UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

PROJECT OF 42913

UTS:
ENGINEERING AND INFORMATION TECHNOLOGY

PROJECT AIM

- Apply analysis tools and algorithms to solve real-life problems
- Leave you well-qualified to do network science research

TEAM

The students are required to work in groups of 2-4 people.

Please

- ❑ Start to explore the possible topics and establish/join a group **as soon as possible**.
- ❑ Nominate a group leader, who will send the instructor an email about the group members as well as the proposed project topic if available. (by Week 5)
- ❑ **Start the project ASAP !**

Important Date

- **Fri, Week 5**: The groups should be formed with a group leader, who will send an email about the group members to instructor and a group id will be allocated. The research topic of each group needs be approved by the instructor if not the suggested ones.
- **Fri, Week 12**: Group project due. (Submit via Turnitin)

PROJECT CONSULTATION

- Encourage group discussions in the Lab.

  In the lab, we will leave around 30-45 minutes for group discussions (instructors and tutors may join)

- Make an appointment with tutors and instructors

- We will have an online consultation 4:00pm – 5:00 pm every Tuesday.

HOW TO PICK A TOPIC?

- In general, a topic can fall into one or more of three categories
  - **Network analysis and visualization**: Analyze an interesting network from different aspects, such as degree distribution, network centrality, community detection, network evolution and graph visualization.
  - **Algorithm**: A scalable implementation of an algorithm for processing massive graphs that solves real problems.
  - **Application**: Develop a novel application to offer a new function in real-world problems based on network analysis.
- 7 suggested topics. Otherwise, need the approval of the instructor

EVALUATION

- Students should have a publishable or near-publishable reports for their projects.

- The report may include abstract, introduction, related work, dataset, methods (e.g., algorithms or network metrics used), results (e.g., experimental report, analysis and visualization), and conclusion.

EVALUATION (CONT.)

The project will be evaluated based on:

- The technical quality of the work: Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?
- Significance: Did the authors choose an interesting or a “real” problem to work on, or only a small “toy” problem? Is this work likely to be useful and/or have impact?
- The novelty of the work, and the clarity of the write-up
- Presentation of the results
- Well-formatted, well-organised, spell-checked and grammar-checked documents
- **Plagiarism check**
- **By default, members in the same group will receive equal marks**. If some of the group feel that other members are not contributing, the instructor should be informed and a group meeting should be held to produce a solution. In extreme cases a group member may be asked by the tutor to withdraw from the subject, do extra work or accept a lower mark. *No complaints about group operation will be considered after the project has been handed in.*

EVALUATION (CONT.)

1. Technical Quality (**40%**)

Are the results technically sound?
Are there obvious flaws in the conceptual approach?
Are claims well-supported by theoretical analysis or experimental results?
Are the experiments well thought out and convincing?
Will it be possible for other researchers to replicate these results?
Is the evaluation appropriate? Did the authors clearly assess both the strengths and weaknesses of their approach?

2.Quality of Writing (**30%**)

Is the paper clearly written?
Is there a good use of examples and figures?
Is it well organized? Are there problems with style and grammar?
Are there issues with typos, formatting, references, etc.?

3. Novelty and Significance **(30%)**

We will recognise and reward papers that propose genuinely new ideas. Novel combinations, adaptations or extensions of existing ideas are also valuable.
Is this a significant advance in the state of the art?
Is this a paper that people are likely to read and cite?
Does the paper address an important problem?

SELECTED TOPICS

- We offer **7** selected topics

  - ❑ Social Network Analysis using Large Language Models
  - ❑ Friend recommendation on social networks (e.g., Facebook)
  - ❑ Movie recommendation
  - ❑ POI recommendation
  - ❑ A simple Google search prototype (the foundation of Google Inc)
  - ❑ Cycle detection problem (real-application from Alibaba group)
  - ❑ Maximal Biclique detection (real-application from Alibaba group)

OTHER TOPICS

- Related Tasks from Online Competitions, e.g., **Kaggle,** **Yelp**

- Network Visualization and Analysis. Here are some examples from Stanford course: Analysis of the YouTube Channel Recommendation Network, Network analysis on Startups and VCs and Analysing the development of Wikipedia.

