**“INVESTORS RANKING USING PAGE RANK ALGORITHM”**

A DISSERTATION

SUBMITTED TO

DEPARTMENT OF STATISTICS

SVKM’S NMIMS (DEEMED -TO-BE UNIVERSITY)

IN PARTIAL FUFILLMENT FOR THE DEGREE OF

**MASTER OF SCIENCE**

**IN STATISTICS AND DATA SCIENCE**

BY

SAURABH BHADALE

UNDER THE SUPERVISION OF

DR. K S M RAO

DR. DEBASMITA MUKHERJEE

**SUNANDAN DIVATIA SCHOOL OF SCIENCE**

SVKM’S Narsee Monjee Institute of Management Studies

(Deemed-To-Be University)

Vile Parle West, Mumbai, Maharashtra 400056

NOVEMBER, 2021

**Acknowledgement**

We would like to express our warmest thanks of gratitude to our supervisors Dr. K S M Rao and Dr. Debasmita Mukherjee for providing us the opportunity to do this project on ‘Investors Ranking using Page Rank Algorithm .Their invaluable guidance and advice helped us to carry out the project successfully.

We would like to thank our course coordinator Dr. Pradnya Khandeparkar for providing necessary facilities required for completion of this project.

We would also like to extend our gratitude to the whole Department of Statistics for guiding and supporting us at different stages of our project.

Finally, with great support of our team members and excellent guidance of mentors we are able to complete our project within specific time.

**Abstract**

In a time where the number of start-ups in India is increasing rapidly in every sector, the competition to get the investors for a start-up has also increased. So, it would be beneficial to target an investor/ investment firm who is likely to influence other investors/ investment firms to invest in said start-up. Now, how do we decide which investor can influence other investors in such a way? To answer this question, we have used the method of Page Rank algorithm to rank the investment firms. We consider the investment firms who invested in Indian start-ups that have grown to become unicorns. Considering the data for ‘series A’ and ‘series B’ funding rounds, we create an adjacency matrix and a network diagram that show the links between various investment firms. Using the estimated annual revenue generated by an investment firm, we direct the network graph then create the ‘link matrix’. To start with, we assume the ranks for all the investment firms as equal. Now, using the eigen vector concept of Page Rank algorithm, we carry out the iterations to obtain successive rank vectors. On obtaining the final rank vector, we assign the ranks to the top 15 investment firms. These ranks are based on the influence that an investment firm might have in bringing other investors to invest in a start-up.

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Topic** | **Page Number** |
| **1.** | Introduction to Page Rank Algorithm  1.1 History of page rank algorithm  1.2 Different page rank Algorithm | 1-2  2-3 |
| **2.** | Glossary  2.1 Types of investors  2.2 Concept of investors Rank  2.3 Different Types of Startups  2.4 Different Types of funding | 4-5  5  6-7  7-10 |
| **3** | Objective and scope  Why Page Rank algorithm is used? | 10  10-11 |
| **4** | Literature review | 11-12 |
| **5** | Methodology  5.1 Data collection  5.2 Creation of adjacency matrix  5.3 Network analysis  5.4 Link matrix  5.6 Initial rank vectors and iterations  Outcomes | 12-13  13-14  14-18  18-19  19-21  21-22 |
| **6** | Conclusion | 22 |
| **7** | Future Scope of the work | 23 |
| **8** | References | 23-24 |

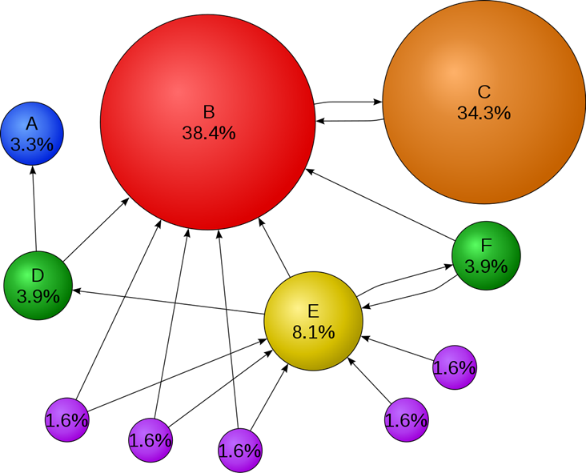
**CONTENT**

**Introduction**

In the last few decades, lots of data have been created and made available through the Internet. Various search engines, like Google, make the data easily accessible with the help of links. To make sure that the relevant data can be accessed within a few clicks, there needs to be a technique to rank the webpages based on their ‘importance’.

The founders of Google, **Larry Page** and **Sergey Brin** introduced the Page Rank Algorithm.[1] The motive behind the Page Rank Algorithm was **“to rank the webpages based on their importance”.** The underlying thought was that the more important webpages are likely to receive more links from other webpages.

Now we see in short how page rank algorithm works with the help of diagram



.

We consider a few webpages, say A,B,C,D,E,F.These webpages consist of incoming and outgoing links to and from the other pages.Webpage B consits of more number of incoming links, hence it has higher page rank .Webpage C is directly connected to webpage B with an incoming link, hence its rank gets boosted.Other webpaes have lesser number of incoming link and consequently, they have smaller page ranks.

