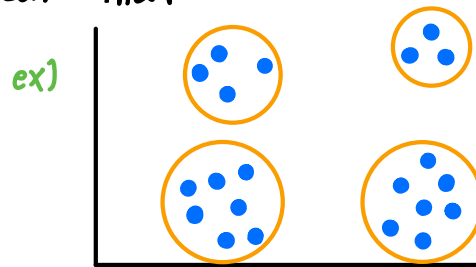A few last notes from yesterday: BFS works for finding shortest paths for unweighted graphs. For weighted graphs, you could us Dijkstra's algorithm.

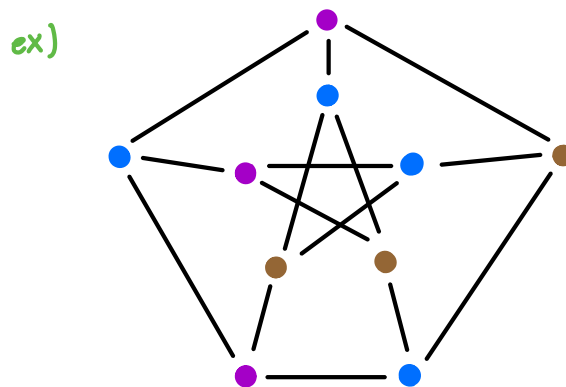possible project idea: Find the "efficiency" of a network through looking at shortest paths.

Other ways to analyze graphs and other graph problems:

1) Clustering methods: techniques that group datapoints. Typically you implement clustering on a bunch of datapoints (AKA nodes without edges between them

ex)

Applications: • determining precision of datapoints
• human genetic clustering
• grouping online shopping items

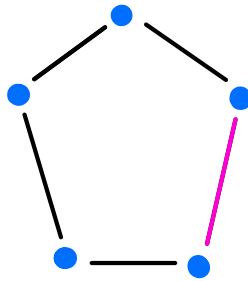2) Graph coloring: color the nodes of a graph so that no two adjacent nodes are the same color

ex)

Applications: • map coloring
• sudoku
• universities scheduling exams

Chapter in the book on the website that covers this problem...
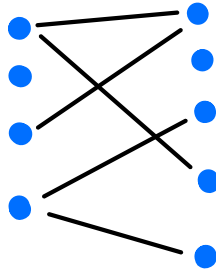
3) Maximum Matching

A **matching** in a graph is an edge set such that each node belongs to at most 1 edge in the matching

ex)

A maximum matching means to maximize how many edges are in the matching.

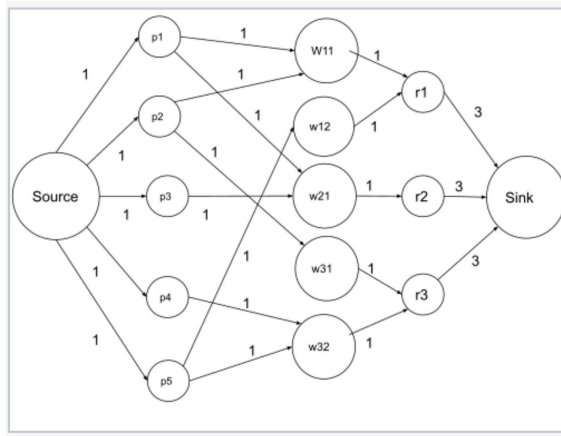Applications: • pairing kidney donors to kidney donor recipients
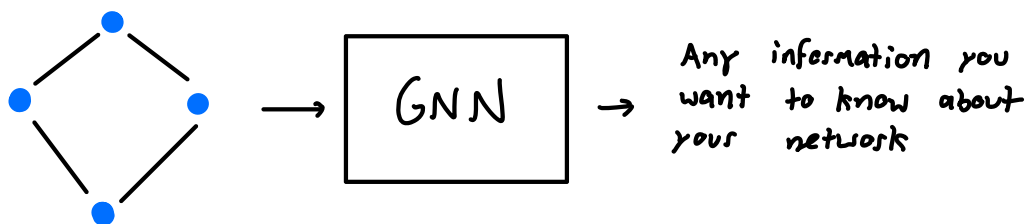

kidney donors   kidney donor rec.

• matching candidates to jobs

4) Maximum flow: finding a flow through a network that obtains a maximum possible flow rate.

Applications: • scheduling flight crews
• transporting people in cars/flights
• baseball eliminations: predict which teams have no chance of making the playoffs

5) Graph Neural Networks: A powerful tool to analyze anything about networks.

Any information you want to know about your network

Applications:
- all of the above

Networks in the Wild

For basically any network you can think of, there's a good chance that a dataset with that data exists already. If not, you might be able to create it on your own.

In a quick search I found:
- Reddit dataset
- dataset of proteins/amino acids
- IMDB dataset: actors and actresses
- AIDS dataset
- Amazon products dataset
- Beer Review dataset

You can also create your own by "scraping" the internet. We'll go over this next Wednesday.

Project proposal due next Friday:
- Need to work in groups of 2-+ (no solo groups)
- proposal should be about 1 page long (between 1/2 page to 2 pages).
- Work on something that has piqued your interest from the class
- The proposal should cover
  1) What problem you are solving or task you are trying to complete
     ex) What are the most important animals in a food web?
     ex) Cluster politicians based on factors x, y, z.
  2) How you plan to solve the problem or task (i.e. what methods).
     ex) degree centrality
     ex) Hierarchical Agglomerative Clustering
  3) What data you need and where you can find that data
     ex) Publically available www.foodweb.com
     ex) Scrape the internet for info about politicians

Let's do an activity to start thinking about project ideas. See the activity worksheet for today.