Class intro

Chrysafis Vogiatzis

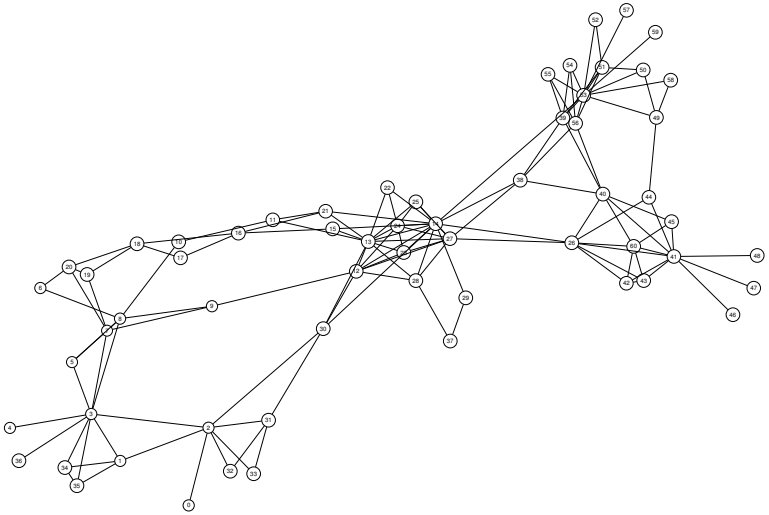
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 1: 08/25/2020

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Welcome to IE 532

In this class, we will study **networks**.



"A collection of points together with lines that connect some subset of the points."

– Wolfram Alpha

We define a mathematical construct called graph G with two sets:

1. a set of "points" (vertices V or nodes N);
2. a set of "lines" connecting "points" (edges E or arcs A).

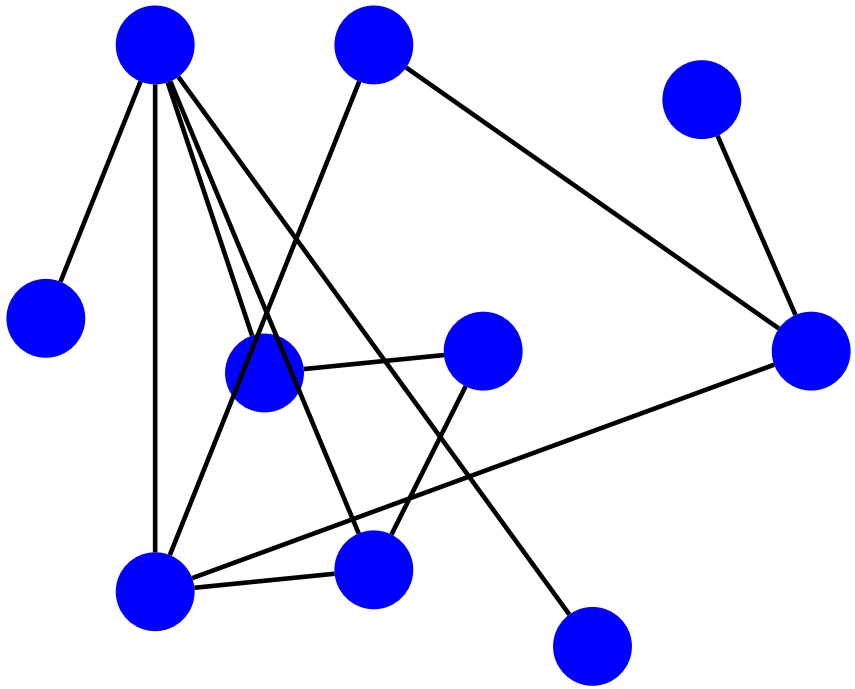
ILLINOIS

Analysis of networks

Networks are the “language” of complex systems:

- social networks;
- transportation networks;
- communications networks;
- power networks;
- biological networks; etc.

In the above, we are presented with **entities** and their **interactions**.



ILLINOIS

Examples of networks: from transportation

Figure: The transportation network of the state of WV, including only major highways. Obtained using `osmnx` (https://github.com/gboeing/osmnx).

ILLINOIS

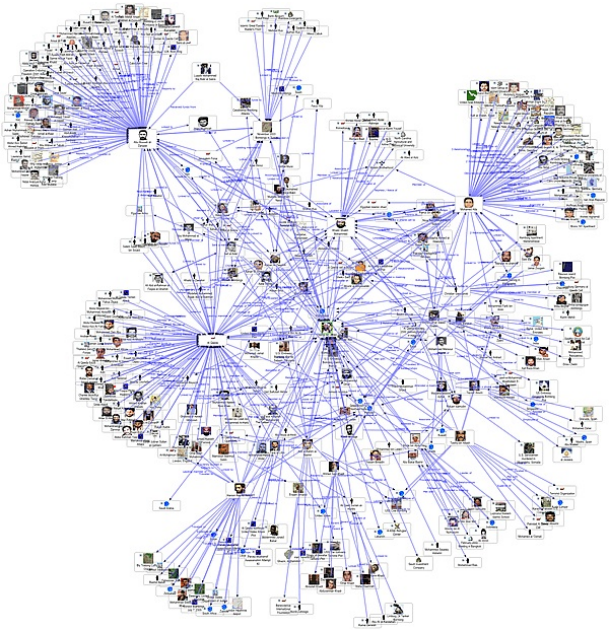
Examples of networks: airport networks

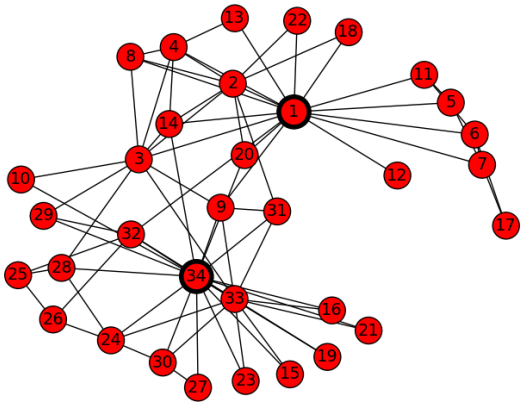
Figure: The airport network (in 2016, as shown in http://www.martingrandjean.ch/wp-content/uploads/2016/05/airports-map.png).

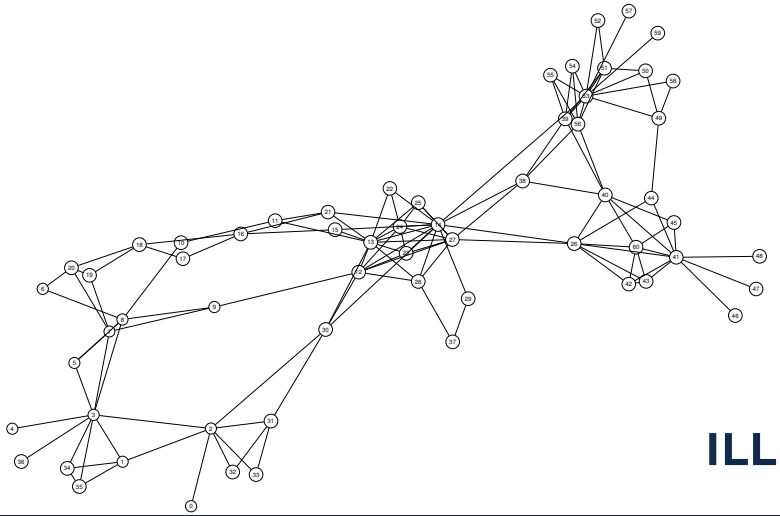
ILLINOIS

Examples of networks: social networks

Both online (facebook, twitter, instagram, email, phone calls) and offline (social interactions, karate club graph, dolphins, 9-11, book graphs).

