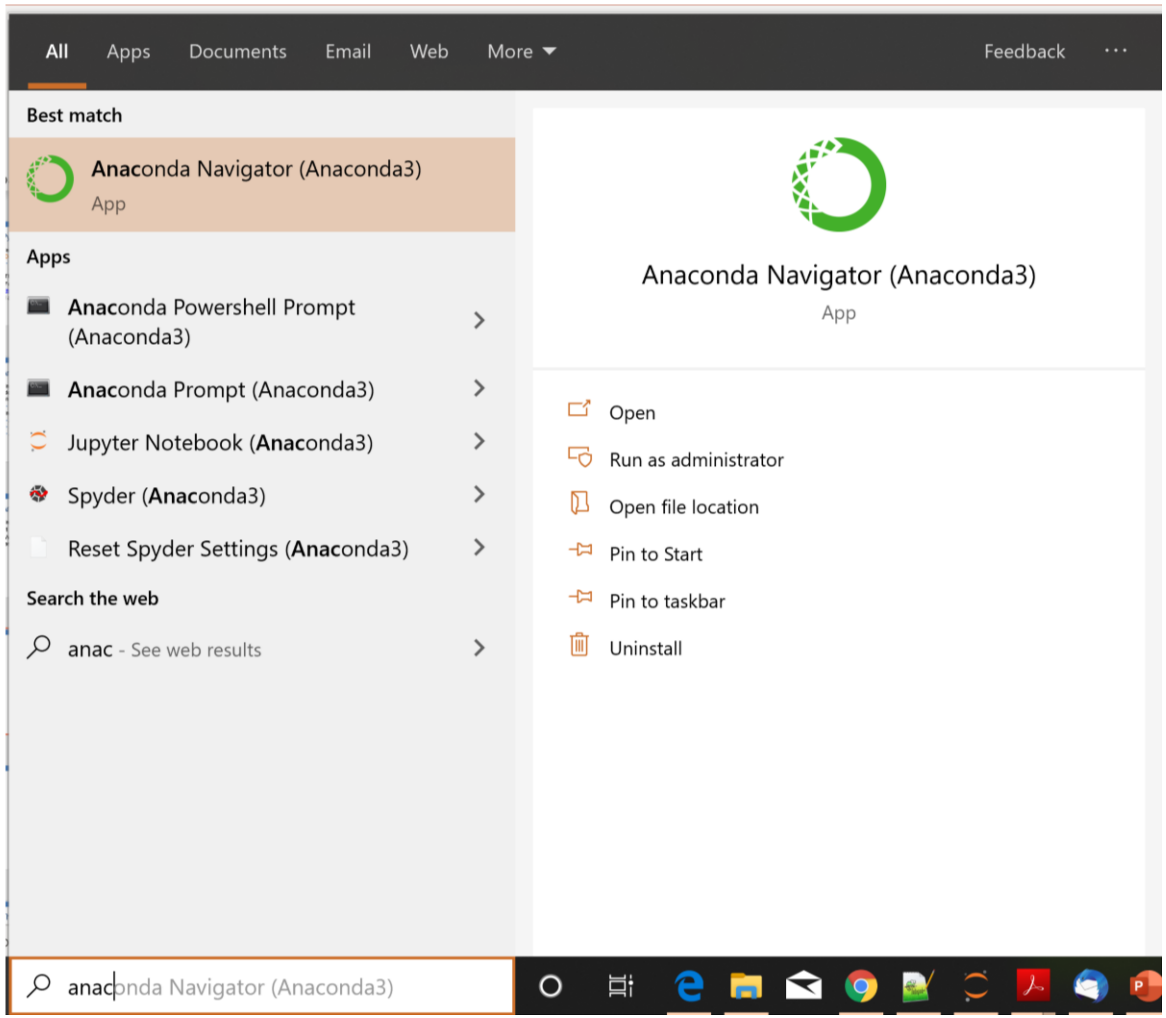
- In the last few years there is an increasing community that creates Data Mining tools in Python
  - Python is overwhelmingly used today for data science tasks
  - It is also heavily used in industry
  - We will use Python for this class.
- There are tons of resources online for Python.
- For an introduction you can also look at the slides of the Introduction to Programming course by prof. N. Mamoulis
- I will assume you have installed Python to your laptop, and that you have a good knowledge of programming in Python.

Anaconda

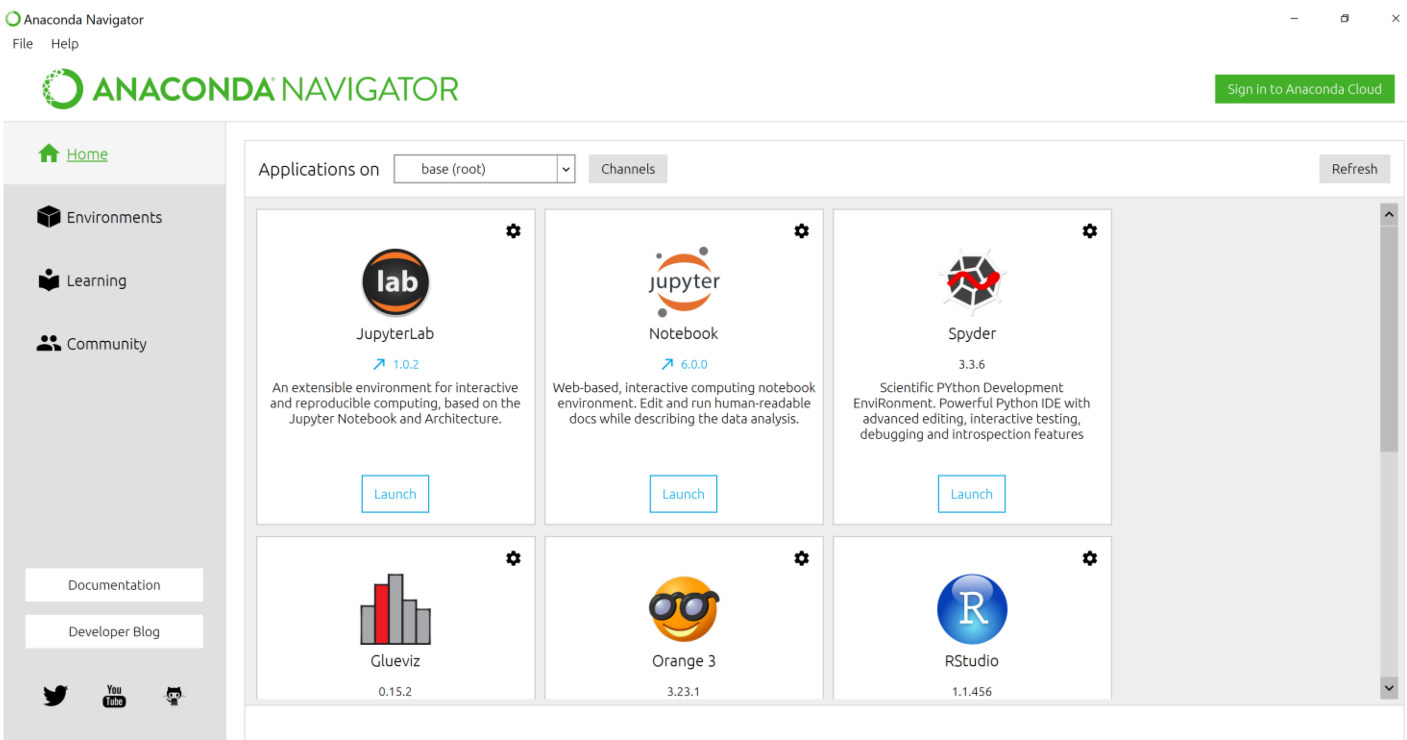
- Installing libraries in Python can be complicated, so you should download the Anaconda Scientific Python distribution which will install most of the libraries that we will use.
  - Use Python 3.0
- Installing Anaconda installs a lot of libraries and also:
  - Anaconda Navigator
  - Jupyter Notebook: An interactive web-based interface for running python.
  - Anaconda Powershell: terminal for running commands

