Aim and Objectives

Aim

My main aim was to explore all the information within the IBM's Twitter account which will give me a clear understanding of the current business situation through this specific social network.

Reason

Analysis of entire Twitter account will allow me to better understand all the connections and information from the dataset. This might be very useful analysis which will allow future forecasting. I will be able to identify the areas which are the most popular as well as trends which are currently within the internal organization.

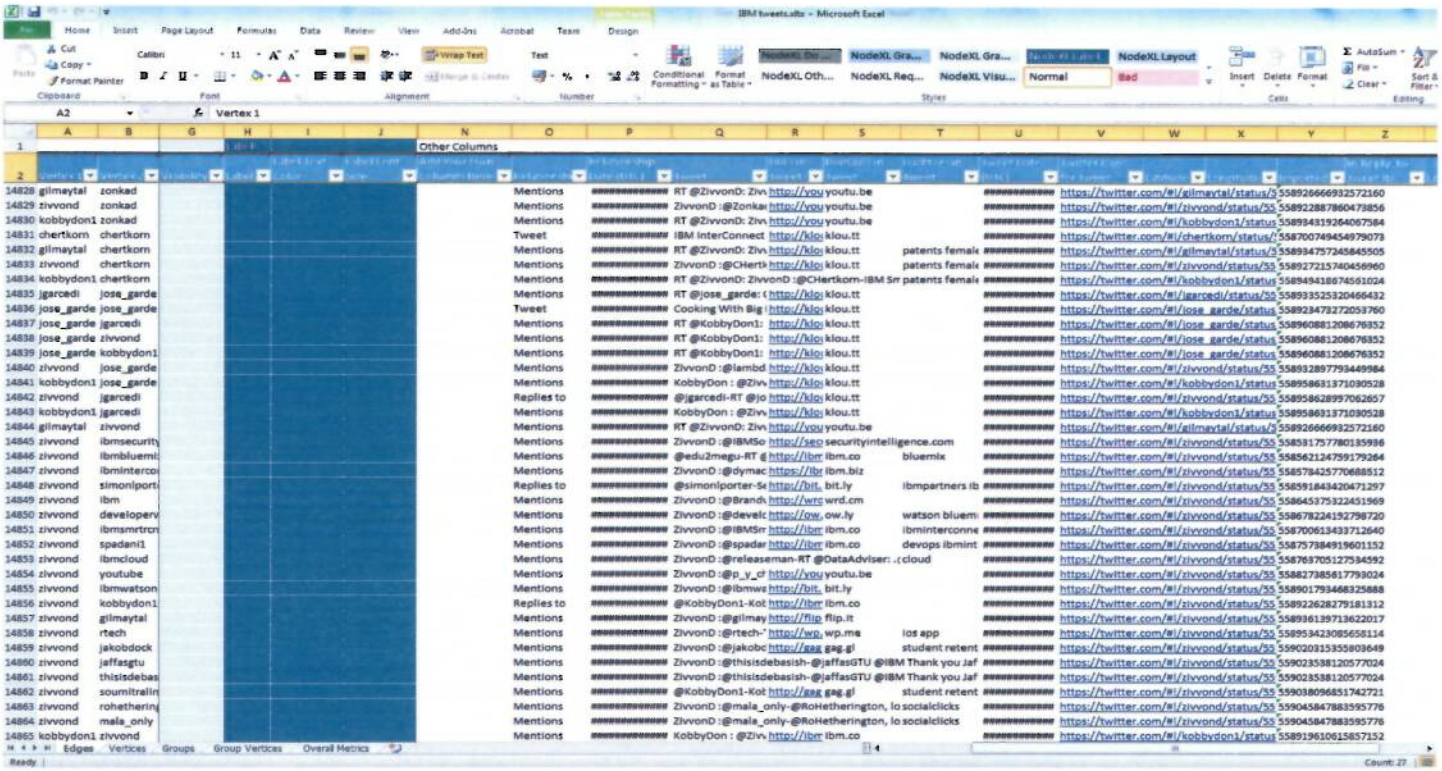
Objectives:

- Collect the data from IBM's Twitter account
- Create a dataset by using the NodeXL (the Microsoft Excel extension)
- Analyse the data
- Visualisation of the progress
- Results and finding
- Future development

Acquiring the Data

For the purpose of collecting the data from Twitter, I have used NodeXL program. This is an extension for Microsoft Excel with an additional functions and columns. NodeXL has a build-in function to gather the data from various sources. I have used Twitter because there was a huge amount of information included about the company which I wanted to explore. As written in the introduction of the document I have chosen the IBM Corporation.
I have gathered more than 20 thousands of various IBM tweets which created for me more than 10 thousands of nodes with more than 15 thousands of edges.

