Master 2 Research - Interaction Specialty

Bitcoin Visualization

Author:
Loïs Saublet

Supervisor:
Petra Isenberg
Tobias Isenberg

Hosting lab:
Aviz, Inria, Saclay

March 2nd – August 28th
2015

Secrétariat - tel: 01 69 15 66 36 Fax: 01 69 15 42 72
email: Murielle.Benard@u-psud.fr

Contents

Summary

In this thesis, I propose a visual analytics tool enabling the end users (they are economists) to explore the data produced by Bitcoin. Bitcoin is a new payment system, decentralized from any authority and is free and open-source. Bitcoin was created in 2009 and has become more and more used in the past two years, this is evident by the fact that one can now pay using bitcoins in more and more locations from bars to universities. This visual analytics tool's metrics are extracted from requirements given by the end users. This thesis also provides an evaluation of this tool. I designed this visual analytics tool for economists who want to determine if Bitcoin behaves and is used as a currency or a commodity. My thesis forms part of a one year long exploration project on the development of Bitcoin. This tool is designed to quickly select an interesting period of time (according to a certain metric) and then compare other metrics at the same period. This tool is composed of a web based front end implemented using Javascript, HTML/CSS and a javascript library called D3.js; and a back end which uses a database set up as a data warehouse.

Keywords

Bitcoin, Visual analytics, Interactive

CHAPTER 1

Introduction

The research presented in this thesis lies in the field of visual analytics. Visual analytics uses interactive visual interfaces to help analytical reasoning. My thesis' goal is to design and implement a visual analytics system of the Bitcoin payment system. My focus is on developing visualizations as an interface to understanding the data. Bitcoin is a cryptocurrency and information on all the transactions made with this currency are publicly available. The use of metrics defined by our expert end users is my thesis' main contribution.

1.1 Motivation

This master thesis is about making a visual analytics tool for Bitcoin for economists to be able to evaluate this new payment system. Bitcoin is a new medium of payment that is becoming more powerful and widely used (the number of transactions per day went from around 5,000 at the beginning of 2012 to around 100,000 since the beginning of 2015), but stays obscure for most people. This is why tools to analyze the transactional data are needed.

The Bitcoin payment system

Bitcoin is a decentralized cryptocurrency based on a peer-to-peer network. It is decentralized because there is no main entity regulating it. As a cryptocurrency, it is a currency using cryptography to secure and control its transactions. Peers use the Bitcoin online payment system to exchange money between them, trusting the system and all its peers to validate these transactions. The Bitcoin protocol was first defined in the original paper [9] written by Satoshi Nakamoto - an anonymous person or group of people. It was launched on January the 3rd, 2009 with the creation of the first Bitcoin.

This new payment system is highly interesting from the point of view of the economists we are working with according to several points. First of all, all the transactions are public. This is not the case for any centralized currency. The public data enables the exploration of this new payment system. Then, this data represents the evolution of the system since its beginning, which is also unique and is of great interest for our end users. Finally, Bitcoin is becoming adopted more widely every day. From shops and bars [7] to universities [6], Bitcoin is becoming a valuable medium of trade. From 2009, the number of transactions per day made with Bitcoin exploded. It was

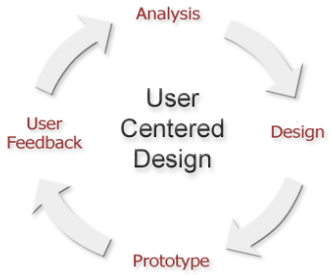


Figure 1.1: Design Life-cycle methodology

oscillating around 100 during the first year to more than 100,000 now with an all time record of 214,500 in July 2015.

1.2 Problem Statement

During my thesis I worked with end users who are economists. They have questions about Bitcoin as a new phenomenon and want to learn how it is used. They want to know if it is used like a currency or a commodity. The visual analytics tool will allow them to interactively explore the Bitcoin system and help them answer this question.
We[1] identified two main goals for this tool. Firstly it should enable the economists to visualize general metrics like the market price of Bitcoin. Secondly, it should give them an overview of the actors of the Bitcoin economy by visualizing its peers and their interactions. The main interest in making this tool is the ability to quickly compare changes and trends over time and being able to change and compare the different metrics as conveniently as possible.

1.3 Methodology

In order to conceive a visualization system for our end users, we followed the interaction design life-cycle seen in Figure 1.1. In close collaboration with another intern who was focused on the analytics and database side of the tool, we used this design method to build the tool. This design method is composed of four iterative steps : Analysis, Design, Prototype and User feedback. In the analysis phase, we first identified expert end users. Then we met with these experts several times in order to learn how a potential visual analytics system could help them, what was

[1]During my thesis, I worked with another intern who took care of the database and the computation of the data. His name is Kofi Manful and he is graduating from Ecole Centrale Paris. The pronoun we is used to refer to him and myself.

their knowledge about Bitcoin or how they currently worked with their tools. From our meeting we gathered the requirements needed for the visualization's first design. The design phase based itself on these requirements to design a useful visualization for our users. Then in the prototype phase we created a first prototype which was then tested by our end users. The whole process then starts again, giving birth to a new iteration of the prototype which eventually became our final visualization. Due to time limitations we were able to go through this cycle only once.

1.4 Contribution

This thesis contribution resides in three parts:

1. The design and development of the visual and interactive component of a visual analytics system to analyze metrics of the Bitcoin transaction graph.
2. Collection, assessment, and application of user requirements from expert economists.
3. The preliminary evaluation of the tool.

More details on these contributions can be found in the conclusion of this thesis.

1.5 Thesis Organization

The following section briefly introduces the contents of this thesis' chapters.

Chapter 2
This chapter covers the related work.

Chapter 3
In this chapter, I present the requirements gathering during the meetings we organized with the end users.

Chapter 4
In this fourth chapter, I describe the solution and summarize the design choices I made.

Chapter 5
This chapter includes a discussion of the extensibility and limitations of my interface. I also present the main feedback I collected.

Chapter 6
This chapter concludes the thesis, I summarize the contributions and present the further work this tool needs.

CHAPTER 2

Related Work

This chapter gives a brief introduction to the Bitcoin payment system and a state of the art of the visual analytics of Bitcoin, the Bitcoin's users clustering and their visualizations.

2.1 Bitcoin

As I explained in the introduction, Bitcoin was created by Satoshi Nakamoto in 2008 [9] and released as a software in January 2009. Bitcoin is a decentralized payment system. It relies on a public ledger called the blockchain, a database of transaction data shared by all users, to store and validate the transactions. A special type of users, the miners, are responsible of the transactions' validations and in exchange they claim a reward. The miners validate transactions by putting them in a block, finding the key to put this block in the blockchain and finally putting the block in the blockchain to spread it across the network. When a miner finds a block, he or she can create a genesis transaction at the beginning of the block, giving him or her a reward. This reward is 25 Bitcoins at the time of writing. These transactions are very important as they obviously encourage miners to mine (vital to the Bitcoin system) and enable the creation of new Bitcoins. This creation of new coins has to be controlled to avoid inflation and that is why the reward's amount is diminishing over time. It was originally 50 Bitcoins and is expected to drop to 12.5 Bitcoins in 2016, halving every 210,000 blocks [4].
Users store the bitcoins they own using addresses. A user can have as many addresses as he or she wants to and it is better for him or her to use a lot of addresses from the point of view of anonymity on the network. The only users that should not change their addresses are the merchants in order to simplify the way they are paid. Every Bitcoin transaction is registered in a public ledger. When a user pays with one of his or her addresses in a transaction, he or she must empty all the Bitcoins it contains—even if it is more than the value of the product or service he or she is buying. The change must be collected using another address.
The Bitcoins transactions are registered in a public, non-encrypted ledger. This means the data is available for visualization purposes. The interesting aspect of Bitcoin for the economists is that its ledger contains all the transactions—all the information—from the beginning of the currency until now. This is a really unique dataset for the economists to study.

