Early Reddit Comments

Carel Kuusk

Motivation

- Reddit is a community aggregator
- Got its start in 2005
- A peek into early evolution of social networks

Dataset

- Reddit comments dataset by u/Stuck_In_the_Matrix
- Scraped via Reddit's public API
- Due to massive total size, only first 2 years

Methodology

- Torrenting the dataset
- Cleaning data
  - Uniform identifiers
  - Remove deleted authors
- Creating a network of co-commentators
- Look at standard metrics of the initial social graph
- Comparing the overlap of different subreddits
- Since data is timeordered, can investigate time-evolution

Results

Thank you for your attention!