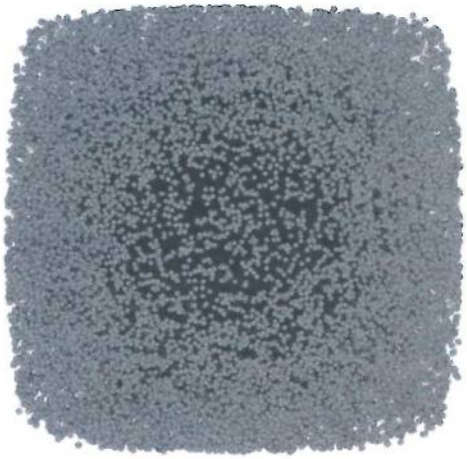
In the following screenshot, you will be able to actually see the Microsoft Excel (with an extension for NodeXL program). There is a focus on edges; however there are also next sets of data such as groups, vertices or vertices groups.



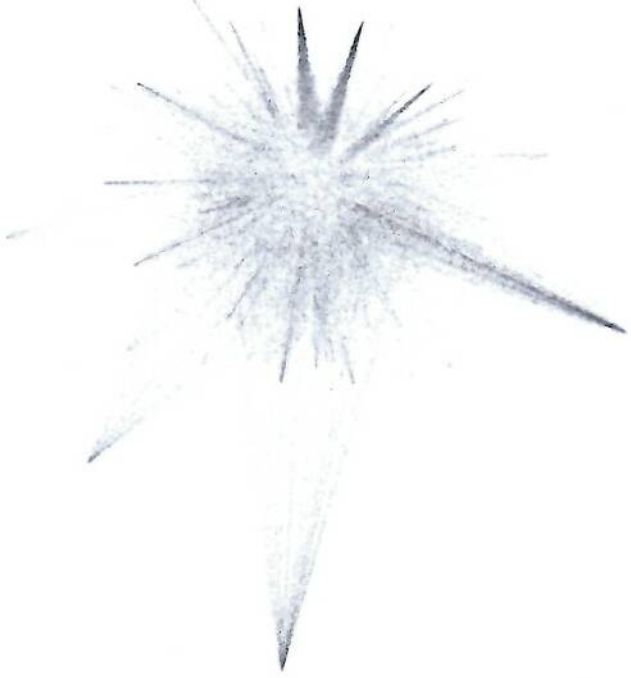
Big Data work

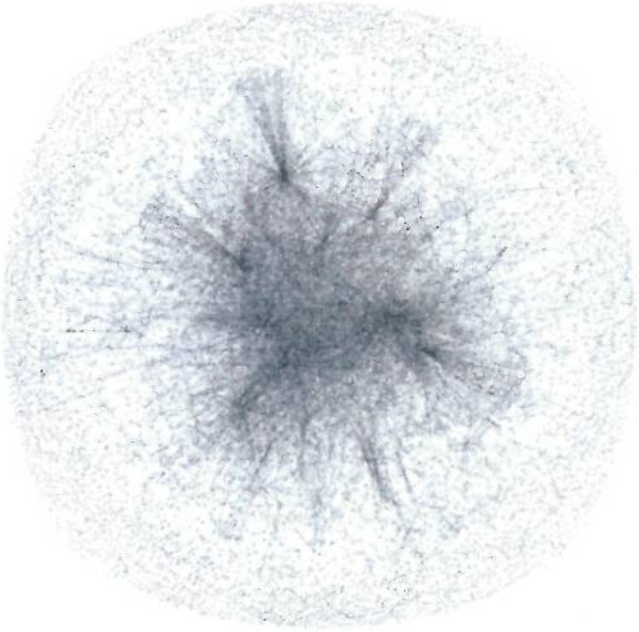
I have used the dataset provided in the previous page of the document to show them in different and more effective ways.

In total, there is 10 869 of nodes which are represented by the dots. Every single dot contains a lot of information about the tweet which has been sent or posted on Twitter account.

