STANFORD UNIVERSITY
Stanford Center for Professional Development

HOMEWORK ROUTE FORM

Stanford Center for Professional Development Student Information

Course No.	CS224W	Faculty / Instructor Name	Jure Leskovec	Date	12/10/2013
Student Name	Anderson Aiziro, Lucas Gomes Silveira			Phone	4252475635,4257853813
Company	Microsoft			Email	aaiziro@stanford.edu, lgomes09@gmail.com
City.	Bellevue	State	WA	Country	USA

Check One: ○ Homework #: ○ Mdterm ◉ Other Final Project Report (Group #59)

The email address provided on this form will be used to return homework, exams, and other documents and correspondence that require routing.

Total number of pages faxed including cover sheet 13

	For Stanford Use Only	
Date Received by the Stanford Center for Professional Development	Date Instructor returned graded project Score/Grade: (to be completed by instructor or by teaching assistant)	Date the Stanford Center for Professional Development returned graded project:

Please attach this route form to ALL MATERIALS and submit ALL to:

Stanford Center for Professional Development
496 Lomita Mall, Durand Building, Rm 410, Stanford, CA 94305-4036
Office 650.725.3015 | **Fax** 650.736.1266 or 650.725.4138
For homework confirmation, email scpd-distribution@lists.stanford.edu
http://scpd.stanford.edu

Last modified October 27, 2008

CS224w:
Social and Information Network Analysis

Assignment number: Final Project Report (Group #59)

Submission time: 18:30 **and date:** 12/10/2013

Fill in and include this cover sheet with each of your assignments. It is an honor code violation to write down the wrong time. Assignments are due at 9:30 am, either handed in at the beginning of class or left in the submission box on the 1^{st} floor of the Gates building, near the east entrance.
Each student will have a total of *two* free late periods. *One late period expires at the start of each class.* (Homeworks are usually due on Thursdays, which means the first late periods expires on the following Tuesday at 9:30am.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than *one* late period after its due date.

Your name: Anderson Aiziro, Lucas Gomes Silveira
Email: aaiziro@stanford.edu, lgomes09@gmail.com **SUID:** aaiziro, lucasgs

Collaborators:

I acknowledge and accept the Honor Code.

(Signed) Anderson Aiziro, Lucas Gomes Silveira

(For CS224w staff only)

Late periods: 1 2

Section	Score
1	
2	
3	
4	
5	
6	
Total	

Comments:

Where is the money? Modeling Bitcoin users

Lucas Gomes Silveira
lucasgs

Anderson Aiziro
aaiziro

ABSTRACT

Bitcoin is a decentralized digital currency introduced in 2008 that enables instant payments to anyone, anywhere in the world. Its increasing popularity has raised many questions on who are the agents in the "bitcoin economy" and what is the nature of the transactions between them. In this paper we use the public bitcoin records to shed some light on these questions.

We categorize users by looking at network features and applying heuristics and clustering algorithms. Then we analyze the flow in terms of transaction quantity and volume over time, identifying what are the most common economic patterns and how they are evolving while bitcoin gets more popular.

Keywords: bitcoin, currency exchange, network analysis, categorization, flow analysis, clustering

1. INTRODUCTION

Bitcoin is a decentralized digital currency that enable near instant payments between parties anywhere in the world. This network was described initially in a paper by Satoshi Nakamoto, and released the MIT license. This peer to peer technology operates without a central authority, using bitcoins (BTCs) as form of payments, and managing transactions collectively across the network.

As of December 2013, bitcoin reached a market capitalization of US$ 10 billion (6500% growth compared to previous year), with 71000 transactions per day (95% YoY growth), and 12 million bitcoins in the network (20% YoY growth). Moreover, political and financial institutions are starting to look into joining and/or regulating the bitcoin network. Examples are: Bitinstant offers to tie user's Bitcoin wallets to Mastercard accounts [9]; Bitcoin Central's partnership with with bank Credit Mutuel Arkea [10]; FinCEN's regulations on virtual currencies [8].

Bitcoin is designed around the idea of using cryptography to control the creation and transfer of money. A pair of public and private keys acts as a digital wallet. Using cryptographic signing all the transactions are validated by a proof-of-work system, and included in the block chain that enforces no double spending.

Even though the transaction data are publicly available, most bitcoin owners are anonymous and operate with multiple public addresses. This has limited the knowledge about bitcoin usage patterns and its evolution.

This work intends to better understand the bitcoin network users and their activity by exploring questions like: "what are the common user behavioral archetypes and how to identify them?", "how does the money flow between these groups?" and "how did the network flow evolve over time?"

Here's how this paper is structured. In Section 3 we describe the dataset utilized for our analysis, and the information we can extract from it. In Section 4 we describe the categories of bitcoin users we will model. In Section 5 we define the features from the bitcoin users that will be used for categorization. In Section 6 we describe the methods and algorithms used for modeling and clustering the users. In Section 7 we apply the clustering results to the bitcoin network, and identify properties of the transactions between different user types, providing us data to answer our initial questions regarding the network. In Section 8 we present a conclusion of the main findings.

2. RELATED WORK

Reid and Harrigan [1] analyze of the bitcoin public block chain. With the goal of showing anonymity is

not guaranteed in the bitcoin network, the article uses different techniques to analyses behavior of user nodes and cross relate personal identifiable data. In this work, we will use the same dataset and a generalized version of the behavior analysis to classify nodes.

