

# Bitcoin Network Project

*Master Finance, Technology, Data*

*By Daniel Deutsch, José Lucas Barretto, Karim Jaber and Ziyad Bekkaoui*

In [ ]:

```
import warnings

import networkx as nx
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from pytrends.request import TrendReq
```

In [ ]:

```
# Filter warnings
warnings.filterwarnings('ignore')

# Matplotlib styles
plt.style.use('ggplot')
plt.rcParams.update({
    'figure.figsize': (15, 7),
    'axes.prop_cycle': plt.cycler(color=["#4C72B0", "#C44E52", "#55A868", "#8172B2", "#CCB974", "#64B5CD"]),
    'axes.facecolor': "#EAEAF2"
})
```

## Introduction

Wikileaks is an international non profit organisation created and lead by Julian Assange. Their objective is to reveal classified leaks and information through anonymous sources. Their main subjects are human right violation, political scandal and corruption in all the countries around the world. One of their financing way is donation and in order to keep their donation anonymous, they created crypto currencies wallet and made their address public in order to allow everyone to transfer funds. The aim of this work is to analyse these known addresses and to extract from it information on the fund management made by Wikileaks and also to identify trends that influence the donations. To do so, we will analyse their two known bitcoin addresses:

- 1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v
- 1MaXZE92yjuy4NYjTspmdWHMRT3jQUcTf4

## Methodology

The methods we will use for our work is to use python and the library Network X. the data provided by the teachers regroup all the transactions on these two addresses from 2011 to 2021. On a second step we will plot graphs to visualise the data and conclude information based on it.

## Data Import

In [ ]:

```
# Import the wikileaks dataset
df = pd.read_parquet('datasets/wikileaks.parquet')

# Import Google Trends data
pytrends = TrendReq()
pytrends.build_payload(['wikileaks'], timeframe='all')
df_trends = pytrends.interest_over_time()
```

## Data Processing

In [ ]:

```
# Process the Wikileaks dataframe
df['date'] = pd.to_datetime(df['time'], unit='s')
df.rename(columns={'PriceUSD': 'price_usd', 'valueUSD': 'value_usd'}, inplace=True)
df.sort_values(by='date', inplace=True)

# Process the Google Trends dataframe
df_trends = df_trends.reset_index()
df_trends.rename(columns={'wikileaks': 'trend'}, inplace=True)
```

## Donations to Wikileaks Over Time

In [ ]:

```
# List of Wikileaks addresses
wikileaks_dsts = ['1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v', '1MaXZE92yjuy4NYjTspmdWHMRT3jQUcTf4']

# Aggregates donations to wikileaks by day
df_agg = df[df['dst'].isin(wikileaks_dsts)].set_index('date').resample('MS')['value_usd'].aggregate(['sum', 'count']).reset_index()

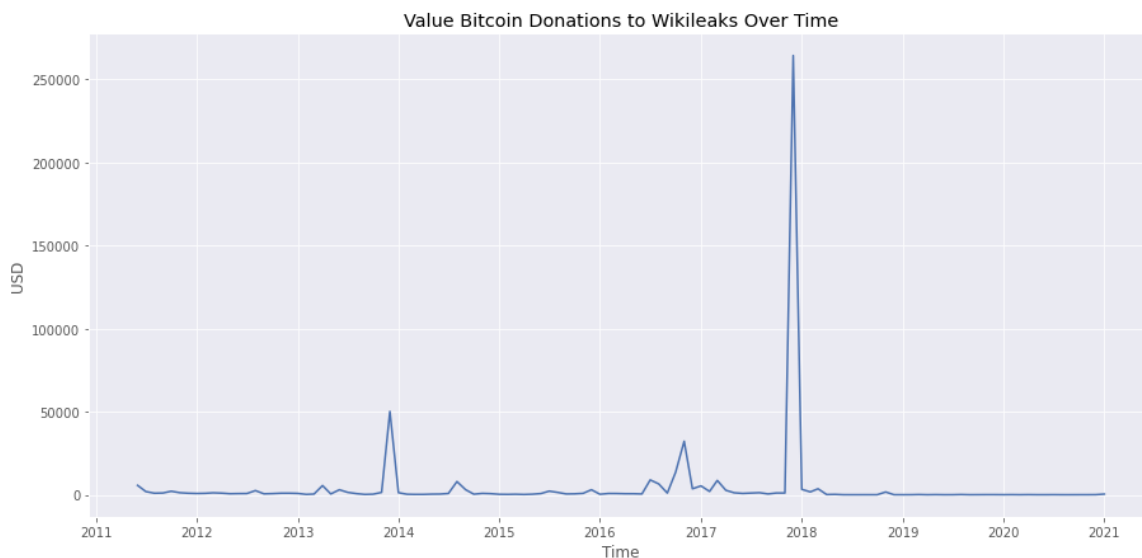
# Merges aggregated results with google trends on date
df_agg = pd.merge(df_agg, df_trends, how='inner', right_on='date', left_on='date')
```

