

BIA 658 SOCIAL NETWORK ANALYSIS
MIDTERM EXAM

NAME - SIDDHARTH PARAAG NILAKHE

There are two parts to this exam. Please note that you should refer to the Golbeck (book/videos) up to week 7 materials. If you use a reference online instead, be sure to include the reference. Wikipedia is not a valid reference.

Part I – Open Ended Questions (50 points total)

Note: Remember your answers should aim to be four sentences in length at minimum. Use more space for your answers as needed. You can hand draw examples if needed. I am a big supporter of sketches.

1. What is the topic you have selected for your semester project and why? **Note:** List the three topics/datasets from the selecting a dataset assignment ONLY if you need them to answer 2-4. Use your primary dataset choice whenever possible and the others in case you get stuck for the examples in 2-4.

Answer:

The three topics that I selected were - Movies, Social Media (Reddit) and Crypto. From these three topics, I have decided to go forward with Crypto as my topic for the semester project topic. There is a simple reason for selecting this topic over others I have been following the crypto market for a long time, and my interest in the field has inspired me to take this topic. I have been trading for 3 years now and I would really like to understand and study the network of the investment fund directory.

2. Explain what makes a social network dataset different than a traditional dataset such as Excel? Provide an example (from the 3 you listed in #1) that compares a social network dataset and a traditional dataset. Be sure to clearly mark the key points.

Answer:

Before understanding the difference between the two let us understand what is a social network dataset - Social Network Dataset in simple words means the data that is collected about relationships and connections. This relationship or connection can be between any two data points within the dataset, which in social network data we refer to as nodes in a network. Now moving to the question which is what makes social network datasets different from traditional datasets such as Excel, it is their focus on the structure of relationships. They show the connections and interactions among network entities rather than individual data points.

Example - In our Crypto dataset, we are more interested and focused on the relationship/connection between a firm and a cryptocurrency. Rather than focusing on individual entities such as only on the individual firm, making it different than a normal dataset. We are more focused on the relationship and different attributes such as the amount invested in the cryptocurrency is what makes the dataset unique and different.

3. Explain what density for use in a Social Network Graph is and how it is computed? Provide two examples (from the 3 you listed in #1) using the density formula (one example with low density and one example with high density).

Answer:

Density in a Social Network Graph is a measure of how interconnected the nodes in a graph are. It basically indicates how many connections actually exist between nodes compared to the total number of possible connections between the nodes.

For an undirected graph, density is computed as follows:

Density = (Number of actual ties) / (Number of possible ties) where,
Number of possible ties = (Number of nodes) * (Number of nodes - 1) / 2

And for a directed graph, density is computed as follows:

Density = (Number of actual ties) / (Number of possible ties) but here,
Number of possible ties = (Number of nodes) * (Number of nodes - 1)

Example -

High density - In our dataset, if many of the firms have invested in several common cryptocurrencies. If the majority of funds have connections (edges) to other funds due to shared investments, this would result in a high density. The formula, when applied to this network, would yield a value closer to 1, indicating strong connectivity.

Low Density - In contrast, consider a network where each firm has a unique set of investments, and there are few or no common investments between them. In this case, the connections between firms would be very few, resulting in a low density. The formula, when applied, would give a value significantly below 1, indicating limited connectivity.

4. Define what is meant by the shortest path. Explain why finding shortest path between nodes is important for social network analysis. Provide an example (from the 3 you listed in #1) of the shortest path.

Answer:

The shortest path simply means the minimum number of steps or edges in a graph required to travel from one node to another. It is just as simple as saying what is the shortest route that will take me from this place ABC to XYZ. Now answering the second part of the question of why finding this shortest path between nodes is important in our social network analysis, one reason is that it helps determine the most efficient route between nodes in the graph. As shorter paths indicate stronger or more direct type of connections, we can say that finding the shortest paths helps in understanding the strength and nature of relationships within the network.

Example - In our Crypto Dataset, Suppose we want to find the shortest path for Alphabit Capital (Firm) and AlphaCoin Capital based on common investments in crypto. Alphabit Capital has the following Investments → Videocoin, StormX, Aelf, and Intimate whereas AlphaCoin Capital has the following Investments → DaatX and Videocoin. Then the shortest path will be through the common investment in Videocoin. This is how we can find the shortest path.

