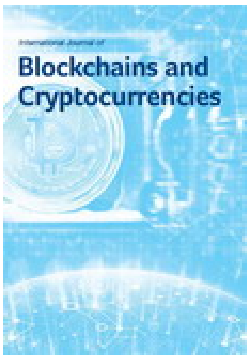


International Journal of Blockchains and Cryptocurrencies

ISSN online: 2516-6433 - ISSN print: 2516-6425
https://www.inderscience.com/ijbc

Advanced techniques in profiling cryptocurrency influencers: a review

Sijie Yu, Jon Padfield

DOI: 10.1504/IJBC.2024.10065113

Article History:

Received:	06 February 2024
Last revised:	31 May 2024
Accepted:	31 May 2024
Published online:	08 July 2024

Advanced techniques in profiling cryptocurrency influencers: a review

Sijie Yu* and Jon Padfield

Purdue Polytechnic,
Purdue University,
Seng-Liang Wang Hall, 516 Northwestern Avenue,
West Lafayette, IN 47906, USA
Email: yu1239@purdue.edu
Email: jpadfiel@purdue.edu
Website: https://polytechnic.purdue.edu/profile/jpadfiel
*Corresponding author

Abstract: Since 2023, the surges in cryptocurrency markets have significantly increased influencer activities on social platforms. However, research on influencers engaged with cryptocurrency-related topics remains sparse. This paper explores recent research accessible from leading academic search engines, focusing on profiling and identifying cryptocurrency influencers on Twitter (X). It analyses scholarly articles that discuss influencers' classification, profiling, and analysis based on various platform statistics, psychological features, tweet content, social connectivity, and crypto price fluctuations. Additionally, the paper explores the emerging decentralised SocialFi platforms that evolved from Twitter (X), examining the unique monetisation models that shape influencers there. Through an extensive review of relevant research, this paper furnishes business and legal leaders with a robust technical framework to identify and understand cryptocurrency influencers.

Keywords: profiling cryptocurrency influencers; cryptocurrency influencers; Twitter influencers; cryptocurrency Key Opinion Leaders; SocialFi Key Opinion Leaders.

Reference to this paper should be made as follows: Yu, S. and Padfield, J. (2024) 'Advanced techniques in profiling cryptocurrency influencers: a review', *Int. J. Blockchains and Cryptocurrencies*, Vol. 5, No. 5, pp.1–18.

Biographical notes: Sijie Yu is pursuing her PhD at Purdue University, focusing on big social data analysis. She also holds a Researcher and Lecturer position at the DigiPen Institute of Technology in Singapore, applying her data science expertise.

Jon Padfield, PhD is an Assistant Professor at Indiana University and an Adjunct Professor at Purdue University, where he teaches graduate classes in Data Analytics and undergraduate classes in Operations Management and Supply Chain Management for the Division of Business. He is also the President of Proffer Brainchild Analytics and Innovation, which offers training and consulting focused on improving quality and productivity.

1 Introduction

The cryptocurrency landscape has undergone a remarkable transformation since 2023, marked by events like the Bitcoin spot ETF approval reported by Wade (2024), a $4.3 billion settlement between American authorities and AZCoinNews (2023), and investor anticipation of an Ethereum ETF as noted by Kharpal (2024). These developments have catalysed significant interest in cryptocurrency, prompting individuals to seek insights from social media platforms, particularly Twitter (X). A small group of Twitter users, known as 'influencers', holds substantial sway over large audiences seeking guidance. Their influence, which we aim to dissect in this research, is profound, often impacting digital coin prices and shaping market trends.

An example of this influence is Elon Musk's tweet about Dogecoin in 2021. As Benson (2022) described, Elon's tweet referring to Dogecoin as 'the people's crypto' with a 'Lion King' photograph caused a substantial rise in its price from under a penny to an all-time high of $0.73, reflecting a gain of over 7200%. Similarly, influencers are critical in marketing products such as non-fungible tokens (NFTs), which benefit significantly from online exposure. Peter (2023) highlighted that influencers' extensive social media reach can effectively broadcast NFTs to potential buyers.

However, the ethical and regulatory aspects of influencers are increasingly scrutinised. For instance, influencers are required to disclose commercial affiliations when promoting tokens. Stark et al. (2023) reported that the SEC filed charges against eight high-profile social media individuals, including Jake Paul and Lindsay Lohan, for promoting cryptocurrency tokens without transparently disclosing compensation. This underscores the importance of regulatory compliance in influencer marketing. Parikh (2023), SEC (2022), and Wang (2022) reported similar cases of FTC and SEC enforcing regulations against influencers' recent 'pump-and-dump' schemes.

Given the substantial influence of influencers on cryptocurrency market trends, regulatory compliance, and user behaviour across platforms, gaining a deep understanding of these influencers is critical. This paper serves as a detailed technical guide for profiling and identifying cryptocurrency influencers and analysing their tweet contents and authorship patterns. It reviews various technology papers, highlighting their use of features, implementation of statistical data analysis, and content analysis through large language models (LLMs). Additionally, this paper delves into the activities of Twitter influencers on decentralised SocialFi platforms on blockchain, emphasising monetisation models used to profile them.

1.1 Research method

This paper aims to comprehensively review recent methodologies for profiling and identifying cryptocurrency influencers on Twitter (X) and Twitter-linked SocialFi platforms. The research method employed in this review is as follows:

1 Search Terms: Primary search terms included 'cryptocurrency influencers', 'profiling cryptocurrency influencers', 'Twitter influencers', 'cryptocurrency Key Opinion Leaders', and "SocialFi Key Opinion Leaders".

2 *Sources*: Relevant literature was gathered using academic search engines such as Science Gate, Google Scholar, and IJBC.

3 *Delimiters*: The search was limited to papers published between 2010 and 2024 to capture recent advancements. Only English-language papers were included.