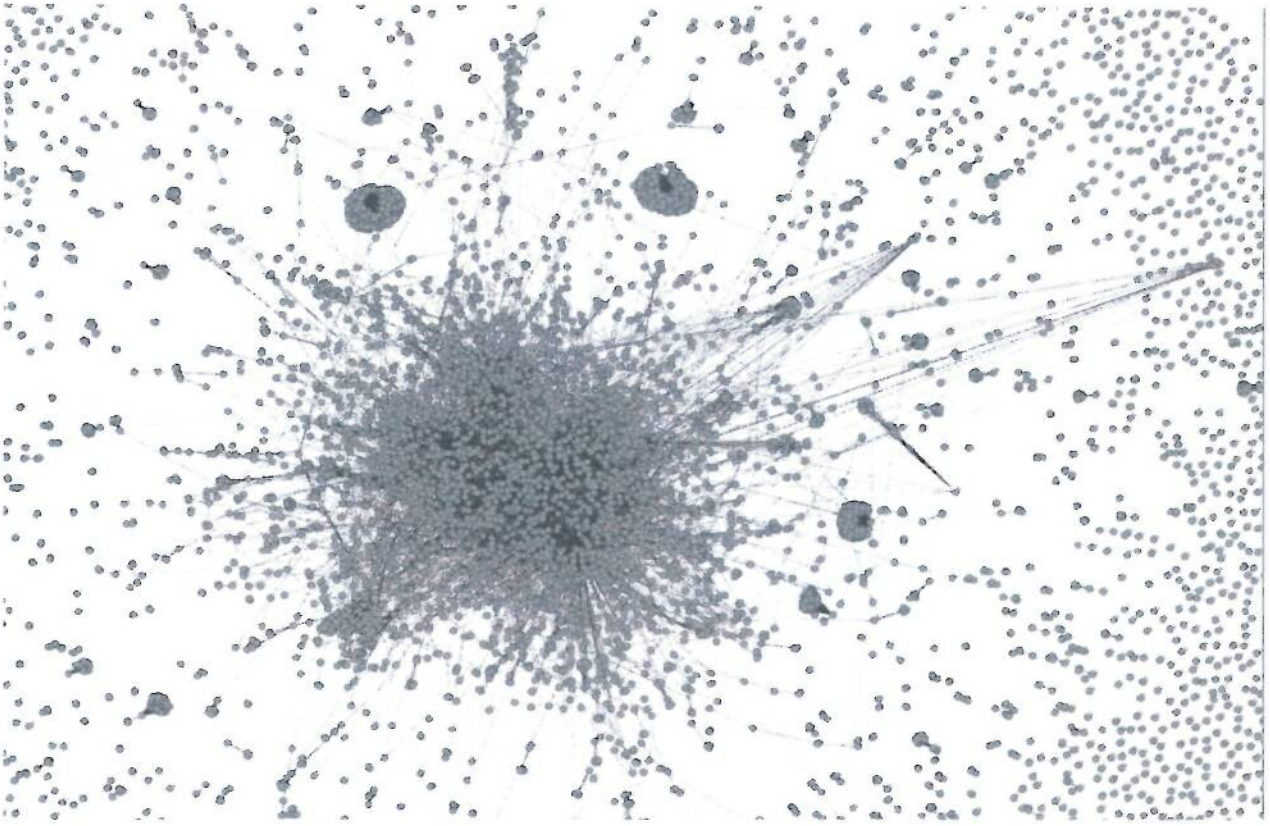


However, the most important for me were connections, the edges. In the following screenshots, you can see clearly that the nodes are connected with each other. In total, there is more than 15 thousands of connection and each node is usually connected with at least one line of the connection. The dataset gained from Twitter has provided me with the "vertices" data as well. In the vertices section, there were described all of the connections within the dataset.

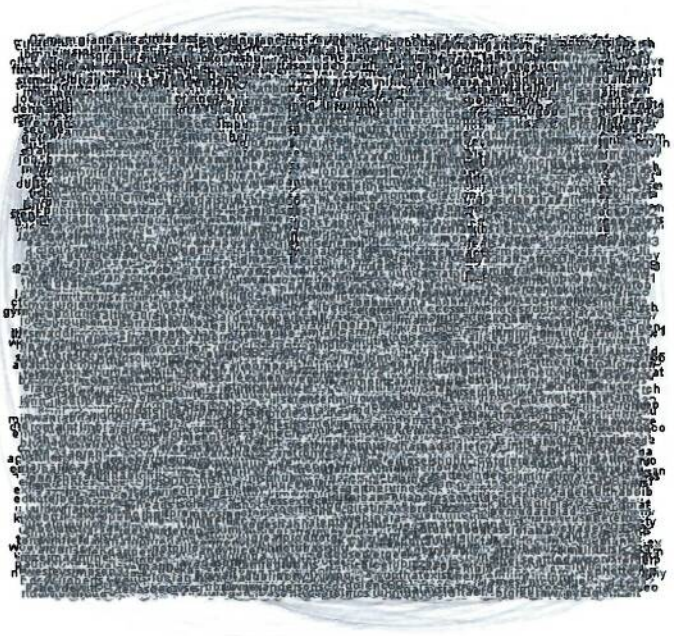




The screenshot below is a closer look at the centre point of the connections. In here, you will be able to see more clearly that there are really a lot of relations. However, there is not always just one connection to each node. There are usually many of connections within the certain parts of the dataset. For me, those central parts were the most interesting as well as important in terms of dataset research.