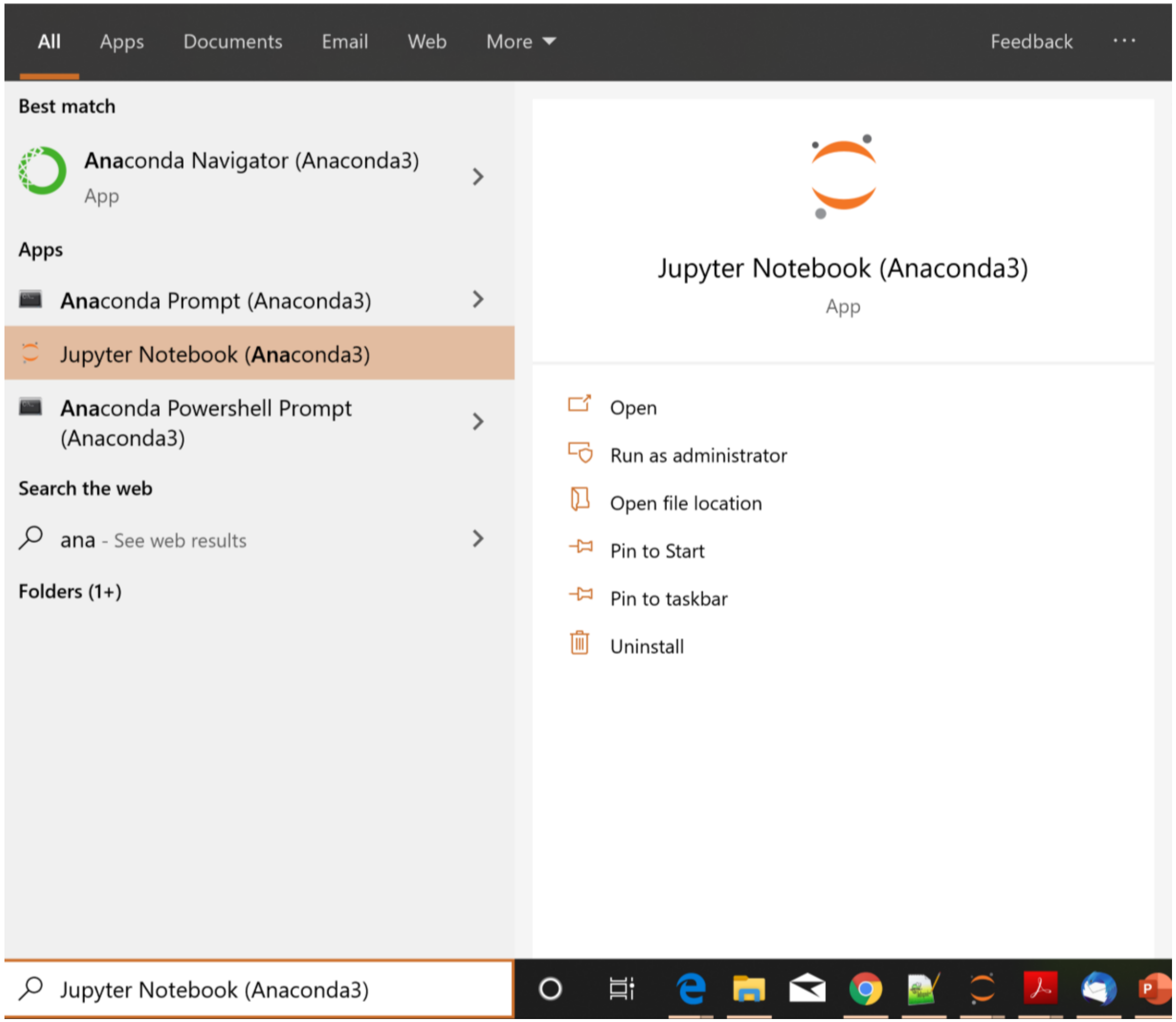
Jupyter Notebook

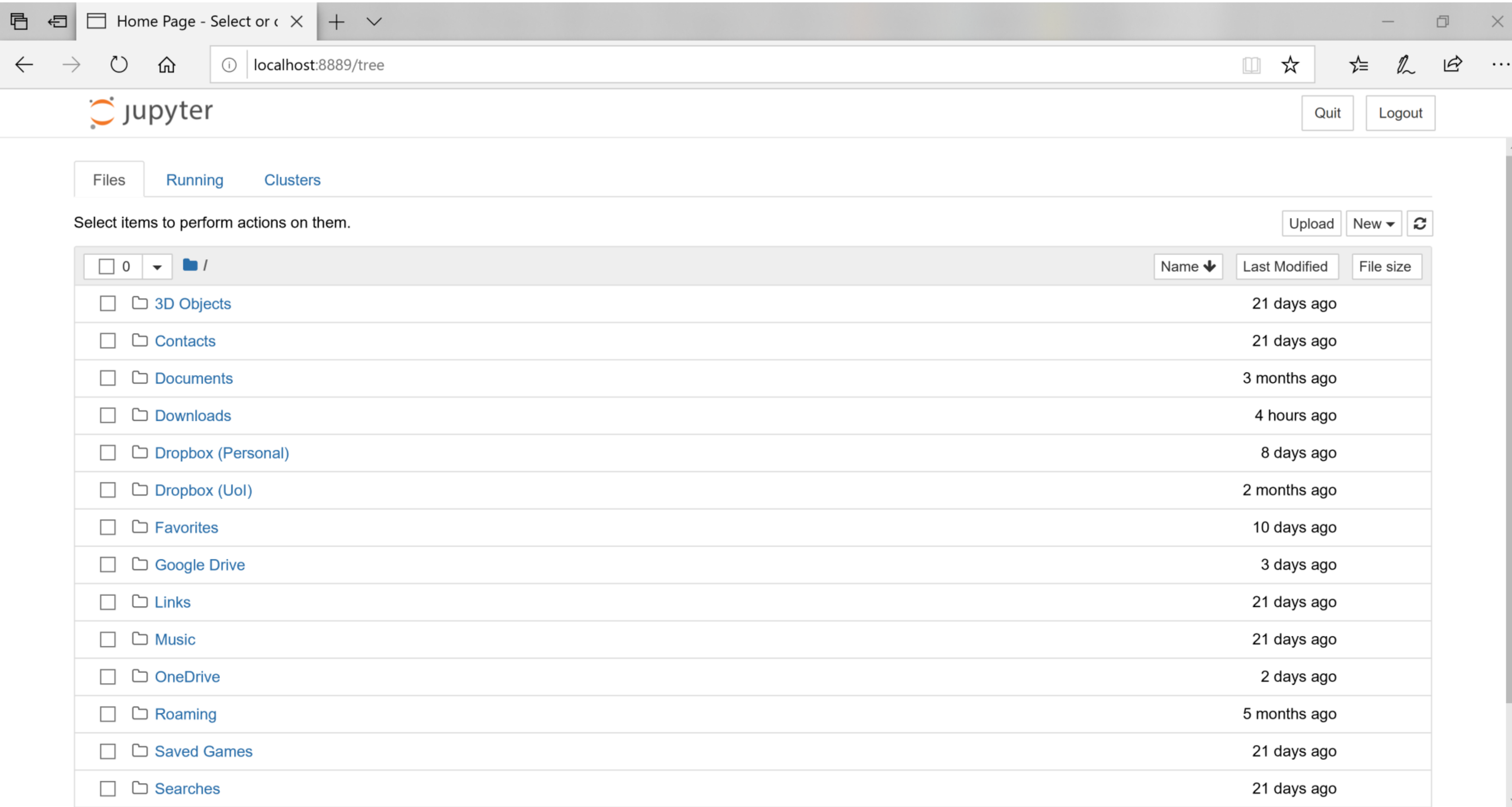
- Installing Anaconda will also install Jupyter Notebook.
- If you wish to install it in a different way, together with the relevant libraries you are free to do so.
- We will use Notebook for our examples and it is required for the assignments.
  - In almost all assignments you are required to submit a Notebook.

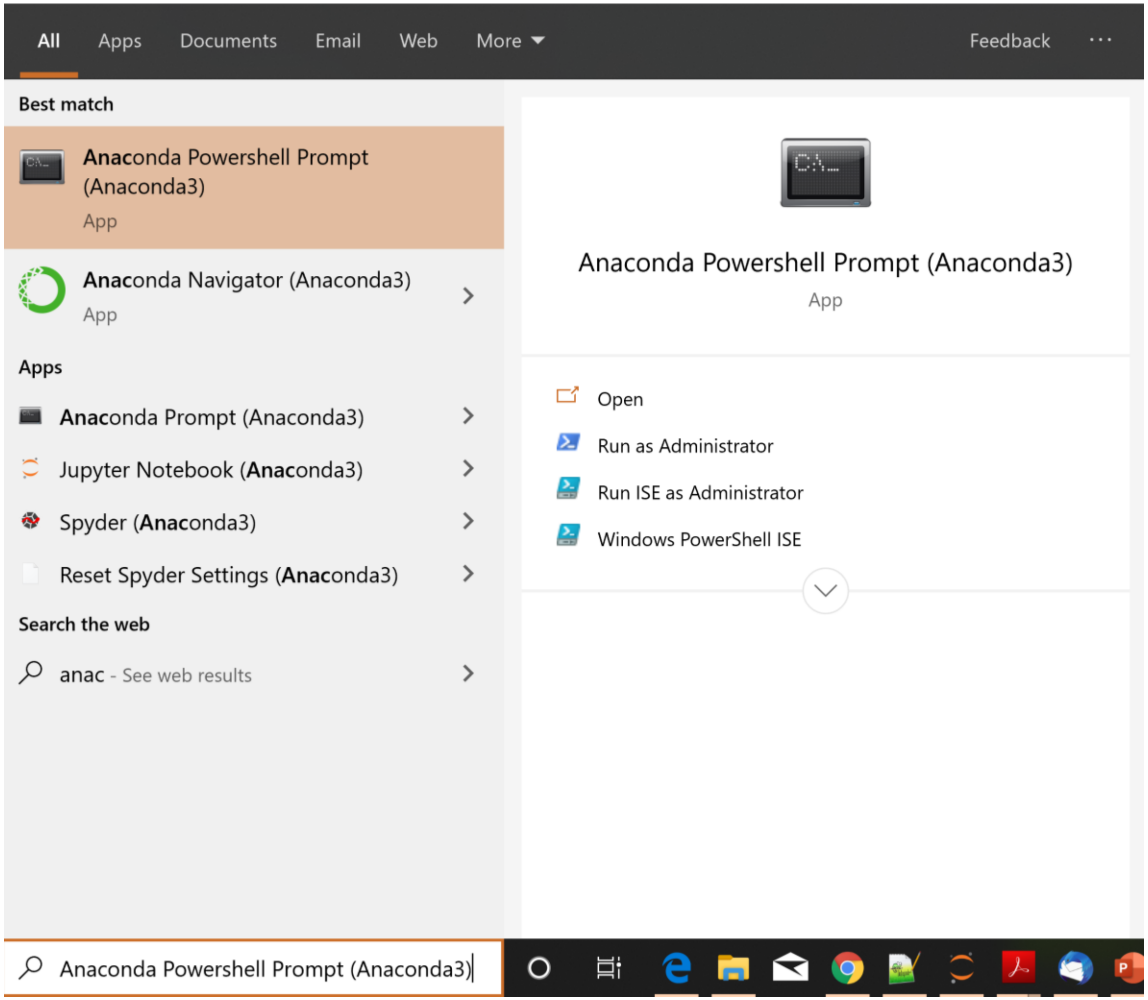

All
Apps
Documents
Email
Web
More
Feedback
Best match
Anaconda Navigator (Anaconda3)
App
Apps
Anaconda Powershell Prompt (Anaconda3)
Anaconda Prompt (Anaconda3)
Jupyter Notebook (Anaconda3)
Spyder (Anaconda3)
Reset Spyder Settings (Anaconda3)
Search the web
anac - See web results
Anaconda Navigator (Anaconda3)
App
Open
Run as administrator
Open file location
Pin to Start
Pin to taskbar
Uninstall
anaconda Navigator (Anaconda3)

