Interactive Visualizations for Citation and Collaboration Networks of Articles on American Physical Society Journals

Yuan Huang

Metadata

The selected data set is the citation and metadata for all the articles published in American Physical Society journals. The data set has been requested from the American Physical Society website (http://journals.aps.org/datasets) and contains over 450,000 articles from journals such as Physical Review Letters, Physical Review, and Reviews of Modern Physics dating back to 1893. The data sets have two parts:

- Citing article pairs: This data set consists of pairs of APS articles that cite each other. For instance, if article A cites article B, there will be an entry in the data set consisting of the pair of DOIs for A and B. This data set will be formatted as a comma-separated values (CSV) file consisting of the DOI pairs, one pair per line.

- Article metadata: This data set consists of the basic metadata of all APS journal articles, including DOI, journal, volume, issue, first page and last page OR article id and number of pages, title, authors, affiliations, publication history, PACS codes, table of contents heading, article type, and copyright information.

Interest

The analysis and visualization of the community structure in the citation and collaboration networks in academia provides numerous benefits for the development of the research field, including building a foundation

for future research through the acknowledgement of past research activities; identify gaps in research for researchers and students; improve the integration between theory and practice, and so on.

With this project, I wish to use various network visualization to answer the following questions: 1) How do different subfields of physics evolve in time? Which subfields are fast growing in recent years? 2) Who are the physicists with most impact in a specified field of physics? How are the top physicists interacting with each other (through citation and collaboration)? 3) What are the top institutions for a specified field? How are the institutions collaborating with each other?

This project can help the users to learn about the research history in physics and identify interesting fields. After identify a subfield which is potentially interesting, the visualizations of the community networks can help the users to find out the top physicists and groups that are working on the selected field, papers with the most impact in the field, the top institutions that the top physicists are working in, and learn about the collaboration relationship in the community.

Analytics

I will first load the data sets into mongoldb and extract the citation network for the articles. Then I will use hierarchical clustering algorithm to cluster the articles into different fields in physics. The citation and collaboration network for the authors, and the collaboration network for the institutions will also be constructed with the metadata and the citation data for articles.

Visualizations

I will provide six major visualizations: 1) a theme river graph (stacked line charts) for different fields in physics to show the development of this field over time; 2) a bubble chart with drill down to show the hierarchical structure in physics research; 3) a network graph with physicists as nodes and citation(directed)/collaboration(undirected) as links to show physicists with most impact in the field and the citation/ collaboration relationship between them; 4) a network graph with articles as nodes and citation relationship as links to show the articles with the most impact in the field and the citation relationship between them; 5) a map with institutions marked as bubbles and linked with each other according to their collaborations on articles; 6) a line chart for the number of articles published and the impact factor as a function of time for different journals in APS.

Interactions

The major visitations will all support selecting, probing and the ability to change the attributes (for example, change the citation network to collaboration network, change the meaning of the size of bubble for the physicists network from the number of paper published to the total cited times). On the left panel of the website I will provide the options to select a subset of the data (for example, select a specific field in physics, select papers from only one journal, select the papers published during a specific period of time, and so on).

Design layout

Time

2007   2017
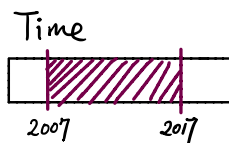
Fields

- ☐ Condensed Matter
- ☐ High Energy
- ☐ Astronomy
- ☐ Interdisciplinary
- ☑ Select All

Journals

- ☐ Rev. Mod. Phys.
- ☐ Phys. Rev. Lett.
- ☐ Phys. Rev. X
- ☐ Phys. Rev.
- ☑ Select All

ThemeRiver   y-Axis: [num. of papers ▲▼]

Legend:
- 🟥 Condensed Matt.
- 🟧 High Energy
- 🟨 Astronomy
- . . .

y

Time
2007 2009 2011 2013 2015 2017

Fields in Physics   link: [num. of collaboration ▲▼]   size: [num. of articles ▲▼]

Legend:
- 🟥 Condensed Matt.
- 🟧 High Energy
- 🟨 Astronomy
- . . .

Clicking on each bubble will triger the drill-down event, showing the subfields in each cluster.

Scientists Network in Condensed Matter

link: [no. of collaboration ▲▼]   size: [cited times ▲▼]

Only show top [10 ▲▼] scientists ranked by [cited times ▲▼]

The scientists' name will be put on top of the bubbles.

Important Articles in Condensed Matter

link: A→B A cited by B.   size: [cited times ▲▼]

Only show top [10 ▲▼] articles.

Legend:
- 🟥 Phys. Rev. Lett. 12,345 (2008)
- 🟧 Rev. Mod. Phys. 11,111 (2007)
- 🟨 Phys. Rev. B 11,100 (2009)
- . . .

Institution Network in Condensed Matter

link: [no. of collaboration ▲▼]

size: [no. of top 100 scientists ▲▼]

Only show top [10 ▲▼]

DATA CENTERS

Illustration figure is borrowed from a website www.centurylink.com.

American Physical Society Journals

Impact Factor

30
20
10

2007 2009 2011 2013 2015 2017   time

Legend:
- 🟦 R.M.P
- 🟧 P.R.X
- 🟥 P.R.L
- . . .

Number of Articles Published

2007 2009 2011 2013 2015 2017   time

Legend:
- 🟦 P.R.B
- 🟧 P.R.A
- 🟥 P.R.L
- . . .