INSTITUTO SUPERIOR TÉCNICO

COMPLEX NETWORKS

2nd Project

Students: *Numbers:*

David Calhas 80980
João Valente 81759

Taught by:
Francisco Correia dos Santos
Alexandre Francisco

2 December 2018

Abstract

The lightning protocol works on top of a scale-free network whose properties are neither well understood nor well defined by the protocol. An analysis of the network, giving special attention to the distribution of the liquidity in present in the state channels, is made. Using a cumulative probabilistic metric whose theoretical ceiling can be defined, one can locate the liquidity of the network and understand its evolution, finding optimizations that increase the usability of the protocol.

1 Introduction

The Lightning protocol [1] defines a distributed system built on top of the bitcoin network and whose goal is to facilitate the transfer of value from A to B.
The network can be represented as a weighted undirected graph, with the edge weights representing transaction volume capacity.
There's a rather important subtlety that won't be taken into account for the sake of simplicity. In the lightning protocol, edges between nodes (also called channels) have an associated state that represents how the committed funds (capacity) are distributed between nodes within a channel. This state dictates the ownership of the funds and can be thought of as beads on an abacus, much like what's represented by figure 1.

Figure 1: Channel with a capacity of 4 and state (Alice: 3, Bob: 1)

The corresponding channel state information is only available to the nodes in the end of each channel but the capacity of the channels is publicly available. This study will consider that the state of a channel between A and B with capacity C is (A: $\frac{C}{2}$, B: $\frac{C}{2}$). In order to simplify the problem channels between the same nodes were merged and their capacities summed.

It is possible to extend the model represented by figure 1 to route payments between nodes connected through a path of in-between channels. Figure 2 exemplifies this type of payment paths.

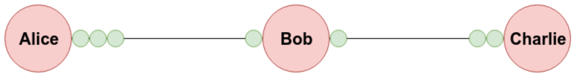

Figure 2: Alice and Charlie are connected through Bob

In this case the initial state of the network can be represented as {(Alice: 3, Bob: 1), (Bob: 1, Charlie: 2)}. If Alice wanted to send 1 BTC to Charlie the state would become: {(Alice: 2, Bob: 2), (Bob: 0, Charlie: 3)}, as represented by figure 3.

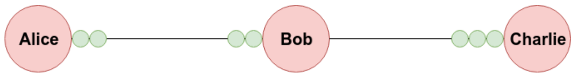

Figure 3: Alice sends 1 BTC to Charlie

This paper will start by analyzing the underlying structure of the network in an attempt to understand how it can shape the functioning of the system it supports. This will be followed by the introduction of a measurement that will help us study the liquidity of the funds in network. This measurement will be used to understand resilience; how the removal of crucial nodes can impact the network and it's usability.
In the end, an analysis of the evolution of the network is made and some proposals for its optimization are discussed.
The analyzed graph in this work corresponds to a snapshot of the lightning network taken using a lightning network client API [2] on October 26th, 2018.

2 Network Structure

2.1 Scale

At the start of this study (October 26th, 2018) the network had 1642 nodes and 8506 edges.

2.2 Degree Distribution

There is a big discussion around the subject of how the lightning network should be structured. It could be wise to try and develop a structured mesh network, since this kind of topology would reduce the existence of hubs, which in turn would increase the resilience and distribution of the network.
One could also argue that the network would benefit from having a scale-free structure so that liquidity would be more concentrated around big hubs and the payment paths shortened.
This section of the paper studies the current structure of the network, starting by plotting the degree distribution present in figure 4.

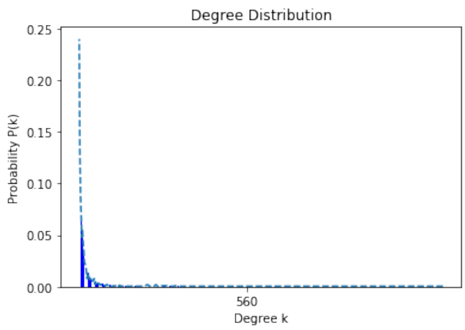

Figure 4: Degree distribution

Using the data plotted in figure 4 one can obtain γ (1) and plot the cumulative degree distribution presented in figure 5.

$$\gamma \approx 2.144 \tag{1}$$

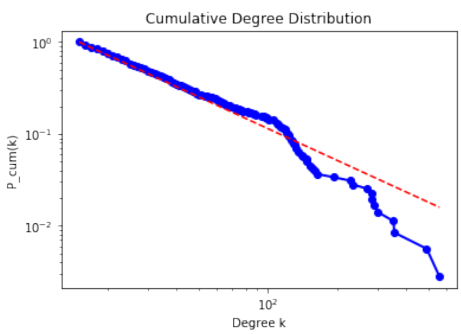

Figure 5: Cumulative Degree Distribution (log/log) scale, obtained from the package PyPl - Python

The power law fitting the cummulative degree distribution is very accurate for nodes with a small degree, however it presents a high degree cutoff as can be seen on the right side of plot 5, this can be explained by the finite size of the network [3].

