Shareholder Structure of Major Technology Companies: A Graph Analytics Study during COVID-19 and Beyond

Julio C. Esquivel*, Ixent Galpin and Oscar M. Granados

Universidad de Bogota Jorge Tadeo Lozano, Bogota, Colombia

Abstract

During financial crises or other unexpected events, investors often seek to include lower-risk assets in their portfolios. Some assets are more sensitive than others to such phenomena. In the equities markets, adjustments tend to be made to the shareholdings of companies that are associated with a higher level of uncertainty. In this work, we explore the evolution of shareholder structure of various well-known companies in the technology sector during the COVID-19 pandemic and beyond. We model, as graphs, shareholder ownership data about twenty US-listed companies between 2020 and 2022. We use freely available tools to explore the bipartite interactions and generate a wide range of topologies that facilitate the identification of how shareholding structures have evolved during the pandemic. In addition, we study the role that some nodes play in the network topology and the process of change that is observed. Our findings include that (1) most investors reduced the amount invested in technology stocks during the pandemic and that these investments tended to bounce back in the post-pandemic era; (2) Vanguard Group, Inc., is the most influential investor in the network; (3) Apple has the highest market capitalization of all technology stocks for all quarters in this study, Microsoft Corp has a significantly lower market capitalization, but a significantly higher number of investors; and (4) While investors for Apple and Microsoft tend to be from London and New York, companies such as Oracle have investors from a variety of locations.

Keywords

Corporate Networks, Shareholder Networks, Financial Structure, Network Topology

1. Introduction

The crisis in early 2020 due to the COVID-19 pandemic affected various sectors of the world economy. Financial markets absorbed government programs to counter the economic crisis in the ensuing months. The technology sector took advantage of the circumstances, leading to accelerated growth in market capitalization due to investments prior to the outbreak of the pandemic [1]. Furthermore, governments' monetary and fiscal programs throughout the world resulted in increased liquidity in financial markets. This situation, in turn, increased the correlation between the performance of various financial markets [2, 3, 4], which led to

ICAIW 2022: Workshops at the 5th International Conference on Applied Informatics 2022, October 27–29, 2022, Arequipa, Peru

*Corresponding author

julioc.esquivelr@utadeo.edu.co (J. C. Esquivel); ixent@utadeo.edu.co (I. Galpin); oscarm.granadose@utadeo.edu.co (O. M. Granados)

0000-0002-7394-4980 (J. C. Esquivel); 0000-0001-7020-6328 (I. Galpin); 0000-0002-4992-8972 (O. M. Granados)

an increase in the valuation of assets in the technology sector. Shareholder support, whether associated with a minority or majority stake, plays an important part in the success of a company [5]. Consequently, companies seek to attract and maintain shareholders to strengthen the company and achieve a collective benefit structurally. This situation is especially true in challenging times such as financial crises, given that shareholder loyalty plays an essential role in a company's resilience so that it can absorb the shockwaves of an economic crisis. Therefore, it is crucial to identify the interactions between companies and investors and understand changes in investor shareholdings.

Several works analyze the evolution of technology companies during crises such as the pandemic [6, 7, 8, 9, 10]. These works mostly focus on the increased use of technology required to address the needs imposed by the health crisis. However, the changes these companies made in their shareholder composition and how their ownership structure evolved during the pandemic have not been explored in depth. Furthermore, we observe considerable variation in Tech companies' share prices during the analysis period. This situation raises questions: what changes do shareholder networks exhibit throughout the pandemic? Were investors loyal? Were the companies successful in attracting new investors? Beyond the pandemic, the World's focus has shifted to the cost of living crisis characterized by high inflation and the outbreak of the war between Ukraine and Russia. Both situations influence financial markets from different angles. Some financial assets were mainly affected by the increase in global inflation and the monetary policies of the governments to counteract it. Others were primarily affected by the war, such as the prices of essential commodities, such as oil, gas, maize, wheat, and nickel, to name a few. The combination of events has resulted in significant volatility in the price of equities in the Tech sector (see Figure 1a), as well as other assets. Also, those situations influenced the price of technological shares like Apple (Figure 1b), Amazon (Figure 1c), Google (Figure 1d), and Microsoft (Figure 1e). First, an affection during the first wave of Covid-19. Second, during the implementation of monetary policies. Third, when the start of warming of the economy by excess financial liquidity policies, and fourth, the combination of the Covid-19 omicron variant, global inflation, and war.

As such, this work aims to analyze the shareholding structure of major technology companies. We aim to study the evolution of the structure of the shareholders of some representative technological companies during the COVID-19 pandemic and the subsequent events that ensued, specifically the high inflation and war. We employ graph analytics techniques for this analysis. We model the relationship between investors and companies as a bipartite graph that allows us to visualize the graph topology representing shareholder structure. By comparing graph topologies, we identify changes in the shareholder structure during the pandemic and beyond.

This paper is structured as follows: Section 2 describes the data set employed, using descriptive statistics, comprising the financial data of companies during the 2020–2022 period. Relevant concepts used throughout the work are also defined. In Section 3, the methods underlying the document are presented. Subsequently, the graph data model is presented, which is loaded to Neo4j[11], a graph database that allows exploration of the nodes and relationships present in the graph. In Section 4, the importance of nodes in the network is studied through centrality algorithms available in the Graph Data Science (GDS) library. Community detection algorithms are also employed to identify groups of nodes in the graph and an analysis of the topological changes of the networks during the analysis period is carried out. Section 5 concludes.