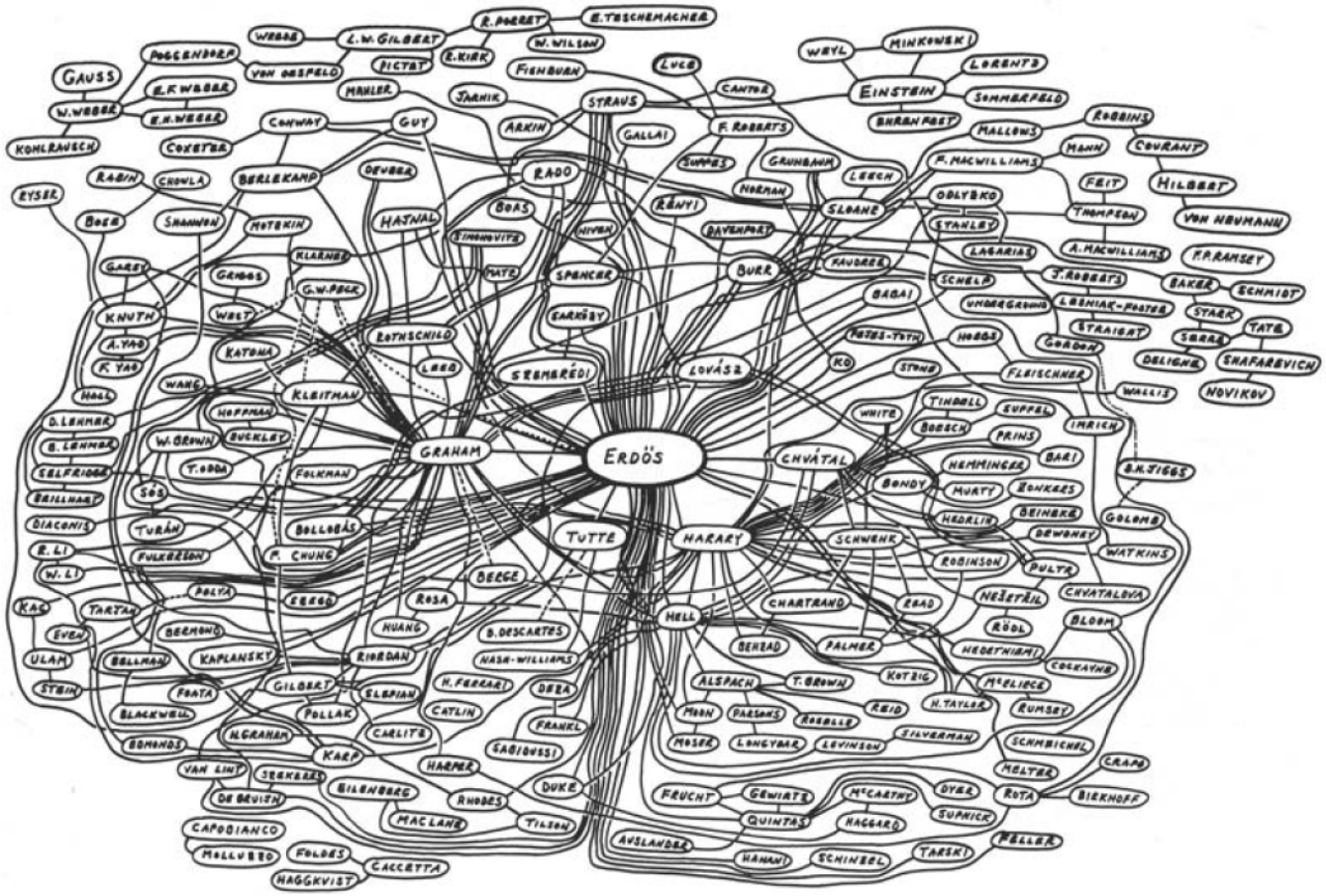






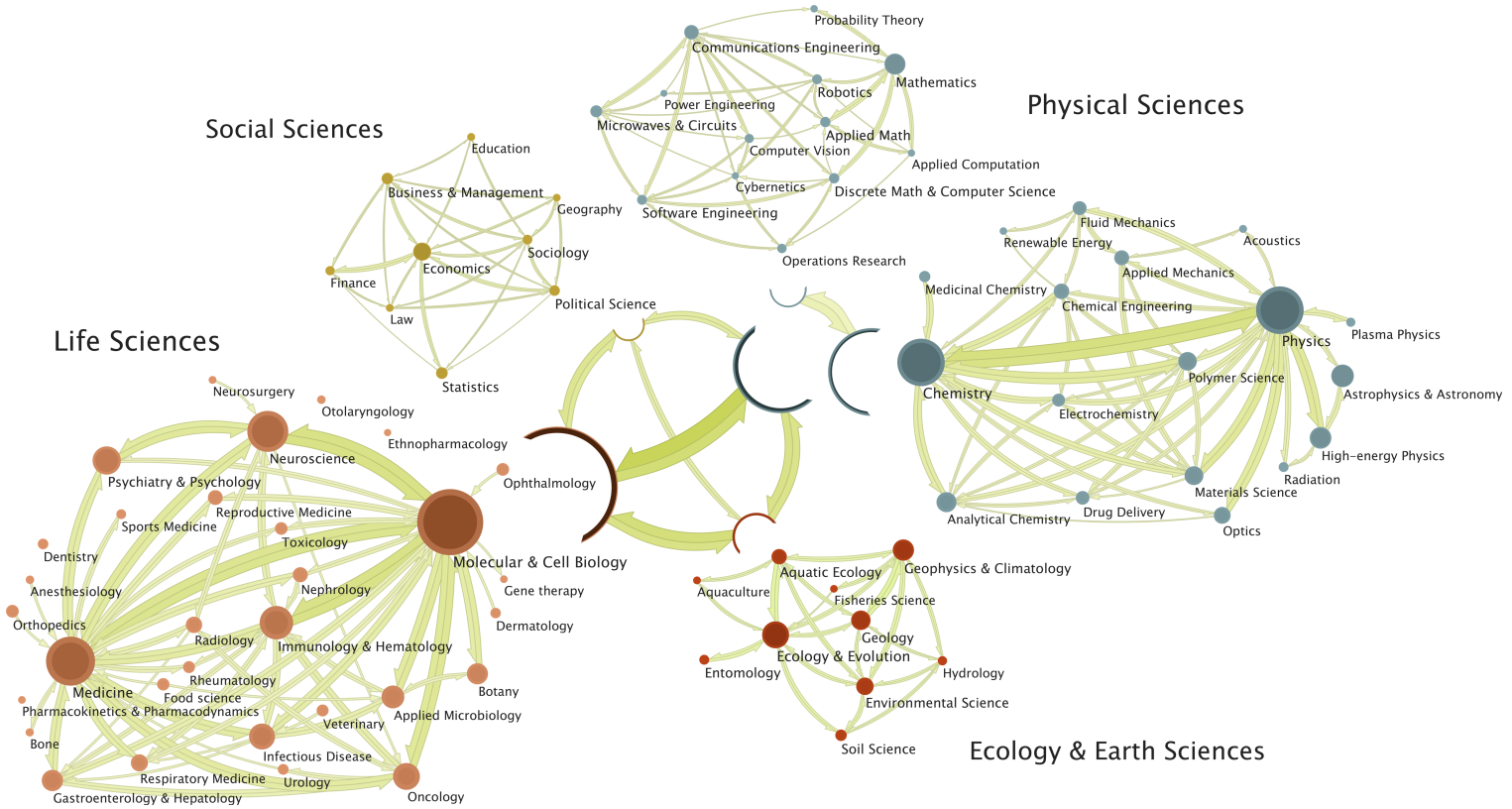
Examples of networks: collaboration networks

Figure: Part of the collaboration network of Paul Erdös (can be seen near the center). Hand-drawn by Ronald Graham. What is your Erdös number?