4 *Inclusion criteria*: Studies focusing on identifying and profiling cryptocurrency influencers on Twitter (X) and its linked SocialFi platforms. Papers discussing statistical features, descriptive data analysis, linguistic content analysis, and network algorithms were prioritised.

5 *Exclusion criteria*: Papers that did not specifically address cryptocurrency influencers from a technological perspective were excluded.

1.2 Manuscript specifications and contributions

The review was systematically conducted using academic databases, with most papers published between 2010 and 2024 to include the latest technical methodologies. The scope encompasses technological papers that directly address the profiling and identification of cryptocurrency influencers, particularly those utilising platform statistics, network algorithms, data analytics, and LLMs. A significant contribution of this work is introducing a comprehensive list of methods combining statistical, data analysis, LLMs, page ranking, and decentralised monetisation approaches. This multidimensional view provides an understanding of cryptocurrency influencer dynamics, enabling the detailed identification of influencers based on platform metrics, content, and monetisation within the cryptocurrency community on Twitter (X).

2 Identification from statistical metrics

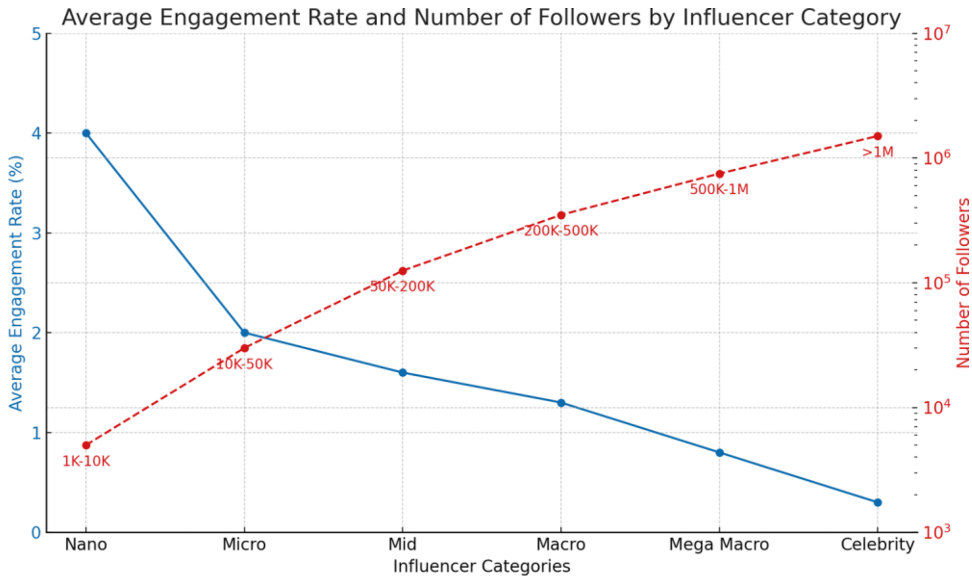
Bevendorff et al. (2023) define non-influencers as users having fewer than 1K followers or 300 posts. The rest are influencers who can be further categorised according to platform statistics. The statistical metrics include follower count, subscriber numbers, or unique visitors to their posts. Pereira (2022) illustrates a simple model to categorise influencers into six tiers using follower counts. The first tier is Nano Influencers, having 1K to 10K followers and the highest engagement rate. Engagement rate is a crucial metric measuring influencer impact used by most marketers. Notably, the engagement rate tends to decrease as the influencer tier increases. Warren (2021) introduces one way to measure influencers using engagement rate: by using audience interaction with the content of tweets. It is calculated as below,

$$Engagement\ Rate = \frac{Total\ Engagements}{Total\ Followers}$$

where Total Engagement can be calculated as the sum of likes and comments, a higher engagement rate suggests a frequent interaction of tweeters with their followers.

The next tier is Micro Influencers with follower counts between 10K and 50K. Mid-tier Influencers have follower counts between 50K and 200K. Macro Influencers have follower counts between 200K and 500K. Mega Influencers have between 500K and 1M followers. Celebrities have over 1M followers, which is an extensive reach. Figure 1 shows tiers, their statistics, and engagement rates. It reveals that the average engagement rate steadily declines as an influencer's follower count increases. Nano Influencers have the highest average engagement rate, with 4% across all posts.

Figure 1 Engagement rate across tiers (see online version for colours)

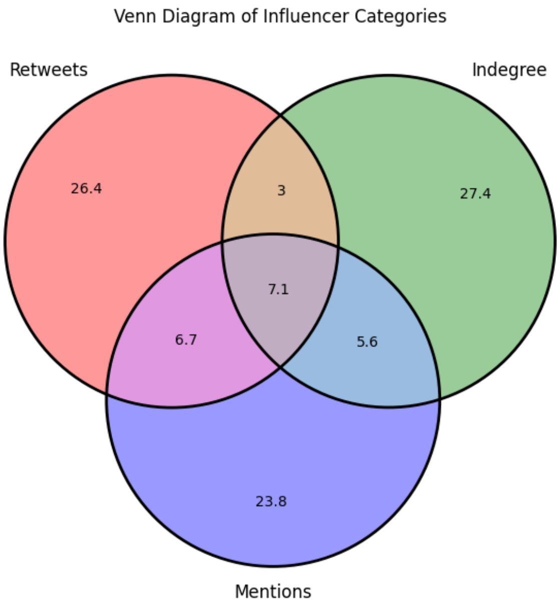


There are more accurate categorisation approaches than the tier model. Cha et al. (2010) used three statistical metrics to categorise influencers: indegree, retweets, and mentions, each highlighting unique influence aspects. Here, indegree indicates the popularity and equals the follower count. Retweets reflect the value of a user's content by how often it's reshared. Mentions, showing name value, count how frequently others reference a user. This study states a good influencer identification algorithm should consider all the factors above. There are interesting observations from the data,

- Influencers with many followers do not necessarily get more retweets or mentions.
- Influence is not granted spontaneously or accidentally, but through continued efforts. Ordinary users can become influencers by concentrating on specific topics and sharing creative, insightful content.
- Indegree is effective in measuring attention from followers through one-on-one interactions.
- Retweet represents influence beyond one's one-on-one interaction domain.
- Mentions mainly were celebrities.