## Bitcoin Donations to Wikileaks Over Time

We firstly made an analysis to study the evolution of BTC donations over time on the graph below. Our finding show that the donations are very different over time in terms of USD value. The main donation period are around 2014, 2017 and 2018. The can cross this data with the reveal made at this period. On 2016, Wikileaks revealed hundred of thousands of hacked email from Hillary Clinton during the presidential election. The pump on 2014 and 2015 maybe linked to the revelation Sony emails which showed that women actresses are less payed than male actors. The biggest pump on 2018 correspond to the period were Julian Assange was under law suit by the US government and had to escape for political protection in Foreign countries and when the UK government decided to refuse to accord him a protection against US extradition.

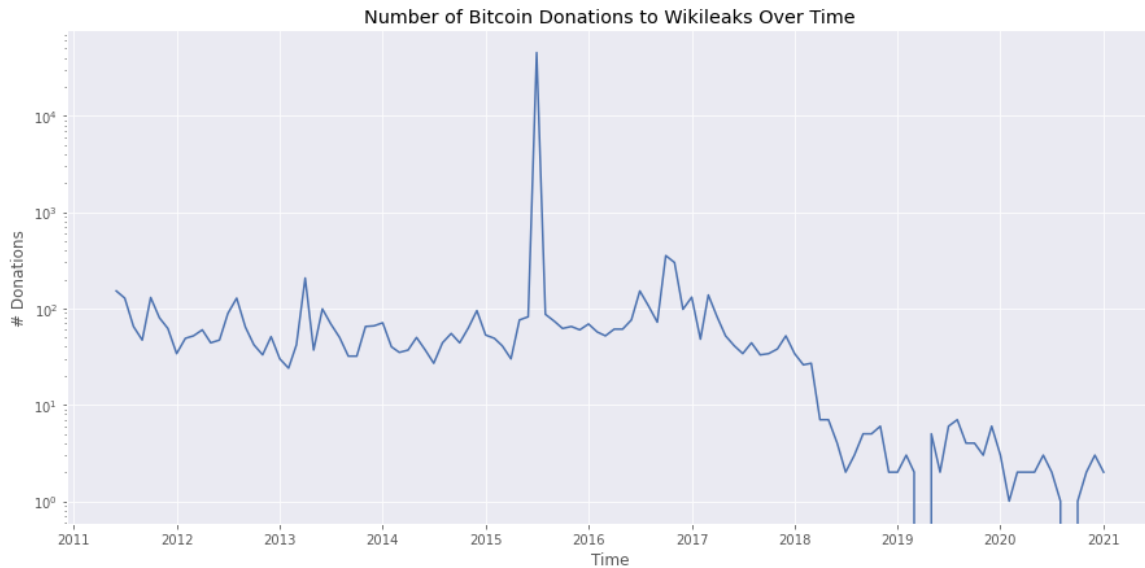
In [ ]:

```
plt.plot(df_agg['date'], df_agg['sum'])
plt.title('Value Bitcoin Donations to Wikileaks Over Time')
plt.ylabel('USD')
plt.xlabel('Time')
plt.show()
```



In [ ]:

```
plt.plot(df_agg['date'], df_agg['count'])
plt.title('Number of Bitcoin Donations to Wikileaks Over Time')
plt.ylabel('# Donations')
plt.yscale('log')
plt.xlabel('Time')
plt.show()
```



## Correlation Between Number of Wikileaks Donations and Google Trends

In order to continue this analysis we decided to cross the evolution of BTC donation with google trends data in order to study if there's a link between the donations and google research volume on wikileaks. Our findings shows that there is no evidence that wikileaks research volume drive the donations amounts.