Chris Farmer, a venture capitalist, developed Investor Rank primarily using a public domain database of venture capitalists and start-up companies, CrunchBase. [1]

In this project, we take on the Investor Rank Problem and use the Page Rank Algorithm to rank the investors that have invested in the Indian Unicorn Start-ups. [2]

**Different Page Rank based algorithms:**

**Weighted Page Rank (WPR) algorithm:**

The more popular a webpage is, the more are the linkages that other webpages tend to have to them. Weighted PageRank algorithm is an extension of the conventional PageRank algorithm based on the same concept.

Weighted PageRank algorithm assigns higher rank values to more popular (important) pages instead of dividing the rank value of a page evenly among its outlink pages. Each out link page gets a value proportional to its popularity, i.e., its number of in links and out links.

To a webpage ‘u’, an inlink is a URL of another webpage which contains a link pointing to ‘u’. Similarly, to webpage ‘u’, an outlink is a link appearing in ‘u’ which points to another webpage. The number of inlinks is represented by **Win(v,u)**and the number of outlinks is represented as **Wout(v,u)**.

**Win(v,u)**is the weight of link (v, u) calculated based on the number of inlinks of page u and the number of inlinks of all reference pages of page v.

Here, Ipand Iurepresent the number of inlinks of page ‘p’ and ‘respectively. R(v)represents the list of all reference pages of page ‘v’. **Wout(v,u)**is the weight of link (v, u) calculated based on the number of outlinks of page u and the number of outlinks of all reference pages of page v.

Here, Op and Ou represent the number of outlinks of page ‘p’ and ‘u’ respectively. R(v) represents the list of all reference pages of page ‘v’.

Based on the importance of all pages as describes by their number of inlinks and outlinks, the Weighted PageRank formula is given as:

Here, **PR(x)** refers to the Weighted PageRank of page x. **d** refers to the damping factor. The PageRank theory holds that an imaginary surfer who is randomly clicking on links will eventually stop clicking. The probability, at any step, that the person will continue is the damping factor. [3]

**HITS (Hyperlink Induced Topic Selection):**

**Hyperlink Induced Topic Search**(HITS) Algorithm is a Link Analysis Algorithm that rates webpages, developed by Jon Kleinberg [15]. This algorithm is used to the web link-structures to discover and rank the webpages relevant for a particular search.   
HITS uses hubs and authorities to define a recursive relationship between webpages. Before understanding the HITS Algorithm, we first need to know about Hubs and Authorities. Given a query to a Search Engine, the set of highly relevant web pages are called **Roots**. They are potential **Authorities**. Pages that are not very relevant but point to pages in the Root are called **Hubs**. Thus, an Authority is a page that many hubs link to whereas a Hub is a page that links to many authorities.[4]

**Eigen Rumor Algorithm:**

The algorithm is designed for ranking information resources provided as blogs or other cyberspace communities, in which the identities of information providers are observable. Unlike generic web pages, a blog site is constructed from a set of blog entries written by a single blogger and the quality of blog entries and topics are dominated by the ability or interests of the blogger. Using this structural characteristic of blogs, the Eigen Rumor algorithm rates a new blog entry or other blog entries that have no in-links according to the past behaviour of the blogger.[5]

**Glossary**

**Types of Investors**

Friends & Family

The first type of investor entrepreneurs should be approaching at the very beginning are friends and family and close personal contacts. At this stage there is very little hard evidence and proof to base a real investment or funding on. They are essentially investing in the idea, and far more importantly - you. These are the people that already know you, like and trust you and believe in you the most. This type of investor may not provide a lot of money. It could be in the range of $1,000 to $200,000. Though if you can’t raise money from this group, other investors are probably going to ask themselves why.

Angel Investors

Professional angel investors are normally approached when it comes to the seed round and beyond. They are willing to fund smaller operations than VCs, may be more flexible in terms, and can offer a lot of value in wisdom and connections.Angel investors can be approached directly online, at live pitch events, and through introductions from other start-up founders.

Venture Capitalist Firms

VCs are the holy grail of investors for fundraising entrepreneurs. They come with the biggest checks, the most power to fuel success and gaining market share and achieving more credibility and visibility. More venture capital firms are looking at and are participating in earlier funding rounds. Though it is much more likely these investors will show up and be secured in Series A, B and C fundraising rounds than earlier. Do note that not all of these firms are created equal. The best match can be influenced by location, the timeline of their funds, their interest and expertise in a certain field, their power to help you get to the next stage and of course, how they treat their founders.

Corporate Investors:

Investing in start-ups carries a variety of benefits for big corporations. Including supporting their own to achieve growth numbers, diversifying assets, and identifying talent and technology which can help them fend off industry changes and fuel revenues and profits. Some have funds to invest in outside start-ups. More are launching their own accelerator and incubator programs and ecosystems for cultivating these opportunities. These investors can be great allies in taking your business to the next level. Though they can be quite different to work with, and any integration or collaboration on sales channels, systems and customer bases needs to be approached carefully and with a lot of patience. Founding entrepreneurs and corporate investors often have completely different styles and perspectives. It’s going to be vital to learn to understand each other and have some boundaries set up when going in, if this is going to be an enjoyable relationship.[6]

**Concept of Investors Rank**

Investor Rank was developed by the venture capitalist Chris Farmer using CrunchBase augmented with data from individual venture capitalist websites (<http://crunchbase.com/>).

