Identifying Beneficial Connection Types in Payment Channel Networks: The Case of Lightning

Daniel Rincon[1]
University of California Berkeley

Eva Yiwei Wu[2]
University of California Berkeley

Sofia Dewar[3]
University of California Berkeley

Daniel Zhu[4]
University of California Berkeley

May 15, 2020

Abstract

The Lightning Network is an innovative public decentralized financial infrastructure that allows users to make instantaneous low-cost digital transactions. It works by allowing participants to create pair-wise payment channels that can then be aggregated to form a network. This network enables participants to send payments to other participants by routing them through mutual connections. The creation of channels and the routing procedure is defined by a protocol that uses cryptography to guarantee payment accuracy, security, and privacy. Nonetheless, for payments to be fast, it is desirable that channels are created in a way that reduces the average number of 'hops' between any two participants, and in doing so the network should not become too dependent on 'hubs' that centralize multiple connections. Given that the Network is open, the way individuals connect is determined by their incentives, the rules defined by the protocol, and the capabilities and information provided by the network's software. Protocol and software designers seek to adjust rules and features in the system to 'nudge' participants so that they form connections that not only benefit them but also benefit the network as a whole.

Our project contributes to existing literature by identifying the three most salient connection types in the network and how they are associated with individual and overall network benefits. We expect this research can guide protocol and software improvements to make the Lightning Network faster and more robust.

Daniel Rincon and Eva Wu equally contributed to all parts of this project. Sofia Dewar and Daniel Zhu contributed to the literature review, interviews and analysis conducted to support the qualitative component of the project.

This work was conducted under a research grant from Ripple's University Blockchain Research Innitiative (UBRI). The authors would like to thank Professors Joshua Blumenstock and Coye Cheshire for their valuable comments.

[1]dsrincon@berkeley.edu
[2]eva.wu@berkeley.edu
[3]sofia.dewar@ischool.berkeley.edu
[4]danielzhu0@berkeley.edu

1 Introduction

1.1 The rise of Lightning

Introduced in 2009 as a peer-to-peer electronic payment system, Bitcoin enabled digital transactions without the need for trusted intermediaries. By using cryptographic signatures to prove coin ownership and leveraging a decentralized (proof-of-work) network of transaction validators, Bitcoin was intended to become a more robust, private and efficient solution to transfer value online [30]. In spite of the challenges the system still faces with respect to privacy and security [14], where it probably lags the most is in its ability to support fast transactions at scale. Confirming a transaction takes at least 10 minutes and can cost on average over $0.40 USD, making it an impractical solution to enable new micro-payment applications or replace existing solutions. Nonetheless, as of March 2020, a significant amount of value was locked into Bitcoin ($161 Billion USD representing 64% of the total cryptocurrency market cap) [2], reaffirming the confidence that a large segment of the market has in its capabilities to transform payments.

With the intention of reducing the latency and cost of executing Bitcoin transactions,the Lightning Network (LN) was proposed in 2016 [33] and launched in early 2018 .

1.2 Payment Channel Networks

The LN is a Bitcoin-specific implementation of a Payment Channel Network (PCN) that allows users to make Bitcoin payments without settling them on-chain, enabling low-cost and nearly instantaneous transactions making use of Payment Channel technology.

1.2.1 Creating a Payment Channel

To create a Payment Channel, a pair of participants, let's call them **Eva** and **Dan**, must deposit funds into a special Bitcoin address, called a 2-out-of-2 multisignature address (2-2-multisig). Creating this funding transaction requires both of them to sign using their private keys. This special address is similar to a shared bank account, users can only withdraw funds if they both provide their private keys and they configure it in such a way that the balance is distributed in a specific way between them. In order to avoid any counter-party risk associated with either Dan or Eva becoming unresponsive once the funding transaction has been created and having their funds locked, they both share 'refund' transactions that allow them to recover the balance that was originally assigned to them. Completing this whole process is known as opening a **Payment Channel**. Given that opening a payment channel is a regular Bitcoin transaction, they must wait at least 10 minutes for validation and pay a fee, but once this is done they can send Bitcoin payments through this channel almost instantaneously and with zero fees.

1.2.2 Transacting within a Payment Channel

To transact Bitcoins within their newly created Payment Channel (or move the state of the Channel forward), Eva and Dan must exchange signed transactions that change the original balance in the 2-2-Multisig. Transactions are electronic messages which they can just send over a private medium. Lets say that the original balance in the Multisig address was 0.5 Bitcoin (BTC) for Dan and 0.5 BTC for Eva. Then Eva wants to send 0.2 BTC to Dan. In order to do so, they simply sign a new transaction that reassigns the balance in the 2-2-mulstisig address to 0.7 BTC to Dan and 0.3 BTC to Eva. It is worth noting here that transaction values are limited by the total value originally deposited in the channel, called the channel's *capacity* and more importantly by the distribution of the values between the participants.

1.2.3 Closing a Payment Channel

Whenever Dan and Eva want to close the channel and 'claim' their balance, they simply agree to send the latest exchanged transactions to the Bitcoin Blockchain and redeem their corresponding balances. In order to avoid that either Eva or Dan using an old transaction to redeem a previous balance from the Blockchain the transactions they exchange create certain timelocks to protect them. For example, if Eva

tried to redeem the old Dan: 0.5 BTC / Eva: 0.5 BTC transaction after they had exchanged the Dan: 0.7 BTC / Eva: 0.3 BTC transaction, she would have to wait a certain amount of time for the transaction to be processed. During this time Dan could audit that she was using an old transaction and settle the channel with the most current one. The real-value of a Payment Channel comes from combining channels from multiple interconnected participants to form a network; this is known as a *Payment Channel Network*.

1.2.4 From Channels to Networks

A Payment Channel Network gets formed when more than two participants get connected via a common Channel. Going back to our example, lets now suppose that Sofia, who has a Payment Channel set up with Dan wants to send a payment of 0.1 BTC to Eva. Instead of creating a new Payment Channel with her, Sofia can route the payment through Dan. This is an overview of how this process works: 1) Sofia asks Eva to generate a secret S and only share it with Dan in exchange for the payment. Additionally Eva uses a Hash H function to hide this secret and shares the hidden version $H(S)$ with Sofia. 2) Sofia then tells Dan that she is willing to pay him 0.1 BTC in exchange for a secret S that matches $H(A)$. 3) Dan goes to Eva and tells her that her he is looking to 'purchase' secret S that matches $H(S)$ for 0.1 BTC. Eva shares S with Dan in exchange for the 0.1 BTC. 4) Dan then shares S with Sofia in exchange for 0.1 BTC. The net result of this process is that 0.1 BTC changed hands from Sofia to Eva. This example is illustrated in Figure 1. In a real world example, Eva would be compensated with a small fee for having facilitated the transaction and if anything had gone wrong half-way through the process, all intermediate transactions would have been reversed. This same mechanism can span multiple 'hops' so that participants that are not directly connected can send payments to each other.

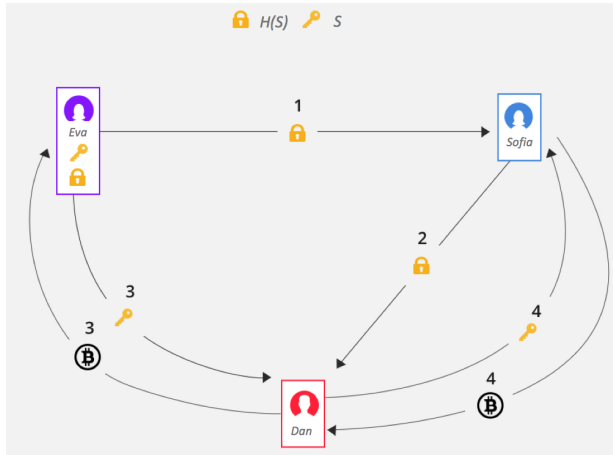

Figure 1: Example of a multi-hop payment. Adapted from: https://www.blockcham.org/understanding-the-lightning-network-part-2-creating-the-network/

1.3 The challenge faced by Lightning

The LN has been live for more than 2 years and currently has over 11,500 nodes connected by more than 36,000 channels, with the capacity to transact over $7.9 Million USD [4]. The mechanisms described above, that regulate the way in which participants interact to send, receive and route transactions are defined by the Lightning Network Protocol, more specifically by a set of specifications know as Basis of Lightning Network Technology (BOLTs) [25]. There are different software clients or 'daemons' that implement the protocol. Both the protocol and most of the available software clients are open-source projects that are maintained by a community of enthusiasts and organizations.

