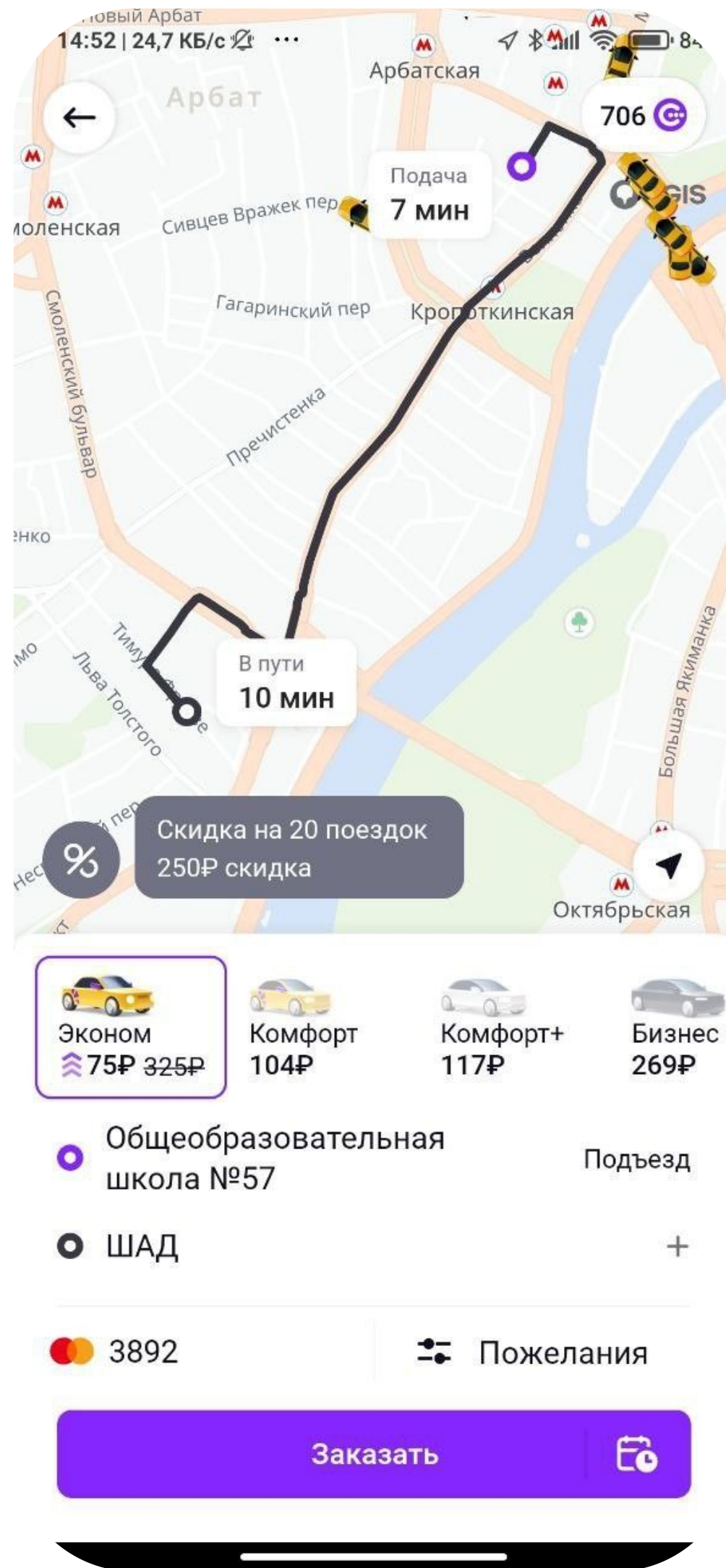


# СИТИМОБИЛ

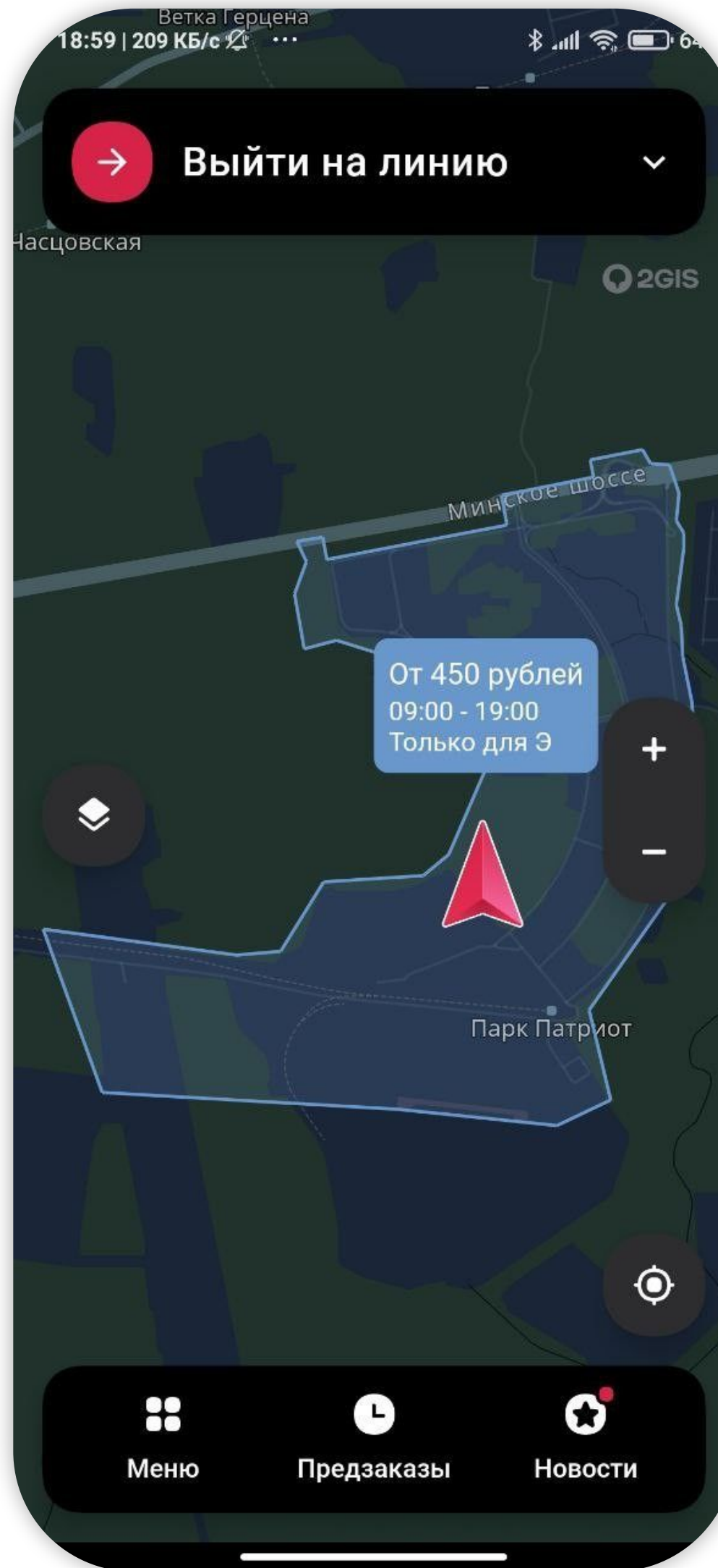
**Поиск мошеннических сообществ  
с помощью алгоритмов  
кластеризации на графах**

