A network analysis of the Litecoin blockchain

Marco Monteiro
Stanford University
marcorm@stanford.edu

Toby Bell
Stanford University
tbell@cs.stanford.edu

Abstract

Today Litecoin is the ninth-ranked cryptocurrency with a market cap of $1.5 billion. In this paper constructs and analyzes two network representations of the blockchain. The first is a user-centric approach that represents user addresses as vertices, and payments between users as edges. The second captures the non fungibility of Litecoin, representing transactions as vertices, and the flow between them as vertices. We discover roles in the user graph, assigning high degree nodes to merchants and exchanges, and low degree nodes to ordinary users. In analyzing the transaction graph we see payments are sparse and irregular, and most importantly not representative of a real economy.

1 Introduction

Litecoin is an altcoin—a spinoff of the well-known cryptocurrency Bitcoin—that self-markets as being a cryptocurrency for more lightweight user-to-user payments. It was launched on October 7, 2011 by Charlie Lee, as a fork of the Bitcoin source code with only a few small changes aimed at making it better suited to its stated purpose of lightweight payments. Today, the collective market cap of Litecoin is $1.5 billion, making it the ninth-ranked cryptocurrency at the time of writing. Since, like Bitcoin and many other altcoins, all transaction data on the Litecoin blockchain is publicly available, its economic behavior constitutes a ripe target for analysis.

In this paper we perform a network analysis of the Litecoin blockchain with two primary goals. The first is to assess the development of the Litecoin currency over time, from when it started in 2011 to the present day in 2018. The second is to compare patterns of economic behavior on the Litecoin blockchain to those of "real" currencies and economies, and assess the current state of Litecoin as a viable monetary instrument. We initially had a third goal of de-anonymizing the network as other work has done for other cryptocurrencies [3] [4] [5], but we ultimately decided that data was not available today for Litecoin. Regarding these two goals, we note that we are neither cryptocurrency experts nor economists, and in many ways our analysis is incomplete. However, we leave many avenues for future work, lead among which is a comparison between the graphical roles of Litecoin addresses and the real-world economic roles of their users.

2 Background

This paper analyzes the Litecoin blockchain from two perspectives: a user-centric perspective and a transaction-centric perspective. In this section we provide a cursory technical explanation of Litecoin, including definitions of several relevant terms, and then we precisely define these two perspectives as two different graph projections of the blockchain.

Litecoin is a currency, and it is denoted by the symbol Ł. It can be held and exchanged in amounts as small as Ł0.00000001—one one hundred millionth of a Litecoin, or one *Litoshi*. Because it is a cryptocurrency, Litecoin has no physical cash, and it is held and exchanged exclusively digitally. The *Litecoin network* is a computer network of connected devices that collectively facilitate the

secure operation of the Litecoin currency. These computers constantly communicate using a common protocol with a distributed, peer-to-peer network topology. Though we will soon define the blockchain as one of the core pillars of this protocol, its precise cryptographic details are not relevant to this paper, and for this work it is sufficient to understand that in following this protocol, the Litecoin network ensures the usefulness of Litecoin as a monetary instrument.

The *Litecoin blockchain* (or simply the *blockchain*) is a public data structure describing every Litecoin transaction that has ever occurred. Every computer in the Litecoin network stores and verifies its own separate copy of the blockchain. The Litecoin protocol ensures that all of these copies remain identical as new transactions occur, thus justifying the term *the blockchain*. As with the Litecoin protocol, the cryptographic details of the blockchain are not important for this paper, other than to note that its design makes it almost impossible for a bad actor to modify it retroactively. This means that no one can spend money they don't have, or retrieve or re-spend money they spent in the past (known as a *double-spend*). The current size of the blockchain is about 20 GB.

A Litecoin *address* is similar to a bank account number, in that it uniquely identifies a monetary account. An example of a Litecoin address is `MTvnA4CN73ry7c65wEuTSaKzb2pNKHB4n1`. Addresses are recorded in the blockchain as the recipients of transactions, and the set of unspent transactions paid to a given address determines that address's balance. Although in general one can expect a loose correlation between Litecoin users and addresses, there is no hard correspondence. A single user can create and use multiple addresses, and multiple users can agree to share a single address. Since the blockchain stores addresses instead of user identities, in this paper we will use an address-centric view of the blockchain to approximate a user-centric one.

