Universiteit Leiden
The Netherlands

Opleiding Informatica &Economie

Unravelling the Connections Between CEO's Online Social Networks and Firm Innovation: An X-Based Analysis

Anna Maria Elizabeth Steenbergen

Supervisors:

Dr. A. Zohrehvand

Dr. A. Saxena

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl 12/06/2024

Abstract

This study explores CEOs' social media engagement and its influence on firms' innovation. Using a sample of CEOs from S&P 1500 firms, this study documents the types of social media profiles that these CEOs follow on X (formerly known as Twitter) and employs Social Network Analysis (SNA) metrics to assess their online connections and patterns of information exchange. The study finds a high followership of news organisations and a high degree of interconnectedness among the CEOs. We argue that this interconnectedness could facilitate the exchange of ideas, knowledge, and insights, fostering a collaborative environment that can drive innovation. Our findings suggest that firms led by more socially interconnected CEOs exhibit higher levels of R&D expenditures and Capital expenditures, indicating a greater emphasis on innovation and technological advancement. This underscores the importance of social media not only as a communication tool but also as a catalyst for knowledge sharing and innovation within organisations.

Contents

1 Introduction

In today's interconnected world, the phrase "It's a small world" often highlights the frequent discovery of mutual connections among people. This phenomenon is driven by factors like social networks, professional and academic networks, increased mobility, and common interests. These extensive connections make the world feel smaller and more connected.

Scholars have devoted great attention to the CEO's social network as it can impact many different outcomes. For instance, Chief Executive Officers (CEOs) who are at the top of their firm must have a well-connected network to make optimal use of business opportunities [1]. We can refer to this capability as social capital, which stands for the accessibility of resources in a broad network of entrepreneurial partners and groups [2]. As part of our investigation into social capital, this research will focus on the online social networks of CEOs on X. These networks provide insights into the specific connections that CEOs build to acquire social capital. By having a larger social network, CEOs gain access to a broader range of information and resources, potentially leading to increased innovation within their firms. This suggests a positive correlation between a CEO's social network size and their firm's R&D and CAPEX spending.

This research aims to unravel the extent to which CEOs' online social networks on X influence their firms' innovation. we operationalise firm innovation as CAPEX and R&D costs because previous literature has shown these spending streams as solid predictors of firm innovation [1]. The CEOs were carefully selected from a pool of 1,500 S&P composite indexes. This stock market index tracks the performance of 1,500 large and mid-cap publicly traded companies in the United States [3]. These CEOs were identified in previous research [4]. The final database concerns around 500 CEOs, ensuring a representative sample from the corporate landscape. By analysing the following landscape of their X profiles we aim to get an idea of their social network and personal connections and if this has a positive impact on their R&D and CAPEX.

Previous research shows that personal connections improve a CEO's access to valuable network information, facilitating innovation by aiding in the identification, evaluation, and exploitation of new ideas [1]. By analysing social networks we can get an idea of the micro and macro levels of a network [5]. If we take this into account we can analyse the relationships between CEOs with several individuals, other CEOs and other firms or organisations. Granovetter [5] also argued the concept of weak ties, which are connections between individuals who are not very close friends or acquaintances. Granovetter [5] argued that weak ties can be more valuable than strong ties for accessing new information and opportunities. It has also been valuable to implement certain Social Network Analysis (SNA) tools into firms to improve collaboration and information sharing. SNA could also identify certain biases in decision-making [6]. Moreover, online social networks are increasingly assuming the role of offline networks and are founded on social connections [7]

Research on CEOs' social networks has shown a positive correlation between a CEO's network size and information gain [1]. However, the relationship between CEOs' online social networks and R&D investment and CAPEX is less clear. This highlights the need for a more comprehensive and rigorous examination of this phenomenon. We want to conclude if having more access to information has a positive correlation with the R&D and CAPEX of the firm. Additionally, utilising the social media platform X to explore the social networks of CEOs also reveals their private, or possibly unknown to the public, social connections and interests in specific fields.

In essence, this paper will address the identified gaps in the literature by employing a comprehensive methodology to analyse approximately 500 CEO's online social networks on X. We will employ SNA techniques to assess the number of connections, centrality, structure, and diversity of CEO networks. By correlating these network metrics with firms' R&D and CAPEX, we aim to shed light on the elusive relationship between a CEO's social capital and R&D/CAPEX investment decisions. Our analysis suggests that firms with CEOs with a better online social network tend to invest more in R&D and CAPEX, leading to increased innovation potential.

2 Related Work

Social Network Analysis (SNA)

Social Network Analysis (SNA) aims to understand social networks and their participants by focusing on the actors and the relationships between them within a specific context [8]. According to Garton, Haythornthwaite, and Wellman [9], a social network consists of individuals or organisations linked by social connections such as friendship, co-working, or information exchange. Research indicates a connection between social networks and social capital. The social capital metaphor is that the people who do better are better connected [10]. Granovetter [5] emphasised the importance of weak ties, or connections who are not close friends or family members, are often more important for information gain than strong ties. He identified that weak ties provide access to information and resources that are not available through strong ties. In addition, by using centrality measures it is possible to identify influential people in social networks.

Limitations and Methodologies in SNA

Research by Zhang et al. [11] highlights the limitations of centrality measures in complex networks. For instance, individuals with many followers on social media are often seen as influential. Nonetheless, the followers could be all clustered together, so the influence is limited to only one community, rather than extending their influence to the whole network [11]. Identifying communities within a network can reveal nodes that are more interconnected [12]. Various algorithms exist to identify these communities, and understanding influence propagation patterns enhances our comprehension of social network structures and dynamics.

Chang et al. [13] explored models like the Independent Cascade (IC) and Linear Threshold (LT) to simulate information diffusion processes. They also examined methods for evaluating user influence, maximising influence, and detecting information diffusion sources, which are crucial for applications like viral marketing and social behaviour prediction. The study highlights challenges in analysing information diffusion due to partial observation and the dynamic nature of networks, suggesting future research directions to address these issues.

Social Network Analysis applied to X

Metrics from Social Network Analysis (SNA) applied to X reveal that the platform initially functions as an information network but evolves into a social network as users gain experience [7]. Myers et al. [7] differentiates between an information network, which is based on information sharing like profiles of news sites or celebrities with no meaningful relationships between users, and a social network, where personal and meaningful real-life relationships exist between users.

In information networks, nodes have high degrees, there is a lack of reciprocity, and clustering is minimal. In contrast, social networks exhibit high-degree assortativity, short path lengths, large clustering coefficients, and a high level of reciprocity. This evolution from an information network to a social network underscores the importance of understanding how user interactions on X can develop into significant social connections.

Several research works have gained meaningful insights by analysing online social networks, specifically X social networks [14, 15, 16, 17].

CEOs Social network and Innovation

Research done by Faleye, Kovacs, and Venkateswaran [1] shows that better-connected CEOs have a greater tendency to invest in corporate innovation. By investigating around 2,366 unique CEOs from S&P 1500 firms by employing instrumental variable two-stage least squares (2SLS) regressions to manage potential endogeneity arising from unobservable heterogeneity [1]. In this research they found that CEOs with extensive networks are significantly more involved in corporate innovation. Their personal connections provide access to valuable information and implicit labour market insurance, which reduces their risk aversion and facilitates investment in risky innovation projects. They used R&D expenditures as a representable measure of the corporate innovation activity of the firms [1]. These interesting findings build a base for this current study, wherein we take the analysis one step further by implementing SNA to CEOs online social network.

Moreover, collaboration networks within firms, with direct ties, indirect ties, and structural holes create multiple benefits [18]. Direct and indirect ties positively impact innovation, though the effect of indirect ties is moderated by the amount of direct ties. Structural holes offer access to diverse information but also expose firms to potential risks.

Furthermore, certain CEO characteristics can help to have a higher firm performance. Research by Manner [19] shows that having a CEO that has traits such as having a bachelor's degree in humanities, a diverse career experience, and being female are positively related to strong corporate social performance.

The Impact of CEO X Activity on Firm Performance

Notably, not a lot of CEOs have an account on X. And if they do, they tend to be not very active on the platform [20]. Nevertheless, earlier research has suggested that if a CEO is socially active on the social media platform X, and has a large amount of followers, creates a positive impact on the short and long-term firm performance [21]. Additionally, if the firm has an active account on X, this has a positive influence on the short-term firm performance [21].

Research by Zohrehvand [22] indicates that CEOs who are active on X take more risks on Mergers & Acquisitions (M&A) deals. If CEOs are more active on social media platforms we see that they tend to be more confident in risk-taking because they have more information about growth opportunities. Social media facilitates both internal and external risk-taking behaviour for CEOs [22]. Nevertheless, increased CEO activity on social media correlates with reduced market trust in their M&A decisions. This is primarily due to the perceived distraction from core business issues, potential for inadequate due diligence, and heightened scrutiny that such activity brings, which collectively undermine confidence in the strategic soundness of their decisions [22].

Recent research by Feng and Wang [23] delves into the construction of social networks and their influence on corporate financing environments based on Chinese listed Companies. Their study explores how social network centrality correlates with variables such as R&D investment intensity and patent performance. By analysing connections within core management teams, they establish a robust positive relationship between social network strength and innovation outputs, emphasising the importance of network centrality, connection quality, and control over information flows [23]. This research highlights that beyond CEOs, the broader network within core management teams plays a critical role in shaping innovation and corporate performance.

Relevance

The exploration of CEOs' social networks, as examined in the related literature, reveals significant insights into their influence on corporate outcomes. Studies, such as those by Feng and Wang [23] and Faleye, Kovacs, and Venkateswaran [1], emphasise the positive correlation between CEOs' social network centrality and innovation outputs within organisations. By analysing connections within core management teams, these studies underscore the critical roles of network centrality, connection quality, and information flow control in fostering innovation. However, some research shows a negative association between an active CEO on X and market trust in M&A decisions [22]. This indicates that there is no clear consensus on whether strong social networks have a universally positive or negative effect on corporate activities — while they may be beneficial for innovation, they might not be advantageous for M&A decisions. Building on these findings, this study investigates how the structural characteristics of CEOs' online social networks on X influence firm performance metrics, specifically focusing on CAPEX and R&D investments.