In [ ]:

```
# Adds the log of the aggregated values to the dataframe
df_agg['log_sum'] = np.log(df_agg['sum'])
df_agg['log_count'] = np.log(df_agg['count'])

# Shows the correlation between each column
df_agg[['sum', 'count', 'trend', 'log_sum', 'log_count']].corr()
```

Out[ ]:

	sum	count	trend	log_sum	log_count
sum	1.000000	-0.006099	0.036456	0.355351	0.101822
count	-0.006099	1.000000	0.001077	0.079223	0.442853
trend	0.036456	0.001077	1.000000	0.448967	0.455833
log_sum	0.355351	0.079223	0.448967	1.000000	0.781307
log_count	0.101822	0.442853	0.455833	0.781307	1.000000

In [ ]:

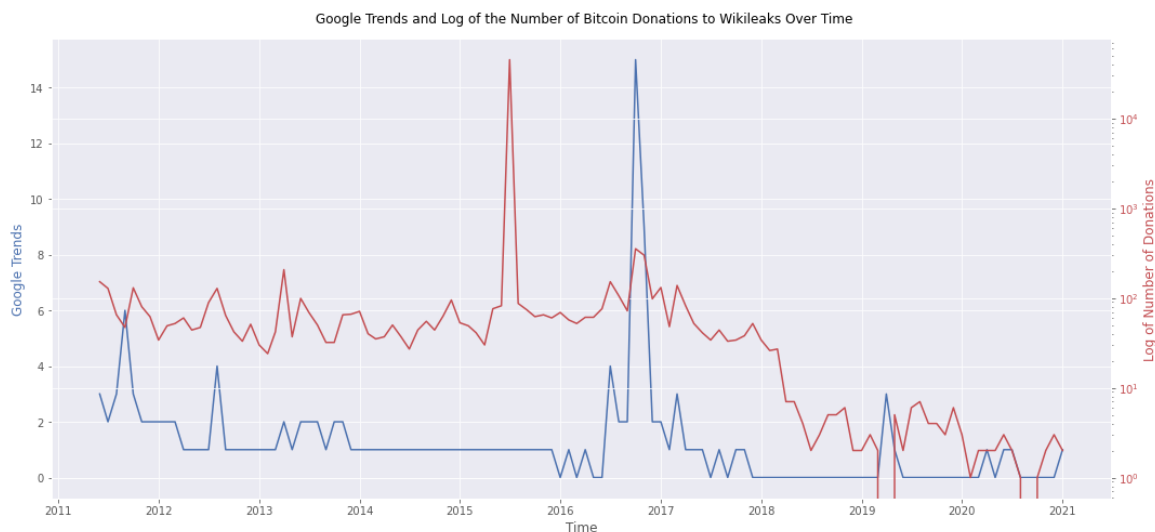
```
fig, ax1 = plt.subplots()

ax1.plot(df_agg['date'], df_agg['trend'])
ax1.tick_params(axis='y')
ax1.set_xlabel("Time")
ax1.set_ylabel("Google Trends", color='#4C72B0')

ax2 = ax1.twinx()

ax2.plot(df_agg['date'], df_agg['count'], color='#C44E52')
ax2.tick_params(axis='y', labelcolor='#C44E52')
ax2.set_ylabel('Log of Number of Donations', color='#C44E52')
ax2.set_yscale('log')

plt.suptitle("Google Trends and Log of the Number of Bitcoin Donations to Wikileaks Over Time")
plt.tight_layout()
plt.show()
```



In [ ]:

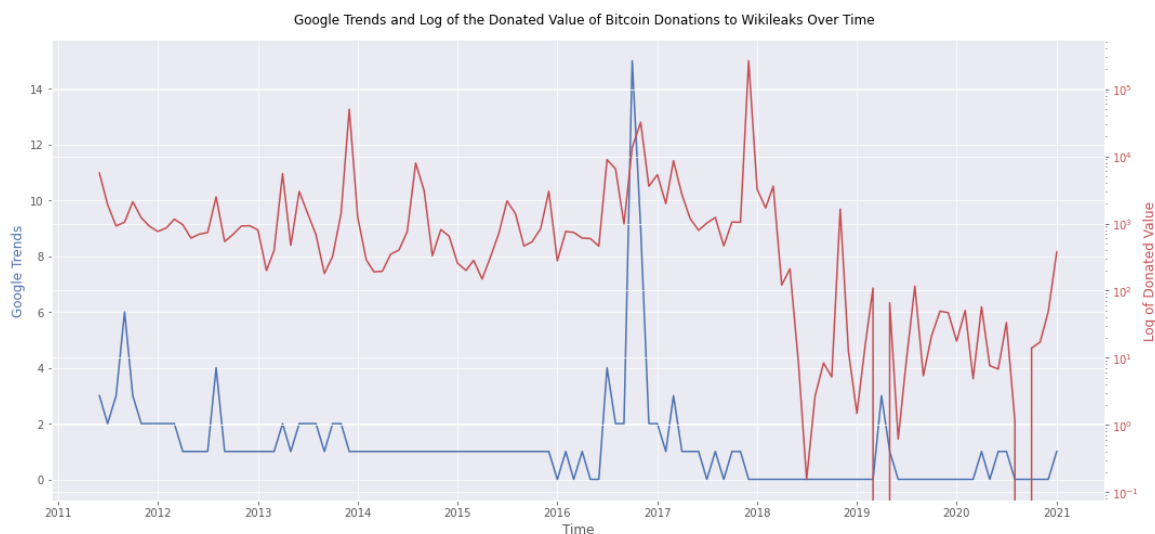
```
fig, ax1 = plt.subplots()

ax1.plot(df_agg['date'], df_agg['trend'])
ax1.tick_params(axis='y')
ax1.set_xlabel("Time")
ax1.set_ylabel("Google Trends", color='#4C72B0')

ax2 = ax1.twinx()

ax2.plot(df_agg['date'], df_agg['sum'], color='#C44E52')
ax2.tick_params(axis='y', labelcolor='#C44E52')
ax2.set_ylabel('Log of Donated Value', color='#C44E52')
ax2.set_yscale('log')

plt.suptitle("Google Trends and Log of the Donated Value of Bitcoin Donations to Wikileaks Over Time")
plt.tight_layout()
plt.show()
```