If two investors invested in the same company at the same time, during a round of funding, the investors are joined by an edge. Second, we infer that there are at least two mechanisms that are used to boost an investor’s rank:

1. bringing a new investor to a company’s rounds of funding boosts the prior investors’ ranks, and
2. if an investor continues to invest in a start-up company across more than one round of funding, then that investor’s rank is boosted.

**What is investor rank system?**

**Rank 1:** (Highest): This investor firm can be termed as the most influential, and would be the preference of various Indian start-ups looking for investors.

**Rank 2** :(Above Average): This investment firm would be the second most influential firm, and should be the target after Rank 1.

**Rank 3:** (Average): Not nearly as influential as Rank 1 and Rank 2, but if the pitches to the top 2 ranked firms fail, the start-ups looking for investors would go for this investment firm.

**Different types of start-ups**

**Minicorn Start-ups**

**Minicorn start-ups** are companies with valuations of more than $ 1 million and they are still on the rise to become a unicorn business. Every entrepreneur has dreams that they will grow "massively" together with their business. And one of them has the luckiest way to make their dreams come true.

Some examples of minicorn start-ups:

* **Cashaa**- a blockchain-powered forex platform
* **Dualitybanx**- a forex product for people who work with multiple currencies
* **Know Your Customer**- the world's only secure, fully integrated platform that provides instant ID verification for both individual and business customers
* **New Gen**- a payment solution provider, international money transfer & multi-channel e-commerce

**Soonicorn start-ups:**

Start-ups have growth potential and the possibility of joining a unicorn club called Soonicorn. **Soonicorn company** is primarily funded and financed by an Angel Investor or a venture capitalist. Based on future forecasts about the industry market and firm valuations so they come up with a valuation for the business. And of course, it always exceeds the real value of the startups. In similar cases, larger organizations acquire startups, and this leads to valuing them above their actual net worth. This is a favorable condition to help startups join these Soonicorn clubs earlier and get closer to a $ 1 billion asset valuation.

Some example of Soonicorn start-ups:

* **Xpressbees** – eCommerce Logistics
* **Rebel Foods** – Foodtech
* **Prato** – Telemedicine
* **Pepper fry** – Vertical eCommerce

**Unicorn**

**Unicorn is something that always gets the attention of investors.**

**Unicorns** are public start-ups that are valued by venture capital one billion dollars or more. This term was created by Aileen Lee, the founder of Cowboy Ventures, and it appeared in 2013. It's worth noting that you'll be able to see a large number of those names look like they're missing from the list books such as Facebook and Amazon. These corporations are currently priced at a terrible number and significantly different from this position.



**Hectocorn**

A tech, financial, or **fintech companies** worth more than $ 100 billion is called Hectocorn. Or there is another name for the company, this corporation is "Super Unicorn". It is not uncommon that one of our familiar names such as Apple, Google, Microsoft, Facebook, Oracle and Cisco are examples of **Hectocorn.**

**Different types of Fundings**

**Series A Funding:**

Once A business has developed a track record (an established user base, consistent revenue figures, or some other key performance indicator), that company may opt for [Series A funding](https://www.investopedia.com/terms/s/seriesa.asp) in order to further optimize its user base and product offerings. Opportunities may be taken to scale the product across different markets. In this round, it’s important to have a plan for developing a business model that will generate long-term profit. Often times, seed startups have great ideas that generate a substantial amount of enthusiastic users, but the company doesn’t know how it will monetize the business. Typically, Series A rounds raise approximately $2 million to $15 million, but this number has increased on average due to high tech industry valuations, or [unicorns](https://www.investopedia.com/terms/u/unicorn.asp). The average Series A funding as of 2020 is $15.6 million.

In Series A funding, investors are not just looking for great ideas. Rather, they are looking for companies with great ideas as well as a strong strategy for turning that idea into a successful, money-making business. For this reason, it's common for firms going through Series A funding rounds to be valued at up to $23 million.﻿[7]

The investors involved in the Series A round come from more traditional venture capital firms. Well-known venture capital firms that participate in Series A funding include Sequoia Capital, Benchmark Capital, Greylock and Accel Partners.

By this stage, it's also common for investors to take part in a somewhat more political process. It's common for a few venture capital firms to lead the pack. In fact, a single investor may serve as an "anchor." Once a company has secured a first investor, it may find that it's easier to attract additional investors as well. Angel investors also invest at this stage, but they tend to have much less influence in this funding round than they did in the seed funding stage.

It is increasingly common for companies to use equity crowdfunding in order to generate capital as part of a Series A funding round. Part of the reason for this is the reality that many companies, even those which have successfully generated seed funding, tend to fail to develop interest among investors as part of a Series A funding effort. Indeed, fewer than half of seed-funded companies will go on to raise Series A funds as well.

## **Series B Funding**

[Series B](https://www.investopedia.com/terms/s/series-b-financing.asp) rounds are all about taking businesses to the next level, past the [development stage](https://www.investopedia.com/terms/d/developmentstage.asp). Investors help startups get there by expanding market reach. Companies that have gone through seed and Series A funding rounds have already developed substantial user bases and have proven to investors that they are prepared for success on a larger scale. Series B funding is used to grow the company so that it can meet these levels of demand.

