ICME-15
7-14 July 2024 • ICC Sydney, Australia
Come and be counted

Konstantin Zuev, Teaching Professor of Computing and Mathematical Sciences, California Institute of Technology, http://www.its.caltech.edu/~zuev/

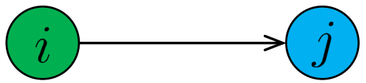
Course-Prerequisite Networks for Analysis and Visualization of University Curricula

Abstract: An academic curriculum is a complex system of courses and interactions between them that lies at the heart of a university and underlies its educational mission. Understanding complex academic curricula is of great importance for providing high-quality education at the tertiary level. This work is dedicated to the study of course-prerequisite networks (CPNs), where nodes represent courses and directed links represent the prerequisite relationships between them. The main goal of CPNs is to represent the flow of knowledge in academic curricula and serve as a key tool for visualizing, analyzing, and optimizing curricula. CPNs can be used by students, faculty, and administrators for detecting important courses, navigating curricula, allocating teaching resources, increasing interdisciplinary interactions between departments, and enhancing the overall students' learning experience. Moreover, CPNs allow to quantify the strength of knowledge flow between mathematics and other areas of study. The proposed methodology can be used for the analysis of any curricula at the tertiary level, and it is illustrated with a network of courses taught at the California Institute of Technology. The network data analyzed in this work and the corresponding publication are publicly available.

Course-Prerequisite Networks

A course-prerequisite network or CPN is a directed network.

- Nodes represent courses
- Links represent course-prerequisite relationships

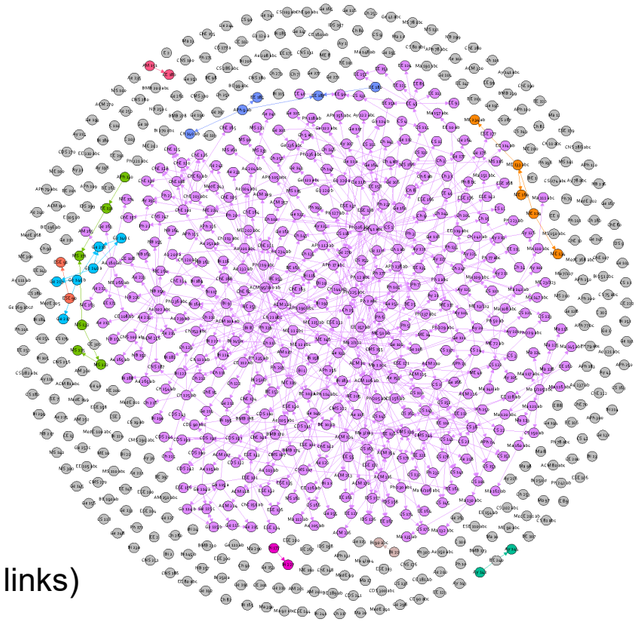


Course i is a **pre**requisite for Course j

Course j is a **post**requisite for Course i

Caltech CPN

- 771 nodes, 772 links
- Gray nodes are isolated nodes
- Groups of colored nodes represent connected components
- There are 10 connected components
- Group of purple nodes is the largest connected component
- The largest connected component (LCC) is the most interesting and nontrivial part of CPN, which represents its complex function and structure
- LCC: 436 nodes (57% of all nodes) and 747 links (97% of all links)



TSG 3.2: Mathematics Education at Tertiary Level

Email: kostia@caltech.edu

The Caltech CPN data is publicly available in the GitHub repository https://github.com/pstavrin/Course-Prerequisite-Networks

Centrality Measures

One of the most interesting and intriguing questions about a curriculum is the following: "Which are the most important courses?" Knowing the most important courses could help to better allocate university resources to provide students with better experiences in these courses and inform students about these courses, so that they can pay special attention to them.

