2024 Abstracts

Columbia Math Undergraduate Summer Research (CMUSR)

Project lead: Rostislav Akhmechet
Mentor: Alvaro Martinez Ruiz

- **Khovanov homology and quantum error-correcting codes**

  Milena Harned, Pranav Konda, Felix Liu, Nikhil Mudumbi, Eric Shao, Tony Xiao

  Abstract: Error-correcting codes for quantum computing are crucial to address the fundamental problem of communication in the presence of noise and imperfections. Audoux used Khovanov homology to define families of quantum codes with desirable properties. In this project, we explored Khovanov homology and some of its many extensions, namely annular and sl(3) homology, in order to generate new families of quantum codes and to establish several properties of codes that arise in this way.

Columbia Summer Undergraduate Research Experiences in Mathematical Modeling (CSUREMM)

Project lead: George Dragomir
Mentor: Vihan Pandey

- **Impact of Subway Centrality on Manhattan High School Performance**

  Caroline Smyth, Kathy Xu, Sandra Zelen

  Abstract: The New York City Metropolitan Transit Authority (MTA) subway system, the largest in North America, serves over three million riders daily. This study explores its impact on public high school performance in Manhattan, hypothesizing that centrally located schools perform better academically. Using graph theory, the MTA system was modeled with stations as nodes and subway lines as edges. Centrality measures for each station were calculated, and public high schools were mapped to their nearest stations to estimate centrality. Initial multiple linear regression analyses of performance metrics against centrality measures yielded low R-squared values, prompting the use of principal component analysis (PCA). PCA identified three principal components: neighbor-driven centrality, distance-driven centrality, and school performance. Clustering these components revealed three groups of schools: low centrality-high performance, low centrality-low performance, and high centrality-medium performance. The study found distinct dominant performance metrics for each cluster: student achievement for low centrality-high performance, strong family-community ties for low centrality-low performance, and rigorous instruction for high centrality-medium performance. Despite different centrality measures, the clusters showed consistent dominant performance metrics.

- **Unemployment: How Consumer Expectations may Affect Teen Workers**

  Elizabeth Casey, Alexandra Ehlinger

  Abstract: This study investigates the predictive relationship between consumer expectations and teen unemployment in the United States. Teens aged 16-19 are a particularly vulnerable segment of the labor force, with their unemployment rates showing more than four times the variance of the aggregate unemployment rate and typically being two to four times higher. Leveraging the OECD's Consumer Confidence Index (CCI)

and unemployment data from the Bureau of Labor Statistics, we model this relationship using three distinct approaches. First, Granger causality tests are conducted to determine if a predictive relationship exists. Second, a Gaussian Generalized Linear Model (GLM) is employed to compare the CCI's predictive power for both teen and aggregate unemployment rates. Finally, the CCI is included as an exogenous regressor in an Autoregressive Integrated Moving Average (ARIMA) model to assess its impact on forecast accuracy based on lagged unemployment rates. Our findings indicate that the CCI is a valuable predictor for both teen and aggregate unemployment rates. Notably, when used alone in a Gaussian GLM, the CCI shows higher predictive accuracy for teen unemployment compared to aggregate unemployment. However, incorporating the CCI as an exogenous regressor in an ARIMA model does not improve forecast accuracy. These results suggest that while the CCI effectively predicts teen unemployment, it does not enhance the accuracy of models that rely on lagged unemployment data for predictions.

- **Controlling Your Attitude is What Gets You To the Stars: Development of an Attitude Determination and Control System for CubeSats via Extended Kalman Filter**

  Alena Chan, Ruimian Zheng

  Abstract: CubeSat attitude control depends on accurate attitude estimates, which can be disrupted by errors in measurement and control systems. This paper presents an adaptive mathematical model that uses the Extended Kalman Filter (EKF) to improve the precision of CubeSat attitude determination and control. We offer detailed instructions for designing, implementing, and simulating a CubeSat Attitude Determination and Control System (ADCS) that includes magnetometers, gyroscopes, and reaction wheels, making it easier to plan CubeSat missions. Our method involves creating and applying a control algorithm that adjusts the satellite's orientation to target a specific celestial body based on predicted attitude data. We also develop a testing setup in MATLAB Simulink, which simultaneously simulates satellite orbits and measurement data in response to control commands. Our simulations evaluates the effectiveness of different EKF approaches in measuring the satellite's attitude and rotation rate. Our results show that the combined Additive and Multiplicative Kalman Filter model achieves a 58.23% reduction in attitude estimation error compared to a model without filtering.