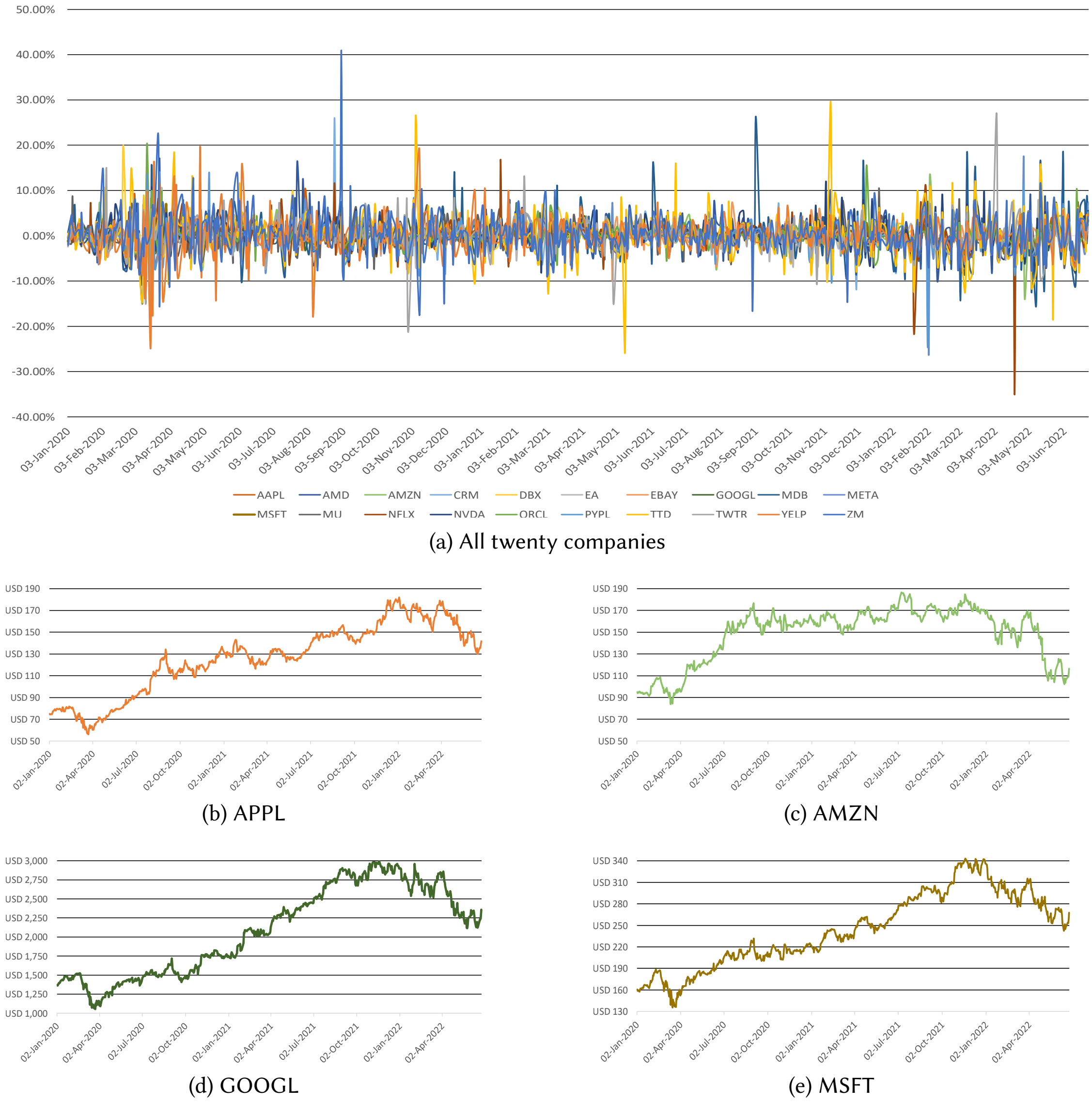


Figure 1: Evolution of shares in the Tech sector, January 2020–June 2022. **(a)** Price volatility for all twenty companies **(b)** Apple (AAPL) share price (including the latest split) **(c)** Amazon (AMZN) share price (including the latest split) **(d)** Google (GOOGL) share price **(e)** Microsoft (MSFT) share price.

2. Materials

2.1. Dataset

The data set was obtained from Refinitiv Financial Platform, a well-known provider of financial market data. Twenty companies from the technology sector listed in the United States markets were selected with their quarterly shareholding structure (Table 1). The data includes investors' participation in the companies during ten quarters beginning in January 2020 and ending in

Table 1
Technology companies selected for this study

Ticker	Company	Category	Shareholder count
AAPL	Apple Inc	Phones & Smartphones	6,710
AMD	Advanced Micro Devices Inc	Semiconductors	3,410
AMZN	Amazon.com Inc	Internet & Mail Order Department Stores	6,618
CRM	Salesforce.Com Inc	Cloud Computing Services	4,366
DBX	Dropbox Inc	Application Software	1,396
EA	Electronic Arts Inc	Application Software	2,764
EBAY	eBay Inc	E-commerce & Auction Services	2,724
META	Meta Platforms Inc	Social Media & Networking	6,016
GOOGL	Alphabet Inc	Search Engines	6,119
MDB	MongoDB Inc	Software & IT Services	1,332
MSFT	Microsoft Corp	Software	7,065
MU	Micron Technology Inc	Semiconductors	3,284
NFLX	Netflix Inc	Online Services	4.216
NVDA	NVIDIA Corp	Semiconductors	5,097
ORCL	Oracle Corp	Enterprise Software	4,150
PYPL	PayPal Holdings Inc	Internet Security & Transactions Services	4,908
TTD	Trade Desk Inc	Software	1,818
TWTR	Twitter Inc	Social Media & Networking	2,579
YELP	Yelp Inc	Online Services	809
ZM	Zoom Video Communications Inc	Software	2,161

June 2022. In addition to the bipartite relationships between investors and companies, the data set integrates the geographic location (city and country of origin) and the type and subtype of investor in which they are classified by Refinitiv. Furthermore, the categories of technology companies such as e-commerce, Semiconductors, Cloud, Software, and IT Services, among others, are shown.

2.2. Data Processing

Data pre-processing is carried out using Python with data manipulation libraries such as Pandas and NumPy. The pre-processing step consists of identifying and correcting or replacing faulty records in the data set. In this way, 77,542 refined records are obtained corresponding to the participation in the twenty companies by the 8,730 distinct investors in the data set (see Table 2). Subsequently, the data in tabular format for all twenty companies are imported into Neo4j using the Cypher language, which provides functionality to convert the data to the graph data model described in Section 3.

2.3. Graph Databases

A graph comprises a set of vertices or nodes, denoting entities, connected by edges, which typically denote relationships between entities [12, 13]. A graph database is a type of NoSQL database [14] that represents data using a graph rather than tables, as traditionally relational databases do. This way, neighboring vertices can be efficiently traversed by following the edges in the graph. The graph model enables information from diverse domains to be represented,

Table 2
Sample of the data set for Apple Inc

Investor	Type	SubType	City	Country	Shareholding_Avg
The Vanguard Group, Inc.	Investment Managers	Investment Advisor/Hedge Fund	Malvern	United States	1,277.06
Berkshire Hathaway Inc.	Investment Managers	Insurance Company	Omaha	United States	912.31
BlackRock Institutional Trust Company, N.A.	Investment Managers	Investment Advisor	San Francisco	United States	701.29
State Street Global Advisors (US)	Investment Managers	Investment Advisor/Hedge Fund	Boston	United States	647.05
Fidelity Management & Research Company LLC	Investment Managers	Investment Advisor	Boston	United States	343.72
Geode Capital Management, L.L.C.	Investment Managers	Investment Advisor/Hedge Fund	Boston	United States	257.66
T. Rowe Price Associates, Inc.	Investment Managers	Investment Advisor	Baltimore	United States	216.83
Norges Bank Investment Management (NBIM)	Investment Managers	Sovereign Wealth Fund	Oslo	Norway	163.84
Northern Trust Investments, Inc.	Investment Managers	Investment Advisor/Hedge Fund	Chicago	United States	127.90
Legal & General Investment Management Ltd.	Investment Managers	Investment Advisor/Hedge Fund	London	United Kingdom	106.40
⋮	⋮	⋮	⋮	⋮	⋮

from biomedical data to financial fraud detection [15]. The graph data model is deemed well-suited for domains that require the representation of complex relationships between entities, as relationships are first-class citizens in a graph Database Management System (DBMS). This is in contrast to the well-established relational model, in which relationships between entities are implemented using foreign keys, and often computationally expensive joins are required to link data from different entities [16].

