Graphs, Networks, and Algorithms

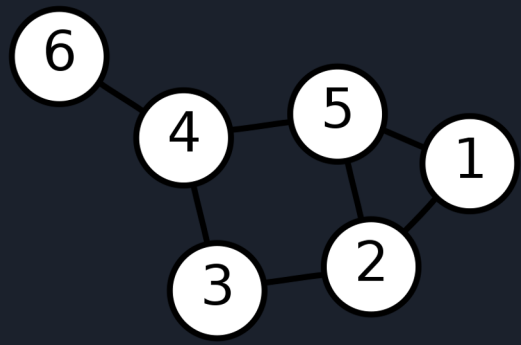
Arindam Pal (arindam.pal@data61.csiro.au)
Data61, CSIRO and Cyber Security CRC
Co-learnium, CSIRO Marsfield
24th September, 2020

Today's Agenda

- What are graphs and networks?
- What is an algorithm?
- How graphs can be used to represent relationships?
- Why graphs are so ubiquitous in mathematical modelling?
- Some applications of graphs and networks in real-world problems.
    - Intelligent transportation networks: Computing shortest and fastest paths and avoiding congested routes in Google Maps.
    - Finding relevant search results in the web graph: Link analysis techniques like Google PageRank and HITS.
    - Social network analysis: Detecting communities, suggesting new friends.
    - Knowledge representation and reasoning: Knowledge graphs and their applications to question answering.
- Conclusion and future directions.

What are graphs and networks?

- Graphs are mathematical structures used to model relationships between objects.
- Formally, a graph $G = (V, E)$ has a set of vertices V and a set of edges E.
- People, places, web pages are represented by vertices.
- Friendships, roads, links are represented by edges.
- Graphs illustrate the power of abstraction and ideas.



What is an algorithm?

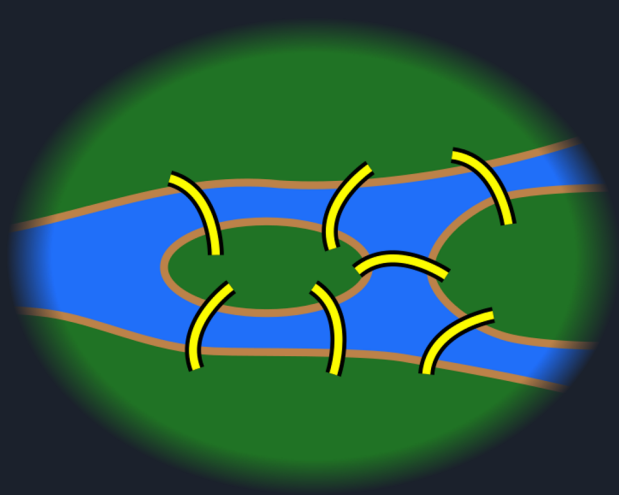
- An ***algorithm*** is a finite sequence of well-defined, programmable instructions, typically to solve a class of problems or to perform a computation.
- The word *'algorithm'* has its roots in Latinizing the name of the mathematician *Muhammad ibn Musa al-Khwarizmi* to *algorismus*.
- Euclid's algorithm to compute the *greatest common divisor* (GCD) of two numbers is the first algorithm designed by a human.
- It appeared as Proposition II in Book VII ("Elementary Number Theory") of his seminal treatise ***Elements***.
- Algorithms are intimately related to *models of computation*, such as the ***Turing Machine***.