Figure 2.1: An example of a chart from the website blockchain.info. The chart plots the number of transactions per day.

Figure 2.2: The visualization of the Bitstamp market price on bitcoinwisdom.com.

2.2 Visual analytics of Bitcoin

Two different ways to visualize the Bitcoin transaction graph exist. This transaction graph can be visualized using charts that plot the market price according to time for example; or using graphs to visualize the users and how they interact with each other. The first representation is meant to visualize financial metrics computed from the blockchain. These metrics can be the number of transactions, the transactions' amount, etc. In this section, I present existing solutions for both types of visualization.

Visualizing the metrics

The metrics I need to visualize are computed from the blockchain of Bitcoin. Several websites propose different web based visualizations for Bitcoin metrics. The website www.blockchain.info/charts gives access to multiple charts on multiple Bitcoin

metrics, as shown in Figure 2.1. These charts are divided on several web pages which is not optimal to compare them. There are also a handful of websites presenting visualizations on the price of Bitcoin on multiple markets in real time. One of them is interesting from the point of view of its interactivity, giving the user the opportunity of hovering the chart with its mouse. This interactivity provides the user the ability to read the exact values at different points in time. An example of these charts can be found at https://bitcoinwisdom.com/markets/bitstamp/btcusd, here visualizing the price of bitcoins on the Bitstamp exchange as shown in Figure 2.2. The only drawback of this website feature is that the hovering is not fixed on the Y values, meaning the user must be accurate to determine the precise Y value according to a certain X.

Visualizing the actors

The first thing to know about the Bitcoin protocol is that an actor in the network has as many addresses as he wants to. This means an actor cannot be identified by a single address. In a transaction, there are one or more addresses as input and one or more as output. Therefore the first step in visualizing the actors of the Bitcoin network is to find the groups of addresses that belong to the same actor. Identifying the actors is different from de-anonymizing the network. We are not trying to find the users (e.g. their personal information) but to be able to identify actors (or groups of addresses). In order to accomplish this, a paper described a set of two heuristics [8].

Heuristic 1 The first heuristic presented says that if a transaction has several inputs, all of them belong to the same user. This rather straight forward heuristic enables the researchers to extract around 5.6 million clusters of addresses belonging to the same user off the blockchain.

Heuristic 2 The second heuristic they used is focused on the change addresses. As I said in section 2.1, if a user uses an address for a payment, he or she must empty the address of all its Bitcoin. If the price he or she wants to pay is lower than this amount a change address - also know as shadow address - is created. This address is interesting because if we can identify it we can link it to the user who made the payment and therefore add one address to its cluster. A change address is defined as follows:

- This is the first appearance of this address in the whole blockchain.
- The transaction it belongs to is not a genesis transaction.
- The address is not in the inputs.
- This is the only address in the outputs that is being seen for the first time.

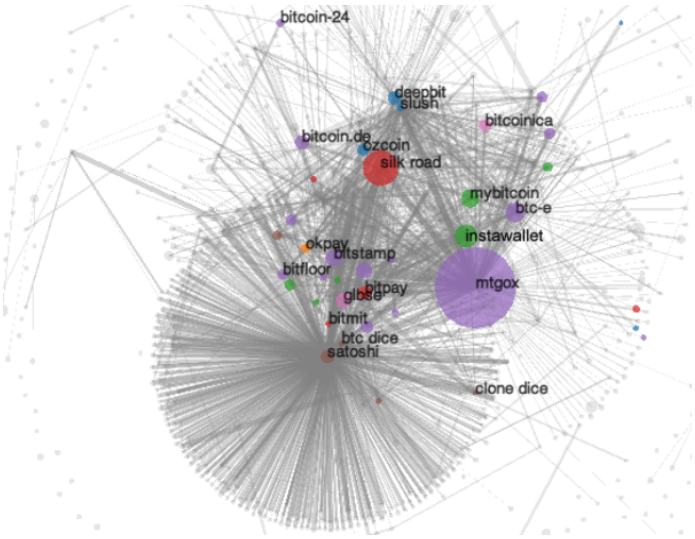


Figure 2.3: Biggest clusters in the Bitcoin network, picture from [8].

> This second heuristic is effective but more challenging and less safe than heuristic 1. It can produce false positives where heuristic 1 cannot.

This paper also introduces a visualization of the biggest clusters of the network, shown in Figure 2.3. This visualization gives an idea of the biggest clusters and how they interact together. The edges represent a minimum of 200 interactions between the two nodes, the colored areas represent the clusters' external incoming values. We can see here that there are around 25 big actors on the network and thousands of small ones interacting with them. We can also state that the biggest actors are either Bitcoin exchange platforms to buy or sell Bitcoin for local currencies (Mt. Gox), gambling platforms (Satoshi dice), mining pools (DeepBit) or marketplaces (Silkroad, Bitmit). A visualization of the addresses is proposed on the Coinalytics website [3]. Coinalytics is a website proposing tools to visualize and explore the blockchain using a graph. The beta part of the website promises tools to explore relationships in the blockchain using a graph based visualization. It says it will be able to plot clusters and their interactions. Right now it is only possible to plot an address and track its previous transactions in a graph based view as shown in Figure 2.4. An interesting point on this website is that the beta has an export functionality to export the explored blockchain data in a CSV file. It is interesting as it was requested by our own end users. I will talk about it in the chapter 3.

A different visualization of the actors in the Bitcoin network is called The Bitcoin Big Bang. It is a visualization proposed by Elliptic - a wallet hosting company [5] [2]. This visualization is graph based and provides a view of the biggest and most influential clusters in the Bitcoin history.

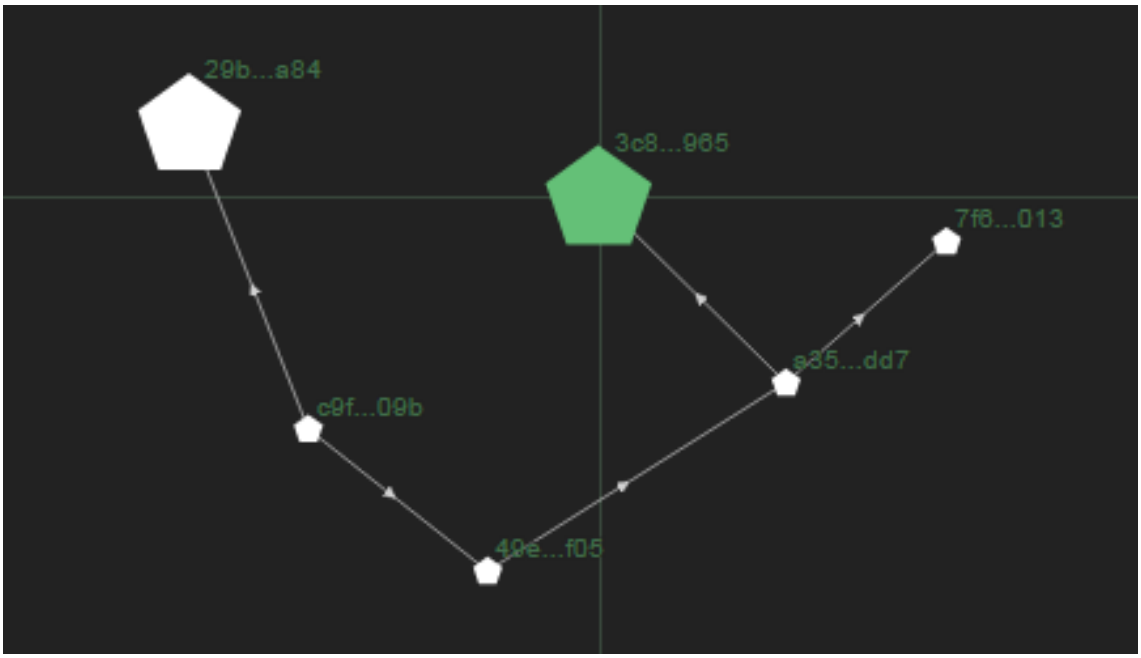


Figure 2.4: Screenshot of the beta version of Coinalytics. It shows addresses and their previous transactions.

