Leveraging social networks for mergers and acquisitions forecasting

Alessandro Visintin[1] and Mauro Conti[1]

University of Padova, Via VIII Febbraio, 2, 35122 Padova, Italy

Abstract. Mergers and acquisitions are pivotal strategies employed by companies to maintain competitiveness, leading to enhanced production efficiency, scale, and market dominance. Due to their significant financial implications, predicting these operations has become a profitable area of study for both scholars and industry professionals. The accurate forecasting of mergers and acquisitions activities is a complex task, demanding advanced statistical tools and generating substantial returns for stakeholders and investors. Existing research in this field has proposed various methods encompassing econometric models, machine learning algorithms, and sentiment analysis. However, the effectiveness and accuracy of these approaches vary considerably, posing challenges for the development of robust and scalable models.
In this paper, we present a novel approach to forecast mergers and acquisitions activities by utilizing social network analysis. By examining temporal changes in social network graphs of the involved entities, potential transactions can be identified prior to public announcements, granting a significant advantage in the forecasting process. To validate our approach, we conduct a case study on three recent acquisitions made by Microsoft, leveraging the social network platform Twitter. Our methodology involves distinguishing employees from random users and subsequently analyzing the evolution of mutual connections over time. The results demonstrate a strong link between engaged firms, with the connections between Microsoft employees and acquired companies ranging from five to twenty times higher than those of baseline companies in the two years preceding the official announcement. These findings underscore the potential of social network analysis in accurately forecasting mergers and acquisitions activities and open avenues for the development of innovative methodologies.

Keywords: Social networks analysis, Merger and acquisition prediction, Twitter analysis

1 Introduction

Mergers and acquisitions (M&A) are strategic business activities that involve the integration of two or more firms into a single entity. These transactions can occur in several forms, including the acquisition of one company by another, the merger of two organizations to create a new entity, or the purchase of a segment of a company's assets by another firm. M&A activities are motivated by a range of factors, such as expanding market shares, diversifying product offerings, reducing costs, or accessing new technologies. These transactions have far-reaching implications for the involved firms, their industries, and the broader economy. In the year 2022, the aggregate value of M&A transactions worldwide amounted to 3.8 trillion dollars [14], underlining the significant role of these activities in the global economy. Therefore, M&A transactions represent a substantial portion of the overall economic landscape and can have an immediate impact on the market value of the companies involved.

For these reasons, the development of predictive models for M&A transactions gained substantial attention from both scholars and stakeholders. Scholars may study M&A activity to gain insights into the underlying economic and strategic factors that drive these transactions. Stakeholders, on the other hand, may try to predict M&A activity to gain an edge in the stock market and potentially generate significant profits. Traditional techniques for the prediction are based on the analysis of numerical fundamentals such as company value, revenue, and profit [19,15,3,22,31]. Statistical methods like logistic regression or support vectors machine are then applied to perform the classification and generate a model. These studies reported promising results, revealing a systematic nature of such transactions. However, their practical application is limited by several factors. Financial data is often sparse and requires extensive manual work to build labeled datasets. Additionally, these datasets are heavily skewed towards negative samples, thus leading to generalization issues and over-fitting. The largest dataset used in the previous works only contained 2394 cases and 61 acquisitions [31]. These aspects hinder the successful creation of an automated method at scale. Methods based on topic modeling and natural language processing [32,18] tried to overcome these issue by leveraging the written text of news articles. These approaches extract valuable features from specialized articles and use machine learning techniques to infer potential players in a merger and acquisition transaction. However, data from articles is very sparse and biased towards larger, best-known companies.

This paper presents an analysis on social network relationship graphs as a valuable source of information regarding merger and acquisition transactions. Many companies maintain official accounts on social networks, and their followers typically include employees and individuals closely associated with the company. By monitoring the evolution of these followers' connections over time, it becomes possible to examine the ties between two companies. Specifically, in the context of a planned M&A, we expect to observe a higher degree of inter-connectivity between the two firms relative to other, unrelated companies. In particular, we except these interactions to happen before the public announcement of the operation, thus becoming a crucial source of information to leverage.

To prove our point, we conducted a study on three major acquisitions made by Microsoft (*i.e.,* LinkedIn, GitHub, and Activision) using Twitter data. To determine the strength of the interconnections between the companies involved, we developed a methodology for recovering the timestamps of user-following events on the platform. We also devised a means of differentiating between employees and random users, and defined the time frame in which they worked for their respective companies. Our results provide evidence in support of our hypothesis, as Microsoft exhibited between five and twenty times more connections than our baseline reference of unrelated companies (*i.e.,* Google and Apple) in the two years preceding the announcement of the acquisition. Furthermore, these connections exhibited clear bursts in the same time period, further reinforcing the validity of our claim.

