Machine Learning with Graphs (MLG)

HW2: Who-To-Follow Link Prediction

In-class Competition

Competition Deadline: **2021.04.09 (Fri.)** 23:59

Report Deadline: **2021.04.11 (Sun.)** 23:59

Submission: Code (.py/.ipynb) and Report (PDF)

HW2: Link Prediction Competition

- "Link Prediction with User Attributes" http://140.116.52.202:5566/
- 3 follower-followee network datasets
  - Edge is follower-followee relationship, **NO direction**
  - Nodes are associated with a list of **attributes**
  - Training and testing data are provided (BUT anonymized)
    - **train.csv** (node pairs), **test.csv** (node pairs), **content.csv** (binary attributes)
- **Goal: predict whether user A will follow user B**
  - Evaluation metrics: average **Precision & AUC** over 3 datasets
  - Tip: supervised approaches may lead to better performance
- Grading: **private leaderboard rank (60%)** & **report (40%)**
  - You can only see the public leaderboard results
  - Private leaderboard results will be released after deadline
- You are competing with last year's top-5 students! ☺

train.csv

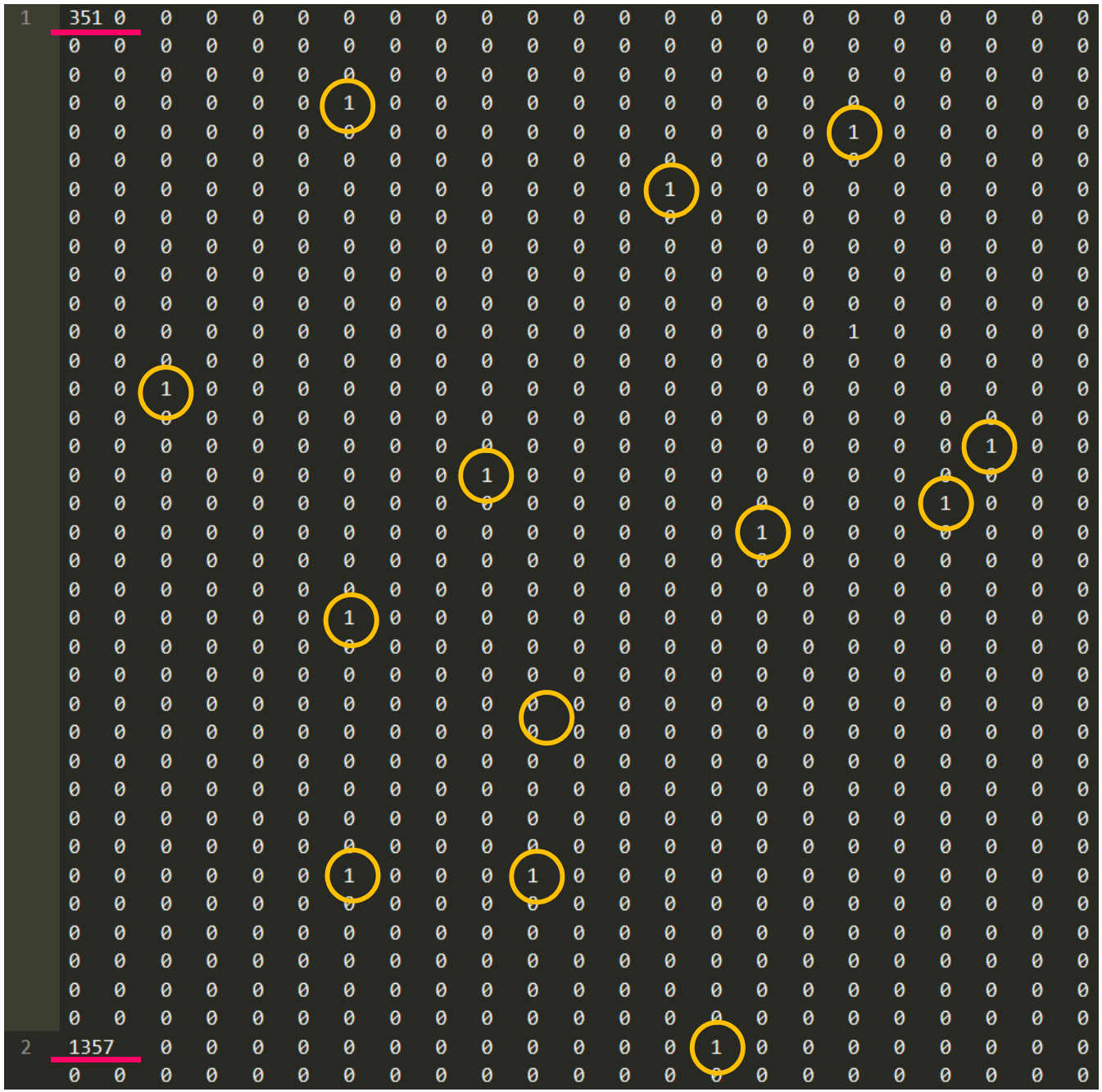
```
id,to,from,label
E10311,2399,2339,0
E10255,2397,1144,1
E10667,854,1726,0
E9395,872,702,0
E5926,2450,1312,1
E485,384,1277,0
E7506,1808,2472,1
E2160,97,1861,0
E7406,2030,2494,1
E5573,682,100,0
E10441,477,1215,0
E4626,918,2028,0
E4251,451,2487,0
E5685,849,161,0
```