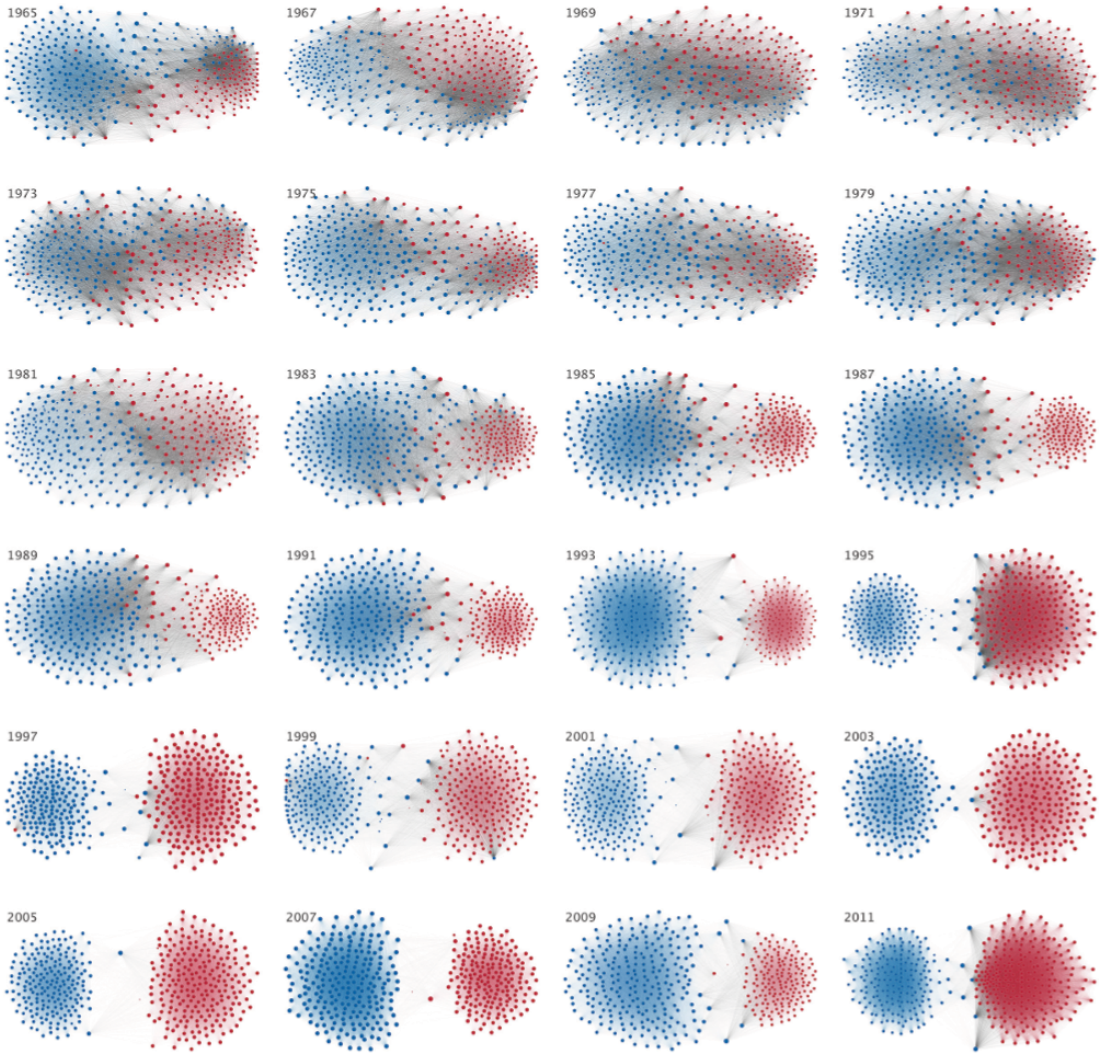
Examples of networks: mapping the sciences

Figure: Relationships between disciplines based on citations. Taken from “Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems” by Rosvall and Bergstrom (2011), PLOS ONE 6(4): e18209. Accessed through this link.



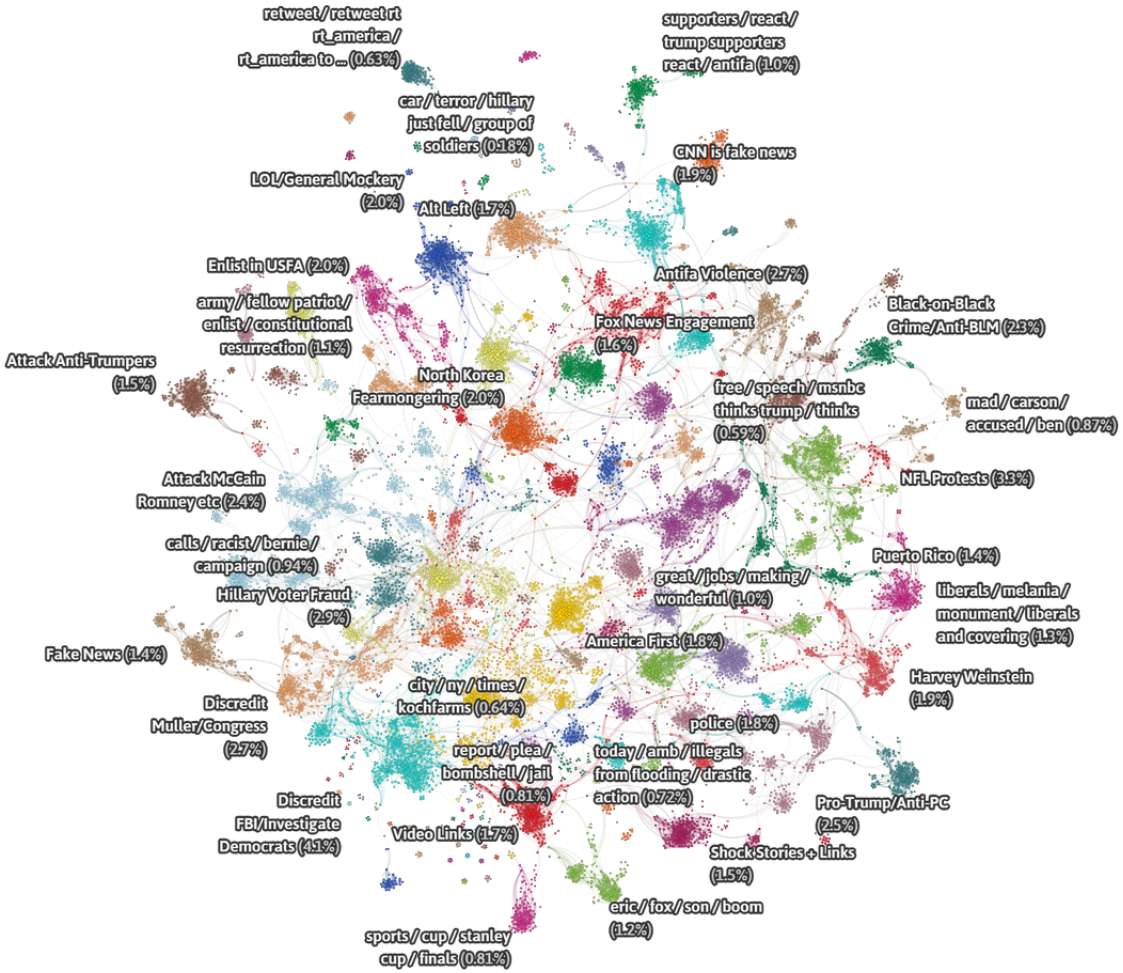
Examples of networks: from politics

Figure: Polarization in Senate voting from 1965 to 2011. Taken from this post.



