CS245-2 Project (2022 Spring)

Assignments and Related Works for Applying Graph Algorithms to Real-World Issues

In this project, you need to choose 1 problem from 2 problems (working on both is yet recommended)

Assignment 1. GNN over Recommendation Senario

Introduction

GNN has performed very well in node classification and link prediction of heterogeneous and isomorphic graphs. In many recommendation scenarios, the trust relationship between users and the essential attributes of products are used, combined with users' purchase records, and using the representation generated by GNN can also achieve good results. However, in the scenarios given by the academic platform, we can find that collaborator recommendations, paper recommendations, and reviewer recommendations for journals and conferences are the main tasks.

For most of the above problems, there are relatively good benchmarks and baselines. However, in the paper recommendation scenario, if it is for scholars conducting scientific research through their collaborators and the paper community where their papers are located, citation recommendation still lacks better datasets and models. A dataset we share here is used in a recommender system. When both the user and product have an associated network, we can extract group features for the user group and product category. Based on the relation between products and the connections between users, it provides users with better and more various recommendations, and this may help to solve the problem of less user behavior caused by the cold start of the recommender system.

Here is a link prediction problem in an academic network. We collected 6,615 authors and corresponding 85,534 papers from top journals in the field of GeoScience as well as citation information of their publications. The collected information is used to form an academic network, and there is a feasible way:

Build a heterogeneous network, which contains two types of nodes, one type of nodes represents authors, and the other represents papers. In this network, each edge between an author node and a paper node means that the authors have read the paper (connecting authors and the papers cited by the papers written by the authors), each edge between two author nodes denotes the co-authorship, and each directed edge between two paper nodes represents the citation relation.

You can propose some other way to form an academic network. Note that what we provide is the author and paper information so far, and it reveals the correlation between work of different authors. Consequently, suppose that you are designing an academic reading recommendation system, and you need to pick out papers related to the author's previous research. This problem can be modelled as a link prediction problem, and your task is to predict each author-paper pairs in the test set based on the information provided. If the paper is recommended to the author, mark it as 1, otherwise mark it as 0.

Data

Will be released soon

Reference

1. 【arXiv 2020】 Graph neural networks in recommender systems: a survey. paper
2. 【arXiv 2021】 Graph learning based recommender systems: A review. paper
4. 【KDD 2018】 Graph Convolutional Matrix Completion (GC-MC). paper
5. 【KDD 2018】 Graph convolutional neural networks for web-scale recommender systems (PinSage). paper
6. 【RecSys 2018】 Spectral collaborative fltering (SpectralCF). paper
7. 【SIGIR 2019】 Neural graph collaborative filtering (NGCF). paper
8. 【SIGIR 2020】 Lightgcn: Simplifying and powering graph convolution network for recommendation (LightGCN). paper
9. 【SIGIR 2019】 A neural influence diffusion model for social recommendation (DiffNet). paper
10. 【WWW 2019】 Graph neural networks for social recommendation (GraphRec). paper
11. 【WWW 2019】 Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems (DANSER). paper
12. 【RecSys 2019】 Deep social collaborative filtering (DSCF). paper

Assignment 2. Key Author on Co-author Network

Introduction

During the pandemic, many scientific researchers have devoted themselves to the pathology, transmission, policy, control methods, and other aspects of the epidemic. In addition to biomedical researchers, there are also many scientific researchers from other disciplines who use their related work to analyze and model. and solve the problems caused by the pandemic. It has spawned many interdisciplinary collaborations, cross-domain projects, and academic achievements.

In scholarly network analysis, there are many issues worthy of our attention. Structurally, the network model is divided into the regular network, random network, small-world network, and a scale-free network. We generally think of it as a scale-free network in the scholar network. In the face of the cooperative relationship between scholars, we often use the clustering coefficient to measure it. By calculating the degree of interconnection between the neighbor nodes of a point, we can know the degree of mutual understanding between scholars. We can use this indicator during the epidemic to know how close the cooperation between scholars is. Related to this is the sparsity of the network, and the cooperation networks in different fields have different sparsity.

Next, we can further analyze the network's topology through different motifs. Some cooperation models are radial, which means that a strong scholar leads some lesser scholars, and some cooperation models are chained, representing scholars. They do not know each other, but they get what they need. There is also a cooperation model that is two radially connected, relying on two influential scholars to cooperate to combine the two big teams.