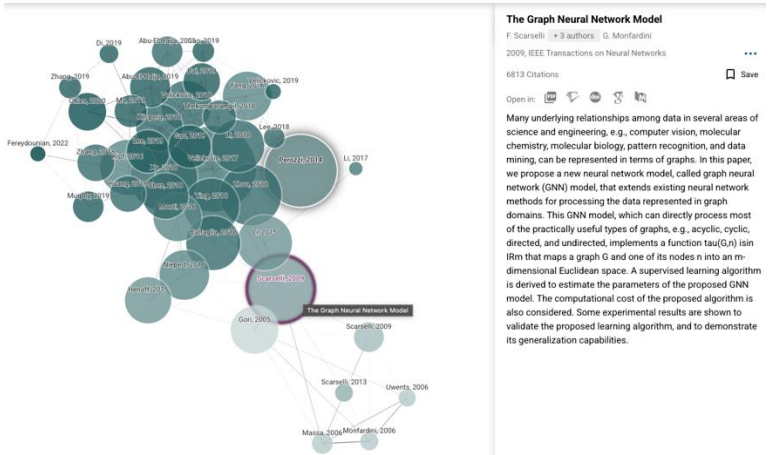
OTHER TOPICS

- Implementation and possible improvement of an existing research paper related to Social and Information Networks. You may find interesting research papers from the following top conferences such as WWW, KDD, WSDM, IJCAI, AAAI, VLDB, and SIGMOD.

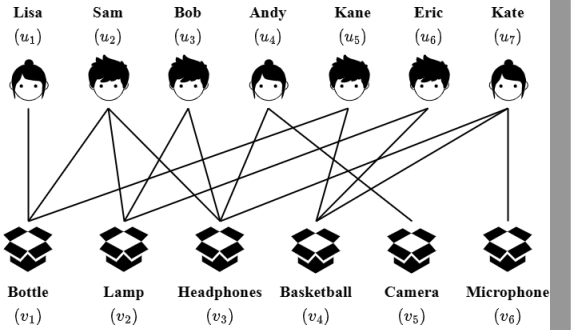
- Design an efficient algorithm to find the shortest path between any two actors in IMDB dataset, similar to Six Degrees of Kevin Bacon (https://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon).

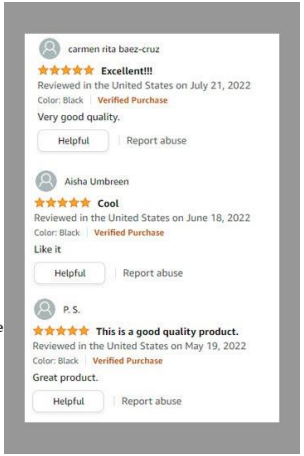
- Topics related to graph node classification (see https://hpi.de/fileadmin/user_upload/fachgebiete/mueller/courses/graphmining/GraphMining-06-NodeClassification.pdf) or graph embedding (See https://cs.stanford.edu/~jure/pubs/graphrepresentation-ieee17.pdf)