Neo4j[1] is a well-known example of a DBMS that implements the graph data model. It is implemented in Java and developed by Neo Technology. It purportedly supports ACID transactions, does not enforce a schema, and offers high availability[16]. Neo4j stores the information by creating a directed graph between the vertices and the connections between them [17]. An undirected graph can also be supported by ignoring the direction of an edge. Each edge has exactly one type, and nodes may have zero, one, or more labels. The declarative Cypher language is used to query or manipulate data in Neo4j. This is inspired by SQL and SPARQL [18]. The Cypher query language makes use of pattern matching for the selection of data from the graph. That is, it allows users to perform queries by specifying sub-graphs to search for within the overall graph[17]. The syntax used to represent the graph's topological features can be represented solely using ASCII symbols.

3. Graph Data Model

3.1. Methods

The relationship between investors and companies shows a bipartite graph structure. A graph G is bipartite if the vertex set $V(G)$ can be partitioned into two disjoint sets U and V such that every link connects a node in U to one in V and two vertices from the same set are not adjacent [19]. That is, U and V are independent sets. In other words, a bipartite graph is a triple $G = (U, V, E)$ where U and V are two disjoint sets of vertices(the top and bottom vertices) and $E \subseteq UV$ is the set of edges [20]. The bipartite network studies usually depend on the one-mode projection of the original network, i.e., given a bipartite network $G(U, V, E)$, where U and V as independent sets, the projection converts $G(U, V, E)$ into GU and GV [21].

Although this projection method is simple and intuitive, it may generate a loss of information, especially when the bipartite network $G(U, V, E)$ is a large graph, and the same situation occurs

[1] https://neo4j.com/

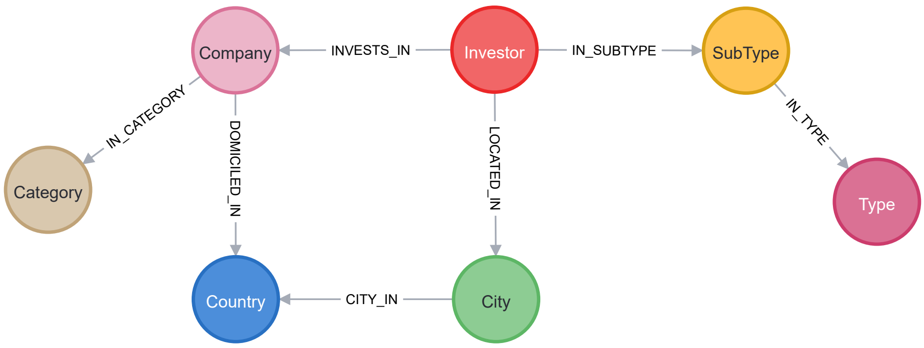


Figure 2: Neo4j data model

with the projection result. Additionally, if we project the original graph into the network of all the investors, the meaningful information may be overwhelmed by the high link density [22]. Another situation is that the weighting of edges is a critical problem in constructing a bipartite network projection. To solve this situation, scholars implemented several methods to use the weighted edge [23, 24], a relevant factor in our dataset. However, the goal of this paper is about the shareholders' dynamics in each company. Thus, we included other data to build the communities, e.g., geographical data like city and country.

3.2. Data Model

The graph data model we designed for the data set in this study is presented in Figure 2. The graph contains nodes labeled Company, which represent the companies in the data set. These are connected by the edge type IN_CATEGORY to the nodes with the Category label, which enables a company to be classified by category. The Company entity is connected to the Country nodes through the DOMICILED_IN relation to determining the Country of origin of the Company. The nodes labeled Investor contain data about the different investors. The information about investors investing in companies is represented through the INVESTS_IN edge type. As a property of this relationship, each shareholder's investment amount in millions of dollars in the quarters of 2020 to 2022 is stored. The average investment value of the ten quarters for each shareholder is also a property of this relationship. Regarding the data relating to the geographic information about investors, the nodes labeled City are connected by the edge LOCATED_IN. The relationship CITY_IN is used to link cities to countries. There is also the classification of the investors given by the nodes labels SubType and Type.

3.3. Graph Exploration

As mentioned previously, Cypher, a declarative query language, is used to query Neo4j databases. The Neo4j Python driver may also be used to pose queries and integrate result sets from within Python programs. The following Python code with embedded Cypher calculates the number of nodes for each label in the data set:

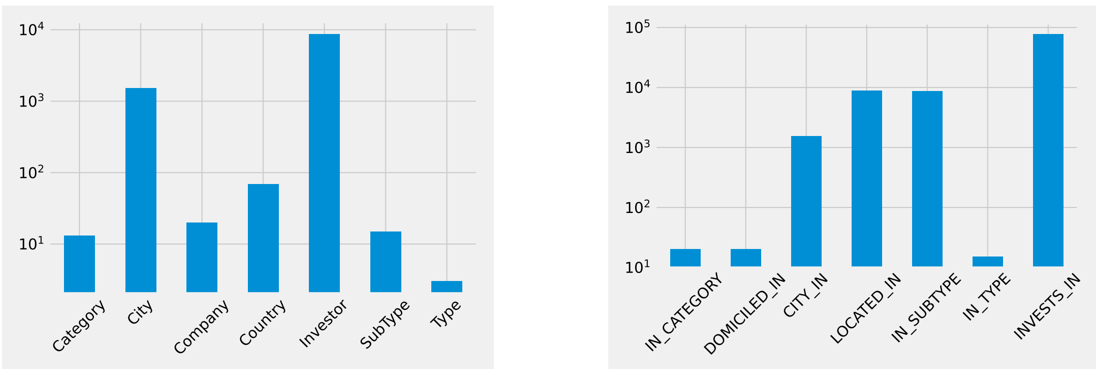
```
result = {"Labels": [], "Count": []}
with connection.session() as consult:
    labels = [row["label"] for row in consult.run("CALL db.labels()")]
    for label in labels:
        query = f"MATCH (:`{label}`) RETURN count(*) as count"
        count = consult.run(query).single()["count"]
        result["Labels"].append(label)
        result["Count"].append(count)

df = pd.DataFrame(data=result)
print(tabulate(df.sort_values("Count"), headers='keys',
               tablefmt='psql', showindex=False))
```

Running this code reveals that the number of nodes per label type is as follows: 3 types, 13 categories, 15 subtypes, 20 companies, 69 countries, 1,525 cities, and 8,730 investors. In an analogous way, the cardinalities of the relations are obtained, as follows: IN_TYPE 15, IN_CATEGORY 20, DOMICILED_IN 20, CITY_IN 1,525, IN_SUBTYPE 8,733, LOCATED_IN 8,789, INVESTS_IN 77,542. Figure 3 shows the cardinalities of the nodes and relationships of the graph.

