



Introduction to Web Science

Tutorial (Assignment 9)

Olga Zagovora



Exercise 1

Name the 6 potential parts you could find in research paper abstracts

1. State the Background and Problem you tackle with your research.
2. Name the methodology you have used.
3. Formulate 1 to 3 precise research question that are answered in your paper.
4. Talk about your unique solution or idea.
5. Demonstrate the results.
6. Conclude with a point of impact.

Exercise 1

Mark all parts you can find in the given abstract.

Hit songs, books, and movies are many times more successful than average, suggesting that “the best” alternatives are qualitatively different from “the rest”; yet experts routinely fail to predict which products will succeed. We investigated this paradox experimentally, by creating an artificial “music market” in which 14,341 participants downloaded previously unknown songs either with or without knowledge of previous participants’ choices. Increasing the strength of social influence increased both inequality and unpredictability of success. Success was also only partly determined by quality: The best songs rarely did poorly, and the worst rarely did well, but any other result was possible.

Exercise 2

List five model choices that stay in conflict with reality
tango

- **frozen network of agents:** The conflict here lies in the assumption, that the number of users stays fixed. But in the real world users sign up and opt out of services like twitter.
- **each about a specific meme:** The agents only see and post memes, where in the real world a diverse set of contents is produced and consumed.
- **allowing posts to survive in an agent's list or memory only for a finite amount of time:** Memory may be recovered by certain trigger events and is not necessarily lost forever.
- **If people use the system once weekly on average, the time unit corresponds to a week.:** The time unit for each agent is fixed. But usage may fluctuate based on vacation etc, which may lead to the rise of memes spread in times of a high occupancy rate of the system.
- **Agents interact on a directed social network of friends/followers.** The Agents have no capability to follow or *unfollow* other agents based on their interest. So the model's edges are static and not in a constant flux like in the real world.

Exercise 2

List five model choices that stay in conflict with reality

Quebec

1. "An agent maintains a time-ordered list of posts, each about a specific meme."
- **One post can contain multiple memes. Also, users see posts ordered by evaluated relevance to them, with the most relevant posts displayed first**
2. "Asynchronously and with uniform probability, each agent can generate a post about a new meme or forward some of the posts from the list" - **People make and share posts with memes in their free time, so long periods of inactivity could be expected (sleep, work), followed by periods in which many posts are created and shared (free time)**
3. "Users pay attention to these memes only." - **usually other content which is competing for the users attention is also on the screen (pictures, videos, advertisements)**
4. "we model limited attention by allowing posts to survive in an agent's list or memory only for a finite amount of time" - **people don't forget the funniest memes as quickly as the ones they found less fun and interesting**
5. "Our basic model assumes a frozen network of agents" - **this means that users neither get new followers, nor expand the list of users they follow. In the real world the network would change very often**

Exercise 3

Write a function that returns the diameter of the given directed network. The diameter of a graph is the longest shortest path in the graph.

What was the diameter of the network?

Why?

What one can measure?

Exercise 3

Main parts:

- Breadth-first search
- Check whether network is connected
- Eccentricity of node
- Diameter

Demo -> ipython notebook

Exercise 3

What is the difference between tree and graph?

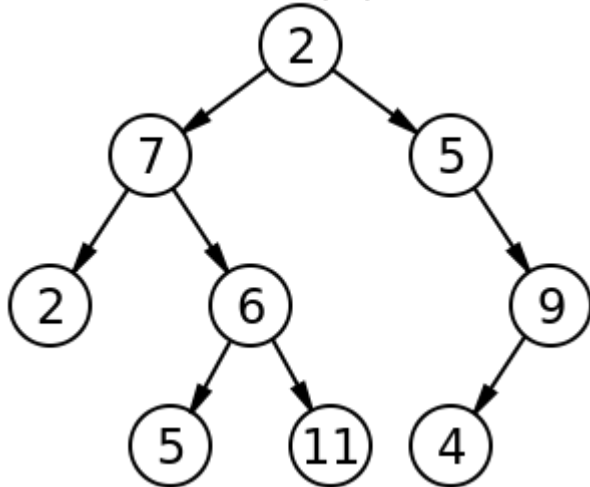
Graph is a set of vertices and edges

Tree is a collection of nodes (starting at a root node) together with a list of references to nodes (the "children"), with the constraints that no reference is duplicated, and none points to the root.

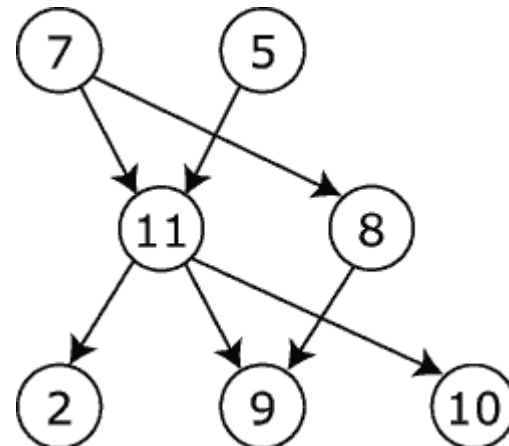
Wikipedia/WWW is a tree or graph?

Graph

Tree

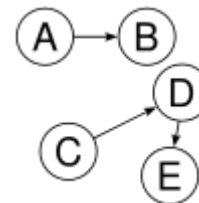
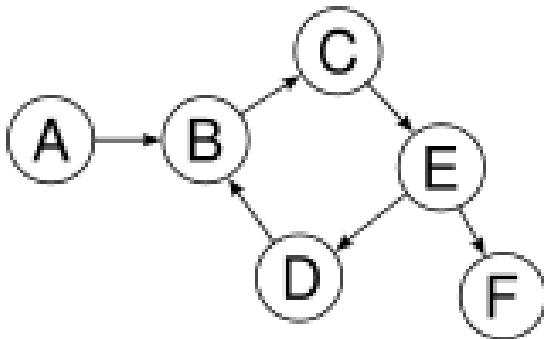
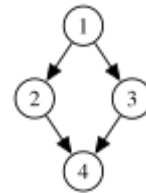


Graph



Exercise 2

Tree or Graph?



Questions?



zagovora@uni-koblenz.de