There are different ways to define the "course importance". Here, we will consider three centrality measures:

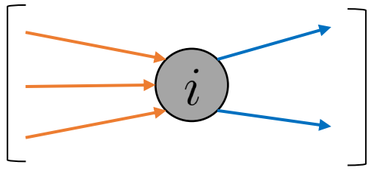
- Out-Degree: measures how fundamental a course is
- PageRank: measures how fundamental a course is and favors more introductory courses
- Betweenness: identifies intermediate-level courses that serve as critical bridges between less and more advanced courses

Degree

In CPNs, nodes have two kinds of degree:

- In-degree of course i is the number of prerequisites i has
- Out-degree of course i is the number of postrequisites i has

$k_{\text{in}}(i)$
Measures how specialized the course is



$k_{\text{out}}(i)$
Measures how fundamental the course is

- The in- and out-degrees of nodes are negatively correlated
- The out-degree is a more suitable measure of course importance

PageRank

The PageRank centrality is a key ingredient of the Google search engine. The key idea is that the importance of a course i is not only about how many postrequisites i has, but also about how important the postrequisites of i are.

Mathematically, the vector π of PageRank centralities of CPN nodes is given by

$$\pi = \left(I - \alpha A K_{in}^{-1}\right)^{-1}\mathbf{1}$$

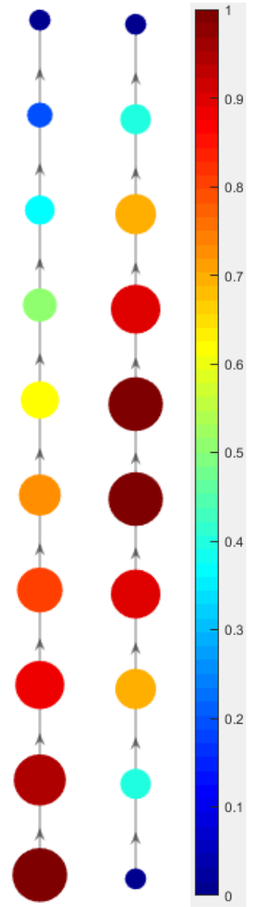
where I is the identity matrix, A is the adjacency matrix of the CPN, $\alpha \in (0,1)$ is the damping factor, $\mathbf{1}$ is the $n \times 1$ vector of ones, and K_{in} is the diagonal matrix with diagonal elements $K_{ii} = \max\{k_{in}(i), 1\}$.

Betweenness

The betweenness centrality quantifies the degree to which a node is located between other nodes. Mathematically, betweenness $\beta(i)$ of node i is defined as

$$\beta(i) = \sum_{s,t \neq i} \frac{\sigma(s,t|i)}{\sigma(s,t)}$$

where $\sigma(s,t)$ is the total number of shortest paths from s to t, and $\sigma(s,t|i)$ is the number of shortest paths from s to t that pass through i.

