IDEAS AND WAYS TO CATCH “STRANGE” AND RISKY BEHAVIOR

ID checks

- Population registry
- Politically exposed person (PEP) (sender or receiver)
- People with criminal, terrorist lists, sanctions, FATF etc. (sender or receiver)
- Black-listed while previous attempts

“Social” connections based analysis:

- Hubs: “influencers” (often receive money) and active “followers” (often send money)
- Large groups of users (number of users, max value of transfers, turnover)
- Long chain of transactions
- Large values of transfers in groups

Behavioral:

- Continuing to send same/different amounts if first attempt was “Cancelled” to same recipient
- Sending same amounts of money
- Sending with the same intervals of time
- High frequency (tiny time since previous transfer, bots)
- Transfer attempt shortly after user created
- Number of transfers (per hour, day, 7 days)
- Unusual location

Risky by location:

- Country in list of “tax heavens”
- High-risk third countries (weak anti-money laundering and terrorist financing regimes)
- Country from EU, G10, continent
- Least developed countries (link)
- Recipient bank with bad reputation (no/low rating, no limits, location, etc.)
- Non-IBAN country

Risky by transaction type:

- Card payments are riskier than bank transfers (online fraud (card-not-present fraud), lost or stolen cards)

Measures implied by law

- EU Directives
- Fifth Anti-Money Laundering Directive (link)

Digital Footprint

- Device info (Device model, applications, geo-location, IP)

RESULTS OF THE ANALYSIS

- Better understanding and "ground truth" knowledge of the problem will help to achieve better results.
- Information containing user IDs that may need further review are stored in the res_Risky2Monitor.xlsx
- Resulting Python code (edd.py) could be modified and applied to analyze same kind of data. To make code more readable some functions were saved to additional files and stored in TOOLS folder.
- The dashboard with implemented and not-yet-implemented ideas is available via Trello link

MEASURES TO CAPTURE POTENTIAL BAD CUSTOMERS

1. New users (risk reducing measures, out-stuff risk assessment)

- To start using Top up from clients personal account
  Reason: user's Bank account was created and approved by bank's compliance.
  Problems:
  - Quality of compliance depends on bank and country standards.
  - Does the transfer include sender info to match data to define account as "personal" (name or document ID)?

  Implementation: Is implemented
- Limit amount (Sum) of first N transfers
  Implementation: Easy to implement
- Automated ID document check
  Implementation: High costs, Different countries -> different documents, stolen identity, fake documents.
- Credit bureau data.
  Implementation: High costs, Different countries -> different bureaus, consent, "No Hit" clients)

2. Regular users (risk)

- Day limit (Sum and number of transfers)
- Month limit (Sum and number of transfers)
- Cash withdrawal limit (Sum and number of transfers)

3. Cutoffs for review triggers

Impact: depends on the
Implementation: Easy
Cost: Low

- Amount of transfer > S
  Reason: Large amounts contain possible risk and should be reviewed especially if sent to high-risk locations
- Amount of transfer < S
  Reason: It could be suspicious that tiny amounts are transferred with such high costs
  Implementation: Easy. Cost: Low
- Number of transfers with 'payment_reference_classification' as 'test'
  Reason: User may try to test the compliance rules with a low amount transfer
- Number of transfers with 'payment_reference_classification' as 'gift'
  Reason: Some countries allow not to include into income money received as "gift", so personal account could be used to hide business activity
- Number Recipients > N (in 7 days, 30 days, 90 days, 181 days)
  Reason: What are the possible economic reasons for this activity? Who are recipients?
- Number of distinct incoming transfers >N (in 7 days, 30 days, 90 days, 181 days)
  Reason: What are the possible economic reasons for this activity? Who are senders?
- Destination is from/to "high-risk locations list" (offshores, high-risk third parties)
  Reason: Possible way to hide money in "specific" jurisdictions
- Long chain of transactions
  Reason: Possible way to hide original source
- User Group turnover > S

Reason: Money relocation that is not a business activity

- Business to Personal transfers
  Reason: If it is not a salary, invoice or other contracted payment, then the purpose should be investigated
- Personal to Business transfers
  Reason: Amount of transfers is not match declared business size
- Number of destination countries > N
  Reason: Financing strange activities