2.3 Betweenness Centrality

It is also important, for the context of the problem, to understand how the degree of a node relates to its centrality, since the network's goal will be transferring value through payment paths, identifying the nodes that appear the most in these paths (and if these nodes can be classified as hubs) can be crucial.

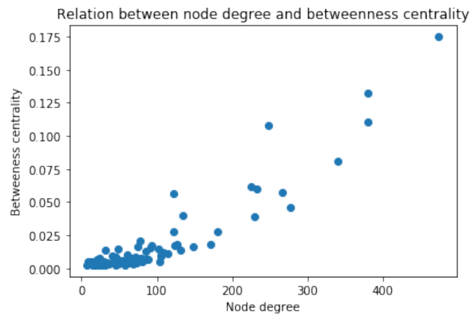

Figure 6: Relation between betweenness centrality and degree

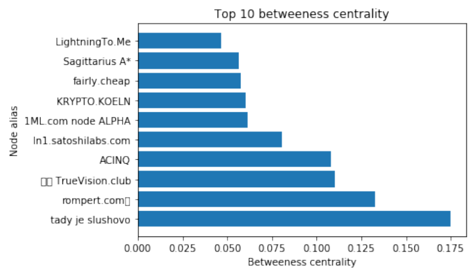

Figure 7: Top 10 nodes by betweenness centrality

According to the plot in figure 6 one can identify that there's a linear relation between a node's degree and its betweeness centrality. As pictured by figure 7 the most central nodes to the network are not private users but instead specialized routing services.
Using this two conclusions it can be said that

the network relies heavily on hubs for its current functioning. In section 4, the paper analyzes this problem from a liquidity-concerned point of view.

3 Flow Study

In this section, the paper introduces a metric that will help study the capacity of the network to make transactions as a function of the transaction volume. The main idea behind the computation of this metric is to find the bottleneck in a certain path, and associate the path with this bottleneck, which can be thought of as the maximum path flow.

3.1 Procedure

The ideal solution for the computation of this metric would be to analyze all the possible paths from one node s, to another node t, store in memory their respective maximum flows/ bottlenecks and compute how many of this bottlenecks are larger than certain predetermined payment volumes. Using this technique, one could plot the probability of the payment being successful for each predetermined payment volume. In section 3.2 a relevant approximation for this solution is presented.

3.2 Shortest Paths Approach

Due to the high number of paths in the network, a brute force approach is not feasible in reasonable time.

Since the technical standard of the protocol specifies that the first payment path that should be tried by a sender is the shortest path between the sender and the receiver, it makes sense to only take into account the shortest paths between nodes instead of the total number of paths. This means that taking only the shortest paths into consideration will yield an exact result if considering the first payment attempt and an approximation if considering every possible payment attempt.

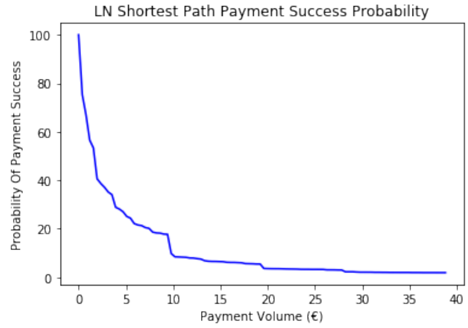

Figure 8: Payment success probability for shortest paths

Figure 8 unveils the result of this approach.

3.3 Upper Bound

It's also interesting to understand that, for a given network, there's an upper bound for the payment success probability metric that can be useful to compare with the actual payment success probability. To reach this upper bound one can take the idea presented in section 3 and imagine that payments from a node s, to another node t, can not be made through one path exclusively but instead can be made through multiple $s-t$ paths. This makes the computation of the upper bound of $s-t$, the same as the computation of the maximum flow between a source s and a sink t.

