ANALYSIS OF SUSPICIOUS ACTIVITY IN ETHEREUM

GUIDED BY DR. PRADHAN

PRESENTED BY BHAVYA

Introduction

- Ethereum openness makes it susceptible to various frauds
- Detecting anomalies in this vast data is required to enable trust and security
- Here, i have done research and analysis on ethereum transactions for finding anomalies
- provided an analytical approach for identifying them
- Detecting these is must because the anomalies might lead to a huge financial loss later on.

Ethereum

- Financial blockchain network
- Two types of account:
  - smart contract
  - Externally owned account (EOA)
- More than only cryptocurrency transfer

Types of accounts

EOA

For Ether transfer and controlled by private keys.

Smart Contract

Controlled by code and cant be executed on its own.

Why two types of account?

EOA lack the ability to store logic and automate process. Smart contract remove need for intermediatries by automation of transfers by predefined code.

Types of transactions

- EOA to EOA Transactions
- EOA to Smart Contract Transactions
- Smart Contract to Smart Contract Transactions
- Smart Contract to EOA Transactions

Previous work

- Did analysis over ethreum network for better understanding of blockchain network mainly ethereum
- Did node classification of accounts present in the network for knowing, how efficient a graph based model is over ethereum and to make it a starting for anomaly detection

Literature Review

Abnormal Transaction patterns like dust transactions, or sudden spike in activity etc

Large amount of cryptocurrency move across multiple contracts without clear reasons

Strange behaviour pattern of contracts during deployment, execution, termination phases.

Irregular time intervals between transactions or contract events

Problem

How can we effectively detect and interpret unusual patterns and anomalies in a decentralized financial system like Ethereum, where the absence of centralized control makes the network more susceptible to malicious behavior, smart contract misuse, and financial manipulation?

what is anomaly?

- Pattern or behaviour which deviates from original data/transactions
- anomalies can arise due to malicious activities, system bugs, automated bot behavior, or even misuse of smart contracts
- Detecting anomalies is essential for identifying potential fraud, security threats, or unintended behavior in decentralized applications

Types of Anomalies

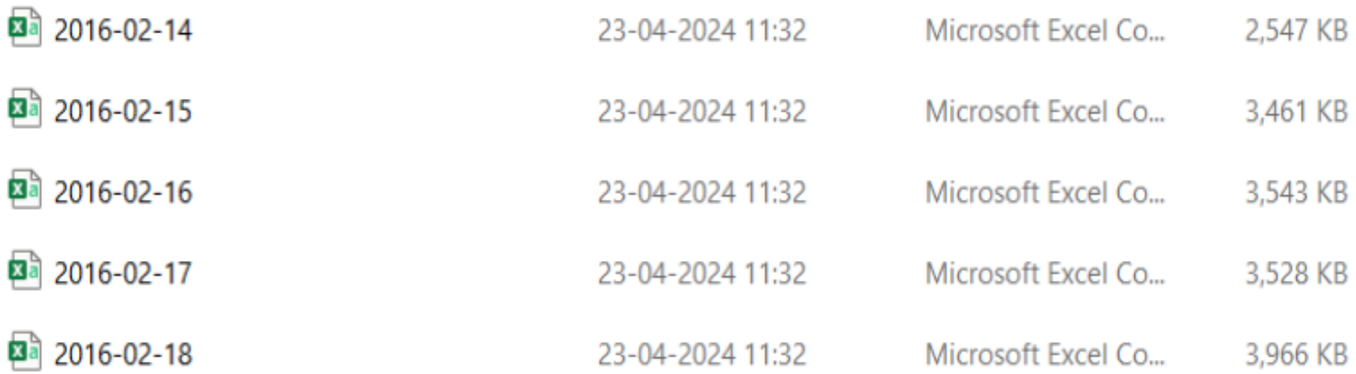
- Accounts taking Ether but never transferring it out
- Bot activity - A sudden spike in new contract deployments from a single address.

- Dual nature of accounts
- Zero value transactions which can be used for signalling
- Unusual number of transactions from an address, potential scam. (money laundering)

Preprocessing

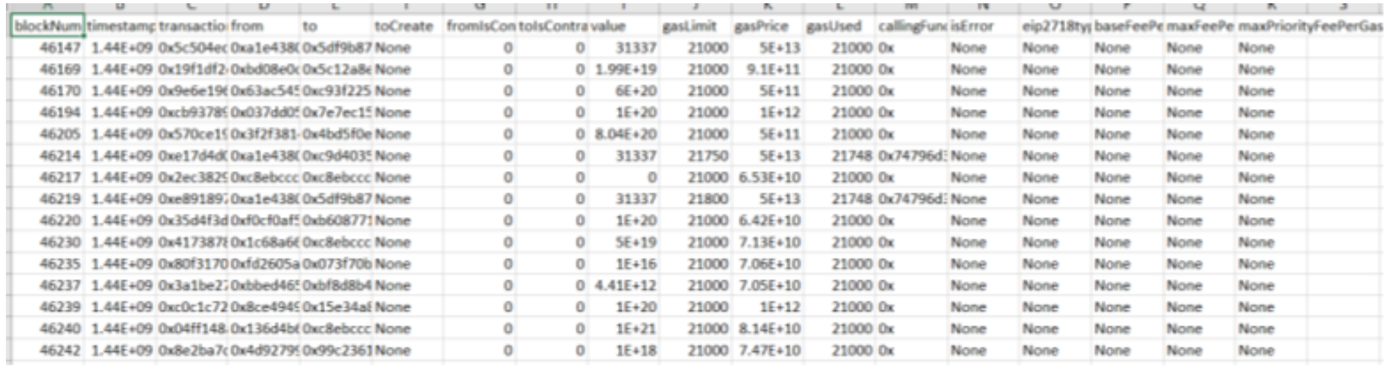
Splitting dataset daywise

- It is done by converting the UNIX timestamp to regular format of year-month-day and we have splitted it into daywise for easy computation.

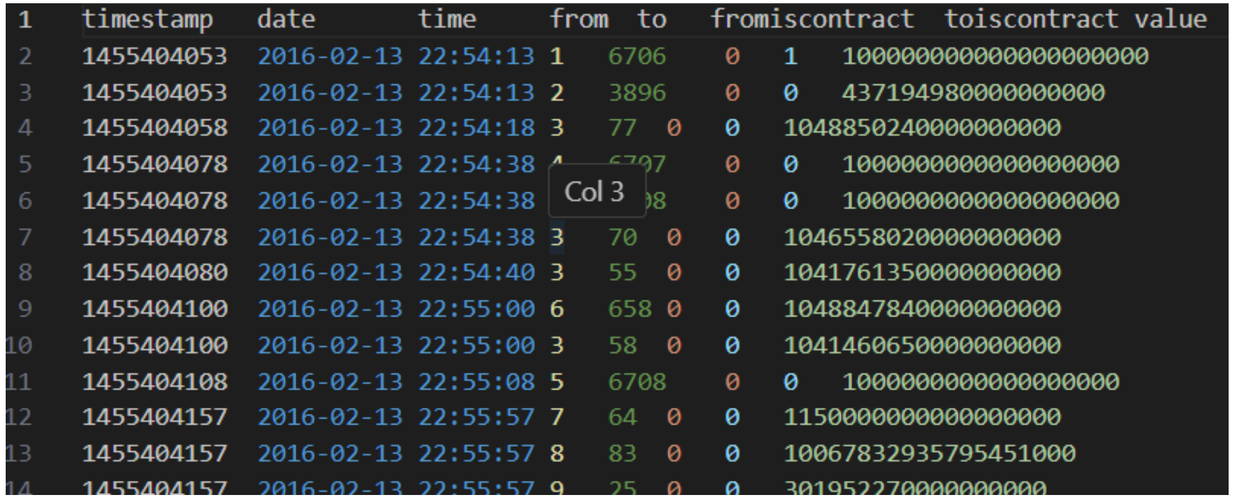


Reduce the storage of files

We have reduced by removing the unwanted columns from the csv files, so that the file size is reduced and might decrease the computation time



blockNum	timestamp	transactio	from	to	toCreate	fromIsCon	toIsContra	value	gasLimit	gasPrice	gasUsed	callingFun	isError	eip2718ty	baseFeePe	maxFeePe	maxPriorityFeePerGas
46147	1.44E+09	0x5c504ec	0xa1e438(	0x5df9b87	None	0	0	31337	21000	5E+13	21000	0x	None	None	None	None	None
46169	1.44E+09	0x19f1df2	0xbd08e0(	0x5c12a8(	None	0	0	1.99E+19	21000	9.1E+11	21000	0x	None	None	None	None	None
46170	1.44E+09	0x9e6e19(	0x63ac54(	0xc93f225	None	0	0	6E+20	21000	5E+11	21000	0x	None	None	None	None	None
46194	1.44E+09	0xcb9378(	0x037dd0(	0x7e7ec1(	None	0	0	1E+20	21000	1E+12	21000	0x	None	None	None	None	None
46205	1.44E+09	0x570ce1(	0x3f2f381	0x4bd5f0e	None	0	0	8.04E+20	21000	5E+11	21000	0x	None	None	None	None	None
46214	1.44E+09	0xe17d4d(	0xa1e438(	0xc9d403(	None	0	0	31337	21750	5E+13	21748	0x74796d:	None	None	None	None	None
46217	1.44E+09	0x2ec382(	0xc8ebccc	0xc8ebccc	None	0	0	0	21000	6.53E+10	21000	0x	None	None	None	None	None
46219	1.44E+09	0xe89189(	0xa1e438(	0x5df9b87	None	0	0	31337	21800	5E+13	21748	0x74796d:	None	None	None	None	None
46220	1.44E+09	0x35d4f3d	0xf0cf0af(	0xb60877(	None	0	0	1E+20	21000	6.42E+10	21000	0x	None	None	None	None	None
46230	1.44E+09	0x417387(	0x1c68a6(	0xc8ebccc	None	0	0	5E+19	21000	7.13E+10	21000	0x	None	None	None	None	None
46235	1.44E+09	0x80f3170	0xfd2605a	0x073f70b	None	0	0	1E+16	21000	7.06E+10	21000	0x	None	None	None	None	None
46237	1.44E+09	0x3a1be2(	0xbbed46(	0xbf8d8b4	None	0	0	4.41E+12	21000	7.05E+10	21000	0x	None	None	None	None	None
46239	1.44E+09	0xc0c1c72	0x8ce494(	0x15e34af	None	0	0	1E+20	21000	1E+12	21000	0x	None	None	None	None	None
46240	1.44E+09	0x04ff148.	0x136d4b(	0xc8ebccc	None	0	0	1E+21	21000	8.14E+10	21000	0x	None	None	None	None	None
46242	1.44E+09	0x8e2ba7(	0x4d9279(	0x99c2361	None	0	0	1E+18	21000	7.47E+10	21000	0x	None	None	None	None	None



	timestamp	date	time	from	to	fromiscontract	toiscontract	value
2	1455404053	2016-02-13	22:54:13	1	6706	0	1	10000000000000000000
3	1455404053	2016-02-13	22:54:13	2	3896	0	0	437194980000000000
4	1455404058	2016-02-13	22:54:18	3	77	0	0	1048850240000000000
5	1455404078	2016-02-13	22:54:38	4	6707	0	0	100000000000000000
6	1455404078	2016-02-13	22:54:38		8	0	0	100000000000000000
7	1455404078	2016-02-13	22:54:38	3	70	0	0	1046558020000000000
8	1455404080	2016-02-13	22:54:40	3	55	0	0	1041761350000000000
9	1455404100	2016-02-13	22:55:00	6	658	0	0	1048847840000000000
10	1455404100	2016-02-13	22:55:00	3	58	0	0	1041460650000000000
11	1455404108	2016-02-13	22:55:08	5	6708	0	0	100000000000000000
12	1455404157	2016-02-13	22:55:57	7	64	0	0	1150000000000000000
13	1455404157	2016-02-13	22:55:57	8	83	0	0	10067832935795451000
14	1455404157	2016-02-13	22:55:57	9	25	0	0	301952270000000000

