CPNs For Understanding Academic Curricula

Jay Gupta Adam Boswell

Stanford University

Introduction

This project concerns itself with the Stanford Course Prequisite Network (CPN) during the 2023 and 2024 academic year. Understanding the relationship between courses is of great importance to a University's educational mission. We employ a range of supervised and unsupervised methods to understand and ultimately and inform the Stanford educational structure.

Data Collection

Data was collected via the ExploreCourses API. The API readily provided the following features: subject, title, description, repeatable, grading basis, units min, units max, final exam, and academic group. Notably, it did not list pre-requisite courses. In effect, we inferred these pre-requisites via the text description.

Results for CS 109: Introduction to Probability for Computer Scientists — 1 courses

CS 109: Introduction to Probability for Computer Scientists

Topics include: counting and combinatorics, random variables, conditional probability, independence, distributions, expectation, point estimation, and limit theorems. Applications of probability in computer science including machine learning and the use of probability in the analysis of algorithms. Prerequisites: 103, 106B or X, multivariate calculus at the level of MATH 51 or CME 100 or equivalent.

Terms: Aut, Win, Spr | Units: 3-5 | UG Reqs: GER:DB-EngrAppSci, WAY-AQR, WAY-FR

Instructors: ; Piech, C. (PI); Sahami, M (PI); Agrawal, A. (TA); Apostolatos, A. (TA); Chen, M. (TA); Eric, M (TA); Goodman, I. (TA); Haaland, C. (TA); Hsu, I. (TA); Lin, P. (TA); Luo, A. (TA); Martinez, P. (TA); Nigam, P. (TA); Puranik, A. (TA); Rao, I. (TA); Salloum, J. (TA); Shapiro, J. (TA); Sun, Y. (TA); Wang, D. (TA); Weiler, A. (TA)

Network Properties

In theory, CPNs are classified as Directed Acyclic Graphs (DAGs) though our CPNs includes erroenous cycles due to difficulty processing the description text. Empirically, CPNs have a large connected component denoted G. They also obey the familiar rules of real world networks. For one, they are Scale Free meaning their degree distributions follow a power law.

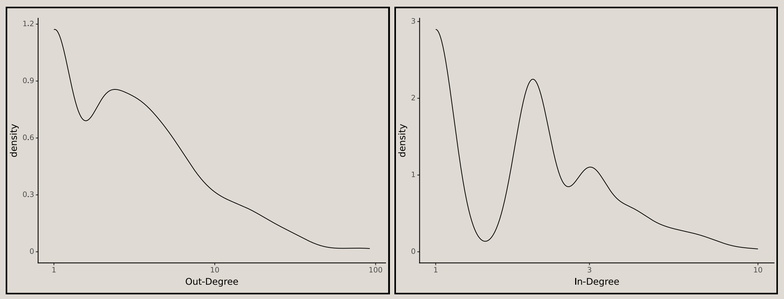

Figure 1. In-Degree and Out-Degree Distributions on Log-Scale

The in-degree is markedly lower than the out-degree suggesting that courses generally require few pre-requisites but open up many other courses in the network

Neighborhood Detection

We ran the KMeans Algorithm to detect clusters based on course features. Text was passed via TFIDF Vectorizer. K was selected using the Elbow Method.

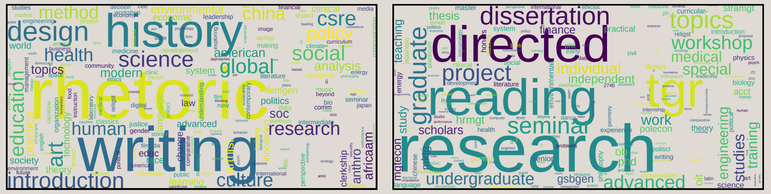

Figure 2. WordCloud of Title Text for Cluster 0 and 2

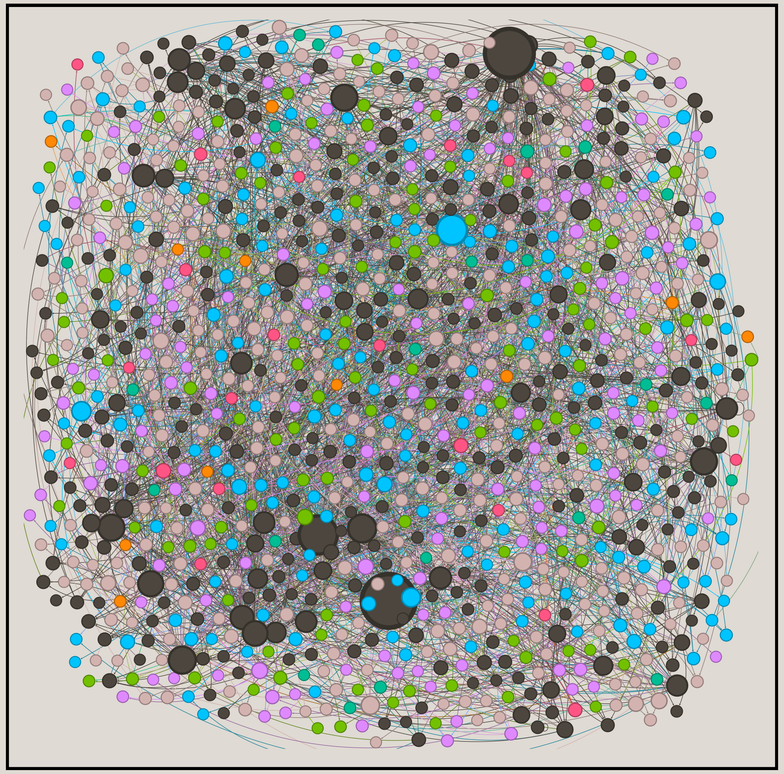

Figure 3. LCC of the Stanford CPN

Cluster 0 Cluster 1 Cluster 2 Cluster 3
Cluster 4 Cluster 5 Cluster 6 Cluster 7

Identifying Bridge Courses

The Betweeness Centrality of node v is the probability that v will appear in the shortest path of two nodes s and t for all such pairs.

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

Intuitively, nodes with large betweeness centrality scores serve as bridges between sources and sinks.

Course	Title	Centrality
MATH 21	Calculus	513
CS 229	Machine Learning	482
PHYSICS 130	Quantum Mechanics	437
CME 100	Vector Calculus	371
MATH 51	Linear Algebra	335

Table 1. Top 5 Bridge Courses in Stanford CPN

Link Prediction

Link Prediction refers to a class of problems that infer missing links (i.e., prerequisites) from an undirected graph.

Source	Target
MATSCI 208	MATSCI 251
CS 173A	CS 246H
BIO 273A	CHEMENG 274
ECON 44	PUBLPOL 19Q
OCEANS 161	BIO 245

Table 2. Top 5 Missing Links in Stanford CPN by Jaccard Similarity

Conclusion

Our analysis of the Stanford CPN visualized how courses are grouped, identified key bridge courses critical for knowledge flow, and detected potential missing links. This work should help enhance future course planning and navigation at Stanford.

References

[1] Preston R. Aldrich. The curriculum prerequisite network: a tool for visualizing and analyzing academic curricula. ArXiv, abs/1408.5340, 2014. URL https://api.semanticscholar.org/CorpusID:8419234.

[2] Pavlos Stavrinides and Konstantin M. Zuev. Course-prerequisite networks for analyzing and understanding academic curricula. Applied Network Science, 8(1), April 2023. ISSN 2364-8228. doi:10.1007/s41109-023-00543-w. URL http://dx.doi.org/10.1007/s41109-023-00543-w.