4. Behavioral (user)

- Unusual behavior for a user (invoice amounts)
- Different locations in same day (IP address, geo-location)
- Number of devices used
- New user location (address)
- New destination country

5. Other risks

- Limit amount of transfers to specific destinations (countries, banks) or currency

HOW TO IMPLEMENT

- Separate measures could be applied for different payment types: cards payments and bank transfers as they contain different risks
- Volume limits could be country-based (as different countries could have different economic conditions (wages, food basket, goods and gasoline prices, home and rent prices, etc.)) or as a general rule for all the transfers
- Rules to capture suspicious activity and to be monitored need to be designed in the way that doesn't create too much cases for review (only rare and uncommon).
- Cut-off can be defined from
  - expert knowledge
    and/or
  - empirical rules based on distribution of values (frequency (percentile) or deviation)

WHAT WAYS TO CAPTURE “STRANGE” BEHAVIOR WERE APPLIED TO THE DATA

- Outliers – extreme values
- Connections – “social” connections between users
- Clustering – groups that are different and not populated

Step 1. DEFINE REASONS WHY NOT TO TRANSFER MONEY

There are several conditions why transactions were cancelled:

- Obvious (97.1% of all not transferred)
  - 'date_request_received' is Null – no money received in advance
    Field Description: Date at which we received the customer’s money
  - Zero invoice amount
  - 'recipient_country_code' is Null – no destination country determined Field
    Field Description: Date at which we received the customer’s money
  - payment_type is Null
    Field Description: Payment method used to upload money
- More complex (2.9% of all not transferred)

Transfer Status

Initial data

payment_status	Number	Share
Transferred	77376	0.77376
Cancelled	22259	0.22259
Pending	365	0.00365

Cleaned data

payment_status	Number	Share
Transferred	77372	0.9914
Cancelled	611	0.0078
Pending	59	0.0008

Step 2. ENRICHMENT

- Exchange Rates (to simplify was taken only single date link) to convert all invoice amounts to EUR
- High-risk third countries with weak anti-money laundering and terrorist financing regimes link
- The EU list of non-cooperative jurisdictions for tax purposes link
- Group of Ten link
- Group of Ten currencies link

Step 3. DATA OVERVIEW (cleaned data)

Countries

- 119 address countries
- 64 destination countries

Currencies

- 20 sender currency
- 41 target currency
- 500 currency pairs

Users

- 143240 total
- 70243 senders
- 74342 recipients

Accounts

- 5608 business
- 72434 personal

Payment types

- 13 different payment types

Banks

- 819 unique counterparty banks from
- 44 countries

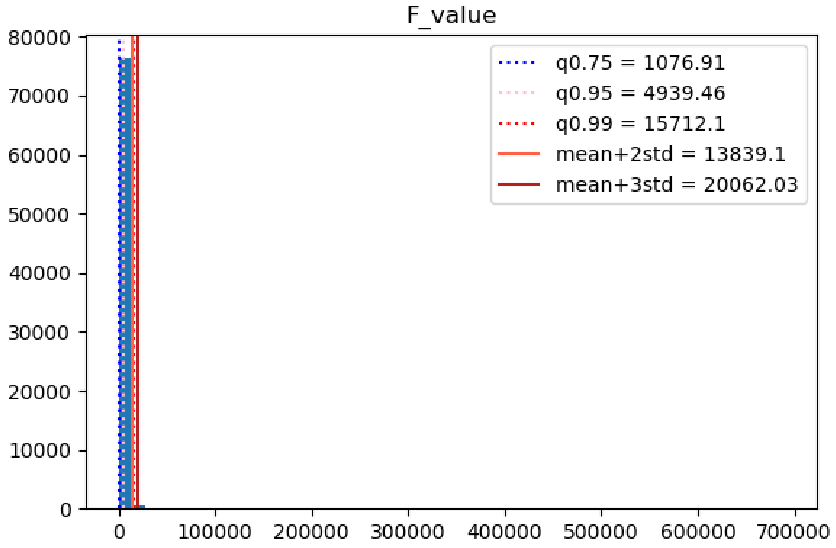
Devices

- 4 different device types

Time

- Users were created from 2011-01-23 13:54:00 till 2016-12-06 05:47:00
- Requests submitted from 2011-02-04 12:28:00 till 2016-12-06 15:03:00