The network itself is open and anyone can join as long as they run a client that is complaint with the protocol. This means that the structure of the network, commonly referred to as the network's *topology* is determined by the uncoordinated actions of participants opening and closing channels. Aside from an auto connect feature implemented in some clients, which seems to be deemed by the community as not very reliable (see Figure 2), participants rely on external forums and node ranking websites [1] to decide

ways in which they can connect. Nonetheless not all of their individual connection decisions are aligned with shaping the network in a way that is beneficial for the network as a whole. If, for example, most participants prioritize having short connections to other nodes (in order to have faster transactions) they will connect to more central hubs that aggregate connections; this can make the network more vulnerable to attacks on these central hubs. On the other hand if most nodes want to become intermediaries, they might connect in ways that 'stretch out' the network and make it less efficient (see Figure 3). Additionally, even if the protocol or the software clients could suggest individual connections that were beneficial to the network, those connections might not be individually beneficial or feasible for the nodes involved. The challenge for designers of the protocol and software clients is to identify connections that are individually beneficial and that help increase the efficiency of the network so that they can implement features that incentivize those connections.

Figure 2: Entry on the Lightning Network Wiki describing issues with the autopilot feature in LND, which is the most popular LN node client. Source: https://lightningwiki.net/index.php/Using_Autopilot_on_LND

1.4 Contribution

There has been considerable work exploring the underlying assurances that allow PCNs to be operational and secure as well as work that characterizes

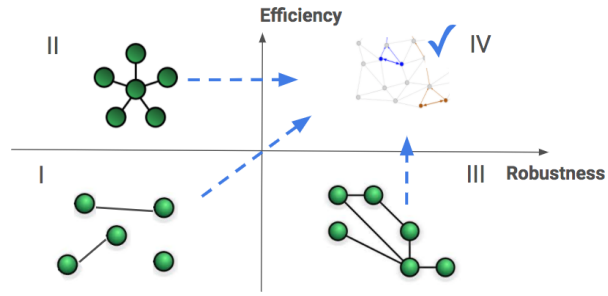

Figure 3: Illustration of the trade-off between efficiency and robustness. I) Represents a highly disconnected network that is neither robust nor efficient. II) Is a highly centralized network that is very efficient but not very robust because it is highly vulnerable to attacks on its central node. III) Is a more robust network that can maintain reasonable connectivity even if most of its nodes are attacked. IV) This quadrant represents the types of network topologies that make Payment Network most beneficial for all users because they combine efficiency, robustness and connectivity.

macro network properties, but thus far no attention has been placed on how the dynamics of network formation impact the overall network structure. The LN is the largest 'real-world' implementation of a PCN and offers an ideal opportunity to study these dynamics.

Additionally given that the LN is a complex large-scale network, existing analytical results from the economic network formation literature do not provide solutions to the challenge described above. On the other hand, results from Network Science that deal with network optimization either assume a top-down control of the network or require the estimation of multiple parameters to potentially run network formation simulations. To overcome these obstacles this paper takes an empirical approach that combines qualitative and quantitative analysis. The contributions of this work to existing academic research on LN and PCNs are listed below:

- Identification of the drivers that motivate participation and shape connection decisions in the LN.

- Creation of a granular dataset of graph snapshots for every step in the evolution of the LN together with key node-level and network-level metrics for every snapshot.
- Identification of the most salient types of connections present in the network with respect to the node attribute that are most relevant to participants.
- Estimation of the statistical relationship between creating specific connection types and individual outcomes as well as network-level outcomes (efficiency and robustness).

We believe it is crucial to study these sociotechnical systems to understand the technical, economic and social factors that influence their formation and evolution because the shape and capabilities of these systems can potentially have profound implications for the equal access to financial services.

2 Related Work

Our work builds on top of existing research that either explores theoretical properties of PCNs or empirically measures the topology of the LN. Our main contribution lies in applying theories of social exchange and collective action to qualitatively explore the motivations behind public node operators and use these findings to define and empirically test a strategic model of network formation. This section describes related work in these areas.

2.1 Properties of PCNs and the LN

The LN [33] was introduced as the first detailed design of a PCN to serve as a solution to Bitcoin scalability. Given it was proposed as an alternative to move transactions 'off-chain', some work, particularly on the theoretical front focused on modeling how the interplay between the conditions of the main Bitcoin blockchain (on-chain fees) and those of the LN could explain network formation. On this line of work [5] modeled network formation as the optimization problem of a payment operator that controls all the nodes of the network and wants to maximize routing revenue given some restrictions. On the other hand, [6] models network formation as a game between different agents that can choose between transacting on-chain vs. creating channels on the LN. From the perspective of the LN economics alone, [12] models network formation as a Bertrand competition model where different 'firms' (node operators) choose an optimal level of connectivity and fees in the network to maximize their utility given an opportunity cost of capital used to open channels. Other theoretical work [10], models network formation as a bond percolation process where nodes preferentially attach to other nodes that exhibit high 'fitness', where the 'fitness' of a node is a function of its capacity and the volume of transactions it processes.

From an empirical perspective recent academic work [28, 38, 35] has focused on measuring the static and dynamic topological features of the network, in particular on the resiliency of the network to attacks that disrupt nodes or channels. These results point out to the fact that the network has a centralized configuration with a few nodes that act as hubs and even though the network is robust to random disruptions, it isn't so to the removal of these highly connected nodes. Given that transactions are not public in the LN the profitability of different nodes is not observable,[11] explore the question of the networks profitability by looking at the current topology and simulating different transaction dynamics. This work concludes that under the best estimates, running a node is not profitable and conjectures that most of the node operators must be acting altruisticalxly.

Existing work has either focused on performing theoretical modeling of the network that is not informed by any insights of actual node operators and is not validated by empirical data or by performing purely empirical measurements of network structure. Our approach is to perform a qualitative exploration of the underlying motivations of node operators and use these motivations to define a strategic model of network formation that can be empirically tested. The next subsections discuss the relevant literature that informs our research.

2.2 Social Social Exchange Theory

In our assessment of the Lightning Network's viability, we consider the social implications of the network's transactions through the lens of social exchange theory. Social exchange theory focuses on the benefits that can be obtained through social interactions. People depend on others for various resources and needs, and they provide them to one another through social exchanges [29]. Within the Lightning Network, social exchanges occur at a basic dyadic level whenever two individuals create a Payment Channel and privately exchange signed transactions with one another.

In the case of these more complex network exchanges, potential power dynamics [15] can develop based upon the centrality of a user's position within the network of Payment Channels. Part of the phenomenon we are trying to observe is to see if users will act selfishly and attempt to maximize their own position of power by establishing themselves as the central node within the exchange network because doing so will allow them to collect fees from relaying transactions to other nodes. The power derived from an individual's structural position within an exchange network is what Mark Granovetter describes as "the importance of structural holes" in which one specific individual is the only connection between multiple separate actors, allowing that particular user to exploit his strategic position and maximize his own gains [19].