This visualization provides an interactive interface to explore the relationships between a selection of the biggest clusters of the network as shown in Figure 2.5.

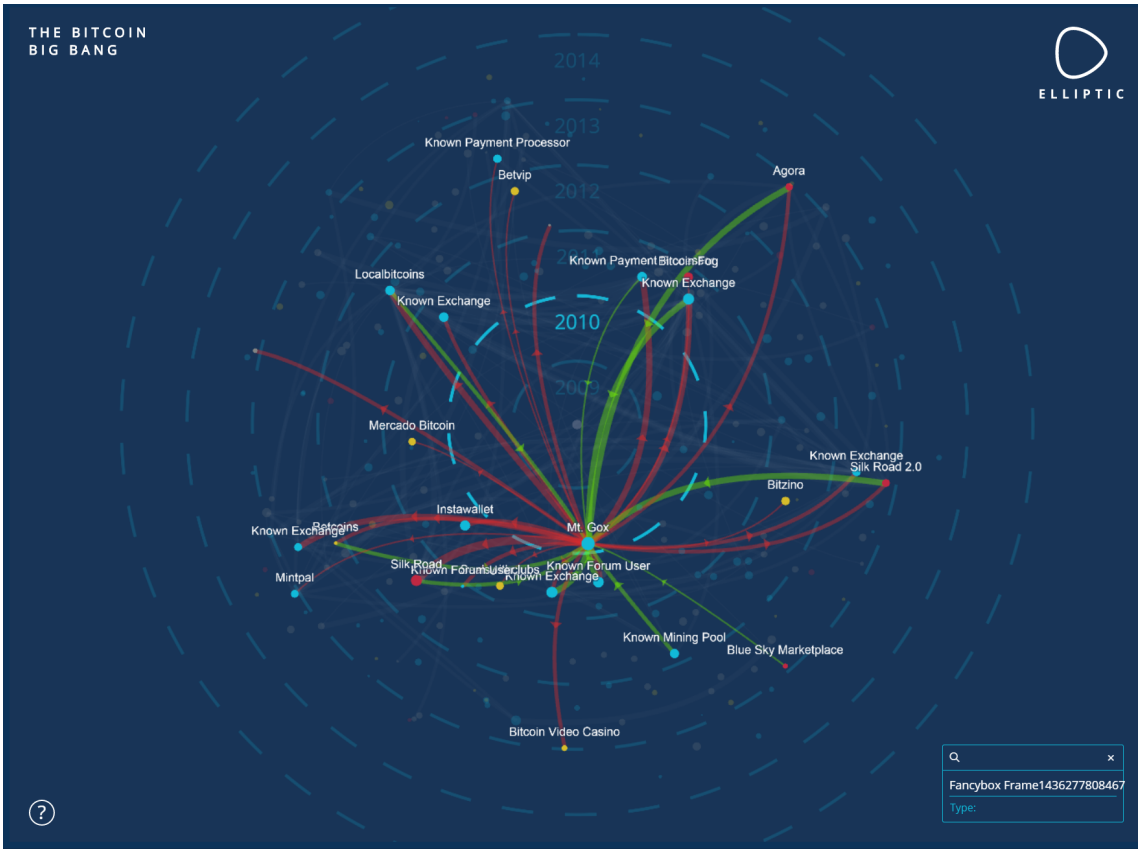


Figure 2.5: The Bitcoin Big Bang visualization, Mt. Gox cluster highlighted

CHAPTER 3

Requirement Gathering

This chapter presents the process by which we obtained the requirements we used to build our solution. It was done through meetings we organized with the end users of our system.

3.1 The users

My thesis formed part of the AIJE-bitcoin[1] project, a one year long exploration project on the development of Bitcoin. The end users are economists who want to learn more about Bitcoin and study its economy. Four of the economists are doing experimental research in economics in the Jean-Monnet faculty of University Paris Sud 11. One of them is a monetary expert and a former trader. The last two are experts in work and innovation economy. They are looking forward to discover this new payment system and to explore its features and see if it behaves as a currency. These users are very specific because they use financial visualization software on a daily basis, using CSV files in order to plot the metrics they are interested in. Most of them also know at least one programming language such as Python.

3.2 Meetings with the end users

During my thesis we had the opportunity to meet with the end users several times.

Preliminary Meetings

At the very beginning of the project, three preliminary meetings were held with the end users. In the first two meetings, we explained the technical background behind the Bitcoin payment system. The last meeting's purpose was to ask how the economists wanted to explore Bitcoin, what metrics they usually used on other currencies, what software they used and how they usually proceeded.
At the end of those meetings we discovered that the main questions the economists wanted to answer were about the users of Bitcoin. They wanted to know how many users there are, how many users hold the monopoly in this economy and if the miners take a big part in this economy. Answering the questions about the users of the

[1] It stands for Analyse informatique, juridique et économique de Bitcoin—Computer science, legal and economical analysis of Bitcoin in English. This project is led by Daniel Augot from Inria Saclay.

Bitcoin system was the most difficult task of this thesis. Firstly we would need to identify these actors (as explained in section 2.1, users can have as many addresses as they want to) and then find the best way to visualize them.
The other demands were for financial metrics with data that can be directly extracted from the blockchain like the number of transactions per day, etc. They also really wanted to see the transition phases. For example, being able to tell when the average transaction value dropped or rose and then figure out why.
In conclusion what the economists wanted to know at the very end is if this new payment system can be considered like a usual, centralized currency. There are two parts to figure out the answer: extract meaningful metrics out of the blockchain and identify the Bitcoin's users.

Showing the design mock up

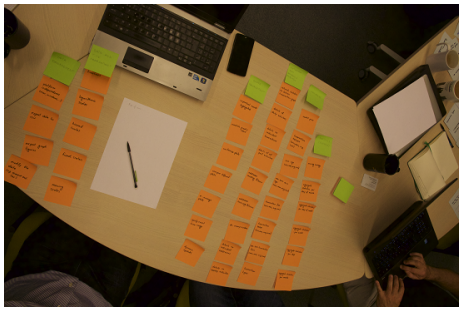
The next meeting occurred on March the 30th, 2015. Its goal was to show the early designs to the end users. The main role of this meeting was to offer them an early mock up in order to make them visualize where we were going. This way we could provide them a concrete representation for them to base their thoughts on. From this meeting we could tell the main focus was still users and particularly miners but a new demand came out. They wanted to have access to selected parts of the raw data. Either by exporting the data or by having the possibility to type SQL commands in. They were asking for this extracting functionality to be able to plot the data with the software they are used to.
This new demand should definitely be offered in the final product as it would easily enable the users to work on new metrics we did not think of.