Building a winning product and growing a team requires quality talent acquisition. Bulking up on [business development](https://www.investopedia.com/articles/personal-finance/090815/basics-business-development.asp), sales, advertising, tech, support, and employees costs a firm a few pennies. The average estimated capital raised in a Series B round is $33 million. Companies undergoing a Series B funding round are well-established, and their valuations tend to reflect that; most Series B companies have valuations between around $30 million and $60 million, with an average of $58 million.

Series B appears similar to Series A in terms of the processes and key players. Series B is often led by many of the same characters as the earlier round, including a key anchor investor that helps to draw in other investors. The difference with Series B is the addition of a new wave of other venture capital firms that specialize in later-stage investing.[7]

**Series C Funding:**

Businesses that make it to Series C funding sessions are already quite successful. These companies look for additional funding in order to help them develop new products, expand into new markets, or even to acquire other companies. In Series C rounds, investors inject capital into the meat of successful businesses, in an effort to receive more than double that amount back. Series C funding is focused on scaling the company, growing as quickly and as successfully as possible.

One possible way to scale a company could be to acquire another company. Imagine a hypothetical start-up focused on creating vegetarian alternatives to meat products. If this company reaches a Series C funding round, it has likely already shown unprecedented success when it comes to selling its products in the United States. The business has probably already reached targets coast to coast. Through confidence in market research and [business planning](https://www.investopedia.com/terms/b/business-plan.asp), investors reasonably believe that the business would do well in Europe.

Perhaps this vegetarian startup has a competitor who currently possesses a large share of the market. The competitor also has a competitive advantage from which the startup could benefit. The culture appears to fit well as investors and founders both believe the [merger](https://www.investopedia.com/terms/m/merger.asp) would be a [synergistic](https://www.investopedia.com/terms/s/synergy.asp) partnership. In this case, Series C funding could be used to buy another company.

As the operation gets less risky, more investors come to play. In Series C, groups such as [hedge funds](https://www.investopedia.com/terms/h/hedgefund.asp), investment banks, [private equity](https://www.investopedia.com/terms/p/privateequity.asp) firms, and large secondary market groups accompany the type of investors mentioned above. The reason for this is that the company has already proven itself to have a successful business model; these new investors come to the table expecting to invest significant sums of money into companies that are already thriving as a means of helping to secure their own position as business leaders.

Most commonly, a company will end its external equity funding with Series C. However, some companies can go on to Series D and even Series E rounds of funding as well. For the most part, though, companies gaining up to hundreds of millions of dollars in funding through Series C rounds are prepared to continue to develop on a global scale. Many of these companies utilize Series C funding to help boost their valuation in anticipation of an IPO. At this point, companies enjoy valuations in the area of $118 million most often, although some companies going through Series C funding may have valuations much higher.4﻿ These valuations are also founded increasingly on hard data rather than on expectations for future success. Companies engaging in Series C funding should have established, strong customer bases, revenue streams, and proven histories of growth.

Companies that do continue with Series D funding tend to either do so because they are in search of a final push before an IPO or, alternatively, because they have not yet been able to achieve the goals, they set out to accomplish during Series C funding. [7]

**In this project we have considered SERIES A, B rounds of funding.**

**Objectives**

1. Ranking investors using Page Rank algorithm and network centrality.
2. To identify the most influential Venture Capitalist firms in India.

**Why Page Rank Algorithm is used in this project:**

There are two reasons to this:

1. **Network quality is a driver of fund returns**. Academics have shown that better networked venture investors perform better. The measure, often referred to as network centrality, is something we can quantify by building a Google PageRank for investors.
2. **An objective measure of network strength**. Every investor says they have a great network. Like “proprietary deal flow”, it is a claim that sounds great, but unfortunately, has largely been unverifiable.

**Literature Review**

In [8] a brief idea is given about how page Rank is calculated by using concept of eigen values and Brief information about random suffer model.

In [9] [Chris Farmer](http://www.crunchbase.com/person/chris-farmer), a VC at General Catalyst Partners, has come up with a method which he calls Investor Rank. Similar to Google’s PageRank orders search results based on how many links each page gets from other sites to calculate the importance of webpages.

Investor Rank looks at the connections between VC firms. Whenever two VC firms co-invest in the same deal, that creates a bond between them. If one VC firm follows another one in a later round, that boosts the rank of the earlier investor.

Chris Farmer revealed the top 10 VC firms based on Investor-Rank. They are:

1. Andreessen Horowitz
2. Sequoia Capital
3. Accel
4. Benchmark Capital
5. Union Square Ventures
6. General Catalyst Partners
7. NEA
8. Kleiner Perkins
9. Khosla Ventures
10. Greylock

In [10] Bryan Sims explains the application of Inverse Page Rank algorithm to the Investor-Rank Problem.

Bryan Sims mined the same Crunchbase dataset as Chris Farmer, produced an adjacency matrix for the coinvestment graph and obtained his own Top 10 venture capitalist rankings. On noticing that the ranking he obtains is different from Chris Farmer’s Top 10 ranking, he proceeds further to use the Inverse-Rank to manipulate the rankings to obtain the same rankings as Farmer.