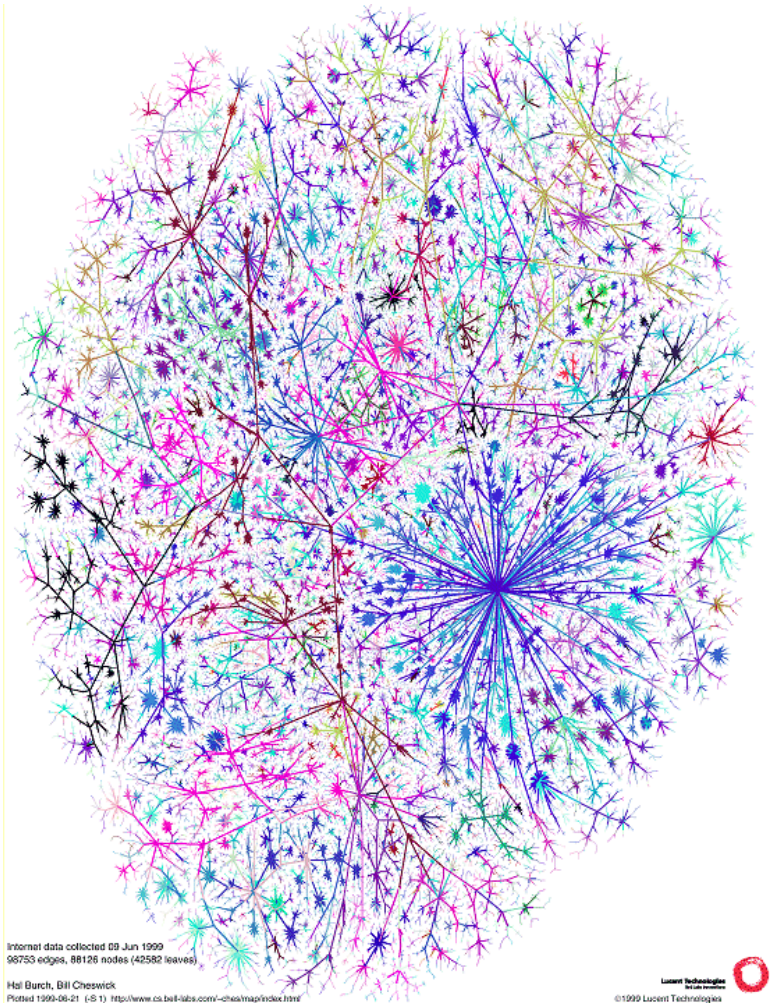
Examples of networks: from reddit

Figure: Network of pro-Trump subreddits during election period 2016.



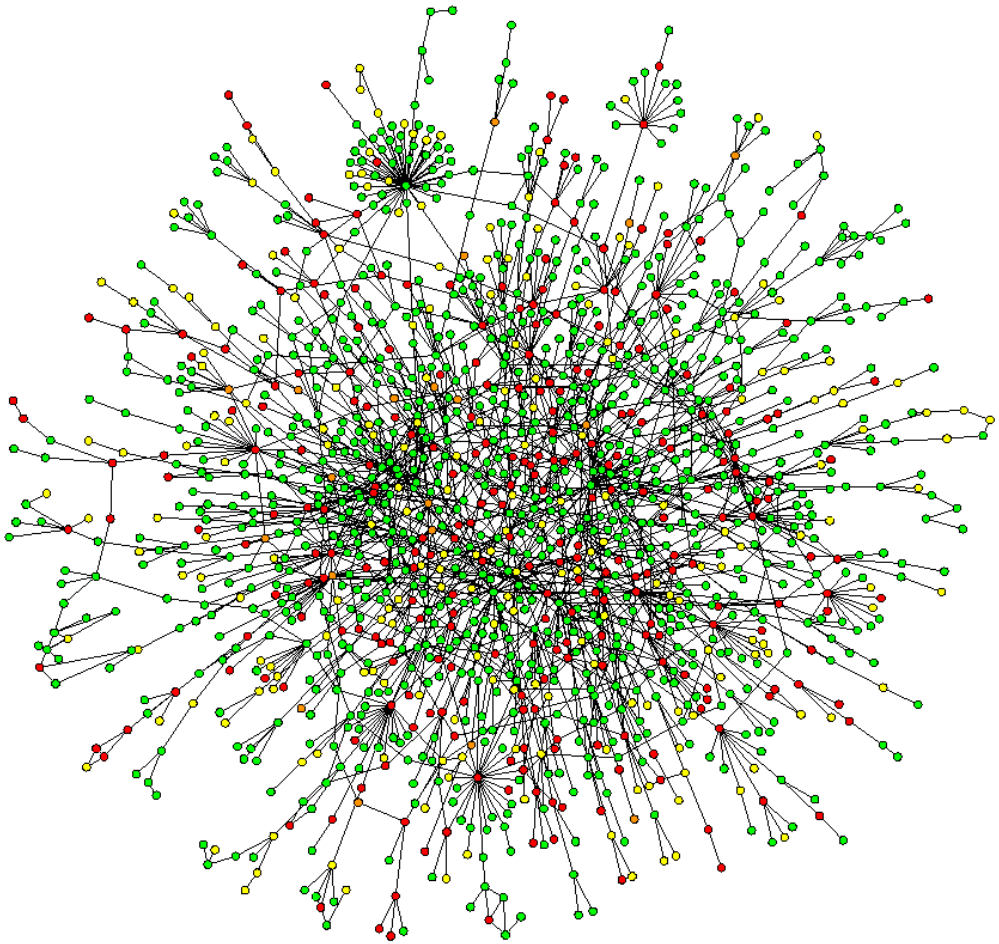
Examples of networks: the Internet

Figure: The directory of the main domains of the world wide web in 1999. Created by Bill Cheswick and accessed through here: http://www.cheswick.com/ches/map/index.html.



Examples of networks: biological networks

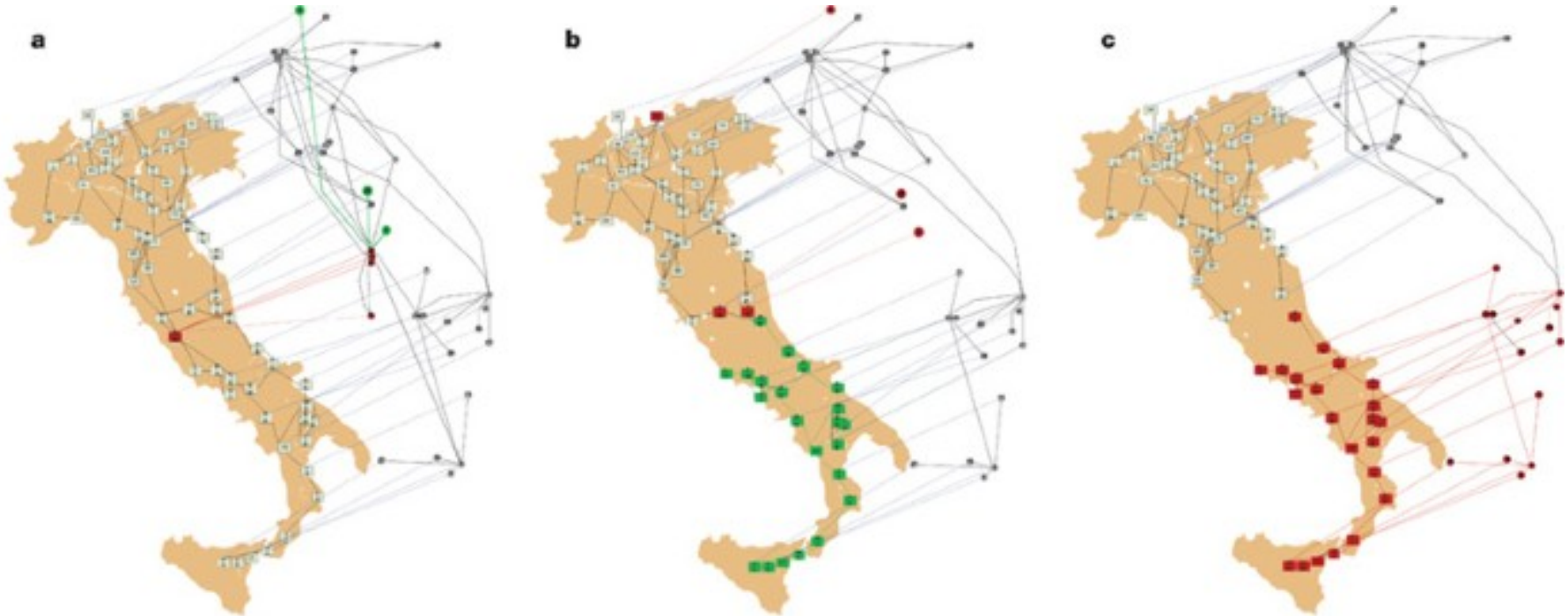
Figure: The protein-protein interaction network of Saccharomyces cerevisiae (source by Hawoong Jeong, KAIST, Korea).



ILLINOIS

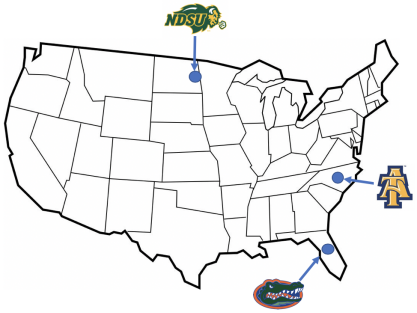
Examples of networks: cyber-physical systems

Figure: The cascading failures that led to the catastrophic blackout of Northern Italy in 2003. Taken from “Catastrophic cascade of failures in interdependent networks" by Buldyrev, Parshani, Paul et al. Nature 464 (2010). https://doi.org/10.1038/nature08932.



