

CSE 522 Visualization Final Project Report

Marvel Comic Universe Data Visualization

Chinmaya Dehury-112077747

cdehury@cs.stonybrook.edu

1.Introduction

Being an ardent fan of comic books and movies I am always curious to know facts and trivia regarding them. So, I chose to present visualization on such a comic book universe. Marvel Universe is a fictional universe which is based on the American comic stories published by Marvel Entertainment. It consists of hundreds of separate universes containing thousands of superheroes. Through this project my aim was to expand our knowledge beyond the Marvel cinematic universe and develop a new perspective for the Marvel universe.

2.Dataset

A quick search on Kaggle returned over many results on Marvel comic book universe. My idea was to create a dashboard of superheroes. So that I could show a brief bio of the superhero and some interesting statistics on the superhero. Unfortunately not many of the datasets had this kind of information. Finally I chose the following data

1. The Marvel Universe Social Network: The dataset contains heroes and comics as nodes. And edges occur between heroes when they occur in the same comic. This was the primary data source for plotting the social network graph.
2. Marvel Superheroes: This data set contained stats and info on each hero in the marvel universe. I wanted this so that I could plot the stats using barchart or pie chart.
3. I also wanted a brief bio on the hero. "Superheroapi" website had just the right info I was looking for. So it became my third source of data.
4. I also wanted to show a picture of the superhero I wanted to show data about. Unfortunately "superheroapi" didn't have the images I was looking for so I had to take help of "superherodb" another website that provided data about superheroes.

3.Data Source

1. <https://www.kaggle.com/csanhueza/the-marvel-universe-social-network>
2. https://www.kaggle.com/danniellr/marvel-superheroes#marvel_dc_characters.csv

3. <https://www.superherodb.com/characters/>
4. <https://superheroapi.com>

3. Project Description

3.1 Technology stack

I used D3 for visualization, Python for data analysis and data collection(through crawling). The front end was made using HTML, css and vanilla javascript.

3.2 Implementation:

1. Data Collection:

I downloaded the social network data from Kaggle. Most important was hero-network csv file that contained all the edges of the social network.

Next I downloaded the superhero data from Kaggle it contained a lot of csv regarding superhero from different publishers. But most importantly it contained superhero stats that I wanted.

But I still needed the biography that I wanted to display in my dashboard. So I crawled the superheroapi website that provided me with the bio I needed.

Next I wanted images of the hero to be displayed along with the info. For this I crawled the superherodb website which got me the images I needed.

Following is a snippet from getimages.py file used for downloading the images:

```
details_file = open("superhero_details.txt")
details_json = json.loads(details_file.read())

for hero in details_json.keys():
    details = details_json[hero]
    if "image" in details and "url" in details["image"]:
        image_url = details["image"]["url"]
        print("fetching: " + image_url)
        os.system("wget -O images/" + hero + ".png" + image_url)
```

2. Data Cleaning and Transformation:

A lot of the data had missing attributes or undefined attributes. For example a lot of minor characters had power statistics all equal to zero. I chose to remove such elements using python.

The hero network file contained edges between superheroes that were present in a comic. But that means there were a lot of repeated edges. Since some superheroes were together in many comics. So I removed the repeated edges and instead gave weight to the edge based on the number of times it is repeated.