Finally, a *transaction* on the Litecoin blockchain is a statement of the transfer of funds. Each transaction lists one or more *inputs* from which it draws funds, and one or more *outputs* to which it sends funds. Each output specifies a recipient user (via their address) and a value to pay them, and each input must refer to an unspent output of a previous transaction. The sum of the values of a transaction's inputs must equal the sum of the values of its outputs. For an example, imagine Alice pays Bob Ł10 one morning, and Bob pays Carol Ł5 that afternoon using the funds he received from Alice. Let us refer to these two transactions as T_1 and T_2, respectively. In this case, T_1 might specify a single output $O_1 = (\text{Bob}, 10.0)$ that declares the transfer of Ł10 to Bob (note that T_1 must have also specified one or more inputs to fund this transfer). Then T_2 would specify a reference to O_1 as its single input, and it would specify an output $O_2 = (\text{Carol}, 5.0)$ paying Ł5 to Carol. However, the sum of a transaction's outputs must equal the sum of its inputs, so T_2 would additionally specify an output $O_3 = (\text{Bob}, 5.0)$ paying the remainder of the funds back to Bob himself. This interaction is illustrated in Figure 1. Finally, note that if Bob were now to attempt to issue another transaction funded by O_1 (*i.e.*, perform a double-spend), the Litecoin network would reject it, since O_1 has already been spent by T_2.

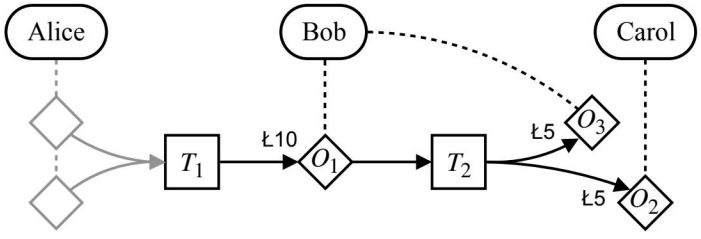

Figure 1: Illustration of a multi-transaction exchange between three users. The light gray vertices simply represent the "one or more inputs" needed for T_1.

Having defined the blockchain, addresses, and transactions, we proceed to our economic network analysis of Litecoin.

3 Methods

3.1 Graph representation

The full structure of the blockchain is quite complex, and even more complex than the structure portrayed in Figure 1, since transactions are further bundled into a Merkle list of connected *blocks* that are added to the chain over time. Given this, for our work it is necessary to create a simplified model of the blockchain that is better suited to network analysis.

We represent the blockchain as a bipartite directed graph of transactions and addresses. The transactions and addresses form the vertices of the graph, and we use directed edges to indicate relationships between them. In particular, an edge from a transaction to an address indicates that the transaction paid at least one output to that address. Similarly, an edge from an address to a transaction indicates that the transaction sourced at least one input from that address. An example of such a graph for the scenario above is shown in Figure 2. This graph, called the *transaction–address graph*, is a significant simplification of the blockchain, but at the cost of completeness we gain ease of analysis. From this graph model we derive two projections, which we then use for our network analysis of Litecoin.

Figure 2: The transaction–address graph for the scenario from Figure 1.

We mentioned previously that we seek to analyze the Litecoin blockchain from two perspectives: a user-centric perspective and a transaction-centric perspective. To this end we create two directed bipartite network projections from the aforementioned transaction–address graph. The first is the *address graph*, defined as a directed graph of addresses, where an edge (A_1, A_2) indicates that there was at least one transaction T_1 with edges (A_1, T_1) and (T_1, A_2) in the transaction–address graph. Since we think of addresses more or less as users, this is effectively equivalent to a "call graph" or communication graph in the directed social network setting. Formally, the address graph may include self-edges, since paying an arbitrary amount to someone will almost always require paying a remainder back to oneself. However, these edges tend not to be particularly meaningful and in practice we often ignore them.