- **Reimagining Manhattan: A Mathematical Modeling Approach to Superblock Developmentt**

  Amy Chen, Cionnie Pineda, Shiva Yeshlur

  Abstract: Introducing superblocks as an urban design and traffic planning strategy in Barcelona has transformed the city's neighborhoods. By creating pedestrian and cyclist-oriented inner-block zones, this approach has reduced air and noise pollution, congestion, and other environmental issues. Current research falls into two categories: qualitative urban planning studies on real-world superblock implementation and quantitative models based on graph optimization and network theory for superblock partitioning. Qualitative studies offer insights into region-specific governance but have limited applicability elsewhere. Quantitative models simulate superblock impacts on traffic networks but often overlook community structures that affect feasibility. Our approach employs a graph-theoretic model that incorporates Point-of-Interest (POI) data into a multi-level, hierarchical partitioning strategy using Leiden Community Detection. We represent Manhattan's grid as a weighted graph with intersections as nodes and road segments as edges. By integrating POI scores from Open Street Maps (OSM) with edge lengths, we create a geospatially- and POI-aware clustering model. We then simulate a superblock growth strategy for Manhattan, using hierarchical partitioning to generate an initial set of superblocks, termed our "pilot program."

- **Rebalancing for Equity Across the Citi Bike Network**

  Timothy Barron, Jarett Reimers, Hyoungjoon Yoo

  Abstract: Bikeshare networks often face challenges with rebalancing, which involves relocating bikes between stations with surplus bikes and those with deficits to ensure adequate availability of bikes and empty docks for all customers. This paper introduces a mathematical model aimed at optimizing rebalancing strategies to address Citi Bike's suboptimal performance in marginalized communities. By analyzing historical data from over 4 million Citi Bike trips, we develop a continuous-time Markov chain model to simulate traffic flow throughout the day. Using this simulation, we apply an iterative algorithm to determine optimal fill levels that minimize customer dissatisfaction across network clusters. Using the optimal fill levels, we employ an integer programming approach to design optimal truck routes for overnight rebalancing. Our findings show that implementing our rebalancing strategies can significantly enhance Citi Bike's performance in socially marginalized areas. The results offer practical and actionable insights for the efficient rebalancing of bikeshare systems, providing a robust framework for future research and potential implementation by bikeshare operators.

- **Predicting M&A Success in the US Retail Industry Using Machine Learning: A Dual Analysis of Financial Metrics and Stock Performance**

  Mergers and Acquisitions (M&A) in the US retail industry often aim to enhance shareholder value, yet the failure rate is as high as 70%. This study provides a dual analysis to improve M&A success prediction. The first part evaluates financial metrics, notably TIC/EBITDA ratios, finding higher ratios are linked to successful deals. The second part examines stock performance, showing pre-announcement stock prices significantly predict M&A outcomes. Integrating these metrics into machine learning models, our research offers a robust framework for M&A evaluation and actionable insights to refine future decision-making in the retail sector.

  Predicting M&A Success Based on Financial Metrics Analysis

  Kathy Chen, Alina Wang

  Abstract: This study develops logistic regression and neural network models to predict M&A outcomes in the US retail sector using 16 financial metrics from Bloomberg. The logistic regression model, with 66.6% accuracy, identifies target return on common equity and acquirer gross margin as key predictors. The neural network model, employing 12 hidden layers and resilient back-propagation, achieved 60% accuracy, addressing class imbalance and non-linear relationships. These models provide valuable insights for M&A decision-making, enhancing the selection of merger targets.

  Predicting M&A Success Based on Stock Performance Analysis

  Jacqueline Lyu, Claire Qu

  Abstract: This study analyzes M&A success determinants in the US retail industry by focusing on stock market performance. Using a dataset of 50 transactions from 2010 to 2020, we applied machine learning techniques, including k-means clustering, t-tests, and logistic regression. Our findings show that pre-announcement stock prices are significant predictors of M&A success. Logistic regression achieved 46.67% accuracy, while the neural network model reached 53.33%. These results provide valuable tools for pre-deal evaluation and post-deal monitoring, though limitations include a small sample size and exclusion of other factors such as market conditions and company health.