Part II – Practical Application (50 points total)

This part requires you to think critically about your dataset. If you are considering changing your dataset, you should use the one I already approved for the exam. You can change your dataset with approval after the exam.

1. Why did you select your dataset for your semester project topic (i.e. what is the problem you are hoping to solve from the dataset you selected)?

Answer:

I selected the crypto investment fund directory dataset because of my three years of experience in trading crypto and stocks. My passion and keen interest in finance-related topics motivated me to explore this area further and conduct a detailed analysis of the investment relationships within the crypto investment fund network.

The dataset was selected to perform a Social Network Analysis to understand the investment relationships in the crypto investment fund network. The problem to be solved is to identify which investment funds have invested in which cryptocurrencies and analyze the patterns and structures within this network.

2. Explain your dataset in terms of basic demographics (descriptive statistics with up to 5 attributes). What type of statistical analysis do you plan to perform and what software will you use?

Answer:

As there are a lot of columns present in the dataset, these are a few of them that I am thinking of including in my analysis for now.

Firm Name	Country	Professional Staff	AUM (Millions)	Firm Type	Investments 1	Investments 2	Investments 3	Investments 4	Investments 5
1confirmation	United States	20	\$450	Hedge Fund	dyDx	Basecoin	MakerDao	0.00%	0.00%
360 Blockchain	Canada	3	\$450	Hedge Fund	Arcology	SV Cryptolab	Pressland	ePIC Blockchain	0.00%
3iQ	Canada	4		Hedge Fund	Ethereum	Bitcoin	Litecoin	0.00%	0.00%
8 Decimal Capital	United States	4		Venture	Bluzelle	0.00%	0.00%	0.00%	0.00%
AI8 Ventures	United States	4	\$55	Venture	Loyakk	0.00%	0.00%	0.00%	0.00%
AI8 Ventures	United States	4	\$10	Venture	Loyakk	0.00%	0.00%	0.00%	0.00%
Akuna Capital	United States	598	\$392	Hedge Fund	0.00%	0.00%	0.00%	0.00%	0.00%
Algo Depth LP	United States	1		Hedge Fund	0.00%	0.00%	0.00%	0.00%	0.00%
Algo.Land	Switzerland	4	\$7	Hedge Fund	FX	0.00%	0.00%	0.00%	0.00%
All in One Coin	Hong Kong	9	\$30	Hedge Fund	Bitcoin	Ethereum	Ripple	Bitcoin Cash	Litecoin
Alluminate	United States	6	\$200	Venture	Coinscore	CryptoPEts	0.00%	0.00%	0.00%
Alphabit Fund	UAE	7	\$20	Hedge Fund	VideoCoin	StormX	AELF	Intimate	0.00%
Alphacoin Capital	United States	7	\$50	Hedge Fund	DAATx	VideoCoin	0.00%	0.00%	0.00%

Firm Name: Names of various investment firms.

Country: The country where each investment firm is based.

Professional Staff: Number of people working for the firm.

AUM (Millions) - Assets Under Management: The total value of assets managed by each investment firm in millions of dollars.

Firm Type: The category or type of each investment firm (e.g., Hedge Fund, Venture).

Investments 1, Investments 2, Investments 3, Investments 4, Investments 5: The top five investments made by each investment firm, which may include cryptocurrencies, assets, or other investment vehicles.

I have planned to perform the following statistical analysis for now-

- 1) Computing measures like mean, median, mode, minimum, maximum, and standard deviation for attributes such as AUM, Professional Staff, and more. This provides a snapshot of the dataset's variability. It tells us a lot more about the data than we know.
- 2) Examining the correlation if any between attributes like AUM and Professional Staff to understand if there's a relationship between the size of assets under management and the number of staff members.