Contributions. The contributions of this article are five-fold:

- we develop a novel technique for extracting connection dates from Twitter, which allows us to precisely measure the strength and evolution of interconnections over time;
- we devise a methodology for discriminating between employees and random users, which enables us to isolate the relevant connections and analyze them more effectively;
- we develop a methodology for identifying the working period of the employees, which enables us to provide more precise results;
- we conduct a comprehensive analysis of three major acquisitions using Twitter data, which provides insights into the dynamics of these transactions and the interplay between the companies involved;
- we identify several potential avenues for further improving our methodology and expanding its applicability to other acquisition transactions.

Collectively, these contributions demonstrate the potential of social network graphs as a powerful tool for gaining a deeper understanding of M&A transactions and informing strategic decision-making.

Organization. This paper is structured as follows: Section 2 provides an overview of previous works and defines the current state of the art in the field of analyzing social network graphs for acquisition transactions. Section 3 introduces Twitter, the Snowflakes IDs, and the methodology employed for extracting timestamps from Twitter connections. Section 4 presents the case studies and describes the methodology used to identify employees and detect their working periods, as well as the results of our analysis. Section 5 discusses the current limitations of our work and potential avenues for future improvements. Finally, Section 6 offers concluding remarks.

2 Related works

Several prior works modeled mergers and acquisitions with the usage of financial and managerial variables. Commonly used indicators are firm size [19,5,3,26,22], market-to-book ratio [19,5,26,23], cash flow [23,3,26], and return on assets [19,22]. Researchers also investigated the forecasting power of patenting data [3] and bankruptcies [24,9,21]. These studies applied a variety of data mining and machine learning approaches: logistic regression [19,5,23,3,22], discriminant analysis [5,23], rule induction [23], decision tree [28], rough set approach [25], support vector machines [24], decision trees [9]. Adelaja et al. [2] built a two-logit model to explain merger and acquisition activities in US food manufacturing. Using firm level data, they derived a target model for predicting the likelihood of a firm becoming a target for an acquisition, and a takeover model predicting the likelihood of a targeted firm being taken over. The reported predictive accuracy was of 74.5% and 62.9% respectively. Wei et al. [31] used ensemble learning algorithm on resampled data to solve the problem of data skewness, resulting in a TP rate of 46.43% on 2394 companies out of which 61 actually got acquired. Olson et al. [21] applied a variety of data mining tools to bankruptcy data, with the purpose of comparing accuracy and number of rules. Decision trees were relatively more accurate compared to neural networks and support vector machines. They reported a classification performance ranging from 0.661 of SVM to 0.948 of J48 decision tree.

More recent works explored the integration of different data sources to improve the scale and forecast capabilities. Xiang et al. [32] used public information available on news websites (*i.e.,* TechCrunch, CrunchBase) to build a topic modeling system. They collected firm information and news articles from these websites and organized them into 22 factual features and a varied number of topic features. Their model achieved a true positive rate between 60% and 79.8%, and a false positive rate mostly between 0% and 8.3%. Li et al. [18] applied graph neural networks to the task at hand. They integrated the database from [32] with novel data points. They reported an improvement over the state of the art, with a true positive rate of 83% and a false positive rate of 7.8%.

3 Retrieving connection dates from Twitter

Twitter [30] utilizes the Snowflake ID generation system [4,11] to organize its data. The Snowflake IDs preserve their respective creation times, thereby enabling the recovery of this information. Consequently, this can be leveraged to construct a time series of profile-following activities.

3.1 Twitter and the Snowflake IDs

Twitter is a micro-blogging service that was launched in July of 2006. It has since become one of the major social media platforms, boasting an impressive active user base of approximately 230 million units [27]. The primary functionality of the service is the ability to post short text messages, known as *tweets*, which are limited to a maximum of 280 characters. These tweets allow users to share their thoughts and opinions with others on the platform. In addition to posting tweets, users have the option to create a short *biography* for their profile and upload a *profile image*, specify a *location* and a *birthday*. One unique feature of Twitter is its unidirectional friendship functionality, which allows users to follow other users without requiring a mutual connection. This results in a distinction between a user's *friend*s and *follower*s. Specifically, a user's *friend*s are those who are followed by that user, while the user's *follower*s are those who follow that user. Consequently, each user on the platform maintains two separate lists of *friend*s and *follower*s in their profile.

Twitter employs the Snowflake ID system to organize data within its platform. A Snowflake ID consists of a 64-bit number that includes a creation timestamp (42 bits), a machine code (10 bits), and a rolling sequence number (12 bits). Every object on Twitter is assigned with a unique ID in this format, including each entry in a user's *friend*s and *follower*s lists. This organization allows for chronological ordering and efficient pagination. Upon retrieving these lists, the data is arranged into batches of profiles. Each batch is enclosed by references to the previous and subsequent batches, known as the *previous cursor* and the *next cursor*. The list is consumed from the