2.3 Collective Action

The literature on collective action and social movements studies how groups of people come together to pursue shared goals. These perspectives provide an initial framework for understanding the cognitive processes that may motivate individual actors to participate in this financial ecosystem and how individuals ascribe meaning to new technological and social structures such as the Lightning network. Furthermore, peer-to-peer systems provide further insight into the way collective structures are maintained and the variety of motivations that produce these emergent information networks.

Prior scholarship has mapped out five key social psychological dimensions of collectives and social movements [39]. One pertinent dimension we expand upon and explore through this research is the motivations experienced by actors within this network. [39], outline two competing schools of thought when it comes to the question of the nature of the decision-making process, with regards to collective action. Rational decision theory assumes "people seek to obtain benefits and minimize costs" and that they "cognitively process information about the likely benefits and costs of various courses of action and then make a conscious choice about their behavior" [39, 18]. Through our qualitative and quantitative work, we will explore whether these assumptions hold true and whether and where non-rational motivations and heuristics emerge.

2.4 Strategic Network Formation

Strategic network formation is primarily concerned with evaluating the conditions under which networks reach stability, the topology of these configurations and the extent to which these configurations are maximize welfare of all the participants in the system (efficiency). Different definitions of stability are used in the context of strategic network formation, from pairwise stability [22], to pairwise Nash stability, to strong stability [21] offer a notion of pairwise stability that allows to reason about the conditions under which agents are motivated to add or sever links in a network given individual utility functions that are dependent on network topology. On the other hand, [22] presents purely utilitarian and welfare optimizing (Pareto) definitions of efficiency that can be used to measure the extent to which a network topology benefits all its users.

Measuring the trade-off between selfishly driven stability and efficiency in network formation was first studied in [23] and in [17] for the context of Internet -like networks, by exploring the "price of anarchy" as the ratio of the cost of the most costly stable network to the cost of the most efficient one. Another measure was suggested by [40] that calculated the "price of stability", which defined it as the ratio between the best possible (efficiency-wise) pairwise stable network

and the most efficient one. These two quantities can be compared in order to study what [22] calls the anarchy-stability gap.

3 Research Methodology

In this study, we rely on a mix of quantitative and qualitative methods to understand Lightning Network's evolution, predict its resiliency and devise design recommendations to strengthen its viability. In our study, we ask the following questions:

RQ1: What motivates node operators to join the network?

RQ2: How does a Lightning Network node operator choose which part of the network to join?

RQ3: How are the different types of pair-wise connections predictive of the Lightning Network's efficiency and robustness over time?

RQ4: Do the types of connections that produce positive network outcomes also benefit individual participants?

3.1 Qualitative Method

We started with a qualitative study of the Lightning Network users to understand their motivations and connection strategies and to complement and inform the quantitative analysis.

We conducted five one-hour semi-structured interviews with different users including enthusiasts of the Lightning Network, business owners who sought to open their own stores on LN, and a business-to-business provider on the Lightning network. The interviews allowed us to establish a qualitative understanding of users' motivations in joining the Lightning Network, their decision making process in creating and closing connections and their perceptions of the Lightning Network.

We conducted the interviews from March 2020 to May 2020 and found participants in two ways: our personal connections and participants in the Lightning Network group on Telegram. All interviews were held on via teleconferences due to COVID-19. We audio recorded and transcribed all interviews using otter.ai.

3.2 Quantitative Methods

3.2.1 Data

For our quantitative analysis, we obtained a dataset from the Bitcoin software consultancy Big Sun [3] that contained all the public channel creations (72,476) and closures (38,355) in the LN between January 20, 2018 and February 13, 2020. It is worth noting that not all the nodes that participate in the network are public and hence our data doesn't allow us provide as a complete perspective of the network. It is estimated that roughly 30% of the nodes in the network are private [34], nonetheless private nodes are likely to be in the periphery the network because they cannot be used by other nodes to route payments hence they don't affect the relevant topological characteristics studied in this work.

To open or close a channel a corresponding transaction must be submitted to the main Bitcoin blockchain and recorded on a block; blocks get created (mined) on average every 10 minutes. The dataset spans over 112,149 blocks out of which 30,543 had channel openings and 11,024 had channel closures. On aggregate a total of 36,544 contained channel decisions (opens or closures), given that some blocks had both open and closure decisions. The amount of decisions per block follows a power law distribution as it can be seen in Figure 5 with the average number of decisions per block being 2.05 (8.70). The dataset included the addresses of the pair of nodes involved in the channel, the block in which the channel was created, its capacity as well as some additional details about the transaction that was used to create it. If the channel had already been closed, additional data about the block in which it was closed, node balances and type of closure was included. We used this data to re-construct block-by-block snapshots of the network and calculate node-level metrics and overall network metrics at every snapshot. We describe the metrics calculated and the methods used in the next sections.

3.2.2 Extracting network snapshots

We created a graph of the LN for every Bitcoin blockchain block that contained channel modifica-

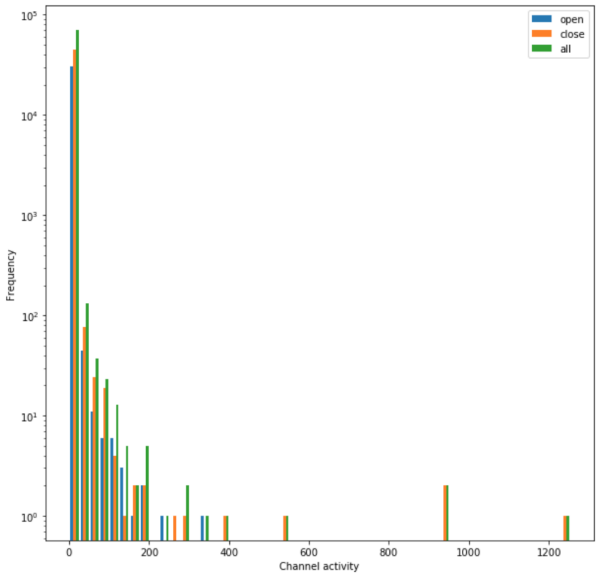

Figure 4: Distribution of Channel opens and closures

tions. Each bitcoin block functions as a timestamp in this context. As described above, only 36,544 blocks in the time period analyzed contained channel modifications, hence we only constructed network snapshots for those blocks, given that the network remained unchanged for blocks without modifications. We denote the set of all blocks for which channel decisions were made as K. For each block k in K we extracted all channel opening and closing decisions up to k and defined a undirected graph $G_k = (N_k, C_k)$ where N_k is the set of all nodes n_i that had joined the network up to block k and C_k is the set of all sets $\{n_i, n_j\}$ where n_i, n_j are nodes that have at least one open channel at block k. For simplicity we denote the set $\{n_i, n_j\}$ by c_{ij}. Given that two nodes can have more than one channel open between them at block k and that every channel is created with a fixed transaction capacity in Satoshis (millionths of a Bitcoin) we defined a node and channel attribute function to map every c_{ij} to the number of channels open between nodes n_i, n_j at time k as: $A_k(c_{ij}) : N_k \rightarrow \mathbb{I}^2$ such that $A_k(c_{ij}) = (chan^k_{ij}, cap^k_{ij})$ where $chan^k_{ij}$ and cap^k_{ij} are the number of channels open between node n_i and node n_j and the total capacity of these channels (in Satoshis) up to block k respectively. Given that for most k the resulting G_k did not end up being fully connected we validated how fragmented the network was at every k. As expected and as is shown in Figure 5, the network fragmentation across time has not been significant and it is reasonable to assume that nearly all of the transaction activity happens only in the largest connected component of the network. Henceforth, when we refer to the network at any given block, we will be referring to g_k.