Ron and Shamir [2] made a quantitative analysis of the bitcoin network based on data from the inception up to 05-12-12. They noticed most of the bitcoins are not circulating. (78% of all created bitcoin at that time were in addresses that never made a transaction. One year later, has this picture changed? We're set to find out.

For node categorization, Pirolli et all [3] applied a linear model on multiple web page features, intrinsic (text similarity) and network (node degrees), and found rules to classify web pages with above random precision. We intend to use a similar approach to create a classification heuristic from the bitcoin user features.

Also for node categorization, we will use the k-means clustering method [12] to partition the nodes into k clusters that we identified from the heuristic. For solving this NP-hard clustering problem, we will use Hartigan and Wong (1979) algorithm [11], which iterates on selecting cluster centers and partitioning the observations in k clusters such that the distance of the observations to their assigned cluster center is a minimum.

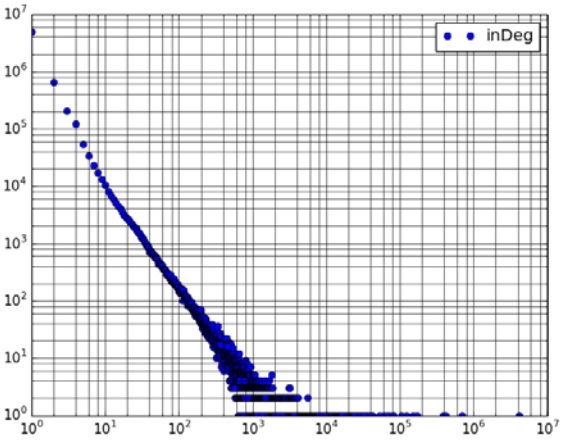
Kleinberg [4] renowned HITS algorithm attributes hub and authority scores to each node. In the web context, a high hub score indicates the page points to many good references about a topic and a high authority score indicates the page contains authoritative content about a topic. Since the algorithm is mainly based on the edges between nodes, we can apply it to the bitcoin network to obtain another feature. Peserico and L. Pretto [5] showed that the convergence of HITS algorithm in RANK can be exponential, which may lead to a long running time to calculate hub and authority scores for the bitcoin network.

3. DATASET DESCRIPTION

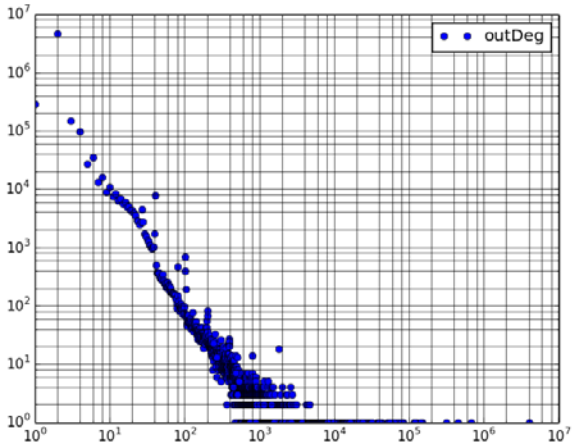
Our dataset is the bitcoin transaction network provided by Ivan Brugere [7] and based on the work of Heid and Harring [1]. It contains all transactions between bitcoin addresses until April 4th, 2013 and grouped public addresses participating in the same outbound side of transaction (as it indicates the same entity controls all these addresses). Some entities will still be represented by multiple groups if they have addresses not linked through transactions.

From this dataset, we generated a network where users are nodes (different public address mapping to the same address are represented by a single node), and transactions represent direct edges between the nodes. 11.1 million self-edges were filtered out as they don't represent monetary flow.

This way, we obtained a power law network with 6336769 nodes (bitcoin users) and 28143065 edges (transactions between users). The network diameter (sampled from 20 nodes) was 1936 and the degree-distributions are shown in pictures 1 and 2.



Picture 1: In-degree log-log distribution for bitcoin network dataset used in this paper



Picture 2: In-degree log-log distribution for bitcoin network dataset used in this paper

4. NODE CATEGORIES

In this section, we present the heuristics used for categorizing the users of bitcoin networks. The categorization will be used for our final goal of studying the usage of the bitcoins in the network. For defining the features we need to extract from the nodes, we need to understand the characteristics of each type of node we want to classify.

To treat this network as an economy with focus on spending and trading behaviors, we will classify the node as:

- Exchanges – trade real currencies and bitcoins .
- Miners – receive newly created bitcoins by creating valid transaction blocks.
- Investors – accumulate bitcoins as a store of value or speculation
- Buyers - send bitcoins in exchange to many nodes for products and services
- Sellers - receive bitcoins from many nodes in exchange for products and services
- Neutral – receive and send bitcoins in the same rate, effectively using bitcoin as a payment medium.

Note that these categories are behaviors that may manifest simultaneously on a node. It's possible for one individual to mine, buy, sell and invest bitcoin with the same address. Therefore we are looking for predominant behaviors.