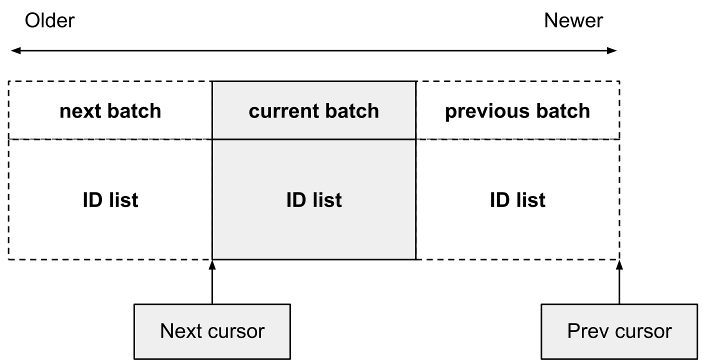

Fig. 1. Twitter employs pagination for organizing the friends and followers list, which are arranged in subsequent batches. Each batch comprises a list of profiles, a cursor to the previous batch, and a cursor to the next batch. The batches are chronologically ordered from the most recent to the oldest.

newest *follower*s to the oldest ones with the *next cursor* pointing to an older batch of IDs. It is important to note that the *next cursor* correspond to the ID of the last profile in the batch. Figure 1 provides a visual representation of this scheme.

3.2 Regressing a conversion formula

As discussed in Section 3.1, Snowflake IDs are timestamp-based identifiers that are utilized to identify each object on Twitter. Specifically, each entry in a *follower*s list is assigned an ID that is arranged in descending order, indicative of a chronological ordering of the list. In this context, the *next cursor* serves as a time pointer that aligns with the ID of the last entry in the current batch. To provide qualitative evidence supporting this observation, we conducted an experiment where we incremented and decremented the *next cursor* to observe how the returned batch would change. The results of this experiment are summarized in Figure 2. Decrementing the *next cursor* (i.e., moving to an earlier time) yields the same batch profiles. Conversely, incrementing it (i.e., moving to a more recent time) also results in the same batch of profiles, but with the last profile of the current batch at the top. This behavior strongly suggests that the *follower*s list is organized as a time series, with the *cursor*s serving as convenient time pointers. When a batch of profiles is requested, the system employs the *next cursor* as a reference point and navigates the time series backward until a sufficient number of profiles are retrieved or the end of the list is reached. Due to this structural arrangement, two consecutive *cursor*s can be leveraged to temporally delimit a batch of profiles. The temporal resolution of this frame is determined by the interval between the two *cursor*s.

Given the above observations, it is feasible to acquire a conversion formula for *cursor*s and UTC timestamps by establishing a mapping between these two values. This is achieved by repeatedly requesting a batch of length one from the top of a *follower*s list and recording the time and *next cursor* values whenever a new *follower* is added. To estimate the actual UTC timestamp, we recorded the time interval

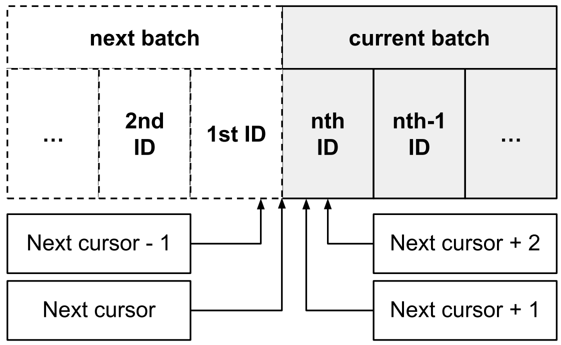

Fig. 2. The cursors within the followers list serve as time pointers. When a cursor retrieves a batch of profiles, the system returns a list starting from the profile that follows the cursor in time. Decreasing the cursor does not alter the output, whereas increasing it generates a list with the last profile of the current batch at the top.

between two consecutive requests. The average time difference between the measures was found to be 1.14s, with a standard deviation of 0.33s. We opted to employ the mean of the two values as an approximation of the actual UTC timestamp. The minor imprecision introduced, in the order of seconds, is negligible, particularly considering the time resolution used in this study, which is in the order of days and months. To expedite the data collection process, we monitored an influential Twitter account (*i.e.,* @Microsoft), which gains several hundred new followers within a few hours. We amassed a total of 732 data points between March 9th and March 12th, 2023. As described in Section 3.1, we utilized the 42 most significant bits of the *cursor*s for estimating a regression line. We employed the RANSAC regressor provided by the *sklearn* package to determine the best-fit line, primarily due to its robustness against outliers. We found the formula for the fitted line to be:

$$utc = 0.004 * (cursor >> 22) - 1746 \tag{1}$$

The first 42 bits of the *cursor*s were obtained by bit-shifting, denoted by $>>$. An excellent fit was achieved, with R^2 values close to 1 up to the tenth decimal place, indicating that almost all of the variability in the dependent variable had been accounted for. The Root Mean Squared Error ($RMSE$) was calculated to be 0.36, further demonstrating the high quality of the fit. The negative intercept value indicates that the reference epoch is half an hour prior to the Linux epoch (ie $1970-01-0100:00:00$). Given the broad time range and minimal error introduced by our method, we conjecture that the actual starting epoch aligns with the Linux epoch and thus eliminate the intercept from the formula. We can then state that incrementing the leading 42 bits by 1 results in a time increase of 0.004s. Put differently, the formula has a sensitivity of 4ms, which is more than adequate for the time resolution required in our study.