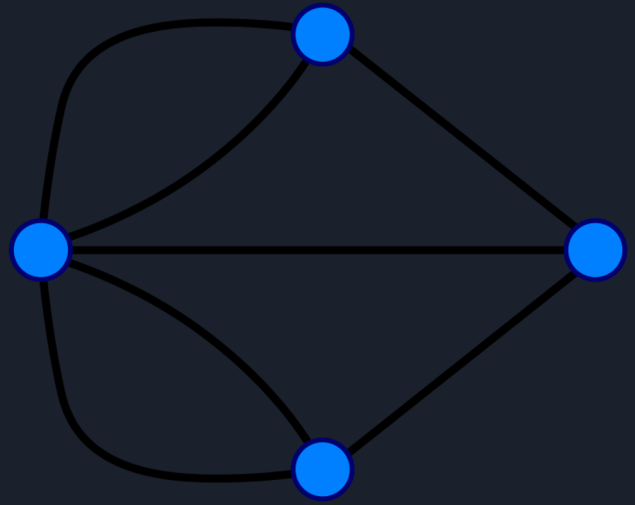
Seven Bridges of Königsberg

- The city of Königsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River.
- It included two large islands Kneiphof and Lomse, which were connected to each other, or to the two mainland portions of the city, by seven bridges.
- The problem was to devise a walk through the city that would cross each of those bridges once and only once.

Map and transformation to graph

7-bridges graph





Euler's solution (that there were none)

- Leonhard Euler proved that the problem has *no solution*.
- If a graph has such a path, it's called an *Eulerian path/tour*.
- For an Eulerian tour to exist, each vertex should have *even degree*.
- This is a *necessary and sufficient condition* (later proved by Carl Hierholzer).
- For the 7-bridges graph, the degrees are 5, 3, 3, 3 respectively.
- So, it can't have an Eulerian tour.
- There are efficient algorithms to construct Eulerian tours.
- No wonder that Euler is considered as one of the greatest mathematicians of all times (along with Gauss and Erdos).

Why graphs are so ubiquitous in mathematical modelling?

- Graphs provide a very powerful tool to represent relational datasets.
- It can be used to go beyond linear data such as text.
- Graphs can be used to model complex problems such as scheduling, routing, social network analysis, planning, game playing, modelling uncertainty and much more.
- Graphs are essential tools in *combinatorial optimization* and *mathematical programming*.
- Powerful methods from *linear algebra* can be used on their adjacency and Laplacian matrices.
- Algorithms and graphs go hand in hand with each other.