Card sorting exercise

The final meeting we organized with the end users was held on June the 23rd, 2015. It was the occasion for us to involve the economists in a participatory design exercise. This exercise is called card sorting. This design exercise is used during the early design phase [11]. It is also a good way to find out which functionalities the users think are the most important as well as suggesting new ones the designer did not think of.
The first step of this exercise consisted of preparing one card for every functionality the product offers. These cards were then distributed to the participants. In our case we already had part of the product implemented. This was why at the beginning of the session, we showed the early prototype running and the functionalities implemented. This gave the opportunity for the participants to think more about new features we did not consider before.
Once the cards were distributed, the participants needed to group them according

(a) The first group of participants taking part in the card sorting exercise.

(b) The second group of participants taking part in the card sorting exercise.



(a) Card groups created by the first group of participants.

(b) Card groups created by the second group of participants.

Figure 3.1: The card sorting exercise session organized with the end users.

to what made the most sense to them, see Figure 3.1. The purpose of this exercise was to find new features and prioritize the ones we had not implemented yet.

This exercise was a success as the participants were very involved in the session. We even had the opportunity to talk about new functionalities with the economists. I present the results of this exercise in Figure 3.2. The features the economists asked for the most were the use of logarithmic scales in the interface as an option as well as the fixing of the scales and the most useful layouts. During the exercise, both groups decided to order the functionalities by importance. This was our basis to order the cards by importance when we merged the two result sets. Based on these results it was easy to plan immediate and future evolutions of the system as shown in Figure 3.2, respectively the blue and orange cards. The most immediate features I had to implement was the possibility to fix the scales and to use logarithmic scales as it was the most asked by the end users. Mostly, the blue cards represent the features I needed to implement before the end of the thesis. At this time, the feature I knew I would not implement was the de-anonymization of the transaction graph as it is impossible in its totality (it is possible to do it partly using known addresses of merchants for example or by crawling Bitcoin related forums looking for users giving away their addresses as it has been done in a PhD thesis [10]) and it would have require much more time than my thesis allowed.

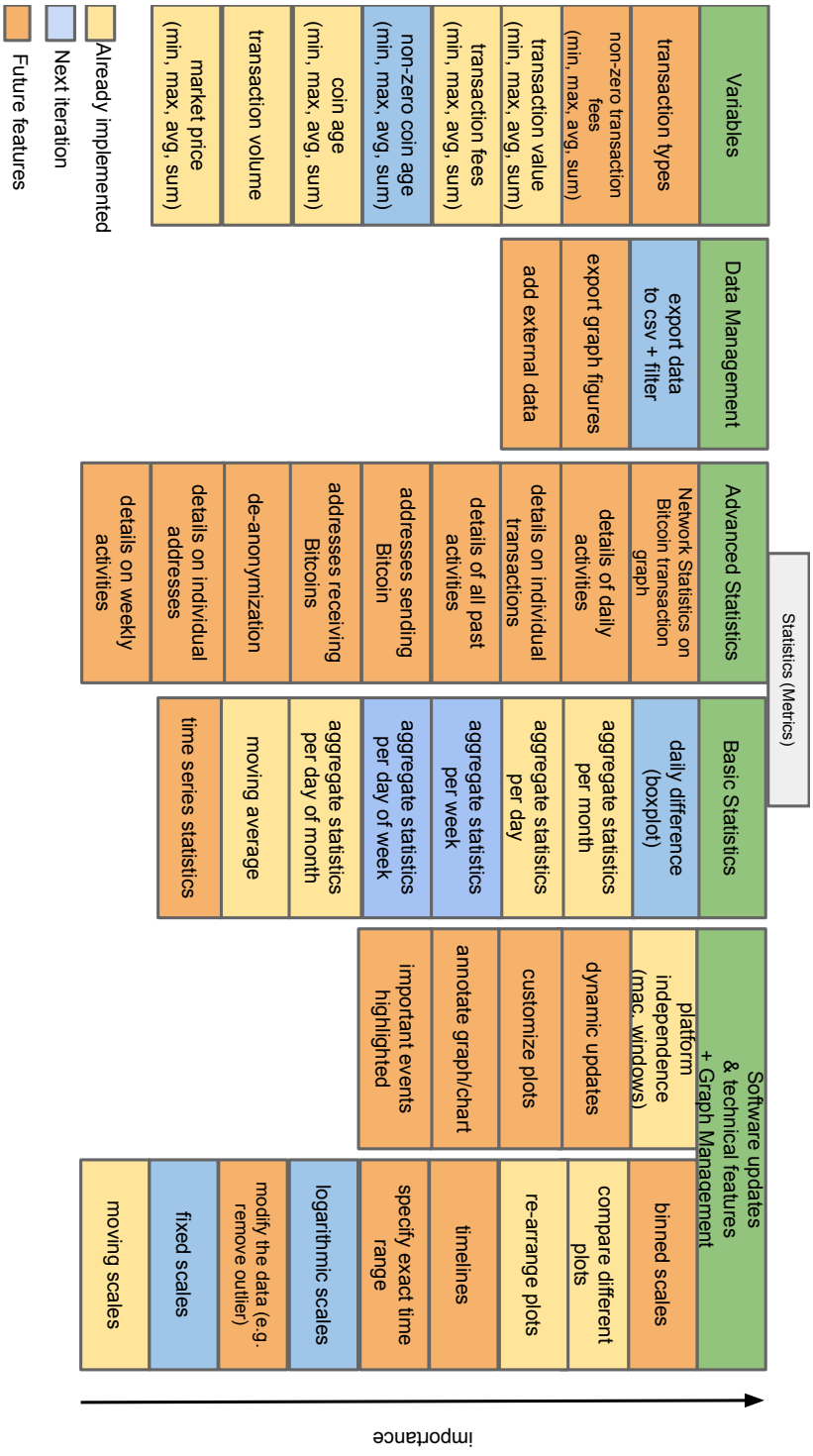


Figure 3.2: Card sorting exercise results. Features we talked about are grouped and sorted by importance.

CHAPTER 4

Solution

In this chapter I present our solution. First I present the raw functionalities of the interface along with the design choices I made and finally I present the evaluation of the interface.

4.1 General Presentation

The solution I propose is web-based and implemented mainly using Javascript and the D3.js library. D3.js is one of the most widely used Javascript libraries for visualization. Along with already implemented useful functionalities (such as behaviors, geography tools or layout tools for example) it ables the creation and manipulation of SVG[1] elements. SVGs are vector graphics made of vector elements described following the XML markup language. SVGs are useful for web-based solutions as it is vector based and therefore does not depend on screen resolution, size, etc.; and are a lot less heavy than non-vector images.

System architecture

As shown in Figure 4.1, the architecture consists of three parts: MIDAS, the web-server and the clients. MIDAS is a super-computer[2]. It is the one hosting the data and the scripts to access the data. As it cannot be made accessible over the Internet we had to dedicate a web server to this task. The web server is very basic and only

[1] SVG stands for Scalable Vector Graphics
[2] MIDAS has 96 processors and 1TB of RAM.

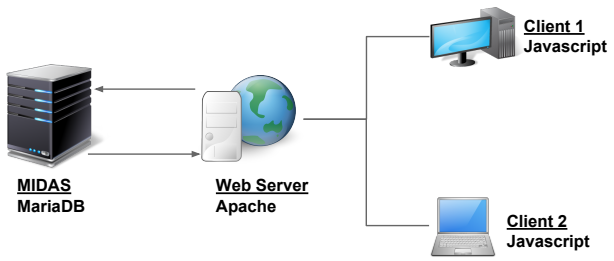


Figure 4.1: The system architecture.

runs Apache in order to deliver the web page to the clients. The clients can then connect to this web server at the address http://bitcoin-viz.saclay.inria.fr and clients run the Javascript. At the moment, the website is not publicly accessible. The database is not connected to the web server at the moment. The queries on the database are not fast enough yet and therefore having the database connected to the visual part would be disastrous for the interactivity of the tool. We plan to connect to it and directly query the data from the web server as I explain it in further work in Section 6.2. We currently use CSV files extracted from the database and stored on the web server to provide the data.

