ISOM 673 - Social Networking Analysis Final Project:
The Marvel Universe Social Network

Rush Bhardwaj, Yi Cai, Meng Wang, Nicholas Yang

INTRODUCTION

In the mid 2017s, the Marvel Cinematic universe (MCU) topped a 12-billion-dollar milestone in box office revenues[6] and recently released its trailer for Avengers Endgame which garnered 48.7 million views on YouTube[7] in a day of release. It is a passion for many comic book fans and a weekend's entertainment for the general public. Our final project is motivated by the inner geek to explore the universe of the marvel heroes and quantify our pre-existing notions about the universe.

Our research question was motivated from our interests and we wanted to analyze the role and influence of each hero in the network and how heroes drive strategic decision making for the studio. In upcoming sections, we will be quantifying the power of the network and see the influence of heroes through differentiation.

Marvel Universe has little over 6,400 heroes (nodes) in the comic world and close to 150 characters in the cinematic universe. This in turns creates a dense network of heroes and there are complexities added to understand the influence of each hero. To begin our analysis, we started with the assumption: heroes have a connection between each other if they appear in the same comic book or same movie. Our first step was to take a look at the properties and characteristics of the marvel network by calculating some network dimensions. After the initial analysis, we created network visualization to analyze the ties of the famous Marvel heroes' network and for that we used Gephi, a visualization and exploration software. Our analysis also includes multiple regression analysis and exponential random graph modeling (ERGM).

DATA DESCRIPTION

The process of collecting, cleaning, filtering and planning data was critical for the analysis of the network. Data was obtained from the given source file on canvas which included the network between all the heroes (574,467 ties) in the marvel universe and the comic they have been a part of. To look into the MCU, we decided to mine data on the web and merge with pre-obtained data for our analysis.

The first dataset collected was the MCU cast[1] from the IMDb public database. This dataset included the name of the actor, the hero played in the cinematic universe and the movie they were a part of. We were then able to merge the MCU cast to the comic network by merging on the hero name attribute. This step of merging came with its challenges. The data obtained from canvas, a hero had many different forms of unique names. For example: Hulk had 11 matching names and, in the case, to find Thaddeus Ross, the name in the universe for the character was "THUNDERBOLT/BILLY CA". Rather than learning new techniques to find similarities, we decided to go ahead and manually make a key of the heroes' names in the MCU and merge and find the cinematic universe in the comic hero-network. We also added the network of GROOT (57 ties) in the dataset by creating a network from the cinematic universe as we could not find GROOT in the comic universe. This helped us map the network and analyze the role of individual (heroes) and their influence.

We were also able to obtain the MCU box office earnings (corrected to inflation) from "Box Office Mojo"[2]. We then created a dataset of the 20 movies and added their box office earnings, release date, ratings[3], phase information[4], producers[5] and awards[4] information. Throughout our analysis we will be using the original and newly created datasets to understand the network and help improve our regression models.

RESEARCH METHODOLGY

DESCRIPTIVE STATISTICS

For the understanding of the Marvel universe, we decided to construct two undirected networks to explore the social network of the heroes. The first is the Marvel comic network, which includes 6,426 unique comic heroes and 574,467 edges. The other network is the MCU network, which includes heroes that have appeared in movie series only. As mentioned in the introduction, the MCU network includes 103 heroes and 224,181 edges. For both the networks our assumptions were: the edge exists if two heroes appeared in the MCU network as well as the comic network. Our primary focus was on the MCU network, but we included the comic universe in our analysis pick up valuable insights of the Marvel universe network. By adopting this methodology, we were able to compare results from the two networks and provide recommendations to the stakeholders and producers about the future of the MCU.

The marvel comic network is one of the most unique networks. It is constructed by interactions of fictional characters but also at the same time, it is very complex and humongous. To understand the network complexities, it is justified to provide some network dimensions and descriptive statistics. The MCU has a network diameter (the largest distance between any two nodes in the network)[9], of 4 units/steps. The average degree (how connected the node is)[9] is 41. Clustering coefficient[9] is 0.55, showing that the two heroes that have collaborated with another hero, are much more likely to collaborate in the future than a randomly chosen pair. The average path length (the measure of characteristic path length completed over geodesics or the shortest path)[9] is 1.90, thus, any pair of heroes can be connected through an average of 2 collaborations. The network has a relatively high Modularity[9] which is 0.12, showing the dense connections between the heroes. The graph density (the relative fraction of edges that are present over total edges minus one)[9] is 0.006. This indicates the comic network is quite sparse and several heroes acquire most of the edge connections. Most of the comic heroes are peripheral nodes and have only a few appearances. Compared to the MCU, the comic universe is much sparser. Table 1 highlights the differentiation of the two network dimensions. The graph density and the cluster coefficient of the comic network are 0.0003 and 0.19 respectively and are smaller compared to the MCU network. The mean distance of the comic network and the modularity are 2.82 and 0.36 respectively. From the table 1, we can demonstrate the sparseness of the two networks. More precisely, the comic universe is sparser than the MCU network.