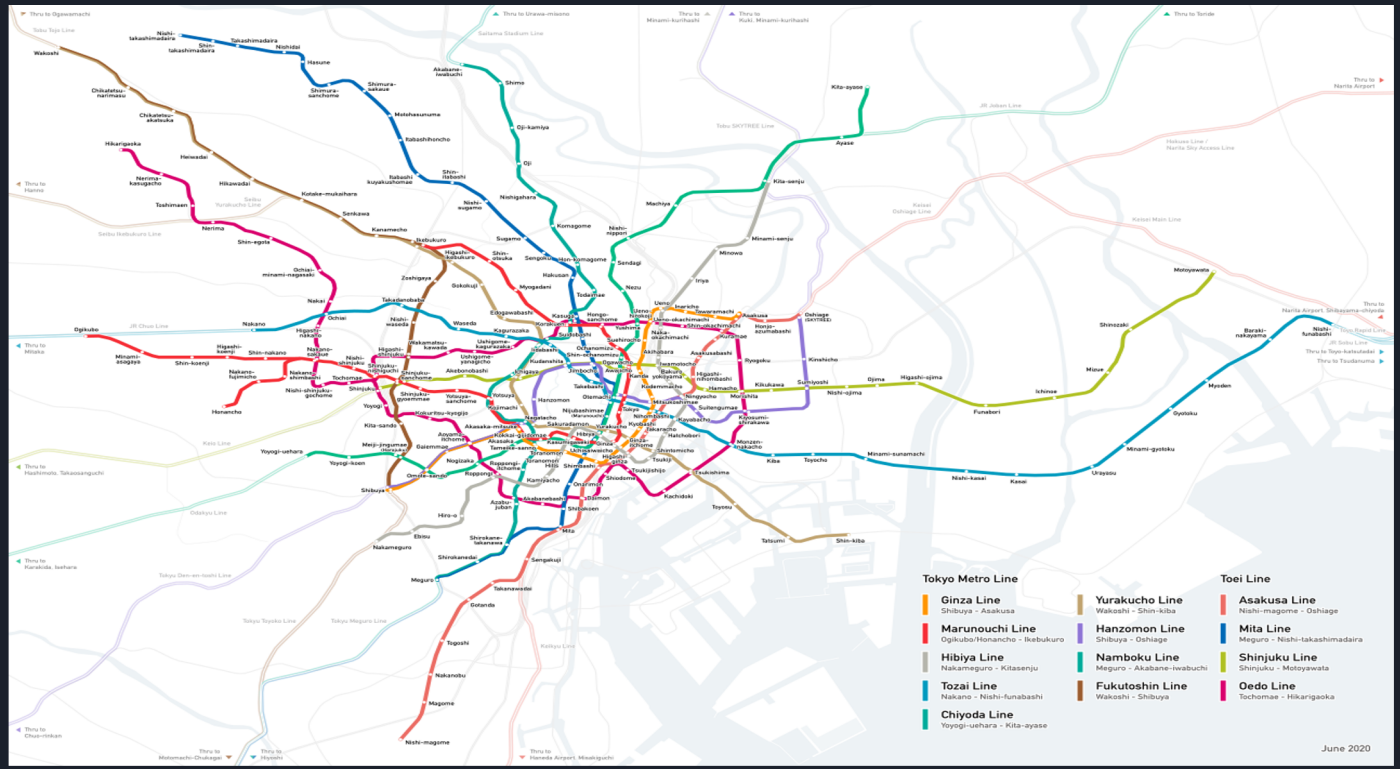
Intelligent transportation networks

- Transportation networks give an ideal playground for applying graph algorithms.
- These networks consist of places, roads, stations, train and bus tracks.
- There are various objectives that an user may have.
- S/he may want to minimize the total distance, time, or the ticket fare.
- In terms of the underlying graph, we are interested to find the shortest, fastest or cheapest paths.
- There are other objective functions such as prediction of the actual travel time, based on historical data and congestion on the roads.

