Automating Course Articulation:
A Deep Metric Learning Framework Using Public Data

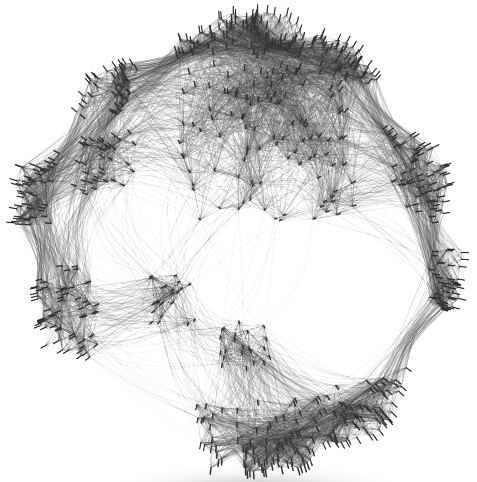
Mark S. Kim

San Francisco State University
Department of Data Science and Artificial Intelligence

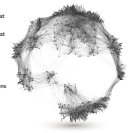
July 9, 2025

The Problem: The Transfer Maze

- The process for determining course equivalency, or **articulation**, is a formidable, largely manual process that creates significant barriers for students [8].
- In California's public system alone, articulation officers at **149 individual campuses** manually negotiate and update agreements [2, 7, 5, 4].
- This task of "bleak combinatorics" is inefficient, slow, and inherently intractable, struggling to keep pace with the needs of a vast and mobile student body [8].
- This is not a niche issue; transferring between institutions has become a normative part of the modern student's academic journey [1].



The Problem: The Transfer Maze
- The process for determining course equivalency, or **articulation**, is a formidable, largely manual process that creates significant barriers for students [8].
- In California's public system alone, articulation officers at **149 individual campuses** manually negotiate and update agreements [2, 7, 5, 4].
- This task of "bleak combinatorics" is inefficient, slow, and inherently intractable, struggling to keep pace with the needs of a vast and mobile student body [8].
- This is not a niche issue; transferring between institutions has become a normative part of the modern student's academic journey [1].

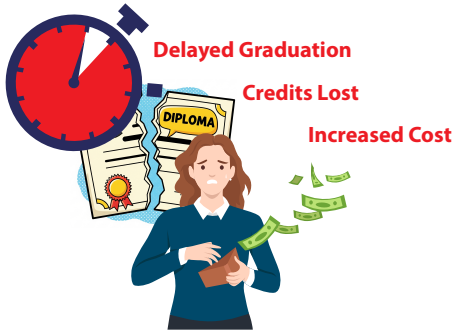


The High Cost to Students & Institutions

Consequences of an Inefficient System

The administrative friction of the transfer process creates a cascade of negative consequences that fall almost entirely on students.

- **Significant Credit Loss:** On average, transfer students lose an estimated **43%** of their academic credits [11, 1].
- **Increased Time-to-Degree:** Lost credits directly delay graduation and postpone entry into the workforce [11].
- **Greater Financial Burden:** Repeating courses increases tuition costs and can exhaust a student's financial aid eligibility [11, 3].
- **Reduced Student Persistence:** The frustration of the process contributes to lower graduation rates for transfer students compared to their non-transfer peers [9].



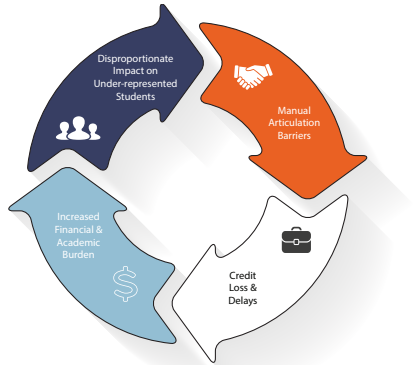
1. This high rate of credit loss often forces students to repeat courses for which they have already received a passing grade.
2. This also increases their overall time in the educational system.
3. This means a process often undertaken to save money can paradoxically result in a greater overall financial commitment.
4. The frustration also has a measurable impact on student morale.

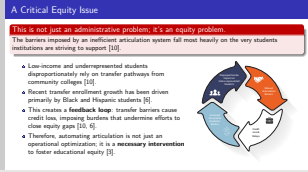
A Critical Equity Issue

This is not just an administrative problem; it's an equity problem.

The barriers imposed by an inefficient articulation system fall most heavily on the very students institutions are striving to support [10].

- Low-income and underrepresented students disproportionately rely on transfer pathways from community colleges [10].
- Recent transfer enrollment growth has been driven primarily by Black and Hispanic students [6].
- This creates a **feedback loop**: transfer barriers cause credit loss, imposing burdens that undermine efforts to close equity gaps [10, 6].
- Therefore, automating articulation is not just an operational optimization; it is a **necessary intervention** to foster educational equity [3].





