Network Science Project 2

We hereby describe what we expect from you for the next few weeks.

1) Project evaluation

Several points shall be inspected for evaluation.

- **Draft**: presentation, novelty, model, related work review, soundness and completeness.
- **Reproducibility & DB update integration**: web Interface, complexity of data scraping, parallel computations and distributed storage
- **Overall Draft Quality & Overall computational part**

Final grade is placed by committee consisting of seminarists and, possibly, lecturer.

2) Description of project

We consider the problem of organizing research data and personal activity of researchers, and more genrally of university staff units. The aim of this project is to understand what services can be provided to researchers and universities based on co-authorship and citation networks obtained from open and semi-open data.

Taking attributive network, we aim to solve several classic machine learning tasks on graph and text information:

- *Multi-label node classification*: marking research areas of researchers and papers, respectively
- *Link prediction* and *Future link prediction* as suggestion system for searching collaborators and citing proper papers
- *Community detection* for studying the process of dense subgraphs and publication activity across such groups for improving cross-discipline collaborations
- *Visualization* of co-authorship network and research activity of departments Citation and alt-metrics prediction, topic-modelling and information retrieval, and much more...

During this task, you shall learn the real pain of data science for obtaining data, making database with update options from heterogeneous sources, creating web interface for the system integrated with popular web services, and solving real-world machine learning problems on graphs via novel techniques, called graph embeddings and graph neural networks.

Relevant Projects:

- https://aminer.org/data
- http://snap.stanford.edu/projects.html
- http://networkrepository.com/

Graph Embedding:

- http://snap.stanford.edu/proj/embeddings-www/

2.1) Conference tool

The main task in early stage is to fetch for data in the big gigantic ocean of the web.

2.1.1) Get data for conferences

Concerning existing conferences, some ressources are useful.

- **Science conference rankings**:
- http://portal.core.edu.au/conf-ranks/
- http://www.conferenceranks.com/
- https://www.aminer.cn/ranks/conf (Hirsch-based approach)
- **Call for papers**:
- http://www.wikicfp.com/cfp/
- https://cmt3.research.microsoft.com
- https://new.precisionconference.com/submissions
- http://www.acm.org/conferences
- https://www.ieee.org/conferences/index.html
- https://easychair.org/

Some code has already been implemented by for some parts of these sites and has to be integrated.

Metadata to be collected (not exhaustive): conference website, dates, location(city, country), ranking, publishing conditions, deadlines, key topics, submission format (template links), page limits, rules for full/short papers (and posters) etc..

2.1.2) Interface and service

The web interface could be similar to representation in http://www.guide2research.com/topconf/ (the latter has a lot of mistakes in linking conference and portal, as well as deadlines)

Add also some tools for integration:

- Google Calendar for time management
- Imported Google.Maps API with deadlines for submitting and travels from relevant lists
- Highlight deadlines in a list
- Highlight important ones by probability to get accepted according to researcher rank
- Highlight Travel intersections and probable chained conferences
- Highlight city/country travel index for not top conferences, as well as Money rate

Add a mailing agent which sends some notifications based on profile (or for conferences manually ticked by researcher). Supported statuses:

1. Interested (add suggestions)
2. Submission rules (template, page limit, overleaf)
3. Deadline update from conference portal
4. Submission status (submitted, notification of acceptance/rebuttal/revision, camera-ready decision)
5. Resubmitting to another section/workshop
6. Logistics (registeration, visa application, hotel, airflight, travel exchange. . .)
7. Follow-up (paper is published, paper is indexed).

2.2) Research profile analysis

Once again, we need to get data first.

2.2.1) Fetch data

Make use of : Google Scholar, Scopus Author, DBLP, and possibly Web of Science Citation Report (some may have obscure API, try your best).

2.2.2) Interface

Ability to spot a university, a research team or an individual researcher.

Constructing Co-authorship and Citation Network, loading journal metadata and quartiles from SJR (Scopus) and JCR (WoS)

2.3) Maintaining

Integrate a updating routine for a long-lasting tool!

3) Research questions

Problem 1

Input: Researcher is defined by a list of papers + Topic modelling extract automated keywords + journal categories from SJR ranking.
Output: Recommend conferences, based on probability to get accepted, ranking travel preferences (under money constraints), etc.

Problem 2

Input: A paper and its authors' profiles (or anything convenient)
Output Predicting author citation, Hirsch, number of co-authors, number of papers in high quartiles, and research impact

Inspired by these:

- http://www.sciencesuccess.org
- http://science.sciencemag.org/content/342/6154/127
- http://barabasi.com/f/442.pdf

Problem 3: Predict funding necessity and its impact on researcher's personal progress

Problem 4: Predicting co-authorship and citations via link prediction or recommender system using graph embedding. https://www.aminer.cn/billboard/id:5965cf249ed5db41ed4f52bf

Problem 5: (Cross-discipline) papers recommendation

Problem 6: Trend analysis for topics

Other problems:

- Prediction to publish research paper based on previous research track
- Evolution of research interests
- Prediction of journal quartiles and JIF/SNIP via time series and evolution of relevant co-author subgraphs.
- Conference PC membership / Journal editorial board and prediction of future publication.
- Expert search based on publications, PC membership in top conferences https://www.aminer.cn/billboard/id:599e47ab9ed5db312e97165d
- Prediction of affiliation change
- Author Disambiguation
- Prediction of staff position, sex, age, location
- Rankings: researchers, universities, conferences, journals
- Research adviser prediction
- Your own question !

4) Task assignment

Main project:

- Scraping data (45%)
    1. Get data
    2. arrange data
- Conference tool (40%)
    1. Interface
    2. AI, suggestion algorithms
- Research profile tool (15%)
    1. AI, suggestion algorithms