	Graph density	Clustering Coefficient	Network Diameter	Mean Distance	Modularity	Average Degree
MCU	0.006	0.55	4	1.9	0.12	40.88
Comic network	0.0003	0.19	8	2.82	0.36	69.77

Table 1 – Marvel cinematic universe and comic universe network descriptive statistics

Moving forward, we reviewed the individual influencers of the MCU. Our review involved a detailed analysis of specific nodes to garner some powerful insights. In terms of degree centrality (how well a node is connected in terms of direct connection)[9], Captain America, Spider Man, and Iron Man are the top three heroes. This is proven by the fact that they are the oldest and most successful heroes. Betweenness (how well situated a node is in terms of the paths that it lies on)[9], a type of centrality measure, can also be defined as the number of times of a node passes through the shortest distance between any two nodes in the network. In terms of betweenness centrality, Wolverine plays the most important role in conveying collaborations (information) to other hero in the network who are not connected with each other. In the MCU, Wolverine acquires the control on the collaborations between different clusters since he is the main connection between the X-men and Avengers. Closeness centrality (how close a given node is to any other)[9] represents the ability of spreading information to the whole network within a shortest time frame. Mister Fantastic is the character who has the ability to reach all other hero as quick as possible, since he is the leader of Fantastic Four and the center of the comic universe. Page rank algorithm, (first used by Google) was utilized to calculate the relative score of a webpage's authority and importance level. Unsurprisingly, Spiderman has the highest page rank centrality since he has the direct connections to other influential heroes such as Iron Man and Captain America. We also notice that all the connected heroes to Spiderman have relatively higher page rank score.

VISUALIZAZTION PROCESS

In order to gain a further understanding regarding both of the networks: the comic network & the MCU, we created network related visualizations with the help of external software - Gephi. With the assumption that edge strength is based on whether two nodes appear in the same comic or movie, we created the following figures.

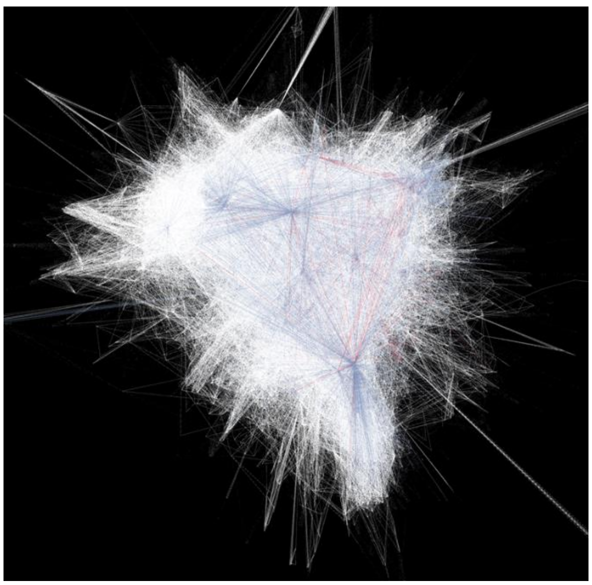


Figure 1 - Marvel comic universe network

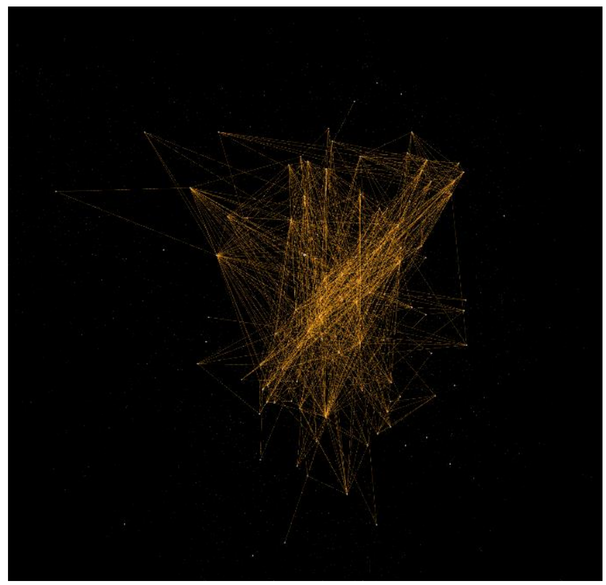


Figure 2 - Marvel cinematic universe network

Figure 1 represents the Marvel comic universe network by ties with different strength. White ties represent weak connections between pure comic heroes, blue ties represent medium connections between comic nodes and movie heroes, and red ties represent strong connections between movie heroes. Figure 2 represents pure MCU network.