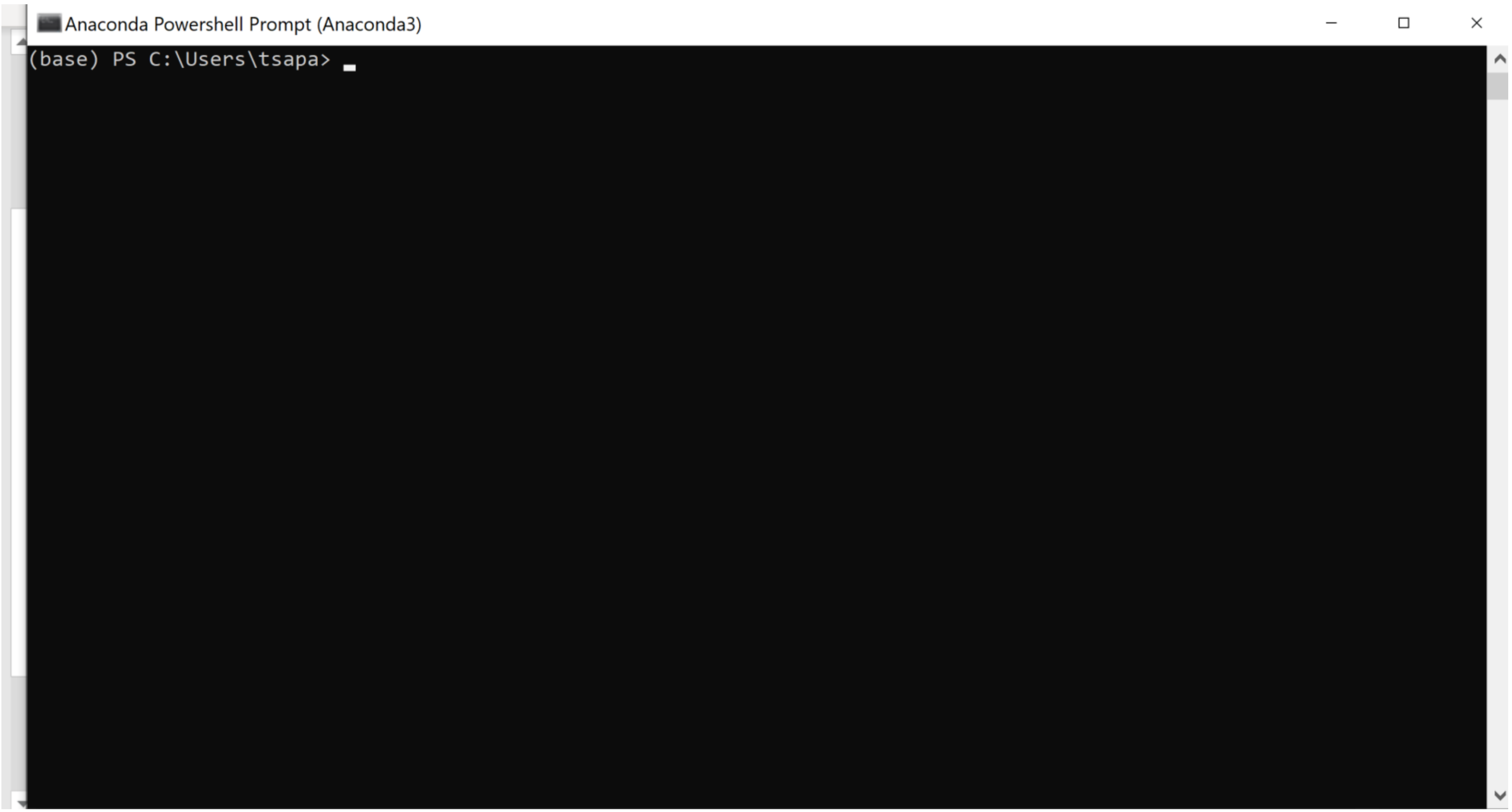
The Anaconda Navigator




All
Apps
Documents
Email
Web
More
Feedback
Best match
Anaconda Navigator (Anaconda3)
App
Apps
Anaconda Prompt (Anaconda3)
Jupyter Notebook (Anaconda3)
Anaconda Powershell Prompt (Anaconda3)
Search the web
ana - See web results
Folders (1+)
Jupyter Notebook (Anaconda3)
App
Open
Run as administrator
Open file location
Pin to Start
Pin to taskbar
Uninstall
Jupyter Notebook (Anaconda3)


Home Page - Select or c
localhost:8889/tree
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
0
/
Name
Last Modified
File size
3D Objects 21 days ago
Contacts 21 days ago
Documents 3 months ago
Downloads 4 hours ago
Dropbox (Personal) 8 days ago
Dropbox (UoI) 2 months ago
Favorites 10 days ago
Google Drive 3 days ago
Links 21 days ago
Music 21 days ago
OneDrive 2 days ago
Roaming 5 months ago
Saved Games 21 days ago
Searches 21 days ago


All
Apps
Documents
Email
Web
More
Feedback
Best match
Anaconda Powershell Prompt (Anaconda3)
App
Anaconda Navigator (Anaconda3)
App
Apps
Anaconda Prompt (Anaconda3)
Jupyter Notebook (Anaconda3)
Spyder (Anaconda3)
Reset Spyder Settings (Anaconda3)
Search the web
anac - See web results
Anaconda Powershell Prompt (Anaconda3)
App
Open
Run as Administrator
Run ISE as Administrator
Windows PowerShell ISE
Anaconda Powershell Prompt (Anaconda3)


Anaconda Powershell Prompt (Anaconda3)
(base) PS C:\Users\tsapa>

Installing Packages

- You can install packages from the Anaconda terminal using the conda install command:
  - `conda install <name of package>`

- For example, Seaborn is a package for Statistical Data Visualization.
  - `conda install seaborn`
- panda-datareader is a package for loading online datasets.
  - `conda install pandas-datareader`

- You can also use the pip install command.
  - `pip install snap-stanford`

Notebooks

- Jupyter Notebook offers an interactive web-based interface for running code.
- The Notebook runs inside a browser.
- It allows you to interact with the code, running different parts of the code
- The results also appear in the browser, so you can have together the code and the results
- You can also add text, commenting on the results.
- We will now see some details on how to create notebooks