Даниил Казанцев  
d.kazantsev@city-mobil.ru

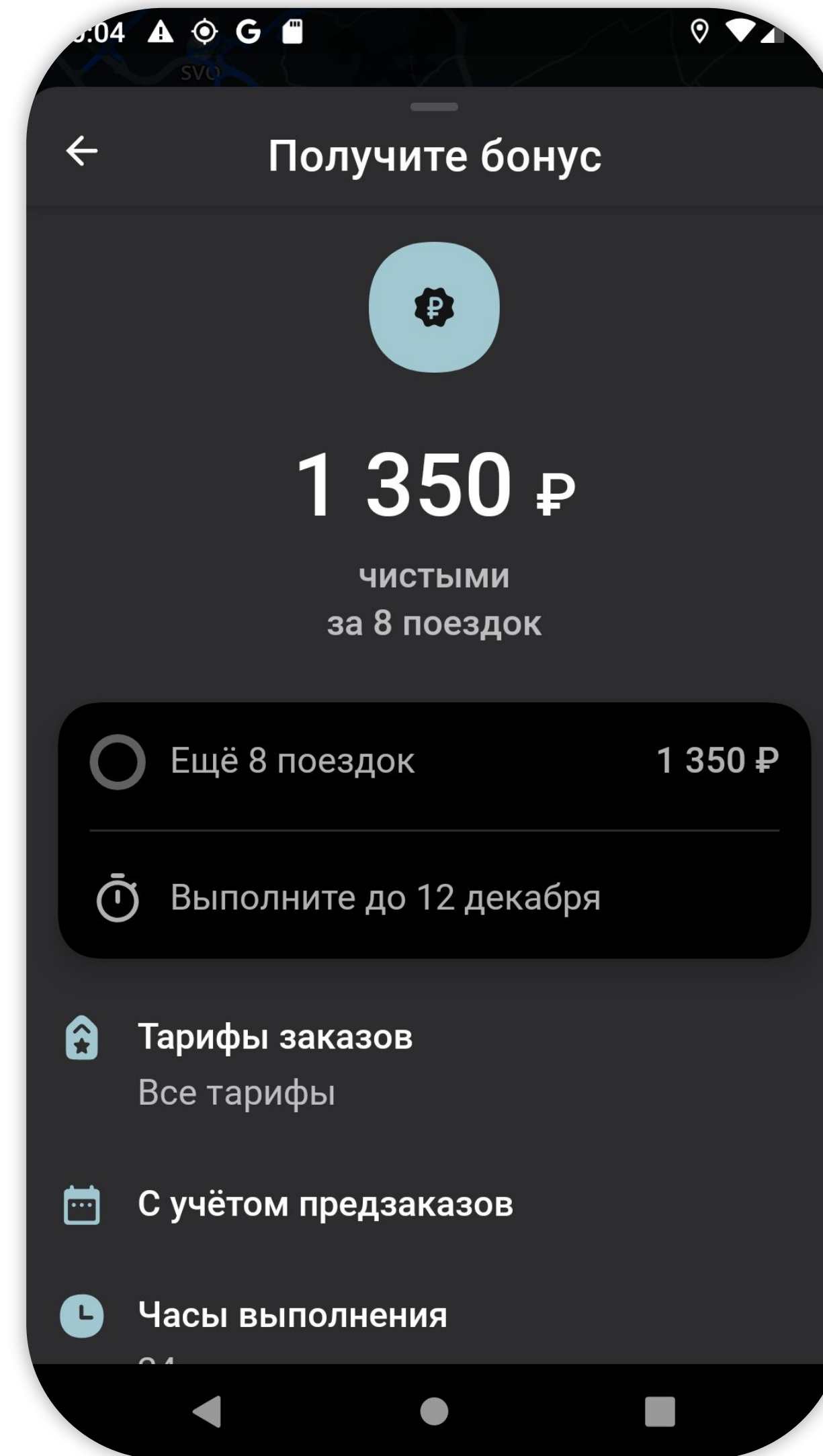
**Как можно заработать  
на агрегаторе такси?  
Зачем нужен Анти-Фрод?**



Клиентские скидки



Геоминимальные зоны



DxGy (Do X - get Y)