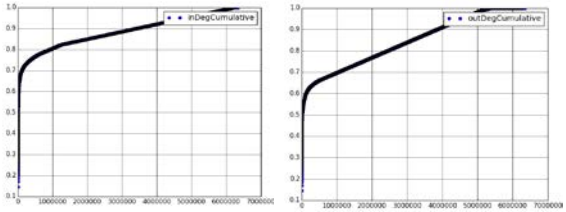
5. NODE FEATURES

With the task of categorizing nodes within those definitions in mind, we consider what would be useful to identify different user groups: how many did he buy and how many he sold, how many times did he buy and how many times did he sell, what is its relation to other users, and number of public addresses held.

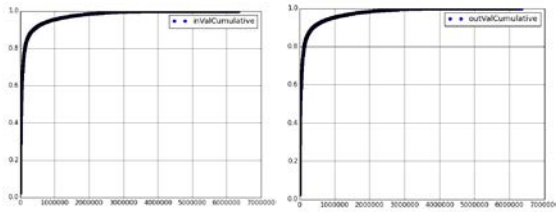
To measure these for each node, we chose to track:

- InDeg/OutDeg: in degree and out degree
- InVal: sum of all bitcoins received
- OutVal: sum of all bitcoins sent
- HubScore: hub result from Kleibergs' HITS
- AuthScore: auth result from Kleibergs' HITS
- NAddress: number of associated addresses

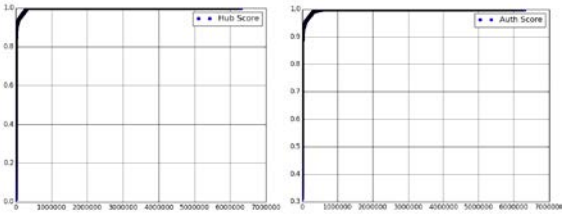
Pictures 3 to 5 show the cumulative distributions for each of these features. All of them show a big concentration of value in the top nodes.



Picture 3: Cumulative distribution of node in and out-degrees (x axis in degree descending order)



Picture 4: Cumulative distribution of node in-values and out-values (x axis in value descending order)



Picture 5: Cumulative distribution of hub and auth scores (x axis in HITS score descending order)

6. NODE CATEGORIZATION

6.1 HEURISTICS

We start by looking at some nodes in the dataset with known behavior (discovered through tagged addresses and transaction history in blockchain.info). Table 1 shows the features collected for known nodes of different categories. Considering the feature patterns in each known node and feature profiles discussed in section 5, we came up with heuristic 1 to categorize the all nodes fitting with the known nodes.

IF (HubRank,AuthRank < 100000 AND NAddress> 1000) THEN EXCHANGE

ELSE IF (OutVal > InVal) THEN MINER

ELSE IF (Inval – OutVal > 10) THEN INVESTOR

ELSE IF (OutDeg >= 1.1* InDeg) THEN BUYER

ELSE IF (InDeg > 1.5 * OutDeg AND InVal > 50) THEN SELLER

ELSE NEUTRAL

Heuristic 1: Linear node classification

One important observation: since our dataset does not include balance and bitcoins issued from mining, the heuristic to classify miners and Investors will not find all nodes with these behaviors.

Table 2 shows how heuristic 1 divided our dataset, fitting with the known nodes annotated categories. As expected in a monetary system, most nodes have predominant buyer behavior, some have seller behavior and very few act as currency exchanges (and other highly transactional roles, like bitcoin game operator.

Bitcoin Node Hash Address	Category	Address count	InVal Value	OutVal Value	InDeg Value	OutDeg Value	Hub Rank	Auth Rank
Mt Gox 1LNWw6yCxkUmkhArb2Nf2MPw6vG7u5WG7q	Exchange	318211	3MM	3MM	689467	444621	408	139
Wikileaks 1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v	Seller (accepts donations)	108	4459	3466	1694	112	1159852	23950
Bitcoin 100 1BTC1oo1J3MEt5SFj74ZBcF2Mk97Aah4ac	Seller (accepts donations)	3	1206	924	255	27	321810	93155
mystery2048 1HWMQv2VYviAgpy6NWNvVg9JhKm4zcMGS5	Buyer (donated bitcoins)	5	22.6590	22.500774	31	116	348714	93803
Eligius Donation 1E1igiusfEjs1pCaGjEERExE9gYcrFwow7	Seller (accepts donations)	522	64275	36175	21371	43586	28801	16653
Unknown 1F92kQ4G2nnLJ1W3B4YTNHqFkxNfF95gnC	Miner (all income from mining)	2	0.0125	50.013	1	2	3544744	4025906
Unknown 1CTgYxMTY5j6SLytKeMsBWAXuUc6yNKcAe	Investor (accumulated 2500 BTC)	10	3312	1122	1024	57	37992	28468
Eligius miner 1EXfBqvLTyFbL6Dr5CG1fjxNKEPSezg7yF	Mixed (mines and does other transactions)	179	1000	1828.81	515	740	12378	11091
SatoshiDice 1dice97ECuByXAvqXpaYzSaQuPVvrtmz6	Online Gaming	1090	3.8MM	3.78MM	4094509	4075472	5	1
Pinball coin 1GRWQA3wLGL9i5aW371Nd1HnF4QaBNEGxG	Online Gaming	4	48.19	42.72	1382	1226	318329	85681