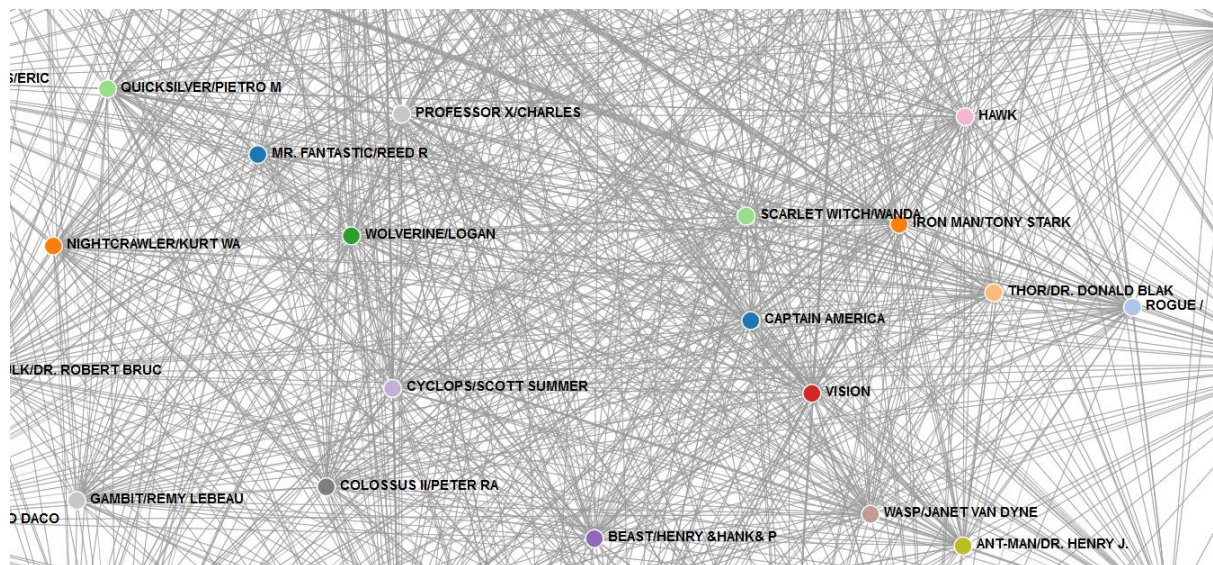
2. Data Aggregation:

Now that I had all the data I needed to join them using some method. Since I had data from 3 different sources. Joining them took a lot of effort since there was no unique ID. The hero names though unique had a lot of spelling mismatch in between the datasets. So I wrote a python script to map a superhero names from one dataset to the other dataset. I stored the map in a csv to be used later in d3.

3. Marvel Social Network:

I used d3 force directed graph to plot the Marvel Social Network. The length of the edges between nodes was set inversely proportional to the edge weight. This resulted in the nodes that were highly connected to come together. And we could see the clusters.

The plot also uses highlighting feature we learnt in the previous mini projects.



Screenshot of the social network graph

I plotted the social network of top 50 superheroes. My dataset contained some 600 heroes but plotting them together in a network was making the plot not readable.

Code snippet from d3 program to plot the force directed layout

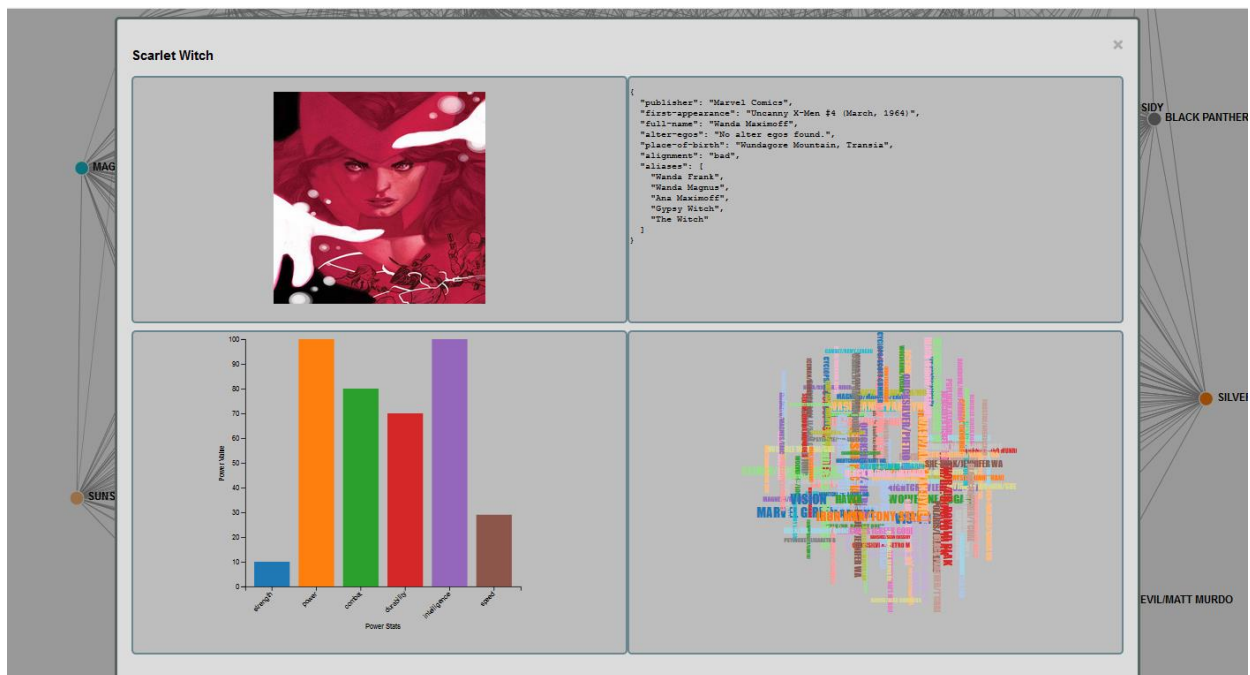
```
var circles = node.append("circle")
    .attr("r", 10)
    .on("mouseover", function(d) {
        d3.select(this).attr("r", 15);
    })
    .on("mouseout", function(){
        d3.select(this).attr("r", 10);
    })
    .style("fill", (d, i) => fill(i))
    .call(d3.drag()
        .on("start", dragstarted)
        .on("drag", dragged)
        .on("end", dragended));

var labels = node.append("text")
    .text(function(d) {
        return d.id;
    })
    .style("font-size", "15px")
    .style("font-weight", "bold")
    .attr('x', 15)
    .attr('y', 3);
```