PROJECT 1: SOCIAL NETWORK ANALYSIS USING LARGE LANGUAGE MODELS



Citation network





E-commerce review network

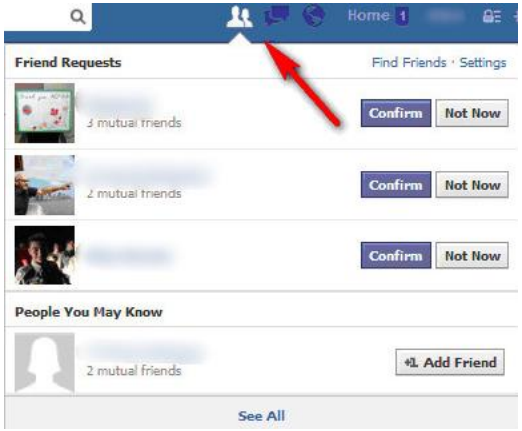
PROJECT 1: SOCIAL NETWORK ANALYSIS USING LARGE LANGUAGE MODELS

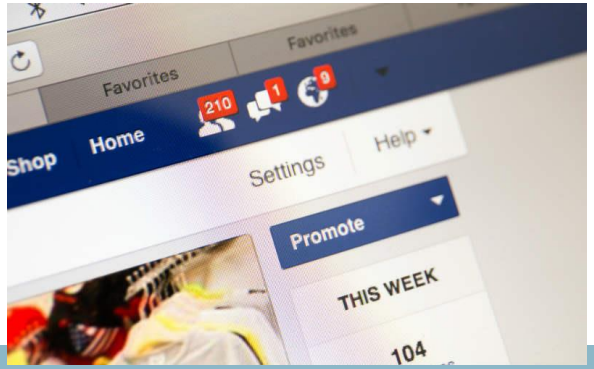
- Analysis of citation networks.
  - Categorizing papers.
  - Predicting possible citation relationship

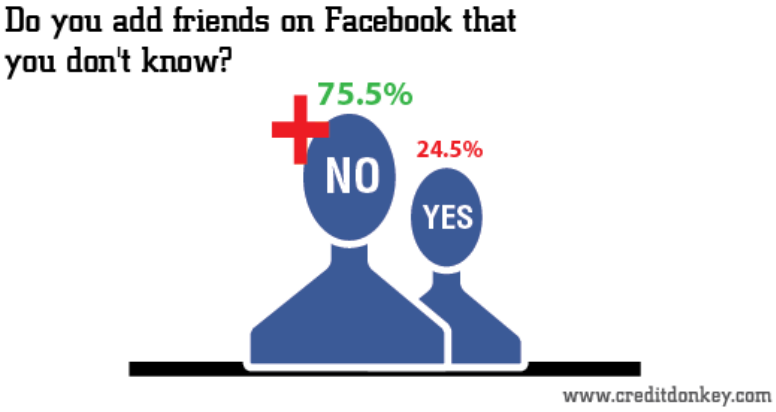
- Fraudulent review detection on e-commerce network.
  - Identifying fraudulent purchases.
  - Identifying fake reviews

- Leveraging the semantic information using large language models!

PROJECT 2: FRIEND RECOMMENDATION ON SOCIAL NETWORK







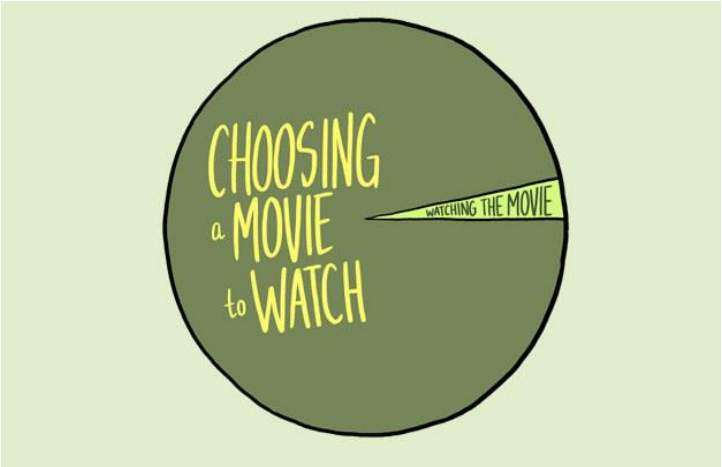
PROJECT 2: FRIEND RECOMMENDATION ON SOCIAL NETWORK

- Predict the future link between two people

- Two datasets will be provided: Facebook network and collaboration network. Only the network structure information is available.
  - ❑ Facebook: about 1m users
  - ❑ Collaboration: about 20k nodes and 200k edges

- You may choose your own dataset

PROJECT 3: MOVIE RECOMMENDATION



PROJECT 3: MOVIE RECOMMENDATION

- Movie and user data will be provided.
- The dataset consists of 100,000 ratings (1-5) from 943 users on 1682 movies.
- You need to use the data to predict the preference of the users and recommend movies for them.