Table 1: Features and categories of known bitcoin accounts

	Count	AvgInVal	AvgOutVal	AvgInDeg	AvgOutDeg	AvgHubRank	AvgAuthRank	AvgNAddress
Exchange	381.00	151099.79	151388.10	21769.76	17309.83	5186.65	1830.02	5776.88
Buyer	4239996.00	268.27	268.26	1.11	3.19	2806654.97	3302219.24	1.04
Seller	91041.00	636.92	636.81	28.15	6.13	2624053.67	1500432.50	8.46
Neutral	1878503.00	21.50	21.32	5.54	2.52	3817057.44	2718979.38	2.17
Miner	74811.00	105.01	223.69	4.32	13.01	3728342.43	5694903.08	2.77
Investor	45618.00	451.64	262.17	40.39	38.21	4846415.25	2982905.17	4.93

Table 2: Features and categories of known bitcoin accounts

6.2. CLUSTERING

The reasonable categorization through heuristics indicates that we have a good feature set for clustering. In this section we explore if we can achieve a similar grouping without using intuitive knowledge of node behaviors.

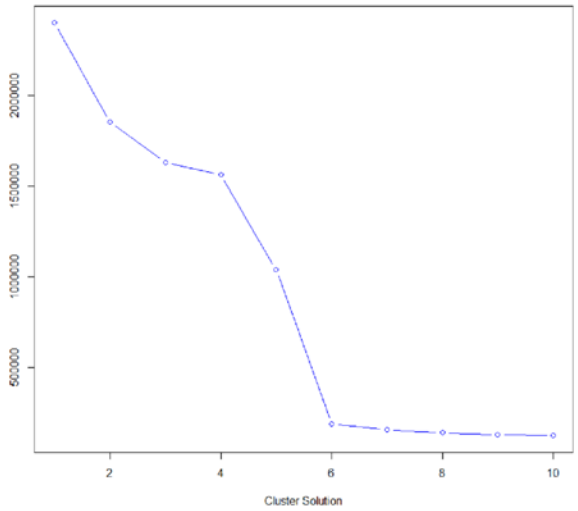
We use the k-means clustering method, with Hartigan and Wong algorithm [11]. Given the set of nodes and k cluster sets $(S_1, S_2, \dots, S_k)$, this method aims to partition the nodes so as to minimize the within-clusters sum of squares of errors (SSE). This can be described by the formula:

$$\arg\min_s \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

Where μ_i is the mean of points in S_i. In our clustering, the points x_j in S_i, will be described by: $(f_1, f_2, \dots f_n)$, where f_i's represents the features chosen in Section 5.

The clustering algorithm has two steps: in the first step, we will find the optimal number of clusters for minimizing the SSE. In the second step, we will split the nodes into clusters.

For the first step, let's consider all possible features: InVal, OutVal, InDeg, OutDeg, Hub, Authority, and NAddress. Iterating over the SSE while changing the number of clusters. Picture 6 indicates 6 clusters is the best choice for minimizing error with less clusters. This result is aligned with the number of clusters suggested by our heuristics.



Picture 6: SSE calculated against number of clusters for Hartigan and Wong k-means clustering algorithm.

We initially assumed that all features together would provide the 6 clustering solution with great differentiation between the clusters. However, this initial approach didn't work as expected (as explained below), and we had to take a different approach. The process to identify the cluster works as follow:

1) Use Hubs and Authorities feature to identify the exchanges
2) Remove Exchanges from working Set. Remove Hubs and Authorities from Feature Set.
3) Use remaining features to identify miners and investors.
4) Remove miners and investors from working Set.

5) Create features that can capture the behavior of Buyers, Sellers, Neutral, as specified in Section 4
6) Create clusters for buyers, sellers and neutral

Analyzing the clustering feature importance of all the features together (table 3), we noticed that: Hubs and Authorities ranks are good for identifying exchanges, but hide the importance of the other features.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
kclust	0	0	0	0	0.07	0
InVal	0.01	0.01	0	0	13.38	0.01
OutVal	0.01	0.01	0	0	13.32	0.01
InDeg	0	0	0	0	18.54	0.01
OutDeg	0	0	0	0	17.7	0.01
hubRank	19.34	36.98	82.23	62.86	15.65	48.97
authRank	80.65	63	17.76	37.13	15.1	51
numKeys	0	0	0	0	6.25	0

Table 3: Feature importance on clustering solution

Using hubs and Authorities features, we obtain our first cluster representing the exchanges of the network, with 405 nodes.

Removing the exchanges from the working set, and removing Hubs and Authorities from the feature set, we repeat the process, and we are able to identify other 2 clusters: miners and investors (table 4).

ClusterId	AvgInVal	AvgOutVal	AvgInDeg	AvgOutDeg
1	3.027371194	3.317458755	14.75128534	18.62862977
2	103.8421271	6.121424254	3.758531223	0.348724714
3	446.8712944	445.7536683	12.51791382	6.419051713
4	1.477668655	1.22900381	27.37753365	1.377308887
5	48.80431428	49.57008929	38.18692074	39.34382923
6	10.94600964	316.184843	0.700188596	9.133237187

Table 4: Average feature values for each cluster – identification of miners and investors

Looking at the average feature values, and based on the heuristics defined in section 6.1, it is clear that miners is represented by cluster 6 (average InVal << average OutVal), while investors are represented by cluster 2 (average InVal >> average OutVal).

Cluster 5 has an AvgInVal < AvgOutVal, but since they are close, we will not classify these nodes yet. This means there are some miners that still need to be found.

