

The logo of KU Leuven, featuring the text "KU LEUVEN" in white capital letters on a dark blue rectangular background.

KU LEUVEN

Modern Data Analytics

Uncovering the Network

Outline

	Part A	Part B	Part C
Three Parts	Introduction	Data Processing	Exploration and Analysis
Details	<ul style="list-style-type: none"> • Background for data selection • Research Questions 	<ul style="list-style-type: none"> • Institutions selection and basis • Data Acquisition • Data cleaning 	<ul style="list-style-type: none"> • How are investment firms connected? • Which companies are co-invested by these big players? • Which companies have the largest number of shares or market values? • Has the connectivity increased? • Is there a change in the network metrics?
Outcome	Get basic Insights & determine the research direction	Obtain the representative and available data	Establish the research method and get the results

Part A - Introduction

Background of data selection

The Securities and Exchange Commission's (SEC) Form 13F is a quarterly report that is required to be filed by all institutional investment managers with at least \$100 million in assets under management in American.

Congress created the 13F requirement in 1975. Its intention was to provide the U.S. public a view of the holdings of the nation's largest institutional investors. Lawmakers believed this would increase investor confidence in the integrity of the nation's financial markets. Firms that are considered institutional investment managers include mutual funds, hedge funds, trust companies, pension funds, insurance companies, and registered investment advisors.

Therefore we selected five top investment institutions, including investment banks and quant funds, for the latest two quarters of analysis and comparison.

Part A - Introduction

Research Questions

1. How are investment firms connected?
2. Which companies are co-invested by these big players? How many companies are co-invested by some of big players?
3. Which investment firms have the largest number of shares or market value of a certain company?
4. Has the connectivity increased?
5. Is there a change in the network metrics?

Part B - Data Processing

Institutions selection and basis

5 representative investment companies

Berkshire
Hathaway

JPMorgan

Bridgewater

Renaissance

TwoSigma

More than 25 % growth for
its shareholders each year

One of the most
famous investment
banks in the world

The world's largest
hedge fund

Leaders in quantitative investing



Renaissance



Part B - Data Processing

Data Acquisition

How: Web scraping with Python from the SEC home and search the EDGAR System

Specify the CIK of the above five companies and run the web scraping code to get the data of two quarters for analysis and comparison.



Use the *xml.etree.ElementTree* package to gather data in the xml files

<input type="checkbox"/> BerkshireHathawayInc-2021q1	2021/5/30 18:18	XML 文档
<input type="checkbox"/> BerkshireHathawayInc-2020q4	2021/5/30 18:20	XML 文档
<input type="checkbox"/> BridgewaterAssociates-2020q4	2021/5/30 18:28	XML 文档
<input type="checkbox"/> BridgewaterAssociates-2021q1	2021/5/30 18:27	XML 文档
<input type="checkbox"/> Jpmorgan-2020q4	2021/5/30 18:40	XML 文档
<input type="checkbox"/> Jpmorgan-2021q1	2021/5/30 18:36	XML 文档
<input type="checkbox"/> RENAISSANCE-2020q4	2021/5/30 18:30	XML 文档
<input type="checkbox"/> RENAISSANCE-2021q1	2021/5/30 18:30	XML 文档
<input type="checkbox"/> TwoSigma-2020q4	2021/5/30 18:32	XML 文档
<input type="checkbox"/> TwoSigma-2021q1	2021/5/30 18:31	XML 文档

```
<informationTable xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.sec.gov/edgar/document/thirteent/informationTable">
  <infoTable>
    <nameOfIssuer>ABBVIE INC</nameOfIssuer>
    <titleOfClass>COM</titleOfClass>
    <cusip>00287Y109</cusip>
    <value>2474794</value>
    <shrsOrPrnAmt>
      <sshPrnamt>22868178</sshPrnamt>
      <sshPrnamtType>SH</sshPrnamtType>
    </shrsOrPrnAmt>
    <investmentDiscretion>DFND</investmentDiscretion>
    <otherManager>4,11</otherManager>
    <votingAuthority>
      <Sole>22868178</Sole>
      <Shared>0</Shared>
      <None>0</None>
    </votingAuthority>
  </infoTable>
</informationTable>
```

Part B - Data Processing

Data Cleaning

```
def parseXML(XMLfile):
    tree = ET.parse(XMLfile)
    root = tree.getroot()
    dict = {'nameOfIssuer': [],
           'cusip': [],
           'sshPrnamt': []}
    for i in range(len(root)):
        dict['nameOfIssuer'].append(root[i][0].text)
        dict['cusip'].append(root[i][2].text)
        dict['sshPrnamt'].append(root[i][4][0].text)
    df = pd.DataFrame(dict)
    df['sshPrnamt'] = df['sshPrnamt'].astype(int)
    df_sum = df.groupby('nameOfIssuer', as_index=False)['sshPrnamt'].agg('sum')
    return df_sum
```

Since the acquired data set is complete and there are no missing values, there is not much data cleaning required. We only combined the number or market value of shares of the same issuer due to some duplications of issuer names.