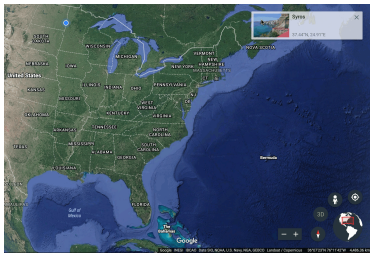
ILLINOIS

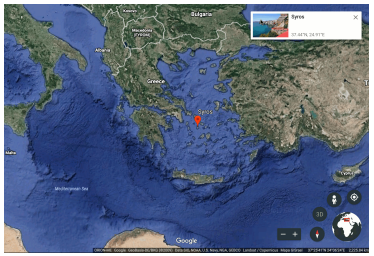
Who I am



- Ph.D. in ISE @ UF.
- Worked at NDSU, NCAT.
- Teaching Assistant Professor in ISE @ UIUC.
- OR, mathematical programming, combinatorial optimization, **network optimization and analysis**.
- Teaching IE 300 (Analysis of Data) and IE 532 (Analysis of Network Data).
- I come from a small island in the Aegean Sea in Greece

What do you mean "in the middle of the sea"?





Who I am

Problems I work on:

- **Network** analysis and optimization.
- Evacuation and disaster management.
- Biological **networks** and phylogenetics.
- Data analytics using **combinatorial optimization**.

Awards I have received:

- Graduate student **teaching** award (2012, UF).
- ISE **teaching** award (2012, UF).
- Now You See Me: Best paper award of the European Intelligence and Security Informatics Conference (2018).
- Best undergraduate **teaching** award (2019, NCAT).
- Best graduate **teaching** award (2019, NCAT).
- Sharp outstanding **teaching** award in Industrial Engineering (2020, UIUC).

ILLINOIS

How do I call you?

Over the years, I've been called or emailed as:

- Professor/Prof. Chrysafis/Chrys;
- Sir/Mister/Mr. Chrysafis/Chrys;
- Man;
- Hey;
- Chrysafis/Chrys;
- Dr. Vogiatzis/Dr. V./Dr. Chrys;
- Greek guy.

and Dude and Bro and...

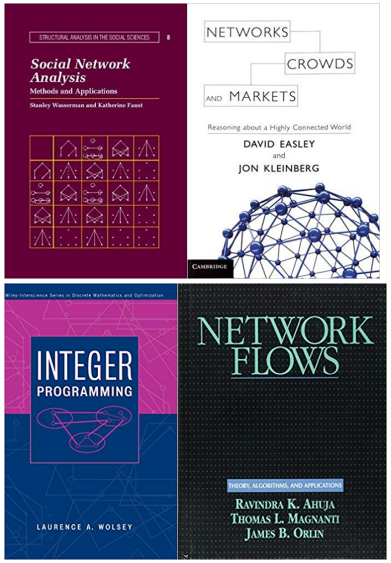
I'm 100% fine with:

- Professor/Prof. Chrysafis/Chrys;
- Chrysafis/Chrys;
- Dr. Vogiatzis/Dr. V./Dr. Chrys.

ILLINOIS

Class logistics

- Class time:
  - TR 2.00pm–3.20pm at ~~TB 206~~ this zoom link.
- Textbooks that I will be using:
  - **Social Network Analysis: Methods and Applications by Faust and Wasserman**
  - **Networks, Crowds, and Markets** by Easley and Kleinberg.
  - **Network Flows** by Ahuja, Magnanti, Orlin.
  - **Integer Programming** by Laurence A. Wolsey.
  - **Highly** recommended, not required.

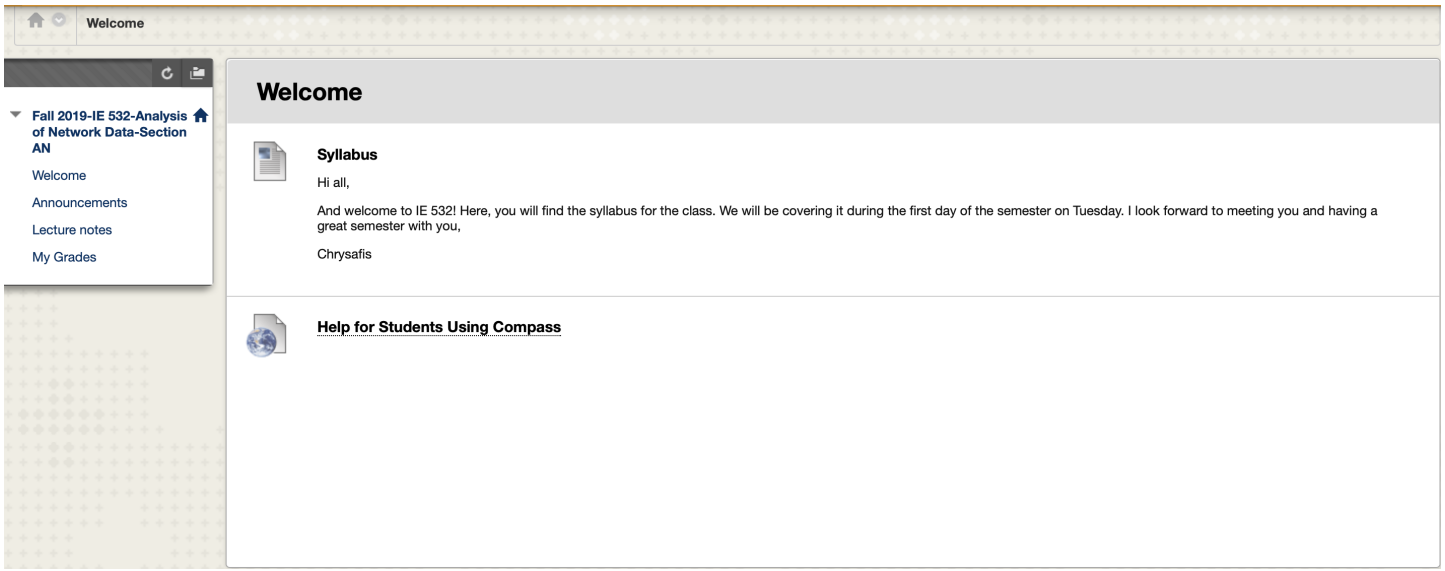


- Software:
  - You will also need to have access to an optimization solver software (`Gurobi`) and a network analysis package (`networkX`) in Python.
  - We will see how to install and use these packages in due time.
- Course materials:
  - All material (lecture notes, HW assignments, examples, etc.) will be posted on Compass 2G.

ILLINOIS

Compass 2g

- http://compass2g.illinois.edu/
- Only available to registered students.



ILLINOIS

Grading

Your letter grade in the class will be calculated based on:

1. Two take home exams: a midterm (20%) and a final (20%).
   - Open books/open notes, but no collaboration between students allowed.
   - The final exam is cumulative.
2. Homework assignments, approximately 5-6 throughout the semester (20%) – collaboration between students allowed and encouraged, so long as it is acknowledged.
3. In-class, small activities, approximately 5-6 throughout the semester (20%).
4. Term project (20%), which can be individual or done in a small group of up to 3 students – more details will be announced later in the semester.

ILLINOIS

What I know about you

...

Help me amend this by filling in the (anonymous) survey online. I will give you 10 minutes to go through it.

The link to the survey is here (becomes active Tuesday 08/25 at 2pm and will stay open until Monday end-of-day):

https://surveys.illinois.edu/sec/493644217

ILLINOIS

Where to begin

- The “Seven Bridges of Königsberg”.
- Leonhard Euler in 1736.
- In the city of Königsberg in Prussia (now Kaliningrad, Russia):

  “Can you start from *B* or *C*, cross every bridge exactly once and end up back in either *B* or *C*?”