Research Question and Hypotheses

This discussion underscores the relevance of understanding the structural characteristics of CEOs' social networks on X. Specifically, the research question guiding this study is: *What insights can we gain from the structural characteristics of CEOs' online social networks on X?*

Understanding these structural characteristics is crucial as they may influence various aspects of firm innovation. In line with this exploration, the hypotheses are formulated as follows:

H1: *The CEOs' online social network centrality is positively associated with the Capital Expenditures of the respective firm.*

H2: *The CEOs' online social network centrality is positively associated with the R&D investment of the respective firm.*

3 Methodology

This section of the paper discusses the methods we used to collect the results. First, it is documented how we collected the data from the X accounts and what applications we used. Second, we discuss how we used SNA methods to the data of the X accounts. Thirdly, we talk about how we compared the SNA results to the financial data of the firms.

3.1 Data gathering

Data gathering built on a dataset previously developed by Dr. Zohrehvand, who identified parameters (e.g. the ID number, name, username, company name, X bio, etc.) that were relevant to our research and are crucial for scraping from the API.

3.1.1 The X API

An API (Application Programming Interface) is a set of rules and protocols that allows different software applications to communicate with each other. X has a developer platform where it is possible to get access to the API. For this study, we used a newer version of the API called X API V2 [24]. Access was granted by Dr. Zohrehvand, who worked with the application in previous research. Due to organisational changes at X, the nature of the API's subscriptions changed as this paper was being developed. Namely, as of February 2023, free access to the API was prohibited, which led to the suspension and cancellation of numerous other studies that were dependent on the API [25]. Fortunately, we kept track of the organisational changes at X, which allowed us to finish our analysis, before the API was put behind a paywall.

Due to the limited access time to the still free accessible API, we were unable to easily make changes or add profiles of CEOs to the data, which affected the reproducibility of our research. Besides, there were several other limitations to using the X API V2 for this research. We could not access the following data of private accounts with our subscription. This meant that we could not include some CEOs or their followers in our analysis. Third, we used Postman as a platform to access and scrape the data from the API [26]. This added an extra layer of complexity and dependency to our data collection process. As it required setting up and managing collections, environments, and variables within Postman. Alternatively, it is possible to write some code or use the sample code from GitHub to access the API directly. Yet, this comes down to the same result, getting access to the data.

3.1.2 Scraping CEOs X Accounts

We made use of Postman to scrape the CEO's X account to get the following data of the profiles. The ID number of these CEOs is the unique identifier, this is what we need to get access to the data of this certain profile. By using the endpoint, "get_users_a_user_id_is_following", of the API, we can get the following data using the ID number of the CEO.

Parameter	Description
created_at	Gives the date and time when the account was created on X
description	Gives the bio description of the profile
entities	Makes the URLs associated with the account visible
id	Gives the ID number of the account
location	Gives the location information of the account
name	Gives the name of the account set by the user
pinned_tweet_id	ID of the tweet that is pinned by the user
profile_image_url	Gives a URL to the profile image of the user
protected	Indicates if the profile is protected or not
public_metrics	Information about the number of followers, following, and tweets
url	Gives the URL of the account of the user
username	The username of the account
verified	Shows if the profile of the user is verified or not
id_number_ceo	The ID number of the CEO

Table 1: Explanation of Used Parameters of X Accounts

Query Parameters	
User.fields	created_at, description, entities, id, location, name, pinned_tweet_id, profile_image_url, protected, public_metrics, url, username, verified
Expansions	pinned_tweet_id
Max_results	1000
Path variables	
Id	id_number_ceo

Table 2: Parameters set in Postman to access following accounts of CEOs

In table 1 and table 2 it is explained what kind of data we scraped and what each parameter means. With using this endpoint there are some limitations to the process. The endpoint has a limit of 75 requests per 15 minutes and a maximum of 1000 results per request. If a CEO was following more than 1000 users, we had to use the next token provided by the API to get the remaining data. We stored the data per request in individual JSON files. This means if a CEO follows 2000 accounts, there are 2 JSON files containing his following data.

3.2 Processing Data into a Dataframe

After scraping 495 CEO X accounts following data, it resulted in accessing 478 CEO accounts. This is because 17 accounts were private or these users deleted their accounts from X. A list of the CEOs' names is stated in A.1. All the data we scraped was saved in individual JSON files per request. Some CEOs are following more than 1000 accounts, so they have multiple JSON files. It is noticeable that these JSON files have a different structure than those of the CEOs who follow less than 1000 accounts. We needed to have a slightly different approach than the other files. To analyse all this data we made a dataframe using Python and a library from Python called Pandas. We made a dataframe consisting of a column called 'id_y' and a column called 'id_x'. Where the column 'id_y' is the ID number of the account of the CEO. The column 'id_x' is the ID number of the account the CEO is following. This way we can easily start to analyse the network and identify patterns. From the original dataset, we also combined the username of the CEO and the account they are following and the company name of the CEO's account.

3.3 Social Network Analysis (SNA)

As presented in section 2, we can see the relevance of SNA. For analysing this network, we follow the cookbook approach for as done in previous works [27, 28]. A social network is represented as a graph, denoted as G = (V, E). Here, V represents the set of nodes, encompassing all the ID numbers from the collected data. The links between these nodes, referred to as edges, are denoted as E. From section 3.2 we can see that we essentially made an edge list. Where all the ID numbers in the database are represented as nodes in the network. Each row in the database shows a directed edge that goes from the source to the target node. In this study, we work with a one-mode network. With the one-mode network, we analyse how all the nodes are connected to one another through one relation meaning followingship on X. These one-mode networks can also be structured as adjacency matrices, and can either be binary or valued [29]. With two-mode networks, we look at how nodes are tied to particular events. Two-mode networks deal with multiple datasets, whereas one-mode networks deal just with one dataset. For this study, we focus on the one-mode network of the (inter)connectedness of the CEOs. This is an unweighted graph because each edge represents a followingship on X and is not weighted with anything else, such as a number of co-liked posts, pinned tweets, or no of common followers/friends.

3.3.1 Centrality measures

To gain insights into the structure of the social network of the CEOs we looked into who are the key influencers in the network [30]. Centrality measures in Social Network Analysis (SNA) quantify the importance or prominence of nodes within a network. The measures explained in this section, help identify key influencers, assess information flow, and pinpoint crucial nodes in social networks.

Using the library NetworkX in the programming language Python [31], we can utilise multiple functions in this library that can be used to assess the key influencers in the network. These nodes are mostly characterised by having high influential power/social capital in the network. We also analysed if there are any nodes that can connect to disconnected groups. We call this capability bridging.

Degree Centrality This centrality measure is a straightforward metric that counts the total number of connections a node has. It serves as a basic indicator of popularity, though it does not distinguish between the quantity and quality of connections [32, 33]. This metric treats all connections equally, whether it is a friendship with the CEO of Apple or with your mother on X. In the context of our research, it indicates which CEOs is following the most profiles on X.

Closeness Centrality This centrality measure uses distance scores to evaluate the connectivity of individuals within a network. Closeness centrality captures the average distance from one node to all other nodes in the network, offering a different perspective from other network metrics [32]. A node with high closeness centrality is well-connected and holds a central position within the network. This means that it can quickly reach other nodes in the network, and that information can flow easily through it. Closeness centrality is a useful measure for identifying nodes that are important for information diffusion and collective action. For our research a high closeness centrality can suggest that the CEO is closely connected to many other nodes, potentially including other influential individuals, thought leaders, or industry peers. This can indicate a strong and diverse network, reflecting the CEO's engagement with a broad range of stakeholders.

Betweenness Centrality Unlike other centrality measures that focus on a node's direct connections, betweenness centrality examines a node's role in facilitating communication and information flow throughout the entire network [29]. This measure identifies nodes that serve as bridges between different parts of the network, allowing information to travel efficiently and effectively. The shortest path between two nodes is known as the geodesic distance, and it is a crucial concept in determining betweenness centrality [32]. In essence, betweenness centrality quantifies how often a node lies on the shortest path between two others. For our research, this centrality measure can be used to identify CEOs who are connectors and who play a key role in connecting different groups of people and facilitating the flow of information between them.

Eigenvector Centrality This measure provides a nuanced perspective on centrality. Even if a node has only a few connections, it can still exhibit high eigenvector centrality if those connections are with nodes that themselves are highly connected [32]. Eigenvector centrality is calculated using an iterative process that involves assigning scores to each node in the network. The score for each node is initially set to 1. Then, the scores are updated based on the scores of the node's neighbours. The more influential a node's neighbours are, the higher the node's score will be. A CEO with high eigenvector centrality is likely to have access to a wide range of information and insights.

Pagerank Centrality This centrality measure is essentially a version of Eigenvector centrality explained above. PageRank centrality is a measure of a node's importance within a network, inspired by the PageRank algorithm originally developed for ranking web pages in search engine algorithms. In the context of social networks, nodes represent individuals or entities, and edges represent connections or interactions between them (such as friendships or follower/following relationships on social media platforms). In PageRank, a node's importance is dependent on the importance of its neighbours. A high PageRank centrality for a CEO indicates a prominent position and significant influence within their network. It suggests that the CEO is well-connected to a diverse range of individuals and groups, including other influential CEOs, industry experts, and thought leaders.

3.4 Structural Analysis

Examining the centrality measures' outcomes provides only a single glimpse into the network's structure. By using community algorithms and finding the interconnectedness of the nodes we can get a more extensive insight into the structure of the network. In network terms, a community(or group) comprises nodes that are more interconnected to one another than with nodes outside the community [32].

