*Roman Khamrin*

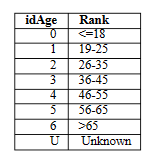
*Nikolai Suanov*

Fraud detection using graph models: banking fraud

There was chosen [dataset](https://www.kaggle.com/datasets/ealaxi/banksim1?resource=download) of synthetic data from a financial payment system. It is based on data of spanish bank and described in a research paper: BankSim: a bank payment simulation for fraud detection reasearch[[1]](#footnote-1). It includes the following variables: step, customer, age, gender, zipcodeOri, merchant, zipMerchant, category, amount, fraud.

First, “step” variable includes data about step of generation – each step represents a day of commercial activity.

Variable “customer” containts unique ID of customers, the same way organized column “merchant” – it contains unique ID of merchants.

Variable “age” is an ordinal and can be decoded according to the following table: 

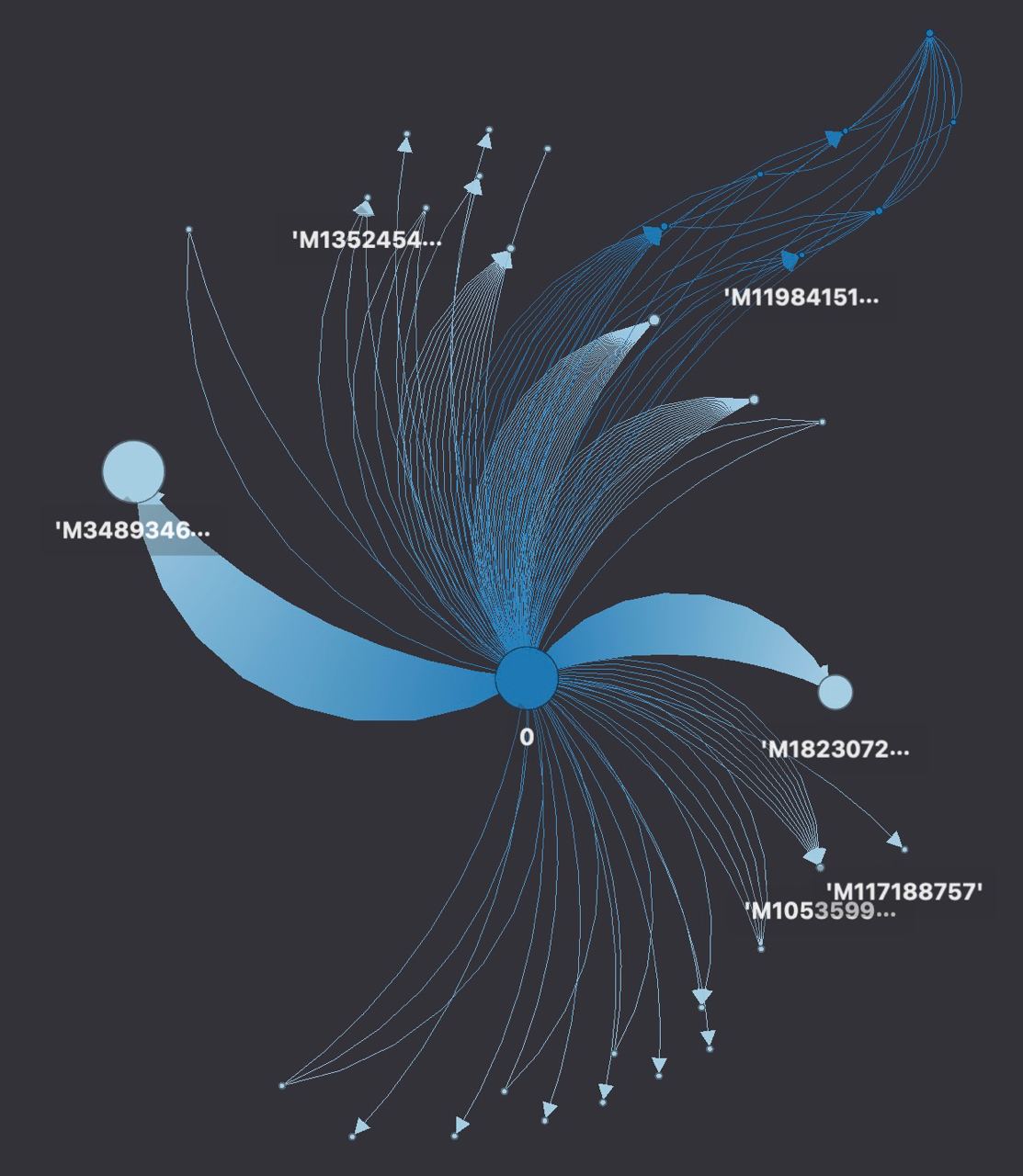
Variable “gender” is binary and contains name ‘F’ in case if the customer is a woman and ‘M’ otherwise.

Variable “amount” contains information about the cost of purchase.

Variable “fraud” is a target variable and contains 0 if transaction is normal and 1 if transaction is operated by fraudster.

*In an email you can find notebook with Exploratory Data Analysis of banking fraud dataset*

Also there were built several graphs which shows in a better way customers, merchants and their interactions.



Merchants by appearing in fraud transactions. 0 and 1 relate to ID of merchant, each edge represents a transaction form sample.

The graph shows the connection between merchants and their appearance in fraud transaction. It’s built using all the transactions from sample so that the ratio between fraud and non-fraud transaction would keep. By checking the [link](https://hub.graphistry.com/graph/graph.html?dataset=9df4a0ae420340e483c1f16b1690a766&splashAfter=false&play=5000&session=922052c6dc0a427f8265ee117ee5a53a) you can see an interactive map, which allows to check every edge (i.e. transaction) from dataset.

By checking this [link](https://hub.graphistry.com/graph/graph.html?dataset=86d21dded13b4afa9bce13cbe548e738) you can see merchants by selling in category. Took only appearing in category, without accounting number of transactions.

The following graph represent the connection between merchants and categories they sell in. We took only unique combination of seller-category, to make it easier to understand. The link also provides an ability to modify the graph, its visualization, etc.

By checking the [link](https://hub.graphistry.com/graph/graph.html?dataset=57bcd256e8f743f3b13fa59a66476c01) you can see customers by selling in category. Took only appearing in category, without accounting number of transactions.

Finally, this graph represents the connection between customers and categories. The methodology was used as in the case of merchant-category graph, the link also provides more tools for graph setting.

1. [(PDF) BankSim: A Bank Payment Simulation for Fraud Detection Research (researchgate.net)](https://www.researchgate.net/publication/265736405_BankSim_A_Bank_Payment_Simulation_for_Fraud_Detection_Research) [↑](#footnote-ref-1)