Col 3

From-to mapping

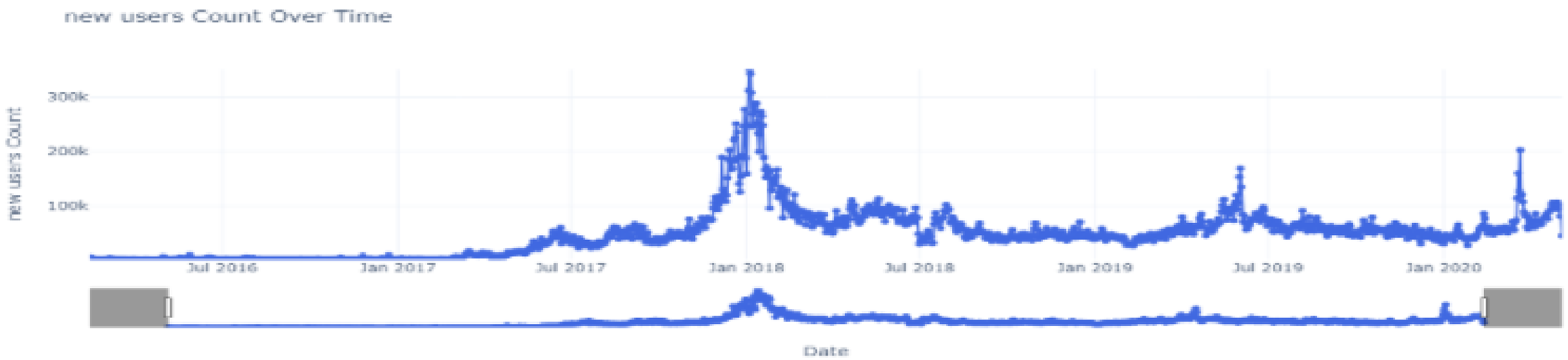
- Now, we need to create a graph for the from and to mapping
- Create dictionary for storing mapped values

from	to
0x39fa8c5f2793459d6622857e7d9fbb4bd91766d3	0xc083e9947cf02b8ffc7d3090ae9aea72df98fd47
0x32be343b94f860124dc4fee278fdcbd38c102d88	0xdf190dc7190dfba737d7777a163445b7fff16133
0x2a65aca4d5fc5b5c859090a6c34d164135398226	0x819f4b08e6d3baa33ba63f660baed65d2a6eb64c
0x2910543af39aba0cd09dbb2d50200b3e800a63d2	0x9e486ad335492959c38a0740cb66c55ad30bb4f0
0x10d5bff7879b7eb5192b3374338bb834981910a8	0xc6c764fc6c1e1211d2b4a06ef2170f660a4512fa
0x2a65aca4d5fc5b5c859090a6c34d164135398226	0x53e0551a1e31a40855bc8e086eb8db803a625bbf
0x2a65aca4d5fc5b5c859090a6c34d164135398226	0x51033f1a1a59cb6a1bf6ca2087a53bd202ac1c83
0x120a270bbc009644e35f0bb6ab13f95b8199c4ad	0x3dc12a32a5abf477e2ec91f6218d0a96150fef99
0x2a65aca4d5fc5b5c859090a6c34d164135398226	0xf4f2c15602b084cae84ea603f75527de19705aa1

-->

1	2
3	4
3	145
3	66
3	66
3	146

- I have plotted new users vs date

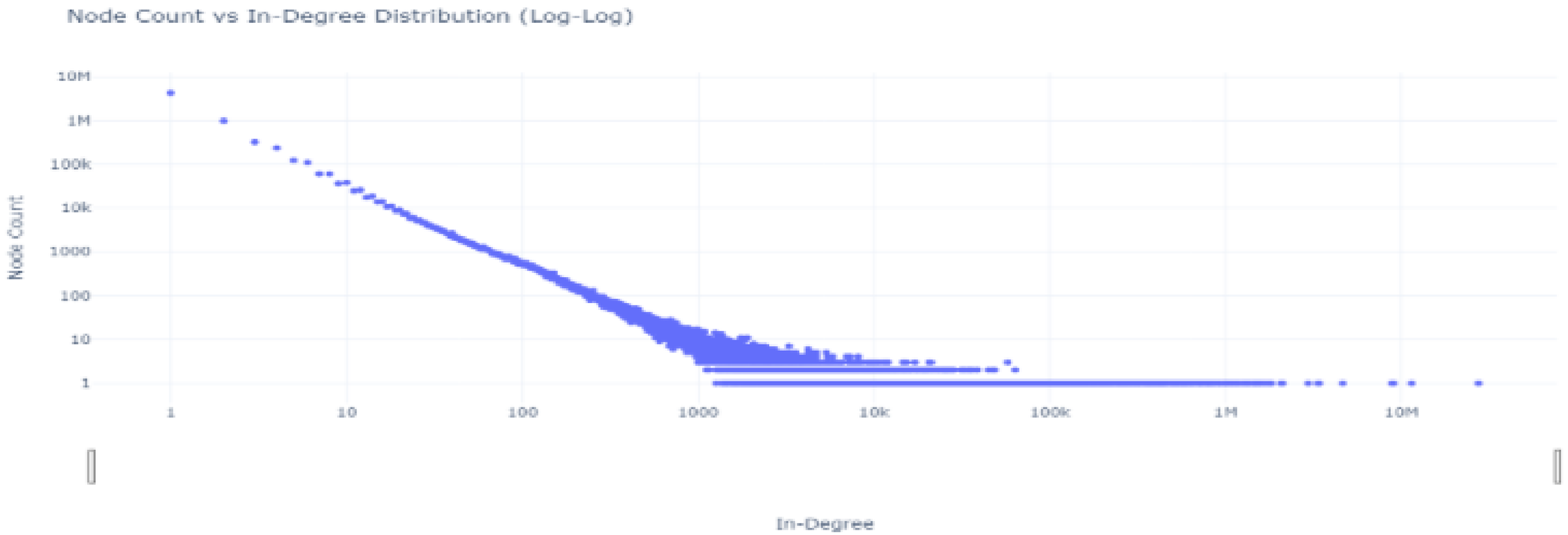


I have plotted number of transactions vs date

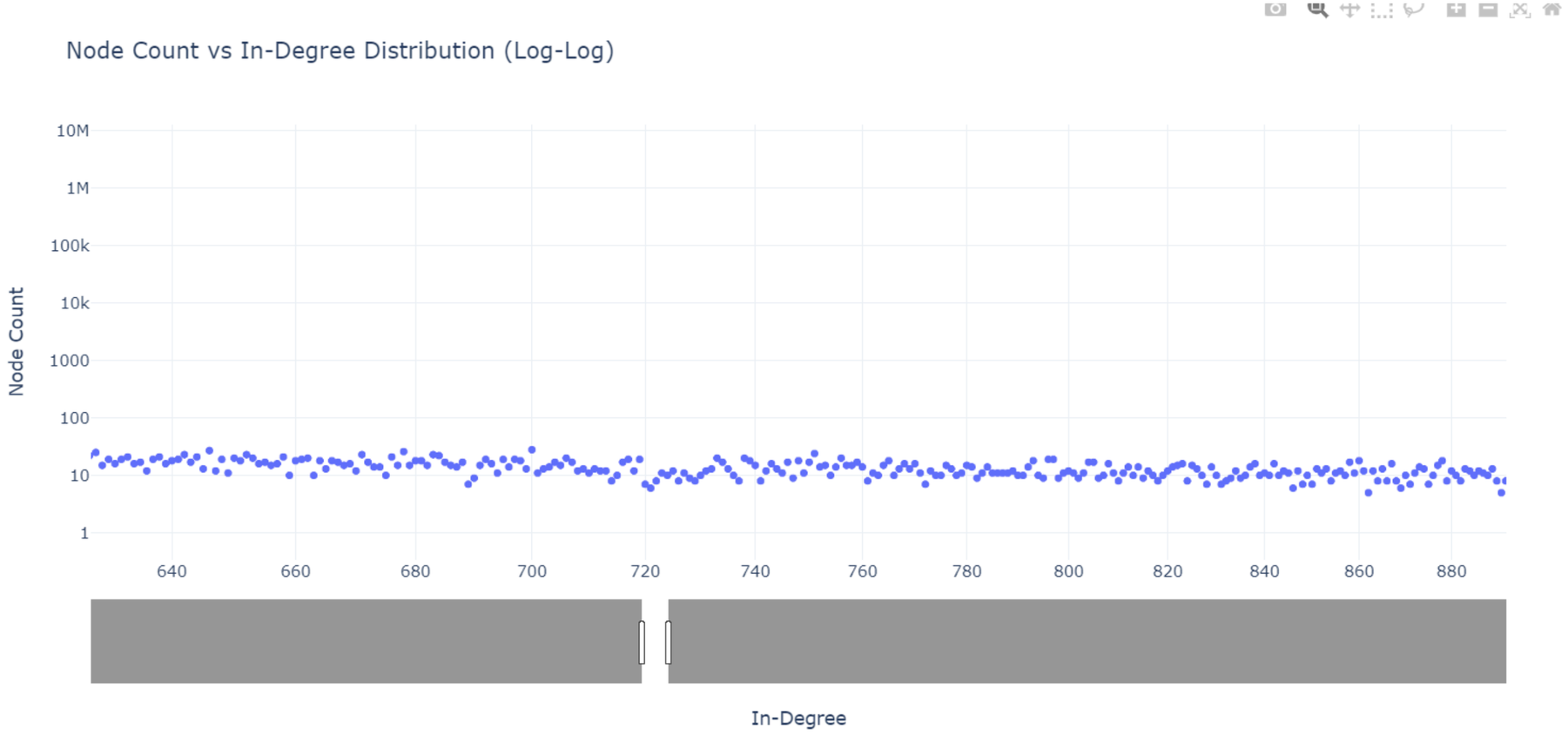
- I have extarcted degrees and weighted degrees for each node
- Extracted nodes which satisfy 4 different cases into separately files

1. In-degree>0 and out-degree=0
2. in-degree=0 and out degree >0
3. weighted in-degree >0 and weighted out-degree=0
4. weighted in-degree =0 and weighted out-degree>0

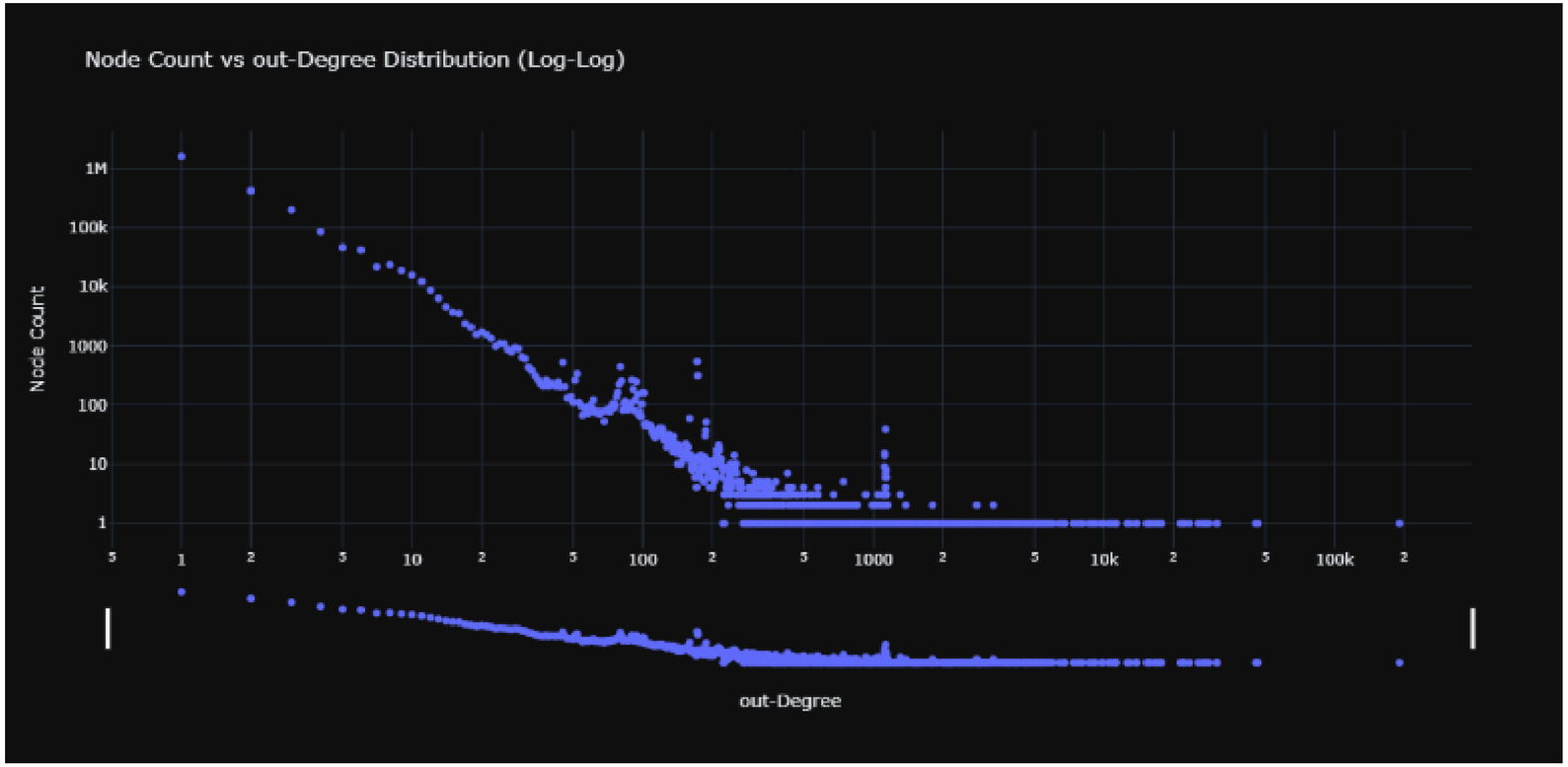
- This is log log plot of the degree distribution of node count vs indegree