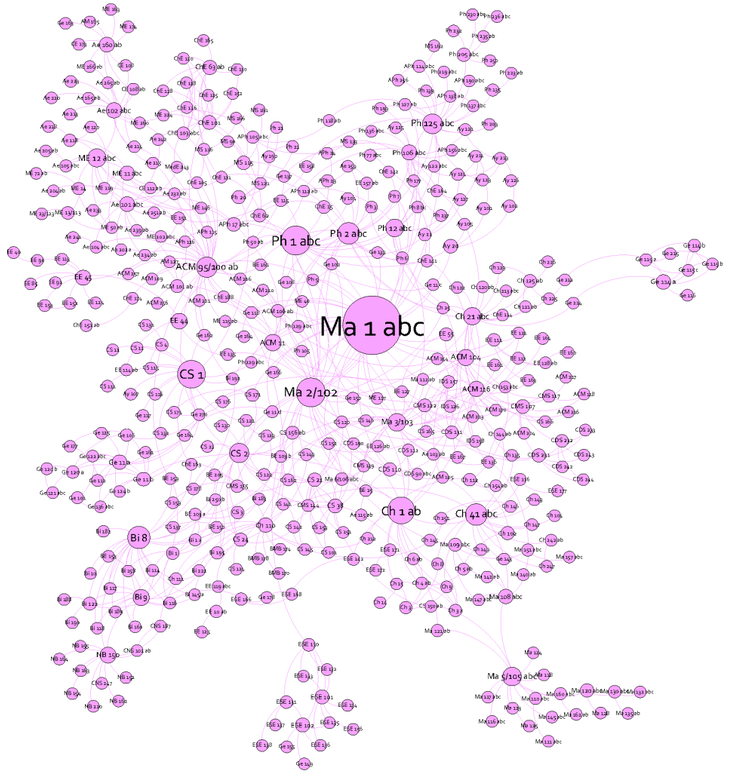


PageRank Betweenness

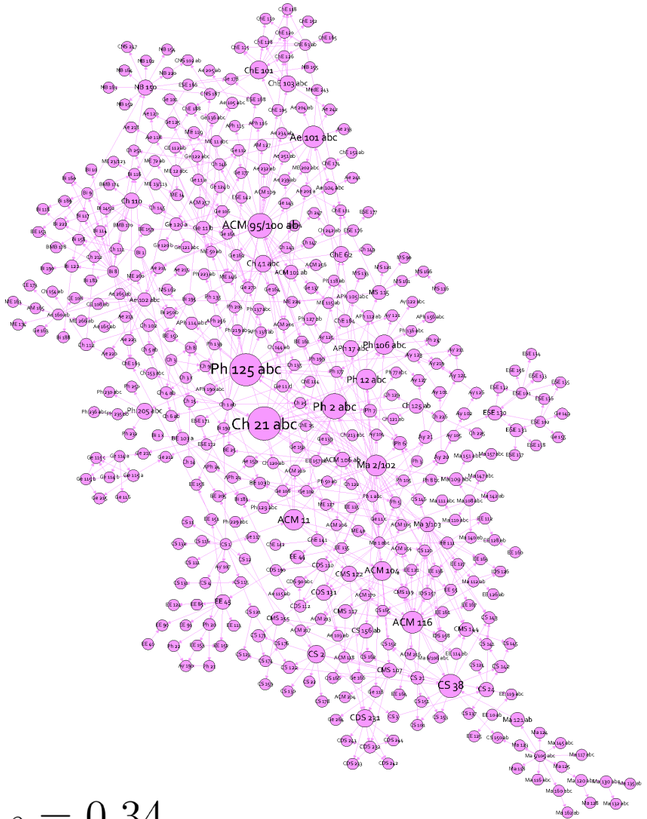
Out-Degree Top 5		PageRank Top 5		Betweenness Top 5	
Course	**Title**	**Course**	**Title**	**Course**	**Title**
Ma2	Diff. Equations	Ma1	Freshman Math	Ch21	Physical Chem
ACM95	Applied Math	Ma2	Diff. Equations	Ph125	Quantum Mech
Ma1	Freshman Math	Ph1	Freshman Physics	Ph2	Sophomore Physics
Bi8	Molecular Bio	CS1	Into Programming	ACM95	Applied Math
Ma3	Intro Probability	Ch1	Freshman Chem	CS38	Algorithms

Details: Stavrinides, P., & Zuev, K.M. (2023). Course-prerequisite networks for analyzing and understanding academic curricula. *Applied Network Science, 8, 9* https://doi.org/10.1007/s41109-023-00543-w