Step 4. OUTLIERS

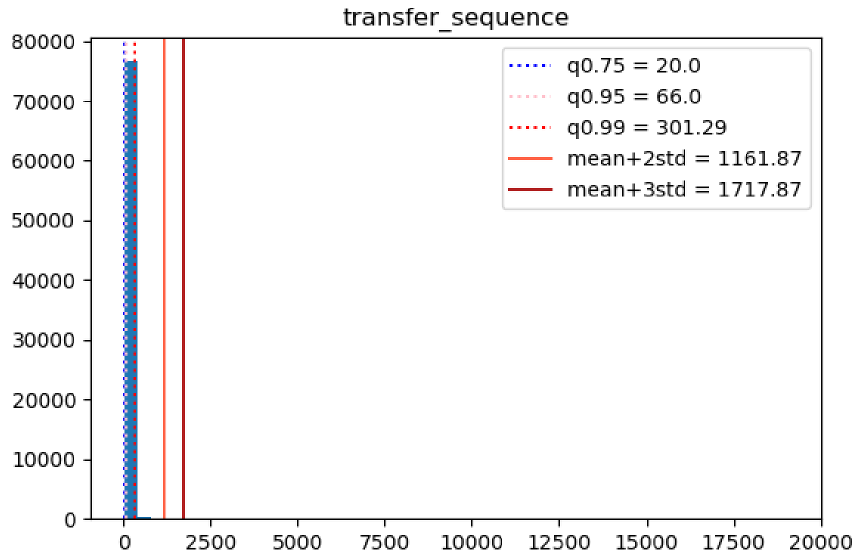


Min	Max	Mean	Std	q1%	q5%	q25%	Median	q75%	q95%	q99%
0.78	688221.6	1393.22	6222.94	10.0	35.0	152.86	425.0	1076.91	4939.46	15712.1

Extreme amount of transfer

user_id	target_recipient_id	request_id	Ccy _send	Ccy _target	Transfer _sequence	F_isBiz	F_value
69cf237499d8ccac9211602c37807d92	cd654f1f3e98b806281200c0c6727323	f550cdf6c57120376657707e4a540810	EUR	GBP	2.0	0	688 221,60
3c45517470b67c633cb8d4694be1a4d1	e5a63fe59c6a8bba164269f2fec8e280	487ea3ae8aab0b2c6d30ad50770c7173	USD	GBP	3.0	0	459 126,86
5e13d1e382d895a5a58b40173eb7abfd	771b1c7dce5c2363071f480433fce6ef	233cf0831cfbb1520c038d1a48fda556	GBP	EUR	6.0	0	274 857,86
62d5122bcb6646b2c0b04142deb0ffc4	bf84b29e361586c24bf7b0ba33de5b89	368ac9f5ec63b80a43e54313ca08aae0	GBP	EUR	3.0	1	267 068,17
8866cefce6886f5dab4d10b52bc3082d	67c1458ed0ab061fb2f8a746cfa76ed3	81c107b099c88e93d13ce26da8daf8ae	GBP	EUR	7.0	0	250 962,78
314ee167f8c7444bcc059a4d10a96999	97c478775fa04f754454dc18c49e50fa	f982ffa7c8dd919faeca6b015645dcb4	GBP	EUR	3.0	0	235 000,00
c6de568175564068cea6cbfa95128c02	5d490d16b7b3bf4d282fb7e8ca3c3998	2cb799409362927e54aa63ae48953758	USD	EUR	3.0	1	280 634,45
3072178819220ff5699a6ef70be68566	00044057f135cc6526be752ad83115f6	5634cdd5f8078fbc22417c38dc348ab6	GBP	USD	6.0	0	190 000,00
ffe9473ac124478429dbc20b40f50e52	c7a62b2a28cbe0c518af810cb614f418	609f0e46d73eb14305a23efac00bca50	EUR	GBP	2.0	0	210 886,00
73413a7e0f5bfd63659ea21cd9acbd51	b1a67cd09905221d55ef61236e1caf1a	b32d789d187e88c98ef80a4393611044	GBP	USD	1.0	0	172 501,68