Figure 2 shows the top 100 influencers ranked by each statistic indicator from the experiment data, with the total normalised to 100%. The overall overlapping is only 7.1%, and fewer influencers are ranked commonly across two or three indicators. The graph proves measures capture the unique characteristics of influencers.

Figure 2 The top 100 influencers ranked by each measure (see online version for colours)



3 Identification from other features

Lichti et al. (2024) used an opinion leader index (OLI) to select influential bitcoin opinion leaders (BOLs). The OLI is computed from statistics and features like audience engagement, niche alignment, reputation, audience reach, activity, and consistency. Lichti also differentiated influencers from BOLs: influencers gain influence primarily through their media presence, and BOLs develop theirs from expertise. However, we review them as the same group of users since many studies often conflate the two. Table 1 lists six psychological features used to assess OLI. A user qualifies as a BOL by meeting at least three criteria.

Based on the features mentioned in Table 1, influencers can be classified into eight categories:

1 Engagement Gurus, listed as the top 50 users according to h-index, e.g., Carl Runefelt. Kozul (2023) described him as a cryptocurrency educator and blogger.

2 Bitcoin Maximalists, listed in Bitcoin maximalists, e.g., Tone Vays. Vays (2014) described him as a cryptocurrency trader, financial educator, and blogger.

3 Crypto All-Stars, listed in crypto influencers, e.g., Vitalik Buterin and Retimuko (2024), highlighted him as the founding father of Ethereum.

4 Millionaire Magnets, with over a million followers, e.g., Elon Musk.

5 Bitcoin Conversationalists, with over 3000 Bitcoin tweets, e.g., Randy Hilarski (2019), a cryptocurrency advocator and educator active across social platforms.

6 Persistent Pundits, active for over nine years, e.g., Jeff Garzik. Garzik (2023) describes him as a cryptocurrency entrepreneur, technologist, and advocator.

7 Confrontational Conversationalists, predominately Bitcoin critics, e.g., Peter Schiff, Tarnishedpath (2023), described him as a stockbroker, entrepreneur, financial commentator, and radio personality.

8 Incognito Influencers, with limited personal disclosure, e.g., PlanB. Edyme (2024) describes him as a mysterious cryptocurrency technologist and commentator.

Among these categories, Bitcoin Maximalists are the most strongly qualified influencers, with the highest total indicator scores, followed by Crypto All-Stars and Millionaire Magnets. Confrontational Conversationalists and Incognito Influencers are the least strongly qualified influencers. Figure 3 summarises each category's average score of BOL criteria.

Table 1 Six Bitcoin OLI psychological features

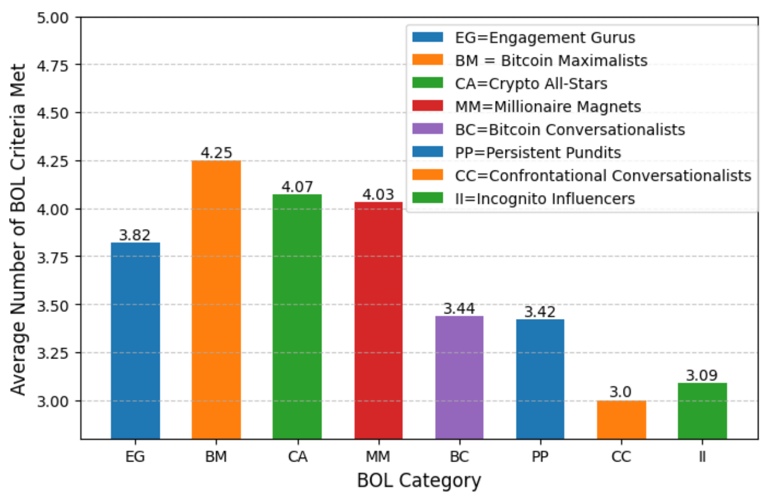
Indicator	*Description*	*Source*
Audience engagement	The annual average Hirsch index (h-index[1]) is at the top 200 and has at least 10K followers	Hirsch index
Niche alignment	Included at least one time as a Bitcoin maximalist[2]	Online references
Reputation	Is listed at least four times as a Crypto influencer[3]	Online references
Audience	Has at least 1M followers	Platform statistics
Activity	Has directly tweeted about Bitcoin at least 100 times	Platform statistics
Consistency	Has directly tweeted about Bitcoin for at least three years	Platform statistics

[1]Hirsch (2005) defines the h-index as a metric assessing academic publications' productivity and citation impact. It is calculated as the maximum value of h such that the given author has published h papers that have each been cited at least h times. Retweets approximate citation.

[2]Bitcoin maximalists are identified from sources like the list of Bitcoin Maxis from CryptoSlate (2024).

3. Crypto influencers are sourced from references like the list of influential people in the crypto space by AlgoBlocks (2023).

Figure 3 The average total met BOL criteria across categories (see online version for colours)



4 Linguistic content analysis

Lichti et al. (2024) also analysed tweet contents and identified eight BOL key topics:

1 money

2 technology

3 future focus

4 power

5 politics

6 risk

7 reward.

They extracted psychometric features from texts to measure writer discourse styles and thematic focuses to classify tweets. The features are analytical thinking, clout, emotional tone (sentiment), cognitive processes, and social dynamics. Analytical thinking indicates influencers can understand the complex technical nature of cryptocurrency. Cloud, emotional tone, cognition, and social processes identify clout or high-status influencers, and they are often emotionally charged. Cognition and social processes evaluate the level of cognitive engagement of social dynamics in contents. Human readers manually extract the above features.