**Как устроен полный цикл  
создания нового продукта?**

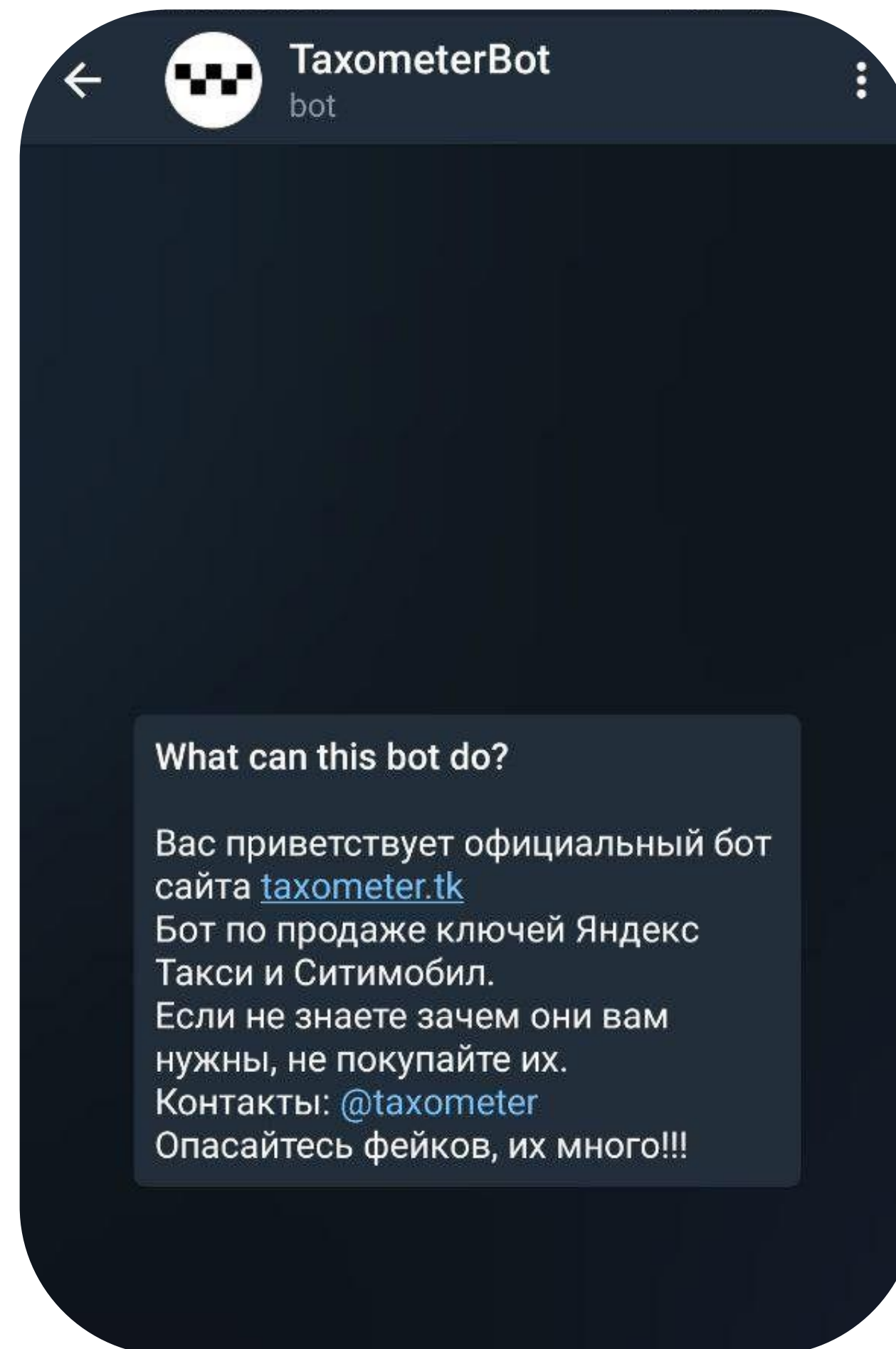


**Кого искать?  
На что смотреть?**





Хакерские фермы



Бот самозаказов

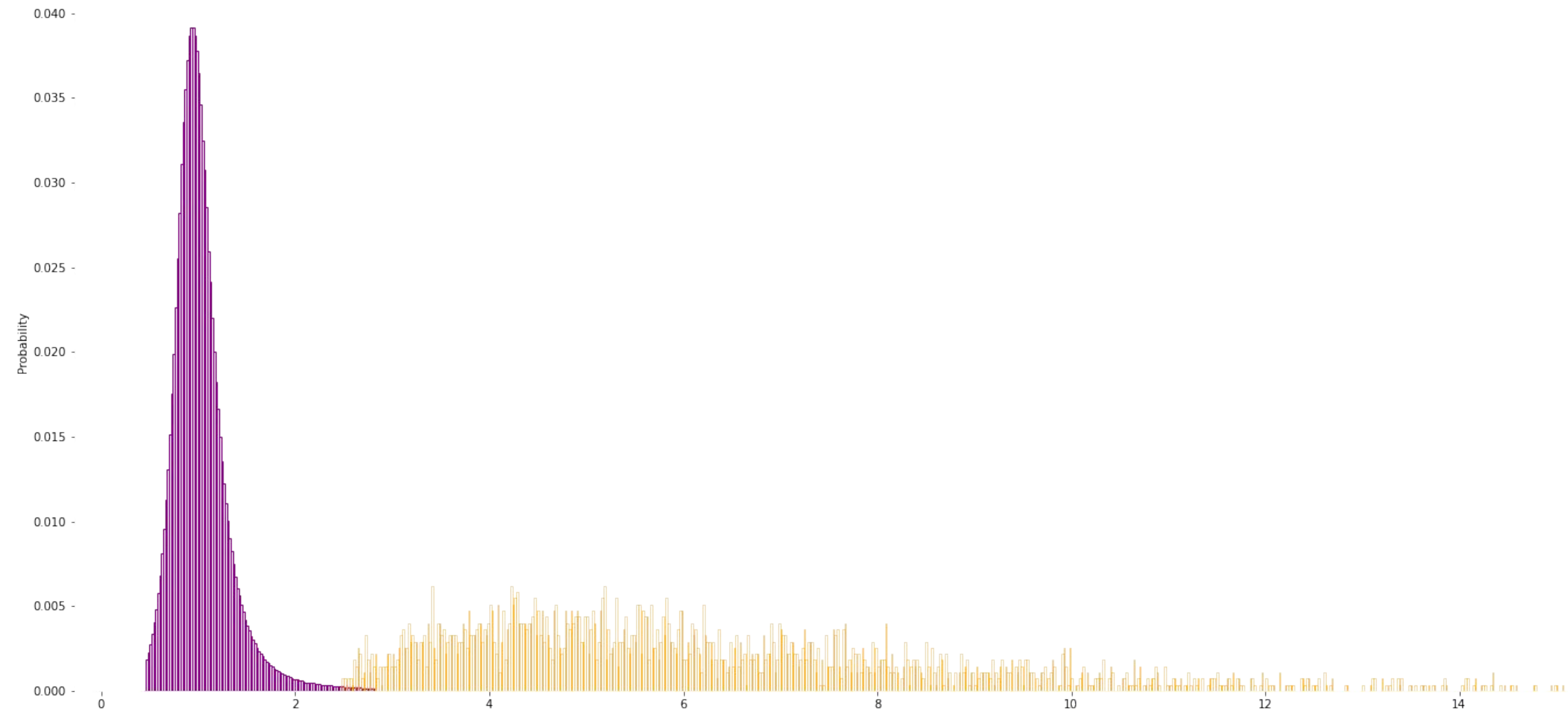


Самозаказы

# Как можно решать?



# Находить фродоёмкие метрики и аномалии в них

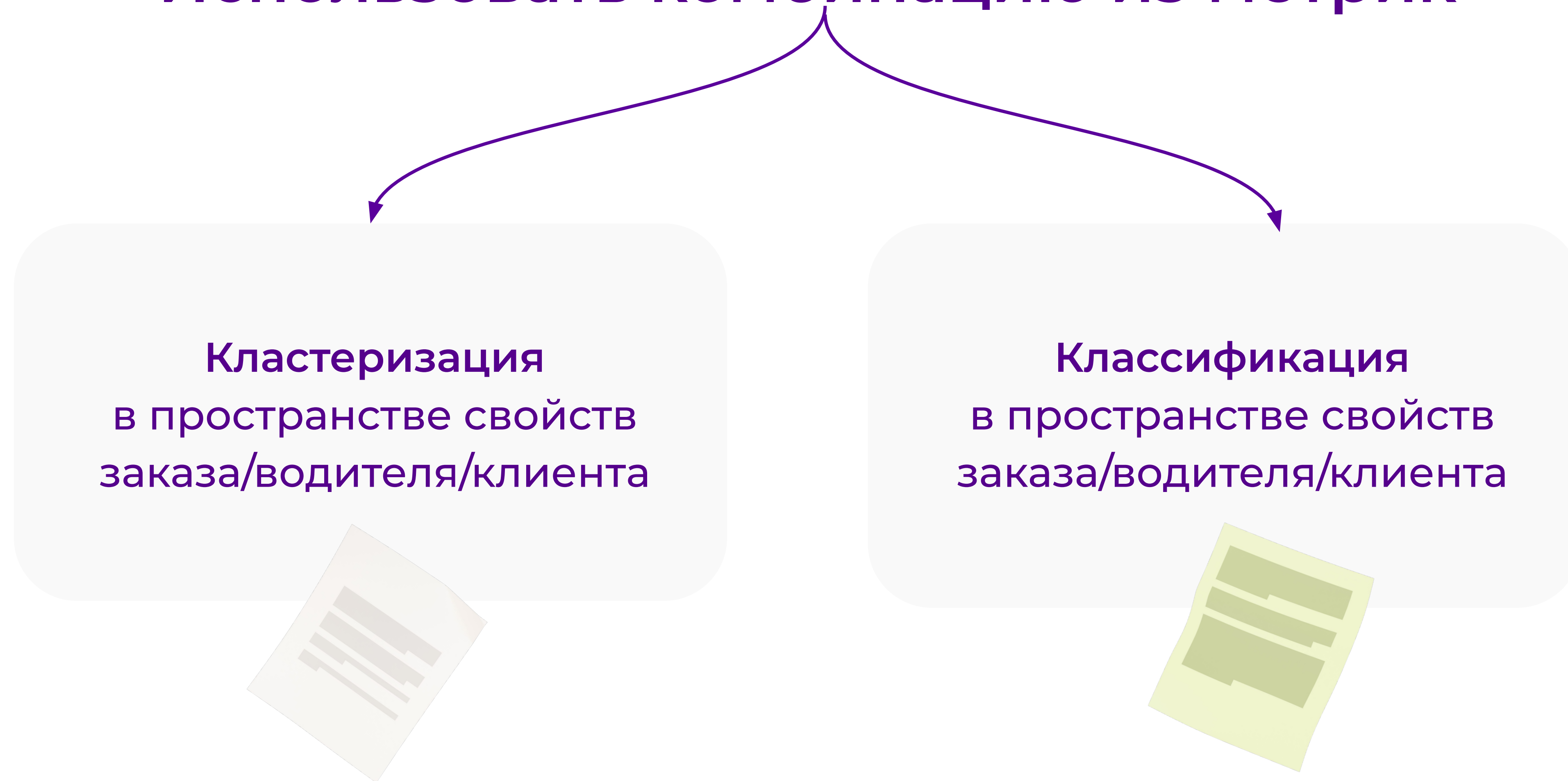


# Вероятностный подход

- ▶ Доля заказов за безналичный расчет  $p\%$
- ▶ Водитель сделал  $R$  - поездок, из которых  $r$  были оплачены по карте
- ▶ Рассмотрим вероятность того, что при  $R$  поездках было  $r$  или меньше поездок за безналичный расчет