Assortativity This measure indicates if the network is equally distributed or not. In the context of network analysis, assortativity refers to the tendency of nodes in a network to connect to other nodes that are similar to themselves in certain respects [34]. Assortativity can be measured in a variety of ways, but it is typically represented by a number between -1 and 1. A value of 1 indicates that nodes in the network are perfectly assortative, meaning that they only connect to nodes that are identical to themselves. A value of -1 indicates that nodes in the network are perfectly disassortative, meaning that they only connect to nodes that are completely different from themselves. A value of 0 indicates that there is no assortativity in the network, meaning that nodes connect to other nodes with equal probability, regardless of their similarity.

Reciprocity This measure refers to the mutual exchange of connections or relationships between nodes. In our case, this will be if the CEOs are following each other. Therefore, a high level of reciprocity means that there are a lot of CEOs that follow each other, this shows a high level of interconnectedness in the network.

K-Cores This algorithm builds on the concept of degree centrality. So, a node is part of a k-core if they have at least a degree centrality of k within that group [29, 35]. As the value of k becomes higher, subgroups will decrease, and vice-versa if k becomes lower, the size of subgroups will increase. K-cores give us information on the hierarchical structure of the network. This implies that if a CEO is in the highest core of the network, they represent the core of the network. There is a core-periphery structure, where the outermost k-cores often represent the periphery of the network and the innermost k-cores represent the core of the network. This implies that if a CEO is in a higher core value, the CEO follows at least k profiles. Or that the node is followed by at least k CEOs. The measure core number is also used to analyse the network even better and get a better look into the hierarchical structure of the social network of the CEO. We use the core_number function in the NetworkX library to get the desired results [31].

3.4.1 Community Detection

Louvain method The Louvain method detects communities in a network through an iterative, multi-step process aimed at maximising modularity. In the first step, all nodes are allocated to their community [31]. The algorithm then iteratively considers moving each node to the community of its neighbours, calculating the change in modularity for each potential move, and moves the node to the community that maximises modularity gain. This process continues until no further improvements are possible. In the second step, the network is aggregated into a new, smaller network where each community becomes a single node, with edge weights representing the sum of original edges between communities. These steps of modularity optimisation and community aggregation are repeated on the new network until no further modularity gains can be achieved, resulting in a hierarchical community structure.

Label propagation communities With this community algorithm, each node is assigned its own label. Then, each node updates its label to be the most common label among its neighbours. This process is repeated until the labels stabilise [31].

Leiden Algorithm This algorithm is an improvement on the Louvain algorithm, the algorithm focuses on refining the partitions obtained from a given community detection method. The Leiden Algorithm is often preferred for its ability to produce more accurate and stable community partitions, especially in larger more complex networks [36].

3.5 Financial Data Comparison

After looking at all the Social Network insights of the CEO's following Data from X. We want to compare this with the Financial Data from each company. There are two important main financial aspects we looked at from each firm. Namely, CAPEX and R&D investments from the financial data. From the financial dataframe, some values were missing. Consequently, we could not compare every social network from each CEO with their designated company.

"The abbreviation CAPEX stands for Capital Expenditures, which refers to the funds used by a company to acquire, upgrade, or maintain physical assets such as property, buildings, equipment, or technology "[37]. These expenditures are typically investments made by a company to increase its productive capacity or efficiency.

"The abbreviation R&D stands for Research and Development, which represents the activities undertaken by a company to innovate, create new products or services, improve existing ones, or develop new processes or technologies" [38]. R&D expenses include costs associated with personnel, facilities, equipment, and materials used in research and development activities.

These financial statistics are interesting to compare the connectivity of the CEOs with because in this way we can compare or invest if having a better social network/social CEO has a positive impact on the corporate innovation of the company. Usually, companies allocate CAPEX towards R&D activities aimed at developing new products, services, or processes. These innovations can enhance competitiveness, open new markets, and drive revenue growth.

3.5.1 Statistical Measures

We have chosen to use OLS regression for each financial parameter. This method evaluates whether a predictor variable, such as a centrality measure, significantly influences the dependent variables, CAPEX and R&D. We proceed under the following model specifications:

$$CAPEX = \alpha + \beta \times \text{Centrality Measure} + \epsilon$$

$$R\&D = \alpha + \beta \times \text{Centrality Measure} + \epsilon$$

In these models, α represents the intercept, β represents the coefficient of the centrality measure, and ϵ represents the error term. We set a significance level of 0.05 for the tests. If the resulting p-value for the coefficient β is less than 0.05, we reject the null hypothesis and conclude that the centrality measure has a statistically significant influence on CAPEX/R&D. Conversely, if the p-value exceeds 0.05, we accept the null hypothesis, indicating no significant influence.

For both CAPEX and R&D, the hypotheses are:

Null Hypothesis (H_0):

$$H_0 : \beta_{\text{Centrality Measure}} = 0$$

This means that the centrality measure has no effect on CAPEX or R&D expenditures.

Alternative Hypothesis (H_1):

$$H_1 : \beta_{\text{Centrality Measure}} \neq 0$$

This means that the centrality measure has a significant effect on CAPEX or R&D expenditures.

These tests were conducted using Python and the 'statsmodels' module Seabold and Perktold [39], which offers comprehensive tools for estimating and testing statistical models. By utilising OLS regression, we aim to understand the extent to which network centrality measures can predict CAPEX and R&D expenditures.

4 Findings & Analysis

4.1 Network Statistics

Measure	**Value**
Number of nodes	107,618
Number of edges	158,247
Average Degree	1.47
Average Out-Degree Source-Nodes	348.56
Number of nodes in the giant component	107,596
Number of edges in the giant component	158,232
Number of connected components	8
Density	1.37
Transitivity	0.00024
Average clustering coefficient	0.024

Table 3: Network Measure Values

From the results in Table 3 shown above, we can see that there are 107,618 nodes in the network and 158,247 edges in the network, which corresponds correctly with the dataframe. The 107,618 nodes represent just around 480 CEOs and the rest are other X users. This could be organisations, companies, newsletters, celebrities, family, friends etc. The average degree is relatively low at 1.47, suggesting that, on average, each node in the network follows approximately 1.47 other users in the community. When we compare this with the average out-degree from the source nodes, meaning the CEO nodes, we see that each CEO follows around 349 other profiles on the platform.

The size of the Giant Component includes almost all CEOs in the network, indicating that a significant portion of CEOs are connected in some way through their following relationships. The presence of 8 connected components may indicate that there are distinct groups of CEOs who may not be directly connected but could share common connections. Looking further at the weakly connected components, we can see that there are a few distinct CEOs who are not prominently active on the platform. This includes ajabbourbki, alfredkahn7, gregebel, kthompsonswi, mark_j_foley, rameshagilysys, and usphchris. It is noticeable that these CEOs only follow a few people on the platform with a maximum of 6 profiles.

The density being greater than 1 might suggest that there are more connections than expected in a random network. This could be due to a high level of mutual following among CEOs or other patterns in the network. The low transitivity indicates a low tendency for the CEO's followers also to follow each other, suggesting a more star-like or hierarchical structure rather than a densely interconnected community. The low average clustering coefficient further supports the idea that CEOs, on average, do not form densely interconnected clusters with their followers.

In summary, the network structure may reflect a scenario where CEOs on X have diverse following patterns, with some common connections but not forming tightly-knit communities. The low transitivity and clustering coefficient suggest a more scattered and less interconnected structure. We will be looking further into this network and we will try to understand the specifics of the network and if its nodes can provide more nuanced insights into the dynamics of CEOs and their connections on X.

4.2 Centrality Measures

Rank	**Degree Centrality**	**PageRank Centrality**	**Closeness Centrality**	**Betweenness Centrality**	**Eigenvector Centrality**
1.	Jack	WSJ	elonmusk	MichaelDell	tim_cook
2.	Benioff	elonmusk	WSJ	elonmusk	elonmusk
3.	MichaelDell	realDonaldTrump	nytimes	Benioff	satyanadella
4.	Gail_Goodman	BillGates	BillGates	lisasu	Jack
5.	Ray_zinn__	nytimes	tim_cook	Brentlsaunders	BillGates
6.	DougConant	Forbes	realDonaldTrump	Jack	Benioff
7.	wallyboston	HP	BarackObama	sundarpichai	sama
8.	adenatfriedman	FoxNews	Forbes	Davidwkenny	pmarca
9.	briandunn	SouthWestAir	TheEconomist	johnlegere	levie
10.	stephenstang	Black_KnightInc	satyandella	dkhos	Barackobama

Table 4: Top 10 Results Centrality Measures

As shown above, table 4 shows the top 10 results from the outcome of the different Centrality Measures. We can see that there is quite some resemblance between the different Centrality Measures but also quite some differences. From a first glance, we can see which nodes are the most central in the network are the biggest influencers in this network and what kind of profiles are followed the most. From table 4, we can observe that newspapers, politicians, influential people and other CEOs are prominent in the network. To take a better look at the resemblance between the different centrality measures and better compare the results with each other we took the results and measured the correlation between the different centrality measures.

The results of these tests are shown in the correlation matrix below in figure 1.

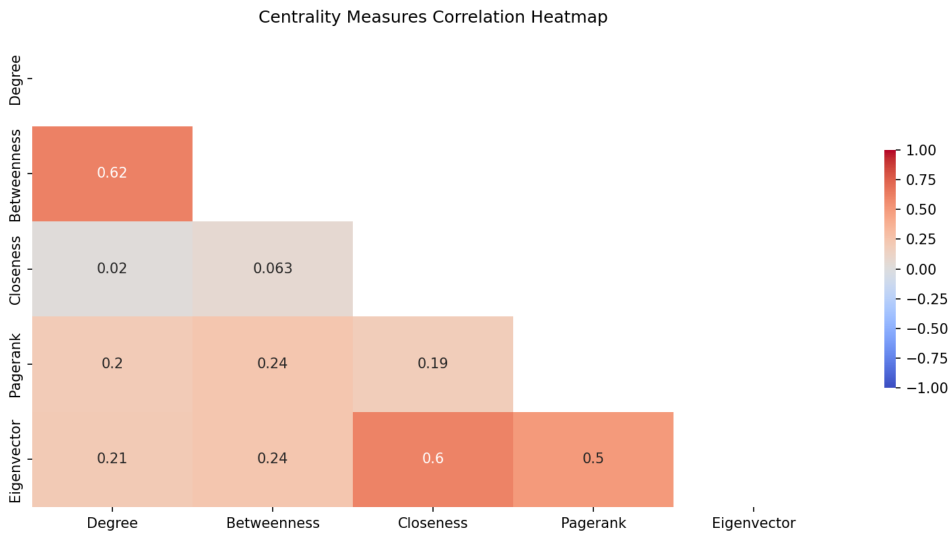

Figure 1: Correlation matrix Centrality Measures

We can see that there is a positive correlation between Degree and Betweenness Centrality, which means that it is likely if a node has a high Degree Centrality it also has a high Betweenness Centrality. There is a moderate positive correlation between the two centrality measures. We can imply that if a node is well connected. It also plays a role in facilitating communication or interactions between other users. As well as acting as a bridge or intermediary in connecting different clusters or groups within the network.