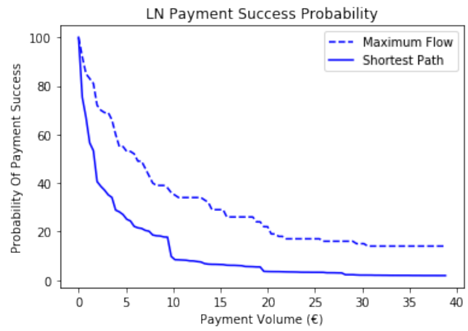

Figure 9: Payment success probability for shortest paths and its upper bound

A comparison between the shortest path payment success probability and its upper bound was plotted in figure 9. Using equation 2 it's possible to reach the network's used capacity, which for this case is 38.48%.

$$UC = \frac{\int_0^\infty shortestPath(V)\ dV}{\int_0^\infty maximumFlow(V)\ dV} \quad (2)$$

4 Liquidity

In this section an analysis of the network's liquidity is made.
Network liquidity is defined as the sum of the capacities of all edges. The main goal is to find out how this liquidity is distributed within the network, as well as to see how this distribution affects the probability of payment success defined in 3.

4.1 Procedure

In order to do this analysis, the top N degree nodes were removed from the original graph. The resulting components, whose number of nodes were larger than C, were analyzed. After this, a maximum flow analysis similar to the one in section 3.3 was made.
The results were then partitioned into four cases, corresponding to $N \in \{0, 40, 80, 120\}$ with $C = 50$.

4.2 Results

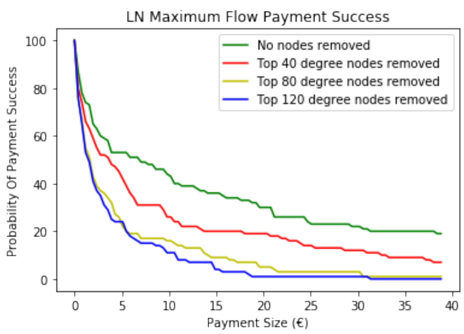

Figure 10: Liquidity plot with high degree node removal.

Figure 10 is the end result of the procedure described in section 4.1. One can immediately notice the following: if there's a version X of the original graph that has more hubs removed than another version Y, then the plot that corresponds to X is always be bounded by Y. All the lines are bounded by the original version of the graph which has nodes removed.
It is also interesting to understand how the length of the payment path changes with the removal of the high degree nodes. Table 1 presents this relation.

Nodes removed	ASPL
no nodes removed	2.90
top 40	4.13
top 80	5.12
top 120	6.24

Table 1: Resulting average shortest path length (ASPL) after nodes removal.

It was found that when nodes with a high degree are removed, the average shortest path length increases.
In summary, the removal of high degree nodes from the original graph impacts the average shortest path length, as well as the liquidity availability of the lightning network.

5 Evolution

The studies realized in this work use a snapshot of the lightning network downloaded in the 26th October, 2018.
In order to see what the future holds, it would be insightful to understand how the lightning network is evolving.

To study this evolution, a snapshot of the network was again taken and downloaded one month later (26th November, 2018).

Immediately, a growth in the number of nodes and edges was observed. The old network had 1642 nodes and 8506 edges, the last downloaded network had 1877 nodes and 12386 edges, a growth of 14.31% for the number of nodes and of 45.61% for the number of edges.
The increase in the number of edges is quite interesting, it may be due to the necessity of making more connections to increase transaction reliability, or perhaps the need to increase potential transactions values by making new connections with higher capacity.

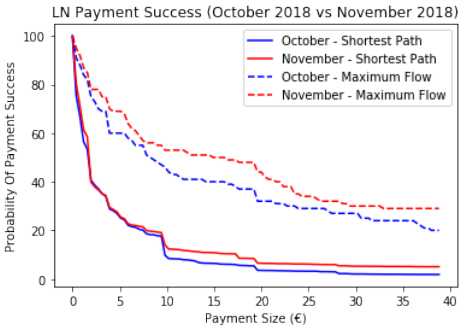

Figure 11: Payment success comparison between the same network, one month apart.

In figure 11, the plots of the payment success probabilities for the shortest path and maximum flow, in both moments in time, are shown together. The red lines (the lines that refer to the network from November) are bounding the blue lines for the biggest flow and maximum flow of the shortest paths, respectively. Table 2 presents the evolution of the network's used capacity.

Network	Used Capacity
26th October	38.48%
26th November	34.42%

Table 2: Used capacity of the network.

The difference between the old and new shortest path threshold capacities was also analyzed. The values obtained are in table 3.

