MOOC Action Network

1. Curation Rationale

Authors and Reference: K.Srijan, X. Zhang, J. Leskovec [1]. The dataset is built on top of KDD cup, 2015 [2].

Purpose: Predicting students' dropout on MOOC platforms. The results are published in [1].

Domain: Interaction network.

Contents: The user actions over objects on MOOC platforms. This is a snapshot of all student-course interaction on MOOC platforms.

Node and Edge Semantics: Each node represents either a student or an object on MOOC platforms. Each edge represents one action a student conducts on an object (videos, answers, webpages, etc.).

Types: directed, multi-graph, unweighted, temporal, bipartite

2. Dataset Collection, Preprocessing and Annotation

2.1 Data Collection

Data collection mechanism: The data is released by XuetangX, a Chinese MOOC learning platform. XuetangX host online courses for students and collected the data.

Raw data description: The raw data contains all students' action on courses including: working on course assignments, watching course videos, accessing course objects, accessing the course wiki, accessing the course forum, navigating to other part of the course, and closing the web page.

2.2 Data Preprocessing

Network construction: *Unknown (this is done by XuetangX)*
Data cleaning: *Unknown (this is done by XuetantX)*
Data filtering: *Unknown (this is done by XuetantX)*
Network transformation: *Unknown (this is done by XuetantX)*
Attribute transformation: Anonymize the name of students, the name of courses, and the timestamp of interactions.
Data splits: Data is split based on time. The first 60% interactions in the timeline are training set, the next 20% interactions are validation set, and the rest is test set.

2.3 Instance Demographics

All students are university students in China.