In [11] application of Page Rank algorithm used to rank Angel Investors, Micro VCs, and Corporate VCs that assesses network strength and quality.

Stock portfolio selection using learning-to-rank algorithms with news sentiment.

In [12] Qiang song and Anqi Liu (authors) developed learning-to-rank algorithms to design trading strategies using relative performance of a group of stocks based on investors' sentiment toward these stocks. They show that learning-to-rank algorithms are effective in producing reliable rankings of the best and the worst performing stocks based on investors' sentiment.

More specifically, they use the sentiment shock and trend indicators introduced in the previous studies, and they design stock selection rules of holding long positions of the top 25\% stocks and short positions of the bottom 25\% stocks according to rankings produced by learning-to-rank algorithms.

After that they apply two learning-to-rank algorithms, ListNet and RankNet, in stock selection processes and test long-only and long-short portfolio selection strategies using 10 years of market and news sentiment data. Through backtesting of these strategies from 2006 to 2014, they demonstrate that portfolio strategies produce risk-adjusted returns superior to the S\&P500 index return, the hedge fund industry average performance - HFRIEMN, and some sentiment-based approaches without learning-to-rank algorithm during the same period.

{\displaystyle PR(A)={\frac {PR(B)}{2}}+{\frac {PR(C)}{1}}+{\frac {PR(D)}{3}}.\,}

{\displaystyle PR(A)={\frac {PR(B)}{L(B)}}+{\frac {PR(C)}{L(C)}}+{\frac {PR(D)}{L(D)}}.\,}**Data Collection**

1. After going through several research papers, we finally settled with **Investor Rank and an Inverse Problem for PageRank by Bryan Sims** for the Investor ranking part.[3]
2. For data collection, we consider the 60 Indian Unicorn start-ups up till August 2021 from Venture Intelligence website.[13]
3. All the investors for the 60 Indian Unicorn start-ups were recorded and the data on the investors and the funding series was collected from Dealroom app and website. [14]
4. For the Estimated Annual Revenue data for various investors considered, we collected the required data from different websites like owler.com and zoominfo.com.
5. The collected data was arranged in the Excel sheets as companies, their investors according to investment round where we have considered Series **A** and **B**.
6. The Estimated Annual Revenue data was arranged in another Excel sheet according to the Investors.

**Methodology**

The Page Rank Algorithm consists of the following steps,

1) **Creation of the Adjacency Matrix:**

We have considered the investors for 60 Indian Unicorn start-ups. There were a total of 135 investors that had invested in these 60 unicorn start-ups.

To arrange the data, we took all the investors and put them each as the row and column headers for the matrix. The rows represent the follower and column represent the followed.

Next, the task was to identify the links between different investors. So, the start-up companies that were common between the two investors filled the intersecting cell. To do this, the conditional formatting tool of MS Excel comes in handy.

After identifying all the links/ edges between various investors, we had to assign the factors 0 and 1 to the pairs of investors. We assign the factor 0 to the cells (pair of investors) that do not have any common investment in any of the considered start-ups. The factor 1 was assigned to any cells that was non-empty or that has some common investment in the considered start-ups. The factor 1 represents having a link between the investors. The number of common investments between any two investors was irrelevant for this algorithm. We just need to consider if there is a link or not.

The Excel tool of Find and Replace was used to fill all the empty cells with zeroes.

Now, any investor that does not have a common investment with any other investor, has zeroes in all its rows and columns corresponding to it. Such investors (nodes) as termed as ‘sinks’ in Page Rank Algorithm. So, all such investors that do not have any link with any other investor, neither incoming nor outgoing, are dropped from the matrix.

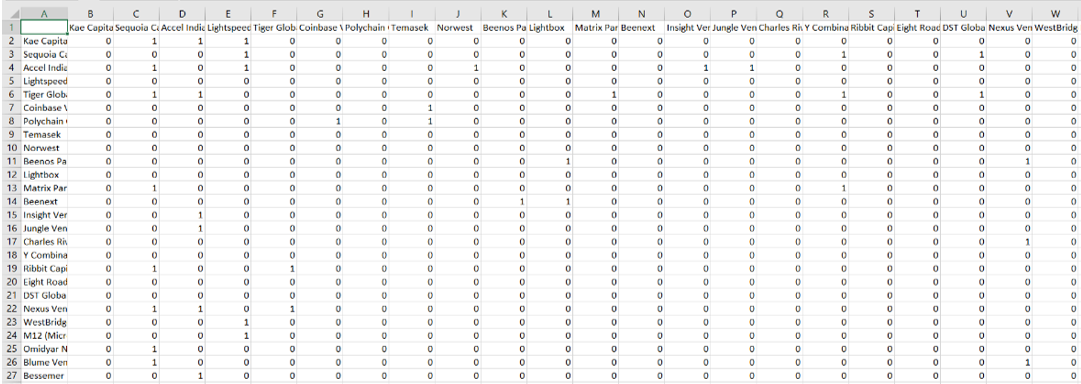
We notice now, that the obtained matrix is symmetric and would give an undirected network graph. But we require a directed network graph in the Page Rank problem. So, the next task is to convert the symmetric matrix into a directed one. For this, we consider the ‘Estimated Annual Revenue’ for all the investors.