Based on the previous results, it is apparent that the graph contains a substantial number of investor nodes compared to other types of nodes. We subsequently pose a query to ascertain which companies in the technology sector have the largest number of them:

```
MATCH (inves:Investor)-[:INVESTS_IN]->(comp:Company)
RETURN comp.name AS Company, count(inves) AS Count
ORDER BY Count DESC
```

The results show that the five largest technology companies (GAFAM or Big Five) have the largest number of shareholders. Microsoft (MSFT) has the largest number of investors with 7,065, (Table 1) followed by Apple (AAPL) with 6,710, Amazon (AMZN) with 6,618, Alphabet (GOOGL) with 6,119 and Meta (META- formerly Facebook) with 6,016. The others have between



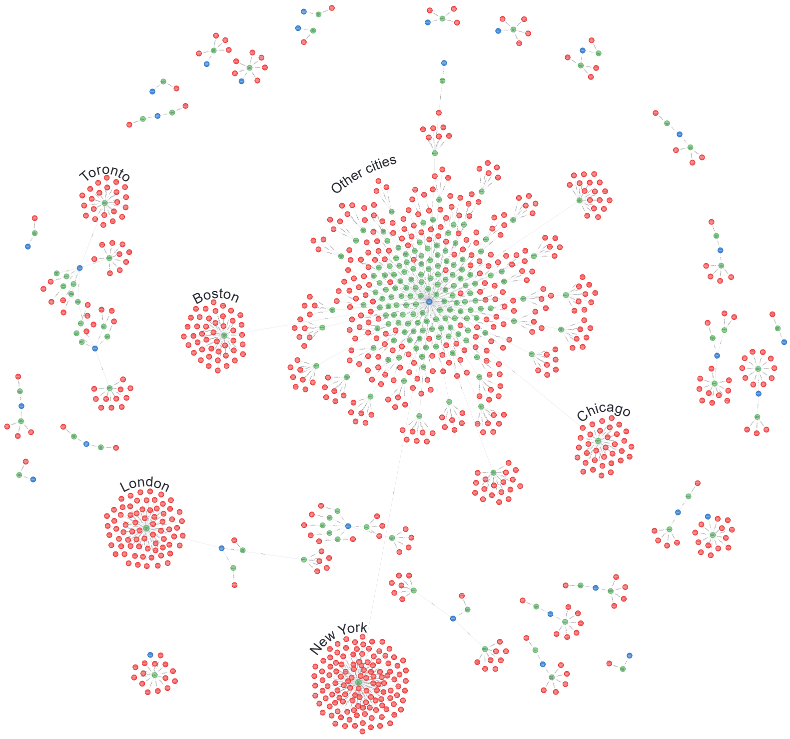
(a) Number of nodes per label

(b) Number of relationships per type

Figure 3: Node and relationship cardinalities

Figure 4: Investor Network for Meta (META)

```
MATCH (i: Investor) -[:INVESTS_IN]->(c:Company), (i)-[:LOCATED_IN]->(ci:City),
(ci)-[:CITY_IN]->(co:Country)
WHERE c.name = "Meta Platforms Inc" AND i.Shareholding_avg > 0
RETURN i, ci, co
```



809 and 5,097 shareholders (see Table 1).

Finally, the Meta network (META) is explored through a query, which shows the relationship between investors, cities, and countries, keeping only investors with a shareholding greater than zero. Figure 4 presents the Cypher query and respective visualization for Investor, City and Country nodes using red, green and blue nodes respectively. We have manually labeled some of the clusters by the city for greater clarity.

This exercise makes it possible to identify the different topologies and their variations during the analysis period for the selected companies. Some topological metrics and results are presented in the next section.

4. Results

Some of the graph analytics algorithms available in the Neo4j Graph Data Science (GDS) library[2] are used to explore the networks of shareholders for each of the companies. The centrality and

[2]https://neo4j.com/product/graph-data-science/

community detection algorithms are executed on a projection of the graph data model, which can be understood as a materialization of a view (or sub-graph) of the overall graph.

4.1. Graph projection

For the graph model described in Section 3, the shareholder participation is stored as a property of the relationship INVESTS_IN, so a relationship projection is performed prior to the application of the GDS library algorithms. The nodes labeled Investor and Company and the relationship type INVESTS_IN are projected, resulting in the projections of a bipartite graph. Several projections are defined using as property the participation of the shareholders in different periods; the start of the pandemic (Q1 2020), the second peak of infections (Q2 2021), the start of the Omicron variant (Q4 2021), the start of the Russo-Ukrainian war and inflation (Q2 2022), as well as for the average of all periods in the study. The following query shows the adjustment of the projection for the average participation for each Investor in the different periods:

```
CALL gds.graph.project(
    'InvestorAndCompany',
    ['Investor', 'Company'],
    {
        INVESTS_IN:{ properties: 'Shareholding_avg',
                     orientation: 'NATURAL'
                   }
    }
)
YIELD
  graphName AS graph,
  relationshipProjection AS INVESTS_INProjection,
  nodeCount AS nodes,
  relationshipCount AS rels
```

The projections' results for the different adjusted periods, and the average participation for all periods of this study, are presented in Table 3.

Table 3
Projection of Investor to Company, including the shareholding over time.