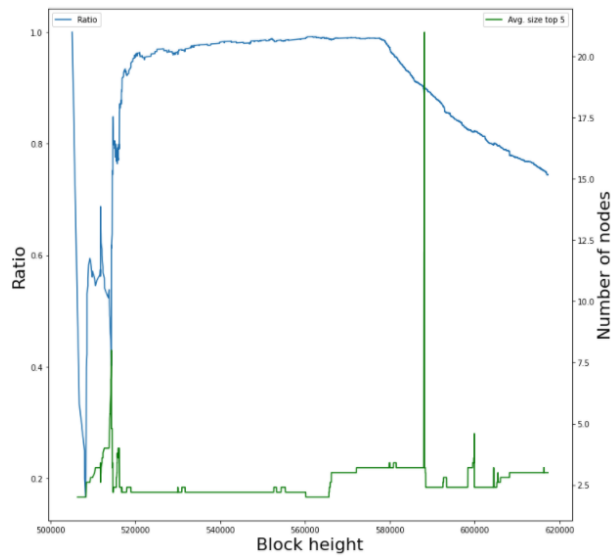

Figure 5: Ratio of nodes in largest connected component g_k over nodes in full graph G_k and average number of nodes in the 5 subsequent (with respect to size) connected components for all snapshots. It can be seen that the ratio of nodes in g_k to nodes in G_k is close to 1 for more than half of the network's life but even when it decreases to values close to 0.8 the average size of the other large components remains close to 2.5, indicating that nodes that don't belong to the connected component are individual isolated nodes or at most sets of few nodes that can't participate in network transactions.

3.2.3 Node level metrics

After analyzing the information provided by participants in the interviews conducted as part of the qual-

itative component of the project (see section below), we identified that participants in the Network used the following metrics to categorize themselves and others in the network: Age, number of channels, capacity of the node's channels and growth in capacity. We then calculated them for each node in every snapshot. In particular, we defined age of a node n_i in the snapshot at block k as the number of blocks elapsed between the block in which that node created its first channel (entered the network) and block k. The definitions of our calculations for number of channels and capacity are described in the previous sub-section. As for capacity growth we calculated it as the change in capacity for that node in the last 3,600 blocks, which roughly represents 25 days. Once these metrics were calculated for each node n_i we normalized them for every snapshot at block k by calculating the ratio between the value of the metric for node n_i at block k and the largest value for that metric in block k.

We additionally identified during the interviews outcomes that nodes expected from their connection decisions: 1) being closely connected to more nodes in the network and 2) increasing their ability to earn routing fees. As a metric for 1) we decided to use closeness centrality, for succinctness we will be calling it *connectedness* and denoting the *connectedness* of node n_i in block k as γ_i^k; for 2) we chose betweeness centrality, which we will be calling *betweenness* and denote the *betweenness* of node n_i in block k as α_i^k. Given the computational challenge of calculating closeness centrality for every node in every snapshot and the fact that we were interested in the relative closeness values of the nodes in a given block k, we resorted to use the estimation procedure proposed in [36]. Similarly, in order to simplify the computation for betweeness centrality we used the current betweeness centrality approximation proposed in [13]. In addition to 1) and 2) our interviews identified other non-structural benefits that are described in our qualitative results section.

For every channel c_{ij} opened in block k_o (and closed in block k_c) in the initial dataset, we appended the rankings of nodes n_i, n_j for age, number of channels, capacity, growth, closeness and betweeness at block $k_o - 1$ (and $k_o - 1$) and also the rankings for connectedness and betweenness at block k_o (and k_c). This allowed us to calculate the change in desired outcomes for each node after a closing and opening decision.

3.2.4 Network level metrics

We calculated network efficiency and robustness of the network over time as follows. Based on [24] we defined efficiency of the network at block k, E_k as the average multiplicative inverse of the shortest path distances between every pair of nodes in the the graph g_k. The higher the efficiency the lower the average distances between nodes, hence transaction can be routed faster. As for robustness, it is particularly relevant in the the case of the LN that the network is resilient to targeted attacks, given that its Small-world topology guarantees significant robustness to random failures. We hence used the results presented in [26] in which various robustness measures are compared and selected and adapted version of the Integral Efficiency defined in [27], given that it proves to be the most consistent one across different topologies and showed to be particularly sensitive to reflect changes in robustness induced by different network formation dynamics. Specifically we defined the robustness of network g_k as $R_k = \frac{E'_k}{E_k}$ where E'_k is the efficiency of the network after removing the top 1% of nodes with the highest degree in graph g_k. We estimated these network metrics every 10 blocks, for a total of 3,652 blocks over the period of study.

3.2.5 Clustering analysis

To extract the characteristics of pair-wise connections that were prevalent in the network channels, for every channel c_{ij} at block k, we calculated the L1 norm distance and averages between nodes n_i and n_j in the channel for the following metrics: age, number of channels, capacity and growth. For example the distance in age between nodes in channel c_ij in block k was calculated as $|Age(n_i, K) - Age(n_j, k)|$ and the average between nodes was calculated as $\frac{Age(n_i,K)+Age(n_j,k)}{2}$. We then used both averages and distances for the metrics above to cluster all the channel opening decisions using K-means clustering and the elbow method. The elbow method indicated that

k=3 is the point where diminishing returns are no longer worth the additional cost. Thus, we identified three salient connection types whose characteristics are discussed in detail in the results section.

3.2.6 Regression analysis

We conducted regression analysis to understand the statistical relationship between decision cluster types and network metrics over time as well as the relationship between decision clusters on pair-wise outcomes. To estimate the first relationship we regressed the difference in Robustness and efficiency between $k-10$ and k ($E_k - E_{k-10}$ and $R_k - R_{k-10}$) against the total number of open and close decisions that belonged to each cluster between $k-10$ and k for all the 3,651 blocks for which we had obtained network level calculations. The results for these regressions are presented in Tables 9 and 8 and discussed in the results section.

To estimate the relationship between belonging to different cluster types and pair-wise level outcomes of opening and closing channels we first estimated the pair-wise outcome for every channel by calculating the average change in conectedness ($\Delta\gamma_{ij}^k$) and betweenness ($\Delta\alpha_{ij}^k$) for nodes n_i and n_j after the decisions in block k as:

$$\Delta\gamma_{ij}^k = \frac{(\gamma_i - \gamma_{i-1}) + (\gamma_j - \gamma_{j-1})}{2}$$

$$\Delta\alpha_{ij}^k = \frac{(\alpha_i - \alpha_{i-1}) + (\alpha_j - \alpha_{j-1})}{2}$$

We then regressed, for every channel opening (and closing) ($\Delta\gamma_{ij}^k$) and ($\Delta\alpha_{ij}^k$) against the dummies corresponding to each of the three cluster types. The results for these regressions are presented in Tables 12 and 13 and discussed in the results section.

We additionally estimated the probability of a channel being open using logistic regression and regressing the state of the channel (open=1, close=0), against the duration of the channel (number of blocks it has been open for) and the cluster to which that connection type was assigned. This allowed us to estimate which cluster types were associated with more stable (long-lasting) channels.

3.2.7 Computational resources

Our data transformation and analysis was performed in Python using the Jupyter [32] interactive computing environment. We made use use of the following libraries: Pandas [31] and Numpy [41] for general data manipulation, NetworkX [20] graph analysis, and Statsmodels [37] for regression analysis.

Given that the process of creating snapshots of the network at the selected level of granularity was computationally intensive but highly parallelizable, we used the Dask Python library [16] for dynamic task scheduling. We set up a data processing pipeline similar to the one proposed by [9] where parallel tasks created by a Dask scheduler and defined for each snapshot were submitted to an AWS Fargate dynamic cluster (40-800 cores) [7], processed and the individual output was stored in an AWS S3 Bucket using the Boto AWS Python SDK [8]. This data was then read back into the Jupyter notebook for analysis. The architecture for this pipeline is depicted in Figure 6.

