Assessing the Impact of ChatGPT on Various Fields

Based on Network Analysis of Twitter Data

Mahsa Shahbazi (2049588), Seyedali Hosseinishamoushaki (2041399), MirMohammadreza Heidarzadnamin (2041430)

July, 2023

Abstract - **This paper talks about new changes in technology, especially in the world of Artificial Intelligence (AI), and focuses on ChatGPT. ChatGPT has caused a lot of discussion, with people having both positive and negative views. Some people are even worried that it could be a threat to people. But there are also many positive views about how ChatGPT could be used in different areas and how it could affect other new technologies. The paper looks at these views and tries to understand which technologies could be most affected by ChatGPT. The authors use Twitter to gather data because it's a big platform for sharing opinions and news. They use Twitter's hashtag system to find tweets about specific topics. The goal of the research is to find important links around a specific topic using hashtags, and to look at the connections between them.**

Keywords – **Artificial Intelligence, ChatGPT, OpenAI, Twitter, Hashtag, Technological Advancements.**

I. INTRODUCTION

In the present era, marked by significant technological advancements, particularly in the realm of Artificial Intelligence (AI), we're witnessing the rise of breakthrough technologies like ChatGPT. This technological wonder represents a critical milestone in the AI landscape and is projected to significantly influence the future direction of AI and emerging technologies.

The introduction of ChatGPT stirred a wide range of reactions, with the initial months seeing a heated debate among the public. Social media platforms became a hotbed for a myriad of perspectives, with some expressing immense optimism, while others showed profound pessimism about the implications of such a powerful AI technology. Some even voiced concerns about the potential threat this could pose to the human race!

Despite the pessimistic outlook, there were numerous optimistic viewpoints about ChatGPT's deployment in various sectors and its potential impact on other tech innovations. We found it intriguing to explore these perspectives and identify which technologies, as per many social media participants, could experience the greatest impact from ChatGPT and related technologies.

For this project, we elected to use Twitter as our primary source of data, given its standing as one of the largest opinion, news, and announcement sharing platforms. One of Twitter's distinct features is its hashtag system, which links tweets from various users under a common theme. This provides us with a streamlined view of tweets relating to a specific topic. Our objective in this research is to pinpoint key relationships around a specific topic using hashtags, and to analyze the links among them, including mentions and the like.

II. METHODOLOGY

Our goal is to discern relationships between hashtags and mentions within tweets by formulating and analyzing graph relationships.

This will hopefully allow us to pinpoint the most significant hashtags associated with a given topic and the user accounts that concentrate on that topic.

The entire procedure unfolds in three phases, as illustrated in the diagram below:

1. Stage of data gathering and preprocessing
2. Isolation of relationships between hashtags and mentions to identify the most crucial nodes in the graph.
3. Exploration and identification of the most important nodes based on the results from the preceding phase.

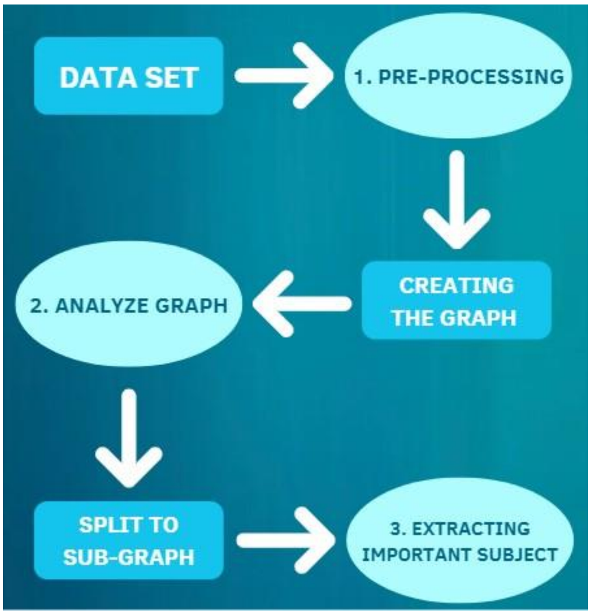

Figure 1. The procedure steps.