2.4 Data Annotation

Each interaction (edge) has annotation on whether this is the last action before the student dropout the course.
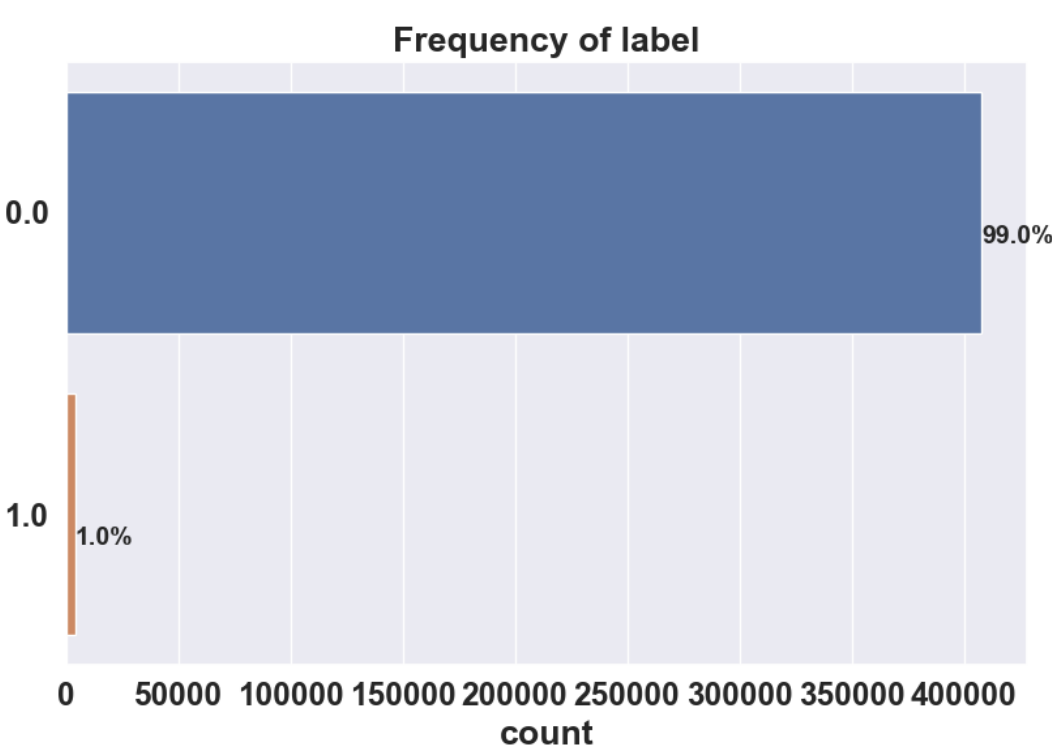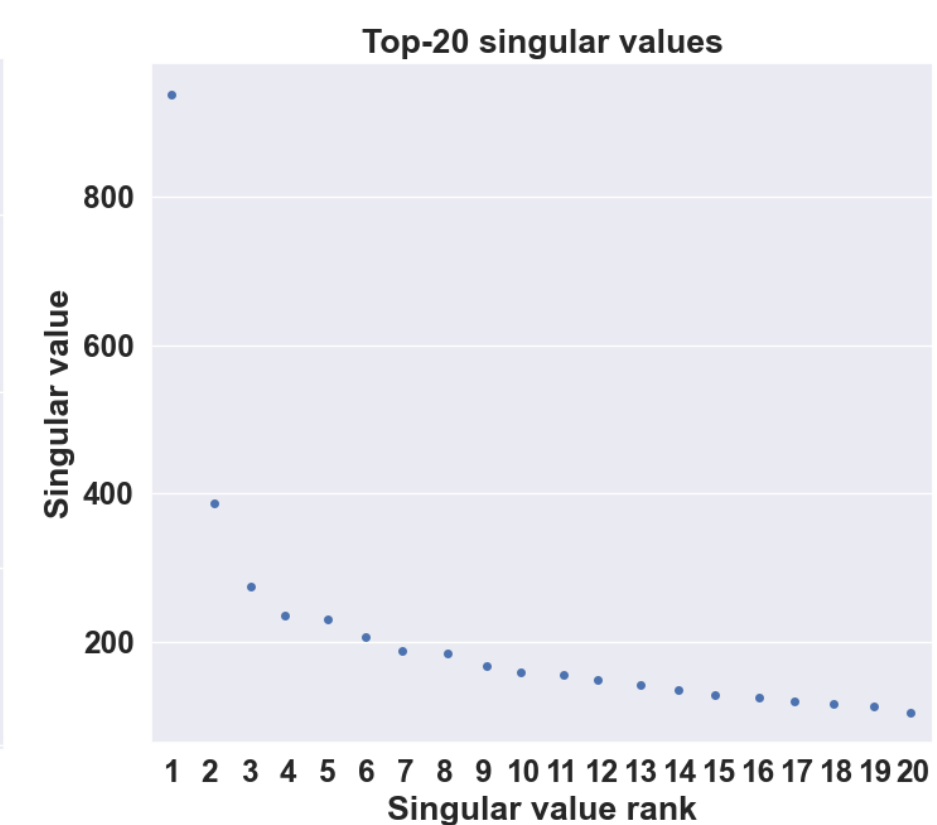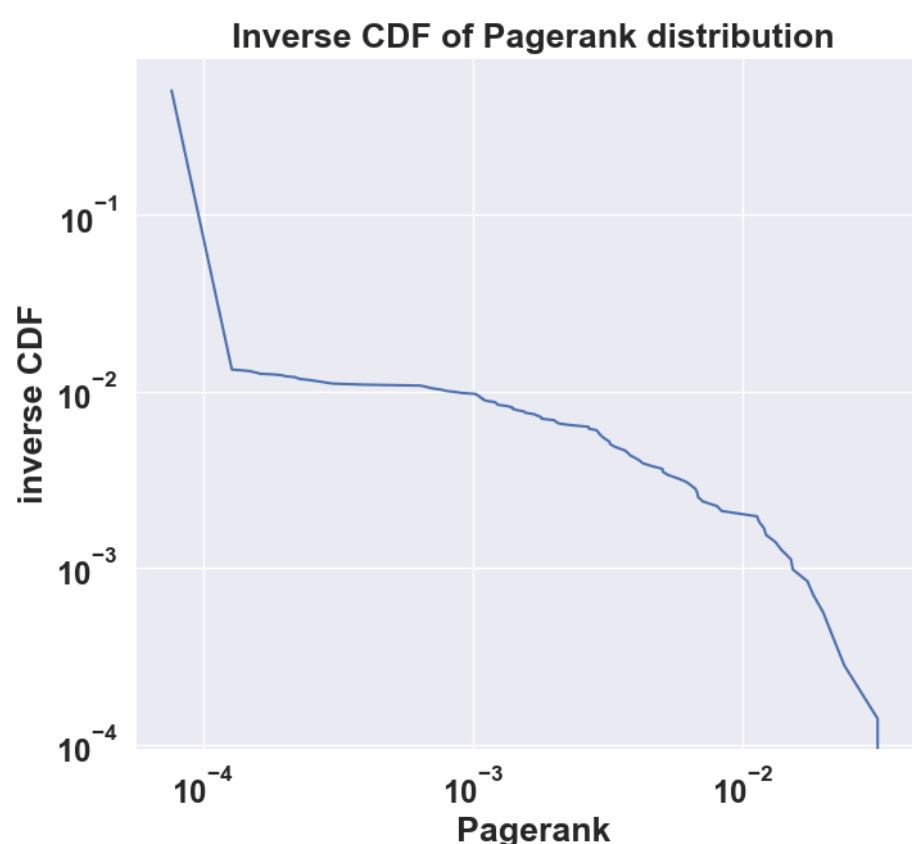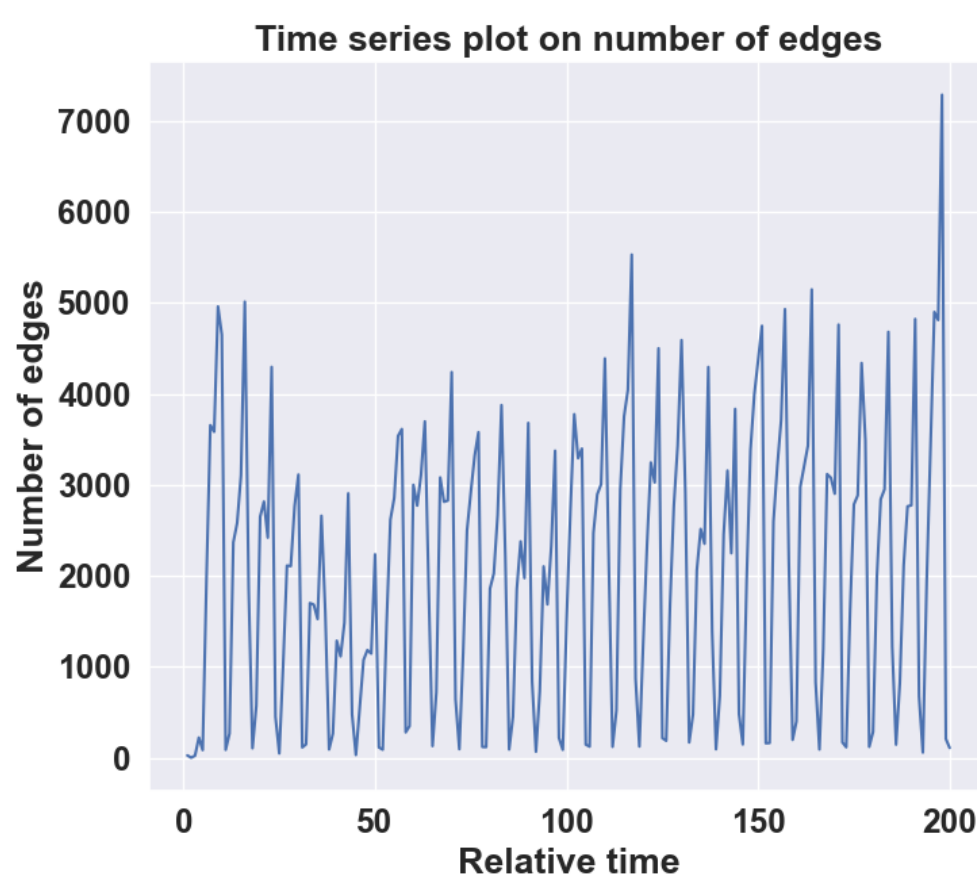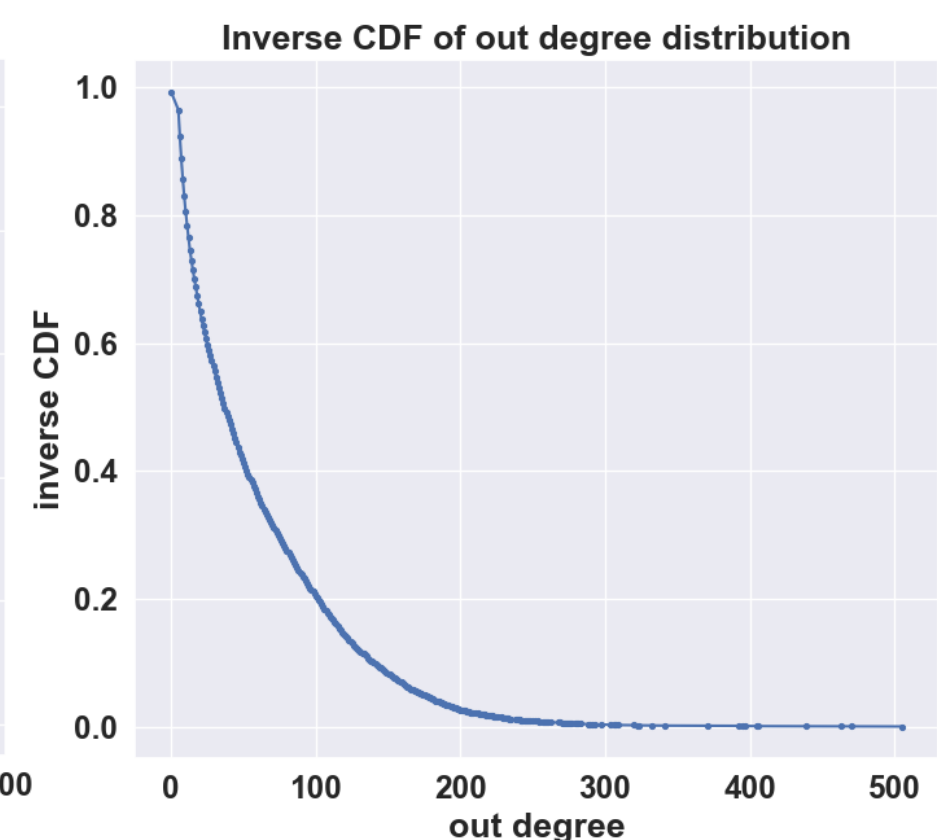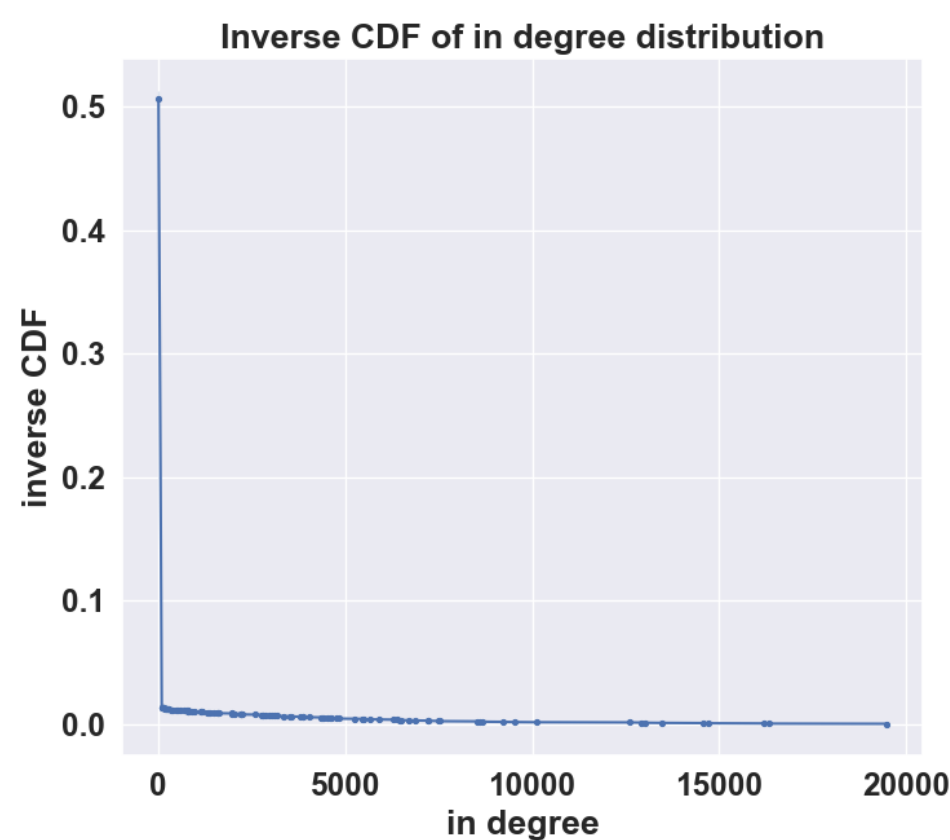
3. Uses

Primary intended uses: Predicting if an edge is the last action before students dropout. The evaluation metric is AUC.

Other uses: Studying students' behavior on online course platforms.

4. Network Statistics

Table 1 Point Statistics

Size of students, objects	7047, 97	Avg. triangle count	0
Size of edges	411749	Avg. clustering coefficient	0
Avg. degree	57.6	Assortativity	-0.24
% in LCC	100%	Algebraic connectivity	1.0
Max k-core	73	Spectral Radius	0
(Tail) power law exponent	N/A		

Inverse CDF of in degree distribution

Inverse CDF of out degree distribution

Time series plot on number of edges

Inverse CDF of Pagerank distribution

Top-20 singular values

Frequency of label

[1] Kumar, Srijan, Xikun Zhang, and Jure Leskovec. "Predicting dynamic embedding trajectory in temporal interaction networks." *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019.

[2] https://www.biendata.xyz/kddcup2015/data//