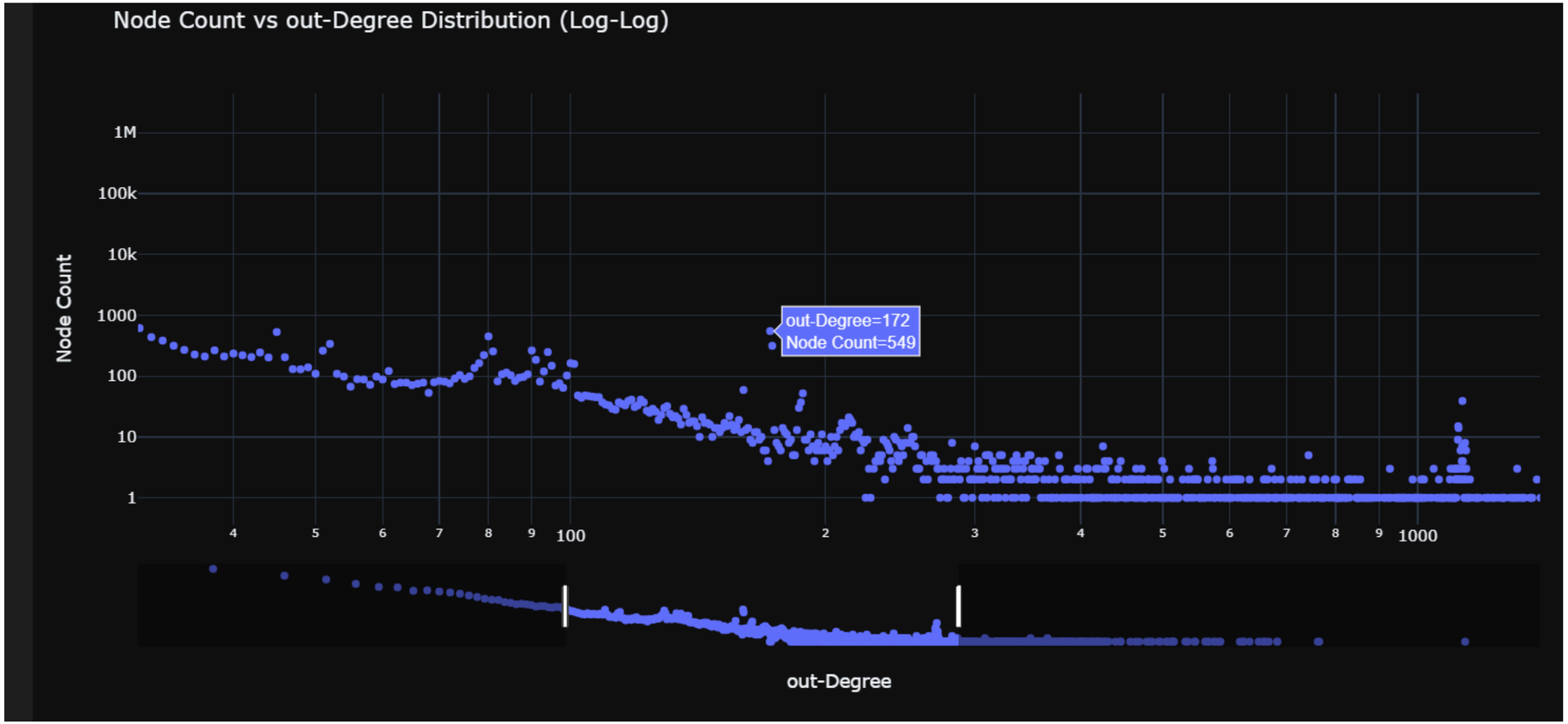
This is the zoomed small parted of the above graph



This is log log plot of the degree distribution of node count vs outdegree



This is the zoomed small parted of the above graph



By comparing these four files i have taken these three conditions for finding the anomaly

1.in-degree>0 and weighted in-degree =0 **case1**

2.out-degree>0 and weighted out-degree =0 **case2**

3.both degree>0 and weighted degrees=0 **case3**

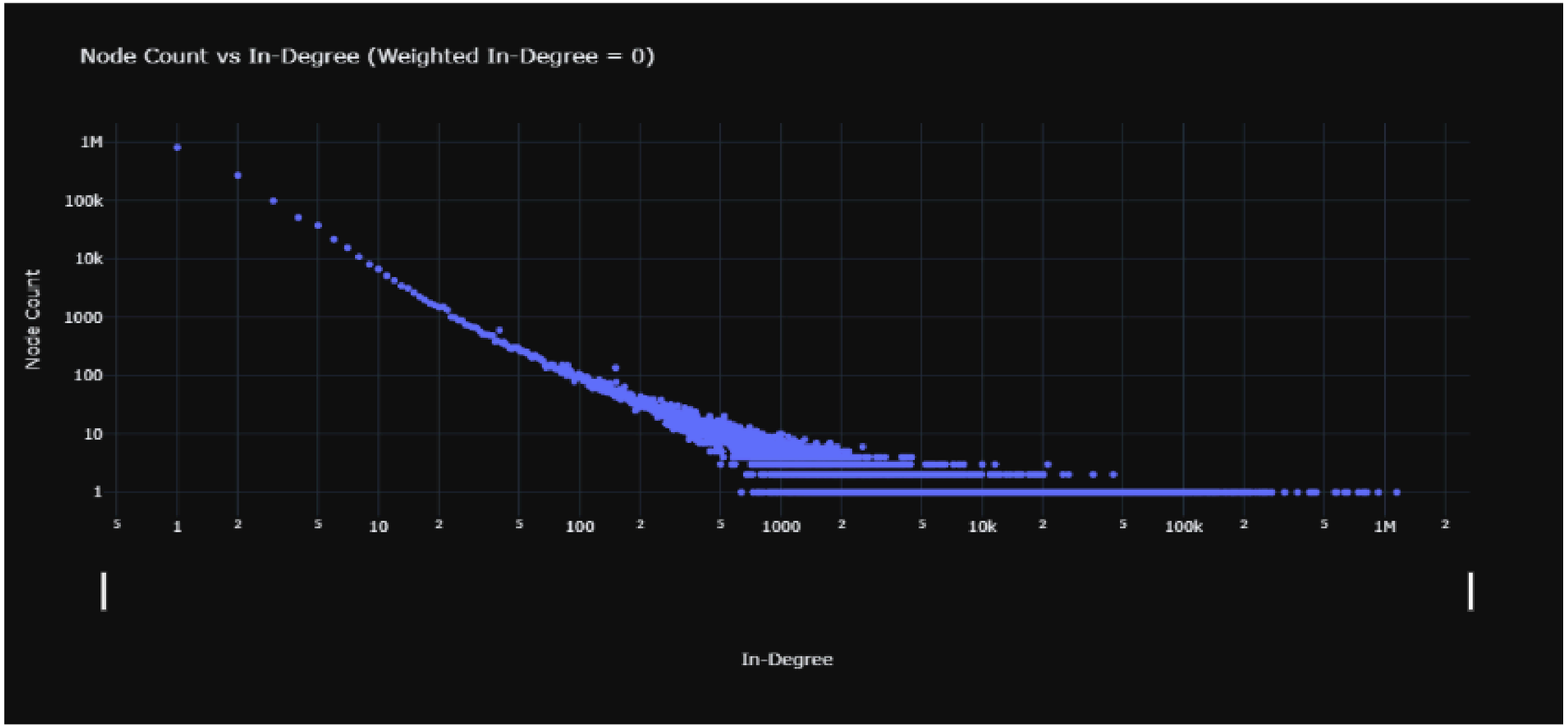
this is the final file which i have extracted

```
node_id in_degree out_degree weighted_in_degree weighted_out_degree contract
295 1 2172 0 296103046448915687669760
841 1 7 0 1404990671143759680
1833 0 21 0 0 0
1913 0 3 0 0 0
1920 2 837 3500000000000000000 0 0
2173 0 3 0 0 0
2888 1 2 500000000000000000 0 0
2951 0 3 0 0 0
3109 0 2 0 0 0
3141 2 2 2800000000000000000 0 0
3528 0 10 0 0 0
3545 4 1 18929668330262868480 0
4507 0 1 0 0 0
4942 1 1 100000000000000000 0 0
```

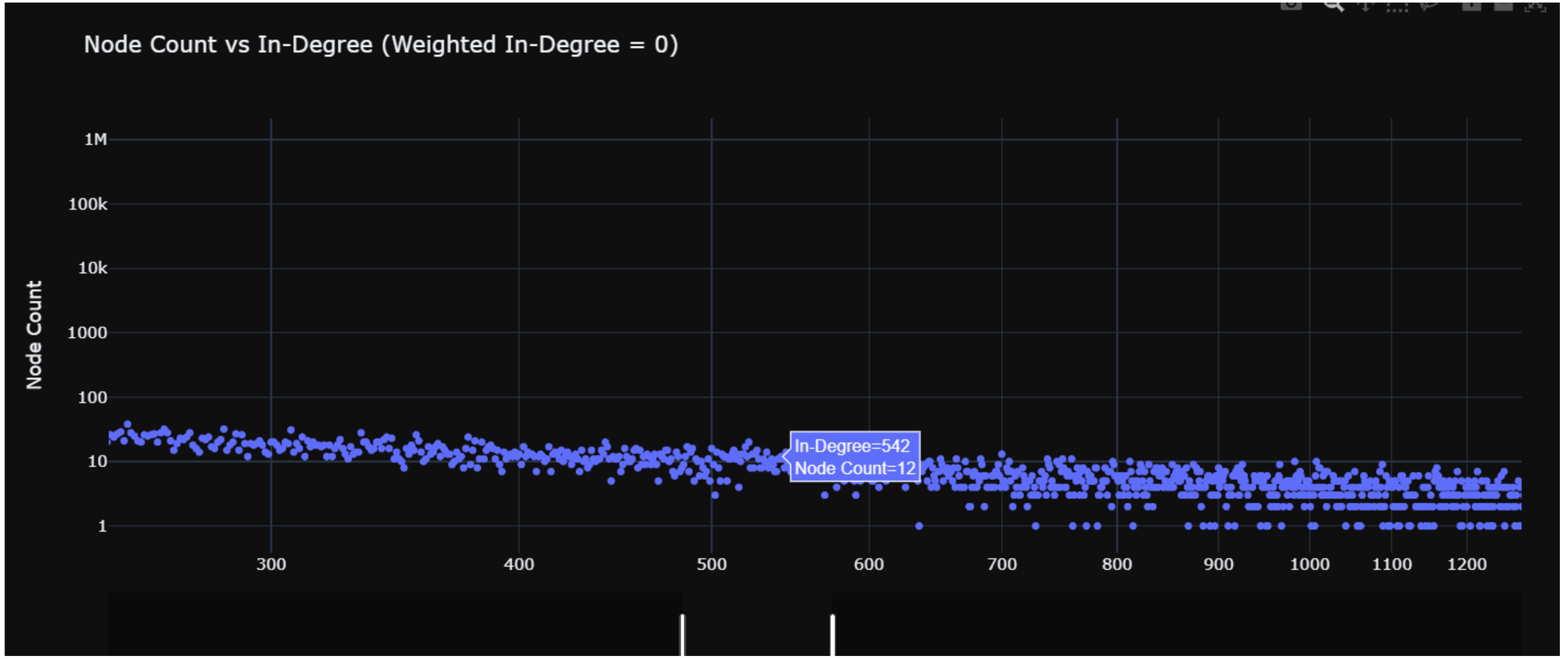
why only those why not these conditions

- High In-Degree and High Weighted In-Degree
- High Out-Degree and High Weighted Out-Degree
- Both Degrees Are High and So Are Weighted Degrees
- Low Degree, Low Weighted Degree
- Non-Zero Degree, Low Weighted Degree
- High Out-Degree, Zero In-Degree
- High Degree, Moderate Weighted Degree
- In-Degree > Out-Degree, and Weighted In > Weighted Out
- Degrees are Equal, Weighted In ≠ Weighted Out

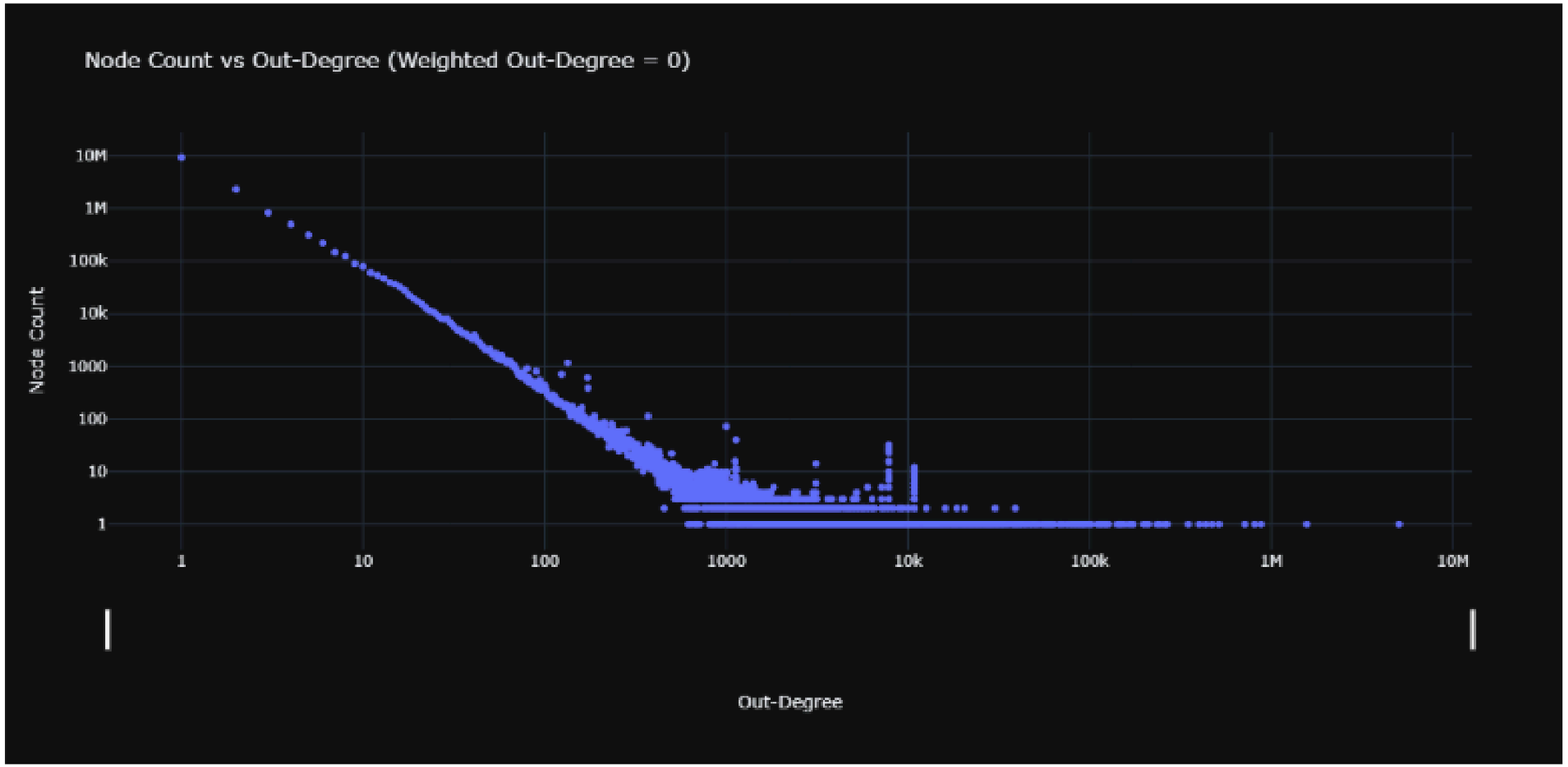
The condition is indegree>0 and weighted indegree=0
Here is the plot in log log scale saying the degree distribution for the nodes



