

# Лабораториска вежба 2 - Предвидување врски и системи за препораки

## Податочно множество - AmazonBook

Податочното множество AmazonBook се состои од 52 643 корисници и 91 599 книги заедно со 2.9М врски кои претставуваат рејтинг, собрани од Amazon. За вчитување на множеството искористете го следниот код:

```
from torch_geometric.datasets import AmazonBook  
  
data = AmazonBook('../data')
```

## Задачи

### Задача 1 - Предвидување на врски со мерки за сличност (15 поени)

Во оваа задача треба да се имплементира предвидување на врски со **Jaccard Coefficient** од графот кој произлегува од податочното множество AmazonBook. Од графот отстранете 20% од врските како позитивни примероци за тестирање. Дополнително, одберете исто толку врски во графот кои не постојат и ќе претставуваат негативни примероци при тестирањето.

Со предвидените и постоечките врски да се пресмета:

- ROC AUC score
  - `sklearn.metrics.roc_auc_score`
- Average Precision
  - `sklearn.metrics.average_precision_score`

### Задача 2 - Предвидување на врски со GNN (15 поени)

Во оваа задача треба да се имплементира предвидување на врски со модел базиран на граф-невронски мрежи. Користете 20% од врските како примероци за тестирање, 20% за валидација, а останатите 60% за тренирање. За оваа задача искористете ја скриптата `gnn_link_prediction.py`. Какви се добиените резултати?

### Задача 3 - Системи за препораки со LightGCN (20 поени)

Во оваа задача потребно е да направите систем за препораки (предвидување на врски) од дадениот граф со користење на LightGCN. Користете 20% од врските како примероци за тестирање, а останатите 80% за тренирање. За јазлите од врските кои претставуваат примероци за тестирање направете предвидување (препорака) со истренираниот модел (испробајте различни вредности за бројот на препораки  $k=1, 5, 10$ ). Потоа пресметајте ги мерките  $precision@k$ ,  $recall@k$  и  $f1@k$ .

$$precision@k = \frac{\text{number of relevant items in top } k}{k}$$

$$recall@k = \frac{\text{number of relevant items in top } k}{\text{total number of relevant items}}$$

$$f1@k = \frac{2 * precision@k * recall@k}{precision@k + recall@k}$$