This counts the same between Eigenvector and Closeness Centrality. Also between these two centralities, we can find a moderate positive correlation. This implies that nodes that are well-connected to other important nodes in the network that have a high Eigenvector Centrality tend to also be close to other nodes in the network, nodes that have a high Closeness Centrality. This suggests that nodes with high centrality measures are not only influential due to their connections to other influential nodes but also tend to have shorter paths to reach other nodes in the network. These correlations may indicate potential advantages such as more efficient communication, better information dissemination, or greater control over the flow of information within the network. However, it is important to note that these are correlational findings, and the directionality of the effects cannot be determined from this analysis alone.

Between PageRank and Eigenvector Centrality there is also a positive correlation. These measures both measure the importance and influence of a node in a network. Nonetheless, they are calculated in different ways as mentioned in 3.3.1. It signifies the node's significance in terms of both influence and connectivity within the network, as well as its strategic positioning and role in maintaining network integrity and information flow.

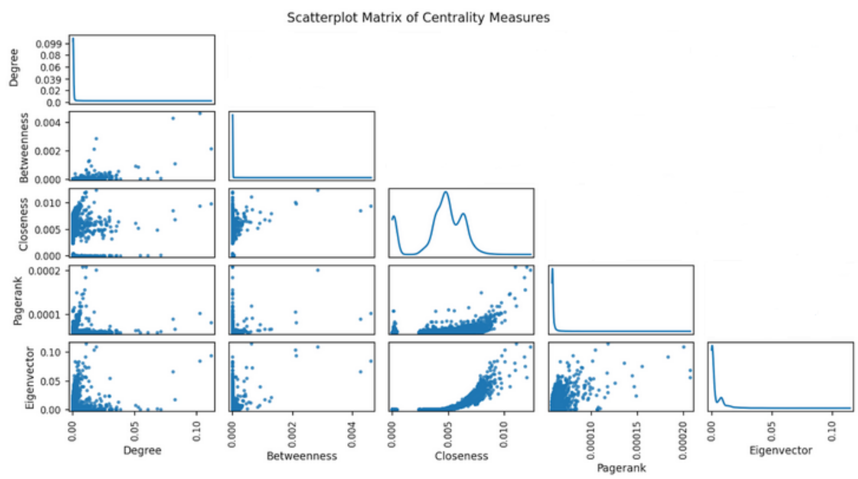

Figure 2: Scatterplot matrix Centrality Measures

In figure 2, we can see that we made a scatterplot matrix of the correlations between the Centrality Measures. This gives us a deeper analysis of the correlation between the different centrality measures. Where we can see that nodes with Betweenness and Degree centrality often exhibit a tendency to correlate, suggesting a close connection between their respective roles within the network. Also, Pagerank and Eigenvector centrality exhibit a close association, indicating that nodes influential according to Eigenvector Centrality are also likely to be influential according to Pagerank Centrality. A similar pattern emerges between Closeness and Eigenvector centrality, suggesting that nodes with short average paths to others, high Closeness Centrality, also tend to be influential, high Eigenvector Centrality.

Looking at the rest of the correlations between the centrality measures relationships appear less clear-cut, suggesting they capture different aspects of node importance within the network.

4.3 Structural Analysis

4.3.1 Assortativity

By doing some structural analysis of the network we can analyse the structure of the network. The assortativity test shows a value of -0.13 on the directed graph, the result being negative shows that the network is disassortative. This suggests that CEOs on X may not have a strong tendency to follow other CEOs with similar characteristics, but rather they might have diverse connections. In the undirected graph, the assortativity results show a value of -0.36.

4.3.2 Reciprocity

The result from the reciprocity test shows a value of 0.00293. This suggests that the connections among CEOs on X are more likely to be one-sided, meaning that if CEO A follows CEO B, it does not guarantee that CEO B will follow CEO A. This could be indicative of one-sided professional connections or a more hierarchical structure in the network. If we look at the exact number of edges in the network that are reciprocated, we get a value of 464 reciprocated edges. One valid argument for why this value is really low is that we collected only one-way directed data from the CEO's X accounts. In our analysis, we constructed a network where each CEO is connected to the accounts they follow. This means we have directed edges from CEOs to the accounts they follow. We did not include the accounts that follow the CEOs or the followers of the accounts that the CEOs follow. By examining the network, we identified the most reciprocated nodes based on the number of mutual connections. MichaelDell emerged as the most reciprocated node, followed by Benioff.

4.3.3 K-Cores

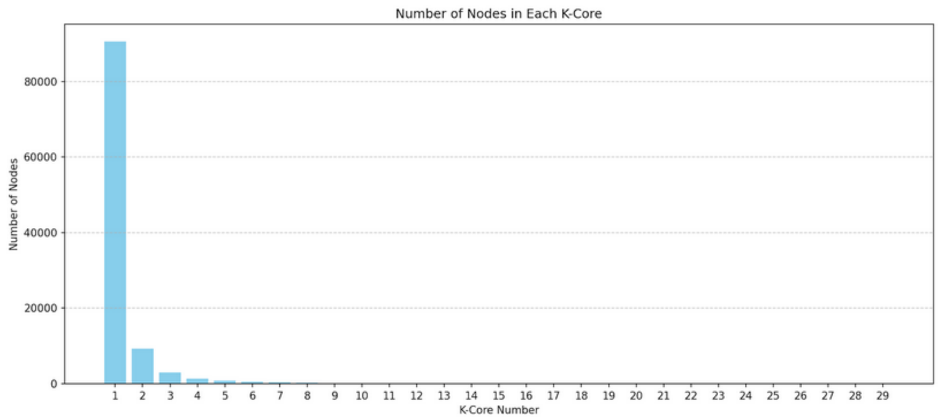

Figure 3: K-Cores distribution of Social Network

As shown in figure 3, we can see the distribution by the K-Core algorithm of the online social network of the CEOs on X. The K-Core analysis shows us a hierarchical structure of the network. We can see that there is a distribution of K-Core from core values from 1 to 29. Where there is a clear high distribution in the first core. By analysing this core further we can find that in this core there are 90,649 nodes. With the total network having an amount of 107,618 nodes. The first core represents around 84% of the network, which is a very large amount of the network. As well as the other first cores we can find large amounts of nodes, k=2 has 9,293 nodes, k=3 has 2,947 nodes and k=4 has 1,328 nodes. In the highest core k=29, the amount of nodes that are in this core is 174 nodes. By further analysing this core, we can find that there are 98 CEOs from the original dataframe in it. The rest of the 76 nodes are the profiles the CEOs are following the most. Profiles with the highest core numbers typically belong to individuals who have a wide reach, active engagement, and strong connections within the network. These profiles consist mostly of influential figures from diverse fields such as politics, media, entertainment, and technology. Since K-Core is in line with degree distribution. We can say that all the nodes in this core have a degree of more or equal to then 29.

As there are a lot of nodes in the network we can easily say to prune these nodes and concentrate on the nodes in the higher shells. Because these nodes are better connected and thus have a higher degree. There is quite a big distribution of the nodes in the different cores, being spread out till k-core = 29. So, it is possible to prune the nodes in the first k-core. With pruning this core we will end up with a dataframe with 16,971 nodes and 6,711 edges, which makes this dataframe a lot easier to handle and analyse visually.

4.3.4 Community Analysis

By analysing multiple Community algorithms, we try to identify the best algorithm fit for this network. By looking at the Modularity value we can analyse the quality of the communities. Modularity is a metric that quantifies the degree to which a network can be partitioned into distinct communities. These values range from -1 to 1, where a higher value suggests a better-defined community structure. NetworkX provides a function to calculate modularity, and we can use it to assess the Modularity value of the communities detected per community algorithm.

Algorithm	**Modularity**	**Number of Communities**
Louvain	0.70	69
Louvain (res 5)	0.69	209
Louvain (res 35)	0.67	672
Label Propagation	0.67	407
Leiden Algorithm	0.72	68

Table 5: Modularity and Amount of Detected Communities per Algorithm

From table 5, we can see that we used different community detection algorithms to evaluate which algorithm detects the best quality communities for this network. In table 5 we can see the Modularity value per algorithm and the number of communities that are detected per algorithm. Noticeable is that every algorithm has a high Modularity value, but they vary much in the number of detected communities. It is not preferable to have as many communities as 407 detected by the Label Propagation algorithm. This algorithm also shows a relatively low value of Modularity. The Leiden algorithm shows relatively the highest Modularity value of all the community detection algorithms we used. It is also noticeable that the Louvain algorithm shows a high Modularity value. Nonetheless, if we tune the algorithm using a hyperparameter, we increase the resolution of the algorithm. This affects the size of the clusters more communities are detected and the Modularity value decreases.

Even so, more research has to be done to analyse the quality of the detected communities per algorithm. In the following section, we used different metrics like cut ratio, conductance and normalised cut to assess the quality of the communities. We only used the three algorithms; Leiden algorithm, Label Propagation and Louvain algorithm to test these measures. Since the results of the higher-resolution Louvain algorithm did not show any promising results.

