Task 1
Team Formation

Computational Social Science Course
JProf. Dr. Claudia Wagner


**Topic 1: Travel**

- **Flickr Travel Pics**
  - Team: Aaron Kohn, Felix Engelmann, Elias Zervudakis, Bibi Nazneen

  - Do users from rich countries travel longer distances?
    - Has flick a bias towards users from rich countries? \( \rightarrow \) is your sample biased?
    - Are users from rich countries more active?
    - Do users from rich countries post more travel pics?
Topic 2: Soccer

- Male and Female Soccer
  - Team: Carina Blüm, Natalie Lang, Philip Kirsch, Mariya Chkalova

- Temporal analysis of gender-differences

- Popularity of male and female soccer players in Google, Wikipedia and News Media (e.g. Zeit)

- Compare textual presentation (sentiment based on dictionaries, tfidf-vectors of male/female players)
Topic 4: Media-Politicians

- German politicians in media
  - Team: Bastian Bernst, Jonas Englich, Ahmad Alsamman, Gerwin Rajkowski
  - Are male and female politicians equally frequently covered by news media (e.g. Zeit)?
  - What is the relationship between media coverage, search interest and gender?
  - How are they presented (textual analysis)

- Make sure that groups are comparable!
  - how can we select equally important politicians?
    - E.g. what functions did politicians have in the past?
Topic 5: Politicians on Wikipedia (~4 people)

Team: Vladyslav Vorontsov, Marco Ehl, Esha Agrawal, Matthias Deisen

Which German politicians are captured on Wikipedia? Does search interest predict existence on Wikipedia?

- Create list of all German politicians between XX and XY
- Analyse search volume of politicians
  - Plot distribution of number of countries from which search volume happens for male and female politicians
  - Plot number of month during which search volume is above threshold for male and female politicians
- Binary logistic regression
  - Outcome variable: article exists on Wikipedia
  - IV: search volume $\rightarrow$ number of countries and month
  - Control: experience (e.g. how often was a politician already part of parliament)
Team: Shiau Chu Heng, Shide adibi Md, Kamal Hossain, Chuyi Sun

To what extent and how do textual online self-presentation of male and female scientists differ?

- Data: collect a sample of conferences and scrape speaker lists (they often contain bio and pictures of invited speakers). E.g. lists like this one can function as seed list https://en.wikipedia.org/wiki/List_of_computer_science_conferences

- Methods: Analyze word vectors for male and female scientists; tfidf
  - Quantify gender difference: cosine-difference between pairs within same group (men-men, women-women) and across groups (men-women)
Team: Slobodan Kocevski, Md Shohel Ahamad, Jabid Ishtiaque, Chiranth Manjunath

To what extent and how do the online self-presentation of male and female scientists differ with respect to pictures?

- Data: collect a sample of conferences and scrape speaker lists (they often contain bio and pictures of invited speakers). E.g. lists like this one can function as seed list https://en.wikipedia.org/wiki/List_of_computer_science_conferences
- Methods: Use automated attractiveness scores, compare them with human ratings
- Plot distribution of attractiveness of male and female speakers; quantify difference between distributions
Topic 8: H-index / Conf Speakers (~5 People)

- Team: Denis Oldenburg, Julian Rogawski, Tara Morovatdar, Orkut Karacalik

- To what extent do h-indices of male and female speakers that were invited to conferences in the same sub-field differ?
  - Data: collect a sample of conferences and scrape speaker lists (they often contain bio and pictures of invited speakers). E.g. lists like this one can function as seed list
    https://en.wikipedia.org/wiki/List_of_computer_science_conferences
  - Extract speakers from recent conferences and collect h-index (or citation counts)
  - Methods: compare h-index distribution of male and female speakers; e.g. KL divergence can quantify difference in distribution
Topic 9: Refugees/Foreigner in Media

- Team: Alexander Schneider, Simon Schauß, Lukas Härtel

- Use list of nationalities to search on Zeit-API or Google n-grams

- How has media presentation/coverage of different nationalities changed over time?
  - Ngrams that contain „syrian“, „iranian“ and so on
  - Words that are highly correlated with „Ausländer“ („foreigner“)
  - Sentiment-words that are correlated with „Ausländer“ (foreigner)

- Foreign criminals versus german criminals?
**Topic 10: Happiness**

- Team: Stefan Strüder, Stefan Strüder, Peter Heuzeroth, Anna Shumylo, Arsenii Smyrnov

- Dataset: sample of geo-tagged tweets
- Method: sentiment analysis
  - Manually assess method (face validity)

- RQ1: which state is most happy according to positive sentiment exposed on Twitter?

- RQ2: How stable is the happiness ranking over time?
Team: Nađa Jeličić, Dhurim Sylejmani, Syed Nabil Afaraz Bukhari

Dataset: MIT and Harvard online courses

RQ: Which factors are related with student‘s success?

Method: Regression model or Matching Method
- DV: grade
- IV: gender, education level, ...
- Control: course
Timeline

- 9.5. Team Formation
- 6.6. Status Report from all teams
- 11.7. and 18.7. Final Presentations
- 25.7. Notebook submission & Presentation
  - NEWS: no final report!
  - Please document and describe your project on git!
What to submit?

- Presentation
  - Phrase a question that you can answer!
  - Describe how you answer the question: data and method
  - Focus on the main findings, describe and interpret them
  - Related Work

- Data (compressed files)
- Git repository with binder
  - Should contain description of the project
  - Python notebooks
  - Explanations of how to run code (if necessary)

- Structure and document your code!
Evaluation

- 50% Exam (25.7.)
- 50% Research Project (Exercise)
  - Task 1 (20%): individual → propose research project: come up with good questions, datasets and explain methods that you want to use to answer questions
  - Task 2 (30%): form teams and work on a research project together; write a small report about the project and present it at the end of the semester
    - Use python and create notebooks
    - Report should only contain 2-4 figures/tables that clearly answer your research question
    - Connect your work with existing research (related work)
Any further questions?

See you next week