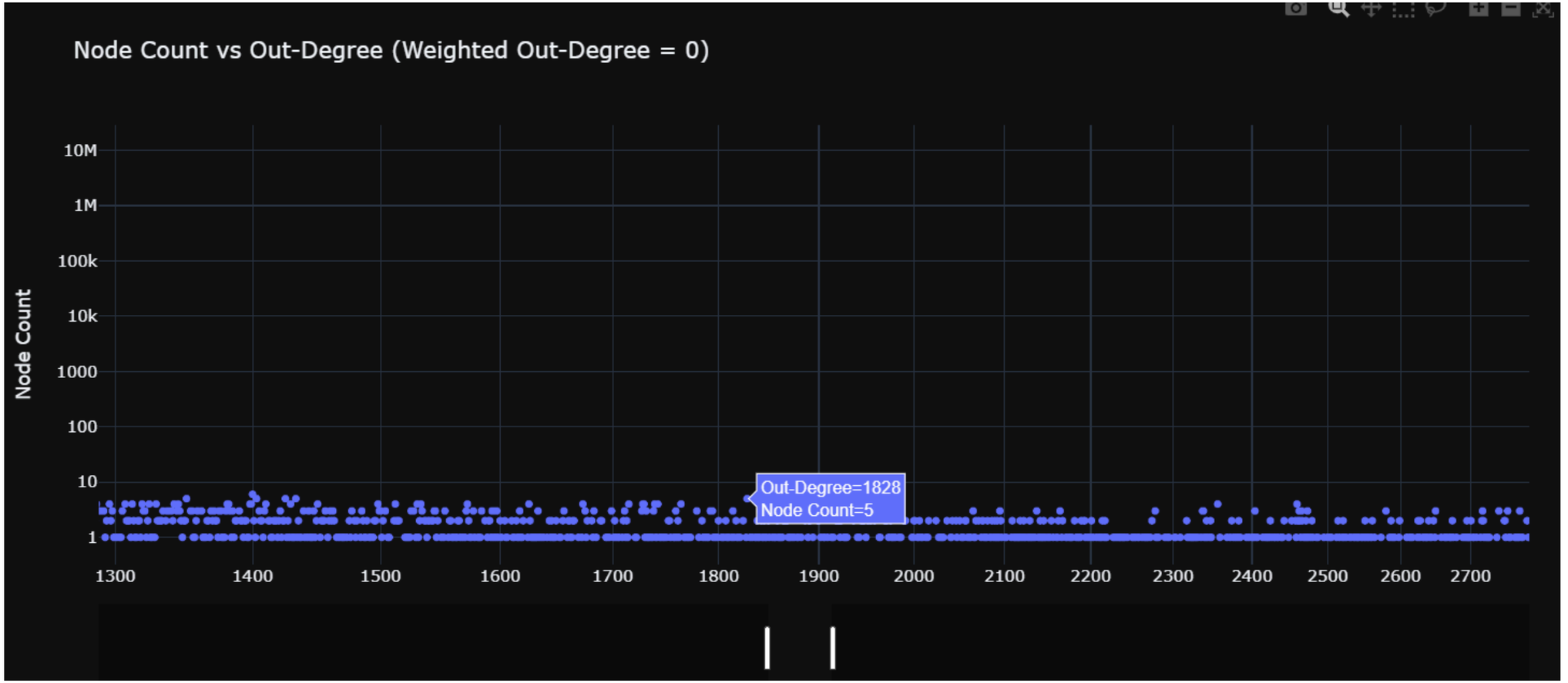
This is the zoomed small parted of the above graph



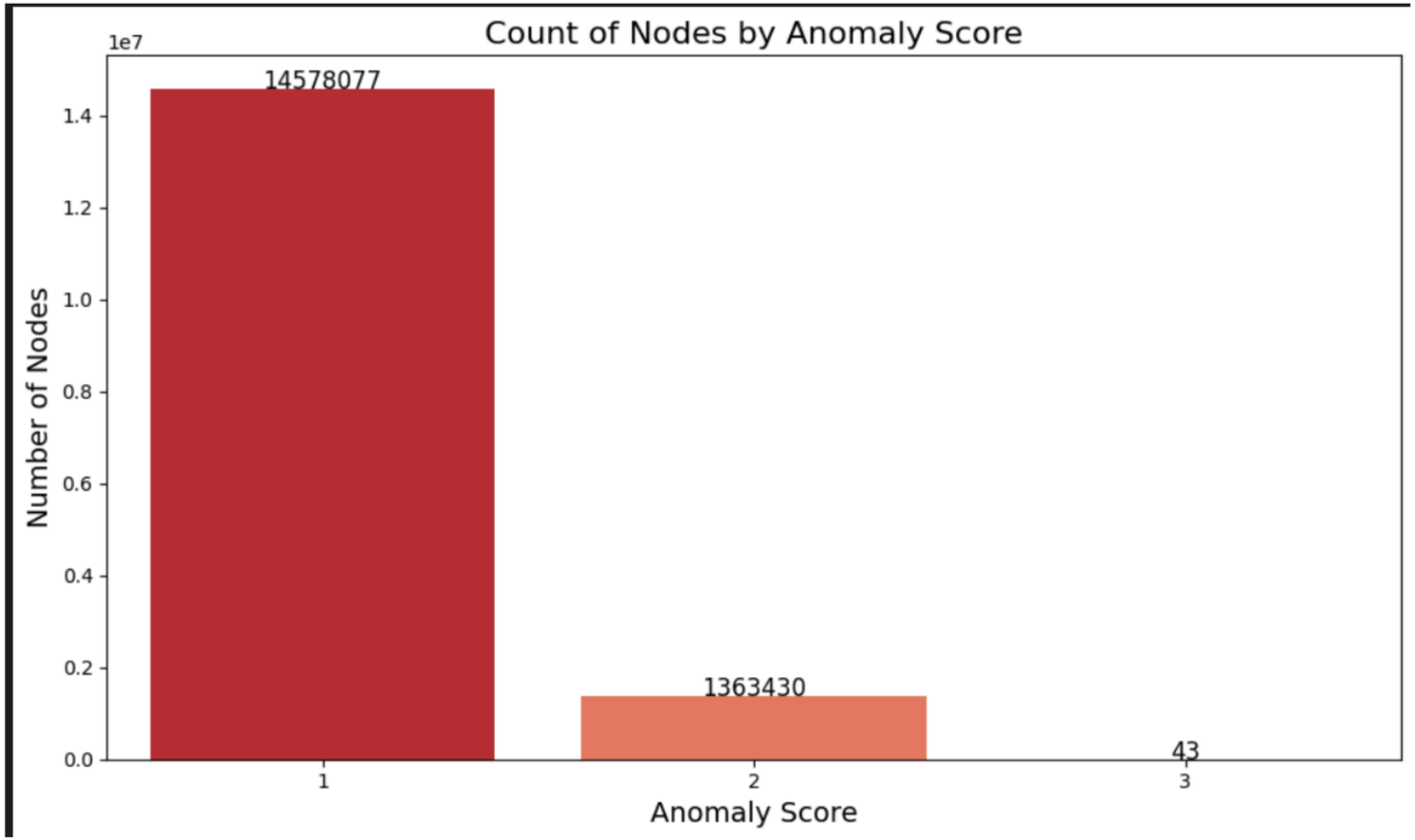
The condition is out-degree>0 and weighted out-degree =0
Here is the plot in log log scale saying the degree distribution for the nodes



This is the zoomed small parted of the above graph



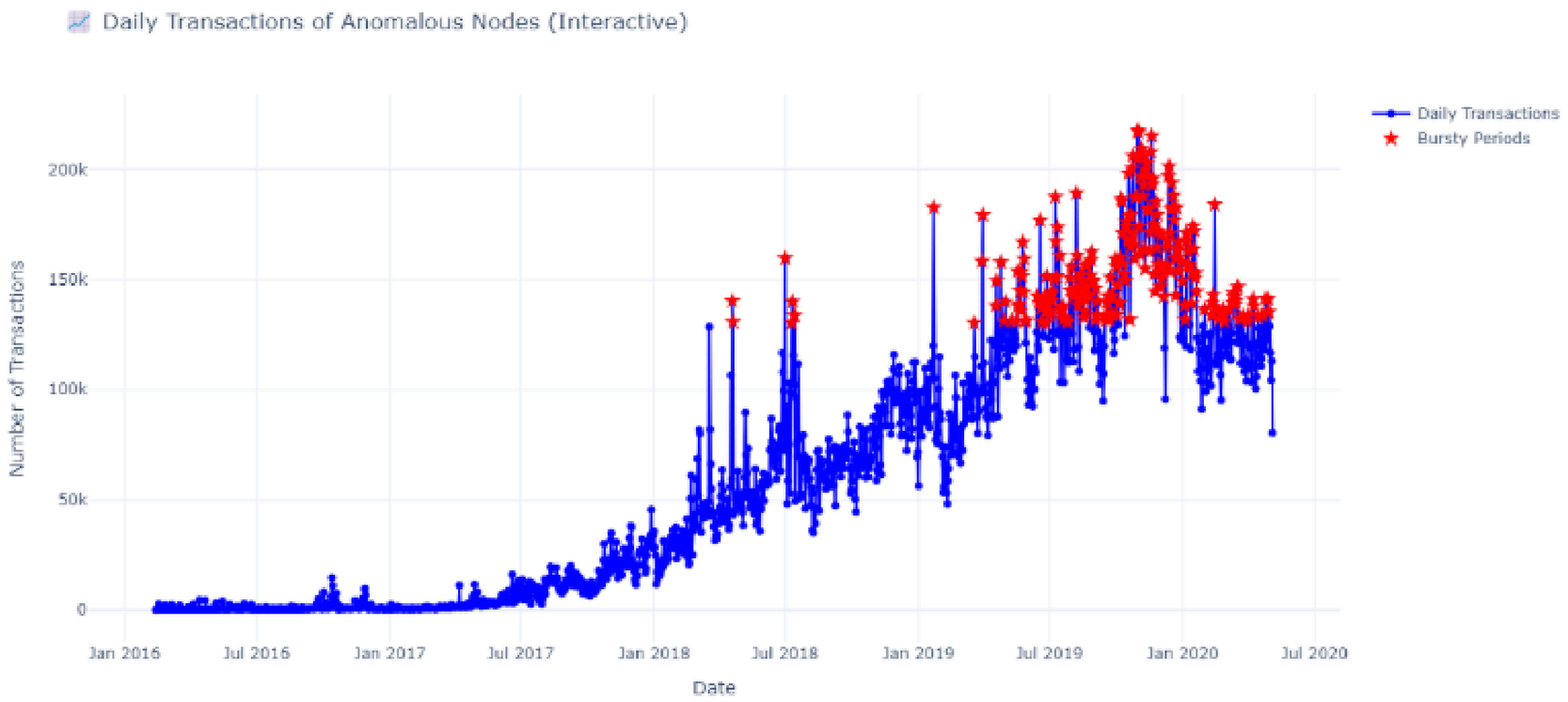
- Assigned scores and extracted nodes with high score as abnormal. Scores are assigned as follows:
  -both in and out degree>0 but both weighted in and out degree =0 then the score is 2.
  -if in degree>0 and weighted in degree =0 then the score is 1.
  -if out degree>0 and weighted out degree=0 then the score is 1.
  -Now, add score to the previous existing one like new score= previous score+ contract value

