

**Spring 2024**  
**DS 5720: Social Network Analysis**  
**Course Project**

**Timeline**

- Project Proposal (10%): Due March 28 by 11:59 pm  
at least  $\frac{1}{2}$  page describing an overview of your project direction, and additional information that you may have including method/application innovations, dataset(s) used, and a weekly timeline for the remainder of the semester. Team size: 2-3 students
- Project Update (5%): Due April 4 by 11:59 pm  
at least 1 page describing your progress that should include some updates such as preliminary results, basic dataset statistics/analysis, refined project plan, an updated weekly timeline for the remainder of the semester, etc., and mentioning any changes to your project since the original project proposal
- Project Presentation (35%): April 16 and 18 (Slides upload: April 25 by 11:59 pm)  
12 minute presentation with 3 minutes Q/A
- Participation in anonymously evaluating your classmates (5%):  
Completed during the presentations
- Code & Documentation (5%): Due April 25 by 11:59 pm  
The code for your project should be reasonably documented.
- Project paper (40%): Due April 25 by 11:59 pm  
3-8 pages with the expectation of 3-5 pages (not counting references)  
Please use the AAAI paper template and follow the general format for writing your papers. You can see papers online (i.e., papers from conferences such as AAAI, ICWSM, KDD, WSDM, etc.) to get an idea of the paper organization if you are not familiar (or ask the instructor/TA for some examples)  
<https://aaai.org/authorkit24-2/>  
<https://www.overleaf.com/latex/templates/aaai-press-latex-template/jymjdgdpdmxp>

**Project Ideas**

1. RecSys Challenge 2024: <https://www.recsyschallenge.com/2024/>  
“The Ekstra Bladet RecSys Challenge aims to predict which article a user will click from a list of articles that was seen during a specific impression. Utilizing the user’s browsing history, session details (like time and device used), and personal metadata (including gender and age), along with a list of candidate news articles, listed in an impression log. The challenge’s objective is to rank the candidate articles based on the user’s personal preferences. This involves developing models that encapsulate both the users and the articles through their content and the users’ interests. The models are to estimate the likelihood of a user clicking each article by evaluating the compatibility between the article’s content and the user’s preferences. The articles are ranked based on these likelihood scores, and the precision of these rankings is measured against the actual selections made by users.”

2. Amazon KDD Cup 2023 competition: <https://www.aicrowd.com/challenges/amazon-kdd-cup-23-multilingual-recommendation-challenge>  
This competition has three separate tasks in the area of multilingual session-based recommender systems. Specifically, next product recommendation, next product recommendation for underrepresented languages, and next product title generation. Note that the competition also provides cash or AWS credit prizes for the best performers.
3. Traditional link prediction as we discussed is to relate two nodes together. However, many times we might want to relate a node to a pair of nodes. For example, you meet a couple in a social network, or wanting to recommend a movie to two friends to watch together. Additionally we can consider the case of this pairwise link prediction in signed networks where we predict how a third node will connect in with an existing link of two nodes to form a triangle.
4. Traditional recommender systems (i.e., business-2-customer (B2C)) suggest items to users, but in customer-2-customer (C2C) ecommerce (e.g., eBay, Taobao, Etsy, etc.) we need not only recommend an item, but also a seller. Develop a GNN model for making these recommendations.
5. Tie strength prediction seeks to predict the strength (i.e., weight) of an unweighted link. Rather than using a GNN to predict links, here we can use them to predict link strengths. This can utilize model changes in attention mechanisms and loss function changes to include strong versus weak tie theories.
6. Given the dataset of an online weight loss application where users can connect with each other, comment on others posts, etc., we can leverage these connections to better predict the weight loss of a user based on a GNN model that accounts for not only a user's information, but that of their local neighborhood.
7. Recently researchers are beginning to acknowledge the importance of understanding broken ties online (e.g., unfollowing, blocking, etc.). Given they can also be used for better performance on tasks such as social recommendations, it is valuable to be able to predict them. We can use a GNN model to predict which links will be broken in the near future, while also predicting the delay time for when these events will happen.
8. Collect data from Reddit, BlueSky, Venmo, TikTok, or another social network site and propose the analysis/prediction of certain trends in the social network data. Note that the social network data collection through BlueSky's API I personally think would be really interesting :)  
Some potentially useful links:  
<https://docs.bsky.app/>  
<https://github.com/MarshallX/atproto>  
<https://www.reddit.com/dev/api/>  
<https://praw.readthedocs.io/en/stable/>

*Again, we note that this are not an exhaustive list, but can be used as seeds for coming up with other ideas/topics. We have datasets for each of the above, but some will require a little more preprocessing from raw data than others and that will be taken into consideration if needing to initially spend more time on processing/collecting the raw data.*