$$\sum_{i=0}^r C_R^i \cdot p^i \cdot (1 - p)^{R-i}$$

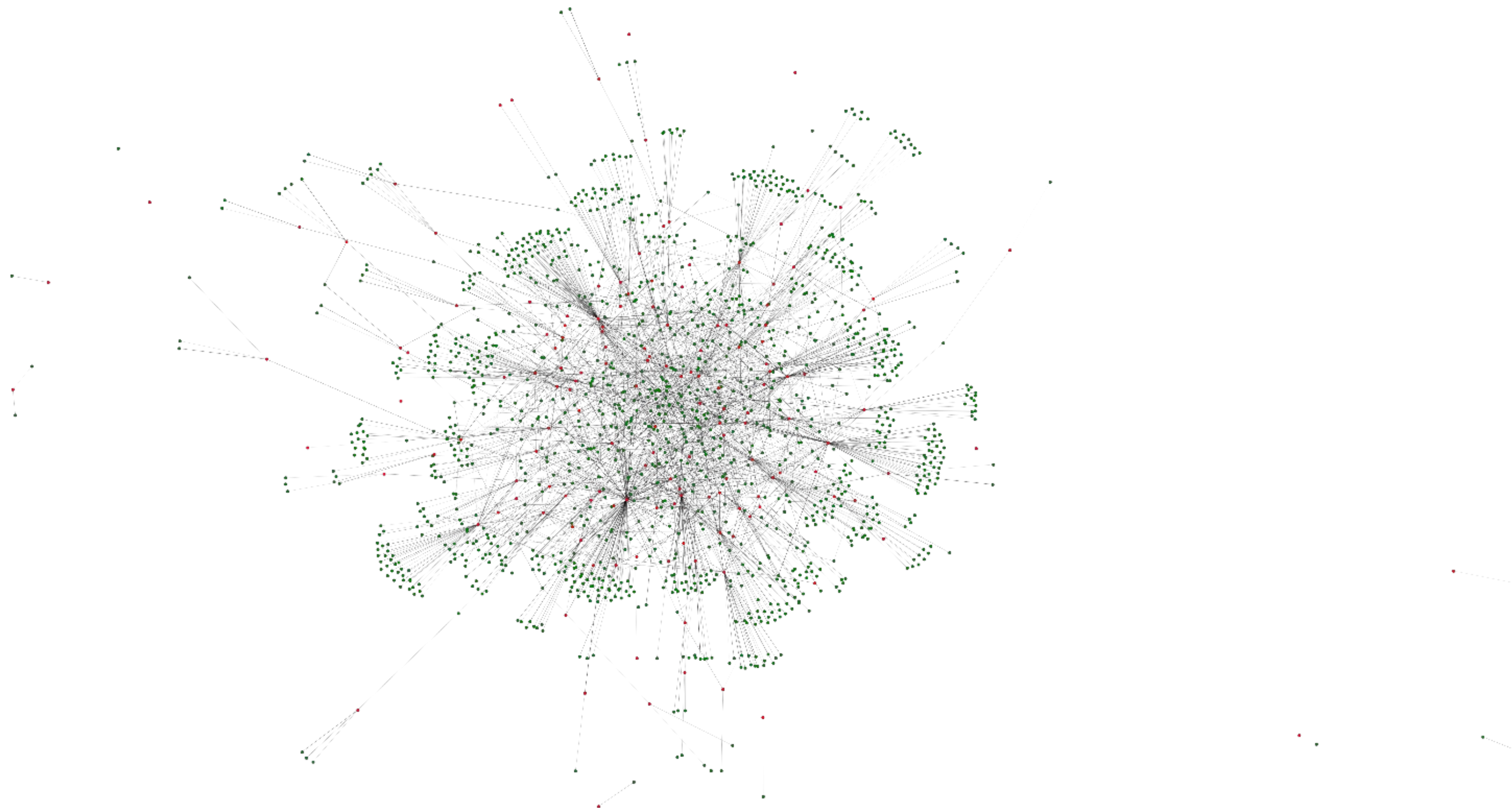
# Использовать комбинацию из метрик



# Как формализовать задачу на графе?



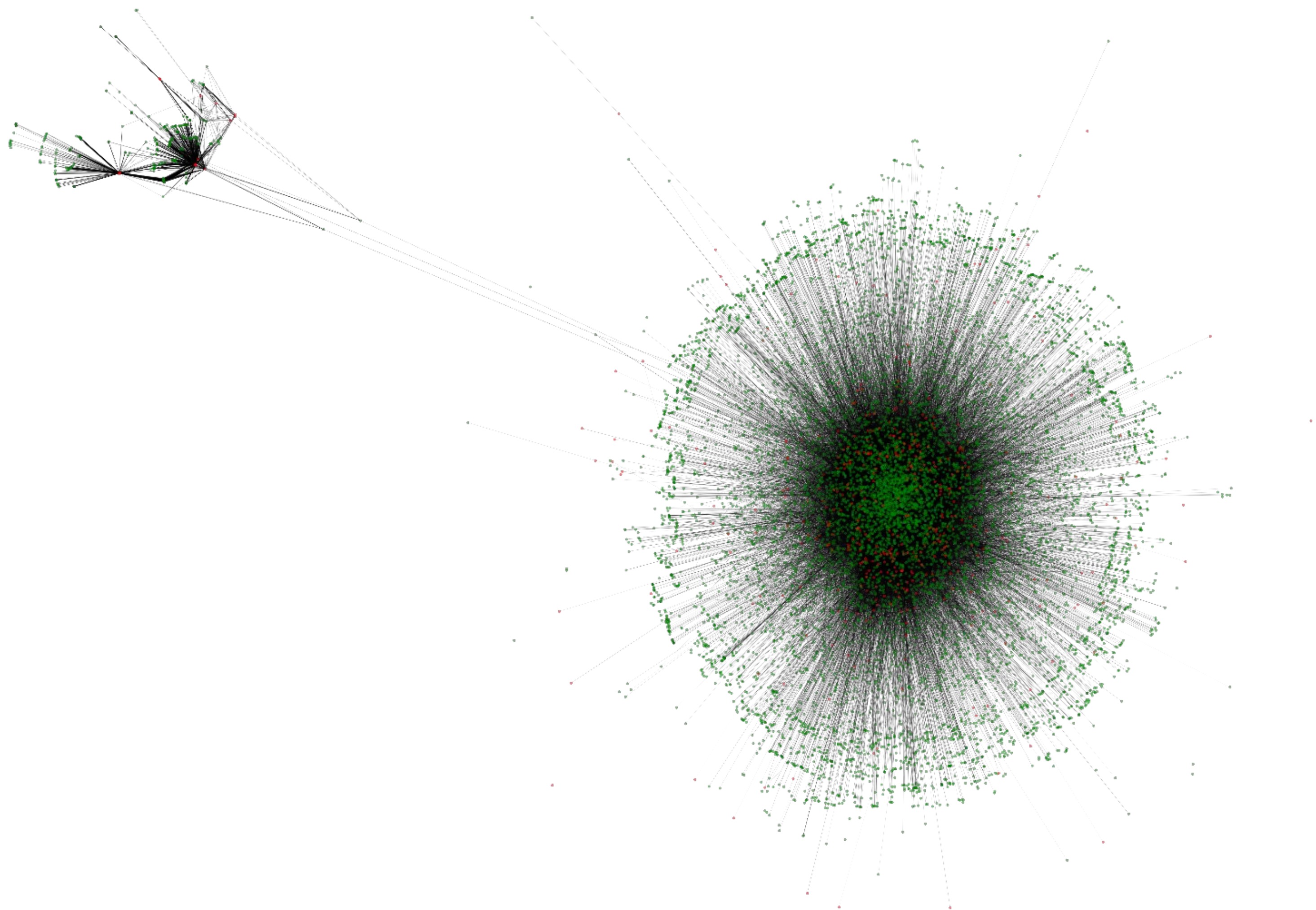
# Граф поездок



**Кто же тогда мошенник?**



# Граф поездок



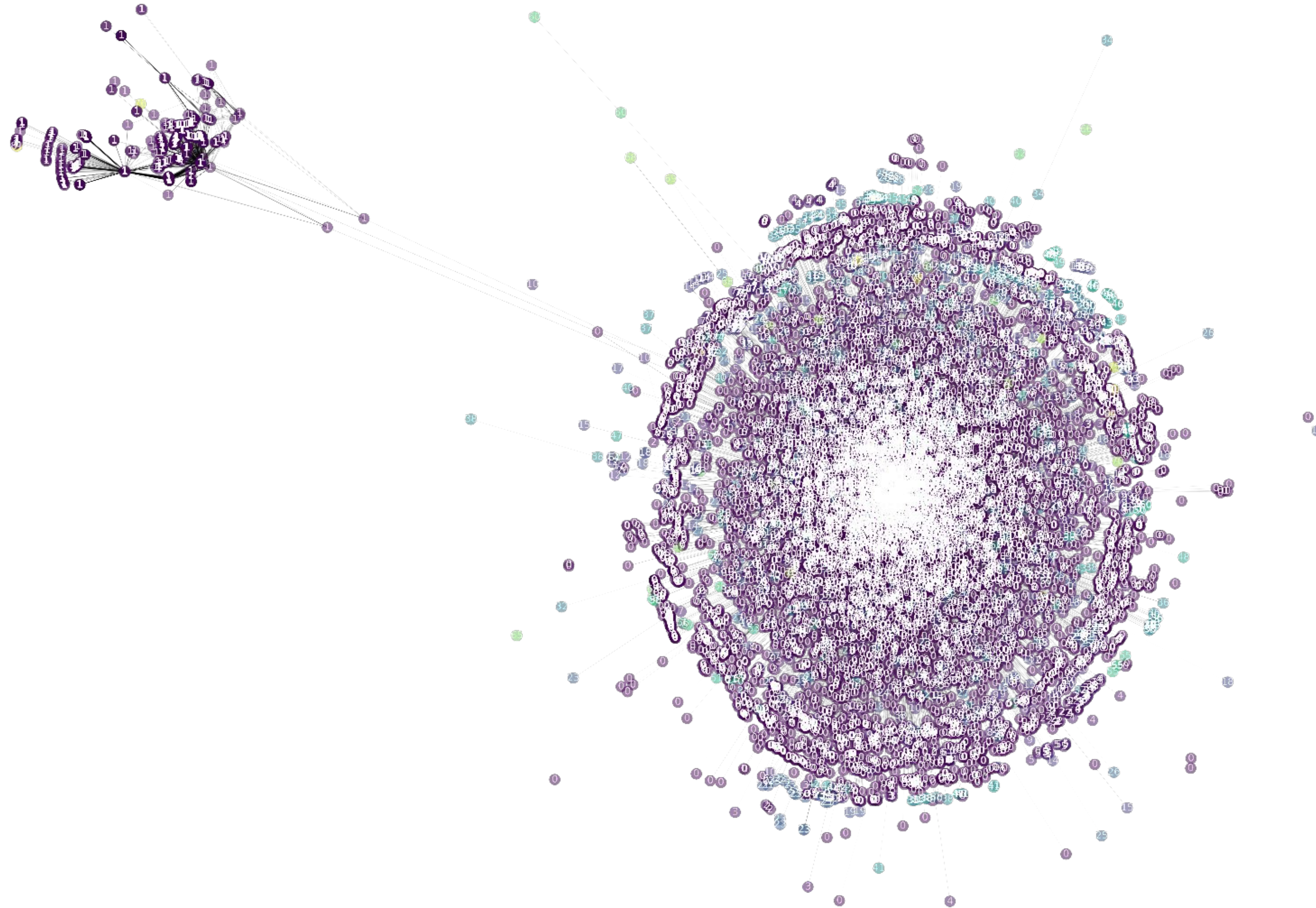
# Как разрезать граф?