4 Analyzing the mergers and acquisitions

In section, we provide a thorough analysis of three acquisitions performed by Microsoft in recent years. After introducing the case studies (section 4.1), we provide the methodology for detecting employees (section 4.2) and their working period in a company (section 4.3). We eventually leverage this information to plot a time series of companies interactions through time and analyze it (section 4.4).

4.1 Case studies

Our study delved into the analysis of three prominent mergers and acquisitions made by Microsoft [20] Corporation, a multinational technology company headquartered in Redmond, Washington, USA. Founded on April 4, 1975, *Microsoft* initially aimed to develop and market personal computer software. Over time, the company diversified its business interests, expanding to various technology products and services such as computer software, gaming, hardware, and cloud computing. The rationale behind our choice to focus on Microsoft lies in its extensive presence on Twitter. Additionally, we were intrigued by the significant financial impact that Microsoft's acquisitions have had on the market.

Acquisition of Linkedin. On June 13, 2016 [6], Microsoft announced that it would acquire LinkedIn [13], the world's largest professional networking site, for 26.2 billion dollars in cash. The acquisition of LinkedIn was seen as a strategic move by Microsoft to expand its offerings in the professional services and social networking markets, and to create new opportunities and innovations for professionals and the broader technology industry. The announcement of the acquisition had a mixed impact on the market value of both companies. Microsoft's stock price decreased by around 3%, which represented a loss of approximately 9 billion dollars in market value. The decrease was likely due to concerns about the high price tag and the potential challenges of integrating LinkedIn's business with Microsoft's existing operations. LinkedIn's stock price, on the other hand, increased by around 47% on the day of the announcement, adding approximately 10 billion dollars to its market value. This increase was largely due to the fact that the price that Microsoft was offering for the company represented a significant premium over its previous stock price.

Acquisition of Github. On June 4, 2018 [7], Microsoft announced that it would acquire GitHub [13], a popular platform for software developers to collaborate and share code, for 7.5 billion dollars. The acquisition of GitHub was seen as a positive move by many in the developer community, as it brought the resources and expertise of Microsoft to one of the most important platforms for software development. After the announcement of the acquisition, Microsoft's stock price increased by around 1%, adding approximately 8 billion dolalrs to its market value. The increase was largely attributed to the potential benefits that the acquisition could bring to Microsoft's developer tools and cloud computing business. GitHub, as a private company, did not have a public market value at the time of the acquisition. However, its previous funding round in 2015 valued the company 2 billion dollars, thus determing an appreciation of 5.5 billion dollars.

Acquisition of Activision. On January 18, 2022 [8], Microsoft announced that it would acquire Activision Blizzard [1], one of the world's largest video game companies, for 68.7 billions. The acquisition of Activision is part of a larger strategy to expand Microsoft's presence in the gaming industry. Microsoft has been investing heavily in its gaming division, which includes the Xbox console and the Xbox Game Pass subscription service. The announcement had a significant impact on the market value of both companies. In the days following the announcement, Microsoft's stock price rose by about 8%, adding approximately 110 billion dollars to its market value. Activision's stock price, on the other hand, jumped by more than 40% on the day of the announcement, adding approximately 27 billion dollars to its market value. This increase was largely due to the fact that the price that Microsoft was offering for the company represented a significant premium over its previous stock price.

4.2 Detecting employees on Twitter

At the heart of our methodology lies the fundamental process of distinguishing between employee profiles and those of random users. To achieve this objective, we rely on the intuition that employees active on Twitter are likely to follow the official account of their employer. As such, we procured the followers lists of the four accounts central to our analysis, in addition to two supplementary accounts for baseline comparison: @microsoft, @linkedin, @github, @activision, @google, and @apple. We used the official API service [29] provided by Twitter to collect the data. We opened 3 developer accounts and interacted with the API using the Python [12] programming language. The collection was performed between March, 12 2023 and March, 19 2023. The data was stored and organized using a *SQLite* [10] database. To detect the employees, we opted for a simple approach based on regular expressions. Regular expressions are a powerful tool for searching patterns in text using sequences of characters. Major programming languages already include a standard library for processing regular expressions (*e.g., re* module for Python). We leveraged on the tendency of Twitter users to summarize their life experiences within the confines of the description field, often including comprehensive accounts of their professional backgrounds. The ensuing examples serve as prototypical illustrations of this practice.

"Principal Design Director @Microsoft"

"Information Technology Professional (currently at @Microsoft)"

"Program Manager @Meta. Ex-@Microsoft, @Twitter"

We created three regex expressions via a semi-automated inspection of a batch of descriptions. While the last two examples are simple variations of a common pattern (*e.g., formerly, former, ex, ex-*), the first one required to collect a list of working positions to prepend to the profile handle. For this purpose, we additionally built a list of 1102 jobs while inspecting the descriptions. Using this approach, we were able to find 4369 employees of Microsoft, 382 for LinkedIn, 408 for GitHub, 818 for Activision, 4115 for Google, and 804 for Apple. The details on the procedure and the regex used are provided in algorithm 1.