Also, from this result it is not clear how to differentiate buyers, sellers and neutral. This happens due to the large span of values of the features InVal, OutVal, InDeg, and OutDeg: Big buyers and small buyers will never be in the same cluster using the SSE clustering method.

To solve this issue, let's define two new features, which can capture the behavior of buyers, sellers, and neutral users:

- $ValueDegRatio: \begin{cases} 1 \; if \; \frac{inVal}{inDeg} > \frac{outVal}{outDeg} \\ 0 \; if \; \frac{inVal}{inDeg} = \frac{outVal}{outDeg} \\ -1 \; if \; \frac{inVal}{inDeg} < \frac{outVal}{outDeg} \end{cases}$
- $DegDiff: \begin{cases} -1 \; if \; inDeg > outDeg \\ 0 \; if \; inDeg = outDeg \\ 1 \; if \; inDeg < outDeg \end{cases}$

Using these features based on the heuristics defined in section 6.1, we expect to find:

- Buyers: $DegDiff \cong 1, ValueDegRatio \cong 1$
- Sellers: $DegDiff \cong -1, ValueDegRatio \cong -1$
- Neutral: $DegDiff \cong 0$

The ValueDegRatio helps us capture the concept that buyers can't spend more than they earn. This will aid in the identification of the missing miners from the previous run of the algorithm.

Removing the miners and investors from the working set, and using the $DegDiff$ and $ValueDegRatio$ features defined, we repeat the process to obtain the following result:

ClusterId	Avg ValueDegRatio	Avg DegDiff	Final Classification
1	1	-1	seller
2	0	-0.175944079	neutral
3	-1	-1	seller
4	0.999984516	1	buyer
5	-1	0.70970971	miner
6	1	0	neutral

Table 5: Average feature values for each cluster – identification of buyer, seller and neutral

From this result, we were able to clearly classify the remaining nodes. Aggregating the results from the clusters, we have the following average values for the features:

Cluster	Count	AvgInVal	AvgOutVal	AvgInDeg
Exchange	405.000	133731.108	133677.516	19882.370
Buyer	4211441.000	271.466	271.494	1.327
Seller	907812.000	97.590	96.756	13.631
Neutral	1062141.000	26.555	26.502	1.738
Miner	144252.000	2.805	60.196	0.464
Investor	81561.000	94.998	2.265	2.663

Cluster	AvgOutDeg	AvgHubRank	AvgAuthRank	AvgNAddress
Exchange	16396.802	726.630	467.138	5170.630
Buyer	3.792	2830096.106	3329770.101	1.080
Seller	3.793	4195063.533	2169230.572	3.916
Neutral	1.435	3402928.502	3163851.745	1.358
Miner	3.823	1914681.410	2932993.381	1.115
Investor	0.161	5482807.812	3542759.429	1.024

Table 6: Average feature values for each cluster on the entire network

From the clustering classification in table 6, 6 types of users in the network we clearly identified, with characteristics similar to the groups obtained through heuristics:

Exchanges: a small portion of the network that participates in a large number of transactions (average InDeg of 19882.370 and average OutDeg of 16396.802)

Miner: users that produce bitcoins. These users don't buy many bitcoins (average InVal of 2.805 and InDeg 0.464), but they sell a lot of bitcoins (average OutVal of 60.196 and average OutDeg of 3.803)

Investor: users that clearly want to accumulate bitcoins. Their average InValue is 94.998 with an average InDeg of 2.265, while the average OutValue is 2.265 with an average OutDeg of 0.161.

Buyer: the majority of the network users. They have an expected value of InDeg < OutDeg. However, they surprisingly have an average InVal and OutVal that are larger than the Sellers.

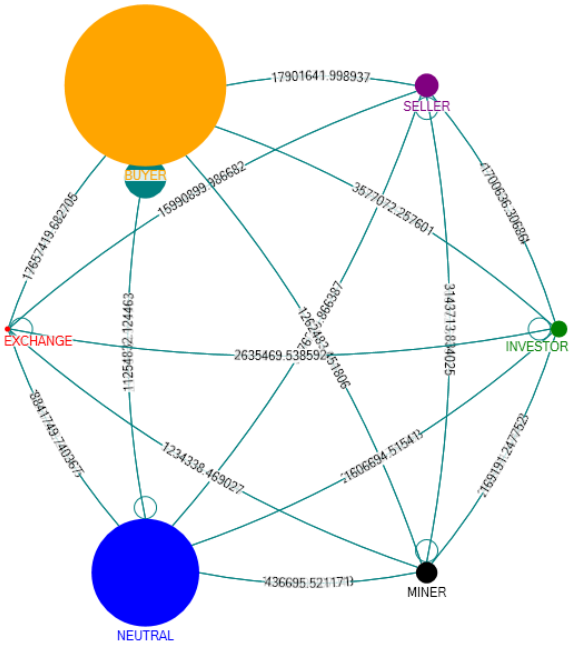
Seller: a fair amount of users. They present the expected behavior of InDeg >> OutDeg.

Neutral: these users can act sometimes as buyers and sometimes as sellers. Their participation in the overall flow of the network is unknown. There is no expressive feature that will define them, since InDeg ≅ OutDeg, and InVal ≅ OutVal.