The idea behind using the Estimated Annual Revenue was that an investor with higher estimated annual revenue is considered to be more successful in the business world. So, a ‘more successful investor’ is likely to influence other investors to invest in a start-up in which they have put their money in.

Now, let us consider two investors, ‘Kae Capital’ and ‘Sequoia Capital’. We notice that Sequoia Capital has a higher estimated annual revenue than Kae Capital. So, the network would be directed from Kae Capital to Sequoia Capital. The cell corresponding to the row Kae Capital and column Sequoia Capital would be assigned the value 1 while the column corresponding to row Sequoia Capital and column Kae Capital would be assigned the value 0. This implies that Sequoia Capital has likely influenced Kae Capital to invest in some start-ups of its liking.

On proceeding further, we get a directed matrix with elements 0 and 1 that we call the ‘Adjacency matrix’. This matrix is large and sparse since there are many investors and only a few collaborate or follow the other investor to invest in a particular start-ups.

Some part of adjacency matrix find below:



**2**) **Netwok Analysis.**

**WHAT IS A NETWORK?**

A network refers to various structures comprising variables, which are represented by **nodes,** and the relationships (formally called **edges**) between these nodes. Network analysis is used to understand the pattern of connection and interaction, also there is a huge loss of information when the complex real-world phenomenon is represented by simple vertex & edge connections.

In the project the variables are investors. So, all the investors are represented as nodes and the only relation between them is the companies they invested in commonly which are their edges, there are some differences in nomenclature in the network literature: nodes are sometimes referred to as vertices, edges are sometimes referred to as links, and networks are also called graphs.

We plot the graph in the R-studio with help of igraph package. For this first we have to get the data sorted in the form of their connections in the adjacency matrix formed, we considered the first investor and found its connections from adjacency matrix and noted the connections separately in an excel sheet similarly this process was carried out for entire data and got the data sorted, and then we used it in our code and the graph was formed consisting of the nodes and their connection as their edges. The code for the same is given below you can refer it and even try it in your own R-script by just installing igraph package.

**R code:**

library(igraph)

adja\_matrix<-graph(edges = c('1','2', '1','3', '1','4', '1','35',

'2','4', '2','17', '2','20', '2','42',

'3','2' ,'3','4' ,'3','9' ,'3','14' ,'3','15','3','33',

'5','2' ,'5','3' ,'5','12' ,'5','17' ,'5','20' ,'5','38' ,'5','42', '5','43',

'6','8', '6','52',

'7','6', '7','8' ,'7','50', '7','51',

'10','11' ,'10','21' ,'10','53' ,'10','56',

'11','53',

'12','2','12','17','12','24', '12','38',

'13','10', '13','11' ,'13','53',

'14','3',

'15','3' ,'15','33' ,'15','45' ,'15','54',

'16','21',

'18','2' ,'18','5' ,'18','42',

'19','39' ,'19','57',

'21','2' ,'21','3' ,'21','5' ,'21','26' ,'21','27' ,'21','40','21','55' ,'21','56',

'22','4',

'23','4',

'24','2',

'25','2', '25','21', '25','34', '25','40', '25','58',

'26','3' ,

'27','26',

'29','28',

'30','28', '30','29',

'31','3','31','14',

'32','5',

'33','15',

'34','2','34','3','34','9','34','17','34','26',

'35','2','35','3','35','4','35','5','35','18','35','20','35','20','35','42','35','43','35','44',