	Conductance		**Cut Ratio**		**Normalised Cut**	
	mean	**std**	**mean**	**std**	**mean**	**std**
Algorithm						
Label Propagation	0.318	0.152	1.060	0.762	0.319	0.153
Leiden	0.999	0.003	5.230	11.880	0.999	0.00045
Louvain	0.351	0.140	1.262	0.798	0.353	0.141

Table 6: Comparison of Algorithms based on Conductance, Cut Ratio, and Normalised Cut

If we look at the above table 6 we can see that overall Label Propagation shows the lowest values of all the measures. The Louvain algorithm is close to the scores of the Label Propagation algorithm. Nevertheless, we have to take into account that the Label Propagation algorithm detected a little bit more than 400 communities, which is a lot.

In comparison with Louvain that only detected 69 communities. As, having too many communities could lead to fragmentation, where small and less meaningful communities are detected, while too few communities could result in oversimplification, where important substructures are missed.

In comparison to the results of the modularity value, where the Leiden algorithm scored the highest, the algorithm here scores the highest values overall with these metrics. However, in comparison to the modularity metric we want to score the closest to zero with these metrics. Seeing that this means that the community that is detected shows better-indicated communities where the communities are more internally cohesive and have fewer connections between them.

If we look at some overall interesting communities detected from the Louvain algorithm which show low overall variables for the metrics. This shows a good partition of the network in communities. It is noticeable that the algorithm detects a few communities with all the metrics equal to 0, these are also the CEOs that are part of the weakly connected compartments mentioned in 4.1. These nodes are not very influential or central in the network as we have seen in the above sections. We can detect multiple communities with extremely low overall scores that are equal to 0 where there is just one source node. We can see that these CEOs are much separated from the rest of the community compared to more interconnected communities with higher scores on the metrics. By definition, these are defined as good communities because they are completely separated from the rest of the network. However, these CEOs are not very influential or central in the network as a whole. There are also some other communities detected by the Louvain algorithm that score lower values yet have only one source node. This is mostly just one CEO who follows lots of profiles on X. And is not much interconnected with the rest of the CEOs of the network. This may be because this CEO has just different interests or preferences or fields of company. For example, the first community of the Louvain algorithm with as source node “2imen”.

Metric	**Value**
Cut Ratio	0.27
Normalised Cut	0.12
Conductance	0.12

Table 7: Community Metric Results of Community CEO 2imen

From table 7, we can see that this community shows overall low on the metrics, which is good. However, it states that this CEO is more distinct in the network and is not really interconnected with the other CEOs in the network. So the algorithm detected overall this community well. However, it does not give us much insight into a distinct community of multiple CEOs. One community that we can find that has the most CEOs in one community with 62 which represents 13% of the network is quite large.

Metric	**Value**
Cut Ratio	2.081
Normalised Cut	0.476
Conductance	0.444

Table 8: Community Metrics result of community with the most CEO’s

The metrics score higher on the values yet are still a well-interconnected network in itself. If we look at the overall mean of source nodes per community = amount of CEO's / the amount of communities = 478 / 69 $\approx$ 7 CEO's per community. With the community detected by the Louvain Algorithm with the amount of 62 CEOs in this community, this is an increase of 885%. So the number of CEOs that are in this community is relatively high.

4.4 SNA and Financial data

Based on the findings above we want to combine the results with the financial data of each firm with the connectivity of each CEO. The financial data used in this section is obtained by supervisor Dr. A. Zohrehvand, as part of a bigger research guided by him. This will give us an insight if having a "better social network" or having a social CEO has a positive influence on the firm's CAPEX or R&D investments. The dataset consists of multiple sections of financial data of the respective firms.

By using the most recent year values of CAPEX and/or R&D investments of the firm, we want to compare this with the connectedness of each CEO on X. It is noticeable that some values of the column R&D investments are missing. Or that not every firm is included in the dataset from the financial data. Besides, some entries are double because multiple CEOs are connected to one firm.

We can identify 335 unique CEOs in this dataset and 285 unique companies in the dataset. There is much overlap between the CEO's and the companies. We choose to see the CEOs as individuals in the network and not merge these CEOs' social networks with each other to the respected company. In view of the fact that we want to inspect the contribution of each individual CEO's network to the company's financial metrics.

By analysing the data a little we can find that the most recent year is 2022 and the oldest year is 2006 from the financial data. After merging and combining the data we get a dataset with 127,516 edges. The values of the CAPEX can be values from 0.015 to 63,645. And for the R&D investments, we can notice amounts of 1 to 73,213. It is noticeable that a lot of values of the R&D investments are missing in the dataset. Only around 74,980 edges have values for their R&D investments and Capital expenditures. It is also noticeable that CEO JeffBezos from the company Amazon.com INC has the highest values for the R&D

investments as well as Capital Expenditures. In table 9 we can see some statistics of the dataframe where there are no missing values of the column Capital Expenditures.

	CAPEX	**R&D**
Mean	1,254.6	1,597.1
Standard Deviation	5,219.6	6,930.9
Count	285	180
Minimum	0.015	0.0
Maximum	63,645	73,213

Table 9: Financial dataframe statistics without Capital Zero

	CAPEX	**R&D**
Mean	1,567.3	1,903.9
Standard Deviation	6,681.6	7,532.4
Count	151	151
Minimum	0.015	1.0
Maximum	63,645	73,213

Table 10: Financial dataframe statistics without Capital zero and R&D zero

From table 9 and 10, we can see a slight difference in the mean and standard deviation of the dataframes. Also, the Standard Deviation of both the values is quite large concerning both Capital Zero and R&D investments. Thus, this indicates that there is a lot of variation between the values in the dataframe.

4.5 OLS Regression: Financial data and Centrality Measures

We now want to analyse if a certain centrality measure affects CAPEX and/or R&D. We analysed the following centrality measures: Degree, Betweenness, Closeness, PageRank and Eigenvector Centrality. We needed to plot these measures again in combination with the CAPEX and R&D variables in the new dataset. Since, not all CEOs have corresponding financial data, unlike the original dataset we analysed. We also use a standardisation of the centrality measures. This allows for more meaningful comparisons and interpretations, as the coefficients from regression analyses become more interpretable in terms of standard deviations.

Centrality	Dependent Variable	R-squared	Coefficient	Intercept	P
Degree	CAPEX	0.005	-10.31	1409.91	0.000
Degree	R&D	0.006	-12.71	1713.79	0.000
Closeness	CAPEX	0.094	1086.91	658.69	0.000
Closeness	R&D	0.080	1086.91	658.69	0.000
Eigenvector	CAPEX	0.096	267.72	440.86	0.000
Eigenvector	R&D	0.091	267.72	440.86	0.000
Betweenness	CAPEX	0.018	13.31	811.83	0.000
Betweenness	R&D	0.010	13.22	913.72	0.000
PageRank	CAPEX	0.134	298.29	462.29	0.000
PageRank	R&D	0.174	476.90	123.61	0.000

Table 11: OLS Regression Results Summary

From table 11 we can see all the different results from the OLS regression. The results of the influence of the centrality measures on the financial metrics show different results. It is noticeable that Degree Centrality shows a statistically significant negative relationship between CAPEX and R&D. With each unit increase in Degree Centrality resulting in a decrease of approximately 10.31 units in CAPEX and approximately a decrease of 12.71 units in R&D. However it is crucial to note that for both these OLS regressions the R-squared value is very low 0.005 for CAPEX and 0.006 for R&D, indicating that Degree Centrality alone explains a very small portion of the variation in CAPEX and R&D.

For Closeness, Eigenvector and Betweenness centrality we can identify a statistically significant positive relationship. Nevertheless, the R-squared value for these centrality measures is again very low. This suggests that these centrality measures explain only a small fraction of the variability in R&D and CAPEX.

Nonetheless, if we look at the centrality measure called PageRank. It shows the highest R-squared value for both CAPEX and R&D. With CAPEX being 0.134 and R&D being 0.174. With the p-value for both dependent variables being 0.000 we can say that the null hypothesis can be rejected and the alternative hypothesis can be accepted instead. This means that the regression analysis indicates a statistically significant positive relationship between PageRank Centrality and R&D/CAPEX. Specifically, higher PageRank Centrality is associated with higher R&D/CAPEX.

For R&D we can see an intercept 123.61. This represents the expected value of R&D when PageRank Centrality is zero. The coefficient estimate of 476.9 suggests that for each unit increase in PageRank Centrality, R&D is expected to increase by approximately 476.9 units.

For CAPEX we can see an intercept of 462.29 suggests the expected value of CAPEX when PageRank is zero. The coefficient estimate of 298.2920 indicates that for every unit increase in PageRank, CAPEX is expected to increase by approximately 298.29 units.

We now get the following formulas:

$$CAPEX = 462.29 + 298.29 \times \text{PageRank} + \epsilon$$

$$R\&D = 123.61 + 476.90 \times \text{PageRank} + \epsilon$$

While the R-squared value is still a small portion of what CAPEX and R&D can be concluded by, there are still multiple reasons why this financial measure can be high. Nonetheless, the results of the OLS regression test with PageRank Centrality support hypotheses 1 and 2. From figures 4 and 5, we can see that the regression line follows a positive relationship between the financial metrics and Pagerank Centrality. The data points in both the figures are quite scattered, this is in line with the relatively low R-Squared value.