# Как разрезать граф?

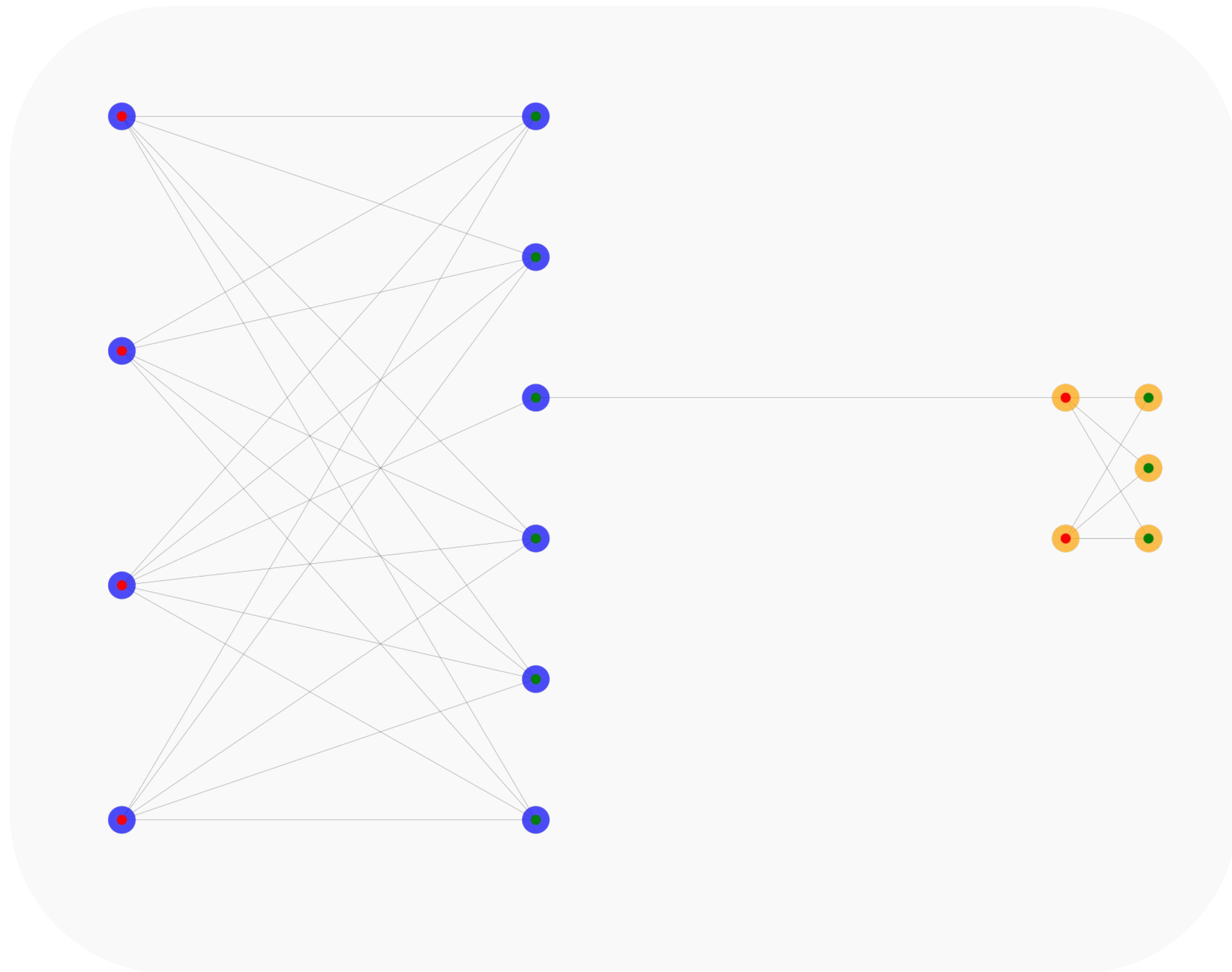
- ▶ Компоненты связности
- ▶ Разрядить графы и построить компоненты связности
- ▶ K-means, Label propagation
- ▶ Louvain algorithm
  - ▶ Алгоритм работает на взвешенных графах
  - ▶ Алгоритм выдает динамическое количество компонент
  - ▶ Асимптотика  $O(n \log(n))$

# Louvain clustering algorithm



**Как понять по компоненте  
графа насколько она опасна  
для бизнеса?**





Замкнутость  
синей компоненты

$$\frac{5 \cdot 4 + 1}{5 \cdot 4 + 2} = \frac{21}{22}$$

Замкнутость  
оранжевой компоненты

$$\frac{3 \cdot 2}{3 \cdot 2 + 1} = \frac{6}{7}$$



Коэффициентом реберного расширения множества  $S$  ( $S \subset V(G)$ ) называется число

$$\frac{|E(S, \bar{S})|}{|E(S, V(G))|}$$

где  $E(S, \bar{S})$  - множество ребер, одна вершина, которых принадлежит  $S$ , а другая нет,  
 $E(S, V(G))$  - множество ребер, у которых хотя бы одна вершина принадлежит  $S$ .