Extreme number of transfers (‘transfers_sequence’)

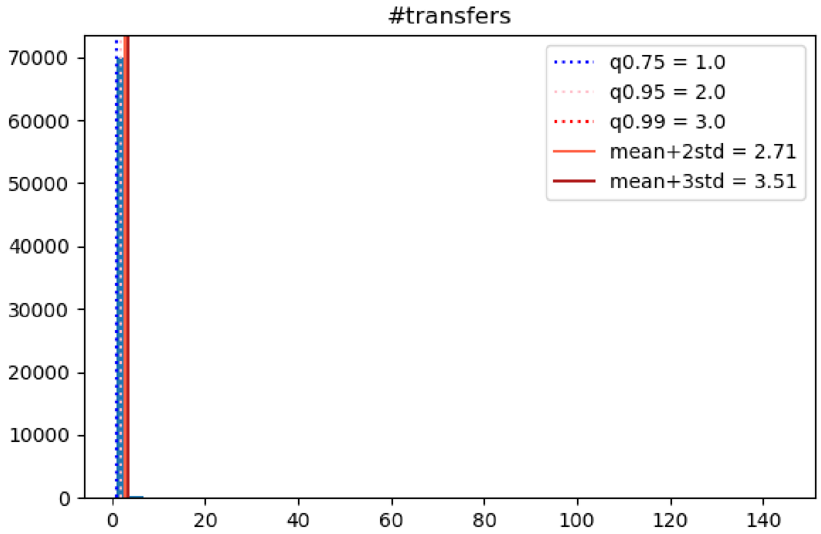


Min	Max	Mean	Std	q1%	q5%	q25%	Median	q75%	q95%	q99%
1.0	19069.0	49.87	556.0	1.0	1.0	3.0	8.0	20.0	66.0	301.29

user_id	target_recipient_id	request_id	ccy_send	ccy_target	transfer_sequence	F_isBiz	F_value
b2afd12d1322929e095bd85468e50a55	183ad425383b741a6f8146efe4adbb33	1152a561ac76e87414c0e2a2d61038ec	GBP	INR	19069.0	1	7.77
b2afd12d1322929e095bd85468e50a55	7ff4eff382a55b0aaa57bcd690eb229e	700aa001ba5544e9292e989b2e2e71b4	GBP	INR	19055.0	1	7.77
b2afd12d1322929e095bd85468e50a55	7f872c308e88f6ffe3657fd98da1ef62	c77a7312053c00b375a09b99a5b516f0	GBP	INR	18844.0	1	7.87
b2afd12d1322929e095bd85468e50a55	7e79a12304eea02bb4511cc42997dcdf	bff8eee0482928cd0dbd366d9866fc5a	GBP	INR	18802.0	1	7.85
b2afd12d1322929e095bd85468e50a55	c143ee59c06bf6a7f5d1b9a4d3018d52	b4f11d2d293afe125b5c94a7c34d0c9d	GBP	INR	18746.0	1	7.85
b2afd12d1322929e095bd85468e50a55	7ab5d82a944edbc7af95b57287929cc6	45e27b5d441e6af025f603f428814c99	GBP	INR	18652.0	1	7.86
b2afd12d1322929e095bd85468e50a55	26b6884a0a9fc8dab803645506bb78a4	ff85cf60bda65e2a948695c821be1c58	GBP	INR	18504.0	1	7.94
b2afd12d1322929e095bd85468e50a55	e543190166661ad276fd2209751e74a2	36490e5a7680774e9f6dd71255703b7f	GBP	INR	18458.0	1	7.94
b2afd12d1322929e095bd85468e50a55	262081b34449a2daa8c8dbac65784372	2997a63d9e88560a0a06a4fe5d74b29f	GBP	INR	18245.0	1	7.44
b2afd12d1322929e095bd85468e50a55	86d82e28ed8cc426e3208bebfbf41694	58251baab3b3d192f502b75626a8fc37	GBP	INR	17817.0	1	7.34

The most extreme values are from the same user_id = b2afd12d1322929e095bd85468e50a55

Extreme number of transfers (present in data)