4.1 LLMs used in influencer analysis

Introduced by Devlin et al. (2018), bidirectional encoder representations from transformers (BERT) represent a groundbreaking advancement in LLMs, utilising transformer architecture to interpret linguistic content. BERT, one of such transformers, uniquely understands words in a context formed by all other words in the text. Its operation involves two primary stages:

1 pre-training on an extensive corpus of text

2 fine-tuning for specific tasks through transfer learning.

This approach has enabled BERT to excel in various NLP tasks, including analysing cryptocurrency-related tweets.

Bevendorff et al. (2023) organised a PAN23 competition, which included research tasks on authorship. One task focused on profiling cryptocurrency influencers based on their tweets. The competition consists of three subtasks: classifying influencers into five categories:

1 null

2 nano

3 micro

4 macro

5 mega.

Classifying tweet intentions into four categories:

1 subjective opinion
2 financial information
3 advertising
4 announcement.

Classifying tweet interests into five categories:

1 technical information
2 price updates
3 trading matters
4 gaming
5 others.

These subtasks comprehensively analyse influencers based on their authorship styles and interests, presenting them as few-shot NLP challenges. Organisers provided baseline models utilising low dimensionality representation (LDSE) by Rangel et al. (2018), character-based Logistic Regression, and a sentence transformer Sentence-T5 by Ni et al. (2021). Most participants outperformed them by fine-tuning BERT or some transformer models.

The top team by He et al. (2023) created a new model: DeBERTaV3 (Decoding-enhanced BERT with disentangled attention). This model fused BERT and RoBERTa using a disentangled attention mechanism, improving model performance. Many BERT variants use self-attention, which takes both tokens in the sequence and their positional information to create one set of attention scores. The scores are weights to adjust the surrounding tokens' contribution to the target token. Distangled attention treats content and positional relationships independently and creates two sets of attention scores. Villa-Cueva et al. (2023) achieved the best result in the second subtask, using an ensemble by an original and an entailment BERT. The entailment BERT was innovatively trained on augmenting samples with their 'entailment' (a similar synthetic tweet) and 'contradiction' (a synthetic tweet from a different category) samples. Li et al. (2023) and Espinosa and Sidorov (2023) approached tasks similarly by fusing BERT and BERTweet. Li et al. also used a contrastive learning objective function to maximise the distance between similar tweets and minimise the distance between tweets from different categories.

Girish et al. (2023) further explored employing a sentence transformer to extract textual features for training the linear SVC model. Siino et al. (2023) enhanced an ELECTRA model, which processes entire input sequences instead of just masked parts. They also augmented tweets with back-translated tweets in multiple languages. This technique was also adopted by Lomonaco et al. (2023), who extended tweets to include their Japanese-translated ones. They also showed that ELECTRA is less accurate than XLNet, developed by Yang et al. (2019). It uses autoregressive prediction on all tokens in a sequence but in a random order. A comparative study by Ferri-Molla and Santamaria-Jorda (2023) evaluated BERT, BERTweet (BERT fine-tuned for English tweets), and RoBERTa. It revealed that BERT is superior to its fine-tuned variants. They also noted that DistilBERT, a compact derivative of BERT, showed comparable results.

4.2 Generative pre-trained transformer (GPT)

Villa-Cueva et al. (2023) used ChatGPT to create synthetic tweets to double the sample tweets. They formed one prompt from one author's tweets and combined generated tweets to create a new synthetic author. Eventually, they made an equal number of synthetic authors, each with the same number of samples in different categories. Giglou et al. (2023) profile influencers using a transformer encoder complemented with prompt templates, a technique prevalent in ChatGPT. Prompt templates structure inputs uniformly, enabling the models to process and respond appropriately. One prompt template is like this, "*Given the following user tweets, determine the profile of this user as a cryptocurrency influencer: tweets: {tweets}*". They sent prompts as inputs to the transformer encoder to tackle the subtasks as few-short problems. A few-shot means no training for the model, only a prompt including a few samples.

4.3 Performance challenge

A notable challenge in tasks is working with small datasets. For instance, one task provides only 32 influencers per label, each with a maximum of 10 English tweets and 380 tweets. This data volume is minimal compared to other research, such as Lichti's work involving 218 Twitter users with 545,711 Bitcoin tweets in total. Bevendorff explained that low-resource tasks are practical in real-world scenarios where tweets stream rapidly with small volumes in processing pipelines. The data statistics for PAN23 tasks are presented in Table 2.

Table 2 Statistics of few-shot task datasets

Task	*Categories*	*Data size*	*Tweets per user*	*Users*
Profiling influencer	5	380	<= 10 tweets	380
Influencer intents	4	722	One tweet	722
Influencer interests	5	548	1 tweet	548

A key challenge in few-shot and low-resource classification tasks is their reduced accuracy and precision. For example, the top-performing team at PAN23 (2023) achieved macro-F1 scores of 62.32 in influencer profiling, 67.12 in identifying influencer interest, and 67.46 in determining influencer intent. The macro-F1 score, crucial for evaluating performance in classification tasks, especially in NLP, highlights this issue. Concurrently, fine-tuning LLMs demands substantial resources. Increasing F1 scores with slightly more data requires GPUs with more memory and proper training of an LLM.

4.4 Other content analysis methods

Merkley et al. (2023) used Google search to select 180 crypto-influencers with tweets mentioning top cryptos. They utilised RoBERTa and FinBERT to conduct sentiment analysis of tweets. They then used Nadam Optimisation with a feed-forward neural network to classify tweets into three categories:

1 buy-recommendation

2 non-recommendation (including hold-recommendation and non-classified, such as news and price updates)

3 sell-recommendation.