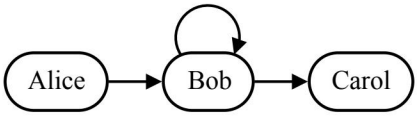

Figure 3: The address graph for the scenario from Figure 1.

The second graph projection we use for analysis is the transaction graph, defined as a directed graph on Litecoin transactions where an edge (T_1, T_2) indicates that an output of T_1 was used as an input of T_2. This graph captures the non-fungibility of Litecoin and the flow of funds through the network, since each transaction is funded by one or more previous uniquely identifiable transactions. This allows us to exactly assess where a particular quantity of Litecoin "came from" or "went to."

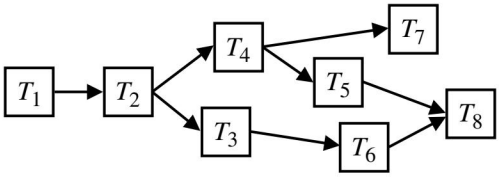

Figure 4: The transaction graph for the scenario from Figure 1 (T_1 and T_2 only), with additional future transactions to illustrate structure.

It is on these two projected graphs—the address graph and the transaction graph—that we perform

our network analysis.

3.2 Dataset

Our dataset for this project is the Litecoin blockchain. By design, the blockchain is public and easy to download (if your Internet connection is up to snuff) simply by joining the Litecoin network. Today, the full blockchain has a size of roughly 20 GB, and it contains over 29.7 million transactions and 2.7 million addresses. For our analysis, we use two smaller subchains from the full blockchain. The first subchain, referred to from here on as Chain A, contains all transactions within a roughly one-year window following the initial launch of Litecoin—between October 7, 2011, and October 12, 2012. The second subchain, Chain B, corresponds to an 8-month window several years later—between August 9, 2016, and April 10, 2017. We will use these two chains as a means for analyzing the difference between Litecoin in an early period and a more mature period of its existence. Cursory statistics for Chain A and Chain B are given in Table 1.

Table 1: Dataset statistics

	Chain A	**Chain B**
Start date	October 7, 2011	August 9, 2016
End date	October 12, 2012	April 10, 2017
Transactions	607,361	997,126
Addresses	606,510	896,858
Volume (LTC)	88.40 million	1.173 billion
Volume (USD)	88.40 million	493.2 billion

4 Results

4.1 The address graph

In this section we provide an analysis of the address graphs from Chain A and Chain B. We refer to these two graphs as Addr-A and Addr-B, respectively. Our primary goal in this analysis is to understand the structure of the address graphs. In the process we discover that many properties of the Litecoin economy are similar to the economy of a small country such as Estonia. [1] The analysis is broken into three sections: a study of the degree distribution, a connectivity analysis, and finally an analysis of induced subgraphs.

4.1.1 Degree distribution

To understand different roles in the address graph, we first plot the degree distributions of Addr-A and Addr-B.

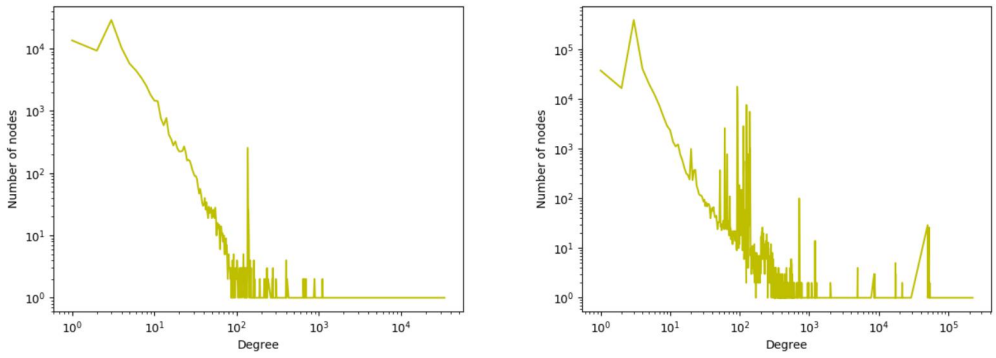

Figure 5: Degree distribution for Addr-A (left) and Addr-B (right).