III. STEP 1

In this research, the dataset we used was compiled by sourcing recent tweets relevant to the topic of ChatGPT via the Twitter API. The dataset, containing 100,000 tweets, encompasses various tweet attributes including the count of likes, references, mentions, tags, and the text itself. Due to the constraints of the Twitter API's free access, these were accumulated over a span of 15 days. The choice to concentrate on popular hashtags associated with ChatGPT was driven by the topic's high visibility and significance, along with its influence on other subject areas.

- **CREATING THE GRAPH**

In order to extract hashtags and mentions from each tweet, we identified corresponding tags for every mention. Subsequently, we viewed each pair of hashtags and mentions as an edge within the graph, where the edge's weight hinges on the frequency of the pair's occurrence in the dataset. Hence, the edge associated with the highest frequency in conjunction with a specific mention carries a greater weight. This denotes a higher co-occurrence frequency of the hashtag and mention pair.

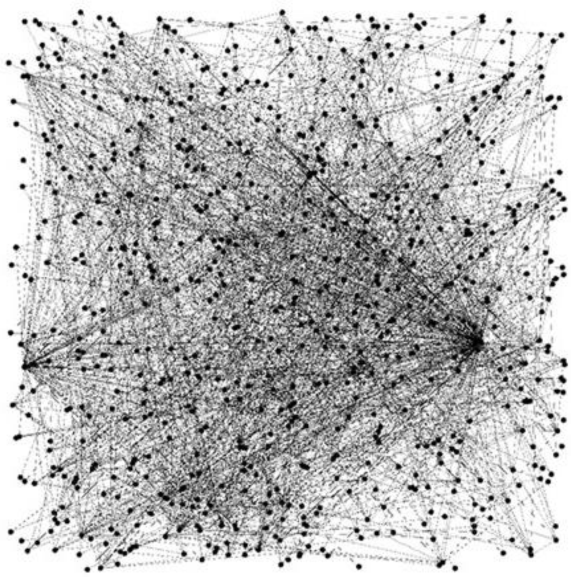

Figure 2. A graph created from the relationships between tags and mentions, which contains 2000 edges and 768 nodes.

IV. STEP 2

To identify the largest subgraph comprising indirectly connected nodes, we utilized the Connected Components algorithm, considering the graph as directed to facilitate this process. Our findings revealed that the largest indirectly connected set encompassed more than 90% of the nodes within the graph. This substantial connected set serves as our search range for identifying the most significant connections. As a result, our graph was partitioned into 15 distinct independent graphs. In the provided figure 3, the red graph represents the main graph, while nodes connected in different colors were deemed insignificant and subsequently removed. The rationale behind removing smaller nodes is that some individuals may have tagged their friends or employed unrelated hashtags that bear no relevance to our primary topic. By eliminating these smaller graphs, we were able to construct a unified graph solely consisting of the larger nodes that were either directly or indirectly interconnected.

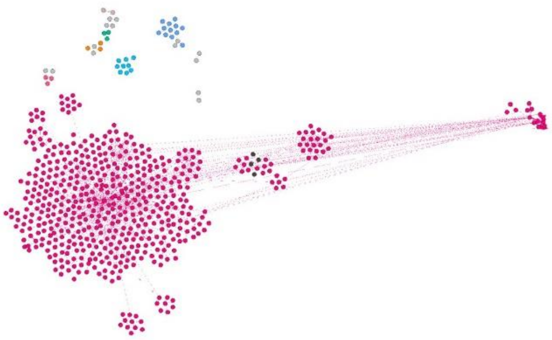

Figure 3. Connected Components Network.

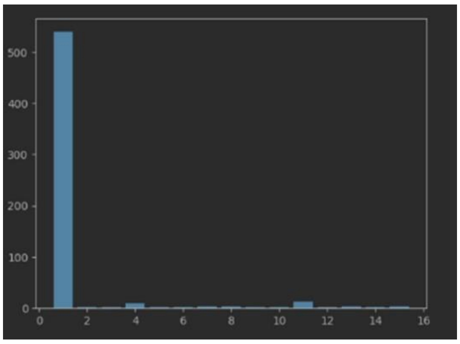

Figure 4. chart of the number of nodes within each set.