Database

As I explained in the previous section, the database is located on a super-computer. This database is run by MariaDB and is set up as a data warehouse. This database contains a copy of the Bitcoin's blockchain up to March 4, 2015. In order to create this database we first parsed the blockchain. This was done by downloading the Bitcoin core application from http://bitcoin.org/en/download. This app lets one download the full blockchain in the LevelDB database format. Then we used the parser of bitcoin-abe [1] to parse the blockchain's database into our SQL database. We do not currently update the database and therefore the data is available from the beginning of Bitcoin, January the 3rd, 2009, until March the 4th, 2015. We plan to update the database daily as I discuss it in further work in Section 6.2.

4.2 Presentation of the visual analytics tool

As shown in Figure 4.2, my solution is composed of three parts: the timeline at the top, the detail charts at the center of the screen and the settings. These three parts are designed according to the requirement I gathered from the end users (cf. chapter 3). The meetings with the economists made me choose this design (based on a timeline to select a time period to then plot the data) as they primarily wanted to be able to focus on interesting periods of time–according to one metric–and then compare this point of interest with other metrics. I present this design in this section.

Timeline

The timeline is located at the top of the interface. As shown in Figure 4.3, the timeline displays a metric specified at the top left (in this case market price in dollars). The user can change the displayed metric through a right click. The chart's period of time is based on the available data in the database, as I presented in section 4.1. The timeline is here to select a time period. The user can brush–click and drag–on the timeline to choose a time period. As soon as the user starts the

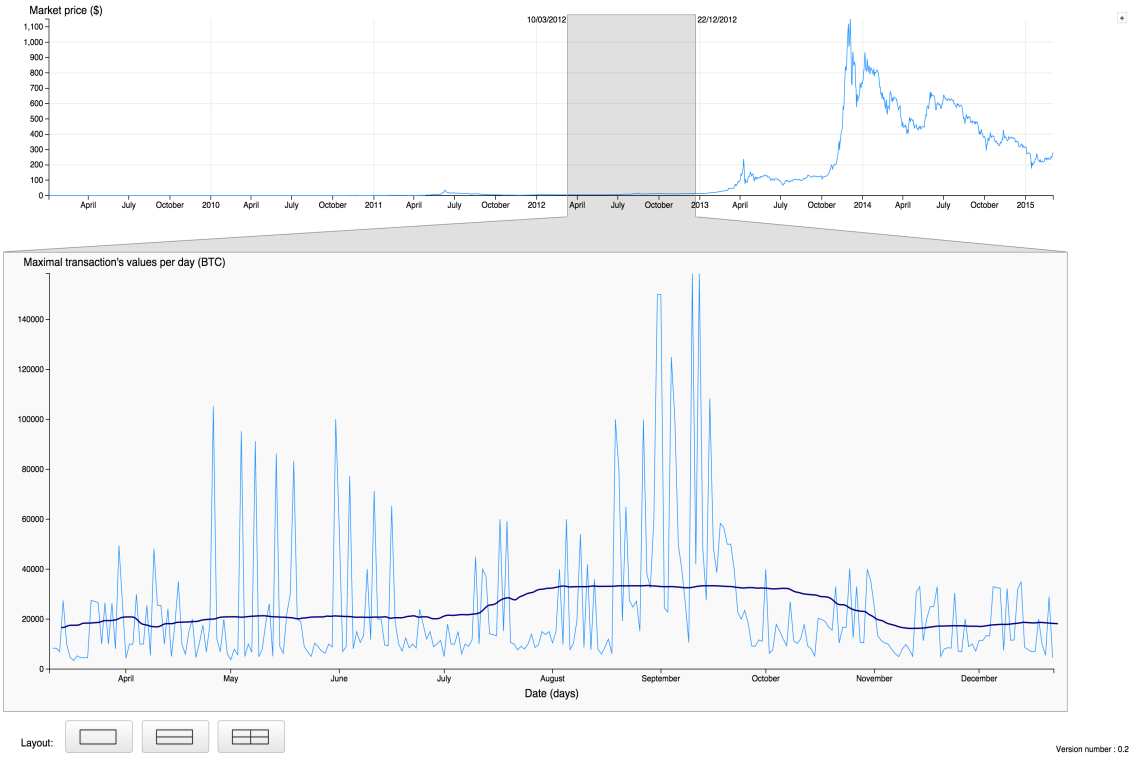
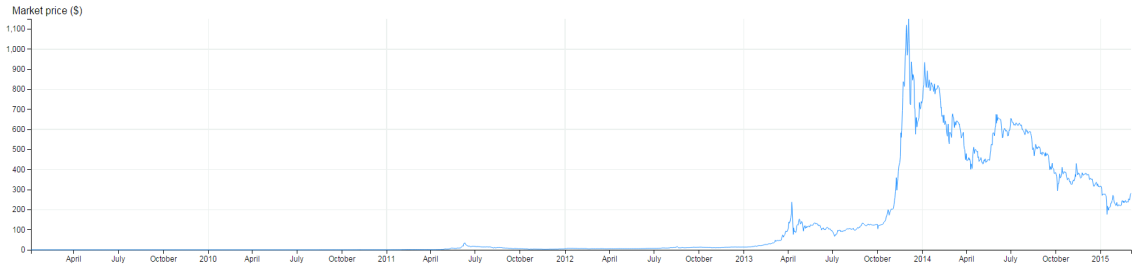


Figure 4.2: The entire user interface, with a selected time period and therefore the detail chart populated.

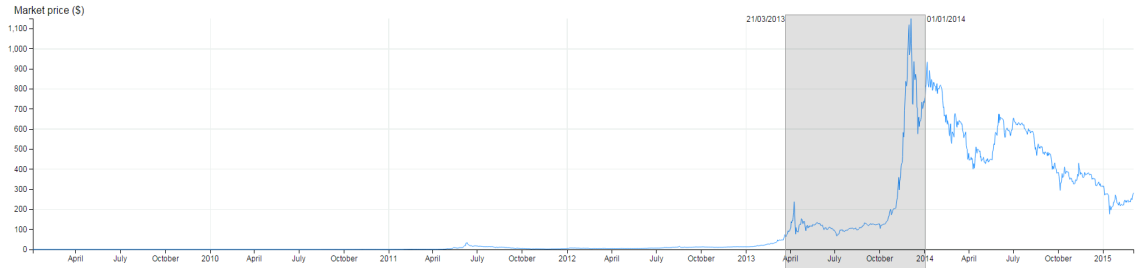
brush the main area–in gray at the interface's center–gets populated by the detail chart(s). The time period can also be chosen in the settings, see section 4.2. Along with the filling of the interface's center chart, a gray overlay is drawn above the timeline. This gray overlay can be modified by clicking and dragging it around to move the same period of time to another time or it can be expanded or reduced by dragging the sides.
The timeline is here for the user to specify a time period to get details on demand. Therefore the time period is the same in the timeline and in the detail charts. That is why the brushed time period on the timeline is associated with these detail charts by a visual link. This link can be seen under the timeline's brush in Figure 4.4.

Explanation of the timeline's design choice

The most important choice in the interface is the use of the timeline. The timeline is used to present one metric globally on the full available time period as well as to select a particular time period to plot the main. The use of a timeline was first motivated by the preliminary meetings with the end users. The economists wanted to be able to focus on an interesting period of time quickly and then compare this



(a) The empty timeline as it appears at the start of the interface.