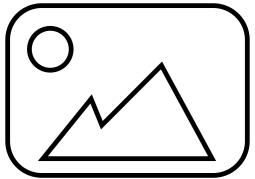
1. These are the very student populations that institutions are striving to support, making transfer efficiency a critical equity lever.
2. The National Student Clearinghouse reported this trend in Fall 2023, highlighting the increasing diversity of the transfer population.
3. This troubling loop means the manual system directly counteracts institutional goals of supporting underrepresented students.

The Goal & My Contribution

The Goal

To develop and validate a novel framework that automates course articulation using only publicly available data.

The resulting system must be:



Primary Contributions

The Goal & My Contribution
The Goal
To develop and validate a novel framework that automates course articulation using only publicly available data.
The resulting system must be:
Primary Contributions

1. For our framework, we achieved F1-scores exceeding 0.97 on the held-out test set.
2. The composite vector combines the element-wise difference of course embeddings with their cosine similarity, which was shown to be a superior feature set in ablation studies.
3. Specifically, our approach avoids using sensitive student enrollment data and the high computational cost and opacity of direct LLM classification.

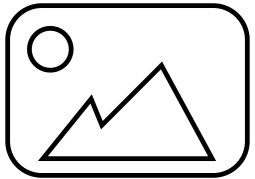
The Goal & My Contribution

The Goal

To develop and validate a novel framework that automates course articulation using only publicly available data.

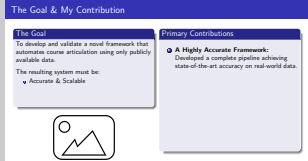
The resulting system must be:

- Accurate & Scalable



Primary Contributions

1. **A Highly Accurate Framework:**
   Developed a complete pipeline achieving state-of-the-art accuracy on real-world data.



1. For our framework, we achieved F1-scores exceeding 0.97 on the held-out test set.
2. The composite vector combines the element-wise difference of course embeddings with their cosine similarity, which was shown to be a superior feature set in ablation studies.
3. Specifically, our approach avoids using sensitive student enrollment data and the high computational cost and opacity of direct LLM classification.

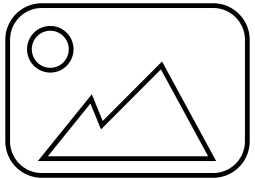
The Goal & My Contribution

The Goal

To develop and validate a novel framework that automates course articulation using only publicly available data.

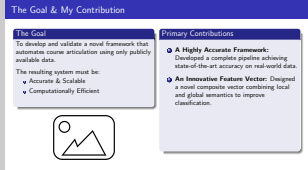
The resulting system must be:

- Accurate & Scalable
- Computationally Efficient



Primary Contributions

1. **A Highly Accurate Framework:** Developed a complete pipeline achieving state-of-the-art accuracy on real-world data.
2. **An Innovative Feature Vector:** Designed a novel composite vector combining local and global semantics to improve classification.



1. For our framework, we achieved F1-scores exceeding 0.97 on the held-out test set.
2. The composite vector combines the element-wise difference of course embeddings with their cosine similarity, which was shown to be a superior feature set in ablation studies.
3. Specifically, our approach avoids using sensitive student enrollment data and the high computational cost and opacity of direct LLM classification.

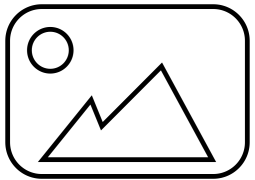
The Goal & My Contribution

The Goal

To develop and validate a novel framework that automates course articulation using only publicly available data.

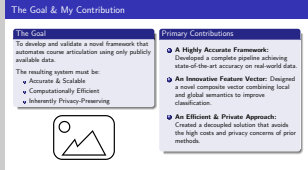
The resulting system must be:

- Accurate & Scalable
- Computationally Efficient
- Inherently Privacy-Preserving



Primary Contributions

1. **A Highly Accurate Framework:** Developed a complete pipeline achieving state-of-the-art accuracy on real-world data.
2. **An Innovative Feature Vector:** Designed a novel composite vector combining local and global semantics to improve classification.
3. **An Efficient & Private Approach:** Created a decoupled solution that avoids the high costs and privacy concerns of prior methods.



1. For our framework, we achieved F1-scores exceeding 0.97 on the held-out test set.
2. The composite vector combines the element-wise difference of course embeddings with their cosine similarity, which was shown to be a superior feature set in ablation studies.
3. Specifically, our approach avoids using sensitive student enrollment data and the high computational cost and opacity of direct LLM classification.