Note that in both graphs, the majority of addresses are connected to fewer than 20 other addresses. These addresses represent regular Litecoin users. Few of them have made more than a handful of payments with Litecoin since the currency's inception. A minority of the vertices have very higher degree distributions. We conjecture that these addresses belong to merchants and exchanges.

Both degree distributions follow an organic decay, apart for the high concentration of vertices around degree 100. This is likely caused by a latent variable not captured in our graph. One possibility is the vertices with degree near 100 all belong to mining pools. Mining pools split the reward from mining a block. Therefore it is likely that all of the addresses in a mining poll form a fully connected clique. It is possible that 100 addresses is the optimal size for a mining pool, explaining the spike in vertexes with degree around 100. An alternate explanation is that when an address mines a block, they receive a small transaction from every address sending a transaction in that block. The vertexes with degree around 100 can represent addresses that mined a block.

Note that the degree distribution of vertexes with degree less than 50 does not change much between Addr-A and Addr-B. This suggests the transaction frequency of regular Litecoin users did not change a lot between 2011 and 2012. However, a few vertexes in Addr-B have much higher degree than vertexes in Addr-A. Assuming these high-degree vertices represent merchants and exchanges, we conclude that in 2017 merchants and exchanges have many more customers than they did in 2011.

4.1.2 Connectedness analysis

As a precursor, we calculate the clustering coefficient of both address graphs.

Table 2: Clustering coefficient of address graphs

	Addr-A	**Addr-B**
Clustering Coefficient	0.098877	0.189887

These clustering coefficients are surprisingly high. In Addr-B, if an address pays two other addresses, there is almost a 20% chance that those two addresses will send money between each other. Also note that the Litecoin graph became more connected over time. We conclude that even though Litecoin is a decentralized payments network, it also represents a social network where two addresses are more likely to be connected if they are connected socially. Furthermore, these clustering coefficients are consistent with that of a real world economy. A study of the payments network in Estonia found a clustering coefficient of 18%.[1]

Next, we try to understand the different components of the Litecoin address graph, similar to the analysis in Broder et al. [2]

Table 3: Components of address graphs

	Addr-A	**Addr-B**
Number of nodes	606,511	896,850
Number of edges	7,155,187	11,356,252
MxSCC	113,706	679,969
MxWCC	606,511	896,850
OUT	11,855	92,620
IN	446,191	80,873

Note that in both graphs the MxWCC contains every node. This is unsurprising, since by design all Litecoin come from a mining reward. An address cannot spend Litecoin that it does not first receive, and therefore any address's payment can be traced back to a mining reward. Note that in constructing the address graph, we consider all mining rewards to come from the same "virtual" address vertex. The Estonia study also found the largest WCC contained very close to 100% of the vertices in the graph.

MxScc 25% 18.7%

4.1.3. Induced subgraphs

To understand the address graphs on a microscopic level we randomly sampled and visualized subgraphs of both address graphs. To create the subgraphs first we sampled random nodes with degree 2. Then we performed a breadth first search with depth 2 of incoming and outgoing nodes, and created the induced subgraph of all visited nodes. Below is a sample of 6 induced subgraphs from Addr-A and Addr-B.