Min	Max	Mean	Std	q1%	q5%	q25%	Median	q75%	q95%	q99%
1.0	144.0	1.11	0.8	1.0	1.0	1.0	1.0	1.0	2.0	3.0

user_id	#transfers	#unique_targets
b2afd12d1322929e095bd85468e50a55	144	144
69fd02c4fbd5bfa6533f7a5eac3bd81c	74	74
0777a4c36ee0b81e85fbad4bfdd23472	50	50
ee97bc9aa9b2e6c7b2908aa16c606f44	43	43
a6e836d9c18562cfd2c574a311908bd0	40	40
19466121d8747bd79d1ec4d109b63c52	37	37
b0c7ae2316c7e8214fd659e4bc8a0dea	24	19
771aac18209b1276a651d3ac808e039a	13	13
e608d060011ee543263f345f9887c6c4	12	12
d422d85d05a3e2982baea7e2190b471d	12	12

Step 5. "SOCIAL" CONNECTIONS BASED ANALYSIS

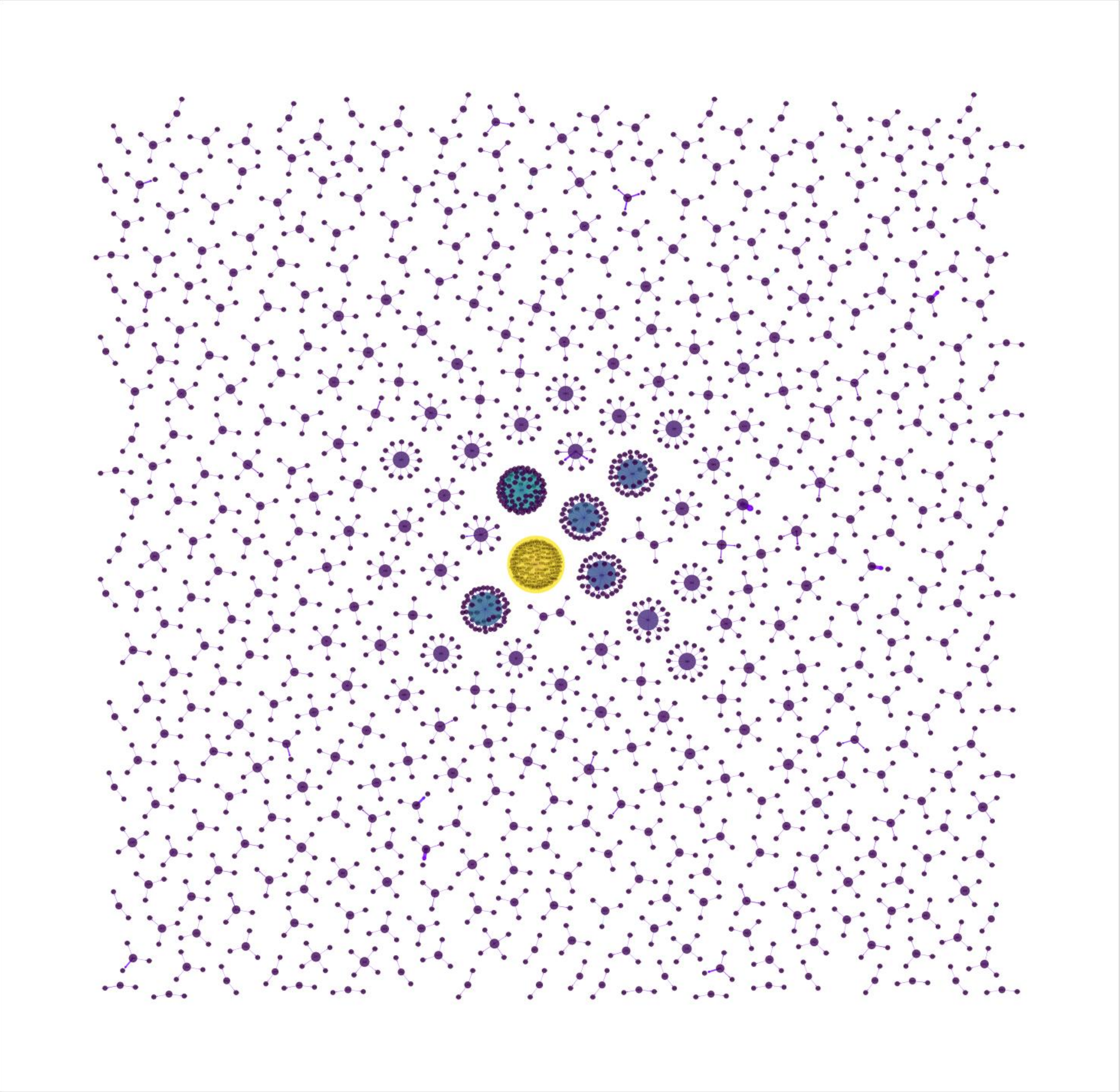
To visualize the structure of the directional graph is build with NetworkX Python package. (On the plot shown depicts users with more than 2 connections (in any direction).