		Shareholding (USD Million)				
Investor	**Company**	**Q1 2020**	**Q2 2021**	**Q4 2021**	**Q2 2022**	**Avg (all quarters)**
The Vanguard Group, Inc.	Apple Inc	1,346.91	1,264.94	1,261.26	1,270.00	1,277.06
Ellison (Lawrence Joseph)	Oracle Corp	1,138.73	1,138.73	1,138.73	1,145.73	1,139.43
Berkshire Hathaway Inc.	Apple Inc	980.62	887.14	887.14	890.92	912.31
BlackRock Institutional Trust Company, N.A.	Apple Inc	749.42	671.03	675.69	676.87	701.29
State Street Global Advisors (US)	Apple Inc	722.24	622.58	633.12	613.85	647.05
The Vanguard Group, Inc.	Microsoft Corp	640.17	610.97	615.95	621.60	620.37
Fidelity Management & Research Company LLC	Apple Inc	359.06	338.32	338.59	338.49	343.72
BlackRock Institutional Trust Company, N.A.	Microsoft Corp	347.00	325.59	333.86	335.63	338.50
State Street Global Advisors (US)	Microsoft Corp	314.77	294.82	302.54	300.10	302.20
Geode Capital Management, L.L.C.	Apple Inc	256.71	254.16	264.35	272.08	257.66
⋮	⋮	⋮	⋮	⋮	⋮	⋮

In Neo4j, although graphs are always directed, it is possible to traverse the connections of a graph according to the direction stipulated in the data model (which is called *natural orientation*, in this case, it would be from the nodes Investor to the Company nodes), or in the opposite direction to that stipulated in the data model (which is called *reverse orientation*, in this case, it would be from the Company nodes to Investor nodes).

4.2. Degree Centrality

Degree centrality is a metric that reflects the importance of the nodes. In turn, it allows the number of incoming and outgoing relations to be measured, taking into account the direction of the relationship in the graph projection. The higher the degree of a node, the higher its degree of centrality, which implies that the entity represented by the node has greater importance [25]. In the following subsections, the degree of centrality algorithm is used to understand the most important nodes within the networks.

4.2.1. Investor Degree Centrality

We compute the unweighted and weighted degree of centrality for nodes labeled Investor in the graph projection InvestorAndCompany. The unweighted degree of centrality counts the number of node connections without taking into account the weights. The query is shown below:

```
CALL gds.degree.stream(
  'InvestorAndCompany'
  )
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Name, score AS degree_centrality
ORDER BY degree_centrality DESC
```

The results of Table 4 show the degree of unweighted centrality of the nodes labeled Investor, which essentially indicates the number of companies that an investor invests in. There are shareholders with participation from one to the total of the twenty companies in the study.

Table 4
Unweighted Investor degree centrality

Name	**Degree centrality (unweighted)**
The Vanguard Group, Inc.	20
Ellison (Lawrence Joseph)	1
Berkshire Hathaway Inc.	3
Capital World Investors	18
Capital Research Global Investors	19
Capital International Investors	17
PRIMECAP Management Company	15
Sanders Capital, LLC	5
Comprehensive Financial Management LLC	4
Edgewood Management LLC	8
⋮	⋮

The following query shows the distribution of the degrees of centrality obtained:

```
MATCH (i:Investor)
RETURN count(i.degree_centrality) AS Count,
avg(i.degree_centrality) AS Ave,
(percentileDisc(i.degree_centrality, 0.25)) AS `25%`,
(percentileDisc(i.degree_centrality, 0.50)) AS `50%`,
(percentileDisc(i.degree_centrality, 0.75)) AS `75%`,
(percentileDisc(i.degree_centrality, 0.90)) AS `90%`,
(percentileDisc(i.degree_centrality, 0.95)) AS `95%`,
(percentileDisc(i.degree_centrality, 0.99)) AS `99%`,
(percentileDisc(i.degree_centrality, 0.999)) AS `99.9%`,
(percentileDisc(i.degree_centrality, 1)) AS `100%`
```

Count	Ave	25%	50%	75%	90%	95%	99%	99.9%	100%
8,730	8.882	4	8	13	18	19	20	20	20

Thus, 50% of the investors only have connections with up to eight different companies. Subsequently, to facilitate comparison, the degree of centrality is normalized. Thus, a weighted projection is used, where the shareholding properties of the relation are considered. In this way, the algorithms of the GDS library calculate the sum of the weights of the relation to determine the degree of centrality of the nodes. The centrality indicator is calculated in each period mentioned in Section 4.1, using the different adjusted projections. In this way, the query is shown to determine the centrality metric for the average participation for each investor in the different periods:

```
CALL gds.degree.stream
( 'InvestorAndCompany',
  {
    relationshipWeightProperty: 'Shareholding_avg'
  }
  )
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Name, score AS
    degree_centrality_weighted
ORDER BY degree_centrality_weighted DESC
```

The results obtained (see Table 5) reflect the weighted degree of centrality of the nodes labeled Investor determined by the properties of the relationship INVESTS_IN. It is observed that the investor The Vanguard Group, Inc. is the node with the highest degree of centrality, both in each period and in the average of the investments for all periods. This is followed by the investor BlackRock Institutional Trust Company, N.A., the second node with the highest degree of centrality. As such, it may be interpreted that The Vanguard Group, Inc. is the most influential investor in the network.

Subsequently, the distribution of the weighted degree centrality values obtained for the average investments indicates that 90% of investors made investments of up to 1.66 million dollars between 2020 and 2022.

Table 5
Weighted Investor degree centrality

Name	Degree centrality (weighted)				
	Q1 2020	Q2 2021	Q4 2021	Q2 2022	Avg (all quarters)
The Vanguard Group, Inc.	3,241.68	3,087.13	3,080.65	3,787.76	3,258.07
BlackRock Institutional Trust Company, N.A.	1,801.86	1,670.75	1,692.31	2,066.42	1,806.21
State Street Global Advisors (US)	1,642.99	1,504.00	1,530.82	1,832.90	1,611.08
Fidelity Management & Research Company LLC	1,257.26	1,149.03	1,099.09	1,325.83	1,211.02
Ellison (Lawrence Joseph)	1,138.73	1,138.73	1,138.73	1,145.73	1,139.43
Berkshire Hathaway Inc.	981.16	887.67	887.67	901.59	914.87
T. Rowe Price Associates, Inc.	768.99	705.37	741.81	1049.21	808.87
Geode Capital Management, L.L.C.	595.45	617.72	642.82	806.67	647.81
Norges Bank Investment Management (NBIM)	396.58	398.65	398.65	427.53	404.86
Northern Trust Investments, Inc.	314.22	288.83	296.17	349.58	313.76
⋮	⋮	⋮	⋮	⋮	⋮

4.2.2. Company Degree Centrality

Using a graph projection similar to the one defined in Section 4.1, but with *reverse orientation* instead of *natural orientation*, the unweighted and weighted degree centrality is computed for nodes labeled Company. For this last metric, the periods defined in Section 4.1 are studied:

The degree centrality for the nodes labeled Company (Table 6) shows that Microsoft (MSFT) has the highest unweighted degree centrality, that is, it has the highest number of connections with the investor nodes. However, the weighted degree centrality suggests that Apple (AAPL) is the most important node, that is, it is the company that received the most investments.