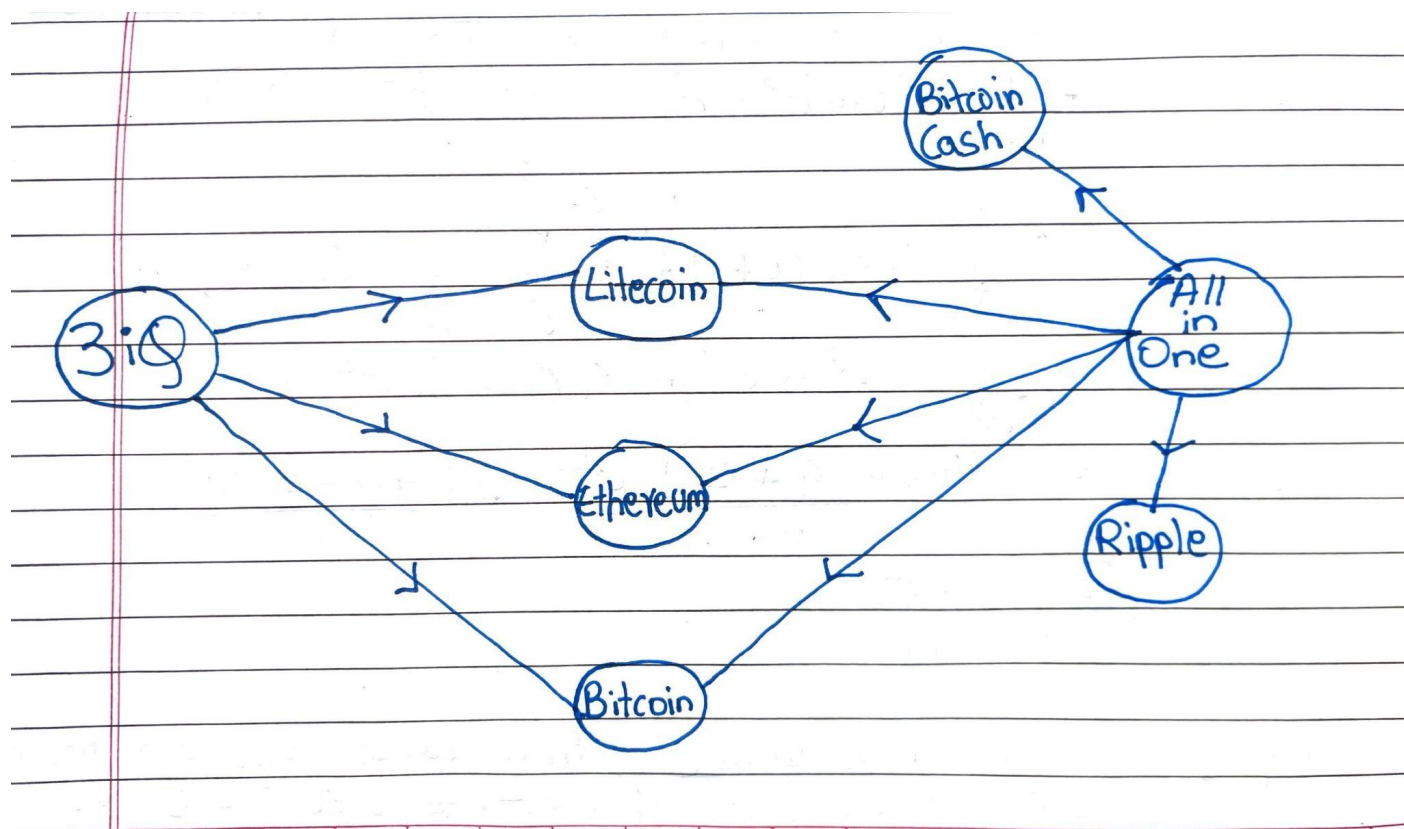
3) Comparing attributes across different groups. For example, you can compare AUM levels between different firm types or analyze how the number of professional staff varies by country.

Software - I will be using R or Python for the statistical analysis. I will also be using Python for basic EDA. To analyse and create the social network graph, I will be using Gephi.

3. Draw a sample of the social network diagram of the nodes and edges you selected. You do NOT need to use any software, you can just draw it by hand and paste it into this exam if that helps. Please refer to the early chapters of the nodes/edges. Remember you ONLY want 1 or 2 types of nodes (attributes) and a single edge (attribute). The examples with Tom Hanks and the movies are the image you want with your own node/edge selection. This is early thinking and will change later which is fine.

Answer:

In the diagram, we have considered Firms and different cryptos/investments as nodes, and the investment made by the firms in these cryptos as edges.



Social Network Diagram