Modern Data Analytics

Uncovering the network

Group: Demark

GroupMembers： Hao Zhu (r0816422), Shiqi Wang (r0815604), Xiao Wang (r0819511), Yuexue Jin (r0835821), Xierui Song (r0829539)

Introduction

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. In this dataset, the most important variables are *nameOfIssuer* (name of the holding), *cusip*(a unique identifier for each security) and *sshPrnamt* (number of shares or market value).

Background of data selection

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.

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?

Data Processing

Institutions selection and basis

Since choosing too many investment firms makes the network redundant and complex, it also reduces the efficiency of data processing. We selected five representative investment companies, which are Buffett, JPMorgan, Bridgewater, Renaissance and TwoSigma. Buffett's Berkshire Hathaway Corporation generates more than 25 percent value growth for its shareholders each year on average; JPMorgan Chase is one of the most famous investment banks in the world; Bridgewater is the world's largest hedge fund; Renaissance Technologies and TwoSigma are leaders in quantitative investing. Therefore, it is very representative and meaningful to study the investment network of these five investment institutions.

Data Acquisition

We acquire the data by web scraping with Python from the SEC home and search the EDGAR System[1] . We use the loop statements constantly matched to find the location fields of the XML files in the source code of the web page, then 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. All the data are present in XML format. We will use the *xml.etree.ElementTree* package to gather data in the 13F-HR files.

---

[1] https://sec.report/Document/Search/?formType=13F-HR#results

Data cleaning

Since the acquired data set is complete and there are no missing values, there is not much data cleansing required. However, due to the duplication of issuer names in the data set of the same investment institution, we combined the number or market value of shares of the same issuer.

Exploration and Analysis

1. How are investment firms connected?

Considering how to show the connections between the investment companies, we can draw the investment network of the five selected companies with *networkx* package in Python. As shown in Figure 1, we can see that JPMorgan has invested the largest number of companies and Berkshire Hathaway Corporation is the opposite. The network is connected by some nodes which means the companies are invested by some or all the investment firms. We will find out these companies in the following parts.
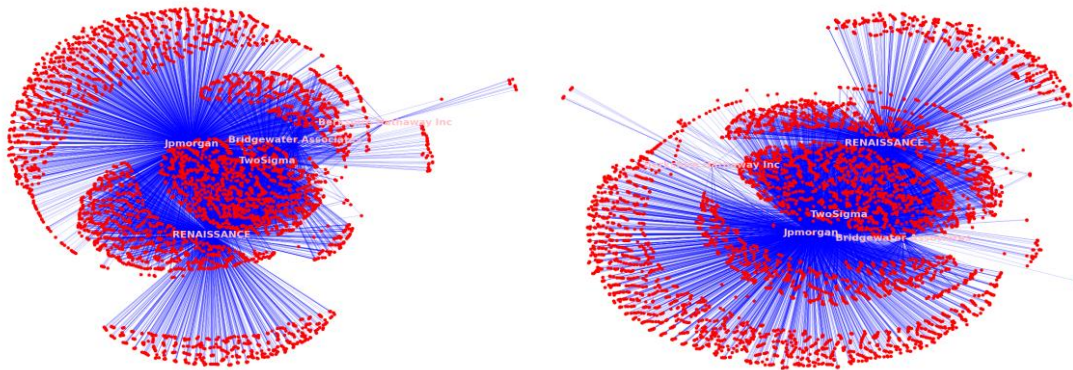

Figure 1. Graph plot in 2020 Q4 (left) and 2021 Q1 (right)

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 paste quarters. Only 4 companies are co-invested in 2020 Q4 while the number became 12 in 2021 Q1. In addition, we can pick two investment institutions and zoom in to see the connection between them. Here, we choose two giants in the quant investment fund industry, Renaissance and TwoSigma, which are representative and comparable, then we apply shortest path (*by dijkstr algorithm)* to find the investment of two institutions on the same enterprise with shortest path to explore their investment commonality. The results show that in 2020 Q4, a total of 1503 companies were co-invested by the two institutions, and this number became 1624 by 2021 Q1, indicating that the investment structures of the two institutions have become more similar. Similarly, we can learn more information about how many companies are co-invested by some or all the five big players as shown in Table 2.

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

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. Shares of APPLE INC in 2021 Q1

We can get the comparison of shares by comparing the weight of edge of the same company, because the size of weights represents the value of SSHPRNAMT. Here, we take a look at five firms and compare their share of Apple in the 1st season in 2021. The results show that Berkshire Hathaway Inc has the largest shares (887135554) of APPLE INC among these five institutions.

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)

For this question, we can make a longitudinal comparison by selecting data from two quarters, namely, the fourth quarter of 2020 and the first quarter of 2021. By using degree centrality, 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 considerably. For instance, the number of companies invested by all five institutions rose from 4 in 2020 Q4 to 12 in the 2021 Q1.

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

We use three metrics to measure the change of network, which are betweenness centrality, degree centrality and density. As for degree centrality, which we mentioned in the previous question, there has been a significant improvement during the two quarters. As shown in Table 5, the betweenness centrality of 2 companies has increased and betweenness centrality of 3 companies has decreased, but the rankings of the five institutions have not changed. The betweenness centrality of JPMorgan is always the highest, which is largely related to the large number of companies it invests in, making it the broadest investor in the investment field. Density is simply the ratio of actual edges in the network to all possible edges in the network, which gives you a quick sense of how closely knit your network is. From Table 6, we can observe that the density shows a slight decrease, which means our network becomes a little bit looser during the two quarters.

You can view the whole project in S3 bucket and GitHub page[link].

References

[1] Networkx_Reference:https://networkx.org/documentation/latest/_downloads/networkx_reference.pdf

[2] Network-Analysis-Made-Simple:https://ericmjl.github.io/Network-Analysis-Made-Simple/

[3] Kadry, S. , & Taie, M. . (2017). Python for graph and network analysis. Springer International Publishing.

[4] James Powell, Matthew Hopkins, in A Librarian's Guide to Graphs, Data and the Semantic Web, 2015

[5] Peter V. Marsden, in Encyclopedia of Social Measurement, 2005