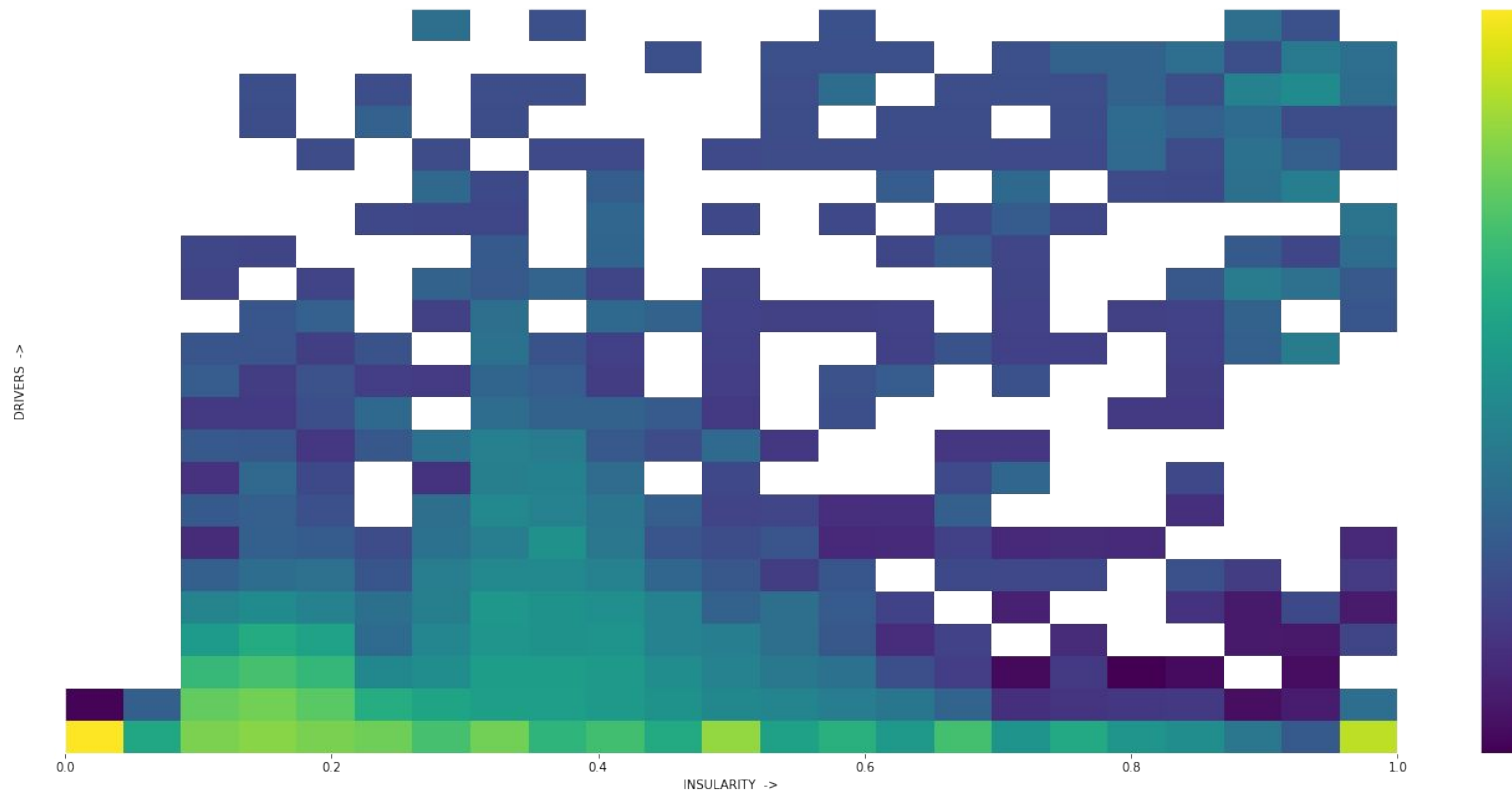
Коэффициентом замкнутости множества  $S$  ( $S \subset V(G)$ ) назовем число

$$\frac{\sum_{e \in E(S, S)} w(e)}{\sum_{e \in E(S, V(G))} w(e)}$$

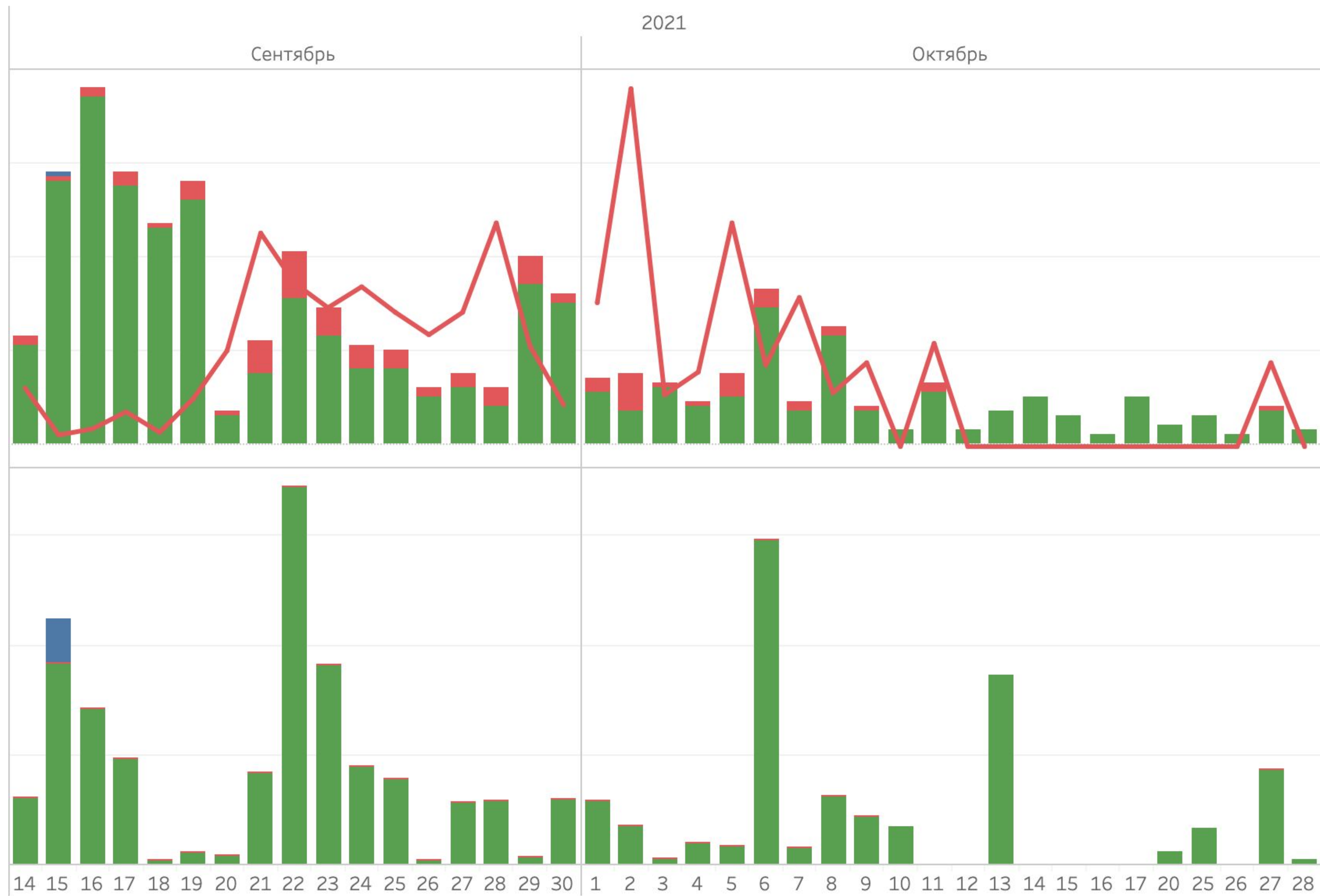
где  $E(S, S)$  - множество ребер, обе вершины, которых принадлежат  $S$ , а  $w(e)$  - это вес ребра  $e$ .