(b) The timeline once the user has brushed the desired time period.

Figure 4.3: The two states of the timeline

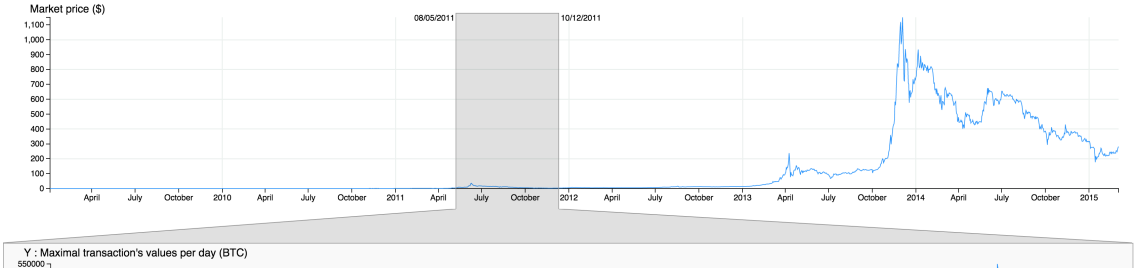


Figure 4.4: The link between the brushed time period and the detail charts area can be seen under the timeline's brush.

point of interest with other metrics to see the correlations and consequences. A timeline is also a good intuitive way to select a time period. I chose not to hide the timeline once the user had selected it as it provides a permanent feedback on the selected time period to the user. Finally, the timeline is generally a good medium to represent and manipulate financial/transactional data as the temporal factor is highly important.

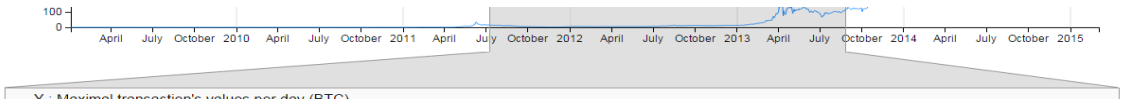


Figure 4.5: Zoom on the link between the brushed time period and the detail charts.

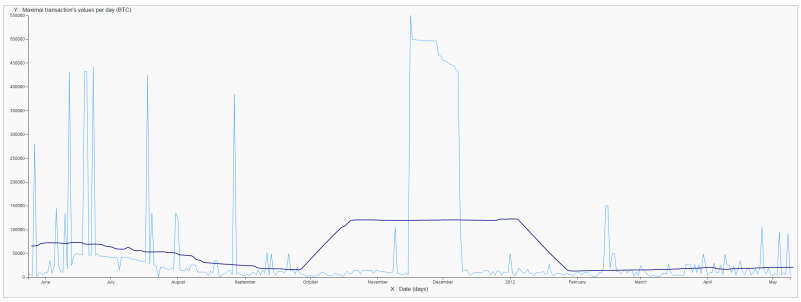
Explanation of the link's design choice

The user selects the time period used for the detail charts on the timeline. Therefore, the grayed area on the timeline represents the same time period as the entire X axis of the detail charts. In order to represent this logical link, I chose to create a physical link between the two areas. This link is represented by a gray trapezoid with its top edge merged with the bottom of the brushed area in the timeline and its bottom edge combined with the top edge of the chart, shown in Figure 4.5. This trapezoid form represents the graphical enlargement of the brushed time period into the detail chart.

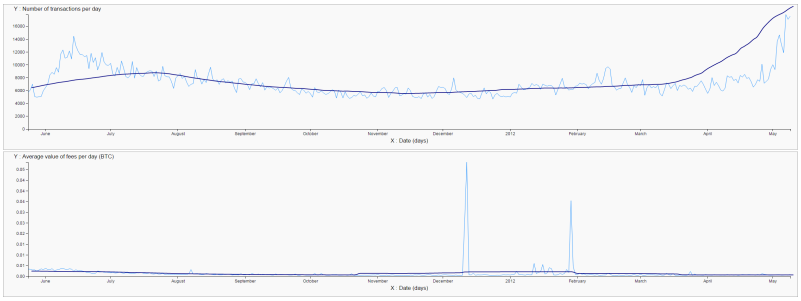
Detail chart(s)

The interface's center is dedicated to the detail charts. In order to populate these detail charts, the user must select a time period. The center area can contain up to four charts as showed in Figure 4.6. These different layouts can be selected through the three buttons located at the bottom of the interface.
These detail charts' goal is to plot the metrics the user is interested in. They can display more metrics than the timeline and are not date restricted on the X axis. The user can change the displayed metrics using a right click on the chart as seen in Figure 4.7. The X and Y metrics' names are displayed on the chart.
When hovering a chart with the mouse pointer the user can see the x value at which the mouse is positioned as well as the value of the metric at this point. The value's date is also displayed on the timeline by a red line as shown in Figure 4.8. This hovering effect can also be seen on the other charts of the current layout. This is very useful for the comparison of multiple charts.
The right click menu is used to change the X and Y axes metrics. The available metrics are:

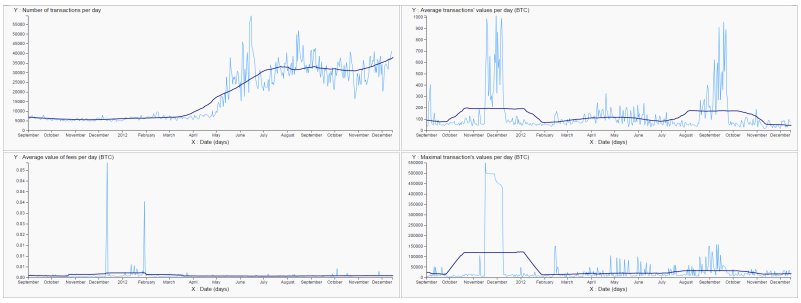
- Date (only for the X axis).
- Market price (in dollars): this is the average price of all the markets, provided by `http://blockchain.info`.
- Transaction volume: number of transactions.
- Transaction value: can be plotted by average, maximum, minimum (non-zero) or the sum of all transactions' values.



(a) Single chart layout: one big chart.



(b) Dual-chart layout: two charts, one on top of the other.



(c) Quad-chart layout: four charts.

Figure 4.6: The different layouts options.

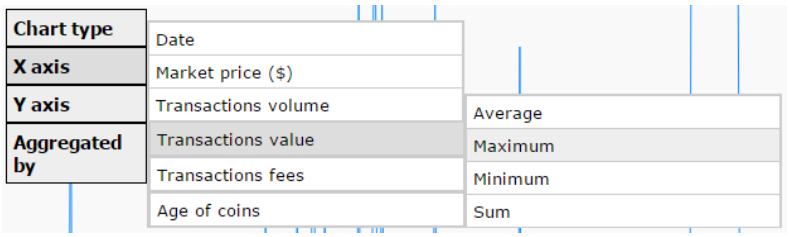


Figure 4.7: The right-click menu, selecting the metric maximum transaction value for the X axis.

- Transaction fees: can be plotted by average, maximum or the sum of all transactions' fees.
- Age of coins: can be plotted by average, maximum or the sum.

This right click menu also gives other chart-specific options detailed in the next subsection.

Settings and layouts

The settings can be found in different places of the interface as shown in Figure 4.9.

Layout buttons

The layout buttons are always accessible by the user. They are shown in Figure 4.9 (b). This setting can be changed using the three buttons at the bottom of the interface.

Settings menu

The settings menu can be found at the top right of the interface. It enables the selection of the time period's start and end dates.

Right click

The right click menu hosts a lot of chart-specific settings. It is used to :

- Change chart type : Line, Bar, Scatter plot, Area.
- Change X and Y axes metrics (presented in Figure 4.2).
- Change the aggregation of the data : day or month. The user can ask for data computed daily or monthly.