They crafted monetisation features to evaluate the crypto investment values from influencer tweets, such as returns surrounding the tweet date, following the tweet date, and within short-window and long-window (e.g., 30 days). Some descriptive features were also introduced, such as self-described expert at Twitter (X), YouTube links, and # mentiones per day. They found that short-term investment gain is correlated with tweets of positive sentiments or classified as 'buy-recommendation'. However, these tweets are followed by significant negative long-horizon returns. The finding echoes concern about influencers' 'pump-and-dump' manipulation. On average, there is no price value in influencer tweets.

Hamza (2020) also analysed the sentiment of tweets from 50 selected cryptocurrency influencers, most from CoinMarketCap. They utilised a simple rule-based algorithm called VADER for sentiment classification, which was invented by Hutto and Gilbert (2014). VADER is claimed to be efficient and effective based on its linguistic features and pre-defined sentiment intensity measures for tokens. Muslihuddeen et al. (2023) extracted enriched features and text embeddings and trained them with statistical models. For example, they extracted the TF-IDF vector, the number of tweets per user, valid and invalid hyperlinks, and cryptocurrency-related terms in texts. Finally, they experimented with statistical models like Random Forest, SVM, Logistic Regressions, and Random Forest. From the experiments, Logistic Regression with Active Learning showed the best performance. However, their solutions were ranked as the lowest in the contest results.

5 Identification using network algorithms

Many researchers, for instance, Alp and Öğüdücü (2018), leveraged the topology of social networks to find influencers, treating users as nodes and their connections as edges. Two main types of algorithms are typically used to identify influencers within sub-networks: graph-based methods with or without nodal features. Nodal features are non-graph attributes associated with nodes in a network. These attributes go beyond edges and provide rich information about the nodes themselves. For example, user profile properties or activity metrics about users.

5.1 Graph-based approach

This approach relies solely on social network graphs' properties to construct diffusion and influence models. Haveliwala (2002) proposed a model derived from PageRank. It essentially determines the importance of a node – a user, according to the importance of connected nodes, i.e., followers. The importance is approximated as a diffusion probability, computed using an algorithm derived from PageRank. Initially developed by Google, PageRank determines the importance of webpages. When applied to the context of influencer analysis, the principle is similar. The model calculates the influence of a topical influencer as below,

$$Pr(i) = \frac{1-d}{N} + d\sum_j \frac{Pr(j)}{L(j)}$$

where

$Pr(i)$	PageRank of Node i
d	damping factor, typically around 0.85. It represents the probability that a user continues following links instead of randomly jumping to another node
N	total number of nodes in the network
$L(j)$	number of links going out of the node j.

Like PageRank, the diffusion probability $Pr(j)$ from the follower j to influencer i is bigger when j itself has fewer followers, that is, $L(j)$ is small. At the beginning, for every user i in the network, $Pr(i)_0$ is initialised as 1, then $Pr(i)_{t+1}$ is updated interactively using $Pr(j)_t$. All users are updated together for each iteration – e.g., after 50 iterations, all user probabilities will converge.

Weng et al. (2010) proposed an enhanced PageRank algorithm, called TwitterRank. In which j is not a direct follower but a user with ‘following’ relationships in the cryptocurrency sub-network. The *following* relationship is a topic-specific random walk in a topic-specific network. A random walk is a stochastic process that involves a sequence of random steps on a network. It simulates a user randomly navigating the Twitter communities, moving from one user to another through connections. It effectively identifies the most influential users based on how likely they will be “visited” during the random walk processes.

5.2 *Graph-based approach with nodal features*

Unlike the purely graph-based approach, this approach incorporates additional features of nodes into the calculation of diffusion probability. Mittal et al. (2020) used a list of metrics for topology, user, and content. Topological metrics include in-degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. The user metric is calculated from likes. Content metrics include mentions, retweets, and mention ratio. Afterward, aggregation techniques combine all metrics. Two aggregation methods exist: positional-based, like Borda Count, and majority voting, like Condorcet approaches. Borda Count gives a node one score for each of its metrics based on metric rank. A node gets a score by adding all its metric scores. Condorcet compares a node with every other node and ranks nodes according to their total wins.

5.3 *Performance challenges*

The primary challenge with network algorithms is their substantial space and runtime complexity. Network algorithms require quadratic space to analyse a user, considering all pairs of user’s connections to calculate diffusion probabilities. This becomes particularly demanding when dealing with influencers comprising thousands of followers. Consequently, computing and continually updating these probabilities for each iteration demands significant computational resources.

6 Key opinion leaders (KOLs) on Friend.Tech

Recently, decentralised social platforms on blockchain, i.e., SocialFi, have gained the populace by allowing influencers to monetise their status. For instance, Friend.Tech has emerged as a notable platform linking to Twitter (X). Ho (2023) explains how Friend.Tech's (KOL) accounts correspond with Twitter influencers, enabling them to tokenise their influence. Users can purchase a key associated with a KOL token, granting them direct conversation with the KOL. These keys are subject to dynamic pricing, depending on the quantities users hold.

Guidi and Michienzi (2022) review technologies used in SocialFi and categorise SocialFi as a groundbreaking fusion of social networking, finance, and blockchain technology. According to Liu et al. (2023), Friend.Tech gained significant attention at a launch, attracting 139,000 users in August 2023. The authors define influencers as prominent users who disclose their Twitter accounts as KOLs and earned protocol revenue exceeding 0.1 ETH, i.e., $0.1 \times \mathrm{USD} / \mathrm{ETH}$. Within the first month, 2553 influencers were identified, deriving Protocol revenue (in ETH) from key transactions – initial sales by KOLs and subsequent resales by holders. The platform employs Bonding Curves to structure the pricing mechanism. Initially, key prices escalate rapidly to draw user interest. When reaching a specific volume of sales, the price grows slower.

The Bonding Curve formula for purchasing a key is defined in equation (1),

$$y = \frac{x^2}{16000} \tag{1}$$

where y represents the buying price, x denotes the total key circulation, which increases when new keys are purchased and decreases upon resale.