Changing the notebook default directory

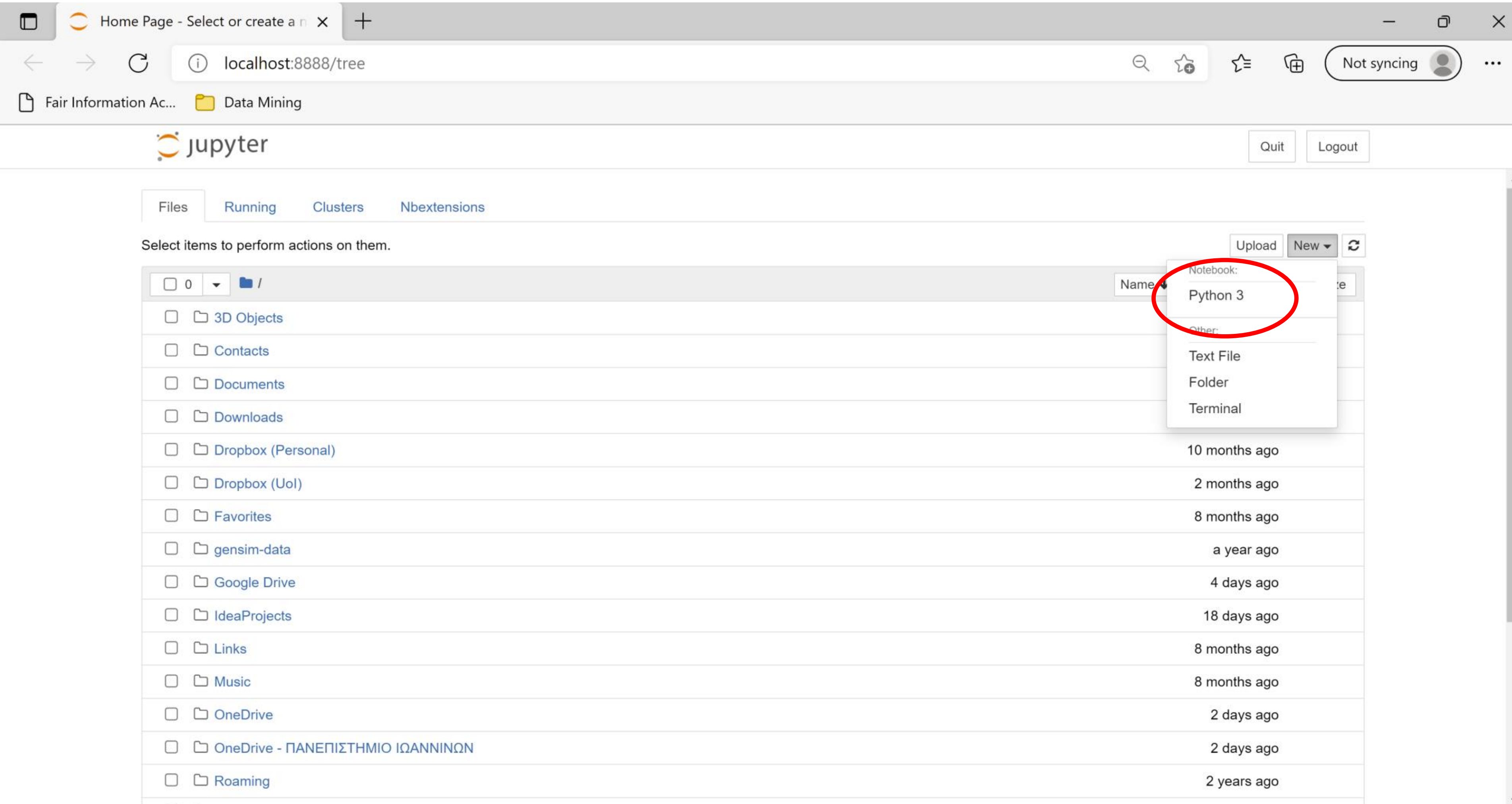
- This used to be important before but now Jupyter Notebook takes you to the home directory

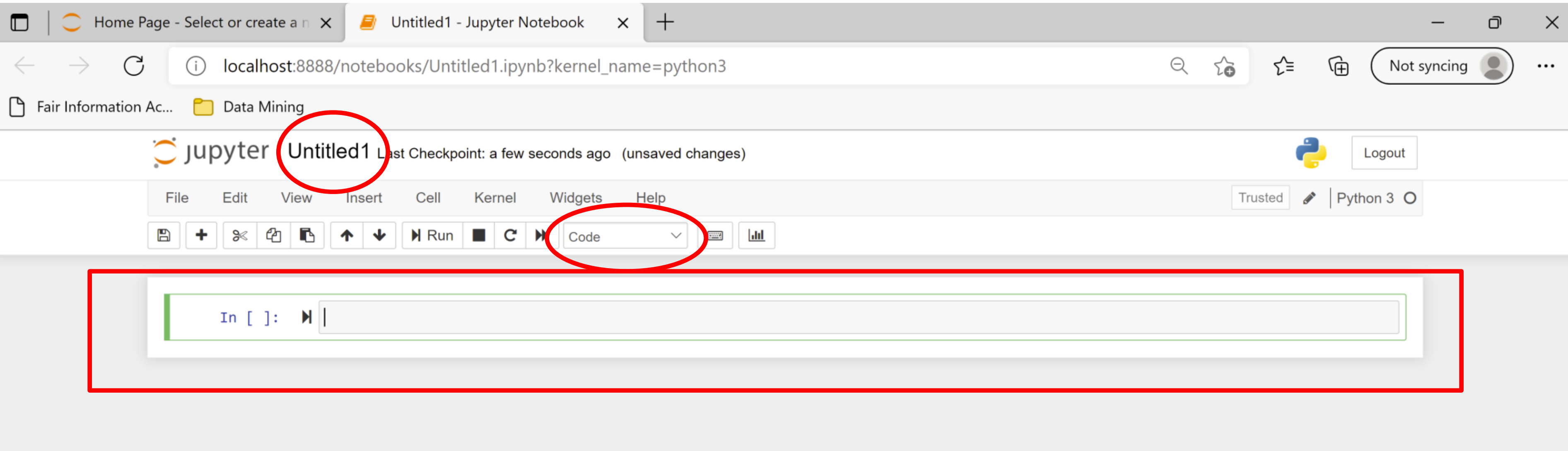
- From the Anaconda terminal type the command:
  - `jupyter notebook --generate-config`

- This will generate `.jupiter/jupyter_notebook_config.py` file under your home directory.