Our processed data is publicly accessible in this AWS S3 bucket https://ln-strategy-data.s3.amazonaws.com and our preprocessing and analysis code can be found here: https://github.com/dsrincon/ln_strategy.

4 Results

4.1 Qualitative Results

Our qualitative insights offer a more nuanced and detailed look at the motivations and decision making processes guiding individual actors in the network and help inform our quantitative analysis. The findings are described below.

4.1.1 Motivations for Joining the Lightning Network

One of the primary goals during our interviews was to understand the motivations behind why individuals chose to join the Lightning Network and engage in transactions on this financial system. We identified four salient motivations: 1) Business and fi-

Figure 6: Data pre-processing pipeline architecture

nancial interests 2) Gratification from social connections and promoting participation in the network 3) Community values of privacy, transparency and self-governance.

Business / Financial Interests One interviewee described getting paid via SatsforLikes for opening a node.

> "When you get someone else that opens a channel with you, it gives you capacity because [...] the channel state starts with that imbalance towards you. And when you get an incoming channel, you can ask capacity to forward and to send to that channel. So, because of this, there is motivation to pay someone so that they open a channel to you. Yes, it was. It was a motivating factor [...] but for that, I needed to actually have my own node open, not necessarily public like it is." (P4)

Another interviewee discussed why from a business standpoint, he wanted to acquire valuable position in the network.

> "Whereas, if you're setting up a node and you're really opening channels with people, that's kind of a business, you know, you're managing liquidity and you want a position in the lightning network, you're part of the infrastructure. And so I want to make sure that I am connecting to real infrastructure, not somebody that's, you know, playing with an app on his computer." (P5)

Gratification from social connections and promoting participation in the network Most of the users we interviewed said that routing fees and financial incentives are not their primary motivators. In fact, many simply use the default routing fee, (some even set it to zero) and they do not make any significant profit from their routing activity on Lightning. Most of them make less than a penny a month and don't derive any utility from the gain as it is so small. Particularly prevalent among the smaller actors in the network, we found they were less driven by extrinsic motivators. Instead, these individuals seem to be more intrinsically motivated by their collective enthusiasm and belief in the Lightning Network as a viable financial network that can provide a better alternative to our current financial systems. More importantly, they want to share the potential of the Lightning Network with a much wider audience. One user created a video describing daily life using Bitcoin and the Lightning Network while another user covers cryptocurrency news through his podcast. Their passion and interest in growing the network and educating people about the network drives them to join Lightning and form connections of exchange.

One interviewee mentioned that it felt good psychologically when he saw more people route through him or when he received a payment:

"It feels like a game." (P3)

Community values of privacy and transparency Several of our interviewees also mentioned that using Bitcoin and the Lightning Network offers a more secure and transparent mechanism for monetary transactions than that of our current financial system. There seems to be a prevailing distrust in our current financial infrastructure that drove individuals to turn to the Lightning Network as a better, decentralized system allowing users greater independence in their transactions and greater self-governance among themselves. One interviewee also discussed why auditability (transparency) of transactions and payments is another critical feature of the network that appealed to him.

> "[it is] human nature to be corrupt and institutions decay and people find ways to make money, so it's important to be able to self-audit and monitor in Lightning" (P3)

4.1.2 Decision-making Process in Creating and Closing Payment Channels

Within the Lightning Network, the decision to open or close a channel is determined by users' motivations, the rules defined by the network protocol, the capabilities and information provided by the software they use to participate, the information users gather from external sources such as 1ml.com and Lightning Network explorers and direct communications with other users.

Direct Communication Builds Trust

In the Lightning Network, node operators can decide with whom they would like to open channels. Our research revealed a consistent theme: individuals often communicate directly with node operators as a way of assessing whether they wish to open a channel with them. In other words, the infrastructure of the network itself and even the analytical tools (1ML.com, etc.) lack sufficient context for such decision making. Node operators seek additional information through a variety of communication platforms. These include Telegram, Slack, Discord and Reddit. They prefer to connect to those in the Lightning Network with whom they have had previous interactions. One interviewee mentioned that initially he connected with his friends because he could trust them, and he felt reassured that he could count on them to give him his money back if something catastrophic ever happened to his node and it failed. Two other interviewees both mentioned that they liked to have some form of interaction with people via messaging on Telegram to get a sense of what the other person's goals are in setting up their node before initiating a connection or accepting a connection request. Three out of five interviewees feared that if they connect to a random stranger, they would not know for certain if this stranger is going to be reliably online all the time, so if they send this new connection a payment and they went offline, then that money is stuck and they could not get it back until either the person comes back online or after the time lock established by the protocol (see Introduction) had passed.

Unreliable Behavior and Cost to Reputation

When it come to closing channels, our interviewees have mentioned that they will close channels mostly with nodes that have remained inactive or are inconsistently online and offline, so those nodes may be viewed as having a bad reputation and as people close channels with them, these nodes become even more ostracized in the network. P3 also mentioned that actors within the network have a sense of a node's reputation, and said that his own node was growing in popularity and gaining a good reputation as he saw more and more people connect and route payments through him. However, once his node was shut down for a week and then restarted, he wasn't able to route as many payments as before because his reputation had suffered and the community no longer saw him as the same reliable node as before.

> "the way that the lightning routing works, it's kind of a gossip network where nodes tell each other, which nodes are connected to who. And that way they're able to route payments through each other. And if you are a member of that gossip as a route, and you're not actually online, a lot of time, I guess energy will be wasted, trying to route through you and failing, because you're not reliably online." (P3)

4.2 Quantitative Results

As described in the Methodology section, after collecting insights from our interviews with participants we conducted k-means clustering on our augmented decision dataset and analyzed the relationship between different decision types and both network-level and node level outcomes. Our findings are detailed below.

4.2.1 Types of Payment Channel Connections in the Lightning Network

As described in the methodology section we calculated the averages and L1 norm distance between nodes when they opened a channel for the following metrics: age, number of channels, capacity and growth. The elbow method in clustering analysis of payment channel openings indicated that the point where diminishing returns are no longer worth the additional cost was when clusters=3. The characteristics of the three clusters are as follows: $cluster_0$ has low averages and differences across all metrics indicating these channels are between participants with low rankings. The defining characteristics of $cluster_1$ are its higher differences and averages between n_i and n_j in capacity and capacity growth, whereas the defining characteristics of $cluster_2$ are its higher differences and averages between node pairs in age and number of channels (see Figure 7.) Our interpretation of the clustering results is that $cluster_0$ aggregates channels between smaller nodes; $cluster_1$ aggregates channels between the small nodes and nodes that we dubbed as "the new rich" (younger in age and have fewer number of channels, but they hold a high-level of liquidity and are growing faster in their level of liquidity); and $cluster_2$ aggregates channels between small nodes and we dubbed as "the old money" (older in age and have extremely large number of connections, but they hold smaller amount of liquidity compared to the "the new rich" as well as adding liquidity at a much lower rate. To categorize the connection types from statistical results, we named these connections as between 1) The 99% and the 99% 2) The new rich and the 99% 3) The old money and the 99%.

4.2.2 Implications of Node Operators' Connections on Network

We found connections between the 99% are not only conducive to a robust network (which means they make the network more reliable and stable), they are also positively correlated with the efficiency of the network, though to a lesser extent compared to the other connection types (see Figure 8, 9 for regression results). Connections between the well-connected and the 99%, are highly positively correlated with the efficiency of the network, but they are negatively correlated with a robust network, which means they tend to make the network more fragile to attacks. We found that the connections between the 99% and the new rich did not have statistically significant relationship with either of the network properties.

4.2.3 Evolution of the Lightning Network

We analyzed how the different types of pair-wise connections are related to the Lightning Network's robustness and efficiency throughout its evolution. Our analysis revealed four phases of the Lightning Network's history (see Figure 10.)