Algorithm 1 Detecting employees using regex expressions

```
procedure CleanDescription(d)
    # transform to lowercase
    d ← d.lowercase()
    # keep only alphanumeric chars, spaces, _, and @
    d ← d.remove([^a-z0-9 _@])
    # substitute multiple spaces with single ones
    d ← d.substitute(\s+,\s)
    return d
procedure FindCurrently(d)
    # catch all variations of currently
    r ← current(?:ly)?(?: at| @)?((?: @\w+)+)
    return d.find(r)
procedure FindPreviously(d)
    # catch all variations of previously (splitted string due to length)
    r1 ← (?:former(?:ly)?|fmr|prev(?:ious(?:ly)?)?|ex)
    r2 ← (?: at| @)?((?: @\w+)+)
    r = concatenate(r1,r2)
    return d.find(r)
procedure FindJob(d, jobsList, profileHandle)
    # check presence of job + Twitter handle for each job
    for job in jobsList do
        s ← concatenate(job, profileHandle)
        if s in d then
            return True
    return False
procedure DetectEmployee(d, jobsList, handlesList)
    d ← ClearDescription(d)
    if FindCurrently(d) then
        return True
    if FindPreviously(d) then
        return True
    # check jobs for each Twitter handle of interest
    for handle in handlesList do
        if FindJob(d, jobsList, handle) then
            return True
    return False
```

4.3 Determine the working period

The application of regex expressions facilitated the identification of employees on Twitter. Nevertheless, given the dynamic nature of the labor market, individuals may switch jobs or fail to update their employment status in their profile descriptions. To address this limitation and refine our results, we developed a methodology to determine the actual period in which an individual was employed by a specific company. We analyzed again the connection graph and examined the mutual links between employees to establish their overlapping tenure within the organization. In Figure 3,

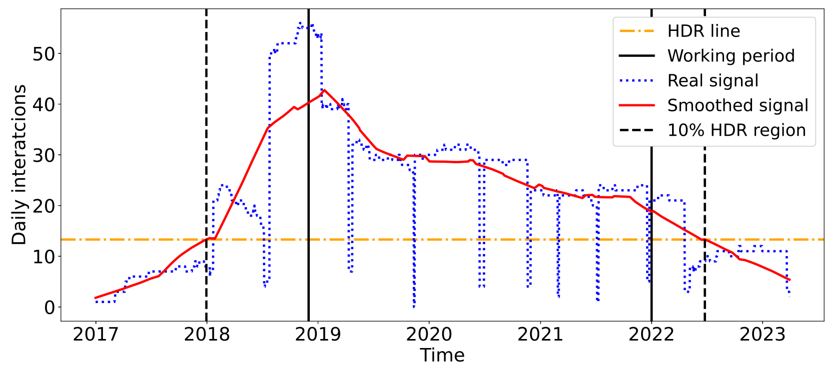

Fig. 3. Daily interactions of an employee with the others. The 10% HDR region provides a good approximation for the working period.

we present an illustrative example of a Microsoft employee's activity on Twitter during the period spanning from December 2018 to December 2022. The dotted blue line represents the actual trend of the data, which is primarily concentrated within the individual's work period, as indicated by the vertical, solid black lines. To estimate the employee's tenure, we first applied a smoothing technique to the curve by utilizing an averaging filter with a window size of 360 days, which corresponds to approximately one year. This choice was guided by the common assumption that people generally do not switch jobs within this temporal frame. We then calculated the 10% highest density region (HDR [16]) on the smoothed curve. The HDR corresponds to the intervals in which the density of a function is mostly concentrated. Specifically, the 10% HDR represents the regions in which 90% of the density is found. In the context of our study, higher densities indicate a greater number of interactions between the employee in question and her colleagues on Twitter. The dashed black lines represent the estimated working period obtained through this methodology. The results demonstrate good accuracy, with a deviation of 11 months for the starting point and 6 months for the ending point. To ensure the reliability of our findings, we applied this methodology to all employees who had at least 5 mutual connections with other employees. Under that threshold the results became less reliable. By applying this technique, we were able to derive the working periods for a significant number of employees: 2309 for Microsoft, 2294 for Google, 244 for Apple, 496 for Activision, 327 for Github, and 144 for Linkedin.

4.4 Results analysis

In Figure 4, we observe the monthly interactions between Microsoft employees and those of acquired companies, considering only the connections that occurred within the working periods detected in section 4.1. To provide a reference baseline, we also included the interactions between employees of unrelated companies and those of the acquired ones, (*i.e.,* Google and Apple).Both the graph for Activision and Github show a larger number of connections with the acquirer before the public announcement of the deal. Notably, the graphs for Activision and Github exhibit a significantly larger number of connections with Microsoft before the public announcement of the acquisition deal. Specifically, in the case of Activision, Microsoft had respectively 5 times more and 20 times more connections with respect to Google and Apple in the two years preceding the announcement (515, 107, and 25 connections). A similar trend is observed for Github, where Microsoft had 6 and 17 times more connections than the reference companies (121, 18, and 7 connections). An interesting observation about these two graphs concerns the trend of the lines. In the case of Activision, the graph for Microsoft displays a positive trend that begins around 2018, with a sudden jump in the middle of 2020. This trend suggests that the acquiring firm's interest was steadily growing and peaked a couple of years before the public announcement of the acquisition. Conversely, in the case of Github, the graph for Microsoft exhibits a sudden jump in the first half of 2017, a year before the public announcement. This trend suggests that the decision to acquire Github was more sudden and less predictable than in the case of Activision.