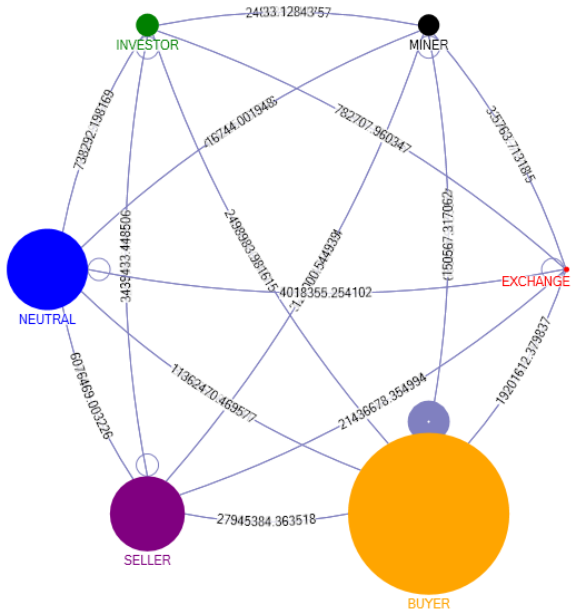
7. FLOW ANALYSIS

In this section we explore the behavior of bitcoin actors and the economic dynamics of the network by analyzing the flow of bitcoins between the different groups identified in section 6.

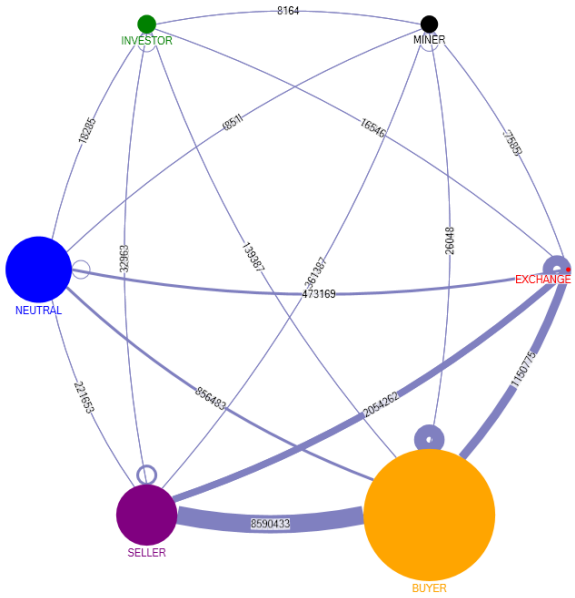
In pictures 7 to 10 generated with NodeXL [13], we visualize the value flow and transaction count between user groups. It's noteworthy that groups obtained through heuristics and clustering showed similar patterns. We can see that buyer-buyer transactions concentrate the majority of value, dwarfing all other kinds of transactions. In terms of transaction counts, buyers continue to be the main cluster, but buyer-neutral, buyer-seller and buyer-exchanges transactions counts are also considerable.



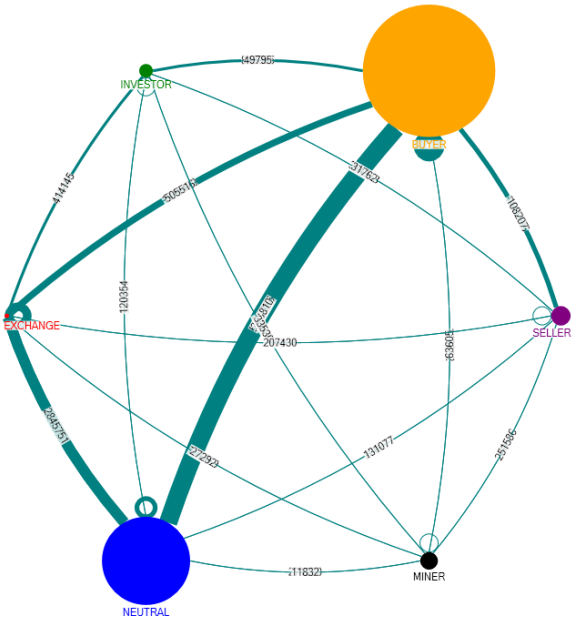
Picture 7: Value flow between node categories obtained by heuristics



Picture 8: Value flow between node categories obtained by clustering



Picture 10: Transaction count flow between node categories obtained by clustering



Picture 9: Transaction count flow between node categories obtained by heuristics

Analyzing the different clusters separately, we will identify properties from each cluster that will help us model the bitcoin network.

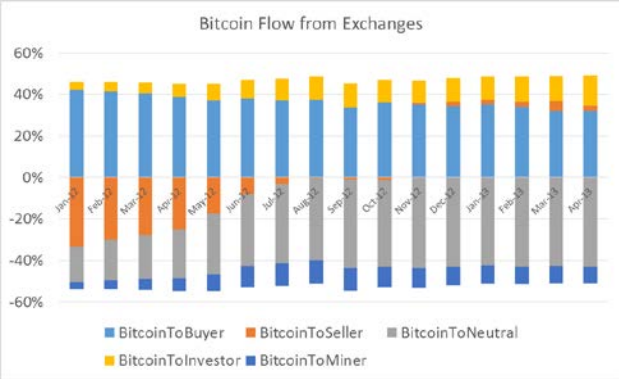
To aid in the analysis of their behavior, we have ordered the bitcoin transactions by date, and taken monthly snapshots of the feature values from Jan 2012 to Apr. 2013.