Part C - Exploration and Analysis

1. How are investment firms connected?

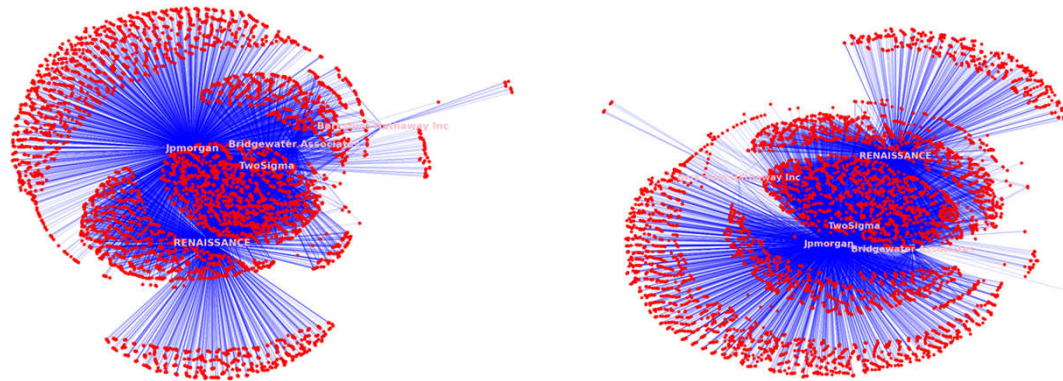


Figure 1. Graph plot in 2020 Q4 (left) and 2021 Q1 (right)

To show the connections between the investment companies, we can draw the investment network of the five selected companies.

The network is connected by some nodes which means the companies are invested by some or all the investment firms. We can see that JPMorgan has invested the largest number of companies and Berkshire Hathaway Corporation is the opposite.

Part C - Exploration and Analysis

2. Which companies are co-invested by these big players? How many companies are co-invested by some of big players?

co-invested companies in 2020 Q4		co-invested companies in 2021 Q1	
ABBVIE INC		AON PLC	APPLE INC
KROGER CO		AXALTA COATING SYS LTD	BIOGEN INC
LIBERTY GLOBAL PLC		BRISTOL-MYERS SQUIBB CO	COCA COLA CO
US BANCORP DEL		JOHNSON & JOHNSON	KROGER CO
—		MONDELEZ INTL INC	UNITED PARCEL SERVICE INC
—		VERISIGN INC	VISA INC

Table 1. Co-invested companies

As shown in Table 1, we can clearly find that co-invested companies by the five big players have changed a lot in the two quarters. Only 4 companies are co-invested in 2020 Q4 while the number became 12 in 2021 Q1.

Part C - Exploration and Analysis

2. Which companies are co-invested by these big players? How many companies are co-invested by some of big players?

co-invested by who?	Number in 2020 Q4	Number in 2021 Q1
Five Big Players	4	12
TwoSigma and RENAISSANCE	1503	1624
Jpmorgan and RENAISSANCE	2496	2499
Jpmorgan and TwoSigma	2028	2355
Jpmorgan, RENAISSANCE and TwoSigma	1404	1549

Table 2. Number of companies co-invested by some of big players

Two giants are selected, Renaissance and TwoSigma. We use dijkstr algorithm to find the same enterprise invested by these two institutions. The final results show that in the fourth quarter of 2020, a total of 1,503 companies were co-invested by the two institutions in the shortest path, and this number became 1,624 by the first quarter of 2021, indicating that the investment structures of the two institutions have become more similar (Commonality promoted).

Part C - Exploration and Analysis

3. Which investment firms have the largest number of shares or market value of a certain company?

Fund	Company	Shares
Berkshire Hathaway Inc	APPLE INC	887135554
JPMorgan	APPLE INC	137145874
TwoSigma	APPLE INC	2615623
RENAISSANCE	APPLE INC	1237532
Bridgewater Associates	APPLE INC	2803

Table 3. Shareholding of APPLEINC in 2021 Q1

For example, APPLE INC was invested by all the five investment firms in 2021 Q1. By using the weight of edge in the graph (the number of shares), we can know how many shares does each investment firm hold for APPLE INC. The results show that Berkshire Hathaway Inc has the largest shares (887135554) of APPLE INC among these five institutions in 2021 Q1.

Part C - Exploration and Analysis

4. Has the connectivity increased?

Number of investment institutions	Degree Centrality (2020 Q4)	Number of companies	Degree Centrality (2021 Q1)	Number of companies
1	0.000185	2138	0.000176	2264
2	0.000369	1669	0.000351	1704
3	0.000554	1325	0.000527	1482
4	0.000738	279	0.000702	231
5	0.000923	4	0.000878	12

Table 4. Degree centrality of five institutions in 2020 Q4 (left) and 2021 Q1 (right)

By calculating and grouping by degree centrality, we can tell how many companies are invested by one or several companies. Thus we can measure the connectivity of different investment institutions and make a comparison.

It is clear from the Table 4 that connectivity has increased, as the number of companies co-invested has increased. For instance, the number of companies invested by all five institutions rose from 4 in 2020 Q4 to 12 in the 2021 Q1.

Part C - Exploration and Analysis

5. Is there a change in the network metrics?

Institutions(2020 Q4)	betweenness centrality	Institutions(2021 Q1)	betweenness centrality
JPmorgan	0.7013	JPmorgan	0.7255
RENAISSANCE	0.2999	RENAISSANCE	0.2539
TwoSigma	0.0939	TwoSigma	0.1090
Bridgewater Associates	0.0044	Bridgewater Associates	0.0031
Berkshire Hathaway Inc	0.0033	Berkshire Hathaway Inc	0.0025

Table 5. Betweenness centrality of five institutions in 2020 Q4 and 2021 Q1

Quarters	Density
2020 Q4	0.000721
2021 Q1	0.000684

Table 6. The density in 2020 Q4 and 2021 Q1

Part C - Exploration and Analysis

5. Is there a change in the network metrics?