- **DEGREE DISTRIBUTION**

To evaluate the significance of a node, we employed the degree distribution approach, which involved examining the distribution of input and output edge weights to a node. Nodes with higher values in this distribution can be regarded as our target nodes of interest.

By analyzing the degree distribution chart, we observed that the average degree distribution was approximately 20. In the accompanying word cloud, nodes with darker colors and larger sizes indicate higher degree distributions. To gain a more comprehensive understanding of this matter, we conducted separate analyses of the input and output degree distributions for hashtags and mentions.

In constructing the graph, we employed a directed graph structure where hashtags are linked to mentions, with no direct connections between hashtags themselves. Consequently, we examined mentions in the input degree distribution analysis and hashtags in the output degree distribution analysis.

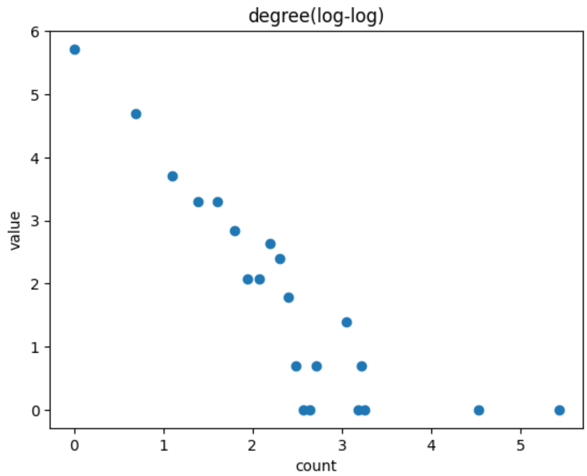

Figure 5. Distribution chart of the variation of degree values.

Based on the word cloud of mentions, it is evident that user accounts such as OpenAI and Trust Wallet exhibit higher input degree distributions compared to other user accounts. This suggests that these accounts have been mentioned more frequently in the dataset. Similarly, upon examining the word cloud of hashtags and the output degree distribution chart, we can observe that ChatGPT2, GPT, and Artificial Intelligence are among the top output degree distributions, indicating their prominence and frequency of occurrence in the dataset. The specific values for each of these nodes are presented in the table below. Through the analysis of the degree distribution, we can effectively identify the nodes corresponding to important hashtags and user accounts.

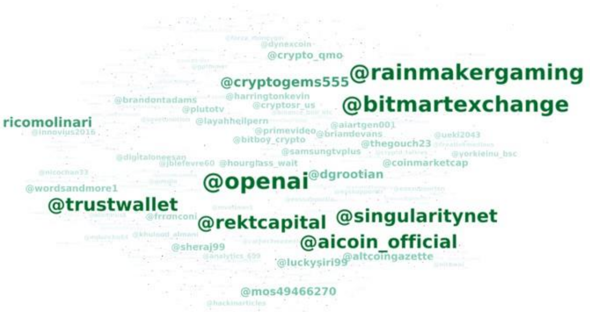

Figure 6. Displaying word cloud based on in-Degree, where the bigger size of the words indicates the higher score.

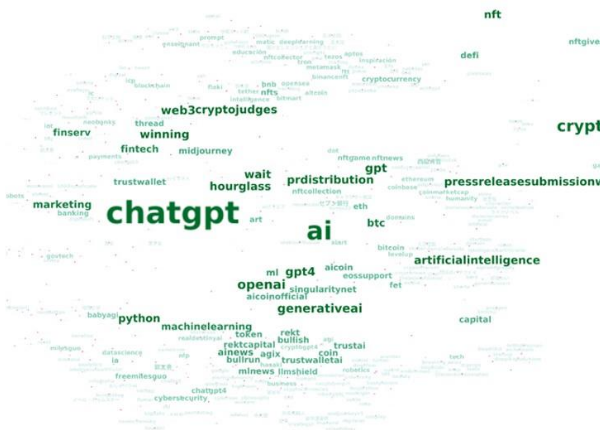

Figure 7. Displaying cloud words based on Out-Degree, where the bigger size of the words indicates the higher score.

