

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 dataset. 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 your project. The list below includes the paper for each project.

Papers also vary in difficulty and scope. For experimental papers, that just report results of experimental studies, we expect that you just present and explain the main findings. Since such papers require less effort, you will be asked to present 2 such papers.

You also need to create a web page for the project (including the final report and dataset used).

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

Deliverables and Timeline:

- *Week before Christmas vacations*
 - A two-page project proposal outlining what you plan to do. This should include the topic (and papers) of your presentation
 - A 15' presentation of the project proposal (on Wednesday 18/12)
 - First version of the web page (include the link in the project proposal)
- *Week after Christmas vacations*
 - A 20' presentation of 1-2 research papers related to your project (presentations tentatively scheduled for 13/1)
- *End of January*
 - The source code of your project
 - A final report describing your project (the report needs to follow a specific format)
 - Final version of the web page

Topic 1

Structural diversity based on network embeddings

Project:

The main goal is to explore embeddings for defining network diversity. The idea is to define the diversity of a node based on the embeddings of its neighbors, for example, you may define the diversity of a node to be equal to the average distance between the embeddings of its neighbors. Then, study the correlation of the embedding-based diversity of a node with other metrics, such as the degree, clustering co-efficient, or PageRank of the node. Since this is an experimentally-oriented paper, use at least 3 different network embedding methods (e.g., node2vec, DeepWalk, GraphSage), 5 different real graph datasets and potentially a number of synthetic datasets (e.g., forest fire, preferential attachment, random graphs). Optionally, try to relate structural diversity as defined in the papers to be presented with the embedding-based definition.

Papers:

Johan Ugander, Lars Backstrom, Cameron Marlow, Jon M. Kleinberg: *Structural diversity in social contagion*. Proc. Natl. Acad. Sci. U.S.A. 109(16): 5962-5966 (2012)

Xin Huang, Hong Cheng, Rong-Hua Li, Lu Qin, Jeffrey Xu Yu: *Top-K structural diversity search in large networks*. VLDB J. 24(3): 319-343 (2015) (you are asked to study/present only Sections 1, 2, and 3.)

Topic 2

Content homophily in a real social network

Project:

The main goal of this project is to measure content homophily in a real social network. One way to formulate this problem is to test whether friends in a social network post similar content. Choose a social network, e.g., Twitter, and a set of users. Then, construct the ego network of these users and collect the posts of the users belonging to these ego networks. Use word2vec to define similarity between posts. Check the similarity between pairs of users that are friends and pairs of users that are not friends. You can further check whether friends endorse (for example, retweet) similar content. Choose the set of users, using a specific criterion, for example, choose newspapers (or, politicians) with different political orientations, football teams, or players, etc., so that you may be able to draw more general conclusions.

Paper:

Marina Drosou, H.V. Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. *Diversity in Big Data: A Review*. Big Data. Jun 2017

Topic 3

Fairness in a real social network

Project:

The main goal is to measure fairness in a real social network, in particular, Github. Construct a number of friends-networks from Github and use existing software to determine the gender of people in these networks. Consider different ways to rank nodes in the constructed networks, for example, based on degree, PageRank, Personalized PageRank, centrality, etc. Test parity fairness: is the percentage of women in the top positions close to the percentage of women in the whole population? Then, assume a diffusion process in these networks using the IC (Independent Cascade) model with a small number of seeds. Test parity fairness in the affected nodes. You are free to select seeds randomly, using some heuristic, e.g., highest degree, or, any other algorithm.

Paper:

Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer and Markus Strohmaier. *Homophily influences ranking of minorities in social networks*. *Scientific Reports* 8 (2018)

Topic 4

Fairness-aware team formation

Project:

Assume that there are one or more protected attributes (e.g., sex, race) associated with each individual and that the decision whether a particular individual will be a member of the team should not be affected by the value of these attributes. Provide a definition of fair team formation. This is a research-oriented project, so, you will need to be creative. For example, you may consider individual fairness and ask that two individuals with similar skills should have similar probabilities of being selected as members of the team. Another alternative would be group- or, parity-based fairness in which case we ask that each group is proportionally represented in the team.

Paper:

Anirban Majumder, Samik Datta, K. V. M. Naidu: Capacitated team formation problem on social networks. KDD 2012: 1005-1013

Topic 5

Cultural Dynamics

Project:

In class we mentioned the Axelrod model for cultural dynamics. The goal of this project is to experiment with this model by performing simulations on real data and try to extend it. Consider a variation where there are correlations between the attributes, and then consider the campaign problem where we want to maximize the positive opinion for a specific topic. This is a research project where you will use the Axelrod model as a basis for exploring ideas about different ways of dissemination of opinions and influence maximization.

Paper:

Battiston, Federico et al. "Layered social influence promotes multiculturalism in the Axelrod model." *Scientific reports* vol. 7,1 1809. 12 May. 2017, doi:10.1038/s41598-017-02040-4

Also useful:

Axelrod's original [paper](#)

Flache, A., & Macy, M. W. (2011). Local Convergence and Global Diversity: From Interpersonal to Social Influence. *Journal of Conflict Resolution*, 55(6), 970–995.

Topic 6

Link Recommendation for reducing polarization.

Project

In the the paper by Matakos et al., a metric is defined for measuring the polarization in a social network with opinions. Propose algorithms for the problem of suggesting links to reduce the polarization metric. The recommendations should take into account the probability of a recommendation to be accepted.

Paper

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. *Reducing Controversy by Connecting Opposing Views*. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17). ACM, New York, NY, USA, 81-90.

Original paper:

A. Matakos, E. Terzi, P. Tsaparas. "Measuring and Moderating Polarization in Social Networks", ECMP/PKDD 2017

Also useful:

Xi Chen, Jeffrey Lijffijt, and Tijl De Bie. 2018. *Quantifying and Minimizing Risk of Conflict in Social Networks*. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). ACM, New York, NY, USA, 1197-1205.

Topic 7

Polarization and extremism on Reddit.

Project

The goal of the project is to study controversy, polarization and extremism on Reddit. Obtain Reddit data on controversial categories, construct the discussion network of who replies to whom and use the approaches in the related paper to see if there is controversy. You can use data beyond the network, such as the content, or the frequency of engagement of users into controversial debates to understand the origins of controversy and the extremism. The following paper quantifies controversy. It is dense and long, so you do not need to study/present the whole material in it.

Paper

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, Michael Mathioudakis: *Quantifying Controversy on Social Media*. ACM Trans. Social Computing 1(1): 3:1-3:27 (2018)

Topic 8

Embeddings of temporal graphs.

Project

The goal of the project is to study embeddings of temporal graphs, that is, graphs that change over time with the addition or removal of edges. This is a research project that will be fully formulated as you gain more knowledge on the topic. The goal is first to understand the existing literature (the related paper is a good starting point), and then explore ideas on the information contained in the embedding of the “delta graph”, or in the time series defined by the embeddings of the same node.

Paper

Uriel Singer, Ido Guy, and Kira Radinsky: *Node Embeddings over Temporal graphs*. IJCAI 2019 ([full version](#) in the arXiv)

Topic 9

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.