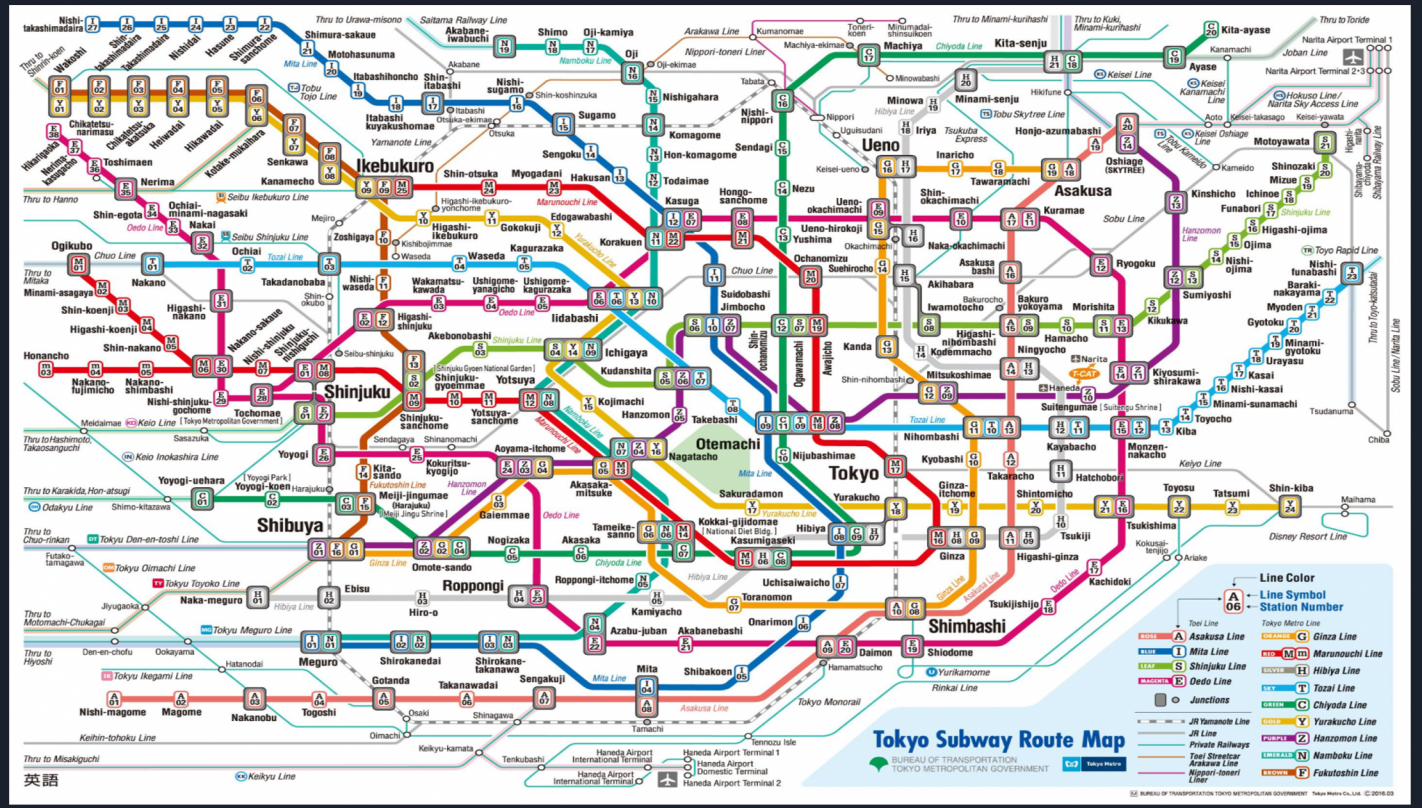
The Tokyo subway

- Greater Tokyo has the most extensive urban railway network and the most used in the world, with 40 million passengers in the metro area daily.
- Tokyo subway is a part of the extensive rapid transit system in the Greater Tokyo area of Japan.
- There are two primary subway operators in Tokyo:
  - Tokyo Metro
  - Toei Subway
- Tokyo Metro operates 179 stations on 9 lines and 195.1 kilometers of route.
- Toei Subway operates 99 stations on 4 lines and 109.0 kilometers of route.
- There are many other lines, including the famous *JR Yamanote Line* connecting most of Tokyo's major stations and urban centres.
- It’s a very sophisticated and complex transportation network.

Tokyo subway map



Tokyo Metro and Toei Subway



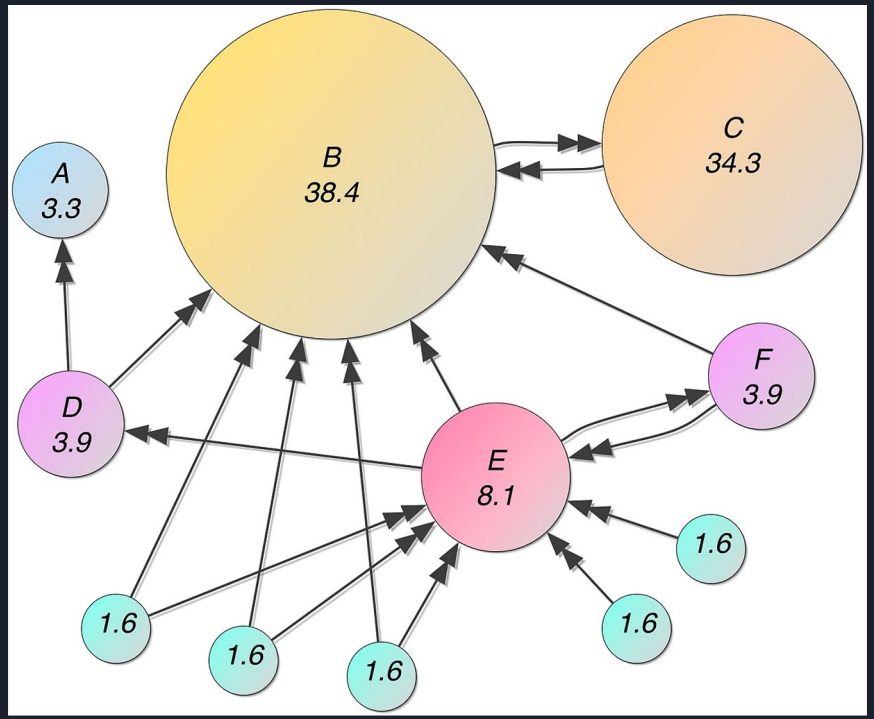
Finding relevant search results in the web graph

- A web user wants to find relevant web pages in response to a search query.
- For example, if someone searches for “electric cars” s/he should be shown results about electric cars and their manufacturers, such as Tesla.
- A search engine is the component which searches through billions of web pages and returns the relevant web pages.
- There are two important tasks here.
  - *Find* all the web pages related to the search query.
  - *Rank* the web pages in descending order of relevance (importance).
- This a highly challenging and non-trivial problem.
- Analysis of the HTML code in a web page is not good enough.