Interesting properties were found while analyzing the features evolution over time are such as: change of bitcoins volume transitioning between clusters, direction of bitcoins flow between clusters, and the increase/decrease of bitcoins accumulated.

Exchanges:

- Do not accumulate bitcoins. It is used as a medium to transfer bitcoins between different clusters.
- Receives bitcoins from Miner and Neutral Users. Sends bitcoins mainly to buyers
- One interesting fact is that over the last year, the exchanges stopped receiving

bitcoins from Sellers, and started receiving it mostly from Neutral Users.



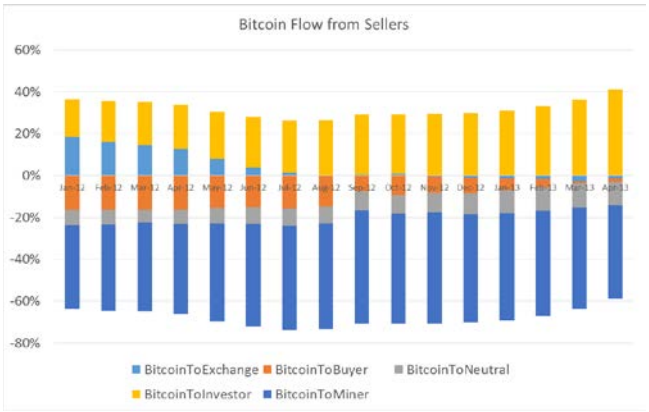
Picture 11: Bitcoin Flow evolution from Exchanges to other clusters (Jan 2012 – Apr 2013)

Buyers:

- The number of transactions from Buyers to Sellers is the largest, as we would expect
- The volume of transactions from Buyers to Buyers is very large (around 80% of the entire volume of transactions). Also, the average transactional value between buyers (287.59 bitcoins) is extremely high compared to average transactional value between buyers and sellers (3.25 bitcoins).
- The observation above is highly likely to happen due to an incomplete aggregation of public keys to the same user. Buyers transitioning bitcoins between accounts will count as transactions from different users.
- Another option would be an error in the clustering solution. However, considering the valuation of bitcoins in USD in April 2013 ($95 USD), an average transaction value of 287.59 bitcoins is not realistic.

Sellers:

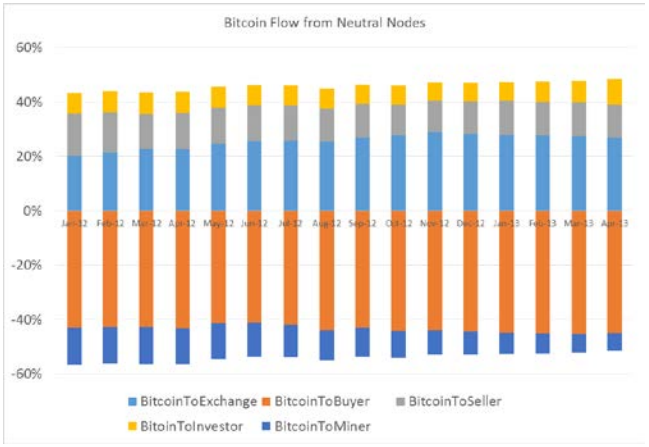
- Receive bitcoins mainly from miners. Also receive bitcoins from Buyers.
- Sends bitcoins mostly to buyers and exchanges, at a similar rate that it receives bitcoins from these entities.
- In the year evolution, we can notice a pattern change, where the flow of bitcoins from buyers to sellers is not existent anymore
- It is also a source of bitcoins for investors, with a perceived flow of bitcoins from sellers to investors



Picture 12: Bitcoin Flow evolution from Sellers to other clusters (Jan 2012 – Apr 2013)

Neutral Users:

- Receives bitcoins from buyers and miners, and send them to sellers, exchanges, and investors
- Does not accumulate bitcoins. It is used as a medium to transfer bitcoins between other clusters: mainly between buyers to sellers and buyers to exchanges

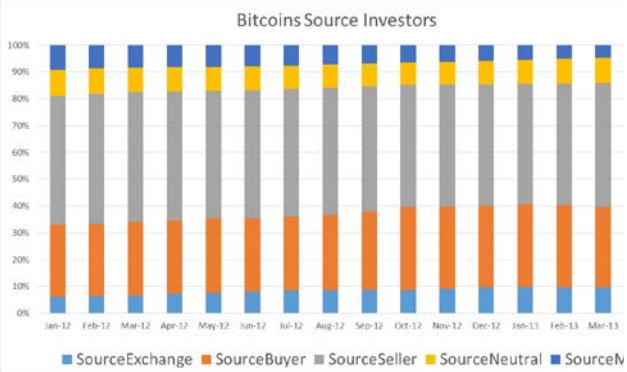


Picture 13: Bitcoin Flow evolution from Neutral Nodes to other clusters (Jan 2012 – Apr 2013)

Investors:

- As the popularity of the bitcoin network increases, more users start seeing value in retaining bitcoins. This trend can be observed by increase in In-volume and In-transactions from the graphics below.
- Investors are obtaining their bitcoins mainly from buyers and sellers. This is an interesting fact of the bitcoin network. Following a market behavior, investors were expected to have their main source of bitcoins from exchanges.

