DissertationStructure

Contents

Abstract

- What I did:
  - Investigation of the subnetworks based on citee
  - Comparing the differences in structure.
  - Showing that they are structurally disperate.
    - No correlation
    - Temporal changes in order occur in one network not the other.

Background

Literature review

- Intro to scale free networks
  - Expand on this section further
  - More focus on the history / core texts
    - Price
    - Barbarasi
    - Newman
- Citation networks / Graph theory
  - Include more primary research on academic citation networks.
  - Contrast between patent and citation networks mainly already present in this section from lit review.
- Innovation Evolution
  - Talking about the study of patent systems as innovation / evolutionary systems and other unique ways of looking at it, not necessarily as a citation network.
- Patent citation networks (drilled down)
  - This is the research closest to what I'm doing.
  - Talk about valverde paper here and the other similar papers.
- Conclusion

Summary of related work

- Valverde and similar studies
- Debate between log-normal and power-law distributions
  - Difficulty in proving
  - Sectors where this has been proven one way over the other
  - Debate within citation networks
  - Debate within Patent networks

Research

Pipeline

- Raw data
  - Different schema, xml, txt, sgml (different versions within that)
  - Different variables being captured.
  - Size and scope of the data.
- Parsing
  - Schema of parsed data.
  - Methodology / flowchart for parsing function
  - How different schema caused problems / dictated how parsing function had to work.
- Cleaning
  - Different things checked for / cleaned.
  - Things not cleaned / why
- Mongodb / processing
  - Why use database
    - Big Data, processing required grouping and processing data in ways it was not already ordered.
  - Why choose mongodb
    - MapReduce, designed for BigData, javascript backend language makes ease of use.

- (r used for final analysis)

Reproduce Valverde

- Number of patents vs. time
  - Cumulative number of patents vs time.
  - Fitting the distribution (power-law vs exponential)
- Order of citations vs. time
  - Fitting the distribution (pl, exp, pois, log-normal)

Comparing Cited by Examiner to Cited by Other

- Splitting the data into two (Cited by Examiner and Cited by Other)
- Repeat the above for the two seperate datasets
  - Patents vs. time
    - Cumulative
    - Fitting distribution
  - Order of citations vs. time
    - Fitting the distribution
- Correlation
  - Scatterplot
  - Pearson correlation coefficient
  - Average of A for given B.
- Mean degree by year
-

Resutls

Conclusion