Additionally, yes, really each of the nodes contain many of the information.

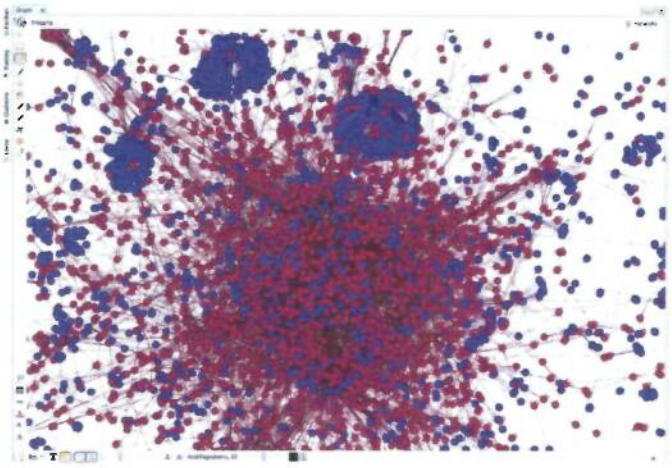


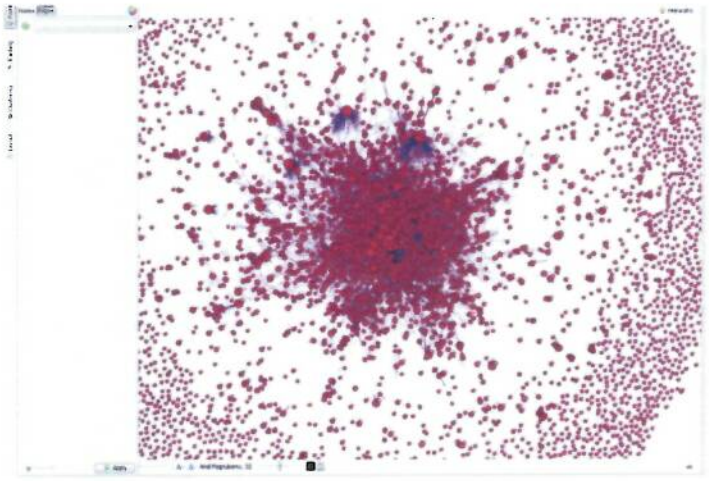
Analysis with Visualisation

The aim of the analysis was to overview entire situation of dataset, to focus popular data information, popular areas of interest. All the outcomes from the analysis would be useful to be able to predict which of the areas the IBM Corporation should generally focus, to be even more popular.

Just to look into the dataset and visualise huge dataset wasn't enough for me. I wanted to more understand the current IBM situation through the IBM's twitter account. Therefore, I started to **explore** through various analysis. Those analyses allow me to better focus really important nodes. Better understanding of the most popular nodes will help me to recognize the most common areas of interest as well.

I started with running "**modularity**" analysis which gives me clear understanding about which data are connected together and how they interact in terms of the various areas. Therefore, this brings me to the idea to determine which areas are the most common and what they actually mean.