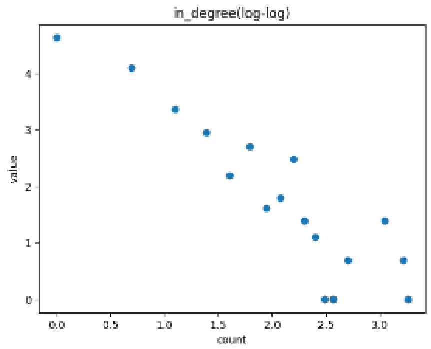

Figure 8. In-Degree distribution chart.

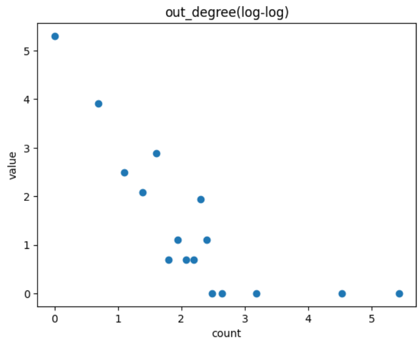

Figure 9. Out-Degree distribution chart.

Id	Label	Interval	In-Degree
@openai			26
@rainmakergaming			25
@bitmartexchange			25
@aicoin_official			21
@trustwallet			21
@rektcapital			21
@singularitynet			21
@enricomolinari			15
@cryptogems555			15
@dgrootian			13
@mos49466270			12
@luckysiri99			11
@altcoingazette			11
@crypto_qmo			11
@wordsandmore1			10
@sheraj99			10
@coinmarketcap			10
@thegouch23			10
@aiartgen001			9
@frronconi			9
@bitboy_crypto			9
@hourglass_wait			9
@harringtonkevin			9
@brandontadams			9

Table 1. Score of in-degree for each node.

Id	Label	Interval	Out-Degree
chatgpt			228
ai			93
crypto			24
openai			14
generativeai			12
web3			11
gpt			11
gpt4			11
artificialintelligence			10
hourglass			10
wait			10
prdistribution			10
pressreleasesub...			10
winning			10
cryptojudges			10
nft			9
python			9
marketing			8
btc			8
finserv			7
fintech			7
machinelearning			7
ml			6
ainews			6
defi			5
metaverse			5
midjourney			5
trustai			5
trustwalletai			5
aicoin			5

Table 2. Score of in-degree for each node.

- **PageRank**

PageRank is indeed a crucial algorithm in graph analysis, originally developed by Google, for determining the most influential nodes based on their frequent references. By employing PageRank, we can identify nodes that receive the highest number of referrals. The PageRank value assigned to each node indicates its significance and score in terms of referrals. Consequently, nodes with higher PageRank scores are more likely to be referenced by other nodes.

In our study, we applied the PageRank algorithm to the nodes within the largest component obtained in the initial step, represented by the red graph in the figure 3. By combining PageRank analysis with this largest component, we were able to select the nodes with the highest PageRank scores. The image below demonstrates that the nodes with high PageRank, denoted by their larger font size, are situated within our largest component.

Figure 10. Displaying word cloud based on PageRank, where the bigger size of the words indicates the higher score.

Id	Label	Interval	PageRank
chatgpt			0.119964
ai			0.034471
@nickypromotes			0.0272
@digitaloneesan			0.016084
@enricomolinari			0.011813
@openai			0.011588
@luckysiri99			0.009494
@mos49466270			0.008808
crypto			0.008663
@altcoingazette			0.007749
@ueki2043			0.007155
@savetonotion			0.007071
@wordsandmore1			0.007035
@rainmakergaming			0.006987
@bitmartexchange			0.006911
@emadciler			0.006603
@hackinarticles			0.006318
nft			0.005985
@khulood_almani			0.005773
@dgrootian			0.00529
@mdancho84			0.005205
@yorkieinu_bsc			0.005048
@eossupportcn			0.004901
@aiartgen001			0.004599

Table 3. Score of page rank for each node.