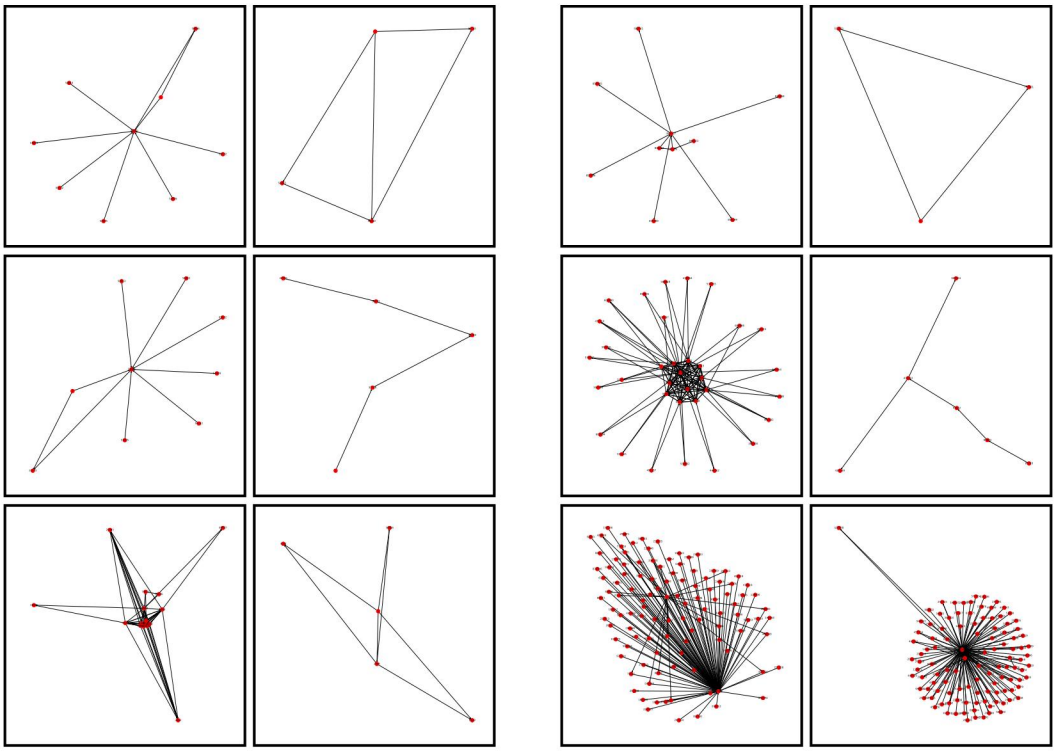

Figure 6: Induced subgraphs from Addr-A (left) and Addr-B (right).

Each of the subgraphs either contains a single digit number of nodes, or over 100 nodes. Each of the subgraphs with over 100 nodes has at least one node with very high degree. These central nodes likely are a merchant or currency exchange because they are accepting payments from hundreds of users. In the analysis of degree distribution we noted the unusually high concentration of nodes with degree around 100. Some of these nodes appear in the sampled subgraphs from Addr-B.

By randomly sampling subgraphs we find a higher concentrations of high degree nodes in Addr-B than Addr-A. This supports the hypothesis that over time more merchants and exchanges joined the Litecoin blockchain.

In the randomly samples subgraphs we observed a surprisingly high number of small communities not connected to a high degree node. This suggest people are doing more than buying Litecoin on exchanges, and using it to pay merchants. Ordinary Litecoin users are actually transferring money between themselves. It is likely these users either know each other in person, or have communities through an online channel.

4.2 The transaction graph

In this section we provide an analysis of the transaction graphs from Chain A and Chain B. We will refer to these two graphs as Txn-A and Txn-B, respectively. Our focus in this analysis is, as for the address graph, on the development of the Litecoin network during the time between Chain A and Chain B, and find a large increase in the interconnectedness of the transactions during this time. We also compare properties of the Litecoin transaction graph to properties of "real money."

4.2.1 Degree distribution

The degree distributions of Txn-A and Txn-B are shown in Figure 7. These plots show the undirected (total) degree of each transaction, but we note that the distributions of in-degrees and out-degrees are nearly identical. Both transaction graphs' degree distributions follow the common power-law decay pattern, which is arguably fairly surprising in this case. Especially in the out-degree case, having a degree of 100 indicates that a single transaction was used to distribute funds to 100 different addresses (or at least in 100 different chunks). It is initially surprising that we see transactions with more than a very small number of out-degrees, since in normal spending patterns with "real" money we usually think of paying only one person at a time. It is possible that various Litecoin wallet software implementations, which users use to issue transactions, elect to source many different inputs and produce many small outputs when producing transactions, but this is speculation on our part, and we leave this as an open direction for future work.