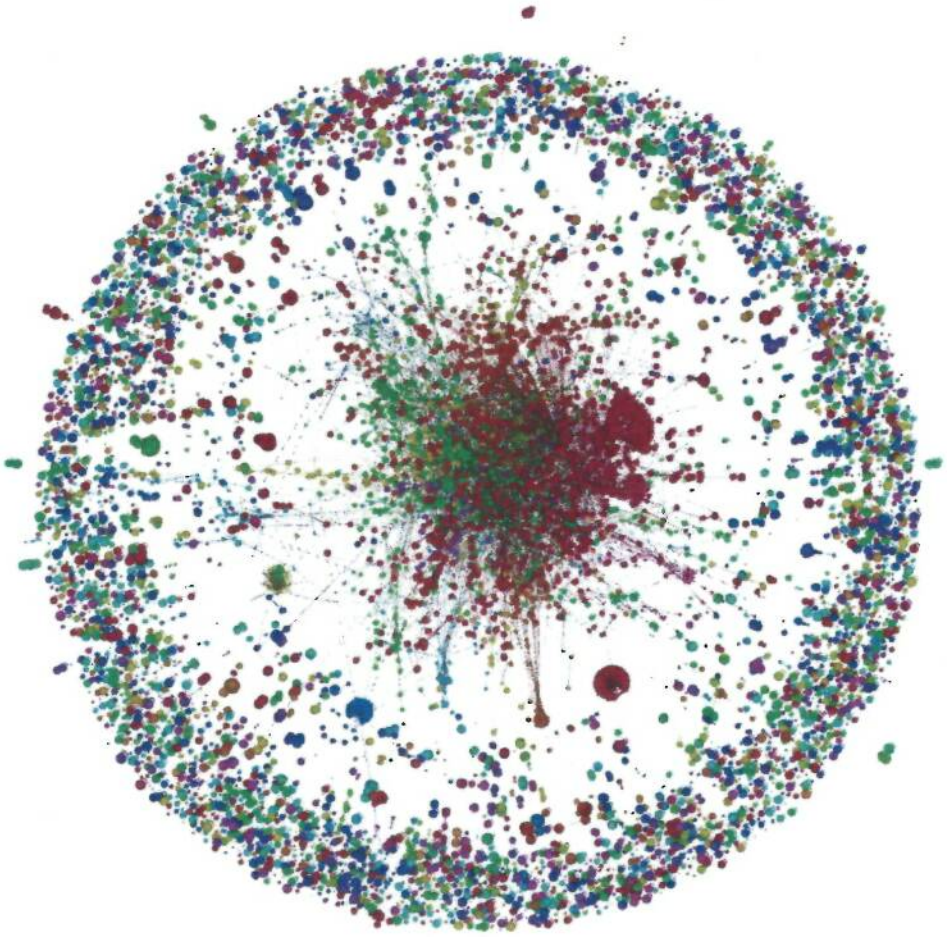




Nodes with the most connections were vital for me as this was clear sign that the information contained within the node will make a huge effect on the overall dataset. Therefore, I run the "**average degree**" analysis as well with the modularity degree. This provides me with such an important numbers of relations (edges).

The bigger and brighter node (dot), in this case means, that it has got a lot of connections/relations with other nodes.

After finishing all of the previous activities, I was able to determine the areas which took place in the IBM's twitter account. The final outcome looks like colourful eye:



What changed here is mainly the position of the nodes (created manually) as well as colours. I have defined the primary colours such as red, green and blue to make the overall outcome clearer and more effective.

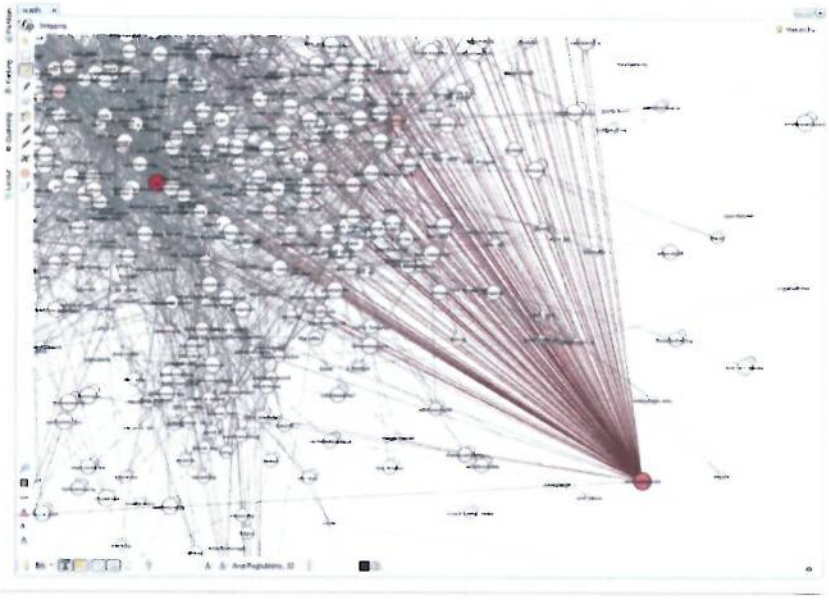
However, the most important question was what actually the colours mean.

The Gephi program is not just able to visualise the current project, but it is also able to provide great information based on the modifications made. After a small research of the information provided, I am certain about the 3 main areas within the dataset.

The red colour represents **advertisement** being posted by IBM team or any different person, but the topic of the post was always related to the IBM subject. Following by green colour which is primarily the news about **recently released technologies** by the IBM. Lastly, blue which were the news about the **IBM events** taking place all around the world.