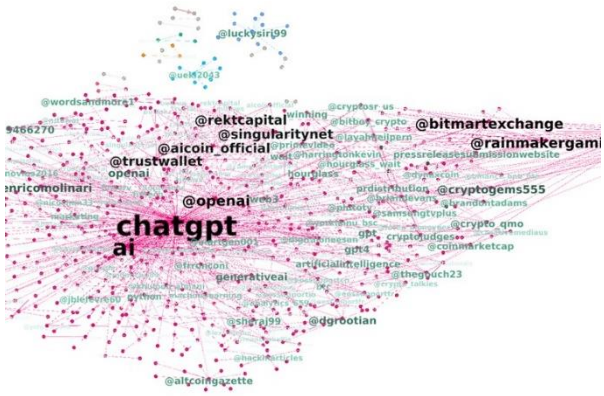

Figure 11. Displaying word cloud on the master Graph.

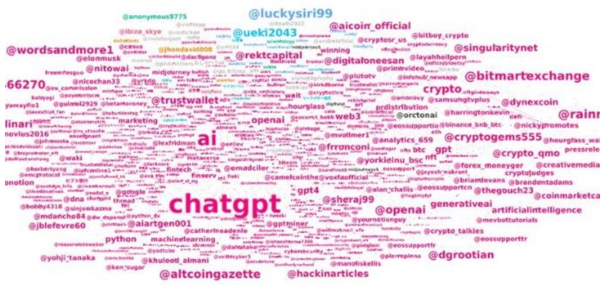

Figure 12. Combining PageRank with connected components highlights related nodes by color and demonstrates their importance through word size.

Upon observing the depicted image (Figure 12), we can ascertain that nodes of the same color are contained within the largest connected set. Consequently, we regard these topics as the most influential ones associated with ChatGPT, as they possess interconnections. Examples of such topics include AI, crypto, GPT4, @altcoin, @bitmarket, and @trustwallet.

However, it is worth noting that nodes like @lucksiri99 and @ueki2042 possess high PageRank scores but do not belong to the chosen connected component. As a result, they are not relevant to our selected topic.

- **MODULARITY**

We used a module to cluster the nodes of the graph based on the weights of the edges. Clustering based on weights allows us to separate the largest clusters and select the most important nodes within each cluster to combine with the largest component set. The cluster size distribution chart shows that all nodes are clustered into 36 clusters, with an average of 15 nodes per cluster. However, the third cluster has different sizes than the average clusters, with more than 50% of the nodes in this cluster.

Figure 13. Diagram of the distribution of nodes within each cluster (Randomize: On, Use edge weights: On, Resolution: 1.0)

In this stage, we focused on identifying the five clusters that encompass the greatest number of nodes. These clusters were chosen based on their higher node count compared to the average node count across all clusters. Upon examining the selected nodes within these five clusters, we found that three clusters stand out as the most important:

1. Red Cluster: This cluster consists of nodes such as OpenAI, Artificial Intelligence, GPT4, and Python.

2. Green Cluster: The nodes within this cluster include Web3, Crypto, and Marketxchange.

3. Purple Cluster: This cluster comprises nodes like Trustwallet, Trustai, and Tokenai.

Figure 14. Five clusters that encompass the greatest number of nodes.

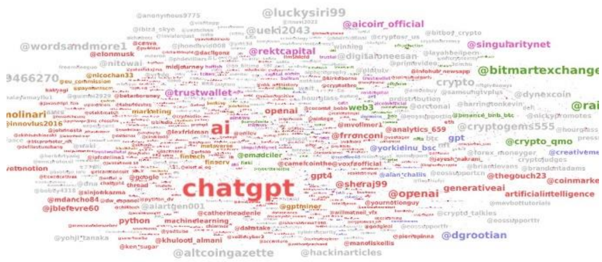

Figure 15. Word cloud were color-coded based on the clustering results to aid in understanding the members of each class.