1) Phase I - "the Stone Age" Phase I corresponds to March 2018 to May 2018. During this period, network efficiency is low but increasing, while network robustness is relatively high albeit decreasing because nodes are decentralized and removals of selective nodes results in insignificant impact on the network. The most prevalent type of connection during this phase is "the 99% and the 99%", potentially because most nodes are small in all measures and there was less inequality among nodes (see Figure 11.)

2) Phase II - "the Golden Age" Phase II roughly corresponds to May 2018 to November 2018. This phase of the LN evolution is characterized by increases in both the efficiency and robustness (thus, we dubbed it "the golden age",) indicating that it is possible to enable faster transactions without relying too much on the centralized participants. The most prevalent type of connections in this period is between the 99%, which is the same connection type we identified as beneficial to both network robustness

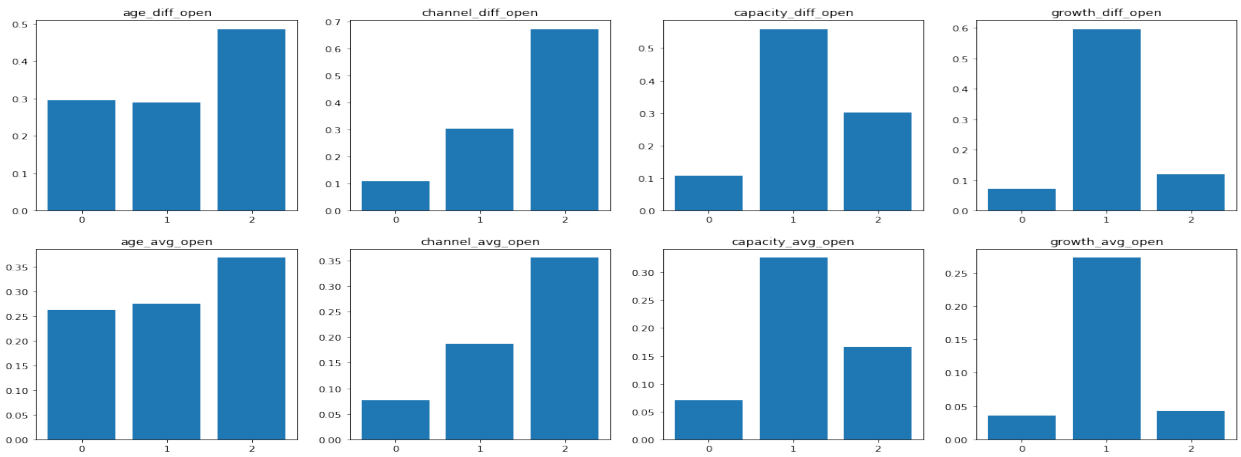

Figure 7: Clustering Analysis: Differences and Averages in Node-level Metrics Between n_i and n_j of a Channel for the Three Decision Cluster Types

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             efficiency   R-squared:                       0.025
Model:                            OLS   Adj. R-squared:                  0.023
Method:                 Least Squares   F-statistic:                     15.54
Date:                Mon, 11 May 2020   Prob (F-statistic):           1.09e-17
Time:                        19:16:18   Log-Likelihood:                 9230.7
No. Observations:                3651   AIC:                        -1.845e+04
Df Residuals:                    3644   BIC:                        -1.840e+04
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                      0.3152      0.001    562.918      0.000       0.314       0.316
open_block_cluster_label_0    8.45e-05   2.81e-05      3.008      0.003    2.94e-05       0.000
open_block_cluster_label_1   6.808e-05   2.09e-05      3.252      0.001     2.7e-05       0.000
open_block_cluster_label_2      0.0005   8.44e-05      5.429      0.000       0.000       0.001
close_block_cluster_label_0     0.0001   2.26e-05      5.043      0.000    6.97e-05       0.000
close_block_cluster_label_1 -8.243e-07   1.49e-05     -0.055      0.956   -2.99e-05    2.83e-05
close_block_cluster_label_2 -2.346e-05    2.2e-05     -1.064      0.287   -6.67e-05    1.98e-05
==============================================================================
Omnibus:                     1581.754   Durbin-Watson:                   0.055
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11311.812
Skew:                          -1.912   Prob(JB):                         0.00
Kurtosis:                      10.729   Cond. No.                         45.0
==============================================================================
```

Figure 8: Linear Regression Table: Channel Types on Network Efficiency

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              robustness   R-squared:                       0.090
Model:                             OLS   Adj. R-squared:                  0.089
Method:                  Least Squares   F-statistic:                     60.28
Date:                 Mon, 11 May 2020   Prob (F-statistic):           1.82e-71
Time:                         19:20:26   Log-Likelihood:                 5484.3
No. Observations:                 3651   AIC:                        -1.095e+04
Df Residuals:                     3644   BIC:                        -1.091e+04
Df Model:                            6
Covariance Type:             nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                      0.5358      0.002    342.949      0.000       0.533       0.539
open_block_cluster_label_0     0.0005   7.84e-05      6.282      0.000       0.000       0.001
open_block_cluster_label_1     0.0001   5.84e-05      2.499      0.012    3.15e-05       0.000
open_block_cluster_label_2     0.0002      0.000      1.014      0.311      -0.000       0.001
close_block_cluster_label_0   -0.0010   6.31e-05    -16.075      0.000      -0.001      -0.001
close_block_cluster_label_1    0.0001   4.14e-05      3.244      0.001    5.32e-05       0.000
close_block_cluster_label_2    0.0002   6.15e-05      4.031      0.000       0.000       0.000
==============================================================================
Omnibus:                      551.928   Durbin-Watson:                   0.168
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             7917.584
Skew:                          -0.196   Prob(JB):                         0.00
Kurtosis:                      10.204   Cond. No.                         45.0
==============================================================================
```

Figure 9: Linear Regression Table: Channel Types on Network Robustness

and efficiency in section 4.2.2.

3) Phase III - "the post-golden age" This phase corresponds to November 2018 to May 2019 and exhibits the expected trade-off between efficiency and robustness where efficiency is rising at the expense of the network robustness. The most prevalent type of connections during their period is between "the old money and the 99%", which corroborates the effect of this type of connection on the network as reported in the previous subsection.

4) Phase IV - "the Dark Age" Phase IV corresponds to May 2019 to February 2020. During phase IV, the efficiency of the network has leveled off and is slowly decreasing as the robustness of the network slips simultaneously. The most prevalent connection type during this period is "the new rich" and the 99%. Our interpretation of this finding is that even though this connection didn't reveal any statistically significance with network metrics, it can be having an indirect impact by crowding out other potentially beneficial connections during this phase (see Figure 11).