What then interest me mostly was: "Which node has actually the most relations and is the most popular?".

I have got already all information needed to find it; therefore, I started my searching. Actually, I have found a few which were very popular. The most interesting was, that they have been connected to each other as well.

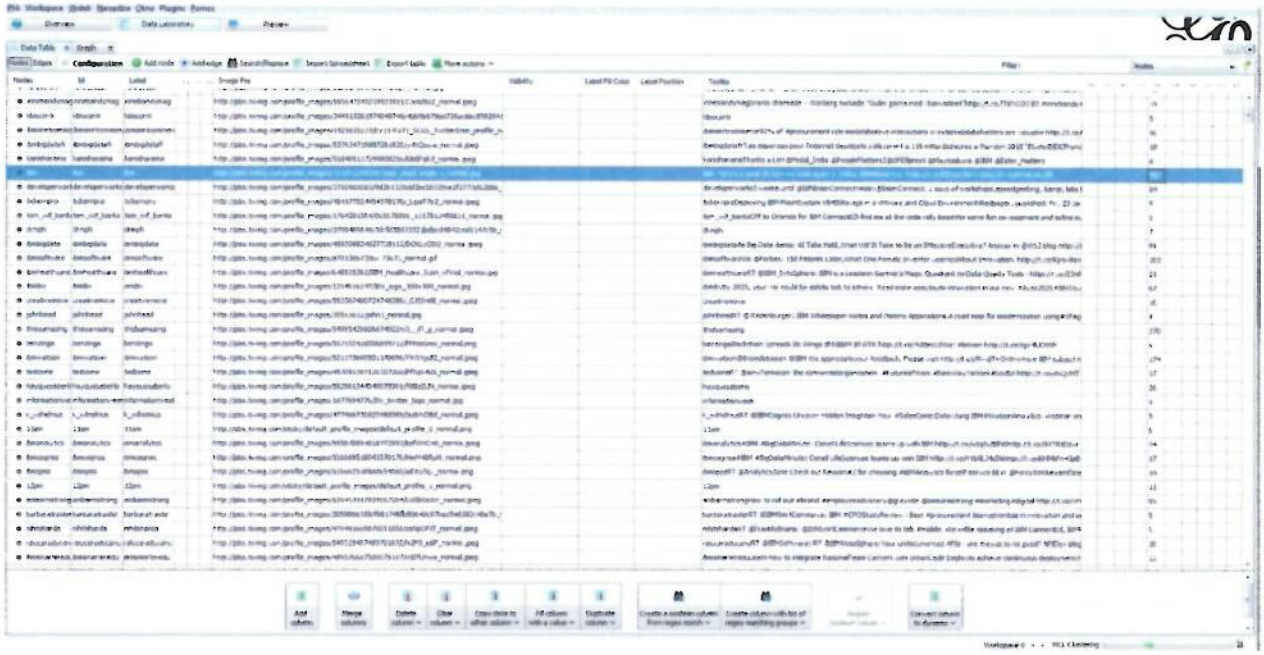


However, one I was looking for was released by "releaseteamltd" account; which I have found out later on, is just the **IBM staff team** who is responsible for posting newest news or interest IBM's activities.