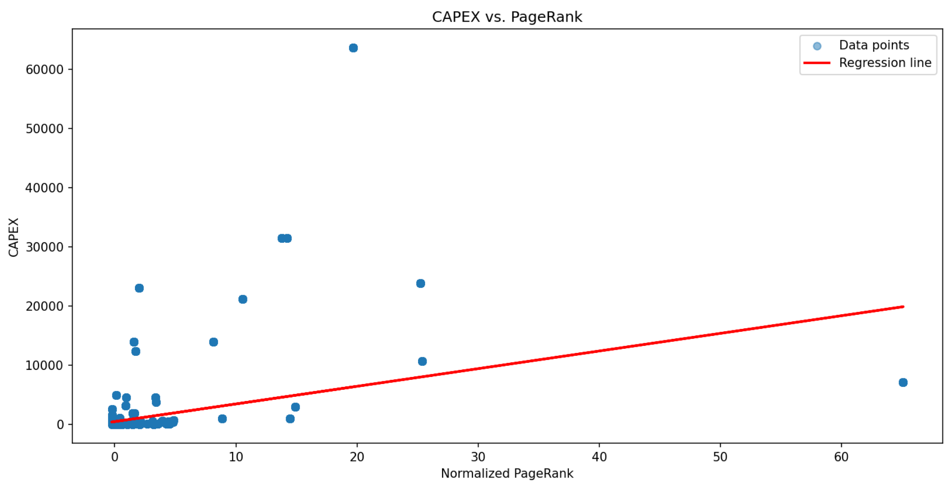

Figure 4: Regression results CAPEX and PageRank Centrality

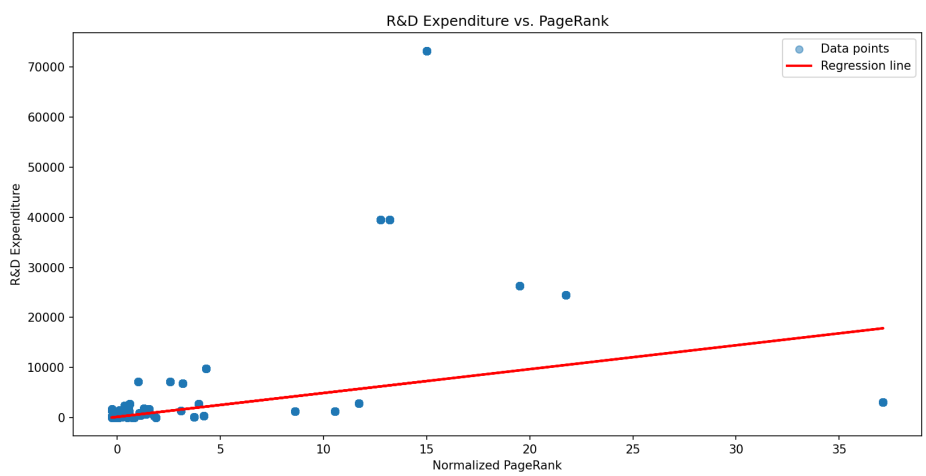

Figure 5: Regression results R&D and PageRank

5 Conclusions & Recommendations

5.1 Conclusions

In this study, we employed Social Network Analysis (SNA) techniques to delve into the X network of CEOs from S&P firms. Specifically, we utilised SNA metrics such as centrality measures, assortativity, reciprocity, and community analysis to uncover insights into the structure and dynamics of the CEO network. Additionally, we integrated financial data to explore the relationship between network centrality and strategic financial decisions, such as CAPEX and R&D investment. Through these analyses, we aimed to provide a comprehensive understanding of the role of social connectivity in shaping leadership influence and corporate innovation.

The network we found consists of 107,618 nodes and 158,247 edges, with CEOs representing a small fraction of the total nodes but being well-connected, as evidenced by their higher average out-degree. The low average degree and clustering coefficient, coupled with high density, suggest that the network is sparsely connected with a star-like or hierarchical structure rather than a tightly-knit community. This is further supported by the presence of a large giant component, which includes almost all CEOs, indicating a significant level of interconnectivity among them. However, the existence of eight connected components and weakly connected CEOs suggests some degree of isolation or selective connectivity among a few individuals.

The centrality analysis identifies key influencers within the network, with notable figures such as Jack, Elon Musk, and WSJ featuring prominently across multiple centrality measures. The correlations between different centrality measures indicate that nodes with a high Degree Centrality often also have high Betweenness Centrality, acting as crucial bridges within the network. This indicates that CEOs who are well-connected to many other nodes in the network also act as key bridges that connect different parts of the network. This dual role allows them to influence a large number of people directly while also controlling the flow of information between various subgroups within the network. Similarly, Eigenvector and Closeness Centralities show moderate positive correlations, suggesting that well-connected nodes are also centrally positioned, facilitating efficient communication and information flow. This indicates that this combination positions the CEOs to effectively distribute information and leverage their connections to other powerful individuals, thereby maximising their influence and access to valuable resources and information.

Assortativity and Reciprocity tests indicate a disassortative and predominantly one-sided connection pattern among CEOs, highlighting the lack of mutual following and suggesting a more hierarchical or diverse connection approach. The K-Core analysis reveals a hierarchical structure with a significant portion of nodes in the lower cores, but the most influential nodes residing in higher cores. This allows for pruning less significant nodes to focus on the core influencers, simplifying the network analysis, and showing that a few influential people occupy the top central position. Community detection algorithms, particularly the Leiden algorithm, demonstrate high Modularity, indicating well-defined communities within the network. However, the varying number of communities detected by different algorithms suggests that further analysis is needed to determine the most meaningful community structures. The analysis of specific communities shows that influential CEOs tend to form distinct clusters, while less connected CEOs remain more isolated.

The integration of financial data with network centrality measures reveals interesting findings. The OLS regression shows statistically significant differences in financial metrics for CEOs with higher centrality measures, particularly PageRank Centrality. This suggests that a CEO's social connectivity on X may positively influence their firm's innovation.

Addressing the research question, *What insights can we gain from the structural characteristics of CEOs' online social networks on X?*, our findings indicate that the structural characteristics of CEOs' social networks on X are instrumental in understanding their influence on firm performance. Specifically, we find that centrality measures such as PageRank are strong indicators of a CEO's potential to drive CAPEX and R&D investments, thereby supporting our hypotheses:

H1: *The CEOs' online social network centrality is positively associated with the Capital Expenditures of the respective firm.*

H2: *The CEOs' online social network centrality is positively associated with the R&D investment of the respective firm.*

In conclusion, the X network of CEOs is characterised by a sparse, hierarchical structure with key influencers playing central roles. The centrality measures provide valuable insights into the network's dynamics, highlighting the importance of well-connected CEOs. The positive relationship between network centrality and financial metrics suggests a link between a CEO's online influence and their firm's innovation. These findings underscore the significance of social networks in understanding leadership influence and corporate strategy, paving the way for further research into the impact of social connectivity on organisational innovation.

5.2 Recommendations

To benefit from these insights, it is recommended that CEOs enhance their engagement on X by interacting with a diverse range of profiles, including other CEOs, industry leaders, and influential personalities. This can have positive results for their firm. As mentioned by previous research of Feng and Wang [23], it is not only the CEO's network that can have a positive impact on the business results of the respective business. Nonetheless, directors, supervisors and executives' social network extending the "CEO network" to the "core management team network" can have a positive effect on the innovation of a business [23].

Focusing on high-impact nodes within the network is crucial. By using centrality measures like Eigenvector and PageRank, companies can identify and engage with these key influencers. This can foster valuable collaborations and partnerships. Companies should map their CEO's social network to pinpoint strategic connections and leverage these for business benefits.

Integrating social network analysis into strategic decision-making processes offers valuable insights into how CEO connections influence business innovation. By analysing the CEO's network, companies can identify collaboration opportunities, strategic partnerships, and knowledge-sharing channels, leading to increased innovation and competitive advantage. This approach aids in better decision-making, resource allocation, and risk management, driving higher CAPEX and R&D investments and improving financial performance. Regular monitoring and adaptation of network structures ensure companies stay agile and responsive to new opportunities and challenges, leveraging CEO connections for enhanced business outcomes.

5.3 Further Research

Reflecting on our study, we focused on examining whom CEOs follow on X. For future investigations, exploring the broader core management team beyond just the CEO presents a more comprehensive view of a firm's representation. Additionally, it is important to note that access to the X API is no longer freely available. Consequently, extending or initiating new research using X for SNA may pose challenges in terms of reproducibility and accessibility. However, exploring other online social media platforms such as LinkedIn or Instagram could provide other valuable insights into CEOs online social networks. Next to that, while the R-squared value is still relatively low between the positive relationship of PageRank and R&D and CAPEX, further research is needed to explain this relationship better. For example, by using control variables that might influence R&D and CAPEX, like company size, revenue, industry sector etc.

One other recommendation is including longitudinal studies to assess the long-term impact of CEO online social network positions on firm innovation. Noticeably, the data has been scraped in February 2023, since then there have been multiple changes to the online social network of the CEOs. For example, one CEO called 'Jack' who first followed the most profiles in this network, now only follows 3 profiles on X. Next to that, analysing CEO networks within specific industries can reveal unique characteristics and inform tailored strategies. Exploring the role of external factors like market trends and economic conditions can offer a deeper understanding of influencing variables. Comparative studies of CEO networks across different social media platforms can identify the most effective channels for engagement. In-depth community analysis can provide nuanced insights into the dynamics of CEO networks while investigating the impact of CEO demographics can help design targeted interventions to enhance connectivity. Research on network pruning techniques can optimise the balance between network size and the quality of insights. As we have seen in this study, not all connections are required to analyse the network as a whole.

Additionally, it is crucial to expand the scope of social network analysis to include other key individuals within the company, such as CFOs, COOs, and other top executives. By understanding the social network positions of these influential figures, companies can gain a more comprehensive view of the internal dynamics that drive strategic decisions and overall firm performance. Investigating how the interactions and connections among various top executives influence CAPEX, R&D investments, and other strategic outcomes can provide deeper insights and more robust strategies for enhancing organisational effectiveness.

By implementing these recommendations and pursuing further research, companies can better leverage the power of not only CEO social networks but also the networks of other critical executives to drive business success and gain a competitive edge in the market. In the realm of research, exploring more on how the network positions of executives influence innovation, market expansion, and crisis management. Analysing the interplay between centrality measures and business performance can lead to precise models for predicting the impact of executive connectivity on company success, informing targeted strategies to harness leadership's social capital effectively.

References

[1] O. Faleye, T. Kovacs, and A. Venkateswaran. "Do Better-Connected CEOs innovate More?" In: *Journal of Financial and Quantitative Analysis* 49.5–6 (2014), pp. 1201–1225. DOI: 10.1017/s0022109014000714.