A different perspective emerges from the graph of Linkedin, which does not display a clear burst of activity preceding the acquisition announcement, resulting in a false negative outcome. This finding may be attributed to the low resolution of our method, which may not be sufficient for detecting lower levels of interactions. Another possible explanation could be that the available data may not be sufficiently informative in certain cases. Our methodology relies on the data available on Twitter, which provides a partial view of the world and may not capture all the relevant interactions. Interestingly enough, an article presents a compelling argument that Google had also expressed an interest in acquiring LinkedIn but later retracted its offer [17]. According to the article, the involved companies presented their interest only a couple of months before the official announcement. This temporal proximity between the expression of interest and the public announcement may explain the relatively low signal observed for both the companies in the graph.

Overall, the findings of this study highlight the potential of our methodology, as evidenced by the observed trends. Specifically, our results demonstrate that companies engaged in the acquisition process tend to participate in a dense network of relationships well in advance of any official announcement. Furthermore, our analysis of social media graphs provides a means of detecting these connections. In Section 5, we will delve further into the implications of these results and explore potential avenues for improving the efficacy of our methodology.

5 Discussion

In this section, we provide some insights on the impacts of our work and potential improvements on the core features of our methodology.

Impacts of the findings. Our developed methodology has significant implications for the research on mergers and acquisitions. One of its crucial advantages is its scalability. All of its components are automated except for the creation of a regex system, which was a one-time semi-automated task. In contrast, previous methods relied on well-formed financial databases of labeled entries, which require extensive manual work to create and maintain. The machine learning-based solutions that