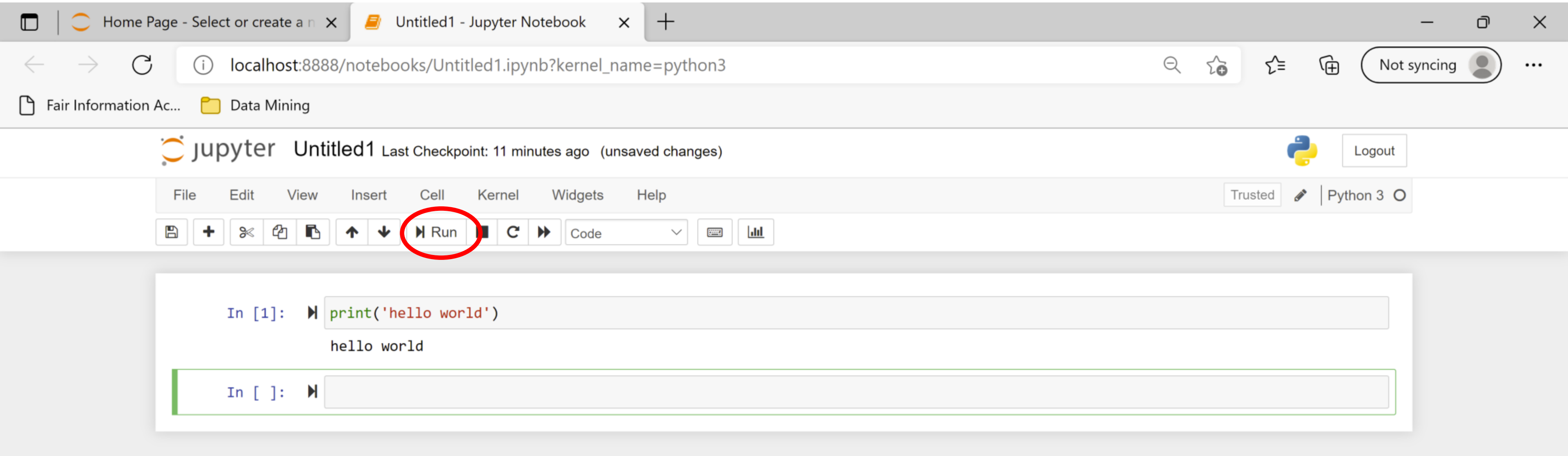
- Find, un-comment and modify the line

`# c.NotebookApp.notebook_dir = ''` in the config file to point to the desired directory

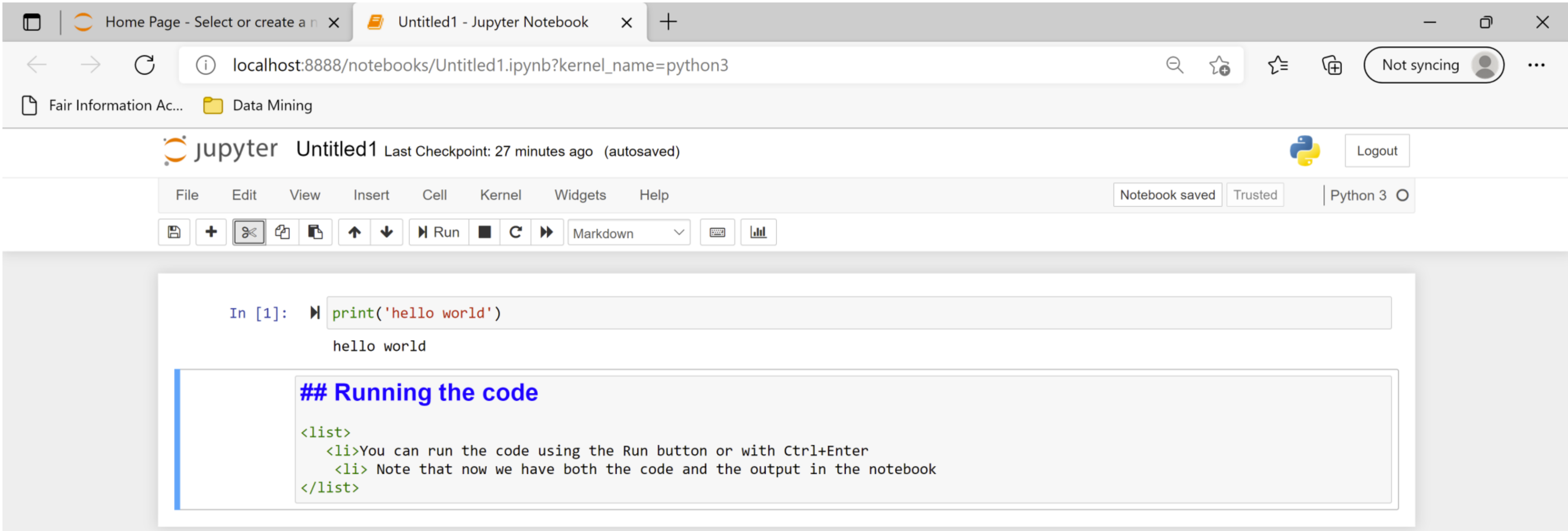

Home Page - Select or create a n
localhost:8888/tree
Not syncing
Fair Information Ac...
Data Mining
jupyter
Quit
Logout
Files
Running
Clusters
Nbextensions
Select items to perform actions on them.
Upload
New
0
/
Name
Notebook:
Python 3
Other:
Text File
Folder
Terminal
3D Objects
Contacts
Documents
Downloads
Dropbox (Personal)
10 months ago
Dropbox (UoI)
2 months ago
Favorites
8 months ago
gensim-data
a year ago
Google Drive
4 days ago
IdeaProjects
18 days ago
Links
8 months ago
Music
8 months ago
OneDrive
2 days ago
OneDrive - ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
2 days ago
Roaming
2 years ago



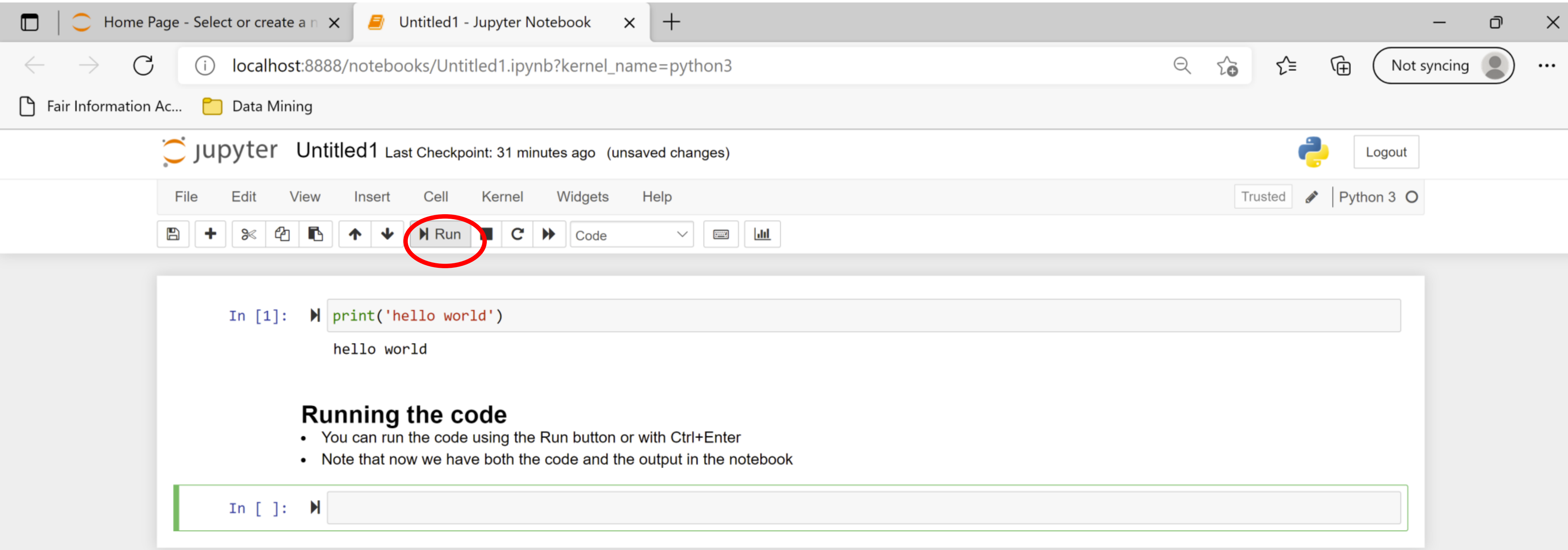
- The notebook is organized in cells
- In each cell you can write either code or text
- The default behavior is code