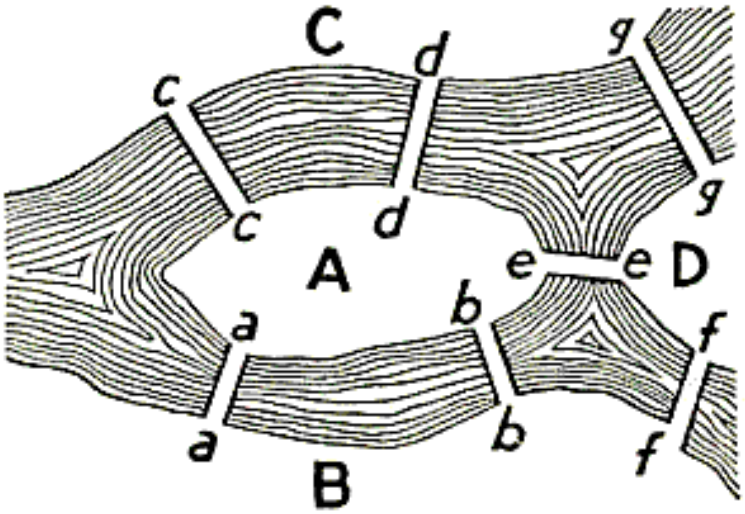
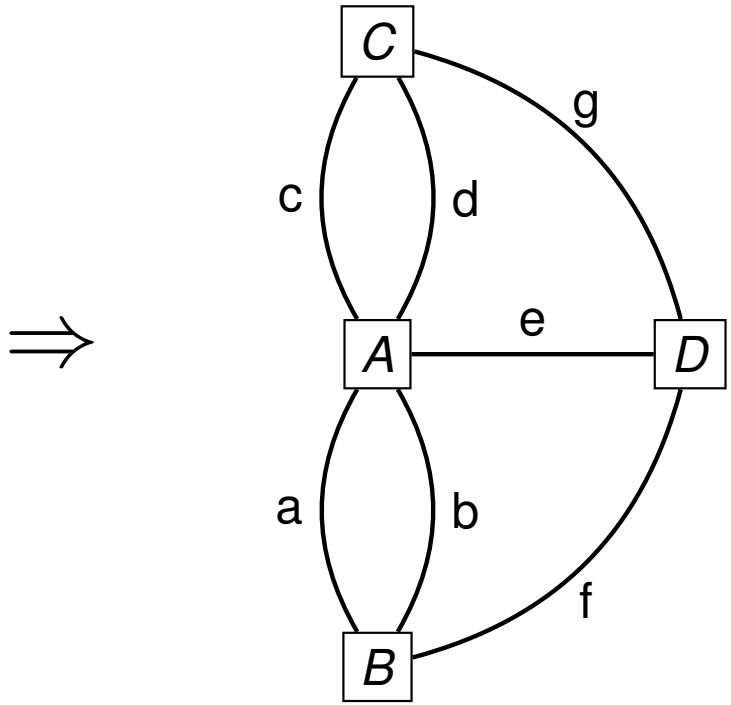


FIGURE 98. *Geographic Map: The Königsberg Bridges.*



- The negative result gave birth to the field of graph theory.

ILLINOIS

Fundamental notation

Given graph $G(V, E)$:

- V: set of vertices (usually labeled $1, \ldots, n$).
- E: set of edges (usually labeled $1, \ldots, m$).
- *undirected edge*: unordered pair of nodes $(i, j) = (j, i)$.
- *directed edge*: ordered pair of nodes $(i, j) \neq (j, i)$.
- *multigraph*: graph with multiple edges connecting two nodes.
  multi-edge (i, j).
- *self-loop*: an edge (i, i).

A graph without self-loops or multi-edges is **simple**. Most graphs in our class will be simple.

ILLINOIS

How to define a network

In general, you need to provide three things:

1. What are your nodes?
2. What are your edges?
3. What are your flows?

For example, in a water distribution network:

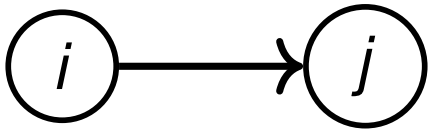
1. Nodes = homes and offices
2. Edges = pipes
3. Flow = water

Extra observation: water flow has a *direction*.

ILLINOIS

Directed/undirected

A directed edge has two **endpoints**: node i and j.



Terminology:

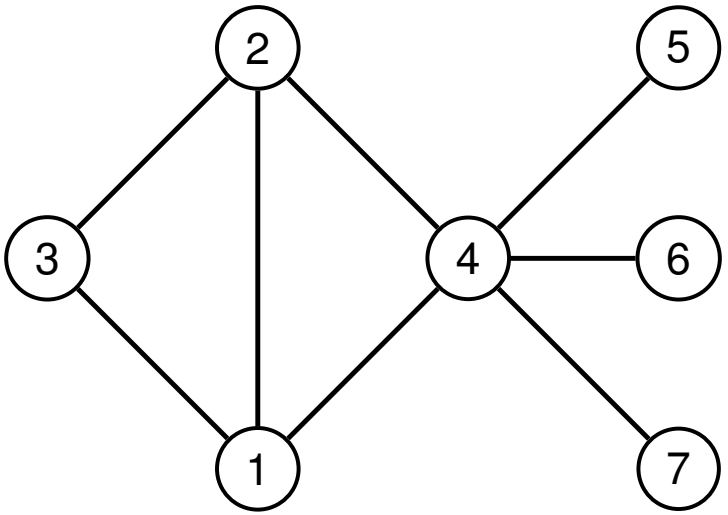
- i is called the tail and j the head of the edge;
- we also say that edge (i, j) emanates from node i and terminates at node j;
- edge (i, j) is outgoing from node i and incoming to node j;
- edge (i, j) is incident to nodes i and j;
- nodes i and j are adjacent.

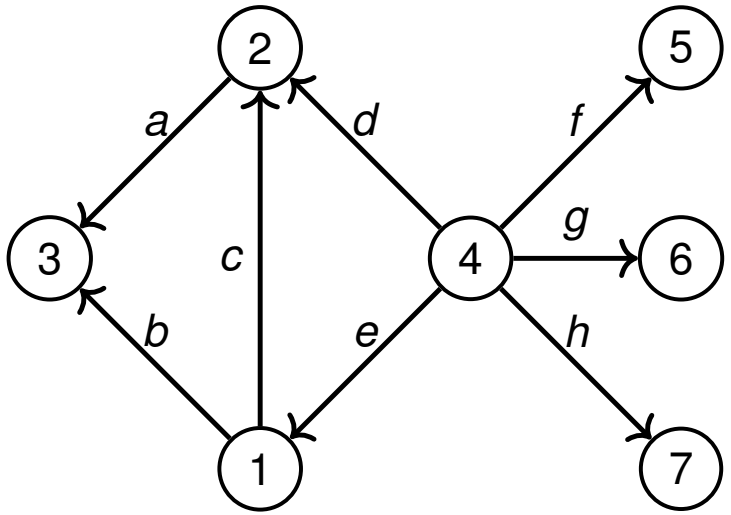
ILLINOIS

Directed/undirected

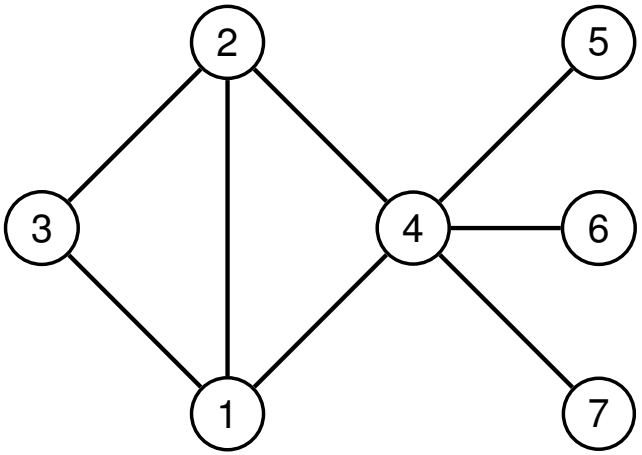
Graphs can be directed or undirected:

1. undirected graphs: with symmetric (reciprocal) edges.
   Typically represented by an **adjacency** matrix A, where $a_{ij} = 1$ if nodes i and j are connected by an edge.
2. directed graphs: with directed edges.
   Typically represented by a **node-arc** incidence matrix A, where for every (*directed*) edge $e = (i, j)$, we have $a_{ei} = 1$, $a_{ej} = -1$, $a_{ek} = 0$, for $k \neq i, j$.