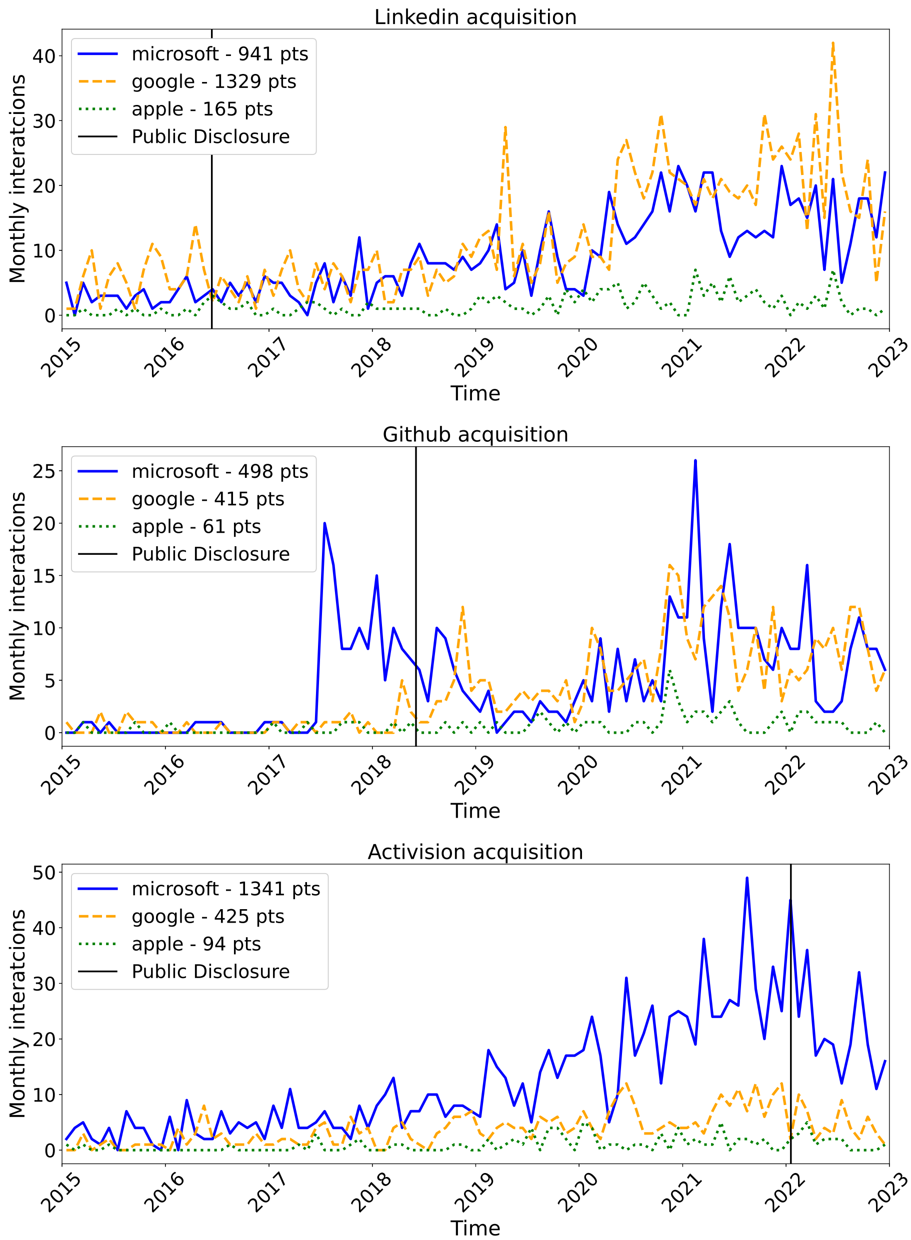

Fig. 4. Monthly interactions between companies over time. Each graph plots both the interactions between the players of an M&A operation and two baselines with the connections with unrelated companies. The trends for Activision and GitHub suggests that the interactions in a social graph can effectively help in the forecast of a M&A transaction.

use article data also suffer from sparse data and bias towards larger companies. In contrast, our methodology only requires a company to have a presence on a social platform, making it more accessible and less biased. However, the speed at which data can be collected from the Twitter website poses a real challenge. For example, to time a connection date at the desired resolution, we used a bisection algorithm applied to the retrieval of batches. By iteratively halving the requested batches, we were able to diminish the distance between the previous cursor and the next cursor, thus reaching the desired resolution. While this operation is fundamental to our approach, it is also costly, as we were able to frame followers in the range of hundreds per hour. This aspect raises important questions about financial fairness and unfair competition, as Twitter has direct access to the data and can potentially harness this advantage point to gain an edge on the market.

Improving the detection of employees. In this work, we introduced a methodology for distinguishing employees from random users on Twitter. The proposed approach relies on the analysis of profile descriptions and the use of regular expression patterns. While the method is effective to a certain extent, it is approximative and may miss some employees due to unexpected text variations. To verify our claim, we checked the pool of common followers of the Microsoft employees we identified. Among the top 50 most followed, we found 11 employees who used either a variation of our regex or a different pattern. To address this limitation, one potential alternative could be to use more sophisticated methods, such as machine learning and artificial neural networks. However, we could not use existing Natural Language Processing (NLP) tools, as they are trained on general text and would not work on the specific syntax of Twitter. Furthermore, our task requires detecting a working position and the employer, rendering traditional NLP tools unsuitable. Therefore, a potential improvement to this work would be to develop a specialized NLP model capable of improving the detection of employees on social media platforms.

Improving the detection of working periods. In addition to detecting employees, we also developed a methodology to identify the period in which an employee worked in a company. To achieve this, we leveraged the connections between employees of the same companies to frame a probable working period. Since the mobility of the labor market is high, this piece of data is fundamental in rendering the final results more reliable. However, our methodology is approximative and only provides summary data on the actual working period. To ensure meaningful information, we set the minimum connection threshold at five. However, the goodness of the results for employees with a slightly higher number of connections still lacks precision. While the connections are usually created inside the actual working period, our measure may underestimate the real extent of the working period. One potential solution to this issue could be to improve the employee detection, which would provide more connection data points. Another approach would be to integrate working periods from external sources (*i.e.,* LinkedIn). This would require acquiring the data from the external website (either by collecting it or buying it) and then developing an algorithm for matching Twitter profiles with the specific entry on the external database. This improvement could be considered for future work.

6 Conclusion

In this paper, we presented a novel methodology for forecasting mergers and acquisitions using social networks. Our focus was on three acquisitions operated by Microsoft, which we analyzed through the Twitter platform. To achieve this, we retrieved the connection dates from the followers lists and developed a method for discerning employees from random users. The results of our methodology support our claims, with the connections between Microsoft employees and acquired companies being from five to twenty times higher than baseline companies in the two years preceding the announcement. These findings demonstrate the potential of social network analysis for forecasting mergers and acquisitions. However, our results also raise concerns about financial fairness and unfair competition from companies holding the data. The speed at which data can be collected from social media platforms, such as Twitter, poses challenges and highlights the need for better strategies for gathering data. Companies that have direct access to this data have an advantage over their competitors, which could lead to unfair competition. Overall, our methodology provides a valuable contribution to the field of forecasting mergers and acquisitions, and there is ample room for further research to improve its accuracy and effectiveness.