These scholars who have met and cooperated with two initially unknown scholars are identified as critical scholars, and these scholars play an essential role in scientific research.

We can sometimes think that the scholar cooperation network is a small-world network. Such a network often has a smaller average distance and a more significant clustering coefficient. These indicators are all used to measure a community, but How much influence does a point have on the clustering and connectivity of this community? Currently, only centrality indicators can be measured,

We can use betweenness centrality (in the shortest paths of all pairs of nodes in the network, the shortest paths passing through a node, the more influential the node is), closeness centrality (the average distance a node has from other nodes in the network, the more is smaller, the greater the closeness of the node is), the eigenvector centerline (the importance of a node depends on both the number of its neighbors (that is, the degree of the node) and the importance of each neighbor), degree centrality (the more neighbors a node has, the more important it is), semi-local centrality (count the number of second-order neighbors)

However, this indicator ignores the attributes of scholars themselves. Therefore, it is far from enough to identify a critical scholar. Therefore, we take the opportunity of the big assignment to discuss the classification of this critical scholar with you.

The data we give contains:

1. A scholar's institution, country, and discipline, from which we can know an essential attribute of a scholar.
2. When two scholars come from the same institution and are in the same discipline, the cooperation situation is likely to be intra-group cooperation.
3. When two scholars come from different institutions but are still in the same discipline, the cooperation situation is likely to be the cooperation of disciplines.
4. If two scholars come from different countries and disciplines, then the two scholars are engaged in some interdisciplinary, international cooperation, which is likely to be a specific topic that requires joint exploration by multiple countries.

We divide these scholars into four categories:

1. the heart of the network
2. network switch
3. network router
4. network edge node

The example is as follows;

the heart of the network	network switch	network router	network edge node

please design an indicator that can divide the scholars in the network.

Data

This topic provides a cooperative network of authors who have published more than 500 papers and cooperated more than ten times during the pandemic.

Will be released soon

Reference

5. 【PNAS】Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research. paper
6. 【PNAS】Collaboration and Knowledge Networks: A Framework on Analyzing Evolution of University-industry Collaborative Innovation. paper
7. 【Nature Communication】Dynamics of social network emergence explain network evolution. paper
8. 【PNAS】How humans learn and represent networks. paper
9. 【Financial Science】The Evolving Network of Legal Scholars. paper
10. 【Scientific Report】Ollivier-Ricci Curvature-Based Method to Community Detection in Complex Networks. paper

Attention

The project will be carried out on the **Kaggle** platform. You need to submit results on Kaggle to participate in the performance evaluation and ranking, and you need to submit other materials on Canvas. For project details, data format, evaluation methods, etc., please check the Kaggle competition page.

Don't forget to join a project group on Canvas.

For fairness, we set a few rules for you to obey --- violation against the rules will lead to score deduction:

1. Please do not copy someone else's code. We will run plagiarism check after submission.
2. Please do not download the dataset somewhere else to train your model. We have shuffled the dataset and will reproduce your experimental results using the dataset. A violation would be considered if there is a big gap between your reported results and our reproduced results.
3. Please do not use the pre-trained model. Everything should be built from scratch.

Other references

State-of-the-art models for graph ML can be found at:

1. OGB leaderboard
2. Top ML conferences: To find papers related to graph ML, you can search for the term "graph" across the titles.
    - ICML 2019: http://proceedings.mlr.press/v97/
    - ICML 2020: http://proceedings.mlr.press/v119/
    - ICML 2021: https://proceedings.mlr.press/v139/
    - NeurIPS 2019: https://papers.nips.cc/paper/2019
    - NeurIPS 2020: https://papers.nips.cc/paper/2020
    - ICLR 2019: https://openreview.net/group?id=ICLR.cc/2019/Conference
    - ICLR 2020: https://openreview.net/group?id=ICLR.cc/2020/Conference
    - ICLR 2021: https://openreview.net/group?id=ICLR.cc/2021/Conference
    - KDD 2019: https://www.kdd.org/kdd2019/proceedings/
    - KDD 2020: https://www.kdd.org/kdd2020/proceedings/
    - KDD 2021: https://www.kdd.org/kdd2021/proceedings/

Acknowledgements

Data used in this project is provided by Acemap group.