HITS and PageRank

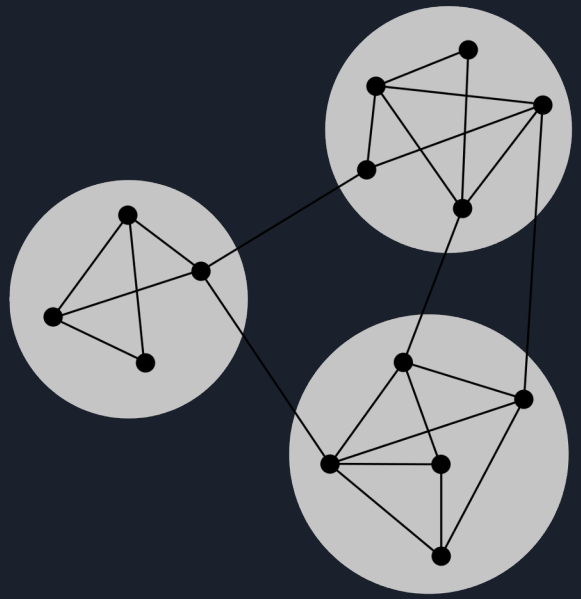
- Two elegant and independently designed algorithms were proposed almost at the same time (1996).
- Jon Kleinberg came up with the ***HITS*** algorithm at IBM Research, which stands for Hyperlink-Induced Topic Search.
- Larry Page and Sergey Brin developed the ***PageRank*** algorithm at Stanford University.
- HITS uses the notions of *hubs* and *authorities*.
- PageRank uses the model of a web surfer doing a *random walk* (*Markov chain*).
- Both of them use an iterative algorithm to repeatedly improve the scores.
- In both cases, the iterative algorithm converges.

HITS and PageRank



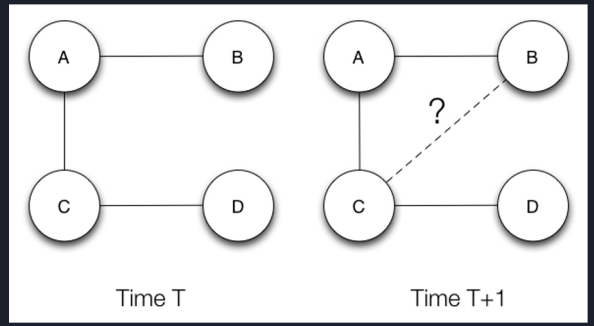
Social network analysis

- In a social network, users are the vertices and their friendships/followerships are the edges.
- There are many important insights that can be obtained from the social network.
- *Community structure* is an important concept.
- It is the occurrence of a groups of nodes in a network that are more densely connected internally than with the rest of the network.
- There are many algorithms for finding community structures.



Link prediction

- Another important problem is *suggesting new friends*.
- This is also known as *link prediction*.
- There are many algorithms for doing link prediction.
- A simple method is to compute the number of common neighbors.
- This is based on the idea that two people with many common friends are also likely to be friends.



