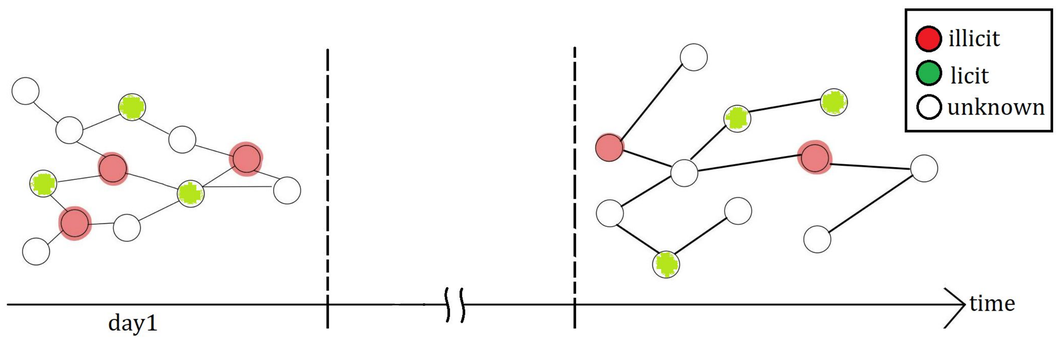
Anti-money laundering on Bitcoin
Data Analysis & Experiment on Various Models

CHUN-YEN LEE 李俊諺; ZHAO-QIAN YUAN 袁肇謙; CHENG-FU WENG 翁丞甫
109-1 Network & Multimedia Lab Final Project

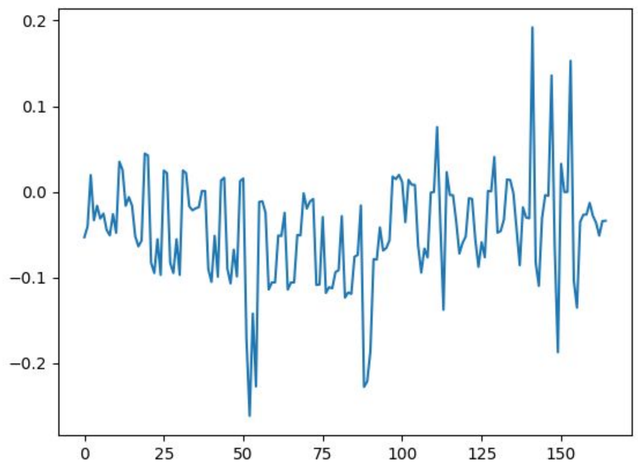
Introduction

Anti-money laundering (AML) regulations play a critical role in safeguarding financial systems, but bear high costs for institutions and drive financial exclusion for those on the socioeconomic and international margins. We contribute the analysis on Elliptic Data Set, to see if any feature highly correlated to illicit transaction.

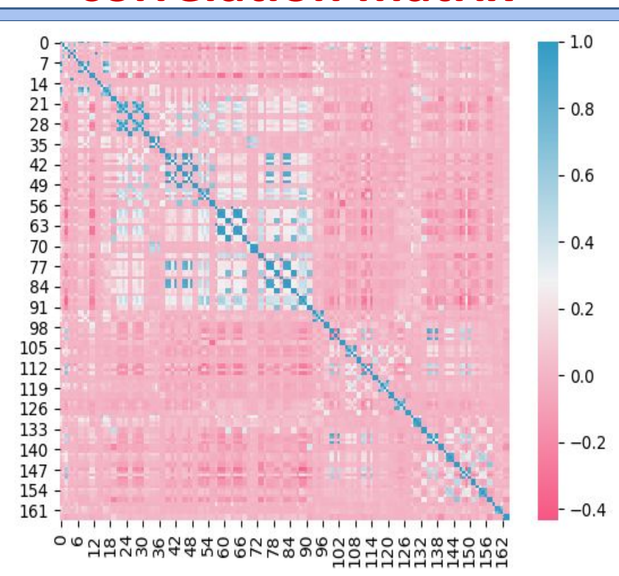
Data Structure



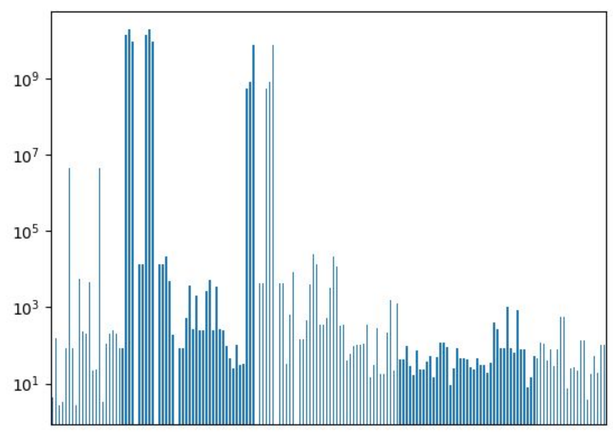
Feature Correlation with Output



Highly Correlated Features- correlation matrix



Highly Correlated Features- VIF



Experiment Replication

Method	Paper F_1	Self-produced F_1
Logistic Regression	0.481	0.529
RandomForest	0.788	0.784
MLP	0.653	0.665
degree_2 Logistic Regression		0.540
degree_2 Polynomial Regression		0.447
Linear Regression		0.257
Logistic RegressionCV		0.676
SVM		0.682

PCA

Method	F_1
Logistic Regression	0.477
RandomForest	0.637
MLP	0.689
degree_2 Logistic Regression	0.590
degree_2 Polynomial Regression	0.489
Linear Regression	0.265
Logistic RegressionCV	0.673
SVM	0.714

Feature selection

Method	PCA F_1	correlation F_1
Logistic Regression	0.502	0.492
RandomForest	0.613	0.779
MLP	0.667	0.283
degree_2 Logistic Regression	0.616	nan
degree_2 Polynomial Regression	0.533	0.640
Linear Regression	0.267	0.156
Logistic RegressionCV	0.660	0.496
SVM	0.714	0.672