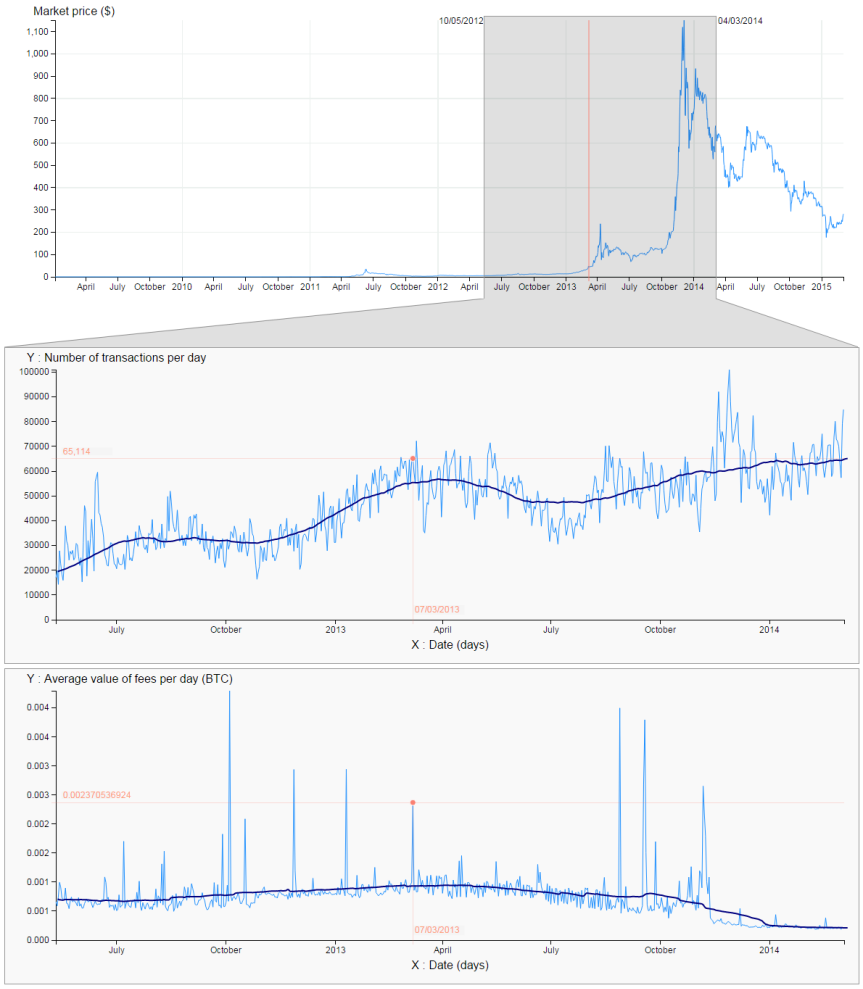
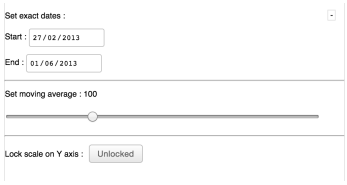


Figure 4.8: The result of hovering a chart with the mouse pointer, here the mouse is hovering March the 7th, 2013.

The right click menu is accessible on every chart. It differs slightly for the timeline where the user can only select the Y axis metric.

Explanation of the layouts' design choice

The interface proposes three different layouts for the detail charts. A single chart layout, a two chart horizontally split layout and a two by two layout. The single chart layout provides a large view of the plotted metrics. The two other layouts are very useful to compare different charts such as different metrics on the Y axis plotted with the same X axis for example.
Originally, there was a fourth layout that was composed of two charts split vertically (two columns, one line). I decided to remove this layout for two reasons. Firstly, I



(a) The settings menu, located on the right of the interface when open.

(b) The layout buttons, located under the detail charts on the left.

Figure 4.9: The different settings of the interface.

had feedback from the end users saying they were not seeing the point of this layout. Secondly, this layout was not useful to compare different metrics as the scale was not necessarily the same on the two charts' Y axes and the two charts were compressed in their width and not in their height, making them look stretched.

4.3 Evaluation

This section presents the feedback on what could be improved in the interface. I collected informal feedback from two different groups of people. The first group is composed of people with a computer science research background. I gathered their feedback at two occasions. At first, I presented the interface at the APVP[3] workshop and received informal feedback at this occasion. Then, at the end of my thesis, I asked two HCI students to give me feedback on my interface as well. The second group of people is composed by the end users of this tool that gave me informal feedback during the card sorting exercise.

Instructions

The main feedback regards the lack of instructions on how to use the interface. At first, the interface is almost completely empty, showing only the chart plotted as the timeline. It is not obvious one has to brush a certain period of time to then populate the center area with charts. One workaround for this issue could be to select the whole available time period at first, so the main area would be populated and the link between the timeline and the center chart would be present and more evident.

[3]APVP stands for Atelier sur la Protection de la Vie Privée or Privacy protection workshop in English. This year it took place in Mosnes, France from June 15 to June 17.

Active chart

Another annoyance in the tool occurs when one changes the layout. The displayed charts are either the default set ones or the ones the user last set on a specific chart. This is not optimal in terms of user experience as when a user changes the layout, it is mostly for comparing the same metric he is looking at with others. This could be solved by using an active chart the user can select before changing the layout.

Hovering several points

One interesting feedback I collected concerns the hovering tool present in the interface. At the moment, when one hovers a chart one can see the X and Y values of the current hovered point. The feedback I received was asking for the possibility to hover more than one point. For example, one can fix one point (rendering it like it was hovered) and still continue to hover other points.

Horizontal layout

Another feedback I received from one of our end users was to have the possibility to add more chart in the two rows and one column layout. This would have the effect to shrink vertically the charts but would give the opportunity to compare a lot more of them based on the same time period that would still be represented by the same width. This is definitely an interesting option for the interface.

Graphic bug

The last recurrent feedback I had concerns a graphical bug appearing either when there are too many charts plotted on the screen (occurring very often on the two by two layout) or when the time period is very large and there are too many data to handle. It happens because I am redrawing the detail chart(s) each time the user moves or resizes the brush selection on the timeline (not waiting the end of the modification). One good workaround for this bug would be to disable the redrawing of the chart while the user did not end the brushing of the timeline (at least on the layouts with several charts).

CHAPTER 5

Discussion

This chapter is presenting the extensibility of the solution as well as its limitations.

5.1 Extensibility

Our proposed solution can be used to evaluate other cryptocurrencies such as LiteCoin or DarkCoin[1]. As it is based on the same principle as Bitcoin with a public blockchain which contains blocks that contains transactions from one address to another it can be visualized using our tool. The metrics used are also useful and meaningful to other cryptocurrencies. This means the exact same interface could be kept to analyze these other payment systems.
In regards to the back end part of out tool, as these cryptocurrencies are based on the Bitcoin protocol, our database could store the same information. Therefore, no change in the database should be envisaged. The only necessary change would be the parsing of the blockchain in our database. As the blockchain's format of those cryptocurrencies is slightly different than Bitcoin, it would be necessary to create a new parser.

5.2 Limitations

The biggest limitation in my solution is the absence of a visualization of the users. This is one of the main demands of the economists and it is very important for them to be able to visualize the actors in this economy as it is an essential basis for them to be able to study Bitcoin. Even if I evaluated the use of BioFabric (discussed in section 6.2) to help visualize a large set of actors, it was not straightforward to design and implement it to be a part of my visual analytics tool.
The second main limitation in the solution at the moment is that the database is not connected with the front end. As I explained earlier in section 4.1, this was first intentional as the queries on a not optimized SQL database would have killed the interactivity of the interface. This is the reason why we began to use preprocessed CSV files exported from our database. It is still not connected due to time limitation as it is not straightforward to make this connection happen. The connection would be based on AJAX[2] queries that are not yet supported by MIDAS. I discuss more about the connection and its benefit in the further work section 6.2.