[2] A. Kaasa. "Effects of different dimensions of social capital on innovative activity: evidence from Europe at the regional level". In: *Technovation* 29.3 (2009), pp. 218–233. DOI: 10.1016/j.technovation.2008.01.003.

[3] S&P Dow Jones Indices. *S&P Composite 1500®*. URL: https://www.spglobal.com/spdji/en/indices/equity/sp-composite-1500/#overview.

[4] A. Zohrehvand. "Fifty Million Followers Can't Be Wrong, or Can They? Effects of Social Media Feedback on CEO Communication". In: *Social Science Research Network* (June 2022). DOI: 10.2139/ssrn.4141444.

[5] M. Granovetter. "The strength of weak ties". In: *American Journal of Sociology* 78.6 (1973), pp. 1360–1380. DOI: 10.1086/225469.

[6] R. Cross, S. P. Borgatti, and A. Parker. "Making invisible work visible: Using social network analysis to support strategic collaboration". In: *California Management Review* 44.2 (2002), pp. 25–46. DOI: 10.2307/41166121.

[7] S. A. Myers et al. "Information network or social network?" In: *Proceedings of the 23rd International Conference on World Wide Web*. 2014, pp. 899–904. DOI: 10.1145/2567948.2576939.

[8] O. Serrat. "Social Network Analysis". In: *Knowledge Solutions*. Springer, 2017, pp. 39–43. DOI: 10.1007/978-981-10-0983-9_9.

[9] L. Garton, C. Haythornthwaite, and B. Wellman. "Studying Online Social Networks". In: *Journal of Computer-Mediated Communication* 3.1 (2006), pp. 0–0. DOI: 10.1111/j.1083-6101.1997.tb00062.x.

[10] R. S. Burt. "The network structure of social capital". In: *Research in Organizational Behavior* 22 (2000), pp. 345–423. DOI: 10.1016/s0191-3085(00)22009-1.

[11] X. Zhang et al. “Identifying influential nodes in complex networks with community structure”. In: *Knowledge-based Systems* 42 (2013), pp. 74–84. DOI: 10.1016/j.knosys.2013.01.017.

[12] I. Himelboim. “Social Network Analysis (Social Media)”. In: *The International Encyclopedia of Communication Research Methods*. Wiley-Blackwell, 2017. DOI: 10.1002/9781118901731.iecrm0236.

[13] B. Chang et al. “Study on Information Diffusion Analysis in Social Networks and Its Applications”. In: *International Journal of Automation and Computing* 15.4 (2018), pp. 377–401. DOI: 10.1007/s11633-018-1124-0.

[14] Ivan Bermudez et al. “Twitter response to Munich July 2016 attack: Network analysis of influence”. In: *Frontiers in big Data* 2 (2019), p. 17.

[15] Michael Paul and Mark Dredze. “You are what you tweet: Analyzing twitter for public health”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 5. 1. 2011, pp. 265–272.

[16] Mariana Macedo and Akrati Saxena. “Gender differences in online communication: A case study of Soccer”. In: *arXiv preprint arXiv:2403.11051* (2024).

[17] Maneet Singh et al. “A bi-level assessment of twitter data for election prediction: Delhi assembly elections 2020”. In: *Companion Proceedings of the Web Conference 2022*. 2022, pp. 930–935.

[18] G. Ahuja. “Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study”. In: *Administrative Science Quarterly* 45.3 (2000), pp. 425–455. DOI: 10.2307/2667105.

[19] M. Manner. “The impact of CEO characteristics on corporate social performance”. In: *Journal of Business Ethics* 93.S1 (2010), pp. 53–72. DOI: 10.1007/s10551-010-0626-7.

[20] P. Capriotti and L. Ruesja. “How CEOs use Twitter: A comparative analysis of global and Latin American companies”. In: *International Journal of Information Management* 39 (2018), pp. 242–248. DOI: 10.1016/j.ijinfomgt.2018.01.003.

[21] W. Ghardallou. “The impact of firms’ and CEOs’ social media usage on corporate performance”. In: *Investment management financial innovations* 18.4 (2021), pp. 21–35. DOI: 10.21511/imfi.18(4).2021.03.

[22] A. Zohrehvand. “Do social media influence CEOs’ strategic decisions? CEOs’ Twitter activity and subsequent M&As”. In: *Proceedings - Academy of Management*. Vol. 2021. 1. 2021, p. 15419. DOI: 10.5465/ambpp.2021.15419abstract.

[23] G. Feng and J. Wang. “The Impact of Corporate Social Network on Innovation: A Mediation Analysis of Agency Costs and Financial Constraints”. In: *Journal of Mathematical Finance* 10.04 (2020), pp. 536–568. DOI: 10.4236/jmf.2020.104033.

[24] Twitter API Documentation. *Docs — Twitter Developer Platform*. URL: https://developer.twitter.com/en/docs/twitter-api.

[25] S. G. Gotfredsen. “Q&A: What happened to academic research on Twitter?” In: *Columbia Journalism Review* (Dec. 2023). URL: https://www.cjr.org/tow_center/qa-what-happened-to-academic-research-on-twitter.php.

[26] Postman API Platform. *Postman API Platform*. URL: https://www.postman.com/.

[27] Akrati Saxena et al. “The banking transactions dataset and its comparative analysis with scale-free networks”. In: *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2021, pp. 283–296.

[28] Jukka-Pekka Onnela et al. “Analysis of a large-scale weighted network of one-to-one human communication”. In: *New journal of physics* 9.6 (2007), p. 179.

[29] C. Prell. *Social network analysis: History, Theory and Methodology*. SAGE, 2012.

[30] A. Saxena and S. Iyengar. “Centrality measures in complex networks: A survey”. In: *arXiv preprint arXiv:2011.07190* (2020). DOI: 10.48550/arXiv.2011.07190.

[31] NetworkX. *NetworkX Documentation*. URL: https://networkx.org.

[32] D. Hansen, B. Shneiderman, and M. A. Smith. *Analyzing social media networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann, 2010.

[33] Akrati Saxena, Ralucca Gera, and SRS Iyengar. “Estimating degree rank in complex networks”. In: *Social Network Analysis and Mining* 8.1 (2018), p. 42.

[34] M. E. J. Newman. “Mixing patterns in networks”. In: *Phys. Rev. E* 67 (2 Feb. 2003), p. 026126. DOI: 10.1103/PhysRevE.67.026126. URL: https://link.aps.org/doi/10.1103/PhysRevE.67.026126.

[35] Akrati Saxena and SRS Iyengar. “K-shell rank analysis using local information”. In: *International Conference on Computational Social Networks*. Springer. 2018, pp. 198–210.

[36] V. A. Traag, L. Waltman, and N. J. Van Eck. “From Louvain to Leiden: Guaranteeing well-connected communities”. In: *Scientific Reports* 9.1 (2019), pp. 1–12. DOI: 10.1038/s41598-019-41695-z.

[37] J. Berk and P. DeMarzo. *Corporate Finance, Global Edition*. Pearson UK, 2019.

[38] J. J. Weygandt, P. D. Kimmel, and D. E. Kieso. *Financial Accounting with International Financial Reporting Standards*. John Wiley & Sons, 2018.

[39] S. Seabold and J. Perktold. “Statsmodels: Econometric and Statistical Modeling with Python”. In: *Proceedings of the Python in Science Conferences* (2010). URL: https://doi.org/10.25080/majora-92bf1922.

A Appendices

A.1 CEO List names

CEO Name	CEO Name	CEO Name	CEO Name
Ayman S. Ashour	Aart J. de Geus	Atul Bhatnagar	Adam M. Contos
Anthony DeChellis	Adena T. Friedman	Ahmad R. Chatila	Anthony M. Jabbour
Alan B. Masarek	Alan David Schnitzer	Alfredo Bala	Albert Bourla
Alexander Lidow	Alfred R. Kahn	Amar Hanspal	Alex A. Molinaroli
Andrea J. Ayers	Andrea Jung	Andres Ricardo Gluski Weilert	Andrew Anagnost
Andrew B. Benett	Andrew F. Puzder	Ann D. Murtlow	Antonio J. Pietri
Antonio Fabio Neri	Aaron P. Graft	Adam David Portnoy	Archie C. Black
Aron J. Ain	Arthur T. Shorin	Arthur Peck	Art Zeile
Adam P. Symson	August A. Busch	Patrick W. Smith	William R. Wagner
Romil Bahl	Bahram Akradi	Barry C. McCarthy	Bassil I. Dahiyat
William D. Jenkins	A. Patrick Beharelle	Benjamin Feder	Marc R. Benioff
Benjamin Wolin	Bethany M. Owen	Robert C. Biesterfeld	William P. Angrick
William L. Ballhaus	William Joseph Hornbuckle	William R. McDermott	William J. Wackermann
Brian Jeffrey Lipke	Brian A. Napack	Robert A. McDonald	Robert A. Kotick
Robert P. Carrigan	Robert C. Paul	Robert H. Swan	Brad D. Smith
Brenton L. Saunders, J.D.	Brett T. Ponton	Brian J. Dunn	Shelley G. Broader

CEO Name	CEO Name	CEO Name	CEO Name
Bruce Dale Broussard	Bruce J. Schanzer	Bryan R. Martin	Brian Slobodow
Bryan J. Wiener	Cheryl A. Bachelder	Candace B. Kendle	Carl Bass
Carlos M. Cardoso	Carol B. Tom	Charles W. Berger	William J. Grubbs
Douglas M. Baker	Michael Dale Hayford	Timothy G. Baxter	Chad Dickerson
Chaim K. Katzman	Charles John Wilder, Jr.	Charles J. Meyers	Kenneth I. Chenault
Christopher H. Franklin	Christopher T. Holmes	Christopher J. Abate	Christopher J. Nassetta
Christopher Oddleifson	Charles H. Robbins	Clifford B. Bleustein	Cabell H. Lolmaugh
Christopher A. Caldwell	Joseph Hugh Moglia	Colin M. Angle	Frank William Conner
	Kimberly S. Lubel	Craig R. Herkert	Craig R. Dahl
Craig Thomas Bouchard	Daniel E. Greenleaf	Damon T. Hininger	Daniel H. Schulman
David A. Wentz	David W. Bernauer	David M. Shull	David Bruton Smith
David E. Flitman	David J. Field	David James Henshall	David Liu
David Gary Neeleman	David C. Novak	David S. Kalt	David Michael Solomon
David W. Hult	David W. Kenny	Dawn M. Zier	David W. Crane
Darren D. Hawkins	Dean M. Drako	Derek J. Leathers	Devin N. Wenig
Diane M. Irvine	Richard Costolo	Dinesh C. Paliwal	Dion J. Weisler
Dirk Van de Put	Daniel J. Oh	Dara Khosrowshahi	Donald L. Blankenship
David D. Ossip	Douglas C. Bryant	Douglas R. Conant	Douglas R. Lebda