Table 6
Company Degree Centrality

Name	Degree centrality (unweighted)	Degree centrality (weighted)				
		Q1 2020	Q2 2021	Q4 2021	Q2 2022	Avg (all quarters)
Apple Inc	6,710	10,753.60	9,771.43	9,743.35	9,682.80	10,002.00
Microsoft Corp	7,065	5,661.58	5,419.21	5,408.00	5,390.45	5,454.89
Oracle Corp	4,150	2,730.44	2,440.21	2,308.70	2,303.48	2,479.28
Meta Platforms Inc	6,016	1,942.35	1,940.62	1,840.34	1,774.26	1,897.03
NVIDIA Corp	5,097	1,869.81	1,774.47	1,770.17	1,747.79	1,796.61
Amazon.com Inc	6,618	374.74	373.19	357.95	7,187.55	1,734.96
PayPal Holdings Inc	4,908	1,014.53	972.71	917.25	873.14	960.63
Advanced Micro Devices Inc	3,410	904.39	851.29	863.57	1,115.37	934.12
Micron Technology Inc	3,284	940.73	925.47	921.66	926.71	931.78
Salesforce.Com Inc	4,366	792.80	782.13	823.09	827.42	801.63
Twitter Inc	2,579	615.14	667.83	649.66	651.91	647.46
eBay Inc	2,724	718.32	641.85	571.39	522.38	628.00
Netflix Inc	4,216	377.92	370.41	373.89	364.87	371.42
Trade Desk Inc	1,818	339.13	296.48	317.25	337.15	321.71
Electronic Arts Inc	2,764	272.72	270.73	265.16	261.72	269.29
Dropbox Inc	1,396	218.64	294.39	257.85	250.78	266.70
Alphabet Inc	6,119	244.97	243.96	264.40	241.76	247.31
Zoom Video Communications Inc	2,161	83.29	176.44	183.09	186.03	151.61
Yelp Inc	809	72.52	72.16	67.73	68.01	71.21
MongoDB Inc	1,332	50.64	63.14	65.00	64.50	59.82

The distribution obtained for average investments indicates that 50% of the companies have connections with a maximum of 3,410 investors and, in turn, receive investments of up to 647.5 million dollars.

	Ave	25%	50%	75%	90%	95%	99%	99.9%	100%
Unweighted	3,877.10	2,161	3,410	5,097	6,618	6,710	7,065	7,065	7,065
Weighted	1,501.37	266.70	647.46	1,734.96	2,479.28	5,454.88	10,001.99	10,001.99	10,001.99

4.3. Community Analysis

There are many large-scale complex networks in the real-world whose structure is not fully understood by some methods [26]. By including the city and the country as other nodes in the shareholders' network, we identify new patterns that bipartite networks could not reveal because of much-hidden information about shareholders' networks that are not easy to detect by simple observation. Although a bipartite network has different methods to identify communities, we decide to use the richness of the data set because we can detect overlapping communities and isolated communities

Communities are groups that are densely connected among their members and sparsely connected with the rest of the network [27, 28]. Habitually, community detection allows nodes in the graph to be grouped into clusters, in such a way that nodes in the same cluster are more closely related than nodes in other clusters [29]. We employ the Louvain method, which allows community detection in large networks. Louvain's modularity algorithm detects clusters by evaluating the density of node connections within a cluster, compared to how connected they would be in an average or random sample [30]. This measure of community allocation is known as *modularity*. Modularity takes values on a scale between -0.5 and 1, where -0.5 indicates a non-modular grouping, and 1 is a totally modular grouping. The algorithm optimizes modularity locally on all nodes present in small communities and subsequently groups them into larger communities.

As with the degree centrality computation described in Section 4.2, we use the InvestorAnd-Company graph projection as the starting point for the Louvain community detection. Unweighted communities are determined using the number of node connections and ignoring edge weights, using the Cypher query presented in Figure 5a. The result defines two communities with 4,450 and 4,300 members respectively, and modularity of 0.104 for the entire graph, with the modularity of the communities in the range [-0.032, 0.104]. with communities in the range [-0.0044, 0.130], expressed by the weights of the average shareholding in the projection of the graph. 96.83% of members are grouped into five communities with populations of 5,793, 1,088, 536, 530, and 526 respectively.

4.4. Network Topology

We have integrated metrics for two of the three companies with the highest degree of centrality, *viz.,* Apple (AAPL) and Oracle (ORCL) in the network visualizations. For each company, we focus on four quarters: $Q1$ 2020, $Q2$ 2021, $Q4$ 2021, and $Q2$ 2022. For each case, we show the number of Investor nodes (N_{IN}), the number of nodes in City (N_{CI}), and the number of nodes

(a) Unweighted

```
CALL gds.louvain.write(
    'InvestorAndCompany',
        {
            writeProperty: '
    louvain_community'
        }
)
YIELD communityCount, modularity,
    modularities
```

(b) Weighted

```
CALL gds.louvain.write(
    'InvestorAndCompany',
        {
            writeProperty: '
    louvain_community_weighted',

    relationshipWeightProperty: '
    Shareholding_avg'
        }
)
YIELD communityCount, modularity,
    modularities
```