[1] These are two cryptocurrencies created from a fork of the Bitcoin source code.

[2] AJAX stands for Asynchronous Javascript and XML and is used to send data to and retrieve from a server asynchronously (without modifying the display or reloading the page).

CHAPTER 6

Conclusions and perspectives

In this thesis I proposed a visual analytics solution enabling our end users to explore the financial data of the Bitcoin cryptocurrency. This tool can be used by its end users to plot the meaningful metrics they asked for at the beginning of this thesis.

6.1 Contributions

This thesis contributes to three parts:

1. The making of the visual and interactive component of a visual analytics system to analyze metrics of the Bitcoin transaction graph.

2. This system was built on user requirement gathered from domain experts. Other visual analytics tools to explore Bitcoin's metrics exist but the metrics I provide are meant to support specific Bitcoin analysis that will be done by the end users.

3. The preliminary evaluation of the tool during the card sorting exercise with the end users (presented in section 3.2) as well as a workshop where I presented the interface and received feedback.

6.2 Further work

In this section, I present the next steps in the development of our solution.

Connecting to the database

One of the most important improvement to obtain a final product is to connect the interface to the database. It is not the case right now as we are using CSV files extracted from the database. This connection to the database would be beneficial as the user could query a lot more from our database and even make his or her own queries using SQL for example. We also thought of having the possibility for the user to query through the titles of the charts. For example, one could first select the function used—average, maximum, minimum or sum—then the metric—volume, number of transaction, age of coins, etc.—and finally the aggregation of the chart—monthly or daily— to finally give the full title and send the query.
The main disadvantage of this connection would be the time it would take to get the results. The database is in a data warehouse form and so is really optimized for

speed but the database is also really big and therefore there will always be a delay between the querying and the displaying of the results. In order to counterbalance this downside, we thought of a cache system. It would store on the web server relevant queries—it could be the most used ones—to then deliver the results directly without contacting the database.

Visualization of the users

The next big step in the development of this product is the development of the visualization of the actors in the Bitcoin network. However, this step requires to be connected to the database. Without the connection to the database, this visualization would take a really large amount of precomputed files. Furthermore, using the data warehouse format of the database, a user could filter a lot of parameters of this visualization. One could select a time period and see the actors interactions over this time period and their evolution by dragging the timeline selector.
The design phase of this visualization has already started and we are thinking of using BioFabric. BioFabric is a novel network visualization technique. It displays the nodes of the network as horizontal lines and the edges as vertical lines. It was primarily invented for biology visualization and is meant to represent a lot more nodes than a graph visualization of a network. An example of how it would render with 80 actors can be seen in Figure 6.1. It displays the actors/clusters as horizontal lines and their interactions others as vertical lines. These interactions represent a certain number of transactions (1,000 for example) between the two clusters. In this figure we can see that cluster 12 is the cluster connected with most other clusters. This visualization is great for showing many clusters and interactions and to extract visual patterns. The disadvantage of this visualization is that for really specific interactions (involving a few clusters), it can be quickly disturbing and counter-efficient. That is why this visualization must be coupled with a more traditional network visualization for fewer clusters.

Update of the database

Another improvement to make this solution better would be to have an update of the database with the last blocks of the blockchain. This could be done daily or hourly in order to process a rather large or small amount of new blocks. In contrast, the main disadvantage of updating the database on a time period that is too short is that we could face orphan blocks added to the database. Orphan blocks are blocks that were in a different fork of the current blockchain but did not get accepted by the whole network. This could be a problem as the information contained in an orphan block could be added to the database and not be valid later on.

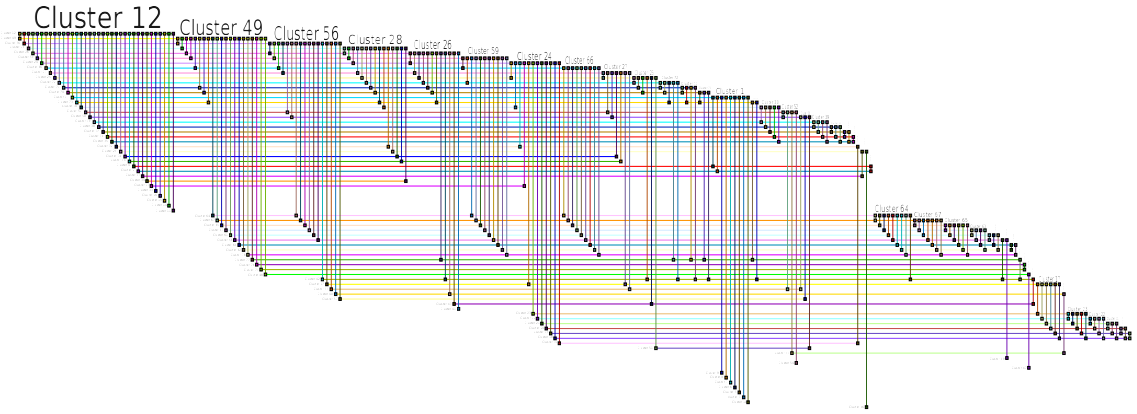


Figure 6.1: A mockup of the visualization of 80 clusters using BioFabric.

Collaborative tools

The last big improvement this product needs is to make it collaborative. From the point of view of the end users, they will not work on the same computer and so it would be useful to have a way for them to work together. We thought of two separate elements to enable the collaborations of our end users.

Annotations

The first one would be to give the possibility to a user to annotate a certain chart at a certain point of time. These annotations would be seen by everyone. They could be used to annotate an event in the Bitcoin history for example or to pinpoint an interesting pattern for the others to see.

Save and Load

The second tool would be the possibility to save a state of the interface (the time period, the metrics, the settings selected) in a file on the user's computer. Users could then re-use these files to restore their previous sessions or share these files with other users to point them an interesting view.

This visual analytics tool is the first block of a complete system that will allow the economists to judge if Bitcoin is behaving as a currency.

Bibliography

[1] Bitcoin-abe. https://github.com/bitcoin-abe/bitcoin-abe.

[2] The bitcoin big bang, visualization of clusters. https://www.elliptic.co/anti-money-laundering/.

[3] Coinalytics, website proposing blockchain exploration. http://coinalytics.co.

[4] Controlled supply, bitcoinwiki page. https://en.bitcoin.it/wiki/Controlled_supply.

[5] Elliptic. https://www.elliptic.co/.

[6] Cyprus university world first to accept bitcoins for tuition, 2013. http://www.rt.com/business/bitcoin-nicosia-university-tuition-060.

[7] The new york bar that takes bitcoins, 2013. http://money.cnn.com/2013/04/08/investing/bitcoin-bar-new-york-city/index.html.

[8] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon Mccoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of bitcoins: Characterizing payments among men with no names, 2013. https://cseweb.ucsd.edu/~smeiklejohn/files/imc13.pdf.

[9] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. http://bitcoin.org/bitcoin.pdf.

[10] Michele Spagnuolo. *BitIodine: Extracting Intellignece from the Bitcoin Network*. PhD thesis, Politecnico di Milano, 2013. https://miki.it/pdf/thesis.pdf.

[11] Donna Spencer. Card sorting: a definitve guide, April 2004. http://boxesandarrows.com/card-sorting-a-definitive-guide/.