'36','6','36','7','36','37',

'37','6','37','7',

'41','21','41','25',

'43','2',

'44','2','44','5','44','3','44','5',

'45','2','45','34',

'46','3','46','5','46','48','46','49',

'48','5','48','21',

'49','5','49','21',

'50','6','50','8','50','52',

'51','50',

'53','8',

'54','15','54','46',

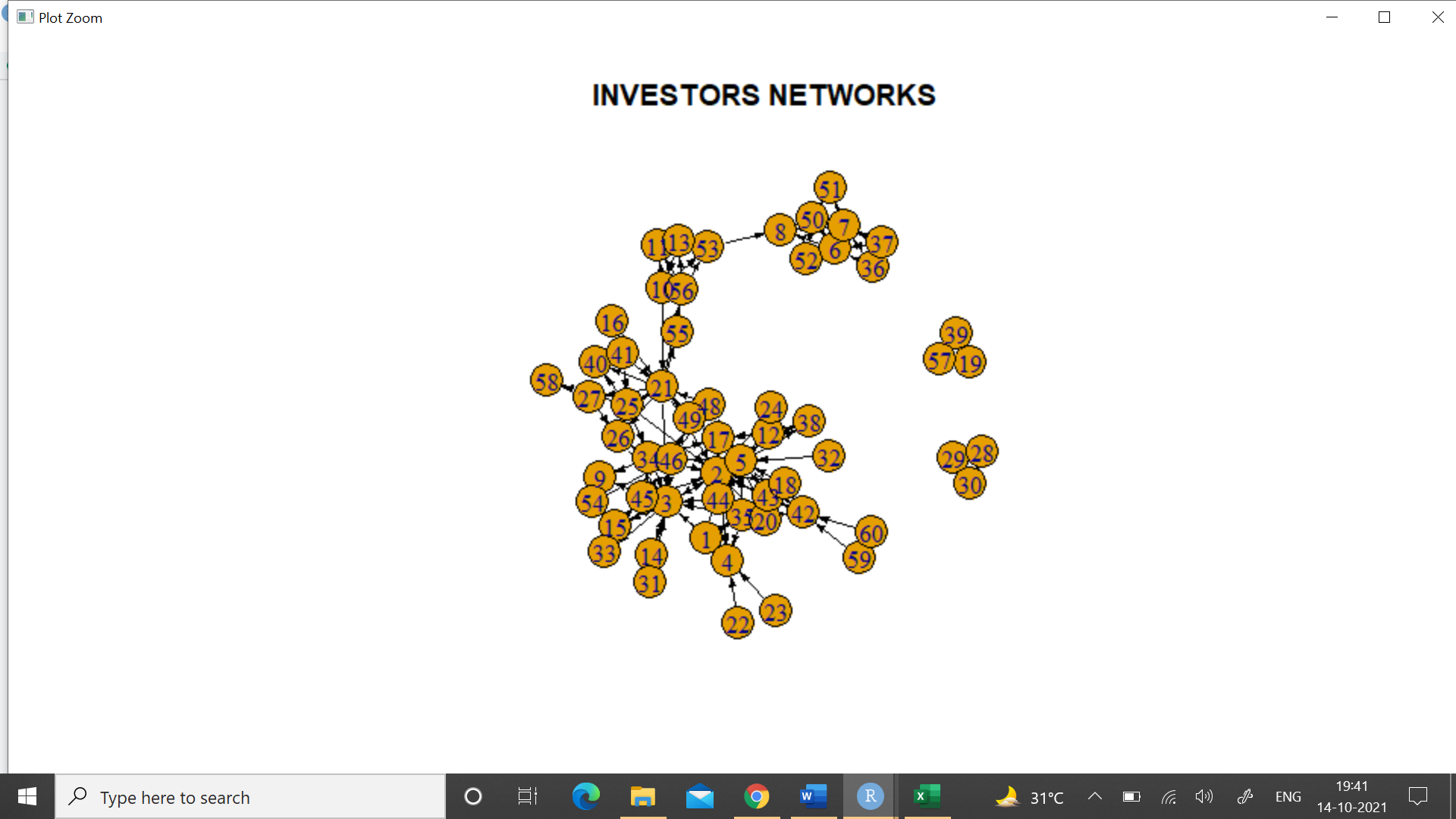
'56','11','56','13','56','53',

'57','39',

'59','42',

'60','42','60','59'),directed = TRUE)

plot(adja\_matrix,edge.color='black',edge.arrow.size= 0.2,vertex.cex=1,layout=layout.fruchterman.reingold, main="INVESTORS NETWORKS")



So, as we can see in the codes the adja\_matrix variable consists of all the edge information and the graph is directed so directed is given TRUE. And hence, we can see the plot.

We can also analyze the ranks on basis of this graph, the weight factor will come in the picture and the concept will be called as weighted page rank algorithm, where the edges with highest weight is bigger in size compared to the rest with lower and same goes for nodes, nodes with more number of connections gets bigger in size compared to the one’s with less connections. The weight can be considered any factor from the dataset. For our dataset we can consider the number of companies common between two nodes, higher the number of common companies more the weight of the edge and similary the connections.

3) **Link matrix (Probability matrix).**

Link matrix can be seen as a matrix of probability that an investor influences another investor to invest in a start-up. Link matrix is a stochastic matrix which has column sums as 1. Each column corresponding to an investor shows the probability of this investor to follow the investor in the corresponding rows.

To get the the link vectors for each investor, we need to divide each link with total number of links for the said investor. Suppose an Investor A has ‘outgoing links’ to 3 other investors B, C and D, so, its link vector would contain the probability 1/3, 1/3 and 1/3 of the rows corresponding to B, C and D respectively. All the other rows would simply contain zeroes. Sum of all the elements in a link vector is 1.

There surfaces another task. There may be investors with incoming links but no outgoing links. If we proceed directly, the resulting link matrix does not conform to the concept of a stochastic matrix (column sums as 1). So, to be fair and separate such investors/ nodes from the ‘sinks’, we distribute the outgoing probability of such nodes equally among all other nodes. So, we assign an outgoing probability of 1/57 corresponding to all other investors except itself.[8]

After link vectors for all the investors have been created, we combine them as columns in a matrix, calling the matrix as ‘Link Matrix’.

To create this Link matrix in our project, we use the R-software.

Using the adjacency matrix that we elaborated on earlier, we perform the following codes to get the Link Matrix.

**R-Code**

#Reading in the Adjacency Matrix.csv file

AM<-read.csv(file.choose(),row.names=1)

#We get a 58x58 Adjacency matrix

#To convert the Adjacency matrix into a Link Matrix

Sums<- rowSums(AM);Sums

Sums<- as.matrix(Sums)

i=1

j=1

for (i in 1:58){

for (j in 1:58){

if (Sums[i]!=0){

AM[i,j]= AM[i,j]/Sums[i]

}else AM[i,j]==0

}

}

AM<- as.matrix(AM)

Link\_Matrix <- t(AM) #Obtained Link Matrix

#To check if the Column Sums of the Link matrix are 1.

check <- colSums(Link\_Matrix)

check

4) **Initial Rank Vector and Iterations**

After obtaining the Link Matrix, we now proceed to the rank vectors.