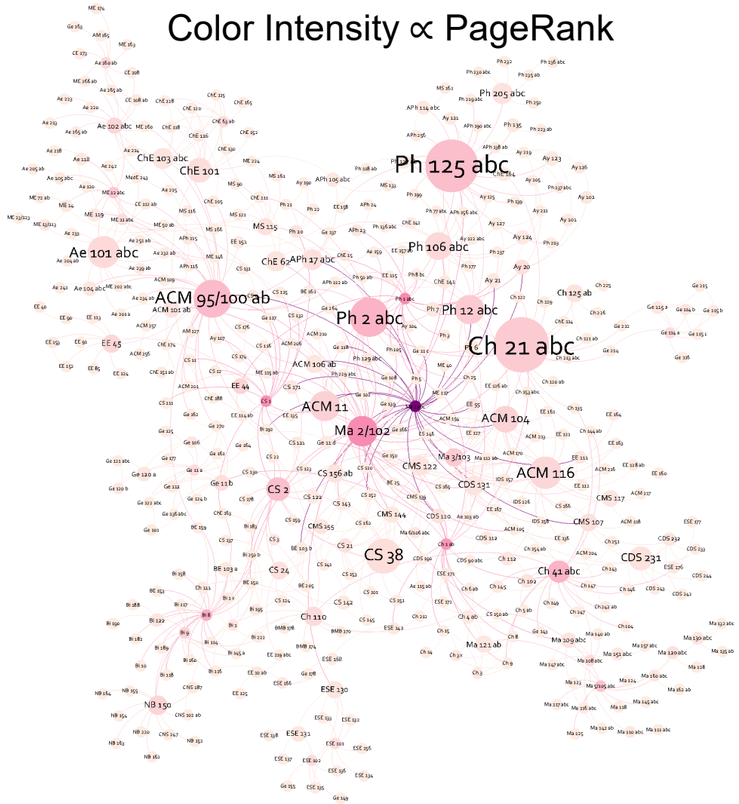
Node Size ∝ PageRank



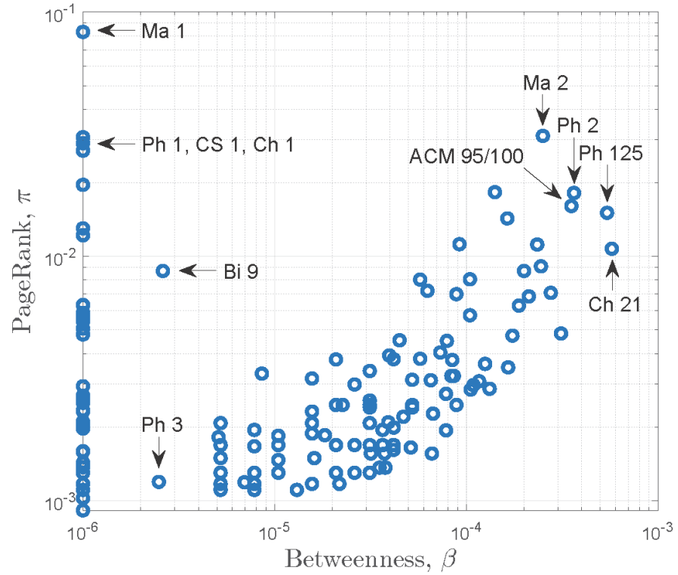
Node Size ∝ Betweenness



Node Size ∝ Betweenness
Color Intensity ∝ PageRank



PageRank and betweenness are positively correlated $\rho_{\pi,\beta} = 0.34$



Courses which are important with respect to one of the measures tend to be also important with respect to the other. Courses that are ranked highly by both PageRank and betweenness play a vital role in the university curriculum. Students should be advised to pay special attention to these courses. Maximizing the quality of these courses is the key step in improving the whole curriculum. Administrators should, therefore, prioritize enhancing the quality of courses with high PageRank and betweenness scores by ensuring that

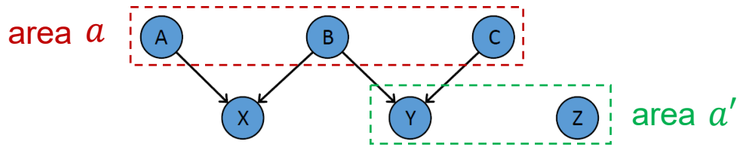
- These courses are taught well (for example, by hiring dedicated teaching faculty) and
- Students are provided with all necessary resources, such as office hours, recitation sessions, and tutoring support.

