Data Mining, Sheet 3

In this exercise sheet we will visualize simulation data for a simple randomized consensus protocol. We are given a system of n agents. Initially, each agent has an *opinion*, either A or B. The goal is that all agents agree on the opinion which is initially supported by most agents.

We assume that agents update their states in discrete time steps according to a simple probabilistic interaction model. In each time step an agent u is selected uniformly at random. The selected agent u observes the opinions of other agents and then updates its state according to simple transition rules. In this model, there are many ways how to solve the consensus problem.

Two-Choices: The selected agent u observes the opinions of two other agents selected uniformly at random. If their opinions coincide, agent u adopts it. Otherwise it retains its original opinion.

3-Majority: The selected agent u observes the opinions of three other agents selected uniformly at random. It then adopts the majoriy among the sample, breaking ties randomly.

Undecided State Dynamics: The selected agent u observes the opinion of one other agent selected uniformly at random. If agent u encounters a different opinion than its own, agent u becomes "undecided". Undecided agents adopt any (decided) opinion they encounter.

Exercise 1

Implement a simple simulation of these consensus processes.

Exercise 2

Run a simulation campaign for Two-Choices, 3-Majority, and the Undecided State Dynamics. Identify interesting properties of the process and generate empirical data for varying numbers of agents n.

- Hint: Simple consensus 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.