Data Mining, Sheet 2

In this exercise sheet we will visualize simulation data for a simple randomized rumor spreading protocol. We are given a system of n agents. Initially, a designated agent has a piece of information, the rumor (it is "informed"). The goal is to spread this information to all other agents (which are initially "uninformed").

The random-phone-call model is simple: it assumes that the system runs in perfectly synchronized rounds. In each round, every agent is activated. All activations are performed atomically and in parallel. Whenever an agent u is activated it opens a connection to a randomly chosen agent v.

There are three simple protocols for information dissemination in this model: push, pull, and pushpull. Assume agent u is activated and connects to agent v.

- push: if u is informed and v is uninformed, v becomes informed as well.

- pull: if u is uninformed and v is informed, u becomes informed as well.

- pushpull: a combination of push and pull.

Exercise 1

Implement a simple simulation of these rumor spreading processes.

Exercise 2

Run a simulation campaign for push, pull, and pushpull. Identify interesting properties of the process and generate empirical data for varying numbers of agents n.

- Hint: Simple rumor spreading processes are used, e.g., by Amazon to synchronize replicated data stores. What might be interesting to them?

- Hint: Think a moment about meaningful values of n to simulate. You are not restricted to a linear scale.

- Hint: It would be interesting to see data for varying numbers of n as well as for the behavior of the process for one specific number of agents n. Can you simulate 10^9 agents in a reasonable amount of time?

Exercise 3

Visualize the data. You are free to choose any meaningful visualization method that you have learned of. Be creative!

Exercise 4

Write a simple LaTeX document that includes your most interesting visualizations. Your document should be "production ready". It should contain a short explanation of the process (feel free to copy-paste it from this exercise sheet) and a quick description of your simulation. Your plots should be placed in figure environments, which should be referenced in your text.

Exercise 5

Reproducibility is a defining feature of science. Prepare a repository for your simulations with all necessary information such that your results can be easily reproduced by your peers.