Network	Mean	SD
26th October	5.70€	19.46€
26th November	11.61€	39.71€

Table 3: Mean and standard deviation of the threshold capacity of the edges.

It can be seen that the mean and the standard deviation increased 103.33% and 104.06%, respectively.

6 Discussion

Results from section 2, more specifically equation 1, point to the fact that the network fits into the class of scale-free networks since $\gamma \in [2,3]$. This is just another example of how real life networks have this kind of properties, which should be taken deeply into account when studying them. Results from section 2.3, where the top betweeness centrality nodes were listed, point to the fact that important nodes for the network are controlled by specialized services that profit from having a a central location in the network. This type of enterprise hubs seem prevail in the network.

Section 3, more specifically figure 8, shows how lighting is limited by its liquidity. The network works well for payments whose volume does not exceed 2€ to 3€, failing for larger amounts. The plot in figure 9 raises an important concern about the lack of payment optimization of the network. The shortest path payment approach that is currently in use by the protocol makes it difficult for the liquidity available throughout the network to be fully used for payments. According to table 2, this problem seems to aggravate as the network grows, making the insertion of new channels and their provided liquidity a suboptimal solution for increasing the usability of the protocol.

The usage of the shortest path approach presented in section 3.2 is debatable. Since the protocol starts by trying the shortest payment path first, the results from this approach end up representing the probability of the payment being successful on the first try. To understand how the probability of success for a random payment path between s and t, a count the number of $s-t$ paths in a graph would be needed, this is a $\#P$-complete problem as shown by [4]. To solve this problem an estimation of the number of $s-t$ paths using the method described in [5] was tested. Unfortunately, the results were not the expected and the approximations diverged very easily, showing big differences between simulations. The PathDistribution Julia package [6] was also tested, having the same the difficulties as the previous method. One can justify this with the fact that this method requires a lot of iterations in order for the results to be a good approximation of the real values. Another reason for the poor estimation of the paths maybe the fact that the scale of the graph is too big for the algorithm. In section 4.2, it was concluded that the removal of large degree nodes increased the ASPL as seen in table 1. This is coherent with the results from section 2.3,

which showed that there was a linear relation between node degree also and betweeness centrality.
Figure 10 plays an important part in understanding how the liquidity is distributed within the network. The sharp drop in the payment success probability when high degree nodes were removed proves that most of the liquidity is available through the hubs, making them a pivotal point for the success of the network's goals and a centralized point of failure which, in the future, could harm lightning's resilience.
Table 3 forecasts a bright future for the evolution of the network, which had a high jump in its average shortest path threshold capacity, making bigger volume transactions more likely to be successful.

7 Future Work

The analysis contained in this work does not emulate the real behavior of the network in the sense that it doesn't model the consequences of the movement of coins within the same channel. Although this information is inherently not public, one could study the behaviour of the network when funds move within channels by locally simulating transactions. There's also a lot that could be done in order to use lightning in a more efficient way. This could mean studying routing solutions that differ from the current shortest path approach or even consider multi-path payments.
Improving the network's resilience, by studying ways of decrease the dependence on the hubs, is also a possible area of interest.
The evolution of the network was analyzed, but only for one month interval time once. In the future, one could take more snapshots of the network and try to model its growth with the objective of predicting how the network will behave in the future.

8 Conclusion

The analysis of the network concluded that lightning is a scale-free network, whose proposition to transfer value from A to B is fulfilled for low volume payments. There are big optimization possibilities for this technology and liquidity could see a drastic improvement should this optimizations be implemented.
Hubs are very important to the well-being of the network, making it vulnerable to failures if there's a general hub malfunctioning or a coordinated attack.
After studying the evolution of the network one can understand that there is real interest in this technology and that it could become a viable implementation for an alternative to a VISA-like payment system.

References

[1] Joseph Poon and Thaddeus Dryja. The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments.

[2] LND API describe graph. `https://api.lightning.community/#describegraph`. Accessed: 2018-12-04.

[3] M Boguñá, Romualdo Pastor-Satorras, and Alessandro Vespignani. Cut-offs and finite size effects in scale-free networks. *European Physical Journal B*, 38:205–209, 03 2004.

[4] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Comput.*, 8:410–421, 1979.

[5] Ben Roberts and Dirk P Kroese. Estimating the number of st paths in a graph. *J. Graph Algorithms Appl.*, 11(1):195–214, 2007.

[6] Pathdistribution julia package. `https://github.com/chkwon/PathDistribution.jl`. Accessed: 2018-12-06.