

Project Topics

Below is a list of possible project topics. Some of these are open-ended, meaning that you are required to come up with a new algorithm or model, and formulate it yourselves. Such projects may require more effort, but they will be also graded based on the effort, as well as the final result. Others are more straight-forward, you would need to obtain a complex dataset and apply algorithms on this data. There are also more theoretical projects, and more practical ones, so you can pick depending on your preference.

You will also have to present in class one paper related with you project. The list below includes the paper for each project.

Projects should be done **in teams of at most two** students.

Deliverables and Timeline:

- A web page with all information related to your project (first version: week before Christmas, final version: end of January)
- A two- page project proposal outlining what you plan to do. This should include the topic of your presentation (first version: week before Christmas)
- A 20' presentation of 1-2 research papers related to your project (presentations are scheduled for 15/1 and 22/1)
- The source code of your project (end of January)
- A final report describing your project following a specific format (end of January)

Topic 1

Project:

Download the Facebook data related to a small number (2-5) users.

Then, use lucene <http://lucene.apache.org/>, a free, open source information retrieval software library to index and search the collected data. The project can be extended to add PageRank in the network of friends to improve ranking.

Paper:

[Chia-Jung Lee](#), W. Bruce Croft, [Jinyoung Kim](#): Evaluating search in personal social media collections. [WSDM 2012](#): 683-692

Topic 2

Project:

Use (existing) Twitter data to find hashtags that are related to the same event. The goal is to understand how different hashtags appear for a given event, and how they evolve over time.

Paper:

Chenhao Tan, Lillian Lee, Bo Pang: The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter. ACL (1) 2014: 175-185

Topic 3

Use the Flickr API

<http://www.flickr.com/services/api/>

to collect datasets from Flickr.

Then perform some analysis on the collected datasets.

Potential topics: identifications of POIs (points of interest), etc

Paper:

Tye Rattenbury, Nathaniel Good, Mor Naaman: Towards automatic extraction of event and place semantics from flickr tags. SIGIR 2007: 103-110

Topic 4

Graph Similarity

Project:

Implement the various similarity measures proposed in the paper below. Propose an extension that takes into account edge and/or node labels, or takes into account the densest subgraph in the two graphs. Report evaluation results of applying the similarity measures on various graph datasets.

Paper:

Christos Faloutsos, Danai Koutra, Joshua T. Vogelstein: DELTACON: A Principled Massive-Graph Similarity Function. SDM 2013: 162-170

Topic 5

Sampling of Graphs

Project:

Propose a method for sampling a graph such that we can measure different properties? For example can we sample and preserve betweenness?

Example paper: Jure Leskovec, Christos Faloutsos [Sampling from Large Graphs](#) (poster) KDD 2006, Philadelphia, PA.

Topic 6

Team formation

Project:

Extend the model and algorithms for team formation described in the paper below for the case of negative edges. Implement and test them.

Paper:

Theodoros Lappas, Kun Liu, Evimaria Terzi: Finding a team of experts in social networks. *KDD* 2009: 467-476

Topic 7

Recommending tags for influence maximization

Project:

Often tags (or, keywords) are associated with resources (such as articles, photos, or posts). A basic criterion for selecting tags is their relevance to the resource. In this project, you will investigate another aspect. Assume that you are given a tag diffusion network that models how the tag influences the propagation of the resource. For example, such a graph may capture how the use of a specific hashtag affects the probability of a tweet to be re-tweeted. Now, the problem is: Given a set of m tags, an estimation of their relevance to the resource and the corresponding tag diffusion network, select the $k < m$ tags that would maximize the diffusion of the resource.

For this project, you are asked to define the above problem more formally study its complexity and propose algorithms for its solution. You should also use an appropriate dataset for evaluating the efficiency and effectiveness of your approach.

Paper:

Robert Jäschke, Leandro Balby Marinho, Andreas Hotho, Lars Schmidt-Thieme, Gerd Stumme: Tag recommendations in social bookmarking systems. *AI Commun.* 21(4): 231-247 (2008)

Topic 8

Bundle recommendations to groups of users

Project:

Previous work has explored recommendations of single items to group of users so as a group-specific optimization objective is met (e.g., the average or minimum relevance of the item to all group members is maximized). There has been also previous research on recommending set of items (e.g., travel packages, or related products) to single users. In this project, you are asked to combine the work on group recommendations with the work on bundle (package) recommendations.

Paper:

M. Xie, L. V. S. Lakshmanan, and P. T. Wood. Breaking out of the box of recommendations: from items to packages. In RecSys, pages 151–158, 2010.

Topic 9

Identifying local events in Twitter

Project:

For this project you would need to download data using the Twitter API, and within this data identify tweets that refer to an event that is localized both in time (something that happened within a specific time interval) and in space (something that happened in a restricted area – e.g. in Ioannina).

Paper:

Ke Xie, Chaolun Xia, Nir Grinberg, Raz Schwartz, Mor Naaman. Robust Detection of Hyper-local Events from Geotagged Social Media Data. MDMKDD 2013

Topic 10

Another option is to suggest a project of your own, based on what you have seen in the class so far, questions you may have thought of, and things that are related to your research area. In this case you should create a project proposal (initially just a paragraph or an idea) and contact us to discuss it.