For selling a key, the formula adjusts slightly as defined in equation (2),

$$y = \frac{(x-1)^2}{16000} \tag{2}$$

Additionally, the platform imposes service fees on transactions. Figure 4 shows the buying key Bounding Curve. This pricing and transaction model underscores the decentralised social platform's dynamic and monetised interactions.

However, heavily monetised social platforms present significant drawbacks. Liu et al. (2023) report that 99.4% of influencers on these platforms have fewer than 100 followers, suggesting that the current pricing and monetisation strategies might deter social interaction. This is counterproductive to the foundational goal of digital social platforms, which is to facilitate open and fair communication among users, irrespective of their ranks or status. Crypto (2024) ranks Friend.Tech as the most significant price fluctuated SocialFi platform over one day.

Figure 4 Bonding curve for buying keys (see online version for colours)

6.1 *Followers-as-a-reputation (FaaR)*

Imani Rad and Banaeian Far (2023) point out that the number of the following users can be considered as a fair measuring parameter of influencers on SocialFi platforms. Instead of counting directly, FaaR should be calculated using AI-based algorithms from the smart contract(s) in the background. This can exclude the vast number of AI-based bots. Almasound et al. (2020) pointed out that smart contracts have been widely used to build blockchain reputation systems to quantify user trustworthiness. This factor is necessary to count true FaaR. Kamboj et al. (2021) proposed a role-based access control (RBAC) to authenticate real Ethereum users from bots.

6.2 *Other SocialFi platform influencers*

Imani Rad and Banaeian Far (2023) extensively reviewed SocialFi platforms, noting that they capitalise on various social elements, including influencers, pages, groups, individuals, posts, assets, and rewards. They highlighted that financial gains do not necessarily correlate with an influencer's impact. Often, the wealthiest individuals on these platforms achieve their fortunes outside platforms. Furthermore, newer platforms such as SocialFiAI and Influencio have been integrated within Metaverse, broadening their scope and functionality. Regional platform Million has been launched to serve the Middle East.

7 Discussions

Cryptocurrency influencers are well-known and active on social platforms. Normal users look up to their expert opinions regarding trends, markets, and policies. However, users lack practical tools to find relevant and well-intentioned influencers. We reviewed and summarised approaches that are accessible to major academic search engines. But there are a few questions we haven't answered so far.

7.1 Open research questions

SEC and various academic researchers have demonstrated that cryptocurrency influencers often distribute biased social messages for several reasons. These include engaging in 'pump and dump' schemes, building personal brands, or misjudging bear markets. Binder (2021) explores prominent scams regarding influencers promoting altcoins and their negative impacts. Given these findings, researchers must develop a clear and actionable definition of a socially responsible influencer who genuinely intends to benefit followers and contribute positively to the cryptocurrency community. Another research question concerns overcoming the current limitations of methods used to analyse influencer behaviour and how LLMs can be effectively combined to accurately profile influencers' actual characteristics. A promising avenue appears to be a hybrid approach that involves fine-tuning pre-trained transformer models on statistical, psychological, and monetary features, which could significantly enhance the profiling of cryptocurrency influencers and the nature of their communications.

On the decentralisation front, the monetisation models employed by SocialFi platforms lack the complexity and sustainability required to foster a thriving digital social community. The cryptocurrency is not transparent and is highly risky, and leaders depend on KOLs to assist users instead of speculative investors. To address this, SocialFi technologies must incorporate social and psychological features to promote genuine KOLs and their community. A more refined profiling method should reward KOLs based on their popularity, engagement, and direct benefits to followers.

7.2 Technological determinism view

Technological determinism suggests that technology development drives social and cultural changes, but not always positively, depending on the designer's intentions. Merkley et al. (2023) identified that market returns of the cryptos specifically mentioned in influencer tweets are associated with significant positive short-term returns but significant long-horizon returns. The finding confirms the SEC's concern regarding influencers' 'pump-and-dump' schemes using their tweets. The mega influencers' posts regarding small market-capped cryptos are the least profitable to users. Hamza (2020) also confirms that influencer tweets are positively associated with market prices in a bull market instead of a bear market.

8 Conclusion

This literature review delved into recent research on the analysis, identification, and classification of influencers, focusing on Twitter influencers discussing cryptocurrency in English. Some literature often interchanges the term 'influencers' with 'key opinion leaders (KOLs)'. Our study highlighted several predominant methodologies: statistical metrics from the Twitter platform, psychological attributes, linguistic content analysis via Machine Learning or LLMs, and graph-based algorithms. Among these, statistical metrics are prevalently used in industry for their simplicity. While enhancing accuracy, psychological properties require manual interpretation. Graph-based algorithms, incorporating both graph theory and nodal features, require spatial databases for data storage and processing. LLMs have improved content analysis in the authorship of tweets. Nonetheless, LLMs primarily focus on text data, often overlooking user connections and dynamics of social community. Future research can enrich pure statistical methods by integrating LLMs with tweets, psychological features, and social features. There's also a growing trend towards employing LLMs for multimodal data analysis, extending beyond textual content to images and videos. Furthermore, this paper synthesises vital knowledge areas in this field, including defining influencers, categorising influencer tiers, and the significance of engagement rates in marketing. It sheds light on classifying topics, intentions, and interests specific to cryptocurrency influencer tweets. It also reviews technologies such as Friend.Tech that allows Twitter KOLs to socialise on the blockchain. We hope these insights and recommendations will inspire future scholars to pursue more profound, more insightful research in this evolving field of influencer analysis.

References

AlgoBlocks (2023) *10 Most Influential People in the Crypto Space*, CoinMarketCap, Retrieved 31 May, 2024, from https://coinmarketcap.com/community/articles/37206

Almasoud, A, Hussain, F.K. and Hussain, O.K. (2020) 'Smart contracts for blockchain-based reputation systems: a systematic literature review', *Journal of Network and Computer Applications*, Vol. 170, pp.102814.