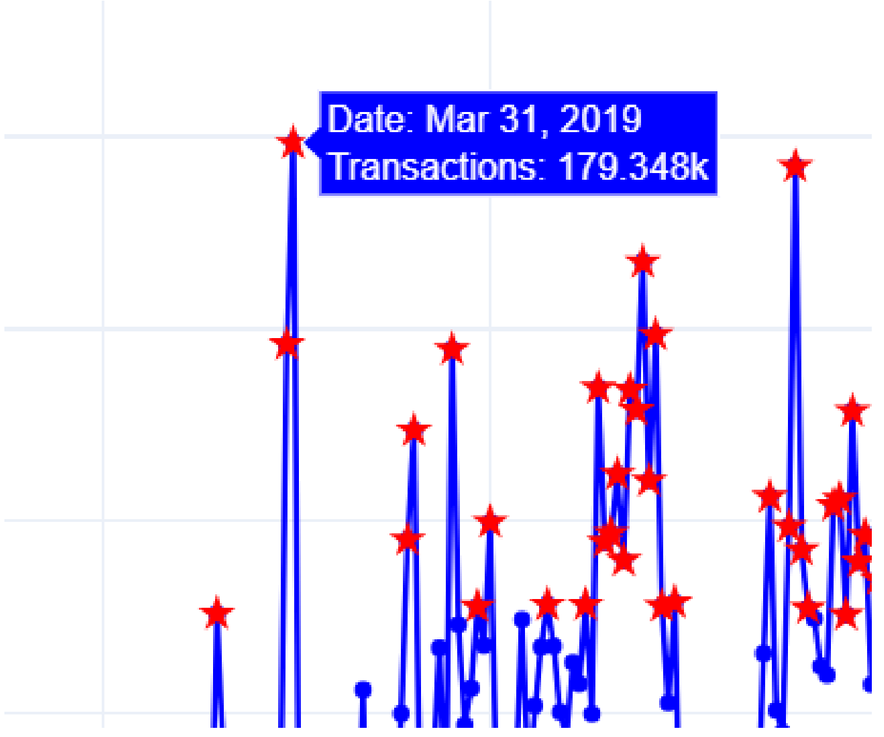


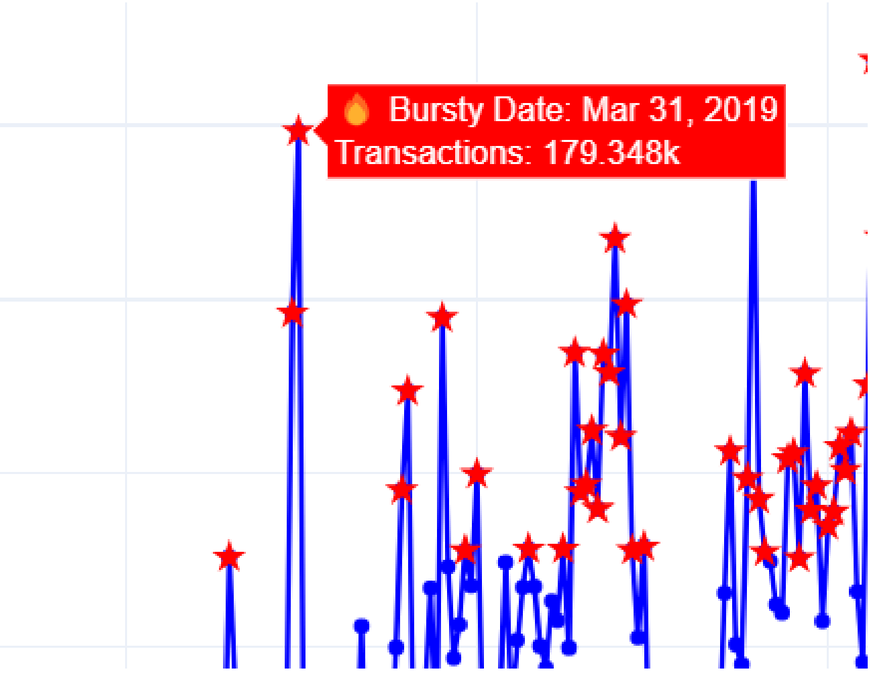
- Observed the transactions of these nodes and plotted.
- Calculated burst periods for these nodes by taking few thresholds
- Burst means finding the high activity time periods for a node when compared to remaining once
- Threshold in this burst activity refers to the count of transactions in which this particular node has greater value than most of the remaining nodes
- I have considered many thresholds for finding the optimal one and found 95% as optimal one.

Result

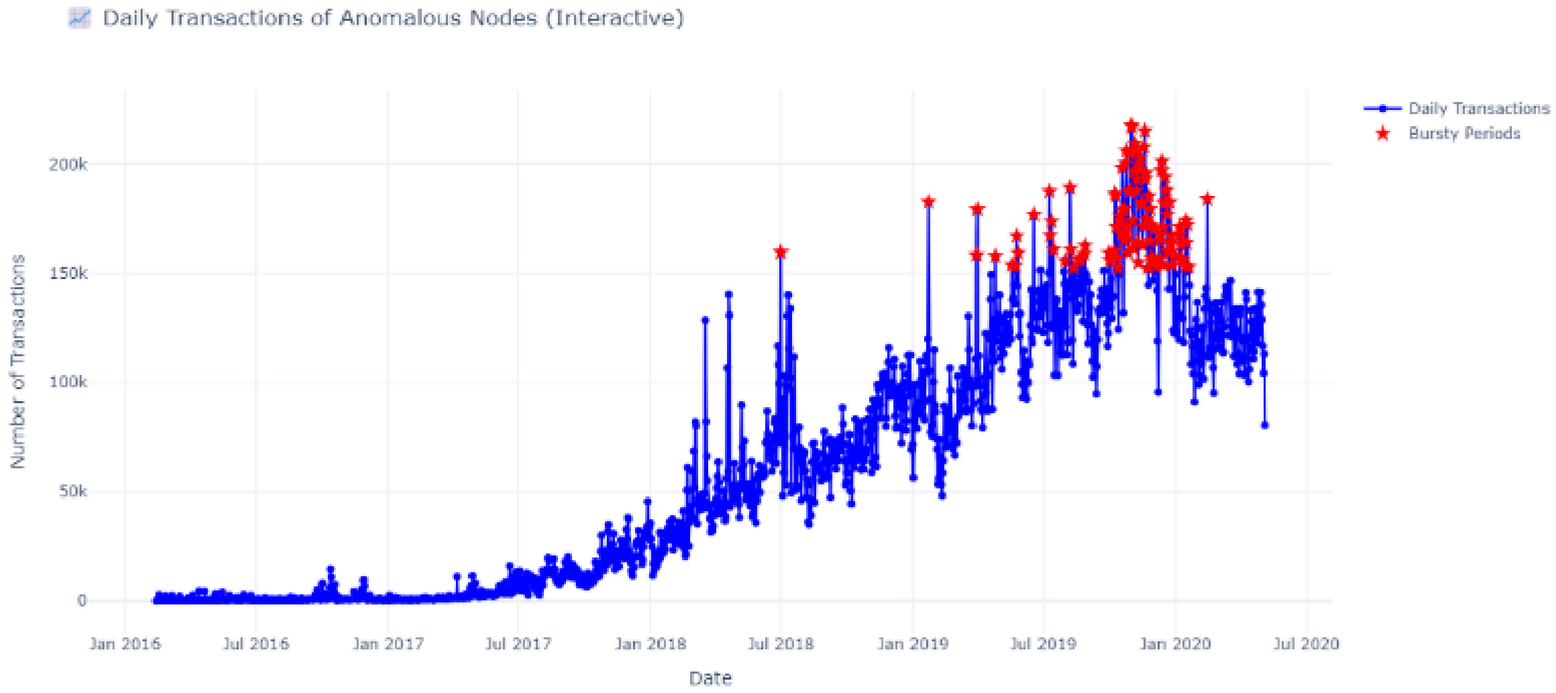
- Here i have plotted the no of transactions vs date for those anomaly nodes only and brust threshold is 85 percent

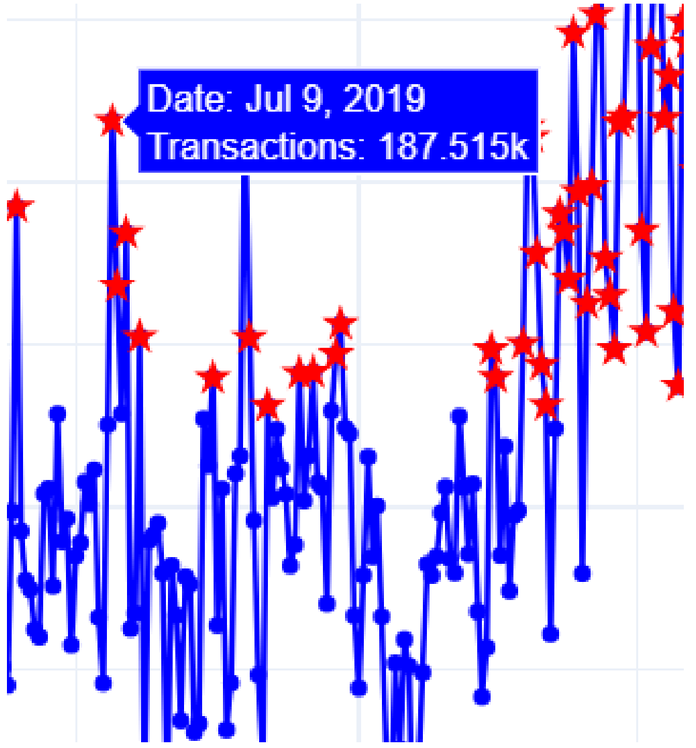


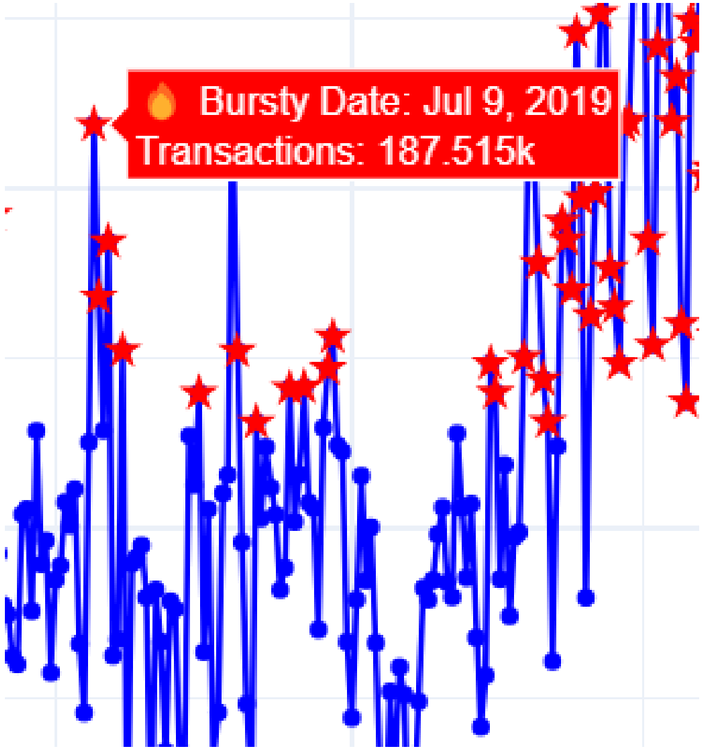

Date: Mar 31, 2019
Transactions: 179.348k


Bursty Date: Mar 31, 2019
Transactions: 179.348k

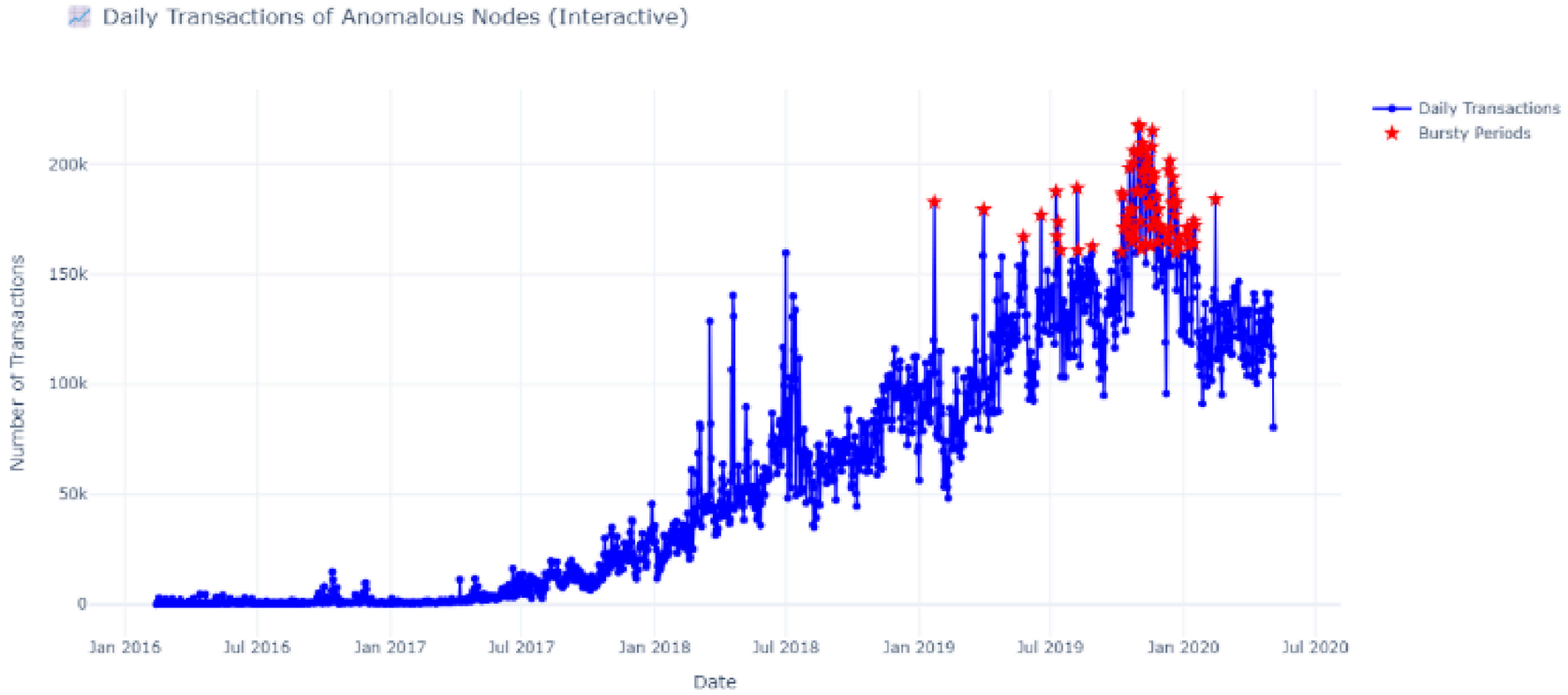
Here i have plotted the no of transactions vs date for those anomaly nodes only and brust threshold is 93 percent