What to pay attention:

1) Large groups
2) Concentration: Size of a ball represents the sum of received and sent transfers number (degree)
3) Huge transfers: Amount of transfer is shown by thickness of an arrow. Thick are high-amount transfers
4) Long chains of transactions

Results

- A number of users were defined that have large amounts of connections.
- Groups with high turnover were discovered.
- "long" user-chains were selected



Users by number of incoming and outcoming links

user_id	links_in	links_out	links_num
b2afd12d1322929e095bd85468e50a55	0	144	144
69fd02c4fbd5bfa6533f7a5eac3bd81c	0	74	74
0777a4c36ee0b81e85fbad4bfdd23472	0	50	50
ee97bc9aa9b2e6c7b2908aa16c606f44	0	43	43
a6e836d9c18562cfd2c574a311908bd0	1	40	41
19466121d8747bd79d1ec4d109b63c52	0	37	37
b0c7ae2316c7e8214fd659e4bc8a0dea	0	19	19

Groups by size

Medians: 2 members | 99 quantile: 3 members

Groups that have more than 6 members: 44

F_groups_size	Number_of_groups
2	64664
3	3610
4	413
5	96
6	34

Extreme group size

F_groups_size	Number_of_groups
38	1
42	1
44	1
51	1
75	1
145	1

Groups by turnover

Medians: 453.56

99 quantile: 17932.25

F_groups	F_groups_size	F_groups_turnover
{'69cf237499d8ccac9211602c37807d92', 'cd654f1f3e98b806281200c0c6727323'}	2	688221,6
{'e5a63fe59c6a8bba164269f2fec8e280', '3c45517470b67c633cb8d4694be1a4d1'}	2	418784,6
{'771b1c7dce5c2363071f480433fce6ef', '5e13d1e382d895a5a58b40173eb7abfd'}	2	311659,4
{'62d5122bcb6646b2c0b04142deb0ffc4', 'bf84b29e361586c24bf7b0ba33de5b89'}	2	302826,7
{'8866cefce6886f5dab4d10b52bc3082d', '67c1458ed0ab061fb2f8a746cfa76ed3'}	2	284564,9
{'97c478775fa04f754454dc18c49e50fa', '314ee167f8c7444bcc059a4d10a96999'}	2	266464,8
{'c6de568175564068cea6cbfa95128c02', '5d490d16b7b3bf4d282fb7e8ca3c3998'}	2	255975,8
{'00044057f135cc6526be752ad83115f6', '3072178819220ff5699a6ef70be68566'}	2	215439,6
{'ffe9473ac124478429dbc20b40f50e52', '6f622501593c39439dc2a84ac330f117', 'c7a62b2a28cbe0c518af810cb614f418'}	3	214217,5

Other ‘social graph’ ideas:

- Banks-based connections – “popular” and “uncommon” banks
- Country based connections - geographical clustering
- Map transfers on a world map to visualize geographical clustering (coordinates need to be added)

Step 6. CLUSTERING

The common idea behind finding strange behavior is to find relatively rare cases (small clusters) inside the data that were separated from general population based on the combination of values.

Two approaches of unsupervised learning were applied to achieve this goal:

- MeanShift clustering (density-based clustering) - number of clusters is defined inside the algorithm. With larger bandwidth the number of cluster will be smaller than with smaller one.
- K-Means clustering - number of clusters needs to be set "manually"

Set of columns

Wider set of columns creates wider diversity of combinations and, as a result, the number of cluster increases.