References

1. Activision Blizzard, I.: Activision homepage. `https://www.activision.com/` (2023), accessed: 2023-01-08
2. Adelaja, A., Nayga Jr, R., Farooq, Z.: Predicting mergers and acquisitions in the food industry. Agribusiness: An International Journal **15**(1), 1–23 (1999)
3. Ali-Yrkkö, J., Hyytinen, A., Pajarinen, M.: Does patenting increase the probability of being acquired? evidence from cross-border and domestic acquisitions. Applied Financial Economics **15**(14), 1007–1017 (2005)
4. Archive, T.: Snowflake github repository. `https://github.com/twitter-archive/snowflake/tree/snowflake-2010` (2023), accessed: 2023-01-01
5. Barnes, P.: The identification of uk takeover targets using published historical cost accounting data some empirical evidence comparing logit with linear discriminant analysis and raw financial ratios with industry-relative ratios. International Review of Financial Analysis **9**(2), 147–162 (2000)
6. Center, M.N.: Microsoft to acquire linkedin. `https://news.microsoft.com/2016/06/13/microsoft-to-acquire-linkedin/` (2016), accessed: 2023-01-08
7. Center, M.N.: Microsoft to acquire github for $7.5 billion. `https://news.microsoft.com/2018/06/04/microsoft-to-acquire-github-for-7-5-billion/` (2018), accessed: 2023-01-08
8. Center, M.N.: Microsoft to acquire activision blizzard to bring the joy and community of gaming to everyone, across every device. `https://news.microsoft.com/2022/01/18/microsoft-to-acquire-activision-blizzard-to-bring-the-joy-and-community-of-gaming-to-everyone-across-every-device/` (2022), accessed: 2023-01-08
9. Cho, S., Hong, H., Ha, B.C.: A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the mahalanobis distance: For bankruptcy prediction. Expert Systems with Applications **37**(4), 3482–3488 (2010)
10. Consortium, S.: Sqlite homepage. `https://www.sqlite.org/index.html` (2023), accessed: 2023-01-01

11. Discord, I.: Discord api reference. https://discord.com/developers/docs/reference#snowflakes (2023), accessed: 2023-01-01
12. Foundation, P.: Python homepage. https://www.python.org/, accessed: 2023-01-01
13. GitHub, I.: Github homepage. https://github.com/ (2023), accessed: 2023-01-08
14. GmbH, S.: Value of mergers and acquisition (m&a) transactions worldwide from 2000 to 2022. https://www.statista.com/statistics/267369/volume-of-mergers-and-acquisitions-worldwide/ (2023), accessed: 2023-01-01
15. Gugler, K., Konrad, K.A.: Merger target selection and financial structure. University of Vienna and Wissenschaftszentrum Berlin (WZB) (2002)
16. Hyndman, R.J.: Computing and graphing highest density regions. The American Statistician **50**(2), 120–126 (1996). https://doi.org/10.1080/00031305.1996.10474359
17. Kuert, W., Mark, B.: Google and facebook also looked at buying linkedin. https://www.vox.com/2016/7/1/12085946/google-facebook-salesforce-linkedin-acquisition, accessed: 2023-01-01
18. Li, Y., Shou, J., Treleaven, P., Wang, J.: Graph neural network for merger and acquisition prediction. In: Proceedings of the Second ACM International Conference on AI in Finance. pp. 1–8 (2021)
19. Meador, A.L., Church, P.H., Rayburn, L.G.: Development of prediction models for horizontal and vertical mergers. Journal of financial and strategic decisions **9**(1), 11–23 (1996)
20. Microsoft, I.: Microsoft homepage. https://www.microsoft.com/, accessed: 2023-01-08
21. Olson, D.L., Delen, D., Meng, Y.: Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems **52**(2), 464–473 (2012)
22. Pasiouras, F., Gaganis, C.: Financial characteristics of banks involved in acquisitions: evidence from asia. Applied Financial Economics **17**(4), 329–341 (2007)
23. Ragothaman, S., Naik, B., Ramakrishnan, K.: Predicting corporate acquisitions: An application of uncertain reasoning using rule induction. Information Systems Frontiers **5**(4), 401–412 (2003)
24. Shin, K.S., Lee, T.S., Kim, H.j.: An application of support vector machines in bankruptcy prediction model. Expert systems with applications **28**(1), 127–135 (2005)
25. Slowinski, R., Zopounidis, C., Dimitras, A.: Prediction of company acquisition in greece by means of the rough set approach. European Journal of Operational Research **100**(1), 1–15 (1997)
26. Song, X.L., Zhang, Q.S., Chu, Y.H., Song, E.Z.: A study on financial strategy for determining the target enterprise of merger and acquisition. In: 2009 IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics. pp. 477–480. IEEE (2009)
27. Statista: Twitter global mdau 2022. https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/, accessed: 2023-01-01
28. Tsagkanos, A., Georgopoulos, A., Siriopoulos, C.: Predicting greek mergers and acquisitions: a new approach. International Journal of Financial Services Management **2**(4), 289–303 (2007)
29. Twitter, I.: Twitter api homepage. https://developer.twitter.com/en/docs/twitter-api, accessed: 2023-01-01
30. Twitter, I.: Twitter homepage. https://twitter.com/ (2023), accessed: 2023-01-01
31. Wei, C.P., Jiang, Y.S., Yang, C.S.: Patent analysis for supporting merger and acquisition (m&a) prediction: A data mining approach. In: Workshop on E-business. pp. 187–200. Springer (2009)
32. Xiang, G., Zheng, Z., Wen, M., Hong, J., Rose, C., Liu, C.: A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on techcrunch. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 6, pp. 607–610 (2012)