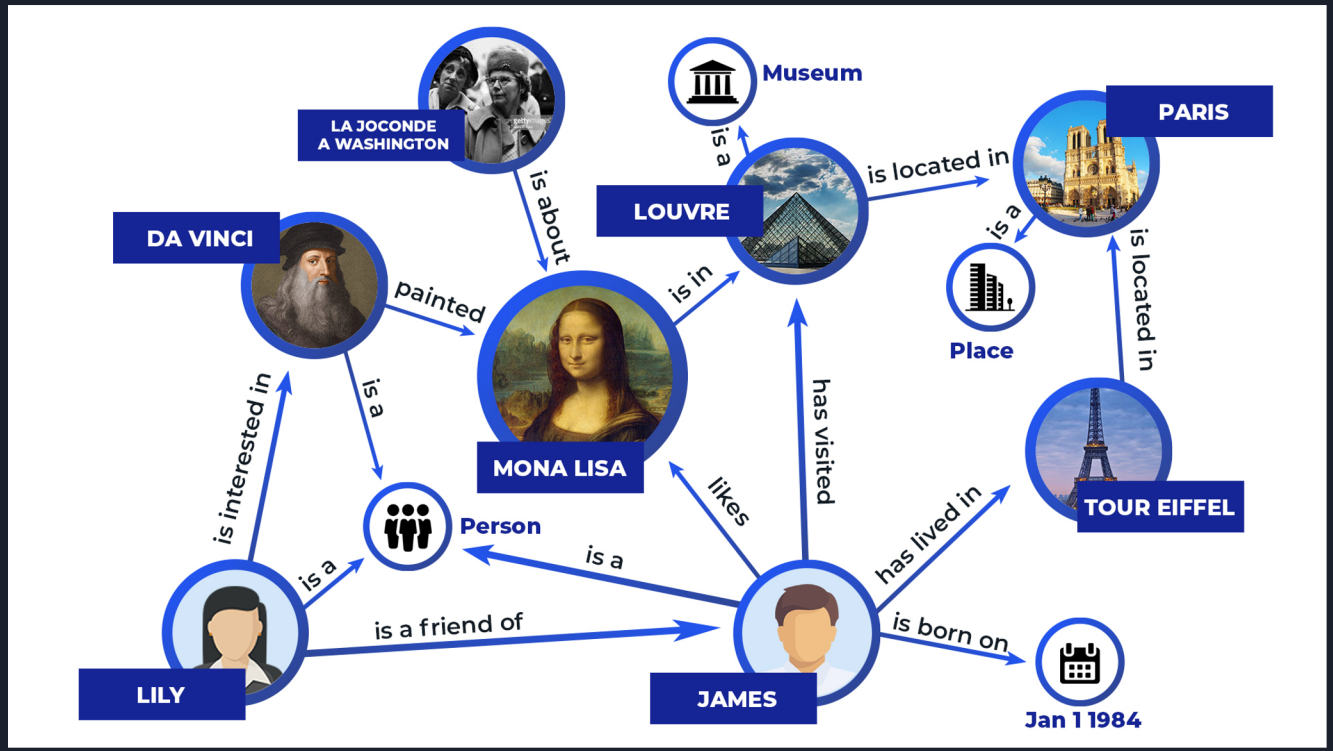
Knowledge representation and reasoning

- How can we represent knowledge in a computer, so that a computer can solve complex tasks such as diagnosing a medical condition or having a dialog in a natural language.
- The algorithm should be capable of doing reasoning, even under uncertain situations.
- Mathematical logic (propositional and first-order) tries to do that.
- A *knowledge graph* is a knowledge base that uses a graph-structured data model to integrate data.
- For reasoning under uncertainty, *probabilistic graphical models* and *Bayesian networks* are very useful.

The Google Knowledge Graph

- The Google Knowledge Graph (GKG) is a knowledge base used by Google to enhance its search results with information gathered from a variety of sources.
- The information is presented to users in an infobox next to the search results.
- In May 2020, GKG has 5 billion entities and 500 billion facts about them.
- Constructing a knowledge graph from an unstructured text corpus is a challenging problem.

Question answering using knowledge graphs



Conclusion and future directions

- Artificial Neural Networks and Deep Learning.
- Academic citation networks.
- Playing games such as *Chess* and *Go* using the *MiniMax* algorithm and *Alpha–Beta pruning*.
- Planning and search algorithms, such as A^*.
- Biological networks like protein–protein interaction networks.
- Propagation of viruses (like COVID-19) in human and computer networks.
- Interdependent networks such as smart power grids.
- Probabilistic graphical models and Bayesian networks.

Questions?