Если у графа нет кратных ребер, то

Коэффициент замкнутости = 1 – Коэффициент реберного расширения

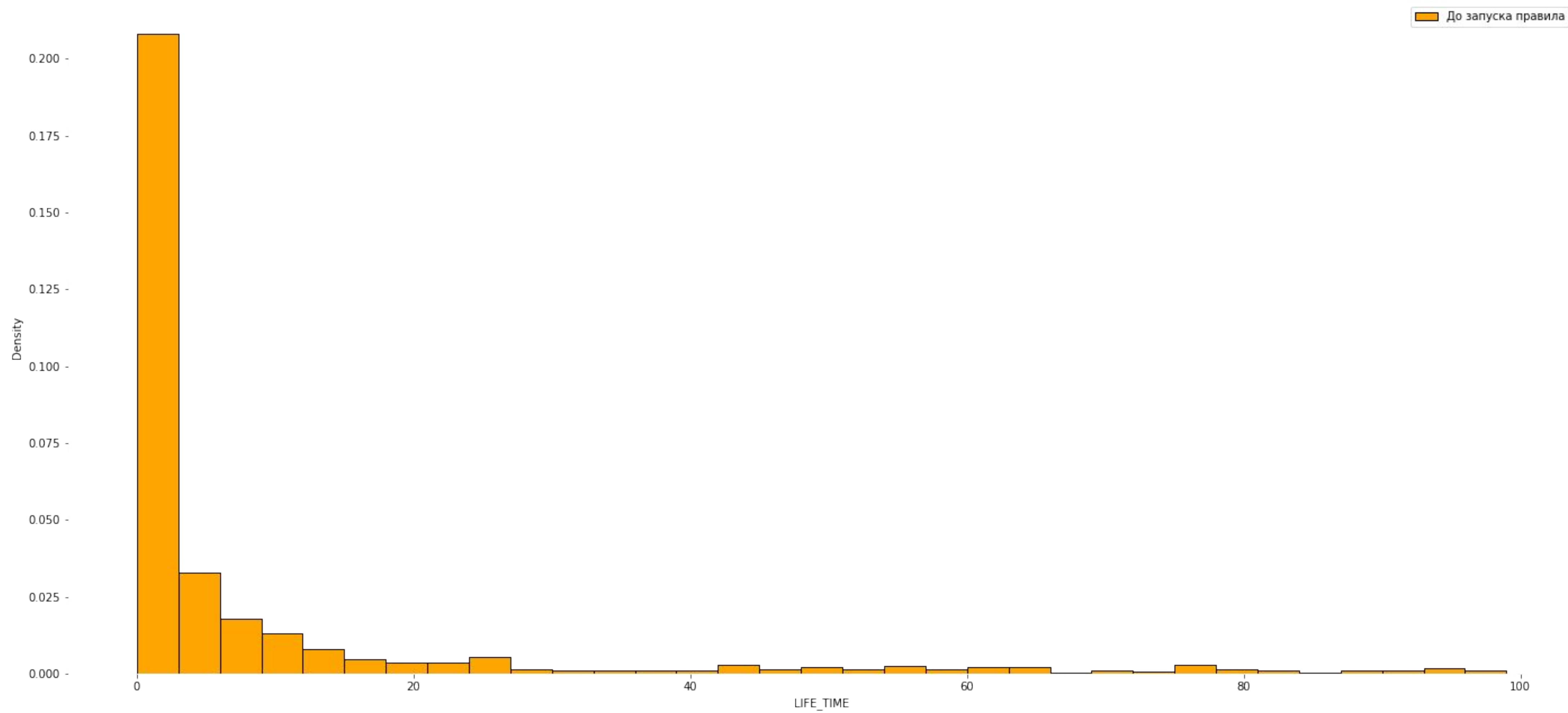


# Как применять санкции?

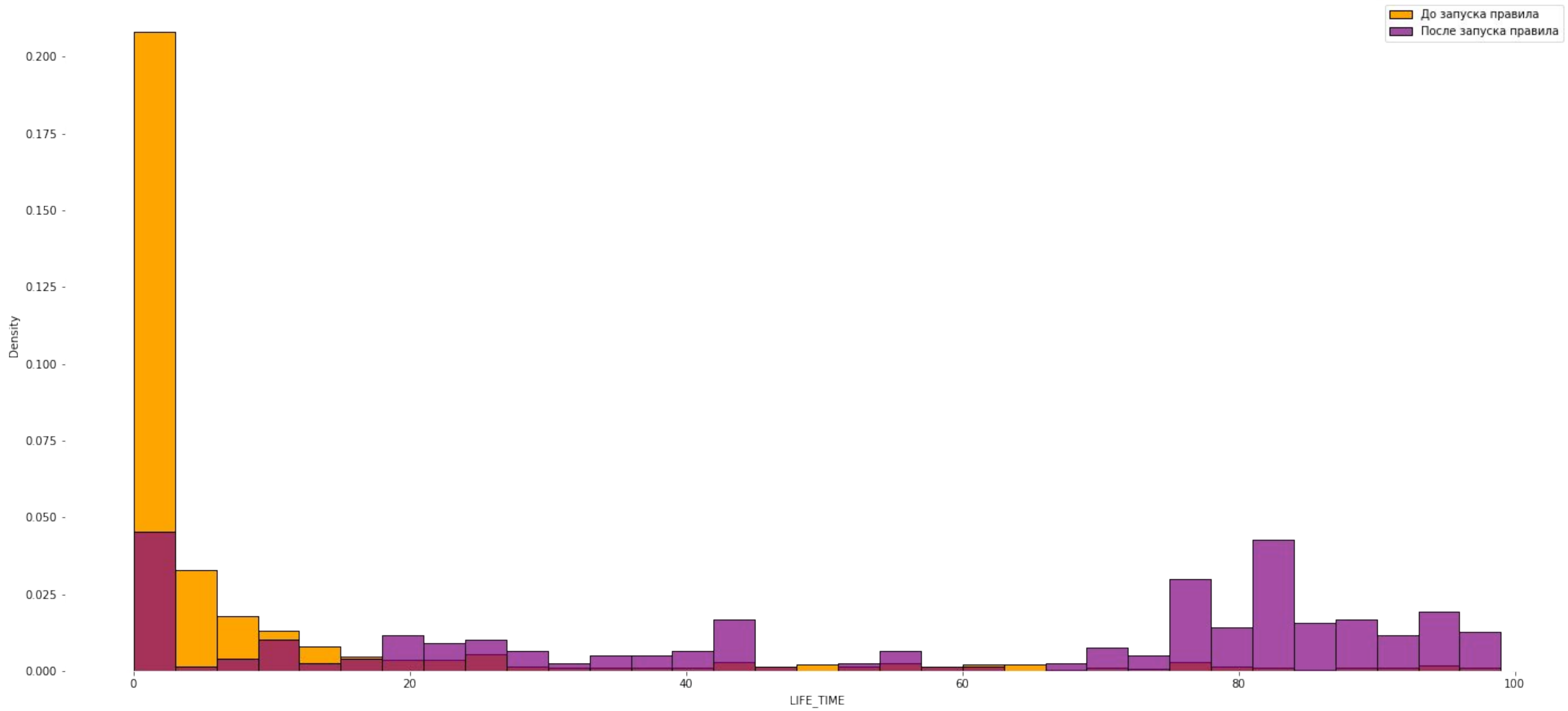




**Как недоброжелатели  
”обучаются” на правилах?**



# Как вылечить ”новых” клиентов?





# СИТИМОБИЛ

**Поиск мошеннических сообществ  
с помощью алгоритмов  
кластеризации на графах**

Даниил Казанцев  
d.kazantsev@city-mobil.ru