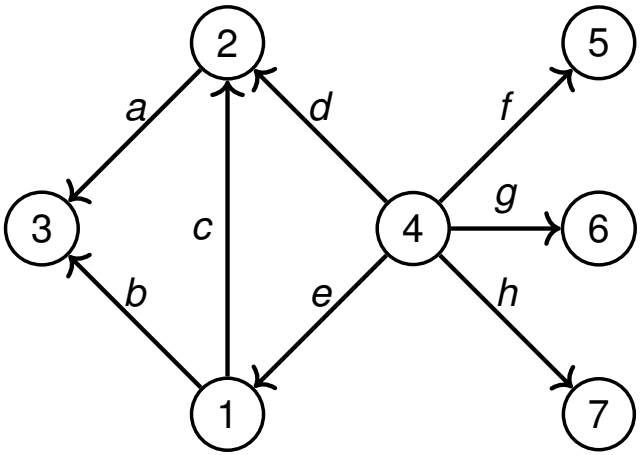




Representing directed and undirected networks



$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$



$$A = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \end{array}$$

ILLINOIS

More definitions

- **Walk** (or chain): a series of edges, such that each edge has exactly one node in common with the previous one without any respect for directions.

$$v_1 v_2 \dots v_k, \text{ such that } (v_i, v_{i+1}) \in E.$$

- **Directed walk**: a version of the walk where we respect all directions.
- **Path**: a walk, without any repetitions of nodes.

$$v_1 v_2 \dots v_k, \text{ such that } v_i \neq v_j.$$

- **Directed path**: a directed version of a path.
- **Cycle**: a path where the starting and ending nodes are the same.

$$v_1 v_2 \dots v_k, \text{ such that } v_1 = v_k.$$

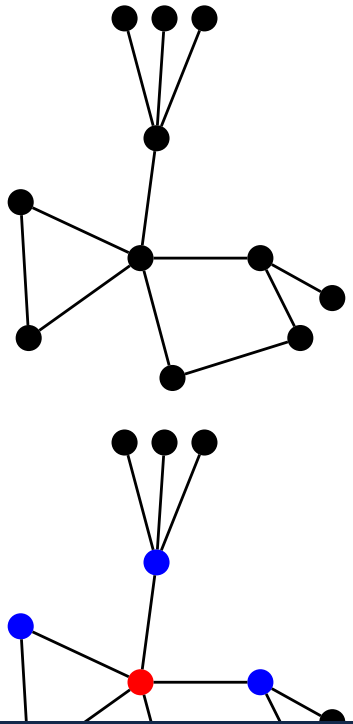
- **Directed cycle**: a directed path where the starting and ending nodes are the same.

A directed graph without any directed cycles is called acyclic.

ILLINOIS

Neighborhoods and subgraphs

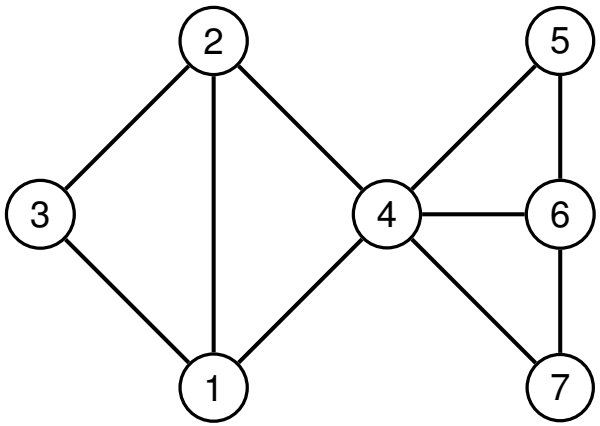
- The **degree** of a node is the number of edges the node is incident to. It can be further divided into **in-degree** and **out-degree** (for incoming and outgoing connections) for directed graphs.
- **Open neighborhood** of a node: $N(i) = \{j \in V : (i,j) \in E\}$.
- Nodes reachable within k hops from i: $N^k(i)$.
- **Closed neighborhood** of a node: $N[i] = N(i) \cup \{i\}$.
- **Induced subgraphs:** For $S \subset V$, $G[S]$ has a vertex set $V[G[S]] = S$ and an edge set $E[G[S]] = \{(i,j) \in E : i,j \in S\}$.
  - For $S \subset V$, $G[S]$ has a vertex set $V[G[S]] = S$ and an edge set $E[G[S]] = \{(i,j) \in E : i,j \in S\}$.



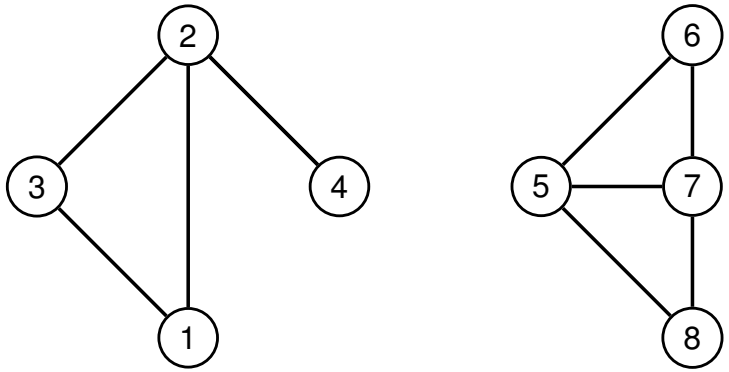
ILLINOIS

Connectivity

A graph is *connected* if there exists at least one path from every node to every other node.



G_1: **connected**
1 component



G_2: **disconnected**
2 components

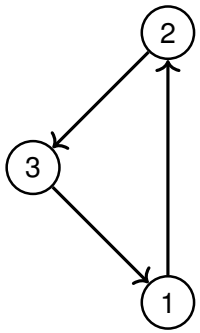
The maximal, connected subgraphs of a graph are its **components**.

- maximal: can't be made bigger by adding more nodes. For any property Π, S is maximal if S has Π but $\nexists i$ such that $S \cup \{i\}$ has Π.
- In G_2 $\{1, 2, 3, 4\}$ is maximal and connected – so it is a component.
- In G_2 $\{5, 6\}$ is not maximal – not a component.

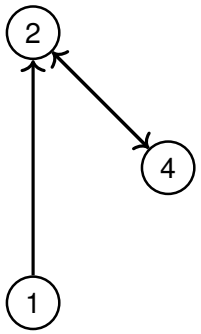
ILLINOIS

Connectivity for directed netwoks

- A directed graph is connected if there exists a path (no direction requirement) from every node to every other node.
- A directed graph is *strongly connected* if there exists a directed path from every node to every other node.



Strongly connected.



Not strongly connected.

ILLINOIS

Eulerian cycle

- *Eulerian path*: a path that includes all edges *exactly* once.
- *Eulerian cycle*: an Eulerian closed path.