From linear algebra, the eigen vector and eigen value concept, we have the equation,

where A is any matrix, v is the eigen vector and λ is the corresponding eigen value.

In Page Rank Algorithm, we proceed with the eigen value 1 i.e. λ = 1, A as the ‘**link matrix**’ and v as the rank vectors.

We apply iterations on the equation to obtain a final ranks.

For (r+1)th iteration, the equation can be modified to,

Here, k is the number of iterations it takes to reach the final rank vector.

Now, as we do not have an initial rank vector, we assume that all the investors have the same ranks. So, the initial rank vector contains all values as 1/58.

Then, as we multiply the initail rank vector with the link matrix, we get the next rank vector . Similarly, we proceed to iterate until two successive rank vectors obtained are identical.

As we obtain two successive identical rank vectors, the iteration is terminated and the last rank vector obtained is used as the final ‘Rank Vector.’

In the Rank Vector, the investor with higher probability is ranked better, i.e. the investor with the highest probability is ranked the ‘first’ and the one with the lowest probability is ranked the last.

To calculate the rank vectors and perform the iterations, we use the R software. The R-Codes used are shown below.

**R-Codes**

#Initial Rank vector with equal initial ranks

v1<- as.matrix(rep(1/58, times = 58))

#Now, we proceed with iteration

#until similar successive rank vectors are obtained

k=1

for (k in 1:100) {

v2= Link\_Matrix %\*% v1

if(v2==v1){

break(0)

}

v1=v2

}

r\_v<- v2

sum(r\_v)

r\_v<-as.matrix(r\_v) #Final Rank Matrix

k #Number of Iterations required

**Outcome**



On applying the R-codes shown in the previous section, we get the ‘Rank Vector.’

The obtained rank vector is sorted in descending order of probabilities and top 15 investors according to our Page Rank Algorithm are shown above.

Now, putting the above output as the ranks, we have,

1. Accel India
2. Sequoia Capital India
3. Jungle Ventures
4. Lightspeed Ventures
5. General Catalyst Partner
6. Temasek
7. SeedPlus
8. Y Combinator
9. Venture Highway
10. Digital Garage
11. DST Global
12. Tiger Global Management
13. SAIF Partners
14. Norwest
15. Nexus Venture Partners

We also note that, in our ranking for investors in Indian-Unicorn Startups, Accel India ranks above Sequoia Capital India. However, in the ranking by Chris Farmer, Sequoia Capital ranks over Accel India. So, we can say that Accel India is much more influential around the Indian Startups as compared to Sequoia Capital, which is more influential overall.

**Conclusion**

From the final Investor ranking obtained using the Page Rank Algorithm, we conclude that,

1. Accel India is the most influential investment firm around the Indian Start-up scenario, that has been a part of the recent Indian Unicorn Start-ups.
2. The second most influential investment firm around the Indian Start-up scenario is Sequoia Capital India.
3. To define the most influential investment firms, we cannot just look at the Estimated Annual revenue for a particular investment firm. Networking is an equally important criterion to determine the success.

**Future scope of the work:**

This study can further be used to rank the investors based on particular sector of start-ups. This may prove to be useful for start-ups to choose the investor and pitch accordingly to convince the investor to invest in their start-ups.

We can also continue the investor ranking using the Weighted Page Rank algorithm to compare the method with the conventional Page Rank algorithm used in this project.

**References**:

1) <http://crunchbase.com>.

2) <https://www.ventureintelligence.com/Indian-Unicorn-Tracker.php>.

3) <https://www.geeksforgeeks.org/weighted-pagerank-algorithm>.

4)<https://www.geeksforgeeks.org/hyperlink-induced-topic-search-hits-algorithm-using-networxx-module-python>.

5) <https://ra.ethz.ch/CDstore/www2005-ws/workshop/wf10/fujimura.pdf>.

6) <https://www.upcounsel.com/types-of-investors>.

7) [2020 Series A, B, C Funding Guide: Averages, Investors, Valuations & How to Get Funding](https://www.fundz.net/what-is-series-a-funding-series-b-funding-and-more#intro).

8) <https://en.wikipedia.org/wiki/PageRank>.

9)<https://techcrunch.com/2011/05/25/top-10-vc-firms-investorrank>.

10) <https://appliedmath.ucmerced.edu/sites/appliedmath.ucmerced.edu/files/documents/PHDdocs/bryansims_msthesis.pdf>.

11) <https://www.cbinsights.com/research/investor-google-pagerank>.

12) <https://www.researchgate.net/publication/315493458_Stock_portfolio_selection_using_learning-to-rank_algorithms_with_news_sentiment>.

13) <https://www.ventureintelligence.com/Indian-Unicorn-Tracker.php>.

14) <https://app.dealroom.co/companies/>.

15) <https://www.gabormelli.com/RKB/HITS_Algorithm>