

Topological-and-Textual Data Analysis of firms' patent portfolios and financial performances

Faculty Member: [Mitsuru Igami](#)

Proposal Description: The rate and direction of innovation is one of the central topics in economics for 60 years, but the “direction” part of the question has rarely been studied despite its potential importance. This project seeks to describe and characterize the dynamic evolution of firms’ patenting activities and financial performances by adapting a new tool from computational topology (the Mapper algorithm, a Topological Data Analysis method), which is a frontier method in applied mathematics and the analysis of high-dimensional data, as well as Natural Language Processing (textual analysis) methods.

*My 20-minute YouTube video presentation is available here:

<https://youtu.be/0LQpJiecCvw>(link is external)

Requisite Skills and Qualifications: The required skills are (1) proficiency in handling text data in Python or R, (2) solid understanding of econometrics (e.g., can you explain why “exogenous variations in data” are crucial for credible “causal inference” and “empirical strategy?”), (3) passion and patience in handling real-world data [even a commercial-grade financial database like COMPUSTAT is pretty crazy and disorganized, so it’s our job to make it useful for rigorous academic research!], and (4) willingness to communicate and cooperate with me on a regular basis.

Don’t worry; You don’t have to be a top coder or mathematician/topologist to assist this project. That said, I do need highly motivated self-starters—someone who can improvise and teach themselves new skills as such necessities arise. If that sounds like you, please apply!

Award: Erkin Asci

Adrian Olsen

Tobin Application Link: [Tobin Application](#)

Project Type: Tobin RA

Project Year: 2022

Term: Spring 2022

Source URL: <https://economics.yale.edu/undergraduate/tobin-ra/spring-2022/topological-and-textual-data-analysis-firms-patent-portfolios-and>