DSC 190 Graph Networks: Final Project Proposal

The domain I would like to explore for the final project is related to cryptocurrencies and blockchains. I would like to do a wallet and transaction analysis on some of the top holders (monetary value) across the top blockchains to see how behaviour of these whales differs across different networks, also some degree and graph statistics and how they vary I would be very interested to compare. I plan to investigate the Bitcoin, Ethereum and Cardano blockchains. In the project I would have the wallets/address represent a node, a wallet could basically be synonymous with a user, the wallet/address is a fair representation of a node because this is where the cryptocurrency they control is 'held', the links/edges would be represented by a transaction or the act of transferring money to another wallet/address, in this project I would want weighted edges to represent the amount of value transferred and in this scenario I would also want a directed graph. There could be a scenario where may an undirected graph could be used if I didn't cary about the value transferred by for our exploration the monetary amount will be important. There are many block explorer websites that contain the addresses of the top N holders of a given blockchain,

Bitcoin: https://bitinfocharts.com/top-100-richest-bitcoin-addresses.html
Ethereum: https://www.coincarp.com/currencies/ethereum/richlist/
Cardano: https://www.coincarp.com/currencies/cardano/richlist/

I know I could webscrape but some of these sites do have an api or csv exports which could make the data collection process much easier. I think the data collection flow would be to first obtain the top N addresses or maybe the top N addresses that contain X proportion of the total supply of the network, this could be done though webscraping,copy/paste off the website or api, the next step would be to get all the transactions for the top N addresses through webscrape or api if available, but I believe many sites have a csv download of all the transactions for an address. The next step would be to figure out how to represent all data, what data structure would I use? I'm not sure yet myself but I'm eager to further explore this and think how would such an approach scale efficiently to larger datasets and not just the top N nodes. The number of nodes would be the top N nodes that control X % of the circulation of coins in a network, this could range from 100 - 10,000 depending on the X % and will vary depending on the blockchain. The number of links I'm honestly not to sure, it's the reason why I'm interested in doing this for the final project, some addresses might only have less than 10 transactions as they just buy and hold, others might have millions of transactions if they are an exchange wallet that is constantly processing transactions every minute.

The questions I'm most interested in asking is really evaluating key statistics and behaviour of the top wallets, even such simple questions as degree distribution in and out would give a lot of insight about these holders, average distance could tell how these holders interact with other users on the blockchain, I'm really interested to see how transaction behaviour changes with time, like in a bull market vs a bear market how to in and cout degree change, in this case could out degree indicate selling transaction and in degree represent buy transaction, I would want to append time data to the edge and maybe for a time frame we drop all edges that aren't in the time frame, how do key graph statistics change, I am super curious about how this changes in a bull/bear market. I would also like to explore maybe a

couple more advanced graph statistics for weighted graphs, I'm not sure what maybe we will learn later on in the class or I could look at another paper and try to apply it but maybe there are other local or global graph statistics that can be applied to this exploration.

The reason why this analysis matters is because if cryptocurrencies and their blockchain networks continue to gain mainstream adoption, the industry will need analytics and research done on these blockchains and the transactions they facilitate. Bitcoin was able to attain a 1 trillion dollar valuation at the peak of the 2021 bull market, and hundreds of billions of dollars of value was transacted on the bitcoin network alone. Due to the nature of blockchains being transparent ledgers, all of this transactional data is available for analysis, with so much money at stake factoring this data into quant models seems like an inevitability and some crypto quant groups and crypto ventures already perform large scale blockchain analysis to help inform their trades. I'm interested to see what indicators I could use from the largest holders of these networks to inform my own trading, what are the largest holders doing with their money, when are they buying and when are they selling. I would imagine with millions to billions of dollars at stake they might be some of the smartest traders/holders out there and perhaps I could learn something interesting.

I like your motivation and enthusiasm for this
topics that are relevant wil be in chapter 9 on communities and centrality measures to identify interesting nodes in the graph.
if you indeed have temporal information then exploring rich get richer phenomena
another thing you could explore are dynamics of events on the graph - are external economic events leading to chain reactions / spreading phenomena on the network (seems you have domain knowledge on this topic to explore 1-2 such events)