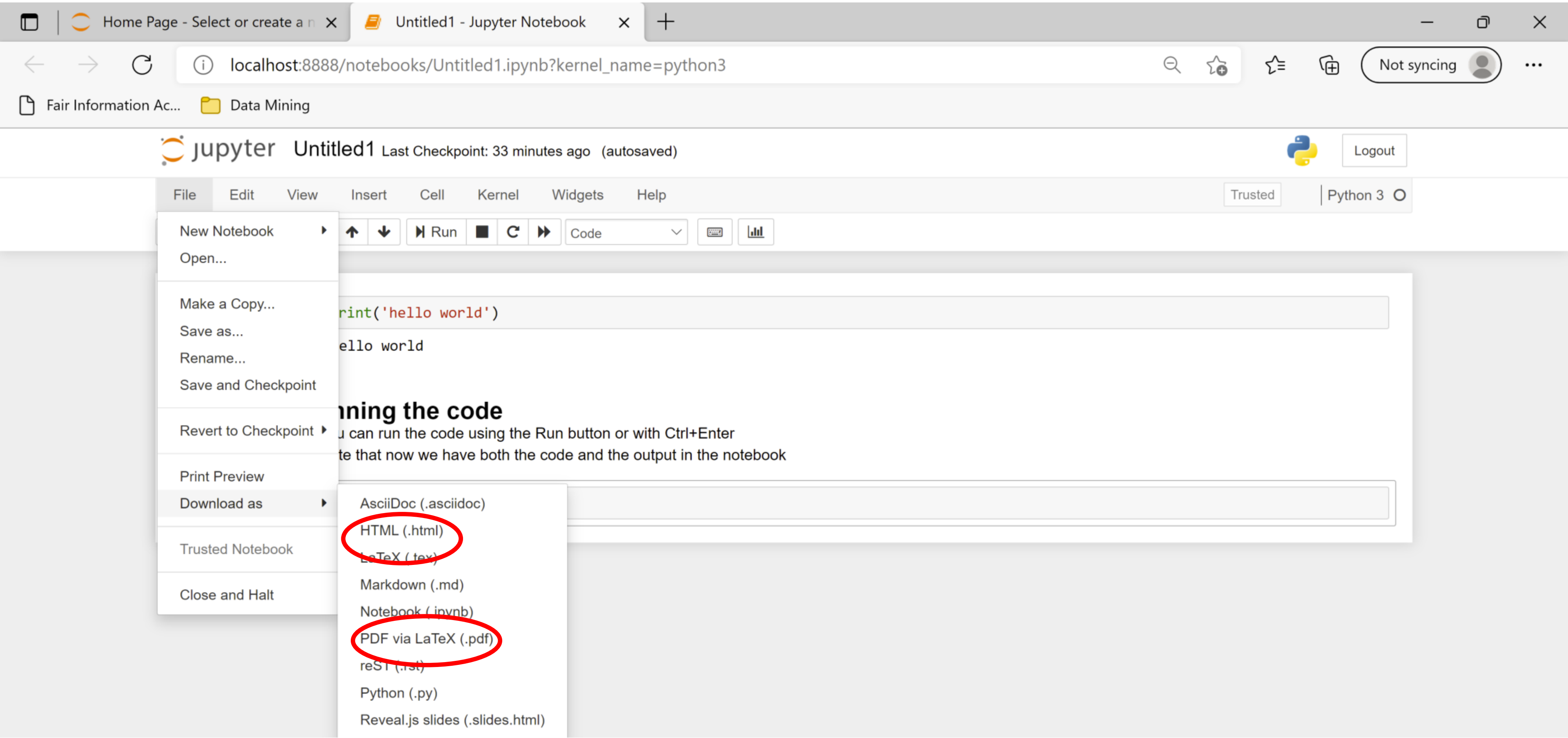


- You can run the code using the Run button or with Ctrl+Enter
- Note that now we have both the code and the output in the notebook



- You can also write text in Markdown language
  - You can combine HTML, and Latex, and there are some other commands
- You can learn more about Markdown by searching online, e.g.:
  - Learn How to Write Markdown & LaTeX in The Jupyter Notebook | by Khelifi Ahmed Aziz | Towards Data Science
- You need to Run the Markdown cell as well


Home Page - Select or create a n
Untitled1 - Jupyter Notebook
localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3
Not syncing
Fair Information Ac...
Data Mining
jupyter
Untitled1
Last Checkpoint: 31 minutes ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code
In [1]:
print('hello world')
hello world
Running the code
You can run the code using the Run button or with Ctrl+Enter
Note that now we have both the code and the output in the notebook
In [ ]:

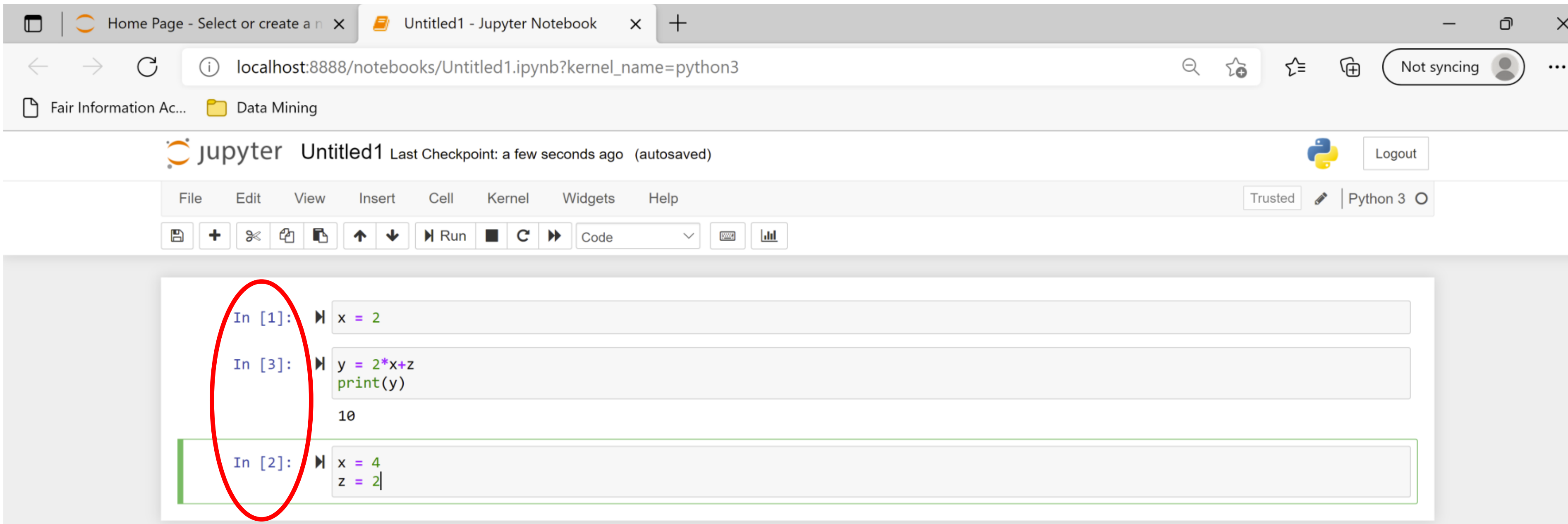


- You can export the notebook into HTML or PDF

Attention!

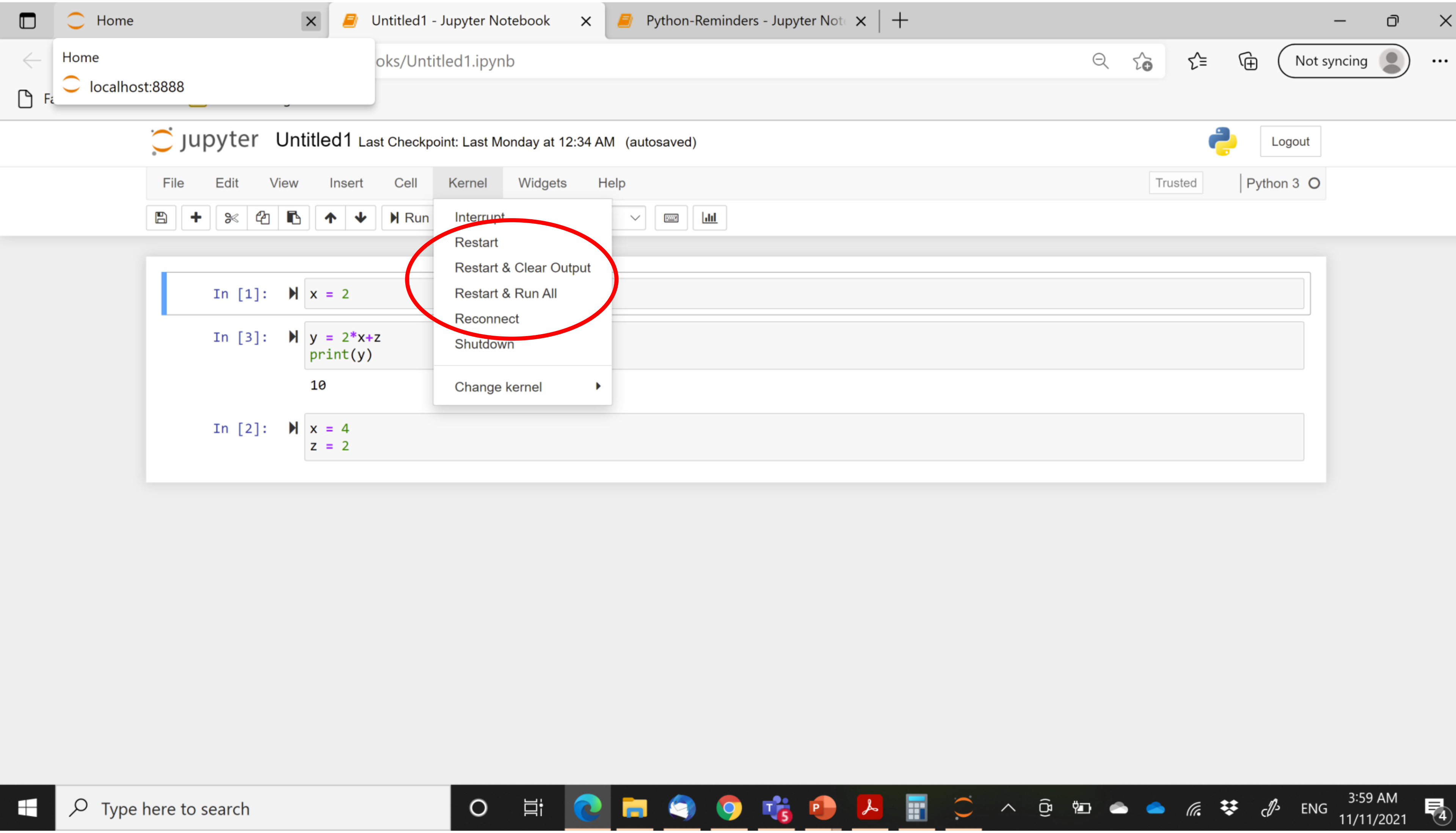
- A notebook is run interactively, each time running a specific cell
- The state of the program remains in memory while the notebook is running
- Each cell has access to the current state of the memory
- You can jump between cells in a non-linear way
- You should be aware of the state of the memory of the notebook when you run a specific cell.

A simple example



- The order in which the cells are executed is shown in the increasing numbers (not always useful)
- The second in order cell is executed third
  - So it as access to z, and uses the value 4 for x

Restarting


Home
Untitled1 - Jupyter Notebook
Python-Reminders - Jupyter Not
Home
localhost:8888
oks/Untitled1.ipynb
Not syncing
jupyter
Untitled1 Last Checkpoint: Last Monday at 12:34 AM (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Interrupt
Restart
Restart & Clear Output
Restart & Run All
Reconnect
Shutdown
Change kernel
In [1]: x = 2
In [3]: y = 2*x+z
print(y)
10
In [2]: x = 4
z = 2
Type here to search
ENG
3:59 AM
11/11/2021

Useful Links

- You can find some more information about notebooks in the tutorials for the Data Mining class
- Useful libraries for network processing:
  - NetworkX
  - Snap