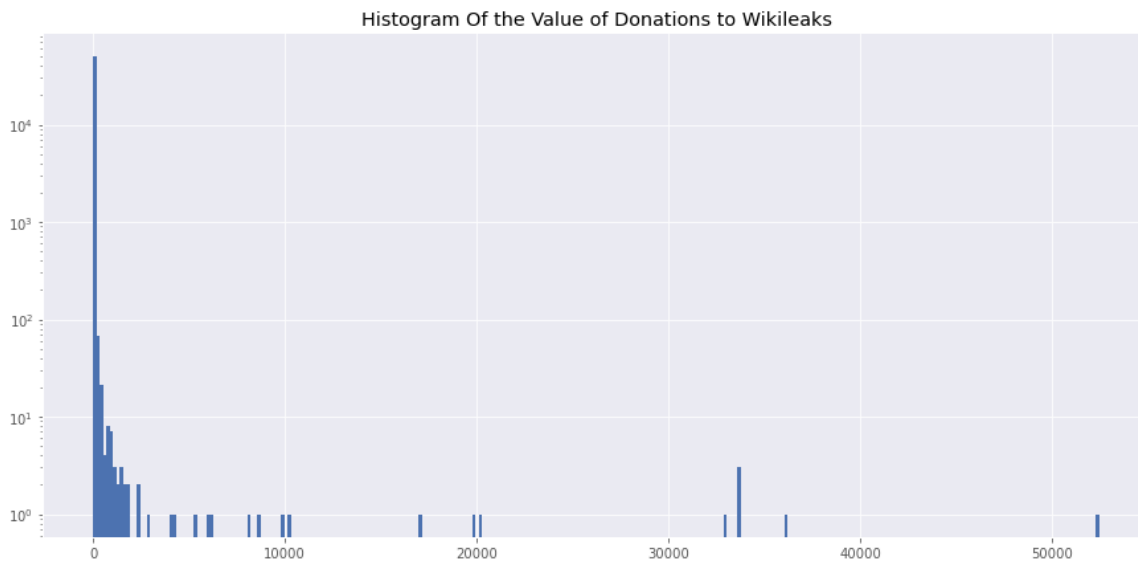


## Histogram Of the Value of Donations to Wikileaks

The next step will be to analyse the distribution of the donations amount in USD and then compare the coming flows and the exit flow from these wallets.

In [ ]:

```
df[df['dst'].isin(wikileaks_dsts)][ 'value_usd' ].hist(bins=300, log=True)  
plt.title("Histogram Of the Value of Donations to Wikileaks")  
plt.show()
```



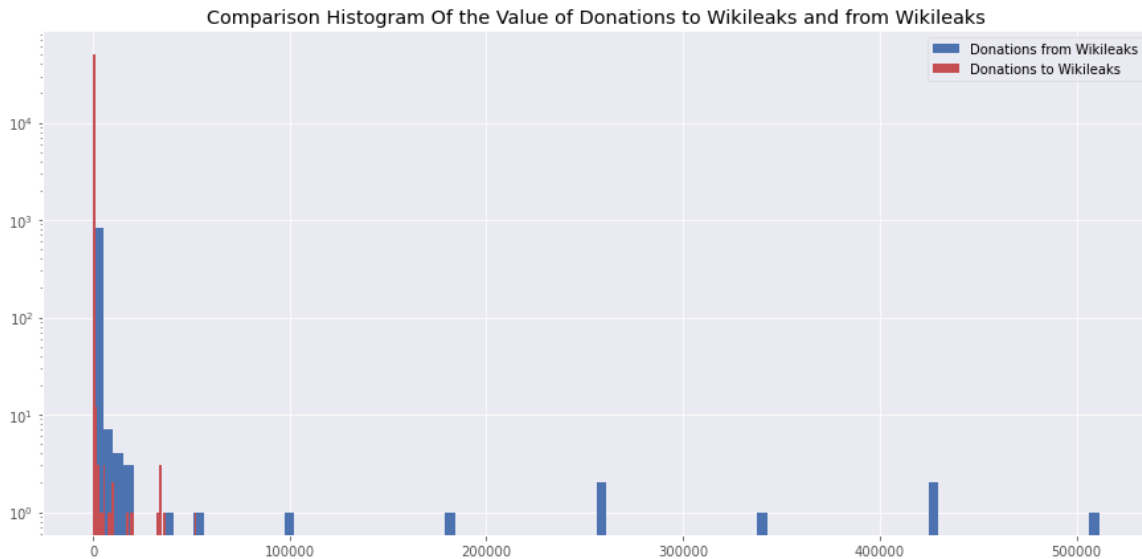
These graph shows that most of the donations are between 0 and 500 USD. However, above 500 USD, the donation are well spreaded with some donations that can represent tens of thousands USD.

## Comparison Histogram Of the Value of Donations to Wikileaks and from Wikileaks

On the next graph were we compare entering and exiting flow, we see that most the exit flow represent large amounts with a maximum of over 500 000 USD. In general, the amount per payments made by Wikileaks is relatively larger than the amounts that it receives through donations. Our hypothesis, is that Wikileaks use their funds to fund whistleblowers or organizations that may help then, but prefer to stay anonymous by receiving payments in Bitcoin.

In [ ]:

```
df[df['src'].isin(wikileaks_dsts)][ 'value_usd' ].hist(bins=100, log=True)
df[df['dst'].isin(wikileaks_dsts)][ 'value_usd' ].hist(bins=50, log=True)
plt.title("Comparison Histogram Of the Value of Donations to Wikileaks and from Wikileaks")
plt.legend(['Donations from Wikileaks', 'Donations to Wikileaks'])
plt.show()
```



## Graph Statistics

We can also verify through the out degree of the Wikileaks nodes that Wikileaks send out coins to a relatively few number of wallets. This can further sustain our previous hypothesis that the company funds collaborators with Bitcoin.