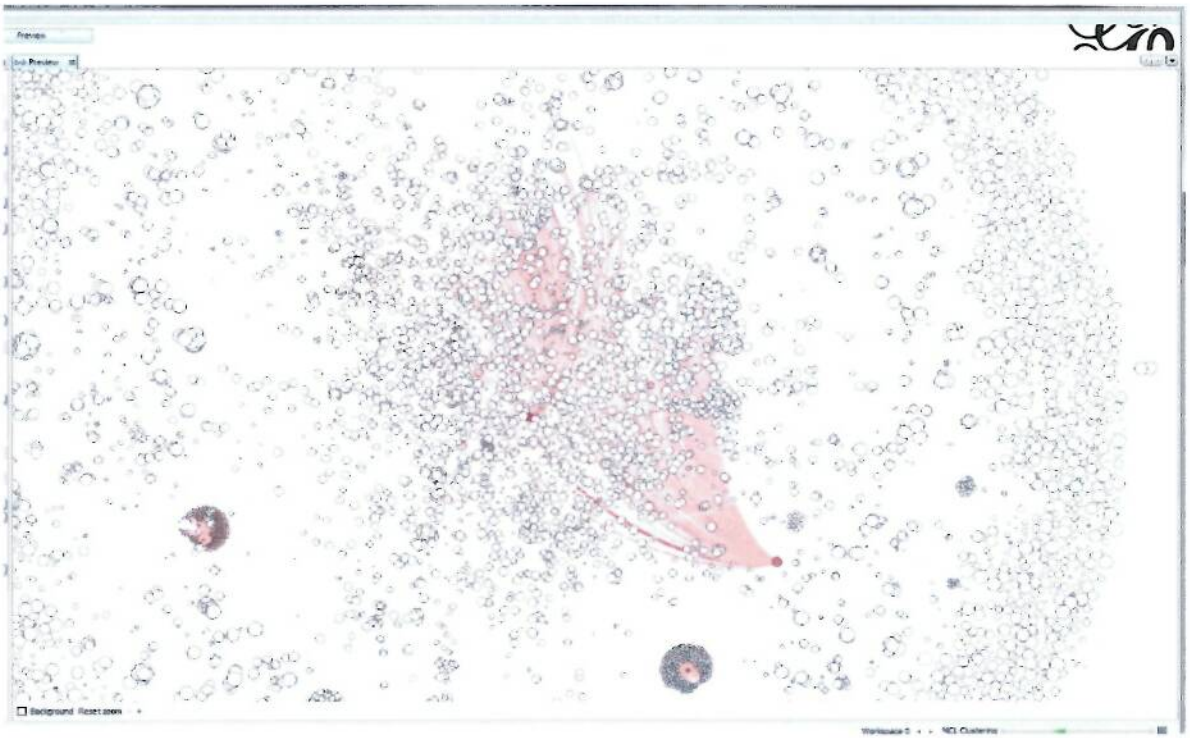
Normal look:

Zoom to find the name of the node (username):

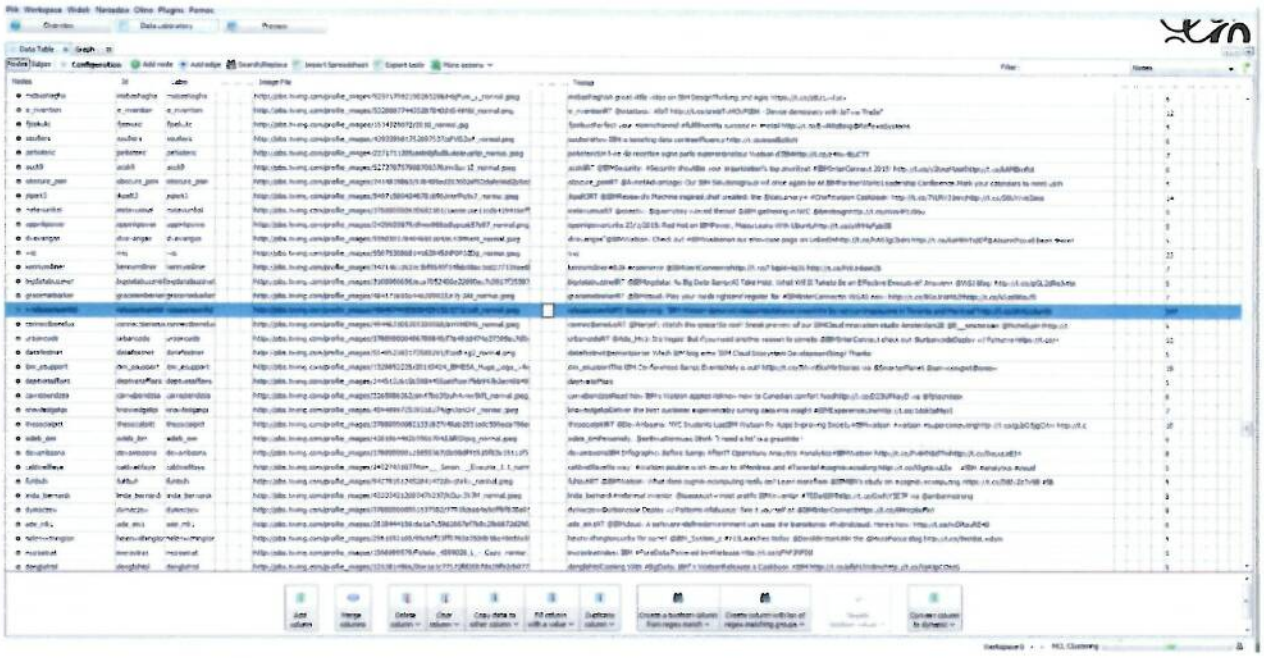
Exact information has been provided by the next part of the analysis within the different area of the Gephi program. In here, you would be able to see some information about the most famous tweet which has been posted to the IBM's Twitter account. It has got exactly 487 of relations and it has got very positive impact on overall users' satisfaction.