binary

test.csv

```
id,to,from
E10559,2323,2673
E4849,81,1634
E3964,2405,1765
E542,2114,498
E331,1013,849
E10471,354,169
E9186,1054,2282
E5576,1004,2322
E9705,2106,918
E7309,2450,1569
E476,911,388
E5217,1056,2570
E7653,415,2022
E1530,760,2592
```

content.csv



upload.csv

```
1  id,prob
2  E10559,0.5
3  E4849,0.5
4  E3964,0.5
5  E542,0.5
6  E331,0.5
7  E10471,0.5
8  E9186,0.5
9  E5576,0.5
10 E9705,0.5
11 E7309,0.5
12 E476,0.5
13 E5217,0.5
14 E7653,0.5
15 E1530,0.5
```

Probability

HW2 Competition Submission

http://140.116.52.202:5566/

- **HW2 Competition submission via our website**
  - Deadline (system): **April 9, 2021 (Fri), 23:59** (close time)
  - Submission count **per day** = 30
    - Do the HW2 as early as possible to have more submissions
  - We have provided the sample uploading format (**upload.csv**)
  - Prediction results of each dataset need to be upload separately
    - That said, you can submit only 10 times for every 3 datasets (10x3=30)
- **Login info**
  - username = password = your 學號 (e.g., R12345678)
  - Change your password at you first login
- **Evaluation columns**
  - **mAUC:** sklearn.metrics.roc_auc_score
  - **mAP:** sklearn.metrics.average_precision_score
  - **Final = (mAUC+mAP)/2**
- **We will provide awards for top 5 students ☺**

HW2 Report+Code Submission

- HW2 Report+Code Submission via **Moodle**
  - Deadline: **April 11, (Sun) 2021, 23:59**
  - Submit your code: **.py** or **.ipynb** (preferred)
  - Submit report (PDF): ≥10 pages (you cannot include code in report)
- Content in the report
  - **1) Introduction** (報告簡述)
  - **2) Methodology**: describe all details of your features & methods
  - **3) Experimental analysis**, along with analysis and insights
    - Systematically compare and report all methods you have tried
    - Explain WHY your prediction is so GOOD or so BAD!
    - Present any insights based on your results
  - **4) Conclusions**
    - Explain the **novelty** of your methods
    - Summarize the findings
    - Point out how to improve in the future
  - **5) Citations** (if you use any methods or papers)

Important Tips & Notes

- **NOT necessary to follow methods introduced in lectures**
  - **Features/methods in lectures are only baselines**
  - → Define your own features
  - → Come up with your own supervised learning methods
- **Find and implement relevant papers/methods in Google**
  - If it is, you must please cite them in your report
- **What if the prediction of my method leads to very bad rank in the leaderboard?**
  - Remember: currently you can only see public leaderboard
  - Leaderboard takes only 60%, you can write a **good** report (40%)
    - E.g. Explain why does your method fail
    - E.g. Analyze when will you method lead to good performance
    - E.g. Compare all method you have tried and analyze each one
    - E.g. Highlight the novelty of your method