In [ ]:

```
print(f"Number of accounts Wikileaks transfers to: {len(df[df['src'].isin(wikileaks_dsts)][ 'dst_identity' ].unique())}")
print(f"Number of transactions made by Wikileaks: {len(df[df['src'].isin(wikileaks_dsts)])}")
```

Number of accounts Wikileaks transfers to: 35

Number of transactions made by Wikileaks: 849

## Network Visualization - Donations to Wikileaks



In [ ]:

```
# group source-destination pairs
df_network = df.groupby(['src', 'dst']).sum()['value_usd'].reset_index()
df_network = df_network[df_network['dst'].isin(wikileaks_dsts)]
df_network.replace(
    {
        '1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v': 'wikileaks1',
        '1MaXZE92yjuy4NYjTspmdWHMRT3jQUcTf4': 'wikileaks2'
    },
    inplace=True)

# select top 100 interactions in terms of value
df_network = df_network.sort_values(by='value_usd', ascending=False).iloc[:500]

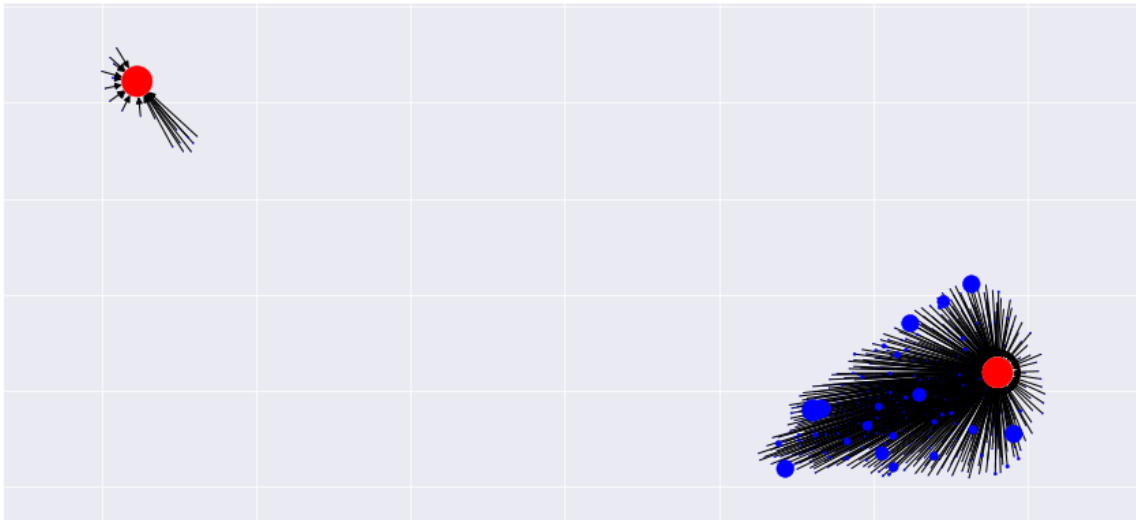
# create graph of donation network
G = nx.from_pandas_edgelist(df_network, source='src', target='dst', edge_attr=True,
    create_using=nx.DiGraph())
```

In [ ]:

```
# plot donation network
size_dict = dict(zip(
    df_network['src'],
    (df_network['value_usd'] - df_network['value_usd'].min())/(df_network['value_usd'].max() - df_network['value_usd'].min())
))

colors = []
sizes = []
for node in G:
    if node == 'wikileaks1' or node == 'wikileaks2':
        colors.append("red")
        sizes.append(500)
    else:
        colors.append("blue")
        sizes.append(250*size_dict[node])

nx.draw_networkx(
    G,
    arrows=True,
    with_labels=False,
    node_size=sizes,
    node_color=colors,
    label='Graphical representation of WikiLeaks donations.'
)
```



The red nodes represent the wallets owned by Wikileaks and the blue nodes represents the donators wallets. The size of the node is proportional to the total of donation. We can see that one of the two wallets is way more used and receive more donations. Surprisingly there is no links between the two wallets, we can conclude that they are managed in a separate way. The wallet who receive the most donations, have a very diverse base of donators, which is in accordance with the distribution amount graph.