CEO Name	CEO Name	CEO Name	CEO Name
S. David Passman	Eileen P. Drake	Joseph Mario Molina	Thurman John Rodgers
Michael R. MacDonald	David J. Wagner	Eddie Capel	Douglas R. Waggoner
Edmund M. Ingle	Ellen J. Kullman	Elmer N. Baldwin	Elon R. Musk
Enrique J. Lores	Enrique T. Salem	Eamonn P. Hobbs	Deborah Dunsire
Eric E. Schmidt	Eric Howard Starkloff	Peter C. Farrell	Francis A. deSouza
Fernando G. Aguirre	Filip J. L. Gyd	Mark Elliot Zuckerberg	Frank M. Mastrangelo
Francis Lobo	Francis S. Blake	Gail F. Goodman	Kieran T. Gallahue
Gary D. Burnison	Gary C. Kelly	Gary Adam Norcross	Gary C. Butler
Gavin D. K. Hattersley	Michael Gavin Isaacs	Geisha Jimenez Williams	Aaron P. Jagdfeld
George R. Oliver	George A. Zimmer	Gerald L. Storch	Gian M. Fulgoni
Virginia C. Drosos	Virginia M. Rometty	Glen E. Tullman	Gregory Q. Brown
Gregory Lorne Ebel	Gregory R. Gianforte	Gregory S. Marcus	G. Tyson Tuttle
Guy Gecht	Graham M. Weston	Richard Brian Handler	Hans E. Vestberg
Hartmut Liebel	Hassane S. El-Khoury	Hessam Nadji	Hikmet Ersek
Howard L. Lance	Howard D. Schultz	Daniel Bruce Hurwitz	Michael C. Dennison
John H. Schnatter	Herve Hoppenot	Indra K. Nooyi	Matthew J. Desch
Irving L. Azoff	James C. Collins	Jack Dorsey	James A. Hyde

CEO Name	CEO Name	CEO Name	CEO Name
James J. Lerner	Jason D. Lippert	John N. Roberts	James F. Brear
Jeremy Burton	J. Clifford Hudson	James C. Ryan	James D. Rollins
Jeffrey P. Bezos	Jeff Rosica	Jeffrey L. McWaters	Jeffrey Adam Citron
Jeffrey M. Ettinger	Jeffrey J. Clarke, M.B.A.	Jeffery W. Yabuki	Jeremy Stoppelman
John Philpin	Joshua G. Silverman	James C. Fish	James R. Fitterling
James E. Heppelmann	James E. Rogers	James M. Taylor	Jonathan E. Lim
Joseph C. Magnacca	Joseph F. Eazor	Joseph B. Burton	Joseph A. Ripp
Joel N. Agree	John Aurelius Addison	John B. Richards	John S. Chen
John L. Flannery	John J. Haley	John W. Kett	John J. Legere
John T. Chambers	Jonathan S. Huberman	Jonas Prising	Jonathan David Klein
Jonathan Oringer	Jory J. Marino	Jose Luis Laparte	Jose Ramon Mas
James Robert Anderson	Jeffrey T. Housenbold	Jonathan W. Gacek	James M. Whitehurst
Karen A. Puckett	Karl McDonnell	Peter Karmanos, Jr.	Stephen Kaufer
Kenneth Dale Naumann	Kenneth D. Denman	Kent J. Thiry	Ken Xie
Keri L. Jones	Kevin R. Johnson	Kirill Tatarinov	Klaus P. Besier
Kevin B. Thompson	Lance M. Fritz	Lance Uggla	H. Lawrence Culp
Lawrence J. Ellison	Lawrence Michael Raffone	Garry O. Ridge	Louis Hernandez
Lip-Bu Tan	Lisa T. Su	Lisa W. Wardell	Andrew N. Liveris
Lawrence E. Kurzius	Lloyd C. Blankfein	A. Lanham Napier	Linda Perneau

CEO Name	CEO Name	CEO Name	CEO Name
Lynn Michelle Jurich	Magid M. Abraham	Michael S. Burke	Milton G. Silva-Craig
Jeffrey A. Joerres	Raul Marcelo Claure	Marc B. Lautenbach	Marec Elden Edgar
J. Mariner Kemper	Mark J. Foley	Mark D. McLaughlin	Mark L. Schiller
Mark D. McClain	Mark C. Miller	Mark D. Okerstrom	Mark Vincent Hurd
Marvin R. Ellison	Mark K. Mason	Matthew V. Booty	Mayo A. Shattuck
Mel Karmazin	Menno Ellis	Michael Rapino	Michael D. Capellas
Michael Saul Dell	Michael W. Yackira	Michel Combes	Micky Meir Arison
Sanjay Mehrotra	Mihael H. Polymeropoulos	Michael P. Gianoni	Michael R. Minogue
G. Michael Sievert	Cindy J. Miller	Mindy F. Grossman	Sanjay Mirchandani
Dwight Mitchell Barns	Michael J. Tattersfield	Michael L. Reger	Montgomery F. Moran
Matthew E. Rubel	Colin Shannon	Mark Aslett	Michael F. Steib
Mary T. Barra	Mark Thomas Bertolini	Bruno Guilmart	George Kurian
Mark R. Goldston	Nicholas J. DeIuliis	Travis Nigel	Gary L. Carano
Nora M. Denzel	Mauricio Gutierrez	Omar S. Ishrak	Richard D. Snyder
Jonathan I. Schwartz	Timothy J. O'Shaughnessy	Judith F. Marks	Gregory S. Clark
Patrick K. Decker	Patrik Frisk	Paul E. Jacobs	Paul S. Galant
Paul A. Hooper	Paul Berthold Kusserow	Paul J. Sarvadi	Pehong Chen
Parker H. Petit	Peter J. Ungaro	Peter W. Quigley	Phaneesh Murthy
Philip C. Mezey	Robert W. Pittman	Paul K. Yonamine	Ramesh Srinivasan

CEO Name	CEO Name	CEO Name	CEO Name
Ramon Luis Laguarta	Randall J. Garutti	Oliver Ratzesberger	Ravichandra K. Saligram
Raymond D. Zinn	Andrew Wilson	Wilmot Reed Hastings	Michael J. Farrell
Richard L. Federico	Richard D. Fain	E. V. Goings	Richard M. Olson
Richard A. Meeusen	Robert M. Lynch	Robert W. D'Loren	Robert A. Iger
Robin Raina	Robert P. LoCascio	Rodney Hershberger	Roderick Christie
Rod R. Little	Roger L. Rawlins	Rohit Kapoor	Rolla P. Huff
Ronald D. Boire	Ron Cohen	Howard C. Root	Roy H. Bubbs
Randall S. Dearth	Keith Rupert Murdoch, AC	John R. Frantz	Robert Thomas Wallstrom
Salomon Sredni	Thomas Sandgaard	Sara A. Greenstein	Sasan K. Goodarzi
Satya Nadella	Michael J. Saylor	Simon Biddiscombe	Scott Freidheim
Scott G. McNealy	Stephen D. Lebovitz	Sean Michael O'Connor	Selim A. Bassoul
Serge Matta	Scott N. Flanders	Scott G. Stephenson	Ronald M. Shaich
Shane S. Kim	Shaun E. McAlmont	Lorenzo Simonelli	Selina Y. Lo
Sheila Lirio Marcelo	Stephen McMillan	Stan Pavlovsky	Stephen S. Tang
Steven R. Beauchamp	Steven H. Collis	Steven M. Mollenkopf	Steve M. Ritchie
Steven A. Ballmer	Steven Alan Sugarman	Steven Humphreys	Steven L. Spinner
Stephen A. Odland	Stephen P. MacMillan	Sudhir Steven Singh	Stratton D. Sclavos
Strauss H. Zelnick	Subrah S. Iyar	Sundar Pichai	Margaret M. Keane
Thomas S. Smith.	Stephan B. Tanda	T. Alex Vetter	Michael Cotoia
Theodore Wahl	Charles M. Swoboda	James Brandon Black	Samuel L. Caster

CEO Name	CEO Name	CEO Name	CEO Name
Thomas A. Fanning	Tilman J. Fertitta	Timothy M. Armstrong	Timothy D. Cook
Timothy D. Hockey	Timothy T. Yates	Tom Klein	Thomas L. Ben
F. Thomson Leighton	Thomas E. Polen	Anthony Aquila	Anthony W. Thomas
Todd J. Vasos	Nallicheri Vaidyanathan Tyagarajan	Myron E. Ullman	Kevin G. Guest
Christopher J. Reading	Samuel J. Mitchell	Vikram Verma	Victoria M. Holt
Victor Lynn Lund	Victor J. Coleman	Vincent K. McMahon	Vivek Y. Ranadive
Vivek R. Shah	W. Alexander Holmes	Wallace E. Boston	Walter William Bettinger
Walter C. Rakowich	Warren E. Buffett	Wayne T. Gattinella	William B. Van Vleet
Walter P. Havenstein	William S. Kirsch	William M. Lowe	William A. Roper
William Mallory Walker	William K. Heiden	William Taylor Rhodes	Wyman T. Roberts
Yogesh K. Gupta	Yuval Wasserman		

Table 1: List of CEO Names from original dataframe