PROJECT 4: POI RECOMMENDATION

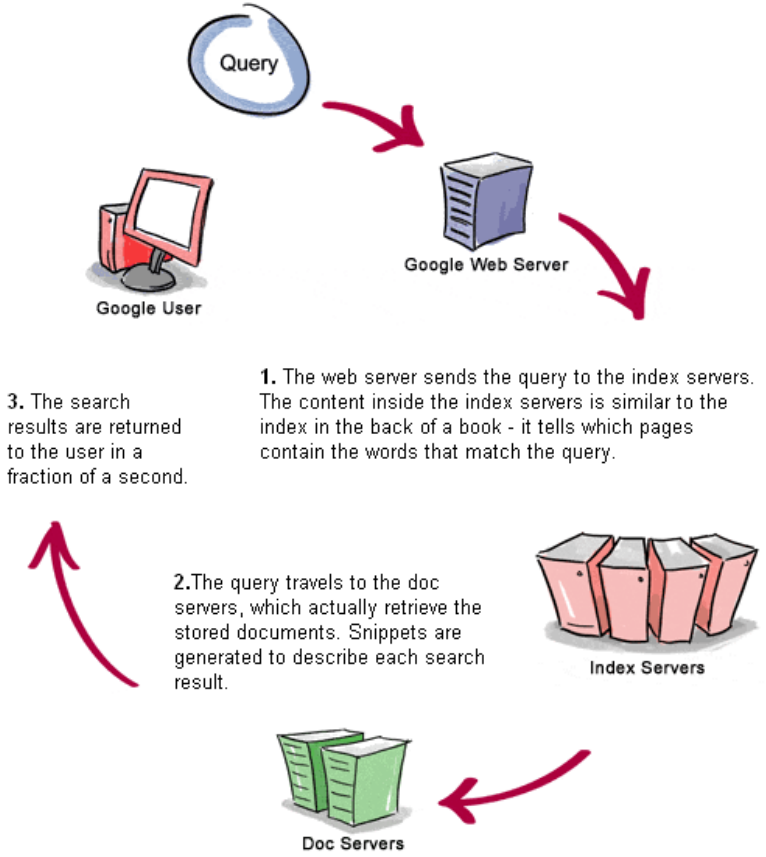
Which restaurant to choose for dinner?

PROJECT 4: POI RECOMMENDATION

- Dataset which includes the following details will be provided:
  - ❑ User detail
  - ❑ Business detail (location, type, tag, rate etc.)
  - ❑ Check in detail (user, business, time, location)
  - ❑ Rate and comment detail

- You need to use this information to recommend point-of-interest for the users.

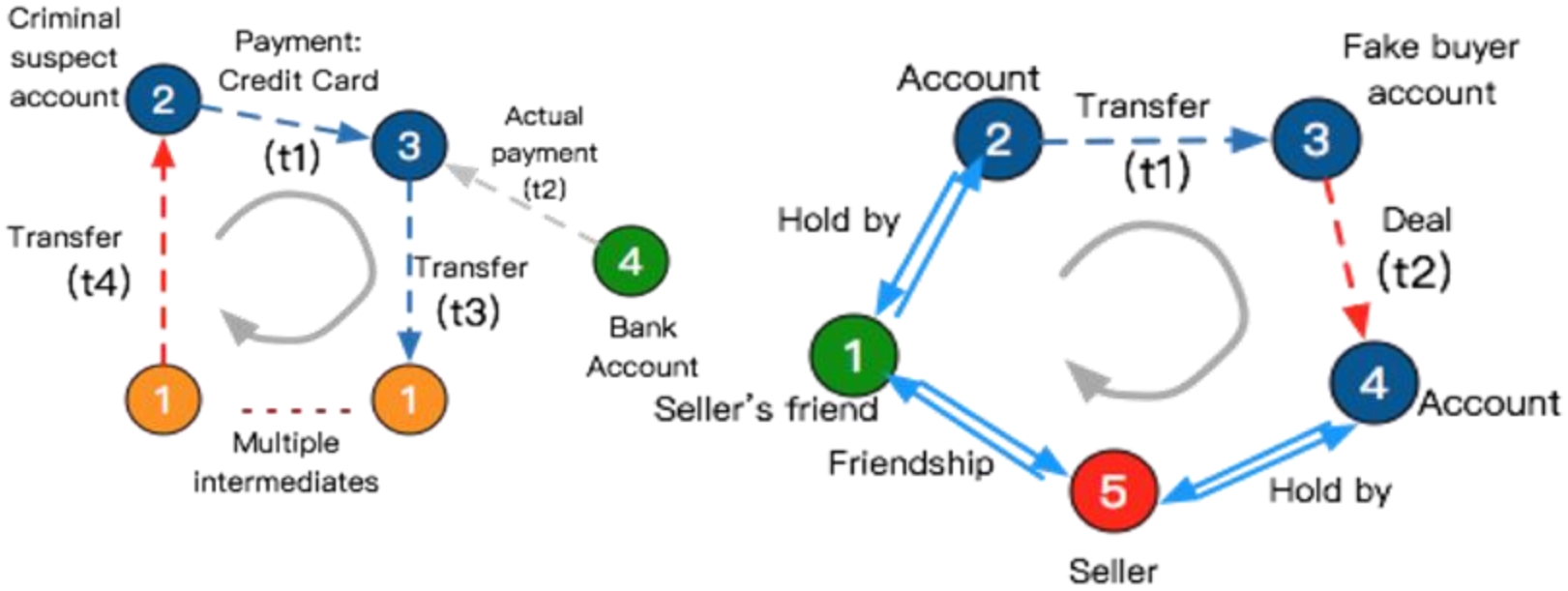
PROJECT 5: A SIMPLE GOOGLE SEARCH PROTOTYPE