Alp, Z.Z. and Öğüdücü, Ş.G. (2018) 'Identifying topical influencers on Twitter based on user behavior and network topology', *Knowledge-Based Systems*, Vol. 141, pp.211–221.

AZCoinNews (2023) *U.S. DoJ pushes For $4 billion Binance Settlement to Halt Investigation, BNB Price Surges to $266*, Binance, Retrieved 31 May, 2024, from https://www.binance.com/en/square/post/328564504874

Benson, J. (2022) *Elon Musk's Dogecoin and Bitcoin Tweets: A Crypto Twitter Timeline*, Decrypt, Retrieved 31 May, 2024, from https://decrypt.co/97015/elon-musk-dogecoin-bitcoin-tweets-crypto-twitter-timeline

Bevendorff, J., Chinea-Ríos, M., Franco-Salvador, M., Heini, A., Körner, E., Kredens, K., Mayerl, M., Pęzik, P., Potthast, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B., Wiegmann, M., Wolska, M. and Zangerle, E. (2023) 'Overview of Pan 2023: authorship verification, multi-author writing style analysis, profiling cryptocurrency influencers, and trigger detection', *European Conference on Information Retrieval*, Dublin, Ireland, pp.518–526.

Binder, M. (2021) *Inside the Shady World of Influencers Promoting Cryptocurrency*, Mashable, Retrieved 31 May, 2024, from https://mashable.com/article/influencers-altcoin-scams

Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, K. (2010) 'Measuring user influence in Twitter: the million follower fallacy', *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 4, No. 1, pp.10–17.

Crypto (2024) *Top SocialFi Tokens by Market Capitalization*, Cypto.com, Retrieved on May 15, 2024 from https://crypto.com/price/categories/SocialFi

CryptoSlate (2024) *Bitcoin Maximalists*, Retrieved on May 15, 2024 from https://cryptoslate.com/people/category/bitcoin-maxis

Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2018) *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, arXiv:1810.04805, https://doi.org/10.48550/arXiv.1810.04805

Edyme, S. (2024) *PlanB's Market Cycle Analysis: Bitcoin Gears up for Skyrocketing Beyond $100,000*, News BTC, Retrieved 31 May, 2024, from https://www.newsbtc.com/bitcoin-news/planbs-market-cycle-analysis-bitcoin-gears-up-for-skyrocketing-beyond-100000

Espinosa, D. and Sidorov, G. (2023) *Using BERT to Profiling Cryptocurrency Influencers*, Working Notes of CLEF.

Ferri-Molla, I. and Santamaria-Jorda, J. (2023) *Profiling Cryptocurrency Influencers with Few-Shot Learning*, Working Notes of CLEF.

Garzik, J. (2023) *Biography of the Bloq's CEO*, Bitcoin Wiki, Retrieved 31 May, 2024, from https://bitcoinwiki.org/wiki/jeff-garzik

Giglou, H.B., Rahgouy, M., D'Souza, J., Oskuee, M.M., Tekanlou, H.B.A. and Seals, C.D. (2023) *Leveraging Large Language Models with Multiple Loss Learners for Few-Shot Author Profiling*, Working Notes of CLEF.

Girish, K., Hegde, A., Balouchzahi, F. and Shashirekha, H.L. (2023) *Profiling Cryptocurrency Influencers with Sentence Transformers*, Working Notes of CLEF.

Guidi, B. and Michienzi, A. (2022) 'SocialFi: towards the new shape of social media', *ACM SIGWEB Newsletter*, pp.1–8.

Hamza, S. (2020) 'The effect of tweets made by cryptocurrency opinion leaders on Bitcoin prices', *Saudi Journal of Economics and Finance*, Vol. 4, No. 12, pp.569–589.

Haveliwala, T.H. (2002) 'Topic-sensitive PageRank', *Proceedings of the 11th International Conference on World Wide Web*, Honolulu Hawaii, USA, pp.517–526.

He, P., Liu, X., Gao, J. and Chen, W. (2020) *DEBERTA: Decoding-enhanced BERT with Disentangled Attention*, arXiv: 2006.03654, https://doi.org/10.48550/arXiv.2006.03654

Hilarski, R. (2019) *Introducing Randy Hilarski, Crypto Daily's Newest Ambassador*, Steemit, Retrieved 31 May, 2024, from https://steemit.com/crypto/@hilarski/introducing-randy-hilarski-crypto-daily-s-newest-ambassador

Hirsch, J.E. (2005) 'An index to quantify an individual's scientific research output', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, No. 46, pp.16569–16572.

Ho, A. (2023) *Friend.Tech*, Crypto.com, Retrieved on 16 May, 2024 from https://crypto.com/research/friend-tech.

Hutto, C. and Gilbert, E. (2014) 'VADER: a parsimonious rule-based model for sentiment analysis of social media text', *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, pp.216–225.

Imani Rad, A. and Banaeian Far, S. (2023) 'SocialFi transforms social media: an overview of key technologies, challenges, and opportunities of the future generation of social media', *Social Network Analysis and Mining*, Vol. 13, No. 1, p.42.

Kamboj, P., Khare, S. and Pal, S. (2021) 'User authentication using blockchain based smart contract in role-based access control', *Peer-to-Peer Networking and Applications*, Vol. 14, No. 5, pp.2961–2976.

Kharpal, A. (2024) *Ether Prices Spike on Hopes of an ETF – but the SEC Signals that's less Clear Cut than Bitcoin*, CNBC, Retrieved 31 May, 2024, from https://www.cnbc.com/2024/01/11/ether-price-spikes-on-hope-of-eth-etf-but-will-the-sec-approve-it.html

Kozul, V. (2023) *What is Carl Runefelt's Net Worth And How did he Make it? A case study*, FounderBounty, Retrieved 31 May, 2024, from https://www.founderbounty.com/what-is-carl-runefelts-net-worth-and-how-did-he-make-it

Li, Z., Han, Z., Cai, J., Huang, Z., Huang, S. and Kong, L. (2023) *Profiling Cryptocurrency Influencers with Few-Shot Learning via Contrastive Learning*, Thessaloniki, Greece.