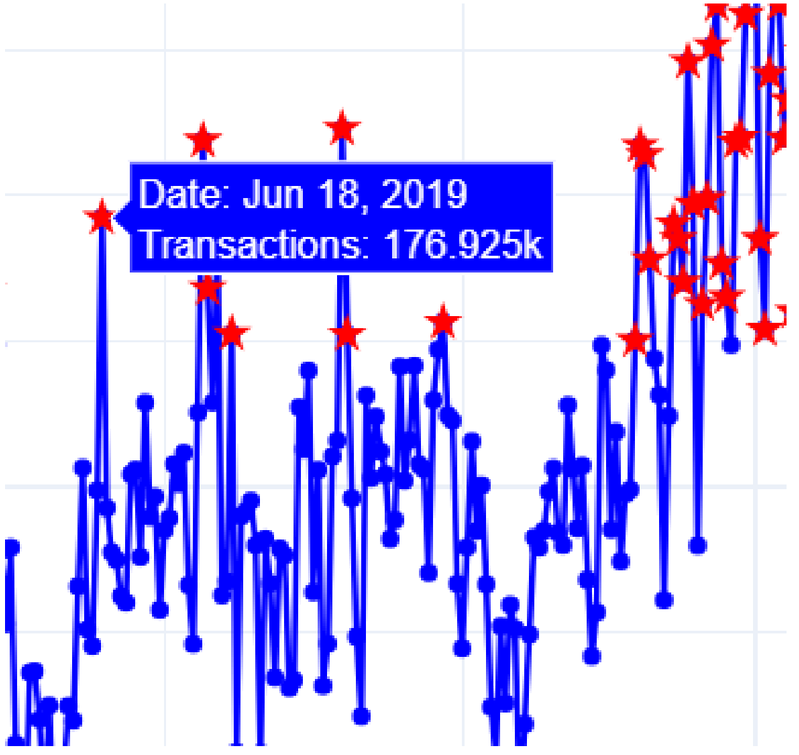


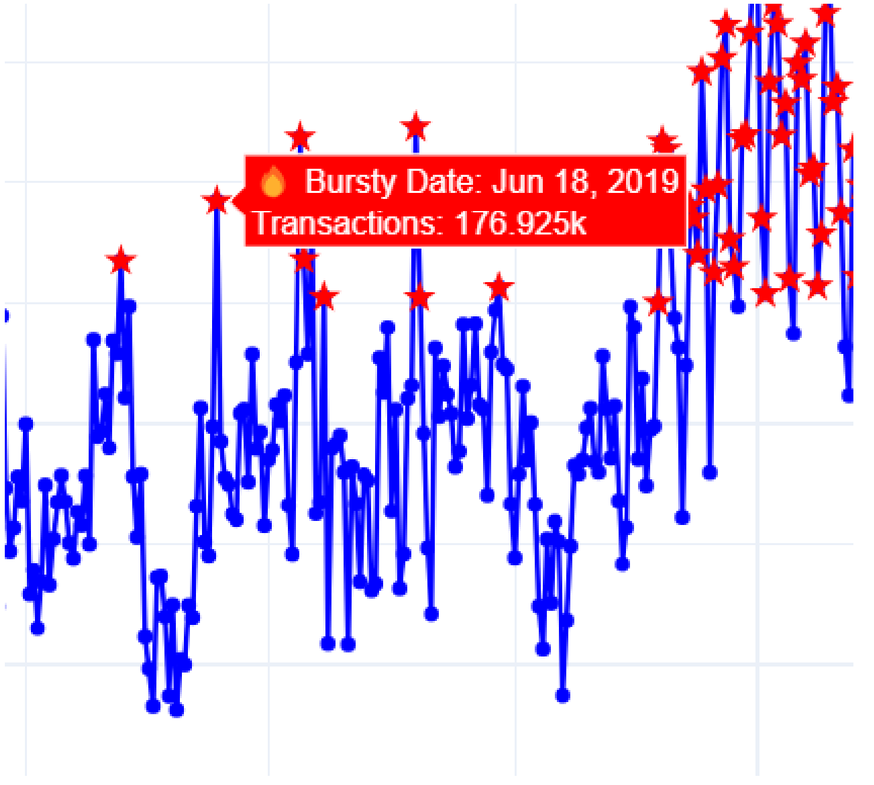

Date: Jul 9, 2019
Transactions: 187.515k


Bursty Date: Jul 9, 2019
Transactions: 187.515k

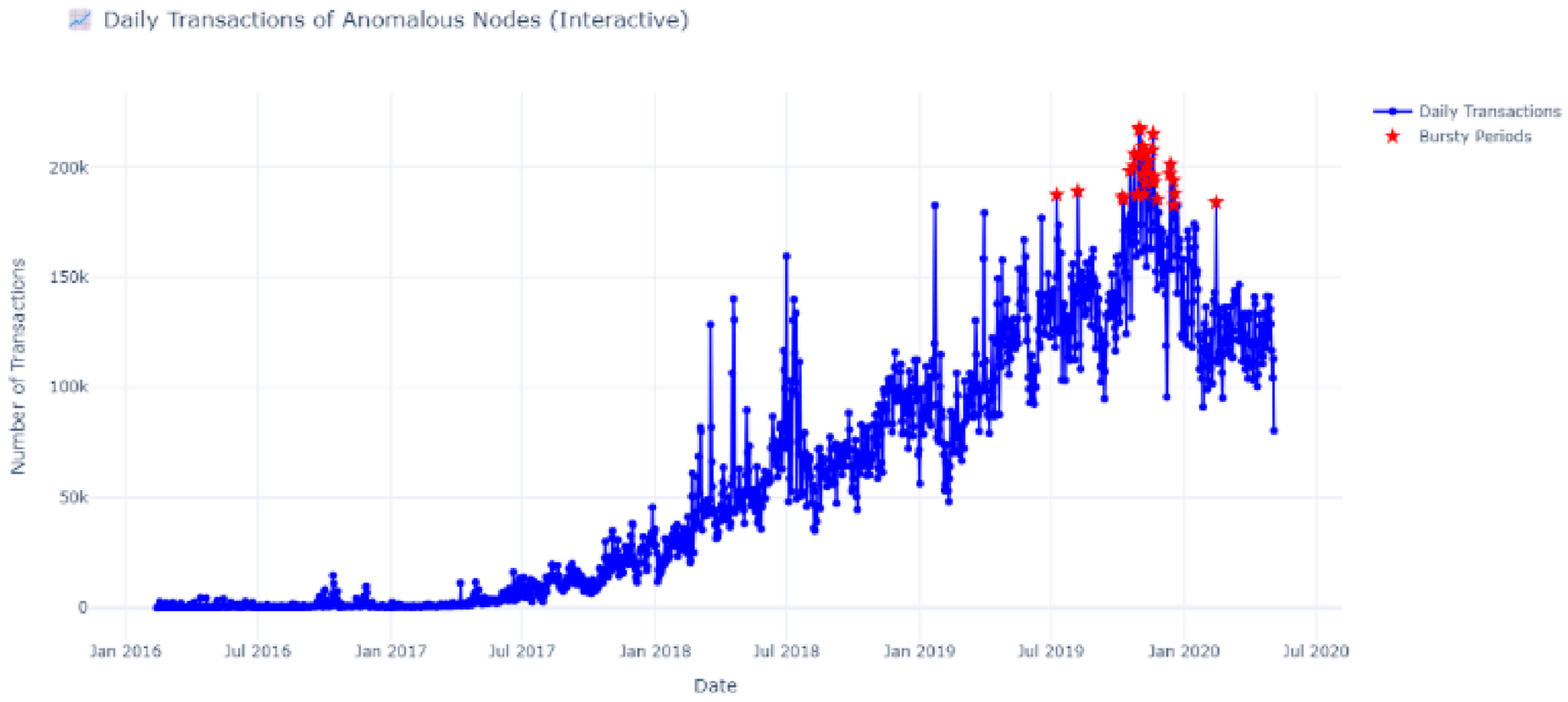
Here i have plotted the no of transactions vs date for those anomaly nodes only and burst threshold is 95 percent

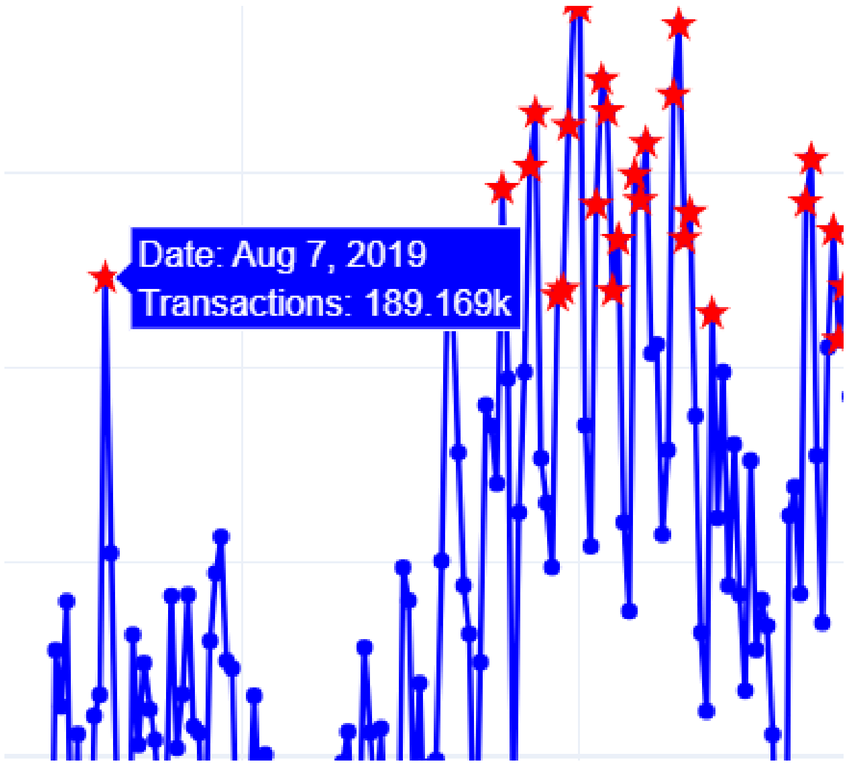


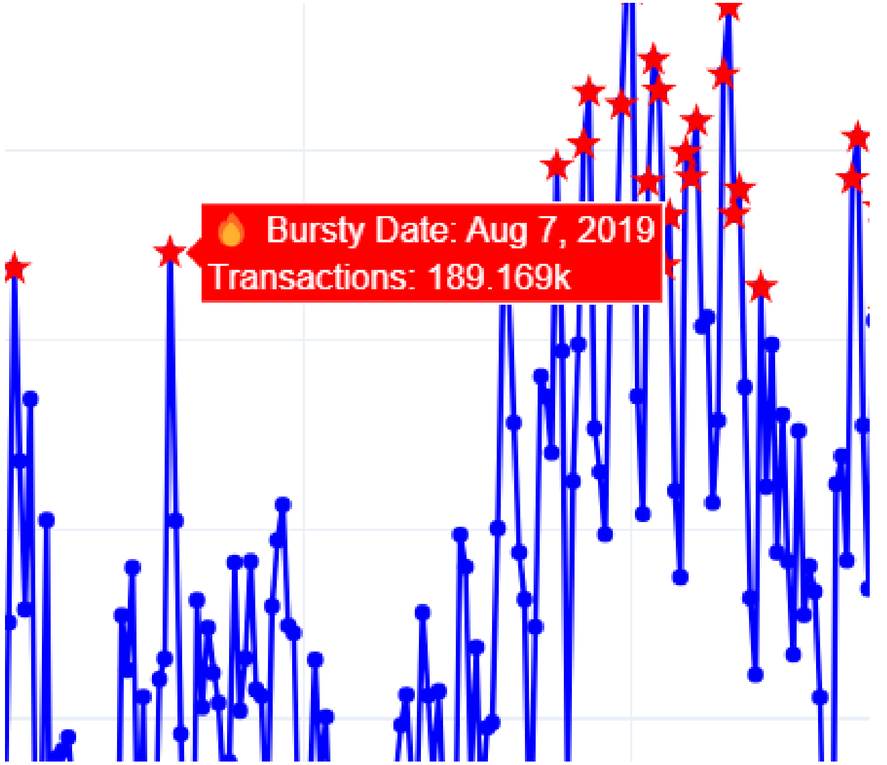

Date: Jun 18, 2019
Transactions: 176.925k


Bursty Date: Jun 18, 2019
Transactions: 176.925k

Here i have plotted the no of transactions vs date for those anomaly nodes only and burst threshold is 98 percent