V. STEP 3

In this step, we construct a new graph by merging the nodes identified in the previous step with nodes possessing the highest PageRank scores, located within the significant clusters obtained earlier. The inclusion of nodes is determined based on their incoming and outgoing degree values obtained in the previous step.

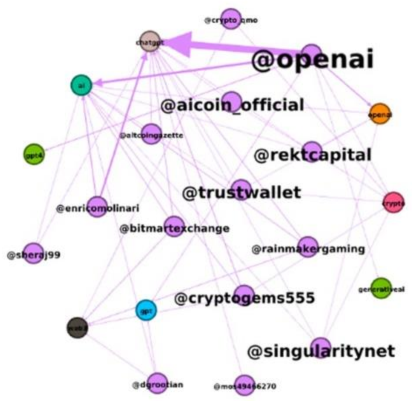

Figure 16. Selected subgraph.

To distinguish hashtags from nodes, we employ the Hubs algorithm. By evaluating the nodes' centrality, we identify those that play a central role in the graph. Through analysis of the resulting chart, we directly determine the centrality of the nodes. Six nodes are recognized as central nodes, with one of them exhibiting the highest centrality. This particular node is chosen as the target node, and we proceed with graph clustering using modularity, without considering the graph's weights. The cluster encompassing our target node is then designated as the target cluster.

Id	Authority	Hub
@openai	0.0	0.387742
@aicoin_official	0.0	0.343766
@singularitynet	0.0	0.343766
@rektcapital	0.0	0.343766
@trustwallet	0.0	0.343766
@cryptogems555	0.0	0.321286
@bitmartexchange	0.0	0.248375
@rainmakergaming	0.0	0.248375
@sheraj99	0.0	0.212249
@enricomolinari	0.0	0.212249
@dgrootian	0.0	0.159538
@crypto_qmo	0.0	0.109037
@altcoingazette	0.0	0.107031
@mos49466270	0.0	0.107031
chatgpt	0.587005	0.0
ai	0.577064	0.0
crypto	0.39988	0.0
openai	0.321417	0.0
web3	0.198128	0.0
gpt	0.099788	0.0
gpt4	0.070698	0.0
generativeai	0.070698	0.0

Table 4. Hub scores for each node.

In our research, we used the Eigenvector Centrality to identify influential nodes and important elements in the network.

Eigenvector Centrality is a measure of the influence of a node in a network. It extends the concept of degree centrality by taking into account not just the number of links that a node has (i.e., its degree), but also the quality of those links. In other words, a node is considered more important if it is connected to other important nodes.

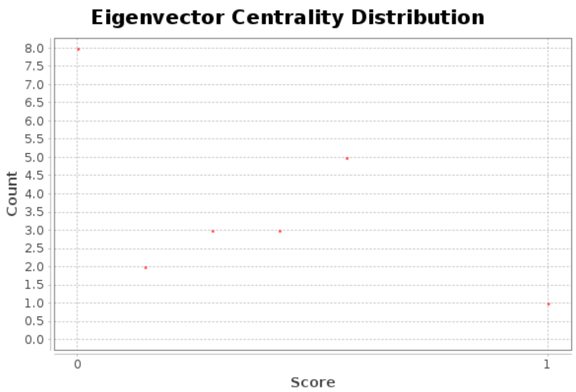

Figure 17. Centrality score distribution diagram (Network Interpretation: directed, Number of iterations: 100)

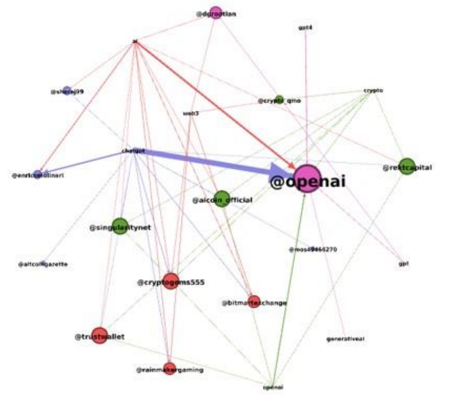

Figure 18. Apply centrality score on nodes to view nodes with higher score in larger size

We used Modularity for this subgraph individually to partition the network into communities of nodes with high internal connectivity and low external connectivity.

