**Blockchain Analysis Proposal**

**COMP 7570**

**By Baha Rababah**

**What is your group, what are your considered articles?**

Group: clustering.

Papers:

Opsahl, Tore, and Pietro Panzarasa. "Clustering in weighted networks." *Social networks* 31.2 (2009): 155-163.

Chawathe, Sudarshan S. "Clustering Blockchain Data." *Clustering Methods for Big Data Analytics*. Springer, Cham, 2019. 43-72.

Meiklejohn, Sarah, et al. "A fistful of bitcoins: characterizing payments among men with no names." *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013.

Ermilov, Dmitry, Maxim Panov, and Yury Yanovich. "Automatic Bitcoin address clustering." *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017.

Zambre, Deepak, and Ajey Shah. "Analysis of Bitcoin network dataset for fraud." *Unpublished Report* (2013).

**What is the aim in your project? What algorithms will you implement?**

Implement a bitcoin clustering technique that able identify the category that the bitcoin address belong to, the categories are: exchange, Minors, gambling, services, etc. I will also try to identify the addresses that controlled by one users.

The algorithm will be build based on some heuristics about common spending and one time change, those heuristics based on analysis user behavior patterns.

**What type of data do you require from the database, on which blockchain?**

We require the blocks dataset in order to build the transactions and addresses graph, based on that graph we will start to implement the algorithm.

**What will be your results about? Addresses, transactions, blocks, clusters, etc.?**

It will be about identifying the categories that the bitcoin address belong, and the group of addresses that is controlled by single user.

**Are your results time dependent? A time dependent result is valid for a specific time period. For example, degree centrality of an address can be computed daily.**

Yes, we need to know the transaction coming in/out the addresses in a certain period of time. For example, the degree of the address node will help to build our project.

**What do we need to store from your results? (for example, the fact that “address a appears in the same cluster with address b” can be stored).**

We need to store the users cluster groups.

**What are your findings that can be visualized by the visualization group?**

Visualization of resulting clusters.

**Can your results be used as input in algorithms of other groups? (This estimate will be updated after you see algorithms that are being implemented by others).**

According to the existed groups no it cannot. Our work will help people who interested in understanding the blockchain network activities, enhancing trading strategies, preventing money laundering.