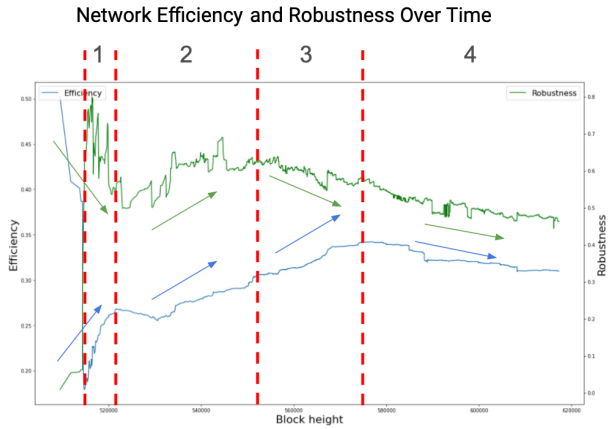

Figure 10: Network Efficiency and Robustness Over Time

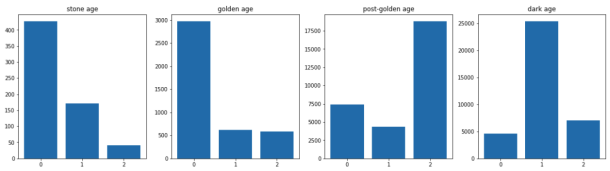

Figure 11: Distribution of Connection Types in Different Phases of the Network Evolution

4.2.4 Implications of Node Operators' Actions on the Pairs of Nodes Involved

Implications of Channel Openings

Opening a channel among all connection types is positively correlated with closeness for the two nodes in the channel (see Figure 12,) the 99 percenters-new rich and 99 percenters-old money payment channels have a larger positive impact on closeness than the 99 percenter-99 percenter connections. We found that the connections among the "99 percenters" are positively correlated with increasing betweenness for the individuals involved in those pair-wise payment channels (see Figure 13). Higher betweeness indicates that nodes that engage in these types of connections increase the chances of getting payments routed through them and the possibility of earning routing fees. We also found that both the 99 percenters - the new rich and the 99 percenters-the old money payment channels are positively correlated with closeness for the involved participants. These two types of connections do not have statistically significant relationship with the betweenness of the nodes involved (see Figure 13). A third dimension in the individual benefit that was salient from our qualitative analysis, that is the gratification from social connections with other node operators and the intrinsic gratification from helping the network grow. However, this dimension was not able to explicitly manifest in the quantitative analysis.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            y_closeness   R-squared:                       0.046
Model:                            OLS   Adj. R-squared:                  0.046
Method:                 Least Squares   F-statistic:                     1749.
Date:                Tue, 12 May 2020   Prob (F-statistic):               0.00
Time:                        00:17:49   Log-Likelihood:                 77151.
No. Observations:               72445   AIC:                        -1.543e+05
Df Residuals:                   72442   BIC:                        -1.543e+05
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
C(cluster)[0]     0.0364      0.000     95.322      0.000       0.036       0.037
C(cluster)[1]     0.0458      0.001     61.708      0.000       0.044       0.047
C(cluster)[2]     0.0866      0.001    114.327      0.000       0.085       0.088
==============================================================================
Omnibus:                    20960.271   Durbin-Watson:                   1.041
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            62905.970
Skew:                           1.508   Prob(JB):                         0.00
Kurtosis:                       6.426   Cond. No.                         1.98
==============================================================================
```

Figure 12: Linear Regression Table: Change in Pairwise Node Closeness regressed against Channel Openings by Connection Types

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           y_betweeness   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     3.880
Date:                Tue, 12 May 2020   Prob (F-statistic):             0.0207
Time:                        00:23:38   Log-Likelihood:                 85970.
No. Observations:               72445   AIC:                        -1.719e+05
Df Residuals:                   72442   BIC:                        -1.719e+05
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
C(cluster)[0]     0.0019      0.000      5.580      0.000       0.001       0.003
C(cluster)[1] -8.188e-06      0.001     -0.012      0.990      -0.001       0.001
C(cluster)[2]     0.0007      0.001      1.029      0.303      -0.001       0.002
==============================================================================
Omnibus:                     8252.579   Durbin-Watson:                   1.746
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            83525.383
Skew:                           0.020   Prob(JB):                         0.00
Kurtosis:                       8.260   Cond. No.                         1.98
==============================================================================
```

Figure 13: Linear Regression Table: Change in Pairwise Node Betweenness regressed against Channel Openings by Connection Types

Implications of Channel Closings

Closing a channel negatively correlates with the betweenness and closeness of the pair of nodes involved

for all of the three connection types. Surprisingly, it is worse to close a 99%-99% channel than a 99%-"the-new-rich" channel, confirming our interpretation that a 99%-"the-new-rich" connection adds little value to both nodes, and removing such a connection does not significantly reduce the connectedness and betweenness for either. The most significant impact in absolute term is from channel closures between a 99%-"an old-money" (see Figure 14, 15.)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       closey_closeness   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.006
Method:                 Least Squares   F-statistic:                     125.3
Date:                Sat, 16 May 2020   Prob (F-statistic):           5.69e-55
Time:                        12:33:27   Log-Likelihood:                 19721.
No. Observations:               38341   AIC:                        -3.944e+04
Df Residuals:                   38338   BIC:                        -3.941e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
C(close_block_cluster_label)[0]    -0.1083      0.001   -117.832      0.000      -0.110      -0.106
C(close_block_cluster_label)[1]    -0.0845      0.002    -49.703      0.000      -0.088      -0.081
C(close_block_cluster_label)[2]    -0.1228      0.002    -67.431      0.000      -0.126      -0.119
==============================================================================
Omnibus:                     6082.073   Durbin-Watson:                   1.682
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9409.667
Skew:                          -1.191   Prob(JB):                         0.00
Kurtosis:                       3.465   Cond. No.                         1.98
==============================================================================
```

Figure 14: Linear Regression Table: Change in Pairwise Node Closeness regressed against Channel Closings by Connection Types

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     closey_betweenness   R-squared:                       0.012
Model:                            OLS   Adj. R-squared:                  0.012
Method:                 Least Squares   F-statistic:                     229.0
Date:                Sat, 16 May 2020   Prob (F-statistic):           1.39e-99
Time:                        12:35:12   Log-Likelihood:                 45652.
No. Observations:               38341   AIC:                        -9.130e+04
Df Residuals:                   38338   BIC:                        -9.127e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
C(close_block_cluster_label)[0]    -0.0094      0.000    -20.082      0.000      -0.010      -0.008
C(close_block_cluster_label)[1]    -0.0027      0.001     -3.138      0.002      -0.004      -0.001
C(close_block_cluster_label)[2]    -0.0284      0.001    -30.707      0.000      -0.030      -0.027
==============================================================================
Omnibus:                     3526.440   Durbin-Watson:                   1.870
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            24677.099
Skew:                           0.110   Prob(JB):                         0.00
Kurtosis:                       6.924   Cond. No.                         1.98
==============================================================================
```

Figure 15: Linear Regression Table: Change in Pairwise Node Betweenness regressed against Channel Closings by Connection Types