Interdependence Analysis

- Provides a bird's eye view of the whole CPN
- Quantifies the strength of knowledge flow between different areas of study
- Identifies the most influential and interdisciplinary areas

The interdependence between two areas of study is defined as follows:

- $\mathcal{A}$ is the set of all areas of study (Math, Physics, Chemistry, etc.)
- $\mathcal{C}_a$ is the set of all courses in area $a \in \mathcal{A}$
- $|\mathcal{C}_a|$ is the number of courses in $\mathcal{C}_a$
- $\mathcal{C}_a^+$ is the multiset of all postrequisites of $\mathcal{C}_a$
- $\mathcal{C}_{aa'}^+$ is the multisubset of $\mathcal{C}_a^+$ consisting of courses whose area is $a' \in \mathcal{A}$



$$\mathcal{C}_a = \{A, B, C\}$$
$$\mathcal{C}_a^+ = \{X, X, Y, Y\}$$
$$\mathcal{C}_{aa'}^+ = \{Y, Y\}$$

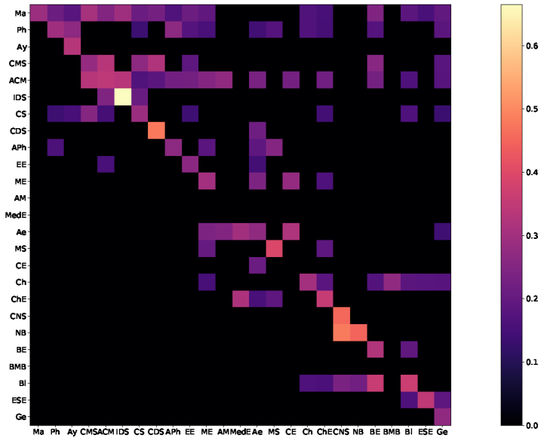
- Measure of dependence of area a' on area a:

$$\underbrace{R_{aa'}}_{\text{Inverse Shanon Information}} = \left(-\log \underbrace{\frac{|\mathcal{C}_{aa'}^+|}{|\mathcal{C}_a||\mathcal{C}_{a'}|}}_{\mathbb{P}(X_a \to X_{a'})} \right)^{-1}$$

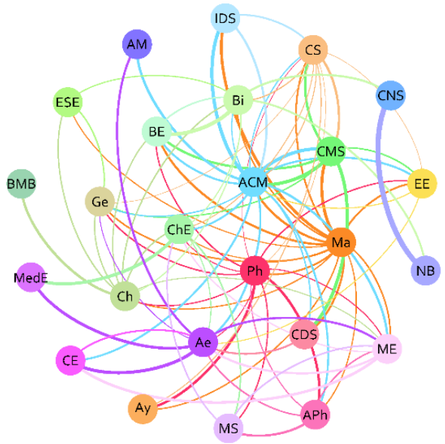
is large $\iff$ dependence of a' on a is strong

probability of a link from a course randomly chosen from area a to a course randomly chosen from a'

Heatmap of Interdependence



Network Representation



- Most influential areas: Mathematics, Applied Math, Physics, Computer Science
- Most interdisciplinary areas: Chemical Engineering, Geology, Aerospace
- Out-degree top 3 areas: Mathematics, Applied Math, Physics
- PageRank top 3 areas: Mathematics, Applied Math, Computer Sciece
- Betweeness top 3 areas: Biology, Applied Math, Aerospace

Summary

This work introduces course-prerequisite networks and shows that a CPN is an indispensable tool for summarizing, visualizing, and analyzing an academic curriculum. It can help to better understand and revamp the curriculum, detect important courses, improve existing and create new courses, meaningfully allocate teaching resources, increase interdisciplinary interactions between various university units, and enhance the overall student learning experience.

Reference: Stavrinides, P., & Zuev, K.M. (2023). Course-prerequisite networks for analyzing and understanding academic curricula. *Applied Network Science, 8, 9* https://doi.org/10.1007/s41109-023-00543-w

Thank you very much for your attention!

Konstantin Zuev
Teaching Professor of Computing and Mathematical Sciences
California Institute of Technology
Web: http://www.its.caltech.edu/~zuev/
Email: kostia@caltech.edu