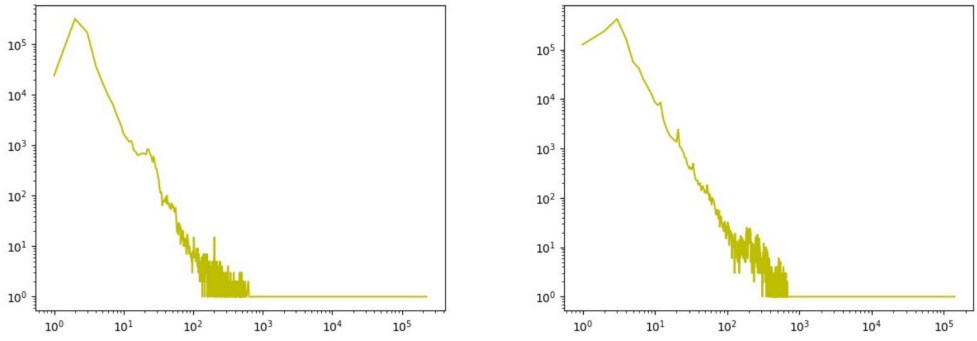

Figure 7: Degree distributions of Txn-A (left) and Txn-B (right).

4.2.2 Reachability analysis

Here, we give a reachability analysis in the style of Broder et al. [2] to assess the interdependence of transactions from Chain A and Chain B, and the extent to which value ultimately flows through other transactions. It is important to note that this kind of analysis depends on the non-fungibility of public-ledger cryptocurrencies like Litecoin does not easily extend to "real" money. We use it as a means of assessing the difference between the structure of transactions in the early period versus the late period in the

Our reachability analysis consists of selecting 500 random vertices (*i.e.*, transactions) from each of the two transaction networks, and then performing two breadth-first searches starting from each of these vertices. The first search follows only inlinks, while the second search follows outlinks. We plot the percentage of vertices reached in each search against the ascending percentile of the 500 starting vertices. These results are shown in Figure 8.

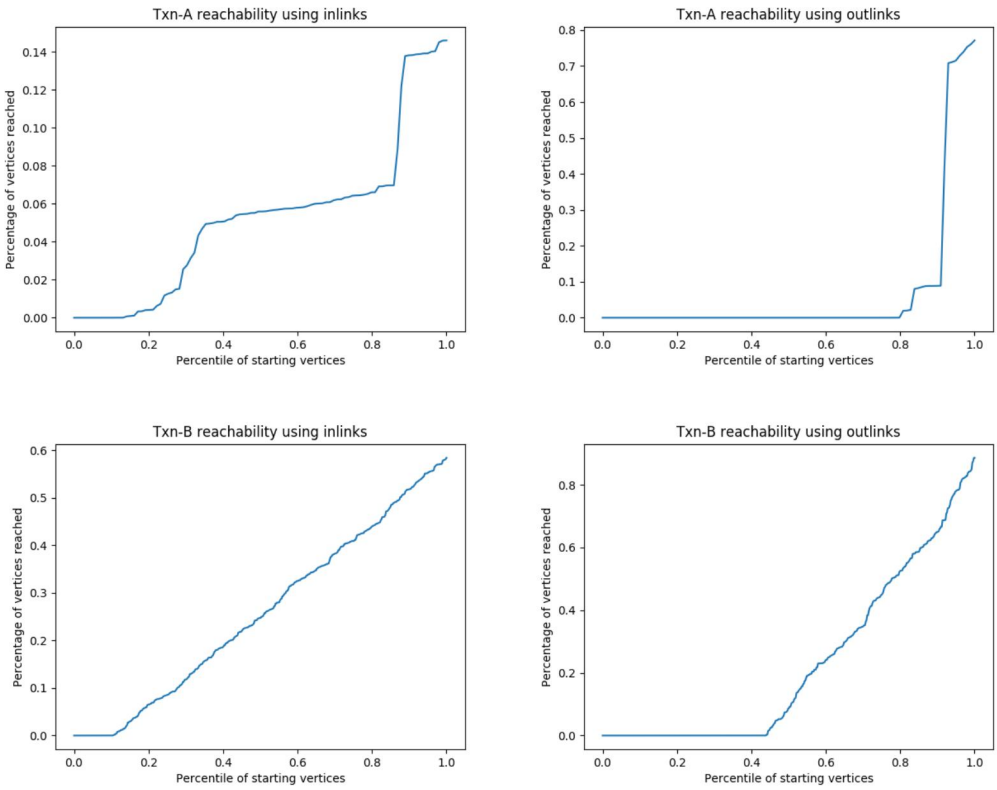

Figure 8: Reachability profiles for Txn-A (top) and Txn-B (bottom).