Parameters

While defining the parameters (number of clusters and cutoffs) the capacity and cost of monitoring need to be taken into account.

Number of clusters

Different approaches were tested (with small and large amount of clusters) and the options with large amount was chosen (126 clusters).

The benefit of having a large number of clusters is that more "uncommon" cases are separated from "general population", that also can help to capture the country-based differences in data and provide more information and ideas that may help to create new rules for further monitoring.

The drawbacks are: it is harder to observe the patterns as there is more data; some clusters contain outliers that can be discovered with simple methods.

Results

Both clustering methods applied to a chosen set of variables resulted in several highly-populated groups and a large number of not-that-populated groups. The latter contain observations (transactions) that are 'different' from the general population and therefore could contain cases that may be treated as 'strange'.

As a result of the analysis of these cases a number of specific rules can be developed.

For example, a transfer that is performed:

- from Asia to Africa via offshore bank
- from Russia to Ukraine via Great Britain
- from Morocco to Morocco via France
- From Ukraine to Ukraine via Latvia

Step 7. HIGH-RISK LOCATIONS

Transfers from/to countries from the list of third countries with weak anti-money laundering and terrorist financing regimes defined by European Commission should be monitored.

Out of 619 transfers that have high-risk location at least as one side 10 have high-risk locations as both sides, 28 were sent, and 601 were received by user from such country.

Transfers that have both sides as High-risk countries

user_id	request_id	addr_country_code	sending_bank_country	recipient_country_code	payment_type	ccy_send	ccy_target	F_value	F_isBiz	transfer_sequence	days_since_previous_req	payment_status
f320ed3552f79c04da5659a20fd54da0	c24517e30e5e6df236a77209c0124884	NGA	GBR	NGA	Bank Transfer	GBP	NGN	138.0	0	5.0	3.0	Transferred
80444512a708779e3e3772abc18498e5	88c395298fa857881ae34012fb046d5d	PAK	GBR	PAK	Cards	GBP	PKR	1200.0	0	14.0	11.0	Transferred
b70ee76b0cdafb6ee6f63d1d08de03aa	4b8309486cb560c942f82a0377d9d6ee	SAU	Other/unknown	PAK	Cards	GBP	PKR	500.0	0	1.0	-1	Transferred
2c77733c83a0433c334d24d755915726	ccbb76b92d3a3c36bf1c88677695dca9	PAK	GBR	PAK	Cards	GBP	PKR	65.89	0	1.0	-1	Cancelled
f6f7c90fc99b426d4976ee1c47a32f82	b249b6cc344182b732a55c23e88eeca6	PAK	GBR	PAK	Cards	GBP	PKR	500.0	0	7.0	13.0	Transferred
11898a3fcfb7973d8b2fc97f8f49c1fa	6462f70a94c1b20ed31bc30404b5423c	PAK	GBR	PAK	Cards	GBP	PKR	500.0	0	5.0	0.0	Transferred
8bd92876be424f6eeb1eb77de7e74bb4	43c524f4cbc8d524d64f68f3d0c1710e	ETH	Other/unknown	NGA	Bank Transfer	EUR	NGN	858.7368	0	9.0	24.0	Transferred
b3ff33e9c71b30d537ef351a498a3ab0	035d803e42cb3abd25ea02cb777f26cd	PAK	Other/unknown	PAK	Direct Debit	USD	PKR	3788.524666	0	7.0	0.0	Transferred
2e555d91e04bab1c4000b1fb651a8422	6219e4ffa5c1130fda50c9f87efd5d75	NGA	GBR	NGA	Cards	GBP	NGN	200.0	0	2.0	21.0	Transferred
a45aa5eed6123131e1b30b48e8dfb7ab	0246a93390fb4ae187b914c548a1cad6	PAK	Other/unknown	PAK	Bank Transfer	DKK	PKR	1120.1765	0	7.0	58.0	Transferred

Extra. ILLOGICAL DATES

Data contains 6 time-date fields. It is not fully clear from the description and my understanding of the processes the logic behind some "uncommon" cases.
For example:

- 5 transfers that have 'date_request_submitted' earlier than 'date_user_created'
- 19 transfers have 'date_request_transferred' earlier than 'date_request_received'

These data needs additional clarification for further analysis.