Date: Aug 7, 2019
Transactions: 189.169k


Bursty Date: Aug 7, 2019
Transactions: 189.169k

Methodology

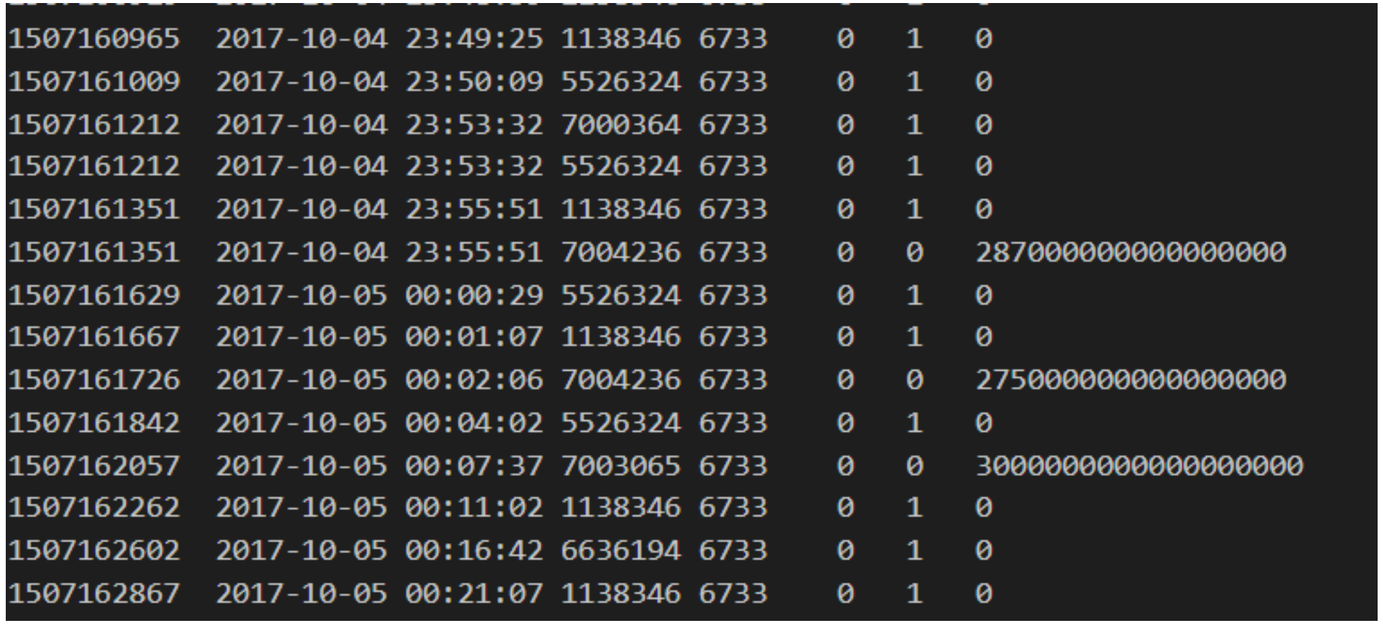
- Have identified dual nature of nodes like checking in between if any node contract status is flipping or not

Node	Contract_Type
6733	2
265660	2
274298	2
274307	2
343726	2
368581	2
406867	2
410730	2
410900	2
417618	2
426376	2

- Extracted such nodes and also transactions of those nodes along with the time period
- here the two same nodes having the 8 transactions with in same nodes in one sec

7	1455850233	2016-02-19	02:50:33	4264	6733	0	1	0
3	1455850305	2016-02-19	02:51:45	4264	6733	0	1	0
9	1455850340	2016-02-19	02:52:20	4264	6733	0	1	0
0	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
1	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
2	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
3	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
4	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
5	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
6	1455850344	2016-02-19	02:52:24	4264	6733	0	1	0
7	1455859506	2016-02-19	05:25:06	6491	6733	0	1	0
3	1455863324	2016-02-19	06:28:44	2993	6733	0	1	0

- here is the transactions that the node is flipping its account and accepting the high values

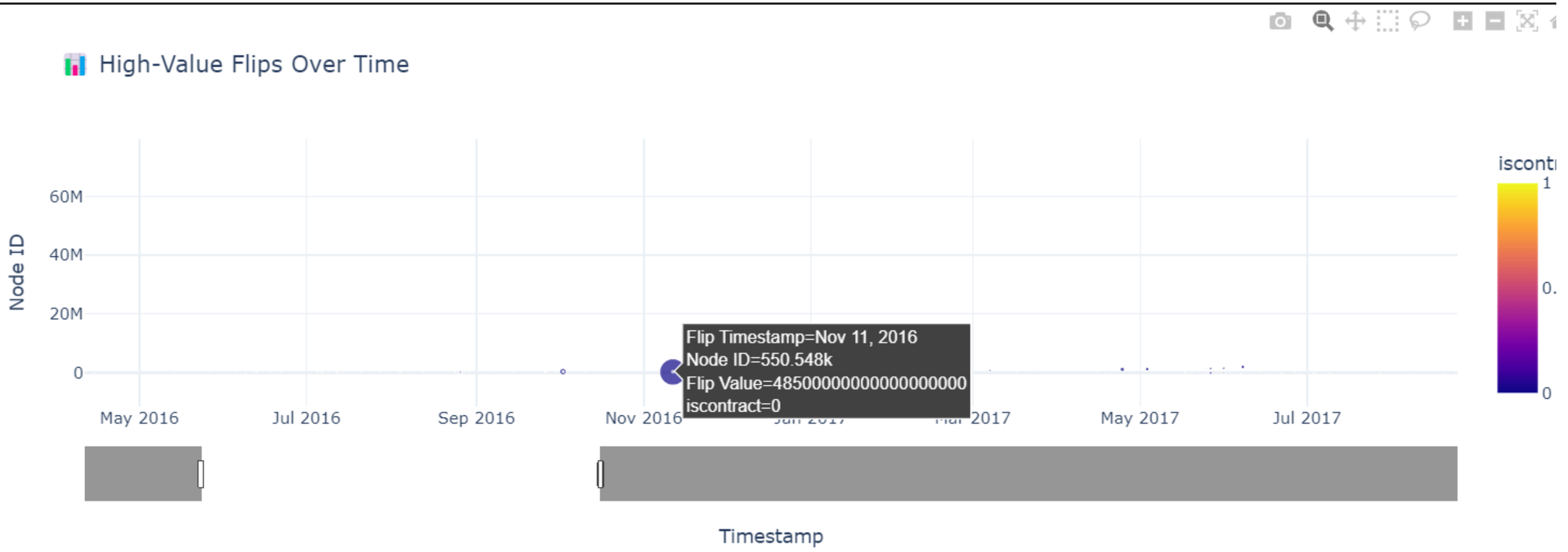


1507160965	2017-10-04	23:49:25	1138346	6733	0	1	0
1507161009	2017-10-04	23:50:09	5526324	6733	0	1	0
1507161212	2017-10-04	23:53:32	7000364	6733	0	1	0
1507161212	2017-10-04	23:53:32	5526324	6733	0	1	0
1507161351	2017-10-04	23:55:51	1138346	6733	0	1	0
1507161351	2017-10-04	23:55:51	7004236	6733	0	0	287000000000000000
1507161629	2017-10-05	00:00:29	5526324	6733	0	1	0
1507161667	2017-10-05	00:01:07	1138346	6733	0	1	0
1507161726	2017-10-05	00:02:06	7004236	6733	0	0	275000000000000000
1507161842	2017-10-05	00:04:02	5526324	6733	0	1	0
1507162057	2017-10-05	00:07:37	7003065	6733	0	0	300000000000000000
1507162262	2017-10-05	00:11:02	1138346	6733	0	1	0
1507162602	2017-10-05	00:16:42	6636194	6733	0	1	0
1507162867	2017-10-05	00:21:07	1138346	6733	0	1	0

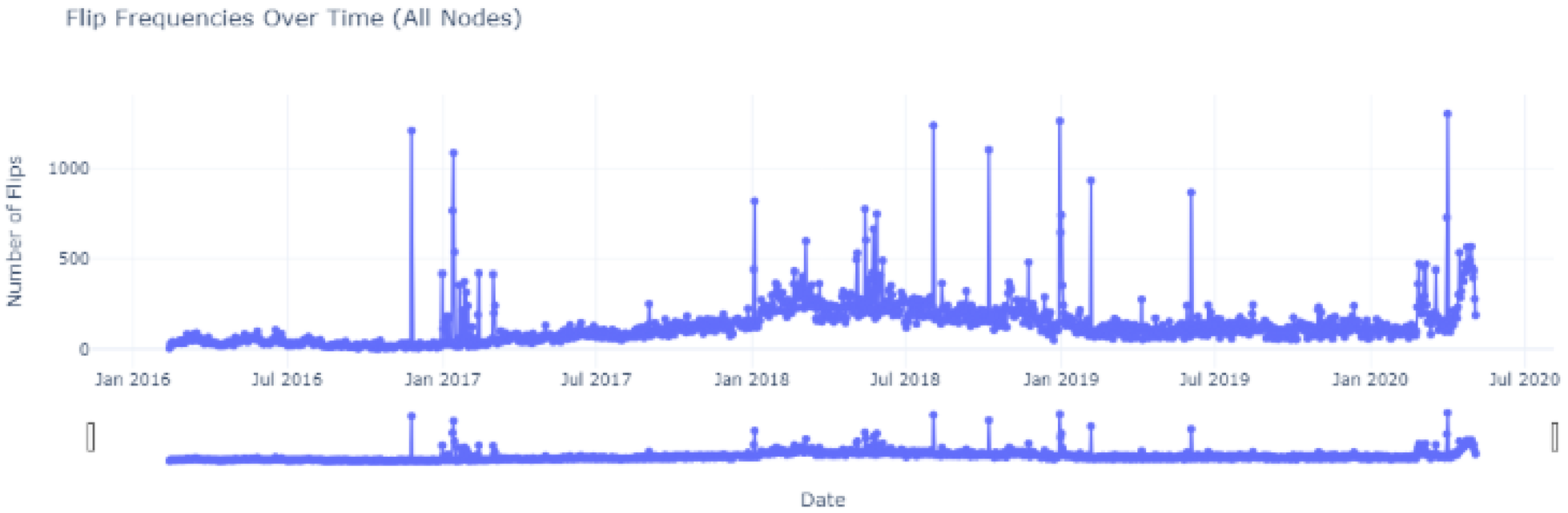
- counted how many times a node flipped in a day
- total flipped node transaction count vs time is plotted
- burst time period is taken for identifying how many times a node flipped in that interval and how many transactions it is doing

Results

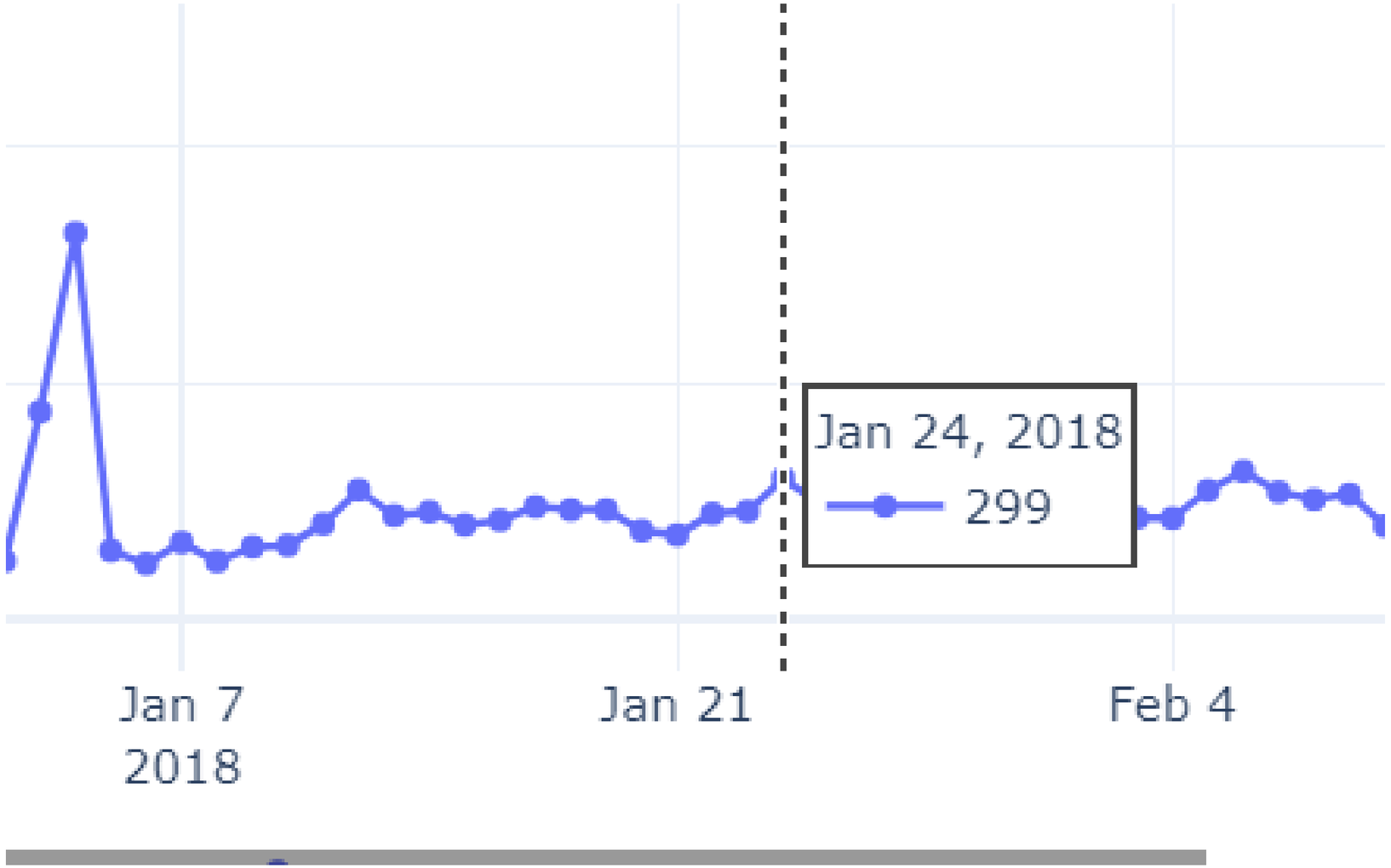
I have checked whether any high value transfer is done through it or not and extracted and also noted how many times a particular node has flipped.