4.2.5 Implications of Connection Types and Channel Age on Channel Status

The probability of a channel being open seems to be affected by the age of the channel. Old channels are more likely to close than new ones. Belonging to 99%-the new rich connection has the highest probability for channel closure, followed by 99%-99% and 99%-the old money. 99%-the old money is often the "lifeline" for 99% because they otherwise loose a lot of connectedness to the rest of the network. The 99%-the new rich connection seem to be more "fragile" than the other types, even the 99%-99% , which is between smaller less well connected participants (see Figure '16.) Insights from the qualitative analysis help us interpret that probably the interpersonal social pressure hold the 99%-99% connections together more strongly than expected by just looking at their structural properties.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                is_open   No. Observations:                72454
Model:                          Logit   Df Residuals:                    72450
Method:                           MLE   Df Model:                            3
Date:                Sat, 16 May 2020   Pseudo R-squ.:                  0.1250
Time:                        23:53:19   Log-Likelihood:                -43835.
converged:                       True   LL-Null:                       -50096.
Covariance Type:            nonrobust   LLR p-value:                     0.000
==============================================================================================
                                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
C(cluster_label)[0.0]     -1.1886      0.015    -81.451      0.000      -1.217      -1.160
C(cluster_label)[1.0]     -1.4895      0.023    -64.617      0.000      -1.535      -1.444
C(cluster_label)[2.0]     -1.0866      0.022    -49.745      0.000      -1.129      -1.044
duration                   4.6723      0.047    100.157      0.000       4.581       4.764
==============================================================================
```

Figure 16: Logistic Regression Table: Connection Types and Channel Age on Channel Status

5 Discussion

Contrary to our hypothesis that the connections between two 99-percenters are conducive to network-level robustness but are unfavorable to network-level efficiency, we found that payment channels between the 99-percenters are not only highly positively correlated with network-level robustness, but also positively correlated with network-level efficiency, albeit the positive correlation coefficient is smaller compared to the 99-percenter and the new rich or the 99-percenter and the old money connections. Connections between 99-percenters also positively correlate with node-level closeness and betweenness for the two nodes involved in the channels. In addition, the intrinsic gratification of social connections to individual participants is most pronounced among the 99 percenters. This dimension of individual benefits was not able to manifest through the quantitative analysis, but we extracted this salient dimension from the qualitative analysis. Promoting the connections between the 99-percenters in the network ben-

efits both the network and the individuals involved in all aspects to different extent. However, the positive correlations of the other two types of connections should not be dismissed. As discussed in the results section, the other two connections types have higher correlation with the connectedness for the individuals than the connections between the 99-percenters, despite a negative correlation of the 99-percenters and the-old-money connections on the robustness of the network. Designers of both the protocol and software clients should consider incentivizing the various combinations of connection types in proportion of their benefit to push the network in the right direction.

5.1 Design Recommendations

How can network architects adjust the different levers at their disposal to promote the optimal combination of connection types through the software client or network protocol? Both the information provided through the software client and the rules defined by the network protocol play significant roles in node operators' channel opening decision process.

Because Lightning Network channel creations and closings happen through software clients, the information provided through them is a major source of information in addition to external sources in helping the node operators evaluate a potential connection partners. Currently, the information provided through the software client limits to business/financial-interest oriented information such as the number of channels a potential partner has. Smaller nodes are at a disadvantage in successfully establishing connections if they do not have the social or personal connections with potential partners through auxiliary platforms such as the Lightning Network Telegram. For network software developers to promote the connections between the 99-percenters that we identified as beneficial to the network in all measures, the developers need to invoke the intrinsic motivations that were most prominent among this group of users. For example, developers can provide information through the software client about the overall benefit a potential connection can generate to the network or to the potential channel partner, users will more likely consider creating a channel due to the gratification of helping the network grow.

6 Conclusion

We studied the Bitcoin's Lightning network using a mix of quantitative and qualitative methods. We uncovered a diverse set of motivations that drive users' channel opening and closing decisions, which are also determined by the information available to the users through the software client, third-party Lightning Network search engines and direct communications between users through online communities. In addition, we found three major connection types in the lightning network and analyzed their effects on both the network and the network participants. Our results inform the developers of the network's protocol and software clients of the types of connections to incentivize. For future work, we intend to identify the most optimal combinations of connections and to translate our findings into concrete software feature recommendations.

References

[1] Lightning network search and analysis engine.

[2] Global cryptocurrency market cap, Mar 2020.

[3] Lightning network online database - big sun, Feb 2020.

[4] Lightning status mar 2020, Mar 2020.

[5] Georgia Avarikioti, Gerrit Janssen, Yuyi Wang, and Roger Wattenhofer. Payment network design with fees. In Joaquin Garcia-Alfaro, Jordi Herrera-Joancomartí, Giovanni Livraga, and Ruben Rios, editors, *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 76–84, Cham, 2018. Springer International Publishing.

[6] Georgia Avarikioti, Rolf Scheuner, and Roger Wattenhofer. Payment networks as creation games. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 195–210. Springer, 2019.

[7] AWS Documentation. *Amazon ECS on AWS Fargate*, 2020.

[8] AWS Documentation. *Boto AWS SDK for Python*, 2020.

[9] Will Badr. Serverless distributed data preprocessing using dask, amazon ecs and python (part 1), Jan 2019.

[10] Silvia Bartolucci, Fabio Caccioli, and Pierpaolo Vivo. A percolation model for the emergence of the bitcoin lightning network. *Available at SSRN 3500229*, 2019.

[11] Ferenc Béres, Istvan Andras Seres, and András A Benczúr. A cryptoeconomic traffic analysis of bitcoins lightning network. *arXiv preprint arXiv:1911.09432*, 2019.

[12] Louis Bertucci. Incentives on the lightning network: A blockchain-based payment network. *Available at SSRN 3540581*, 2020.

[13] Ulrik Brandes and Daniel Fleischer. Centrality measures based on current flow. In *Annual symposium on theoretical aspects of computer science*, pages 533–544. Springer, 2005.

[14] M. Conti, E. Sandeep Kumar, C. Lal, and S. Ruj. A survey on security and privacy issues of bitcoin. *IEEE Communications Surveys Tutorials*, 20(4):3416–3452, Fourthquarter 2018.

[15] Karen Cook, Coye Cheshire, and Alexandra Gerbasi. *Sociological Perspectives on Social Psychology.* 2018.

[16] Dask Development Team. *Dask: Library for dynamic task scheduling*, 2016.

[17] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H Papadimitriou, and Scott Shenker. On a network creation game. In *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pages 347–351, 2003.

[18] Debra Friedman and Michael Hechter. The contribution of rational choice theory to macrosociological research. *Sociological theory*, pages 201–218, 1988.

[19] Mark Granovetter. The impact of social structure on economic outcomes. *The Journal of Economic Perspectives*, 19(1):33–50, 2005.

[20] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[21] Matthew O. Jackson. *Game-Theoretic Modeling of Network Formation*, pages 371–410. Princeton University Press, 2008.

[22] Matthew O. Jackson. *Strategic Network Formation*, pages 153–182. Princeton University Press, 2008.

[23] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

[24] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

[25] Lightningnetwork. lightningnetwork/lightning-rfc, May 2020.

[26] Jing Liu, Mingxing Zhou, Shuai Wang, and Penghui Liu. A comparative study of network robustness measures. *Frontiers of Computer Science*, 11(4):568–584, 2017.

[27] Vitor HP Louzada, Fabio Daolio, Hans J Herrmann, and Marco Tomassini. Generating robust and efficient networks under targeted attacks. In *Propagation Phenomena in Real World Networks*, pages 215–224. Springer, 2015.

[28] Stefano Martinazzi and Andrea Flori. The evolving topology of the lightning network: Centralization, efficiency, robustness, synchronization, and anonymity. *PLOS ONE*, 15(1):1–18, 01 2020.

[29] Linda Molm and Karen Cook. *Sociological Perspectives on Social Psychology.* 1995.

[30] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.

[31] The pandas development team. pandas-dev/pandas: Pandas, Feb. 2020.

[32] F. Perez and B. E. Granger. Ipython: A system for interactive scientific computing. *Computing in Science Engineering*, 9(3):21–29, 2007.

[33] Joseph Poon and Thaddeus Dryja. The bitcoin lightning network: Scalable off-chain instant payments, 2016.

[34] BitMEX Research. Bitmex research, May 2020.

[35] Elias Rohrer, Julian Malliaris, and Florian Tschorsch. Discharged payment channels: Quantifying the lightning network's resilience to topology-based attacks. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 347–356. IEEE, 2019.

[36] Akrati Saxena, Ralucca Gera, and SRS Iyengar. A faster method to estimate closeness centrality ranking. *arXiv preprint arXiv:1706.02083*, 2017.

[37] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[38] István András Seres, László Gulyás, Dániel A Nagy, and Péter Burcsi. Topological analysis of bitcoin's lightning network. In *Mathematical Research for Blockchain Economy*, pages 1–12. Springer, 2020.

[39] David A Snow and Pamela E Oliver. Social movements and collective behavior: Social psychological dimensions and considerations. s. 571–599. *Karen Cook, Gary Fine und James House (Hg.): Sociological Perspectives on Social Psychology. Needham Heights, MA: Allyn and Bacon*, 1995.

[40] Eva Tardos and Tom Wexler. Network formation games and the potential function method. *Algorithmic Game Theory*, pages 487–516, 2007.

[41] S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, 2011.

April 2019 [revised]June 2019 [accepted]August 2019