For project one, Tamika and I will bd reviewing a bitcoin data set on the snap website which rates bitcoin transactions on a trust/distrust level from a scale of 1 to 10. It’s a little bit about how bitcoin transactions are done. It uses peer to peer technology to manage transactions. There are some advantages to doing so. The peer to peer network creates A proof of work system. Users send and receive bitcoins, the units of currency, by broadcasting digitally signed messages to the network using bitcoin cryptocurrency wallet software. The signed transactions validates usage and creates a historical log for the network . All confirmed transactions are included in the block chain. It allows Bitcoin wallets to calculate their spendable balance so that new transactions can be verified thereby ensuring they're actually owned by the spender. The integrity and the chronological order of the block chain are enforced with cryptography.

The bitcoin reward that miners receive is an incentive which motivates people to assist in the primary purpose of mining: to support, legitimize and monitor the Bitcoin network and its blockchain. So little bit about how the bitcoin mining process works. The first step in this process is each minor node collects a new transaction from a blockchain. The miner then finds a proof of work code from the block, once the node finds a proof of work block, it performs mining task and broadcasts the work to all nodes. The receiving node then validates the transaction and only excepts the ones which are valid. When it comes to the transactions, there are two criteria‘s which determines a successful transaction. 1) the fees earned and received rating from a processing Nodes. In our activity, this rating is represented on a scale from 1 to 10. as you can see from the image of the note example below of the dataset below, you see the parent node, Child node, a trust rating and a timestamp to represent a transaction.

Our first step is to load the data into python and create both nodes and edges. Our goal here is to determine degree centrality, eigenvector centrality, and betweenness centrality as it pertains to trust/distrust ratings. We will then perform a t-test to determine the significance between the means of the distrust interest groups.

Here’s a summary of the centrality test we will be performing and how it impacts our results. degree centrality assigns an important score based on the number of links Held by each node. We will determine if it in if the score importance impacts the trust distrust results. Betweenness centrality measures the number of times but no lies in the shortest path between other notes we will analyze if the trust distress scores on a bridge path impacts the outcome of the receiving node.

Closest centrality and Eigenvector centrality scores each node based on the closeness to other notes in the network we will see if closeness impacts distress and trust outcomes. So here we can see a general summary of the date is that there is 1500 data points with four different index columns source target rating and time will primarily be focusing on source target and rating, we removed type timestamp for our analysis. Here is a basic view of the network. And here’s the general information of the network we can see that there are 381 notes with 874 edges. The largest diameter of the note is seven notes and here’s the top 20 nodes by degree.

Here we can see the top 10 betweenness centrality nodes and the corresponding centrality values together with the top 10 eigenvector nodes with the corresponding eigenvector centrality values. The network density value for this particular date if that is a .012 on a scale from 0 to 1 it says that this particular network isn’t dense at all.

So here I developed a graph of the top 250 notes which shows node centrality on a trust/distrust scale. As you can see from the Legend, the distrust trust values correspond with red to green Values. We can quickly determine that the larger nodes are closer to each other and have a positive trust rating than the smaller exterior nodes. Those are nodes Generally suffer from poor trust ratings. We can see two particular notes which have highly negative distrust values.

So let’s perform at Ttest to see the significance of difference between the means of the trust/distrust group. So what we done here is separated groups according to their distress and trust values. Here is a summary of both groups. Here we separated both groups according to the betweeness and eigenvector values from the distrust/trust groups. So let’s see now from a betweenness centrality t-rest off the distrust/trust groups whether the mean value is equal to zero or not equal to zero. We can see that the P value is much lesser than the .05 threshold we therefore reject the null hypothesis. We can conclude that there is in fact a difference between the mean values of the trust and distrust groups.

For our second test, Let’s see if the Eigenvector centrality between the distrust and trust groups equal to zero just like the other test performed. we can see that it is less than a .05 threshold as well. So we can conclude that there is not a difference between the eigenvector centrality between the trusted and distrust groups and therefore we reject the null hypothesis.

In both instances we concluded the The [t-test’s effect size](http://docs.statwing.com/examples-and-definitions/t-test/effect-size/) does not complements its statistical significance. We are unable to use the mean values as a comparison of distrust/trust groups because there is not enough data. Our recommendation is to compile more data to determine if in fact mean analysis can be performed.