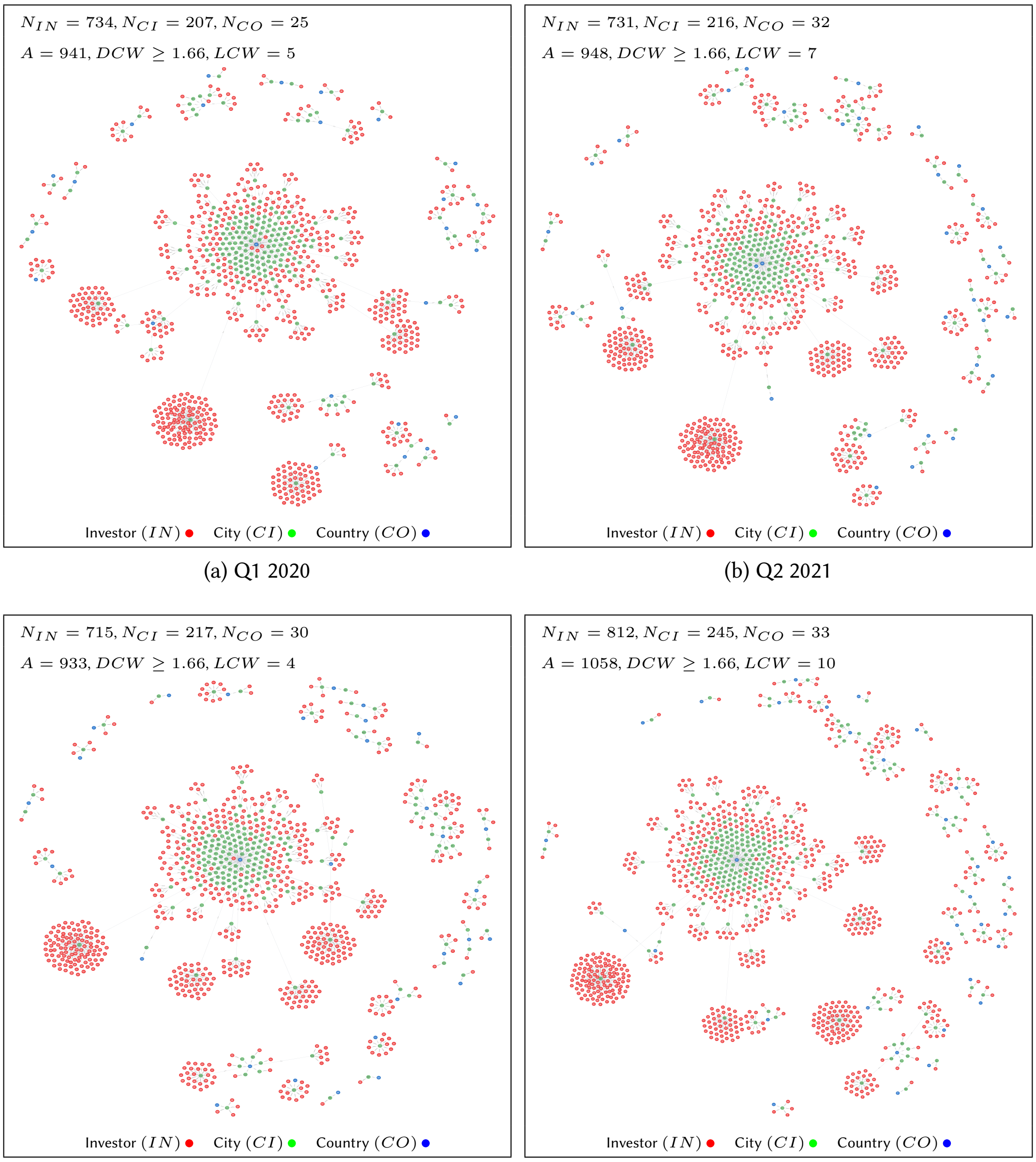
Figure 5: Louvain Community Detection using Neo4j GDS library

in Country (N_{CO}). The number of edges (A), the weighted degree of centrality (DCW), and the number of weighted Louvain communities (LCW). In both cases, we used the investors with a weighted degree of centrality indicating greater investment than 1.66 million dollars.

The Apple network we analyzed contains, for Q1 2020, 734 investors, 207 cities in 25 countries, and 941 edges. By Q2 2021, 731 investors, 216 cities in 32 countries, and 948 edges. By Q4 2021, 715 investors, 217 cities in 30 countries, and 933. While by Q2 2022, there were 812 investors, 245 cities in 33 countries, and 1,058 edges. The results show that topology and the weighted Louvain communities evolved similarly to the stock prices. Until November 2021, the financial markets recovered the impact of the first Covid-19 wave when the financial asset prices were down almost 25% in March 2020. After that, the global inflation fear started to affect the asset prices, and for December 2021, the NASDAQ index had down from its highest level of 16,057 points on November 15 to 15,644 on December 27. Henceforth, technology stock prices continued down, and the NASDAQ index by June 13, 2022, was at 10,798 points, a fall of 30%. However, June had shown one of the best valuation performances of 2022 that would last until August 8, when they began to fall again. As shown in Figure 6, we detected more communities when the prices increase, i.e., the investment appetite changes the community levels. However, some long-term investors did not change their position considerably.

Across each quarter, two crucial groups of investors are apparent: the first community comprises those originating in New York City, and the second, those originating in London. The topology evolution for Microsoft (MSFT) is very similar to Apple's case. There are two communities of investors, one originating in New York and the other in London. The number of nodes and their relationships is lower in the fourth quarter of 2021 compared to the first quarter of 2020.

Figure 7 presents the changes in the topology of the Oracle (ORCL) network for each period in the study. Unlike Apple (Figure 6) and Microsoft, the number of nodes and relationships for this company increased in $Q2$ 2021 compared to the $Q1$ 2020, then decreased, and in the final quarter increased again. Although New York City continues to dominate as a place of origin



(a) Q1 2020

(b) Q2 2021

(c) Q4 2021

(d) Q2 2022

Figure 6: Network Metrics and Topology for Apple (AAPL)

for investors, it is more diversified for Oracle, including communities from Boston, Chicago, London, and Toronto as well (see Figure 7d). A relevant aspect in both cases is that Apple and Oracle had a similar percentage growth in shareholders. However, the community growth for Apple was 150%, while Oracle was 50%. Namely, the investment appetite was more diversified for the first than the second.

(a) Q1 2020 (b) Q2 2021

(c) Q4 2021 (d) Q2 2022

Figure 7: Network Metrics and Topology for Oracle (ORCL)

5. Conclusions

Due to the global health crisis experienced by the COVID-19 pandemic and its variants, as well as different unexpected events such as the Russian-Ukrainian war, various sectors of the world economy were affected, so investors sought to keep their capital in funds of lower-risk investment, generating strategic movements in the structures of the different companies in the market.

In this paper, we focus on particularly examining the evolution of the stock and share data of the largest companies in the technology sector in the United States during the COVID-19 pandemic and beyond; employing a study of some graph analytics techniques. Graphic representations of the bipartite interaction between shareholders and companies are presented, which allow us to determine the evolution of the shareholder structure in different quarters between 2020 and 2022, reflecting the influence of the changes produced by the COVID-19 pandemic and by subsequent events, given that investment decreases each quarter, as shown by the degree of centrality of the figures obtained.

The unweighted investor degree centrality tells us about the number of companies invested per investor. We observe that overall, most investors reduced the amount invested in technology stocks during the pandemic (Q1 2020 to Q4 2021) and that these investments bounced back in the post-pandemic era (Q2 2022). It is also the case that The Vanguard Group, Inc., is the most influential investor in the network, as it invests in all 20 companies and also has the highest shareholding across all companies. On the other hand, the company degree centrality figures show that, while Apple has the highest market capitalization of all technology stocks for all quarters in this study, Microsoft Corp has a significantly lower market capitalization but a substantially higher number of investors.

