Data Description

> Author: daven

lab-1

- author_file_ann.txt
  - Contain an anonymous coauthor network with **two** integers (range from 0 to 6610, each represents one unique author) per line (an edge), which means two authors have cooperated.
- paper_file_ann.txt
  - Contain an anonymous citation network with **two** integers (range from 0 to 79936, each represents one unique paper) per line (an edge), referring that the paper with former id cites the paper with latter id.
- bipartite_file_ann.txt
  - Contain an anonymous bipartite network with **two** integers per line (an edge). The first id refers to an author, and the second id refers to a paper that the author cites.
  - ***In this lab, we can view authors as users while papers can be regarded as items.***
- feature.pkl
  - For a better and reasonable experiment setup, we provide initial features with a dimension size of **512** for each paper, generated by USE (Universal Sentence Encoder), and need to be read with `pickle.load()`.

lab-2

- coauthor_covid.txt
  - Contain an anonymous coauthor network with **three** integers per line, denoted as **(a1, a2, t)**. which means that author ***a1*** and author ***a2*** has cooperated for **t** times.
- author_info_covid.json
  - Contain a dictionary, for each key is an anonymous author id, and each value is the basic information in the format of `dict` for the author.
  - For each author, we provide the nationality of the author and the fields the author has published a paper in. The detailed format and sample are as follows.

```
    # format
    'author_id' : {
        "ct": "country code with 2 letters",
        "ca": [["anonymous field id", count_for_the_papers]]
    }
    # sample
    '0' : {
        "ct": "SE",
        "ca": [
            ["11", 18],
            ["06", 1]]
    }
    """
```

```
    Author with id "0" comes from the country name "SE" and has published "18"
papers in the field "11" and only "1" paper in the field "1".
    """
```

P.S.

For more extra information on the datasets, please contact the author. *e.g.,* the timestamp of each cooperation.