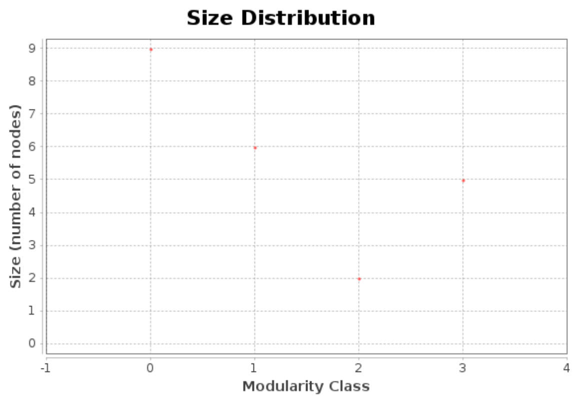

Figure 19. chart of the distribution of nodes within each cluster.

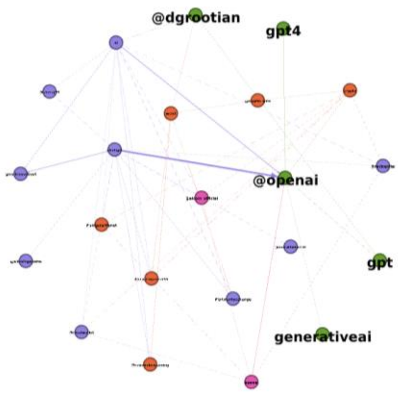

Figure 20. From the set of 22 elements distributed among 4 clusters, we select nodes specifically from the chosen class.

chatgpt	@openai	976
ai	@openai	311
chatgpt	@enricomolinari	253
openai	@openai	136
ai	@enricomolinari	110
gpt4	@openai	87
chatgpt	@cryptogems555	82
ai	@rainmakergaming	65
chatgpt	@bitmartexchange	65
chatgpt	@rainmakergaming	65
web3	@rainmakergaming	65
ai	@bitmartexchange	64
web3	@bitmartexchange	64
ai	@cryptogems555	54
gpt	@openai	47
web3	@cryptogems555	46
chatgpt	@altcoingazette	41
ai	@dgrootian	40
gpt	@dgrootian	40
web3	@dgrootian	40
crypto	@crypto_qmo	39
crypto	@openai	39
web3	@crypto_qmo	39
ai	@aicoin_official	38
chatgpt	@sheraj99	38
openai	@aicoin_official	38
ai	@sheraj99	37
chatgpt	@mos49466270	37
crypto	@cryptogems555	35
generativeai	@openai	33

Table 5. Selected pair (mention, hashtag) edges with value.

VI. CONCLUSION

Based on the outcomes of Step II and the processing conducted in Step III, it is evident that ChatGPT exhibits a substantial direct connection with the domains of artificial intelligence and cryptocurrencies. Analysis of annotated tweets reveals that the most prevalent mentions and hashtags are associated with cryptocurrencies. This indicates that, as per social media users, the innovative ChatGPT technology is poised to exert a significant influence on the future of this field.

Business taxonomy	4064
Entities [entity service]	3411
Technology	3097
Brand	2646
Product	1958
Digital assets & crypto	716
Brand vertical	476
Interests and hobbies'	290
Politician	241
Stocks	194

Table 6. The most significant topics, based on their frequency of occurrence.

Moreover, when examining the selected topic of searching the hashtag #chatGPT on Twitter, the findings reveal a notable emphasis on user accounts and hashtags directly associated with this user account. Noteworthy hashtags in this context encompass 'gpt', 'gpt4', '@openai', and 'generativeai'.