High-Value Flips Over Time
Node ID
60M
40M
20M
0
May 2016
Jul 2016
Sep 2016
Nov 2016
Mar 2017
May 2017
Jul 2017
Flip Timestamp=Nov 11, 2016
Node ID=550.548k
Flip Value=48500000000000000000
iscontract=0
isconti
1
0.
0
Timestamp

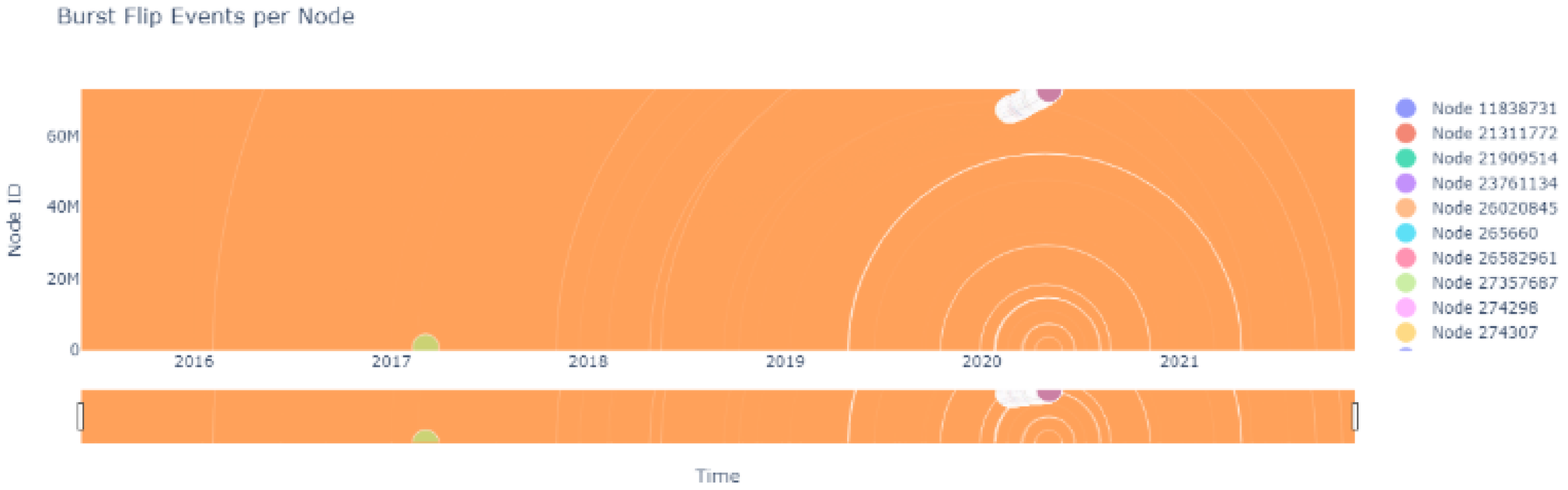
Now i have extracted the number of flips happened overall in a day and plot the graph vs day



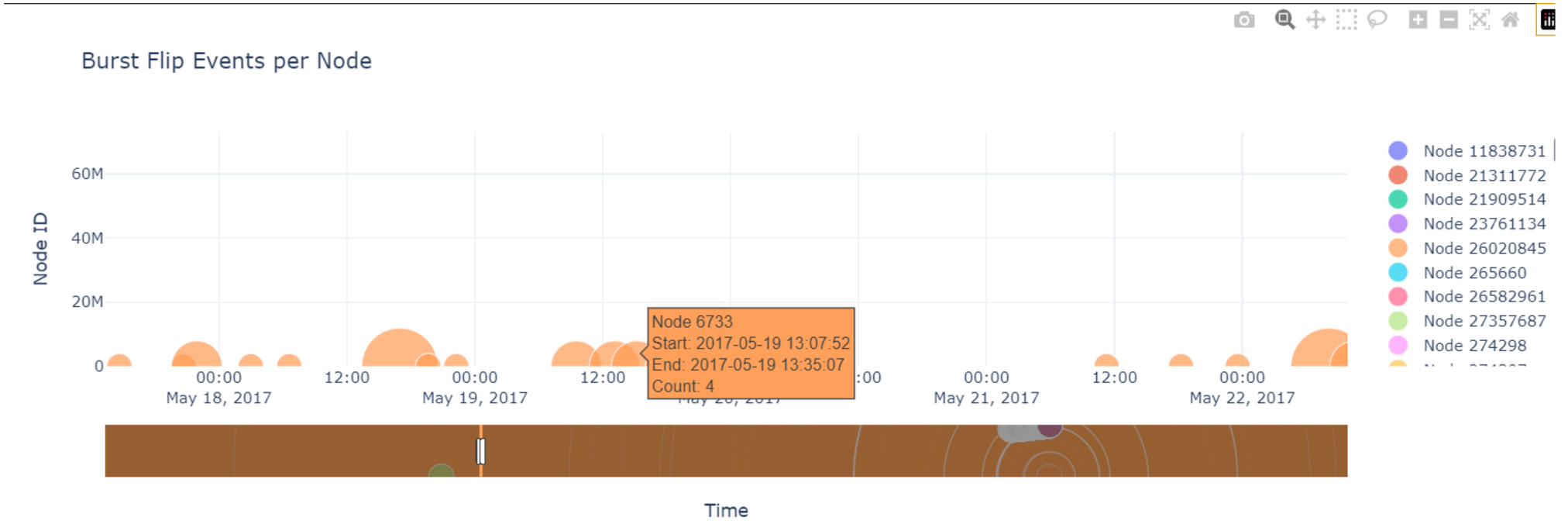
here is the zoomed version of the small part from the above graph



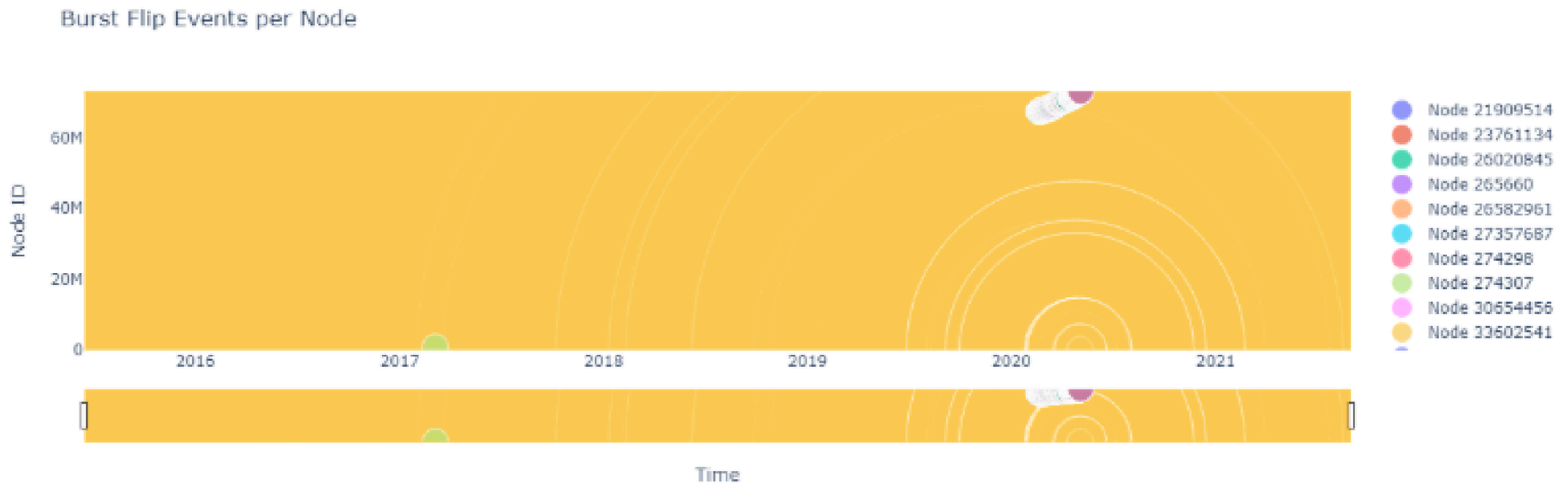
here is the plot for burst periods vs date with threshold for 1hr



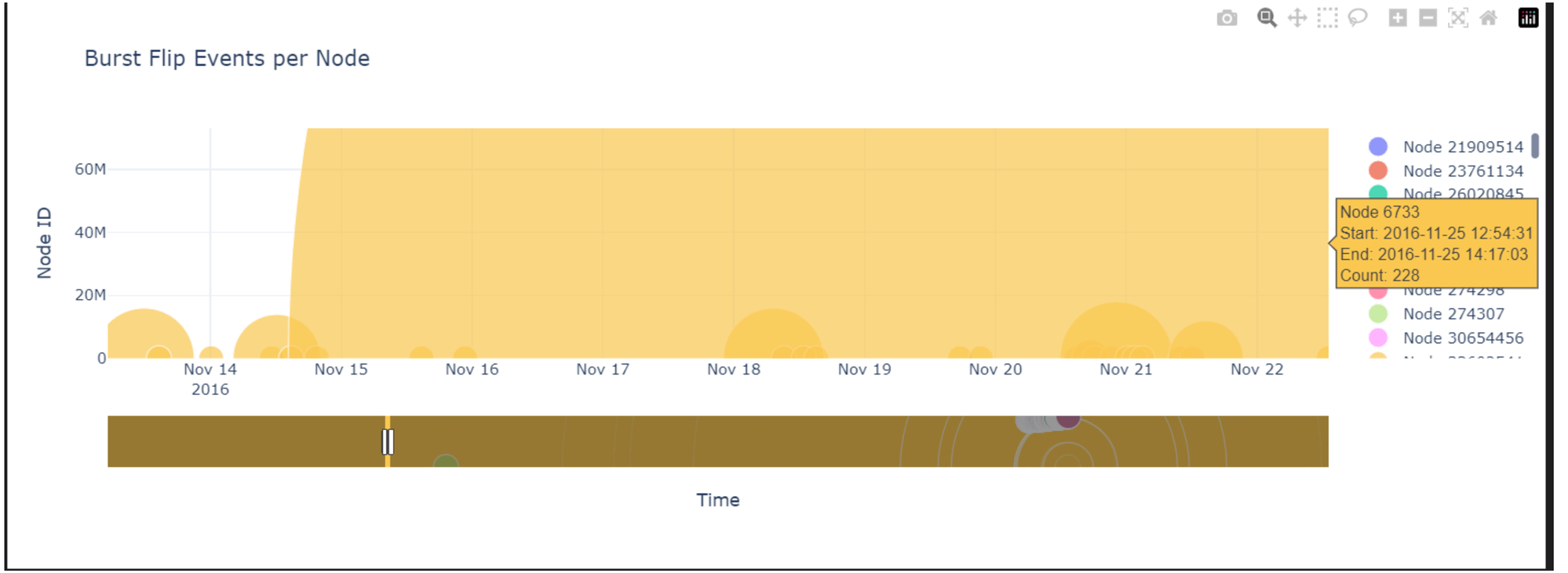
here is the zoomed part of the plot for burst periods vs date



here is the plot for burst periods vs date with threshold for 30mins



here is the zoomed part of the plot for burst periods vs date



Challenges faced

More data columns are required, which made me do the preprocessing from earlier stages

No proper resource papers for the project, which can guide to identify anomalies or to get thresholds

Large data which needs a lot of computational power and time for processing

Understanding and identifing the patterns and no proper resource for dual nature of nodes

Conclusion

- Identifing potentially malcious behaviours and changing the network, so that they wont occur or will be reduction in these kind of anomalies
- Increases the trust and security of the network
- Helps in identification of threats
- Development of more robust fraud prevention mechanisms

Architecture

- Taking the user as nodes, assumed a graphical structure of the ethereum network. Extracted in-degree,out-degree,weighted in degree,weighted out degree for nodes under specific conditions.
- Also extracted nodes that showed dual nature in transactions
- Assigned scores for each abnormal traits present
- According to score, named them as anomaly and have show or marked them in various plots.

I approached in this way to my problem because these transactions been happened can be better visualised in graphical format and the degrees of those nodes/users can speak about how they are interacting between them which help to study abnormal patterns.

Thank you very much!