4. Super Hero Dashboard:

My dashboard consisted of hero image, its bio, a barchart of its power statistics and wordcloud containing the nodes it is connected with. For barchart I used the d3 code I had written in the previous assignment. Word cloud was also displayed using d3. It shows heroes which were highly connected to the chosen hero in larger fonts.

The dashboard is linked to the social network graph which acts as point of navigation. Clicking on a superhero in the social network pops out its dashboard on a modal window.



Screenshot showing dashboard for Scarlet Witch

Below is code snippet for generating the word cloud:

```
var layout = d3.layout.cloud()
  .size([400, 400])
  .words(data)
  .font("Impact")
  .padding(5)
  .rotate(function() { return ~~(Math.random() * 2) * 90; })
  .on("end", draw);

layout.start();

function draw(words) {
  d3.select("#demo1")
    .append("g")
    .attr("transform", "translate(" + layout.size()[0] / 2 + "," + layout.size()[1] / 2 + ")")
    .selectAll("text")
    .data(words)
    .enter()
    .append("text")
    .text((d) => d.text)
    .style("font-size", (d) => (d.size + 5) + "px")
    .style("font-family", (d) => d.font)
    .style("fill", (d, i) => fill(i))
    .attr("text-anchor", "middle")
    .attr("transform", (d) => "translate(" + [d.x, d.y] + ")rotate(" + d.rotate + ")");
}
```

4. Important Observations

Through this project I learned a lot about super hero characters and their relations. Especially amongst the characters that dont fall into well known groups. For example I learnt that

Ice Man a mutant from Xmen is very close to the Fantastic Four Group. Also I got to answer age old questions like who is the strongest Avenger. By comparing the power statistic bar chart of all Avengers it is clear that Thor is the strongest one!

From the social network graph we also learnt which characters are most central to the Marvel universe. For example we can see that all the Avengers are quite important to the Marvel universe and sit in the middle of the social network.

5. Future Work:

Since the social network is highly connected it was sometimes hard to see the edges and to which they are connected. One idea would be to reduce the density of the edges by edge contraction algorithms.

Providing slider for the word cloud to filter the words to show words above a certain threshold.

Making the word cloud clickable and linking the word to its hero dashboard.

6. Related Work:

1. <https://studentwork.prattsi.org/infovis/projects/marvel-universe-visualization/>
2. <http://graphics.straitstimes.com/STI/STIMEDIA/Interactives/2018/04/marvel-cinematic-universe-whos-who-interactive/index.html>
3. <https://www.superherodb.com/>