We note several interesting features of the transaction graphs based on these plots. First, as can be seen in the top-right and bottom-right plots, both Txn-A and Txn-B have a high percentage of vertices that do not reach any other vertices using outlinks—roughly 80% for Txn-A and 45% for Txn-B. These zero-out vertices correspond to unspent money: in the early period (Chain A) 80% of received payments went unspent for the entirety of that first year, and in the more mature period (Chain B), 45% of received payments went unspent for the remainder of the period. This roughly indicates that users tend to hold on to their money for long periods of time, and although the volume did increase in Chain B, it is still far lower than one would expect for a real currency. This is consistent with the general perception that cryptocurrencies are used more as investment instruments than for payments, and is discouraging for proponents of a cryptocurrency-powered future.

Additionally, note the small number of transactions with extremely high outlink reachability. In both Txn-A and Txn-B, the top nodes have an outlink reachability of around 80%. This essentially indicates that 80% of all transactions directly or indirectly received money from a single common source. This can partially be accounted for by the fact that Litecoin (and other cryptocurrencies) use mining to introduce all currency into the network, meaning that all money ultimately comes from a so-called "generational transaction," or mining reward. However, the maximum outlink reachability of 80% is still unnaturally high, since there are over 150,000 distinct generational transactions in both Txn-A, and Txn-B.

A better explanation is supported by considering the inlink reachability of Txn-B as well (the lower-left plot). We see that both reachability plots for Txn-B are fairly linear, with a consistent and smooth increase in reachability as we move through the graph. Furthermore, the maximum inlink reachability—almost 60%—corresponds exactly to the percentage of vertices with nonzero outlink reachability. This points to a rather "narrow" and stable pattern of new transactions, in which the transaction graph grows forward evenly over time, rather than with strong bias towards a particular group of users. How this compares to spending patterns in "real" currencies remains to

be seen, and is difficult to assess because of the aforementioned fungibility of such currencies.

5 Discussion and future work

In our analysis we identified several vertices of interest. In future work one can search the Litecoin addresses of these vertices and see what real world information is available. In this paper we proposed several roles for vertices in the address graphs. (ex. spender and merchant). Real world information from the address of the vertices could, in theory, verify these roles.

We note that in the course of this project we attempted to find public keys of merchants on forums such as Reddit, as well as on Litecoin exchanges such as Coinbase. Despite the principle of decentralization, public keys were very difficult to find, and we found that large merchants intentionally hid their public keys. For example, the travel website Expedia started accepting bitcoin earlier this year, but they only accept payments through Coinbase's payments platform, and Coinbase does not reveal Expedia's public key. In this scenario Coinbase effectively operates as an automated clearing house (ACH), defeating the the purpose of a decentralized currency like Litecoin.

For tractability we limited our analysis to two subgraphs of the Litecoin blockchain network over two separate time intervals. A further work can apply the methods proposed in this paper to the entire Litecoin blockchain.

References

[1] de la Torre, S., Kalda, J., Kitt, R. & Engelbrecht, J (2016). *On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia.* In Chaos, Solitons & Fractals, vol. 90: 18–27. arxiv.org/abs/1602.04352.

[2] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. (2000). *Graph structure in the Web.* In Computer Networks, 33 (1–6): 309–320. sciencedirect.com/science/article/pii/S1389128600000839.

[3] D. McGinn et al., "Toward Open Data Blockchain Analytics: A Bitcoin Perspective."

[4] A. Narayanan and V. Shmatikov, "De-anonymizing social networks."

[5] F. Reid and M. Harrigan, "An Analysis of Anonymity in the Bitcoin System."