The Louvain method yields two communities with 4,450 and 4,300 members, respectively, and a modularity of 0.104 for the unweighted graph. For the weighted graph, we find that 96.83% of members are grouped into five communities. The network topology analyses for both Apple and Oracle reflect the reduction in technology prices during the pandemic and the post-pandemic rebound. We also observe two noteworthy communities of investors in London and New York for Apple. In contrast, for Oracle, we keep that the location of investors is more diverse.

References

[1] P. Ventrici, D. Krepki, H. Palermo, Sector software y la situación respecto de la pandemia de covid-19, El trabajo en los tiempos del COVID 19 (2020).

[2] K. F. Chan, Z. Chen, Y. Wen, T. Xu, Covid-19 vaccines and global stock markets, Finance Research Letters (2022) 102774. doi:https://doi.org/10.1016/j.frl.2022.102774.

[3] Y. Liu, Y. Wei, Q. Wang, Y. Liu, International stock market risk contagion during the covid-19 pandemic, Finance Research Letters 45 (2022) 102145. doi:https://doi.org/10.1016/j.frl.2021.102145.

[4] Q. Zeng, X. Lu, T. Li, L. Wu, Jumps and stock market variance during the covid-19 pandemic: Evidence from international stock markets, Finance Research Letters 48 (2022) 102896. doi:https://doi.org/10.1016/j.frl.2022.102896.

[5] R. S. Thomas, J. F. Cotter, Shareholder proposals in the new millennium: Shareholder support, board response, and market reaction, Journal of Corporate Finance 13 (2007) 368–391. doi:https://doi.org/10.1016/j.jcorpfin.2007.02.002.

[6] F. Santiago, C. de Fuentes, J. Peerally, J. Larsen, Investing in innovative and productive capabilities for resilient economies in a post-covid-19 world, International Journal of

Technological Learning, Innovation and Development 12 (2020) 153–167. doi:10.1504/IJTLID.2020.110623.
[7] E. Popkova, A. Bogoviz, S. Lobova, A. Chililov, A. Sozinova, B. Sergi, Changing entrepreneurial attitudes for mitigating the global pandemic's social drama, Humanities and Social Sciences Communications 9 (2022). doi:10.1057/s41599-022-01151-2.
[8] H. Nam, S. Kim, T. Nam, Identifying the directions of technology-driven government innovation, Information 13 (2022). URL: https://www.mdpi.com/2078-2489/13/5/208. doi:10.3390/info13050208.
[9] K. M. Faqih, Internet shopping in the covid-19 era: Investigating the role of perceived risk, anxiety, gender, culture, and trust in the consumers' purchasing behavior from a developing country context, Technology in Society 70 (2022) 101992. doi:https://doi.org/10.1016/j.techsoc.2022.101992.
[10] C. Watanabe, W. Akhtar, Y. Tou, P. Neittaanmäki, A new perspective of innovation toward a non-contact society - amazon's initiative in pioneering growing seamless switching, Technology in Society 69 (2022) 101953. doi:https://doi.org/10.1016/j.techsoc.2022.101953.
[11] M. Needham, A. E. Hodler, Graph Algorithms: Practical Examples in Apache Spark and Neo4j, O'Reilly Media, 2019.
[12] G. Combariza, Una introducción a la teoría de grafos, XIV Encuentro de Geometría y II encuentro de Aritmética (2003) 565–591.
[13] A.-L. Barabasi, Network Science, Cambridge University Pr., 2016. URL: https://www.ebook.de/de/product/24312547/albert_laszlo_barabasi_network_science.html.
[14] A. Davoudian, L. Chen, M. Liu, A survey on nosql stores, ACM Computing Surveys (CSUR) 51 (2018) 1–43.
[15] R. Kumar Kaliyar, Graph databases: A survey, in: International Conference on Computing, Communication & Automation, IEEE, 2015, pp. 785–790.
[16] E. Lozano, Neo4J: trabajando con grafos, 2021. URL: https://www.paradigmadigital.com/dev/neo4j-trabajando-grafos/.
[17] A. Vukotic, N. Watt, T. Abedrabbo, D. Fox, J. Partner, Neo4j in Action, Manning, 2014. URL: https://books.google.nl/books?id=61GdmgEACAAJ.
[18] J. Pérez, M. Arenas, C. Gutierrez, Semantics and complexity of sparql, ACM Transactions on Database Systems (TODS) 34 (2009) 1–45.
[19] A. Asratian, T. Denley, R. Haggkvist, Bipartite Graphs and their Applications, " Cambridge University Press", 1998.
[20] J.-L. Guillaume, M. Latapy, Bipartite graphs as models of complex networks, Physica A: Statistical Mechanics and its Applications 371 (2006) 795–813. URL: https://doi.org/10.1016/j.physa.2006.04.047.
[21] T. Zhou, J. Ren, M. c. v. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E 76 (2007) 046115. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.046115. doi:10.1103/PhysRevE.76.046115.
[22] N. Du, B. Wang, B. Wu, Y. Wang, Overlapping community detection in bipartite networks, in: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 1, 2008, pp. 176–179. doi:10.1109/WIIAT.2008.98.
[23] Y. Fan, M. Li, P. Zhang, J. Wu, Z. Di, The effect of weight on community structure

of networks, Physica A: Statistical Mechanics and its Applications 378 (2007) 583–590. doi:https://doi.org/10.1016/j.physa.2006.12.021.

[24] M. E. J. Newman, Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (2001) 016132. URL: https://link.aps.org/doi/10.1103/PhysRevE.64.016132. doi:10.1103/PhysRevE.64.016132.

[25] H. Tong, Z. Jia, M. Zhang, J. Qi, Analysis of stock-shareholder associated network based on complex network, Journal of Mathematical Finance 11 (2021) 107–122. doi:10.4236/jmf.2021.111005.

[26] W. Liu, M. Pellegrini, X. Wang, Detecting communities based on network topology, Scientific Reports 4 (2014) 5739. URL: https://doi.org/10.1038/srep05739. doi:10.1038/srep05739.

[27] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113. URL: https://link.aps.org/doi/10.1103/PhysRevE.69.026113. doi:10.1103/PhysRevE.69.026113.

[28] M. E. J. Newman, Modularity and community structure in networks, Proceedings of the National Academy of Sciences 103 (2006) 8577–8582. URL: https://www.pnas.org/content/103/23/8577. doi:10.1073/pnas.0601602103. arXiv:https://www.pnas.org/content/103/23/8577.full.pdf.

[29] S. Fortunato, Community detection in graphs, Physics reports 486 (2010) 75–174.

[30] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (2008) P10008. URL: https://doi.org/10.1088/1742-5468/2008/10/p10008. doi:10.1088/1742-5468/2008/10/p10008.