As mentioned before, there was more of the popular nodes which have influenced the dataset. Please notice, that there is a connection between them:



The second most famous Tweet was one with 347 relations also posed by "releasedteamltd" user. That explain the connection created between those 2 nodes. It was posted by the same user, however, with completely different information. This one was in the green area (recently being released technologies) providing information about the IBM Watson.

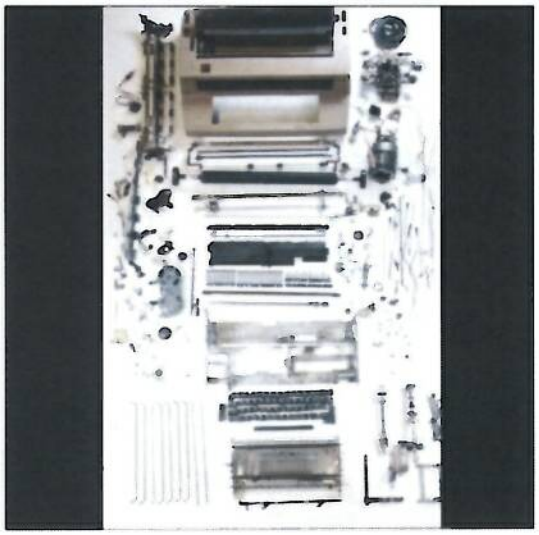


Additional Results and Findings

Key the findings of the project say, that most popular on Twitter are not the issues related to the company; however the biggest volume are the news related to technology and the new technologies that the IBM is releasing.

However, as always, creativity won within the IBM Corporation. From my findings the most popular and the most interesting was with the following video (one with the 487 connections released by IBM team):

Start of the video:



End of the video:

Moreover, as described in previous stages, a big part of the IBM interest comes also with innovations such as IBM Watson. IBM Watson Analytics is the most popular topic within that area.

Additionally, there was also the huge interest in the recent IBM's activities. There was huge attention on the news about the new machine Z13.

Conclusion

By researching and analysing the huge dataset, I was able to determine certain areas which should IBM focus in the future in accordance to satisfy IBM users. The most famous Tweets can also make some direction which IBM release team should consider. All those key points may be used for the future forecasting according to which specific products or technology will many users like.

In my point of view, it's very important to keep track of social sites like Twitter as this might bring some valuable information for the business. IBM is a huge business with many of users, and this is certainly the way how they would be able to find something interesting in terms of satisfactory of current users or potential customers.

Future Development

According to future development, this kind of analysis could be more expanded to give more specific numbers. For instance, I might be able to find exact number of customers who like a certain product or are very interested in a particular area of the IBM. All those information might be very valuable for any big company which is focusing their customers and wish to do their best to satisfy customer needs.

The techniques used by me doesn't need to be used just for the business. This approach might be beneficial in many other areas such as weather forecasting or to analyse certain patterns of natural disasters such as various eruptions or tornados. It might be used for security reasons as well; for mapping various people activities all around the world.

References

Gephi.org. (2015). *Official Releases.* [Online] Available from: http://gephi.github.io/users/download/ . [Accessed 24th January 2015].

Gephi.org. (2015). *Learn how to use Gephi.* [Online] Available from: http://gephi.github.io/users/ . [Accessed 15th February 2015].

IBM. (2015). *We took apart a 1980s #IBMSelectric. Here's what happened.* [Online] Available from: https://instagram.com/p/yNhFxdvhlp/ . [Accessed 19th February 2015].

Koumin. (2015). *Let's Play Gephi: Understand Degree, Weighted Degree & Betweeness centrality.* [Online] Available from: http://matthieu-totet.fr/Koumin/2013/12/16/understand-degree-weighted-degree-betweeness-centrality/ . [Accessed 15th March 2015].

Brittany Kubinski. (2013). *How to Video Social Network Analysis using Gephi and NodeXL.* [Online] Available from: https://www.youtube.com/watch?v=hN3-wTOxrsY . [Accessed 30th January 2015].

Vincent Finat-Duclos. (2015). *A step-by-step introduction tutorial to Gephi.* [Online] Available from: http://www.obviousleaks.org/step-by-step-gephi-tutorial/ . [Accessed 10th February].

IBM. (2015). *Official IBM Twitter account.* [Online] Available: https://twitter.com/IBM . [Accessed 20th March 2015].