Lichti, C., Ademi, E. and Tumasjan, A. (2024) 'Decentralized opinion leadership: A study of crypto influencers in the Twitter discourse on bitcoin', *Proceedings of the 57th Hawaii International Conference on System Sciences*, Hawaii, USA, pp.2853–2862.

Liu, K., Yu, M., Jin, Y., Wang, Y., Yan, J. and Liu, X.F. (2023) 'Tokenomic model of Friend.Tech social platform: a data-driven analysis', *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, Shanghai, China, pp.656–662.

Lomonaco, F., Siino, M. and Tesconi, M. (2023) *Text Enrichment with Japanese Language to Profile Cryptocurrency Influencers*, Working Notes of CLEF.

Merkley, K.J., Pacelli, J., Piorkowski, M. and Williams, B. (2023) 'Crypto-influencers', *Working Paper for the Review of Accounting Studies Research Conference*, Indiana University.

Mittal, D., Suthar, P., Patil, M., Pranaya, P.G.S., Rana, Dipti P. and Tidke, B. (2020) 'Social network influencer rank recommender using diverse features from topical graph', *Procedia Computer Science*, Vol. 167, pp.1861–1871.

Muslihuddeen, H., Sathvika, P., Sankar, S., Ostwal, S. and M, A.K. (2023) *Profiling Cryptocurrency Influencers using Few-Shot Learning*, Working Notes of CLEF.

Ni, J., *et al.* (2021) *Sentence-t5: Scalable Sentence Encoders from Pre-Trained Text-to-Text Models*, arXiv preprint arXiv:2108.08877.

PAN23 (2023) Profiling cryptocurrency influencers with few-shot learning', PAN@CLEF23, Retrieved 31 May, 2024, from https://pan.webis.de/clef23/pan23-web/author-profiling.html

Parikh, R. (2023) *Impact of Celebrity Endorsed Cryptocurrency*, Boston University, Retrieved 31 May, 2024, from https://www.bu.edu/rbfl/2023/05/17/impact-of-celebrity-endorsed-cryptocurrency

Pereira, N. (2022) *5 Different Tiers of Influencers and When to Use Each. Zero Gravity Marketing*, Zero Gravity Marketing, Retrieved 31 May, 2024, from https://zerogravitymarketing.com/the-different-tiers-of-influencers-and-when-to-use-each

Peter, A. (2023) *The Rise of NFT Influencer Marketing: A Game-Changer for Brands*, LinkedIn, Retrieved 31 May, 2024, from https://www.linkedin.com/pulse/rise-nft-influencer-marketing-game-changer-brands-alexx-peter

Rangel, F., Franco-Salvador, M. and Rosso, P. (2018) 'A low dimensionality representation for language variety identification', *International Conference on Intelligent Text Processing and Computational Linguistics*, Konya, Turkey, pp.156–169.

Retimuko (2024) *Vitalik Buterin*, Wikipedia, Retrieved 31 May, 2024, from https://en.wikipedia.org/wiki/Vitalik_Buterin

SEC (2022) *SEC Charges Kim Kardashian for Unlawfully Touting Crypto Security*, SEC, Retrieved 31 May, 2024, from https://www.sec.gov/news/press-release/2022-183

Siino, M., Tesconi, M. and Tinnirello, I. (2023) *Profiling Cryptocurrency Influencers with Few-Shot Learning using data Augmentation and ELECTRA*, Working Notes of CLEF.

Stark, T., Levin, D. and Klein, G.M.P. (2023) *The SEC Continues to Prove it is the Most Powerful Influencer: How to Avoid Touting Charges*, White & Case LLP International Law Firm, Retrieved 31 May, 2024, from https://www.whitecase.com/insight-alert/sec-continues-prove-it-most-powerful-influencer-how-avoid-touting-charges

Tarnishedpath (2023) *Peter Schiff*, Wikipedia, Retrieved 31 May, 2024, from https://en.wikipedia.org/wiki/Peter_Schiff

Vays, T. (2014) *Tone Vays*, Twitter, Retrieved 31 May, 2024, from https://twitter.com/ToneVays

Villa-Cueva, E., Valles-Silva, J.M., Lopez-Monroy, A.P., Sánchez-Vega, F. and López-Santillán, J.R. (2023) *Few Shot Profiling of Cryptocurrency influencers using Natural Language Inference & Large Language Models*, Working Notes of CLEF.

Wade, J. (2024) *Spot Bitcoin ETFs: What are they and How do they Work?*, Forbes, Retrieved 31 May, 2024, from https://www.forbes.com/advisor/investing/cryptocurrency/spot-bitcoin-etfs

Wang, N. (2022) *SEC Accuses 2 Firms of crypto Pump-and-Dump Scheme*, Yahoo!Life, Retrieved 31 May, 2024, from https://sg.style.yahoo.com/finance/news/sec-accuses-2-firms-crypto-162138437.html

Warren, J. (2021) *How Follower Count Impacts Influencer Engagement Rate*, Later, Retrieved 31 May, 2024, from https://later.com/blog/influencer-engagement-rate

Weng, J., Lim, E.P., Jiang, J. and He, Q. (2010) 'TwitterRank: finding topic-sensitive influential twitterers', *Proceedings of the Third ACM International Conference on Web Search and data Mining*, New York, NY, USA, pp.261–270.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J.G., Salakhutdinov, R. and Le, Q.V. (2019) 'XLNet: generalized autoregressive pretraining for language understanding', *Neural Information Processing Systems*, p.32.