PROJECT 5: A SIMPLE GOOGLE SEARCH PROTOTYPE

- You need to implement a simple google search prototype.
- The ranking model MUST integrate PageRank. You may also investigate other features as many as possible.
- A WebSpam dataset will be provided
  - ❑ Contain web pages that could be spam or non-spam.
  - ❑ Contain web link structure.

PROJECT 6: CYCLE DETECTION



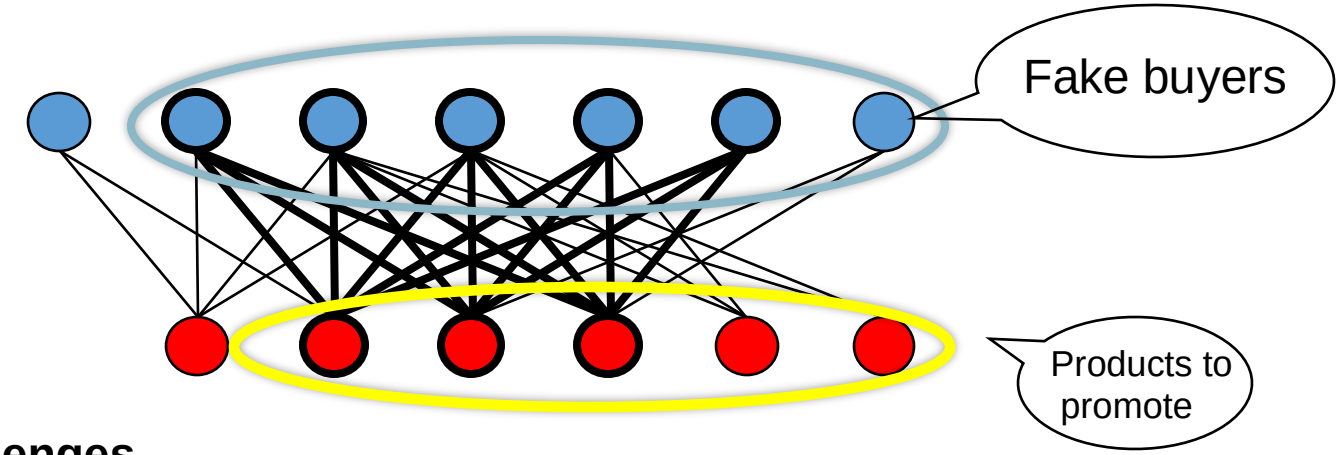
PROJECT 6: CYCLE DETECTION

Given a dynamic graph, for each incoming edge (transaction)

- your goal is to identify the newly generated cycles

PROJECT 7: MAXIMUM BICLIQUE SEARCH

Farm clicking detection (Alibaba **Double 11** festival)



Challenges :

- Billions of buyers and productions, 10+ billions of transactions
- Dynamic data

Solution : Efficient biclique detection on bipartite graph
Outcome: Significantly increase the recall by 40% in Double 11 festival in 2017

PROJECT 7: MAXIMUM BICLIQUE SEARCH

Given a bipartite graph, enumerate the maximum bicliques