CS69000-DM1 (CRN 17239)
Topics in Data Mining

Bruno Ribeiro
Graduate Seminar Course?

- Training assembly line workers:
  - Breadth not depth
  - Need to learn everything
  - Specialization undesirable

- Training knowledge workers:
  - Work as apprentices
  - Focus on tasks
  - Study state-of-art

Ph.D. Seminar
Seminar Goals & Organization

- **Goal: Improve state-of-art**
- You are responsible for course success
- Ph.D.-level course → High expectations

- **Course load (tasks)**
  - Present one paper (at least)
    - Goal: Understand & explain & go beyond
    - I can present some topics
      - Forecasting Complex Processes
      - Experiment Design
  - Submit reviews of papers ONE week before presentation
    - Goal 1: Did you understand the paper?
    - Goal 2: Help next week’s presenters with other viewpoints
    - Review submission: Friday 5pm of week before presentations
  - Projects
    - Goal: Explore research boundaries
Grading

- Paper reviews/Class participation: 30%
  - Grading 2 ways:
    - I grade your reviews
    - Presenters send me subset of least helpful reviews
- Paper presentation: 10%
  - Presentation quality + future work
- Midterm project milestone: 20%
  - Quality of preliminary project
  - How peers evaluate your project (- strategic selection bias)
- Final project 40%
  - Quality of joint project
Schedule

- Aug 25 - Organization
- Aug 27 –
- Sept 1 –
- Sept 3 –
- Sept 8 –
- Sept 10 – Class to discuss project options
- Sept 15 – *(Deadline to choose projects)*
- Sept 17 –
- Sept 22 –
- Sept 24 –
- Sept 29 –
- Oct 1 –
- Oct 6 –
- Oct 8 –
- Oct 15 –
- Oct 20 –
- Oct 22 –
- Oct 27 – Project Milestone Presentations
- Oct 29 – Project Milestone Presentations
- Nov 3 –
- Nov 5 –
- Nov 10 –
- Nov 12 –
- Nov 17 –
- Nov 19 –
- Nov 24 –
- Dec 1 –
- Dec 3 –
- Dec 8 – Final Project Presentations
- Dec 10 – Final Project Presentations
- Dec 12 – FINAL report deadline
Project Dynamics

- Choose a project EARLY
- Projects are individual until Milestone
- Pair students after Milestone (one 3-tuple if odd no. students)
- Merge paired projects

Project pairing algorithm:

- You vote on other projects of interest to you during student presentations
  - Your scores have low entropy? → I will drop highest scores
    (incentive against collusion)
- Probabilistic matching algorithm does the pairings
  (Skinkhorn algorithm)
- Hint: Convince peers with good projects that your project is also awesome
Let’s get to work

- Aug 25 - Organization
- Aug 27 – Experiment Design (Bruno)
- Sept 1 – Obtaining the Data
- Sept 3 – Obtaining the Data
- Sept 8 –
- Sept 10 – Class to discuss project options
- Sept 15 - (Deadline to choose projects)
- Sept 17 –
- Sept 22 –
- Sept 24 –
- Sept 29 –
- Oct 1 –
- Oct 6 –
- Oct 8 –
- Oct 15 –
- Oct 20 –
- Oct 22 –
- Oct 27 – *Project Milestone Presentations*
- Oct 29 – *Project Milestone Presentations*
- Nov 3 –
- Nov 5 –
- Nov 10 –
- Nov 12 –
- Nov 17 –
- Nov 19 –
- Nov 24 –
- Dec 1 –
- Dec 3 –
- Dec 8 – *Final Project Presentations*
- Dec 10 – *Final Project Presentations*
- Dec 12 – *FINAL report deadline*

- Mining & Predicting User Trajectories (2 days)
  - 1
  - 2
- Network A/B Testing (2 days)
  - 1
  - 2
- Data & Graph Streaming (Data Mining)
- Experiment Design
  - Bruno
- Link Prediction & Network Routing
- Factorization & Network Evolution
- Network Utility & Node Matching
- Forecasting Complex Processes
  - Bruno
- Popularity Forecast
- Network Communities
- Obtaining the Data (2 days)
  - 1
  - 2
- Social Choice? (time permitting)