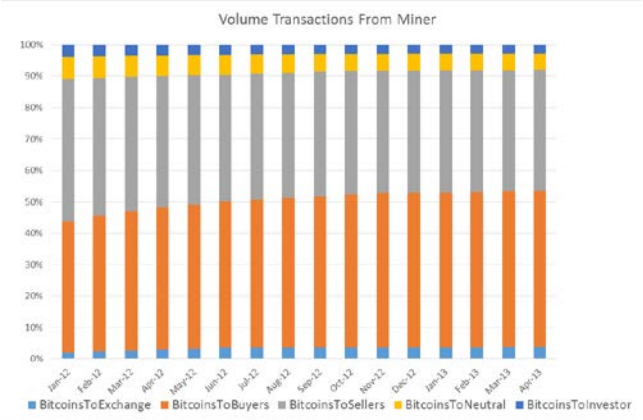
Picture 14: Bitcoin volume and number of transaction evolution (Jan 2012 – Apr 2013)



Picture 15: Origin of the investors bitcoins evolution (Jan 2012 – Apr 2013)

Miners:

- Creators of bitcoins for the network
- The main consumers of the bitcoins generated by miners are the buyers and sellers. This is a very interesting behavior of the bitcoin network.
- This indicates that the miners also present a buyer behavior (interacting with sellers)
- This indicates that the money is being transferred from miners to buyers



Picture 16: Destination of the miners bitcoins evolution (Jan 2012 – Apr 2013)

8. CONCLUSION

We modeled the bitcoin users in meaningful archetypes (buyer, seller, exchange, miner, investor, neutral), and proposed heuristics and a clustering method to classify anonymous nodes into these types using only public network features. In particular, hubs and authorities ranks were a very strong classifier. We had too few known nodes of each type to rigorously validate how good the categorization was, but the similarity of the groups obtained through intuitive heuristics and clustering is a positive signal.

Analyzing the transaction flow between groups of users and its evolution over time, we found that the investor group is increasing its bitcoin balance threefold, a much greater rate than new bitcoins are mined. In the Jan-2012, investor balances represented 25% of bitcoins in circulation where as in Apr-13 the percentage jumped to almost 50%.

There were also a few unexpected, but very interesting findings that provided us insights regarding improvements on the original dataset. The buyers concentrate more than 80% of the transaction values between themselves, with seller and exchange activity being much smaller than we expected (intuitively buyer-seller should be one of the biggest kind of transactions).

The neutral user group was also surprisingly large in size and participation. This was not expected, as intuitively they don't have a big economic role to fill. Also interesting is the fact that miners are the main source of bitcoins for the buyers cluster. Overall, the clustering results obtained here allowed us to confidently define miners, investors, and exchanges.

The anonymity of the bitcoin network, allowing users to control many nodes on our dataset may explain some of the unexpected results. A big part of the buyer-buyer transactions could be simply an internal transfer between wallets of the same user and the neutral nodes might actually belong to more sophisticated users like exchanges and buyers. This needs further investigation, but the high average transaction value in this scenario is a good indicative that this is the problem.

For further research and more conclusive results, we recommend taking the following steps: enhance the underlying dataset by improving the clustering of public keys belonging to the same user, and the definition of a bigger annotated set of nodes that can be used to validate the heuristic/clustering results.

9. REFERENCES

[1] F. Reid, M. Harrigan - An Analysis of Anonymity in the Bitcoin System, 2012. http://arxiv.org/pdf/1107.4524v2.pdf

[2] D. Ron, A. Shamir, Quantitative Analysis of the Full Bitcoin Transaction Graph. Working paper. http://eprint.iacr.org/2012/584.pdf

[3] P. Pirolli, J. Pitkow and R. Rao, Silk from a sow's ear: extracting usable structures from the Web, in: Proc. ACM SIGCHI, 1996. http://www.sigchi.org/chi96/proceedings/papers/Pirolli_2/pp2.html

[4] J. Kleinberg, Authoritative Sources in a Hyperlinked Environment, Proc., in: ACM-SIAM Symposium on Discrete Algorithms, 1998. Available at http://www.cs.cornell.edu/home/kleinber/auth.pdf

[5] E. Peserico and L. Pretto, "HITS can converge slowly, but not too slowly, in score and rank", in Proc.: COCOON '09, Berlin, Heidelberg. http://www.dei.unipd.it/~pretto/cocoon/hits_convergence.pdf

[7] Bitcoin transaction network data. http://compbio.cs.uic.edu/data/bitcoin/

[8] FinCEN. Application of FinCEN's Regulations to Persons Administering, Exchanging, or Using Virtual Currencies,Mar. 2013. www.fincen.gov/statutes_regs/guidance/pdf/FIN-2013-G001.pdf

[9] B. P. Eha. Get ready for a Bitcoin debit card. CNNMoney, Apr. 2012. money.cnn.com/2012/08/22/technology/startups/bitcoin-debit-card/index.html

[10] M. Peck. Bitcoin-Central is Now The World's First Bitcoin Bank...Kind Of. IEEE Spectrum: Tech Talk, Dec. 2012.

[11] Hartigan, J. A.; Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C 28 (1): 100–108.JSTOR 2346830

[12] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297

[13] Smith, Marc A., et al. "Analyzing (social media) networks with NodeXL." Proceedings of the fourth international conference on Communities and technologies. ACM, 2009.