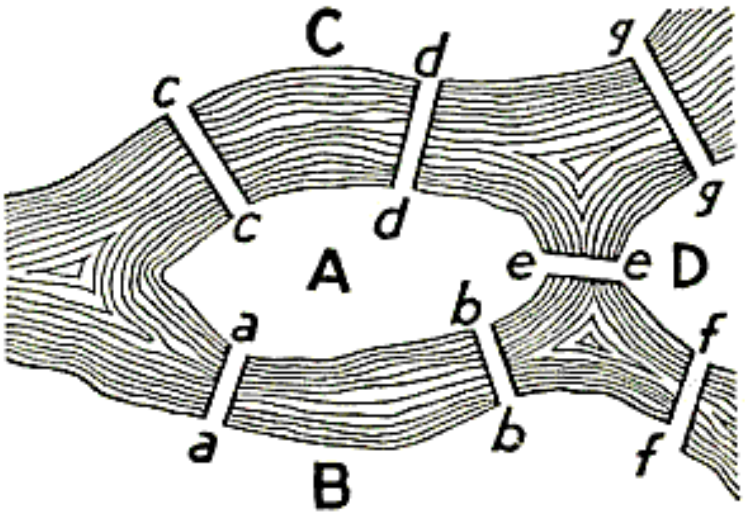
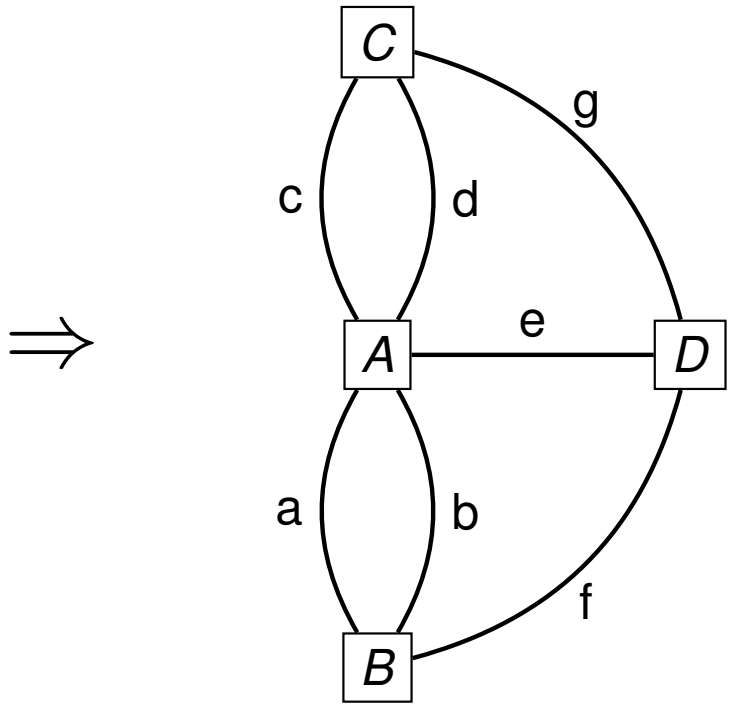


FIGURE 98. *Geographic Map: The Königsberg Bridges.*



Question: Given a graph $G(V, E)$, does there exist an Eulerian cycle?

Question: If there is one, can we find it?

We may similarly define Hamiltonian path/cycle, which use every node once.

Theorem (Eulerian cycle existence)

A graph $G(V, E)$ has an Eulerian cycle if and only if the graph is connected and every node has an even degree.

Proof:

- $\Rightarrow$: easier to show.
  - Assume G is not connected. Then no cycle that uses all edges exists (even allowing for repetitions).
  - Assume G has a node with an odd degree. That node cannot be part of a cycle.
- $\Leftarrow$: constructive proof (due to Fleury, 1883).
  1. Start at any node.
  2. Choose any edge so long as removing it does not disconnect the graph. If no such edge exists, then pick the remaining edge.
  3. Traverse that edge and remove it from the graph.
  4. When done, if there are no edges left, this is an Euler cycle; otherwise, there is no Euler cycle.

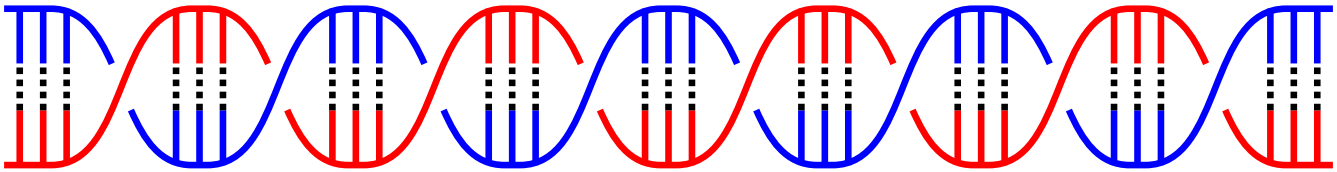
Notes:

a) Correctness? b) Compexity? c) Any *better* algorithms?

From the theorem, we also get that

- an Eulerian path exists if and only if the graph is connected and only two nodes have an odd degree.
- in a directed graph an Eulerian cycle exists if and only if the graph is connected and for every node its in-degree is equal to the out-degree.

Application: DNA sequencing



Remember that DNA consists of 4 nucleotides A, T, G, C in succession.

DNA sequencing: cutting a bigger sequence of DNA into smaller subsequences, each of length k.

$k = 4$: $ATGCCAT \implies \{ATGC, TGCC, GCCA, CCAT\}$.

$k = 3$: $ATGCCAT \implies \{ATG, TGC, GCC, CCA, CAT\}$.

$k = 2$: $ATGCCAT \implies \{AT, TG, GC, CC, CA, AT\}$.

The problem? Reconstructing the full sequence of length n from the available set of $n - k + 1$ subsequences of length k each.

ILLINOIS

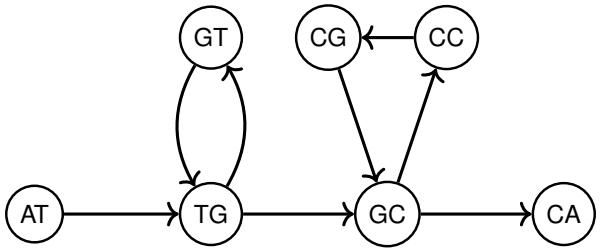
DNA sequencing and Eulerian paths

Let s be the set of subsequences of length k. We construct a graph as follows:

- nodes are all possible length $k-1$ subsequences obtained from s.
- edges between two nodes if the two of them together produce one of the subsequences in s.

The claim? An Eulerian path is a valid, possible full DNA sequence.

For example, say we have obtained the following subsequences:
$\{ATG, CCG, CGC, GCA, GCC, GTG, TGC, TGT\}$. The graph would be:



With an Eulerian path of AT→TG→GT→TG→GC→CG→CC→GC→CA, leading to a sequence of ATGTGCCGCA.

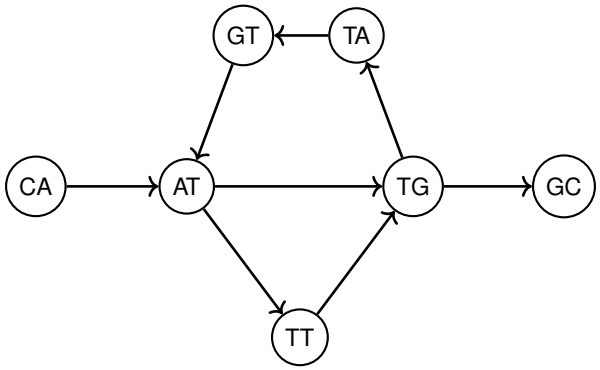
ILLINOIS

DNA sequencing and Eulerian paths

- The Eulerian path is not always unique.
- Actually, the opposite is typically the case.

For instance, consider $s = \{CAT, ATT, ATG, TAT, TTG, TGC, TGT, GTA\}$.

The graph would be:



We may now show that there are two Eulerian paths:

1. CA$\rightarrow$AT$\rightarrow$TT$\rightarrow$TG$\rightarrow$GT$\rightarrow$TA$\rightarrow$AT$\rightarrow$TG$\rightarrow$GC $\Rightarrow$ CATTGTATGC.
2. CA$\rightarrow$AT$\rightarrow$TG$\rightarrow$GT$\rightarrow$TA$\rightarrow$AT$\rightarrow$TT$\rightarrow$TG$\rightarrow$GC $\Rightarrow$ CATGTATTGC.

Hence, we have a 50% chance of getting the correct original sequence.

ILLINOIS

Overview of IE 532

This class is a nice indication of what we'll do in IE 532:

- See many interesting **network problems**.
- Discuss **algorithms** to solve them.
  - correctness, complexity, implementations.

Some topics:

- Shortest paths.
- Network flows.
- Spanning trees.
- Integer programming techniques.
- Network structures.
- Centrality.
- Communities, modularity.
- Graph spectral properties.

ILLINOIS

Next time

1. Coding for the Eulerian path problem (using `networkx`).
   - Will help us get accustomed to network problems and using Python to solve them.
2. Overview of optimization and modeling (with examples).
   - Will take us to at least Lecture 3.

3. Installing necessary material, including:
   - `Gurobi`.
   - `networkx`.
   - `Jupyter`.

ILLINOIS

Before you head out

These are the *Kaliningrad* bridges now.

ILLINOIS