Furthermore, on the basis of Figure 1, we created Figure 3 aiming to show distinctions among peripheral and central nodes in the network with the method - Force Atlas, which assigns heavier weights on the with higher coreness scores. Intuitively, such method will contract central nodes to central positions and expand fewer central ones to peripheral positions. This methodology is a visualization related procedure, which will not affect the nature of the original network. Both medium and strong edges are represented by the color orange.

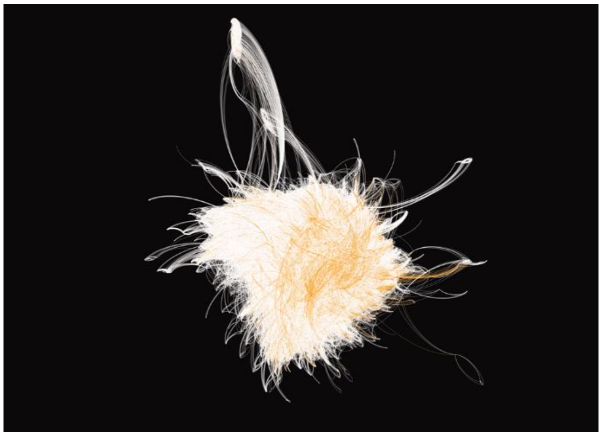


Figure 3 – Marvel comic universe network with Force Atlas adjustment

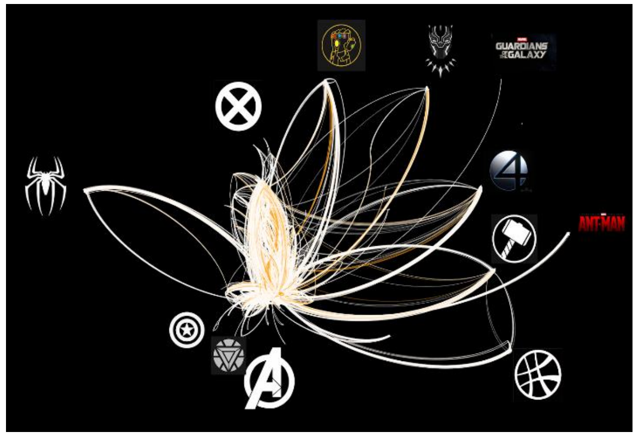


Figure 4 - Clustering with Fruchterman Reingold Algorithm

Finally, with the completion of Figure 3, we constructed Figure 4 for the network clustering realization by applying Fruchterman Reingold Algorithm - a force-directed layout algorithm. In this algorithm, the nodes are represented by steel rings and the edges are springs between them. The attractive force is analogous to the spring force and the repulsive force is the same to the electrical force. The idea was to minimize the energy of the system by moving the nodes and changing the forces between them. Such visualization justified our expectations of the Marvel comic universe as popular comic book series clustered together automatically (More network visualizations can be found in the appendix)[11, 12, 13, 14].

REGRESSION ANALYSIS

As previously mentioned, we created datasets for the analysis of our network. By combining the pre obtained datasets, newly created datasets and the dimensions of the network, we built our dataset for regression called box3_movie, which contains 10 dependent variables- number of movies a hero appears in (movie count), mean IMDb ratings (MN.Rating), mean number of awards (MN.Awards), phase of the movie (Phase), degree (deg), betweenness (betw), closeness (close), page rank (prank), eigen centrality (eigen) and coreness (core) and our target variable - mean box office earning (MN.Earning).

As MN.Earning was a numerical variable, we started our analysis by linear regression. Based on the RSE, adjusted r square and the result from the stepwise attribute selection method, we landed the best performing linear model with 6 variables:

$$lm(MN.Earning \sim MN.Rating + MN.Awards + count + deg + close + eigen, data = box3_{movie}) \quad \ldots (1)$$

To begin with we tried examining the effect of network dimensions on mean earning. Our regression results showed us that neither of the centrality measurement were significant nor the interaction between the measurements. After introducing ratings, awards and movie counts, the significance of closeness and eigen centrality started to appear in our regression. If we tried controlling mean ratings and mean awards, that is, if the quality of the movie is comparable, the hero with higher centrality measurements tends to bring more box office earnings.

Next, we examined the nuances of network effects by analyzing on two narrower network scope: 1) the Marvel movie network and 2) each hero’s connection to one central hero (example: Captain America). By doing so, the significance levels of centrality measurements dropped. A possible explanation for this could be, the heroes that appeared in the movie are most likely to be closer to the center of the network, so while narrowing the scope, the influence of centrality measurement could have weakened.

We further tested our model by incorporating linear panel model by taking movie phases and release dates into consideration:

$$plm\ (MN.Earning \sim MN.Rating + MN.Awards + prank, effect = "time", model = "within", index = "Phase", data = box3\_movie) \quad \ldots (2)$$

In the linear panel model, the effect is “time”, considering the data was collected over time but there is no other dimension that the data run over, so “twoways” is unnecessary. We chose the fixed effect model by selecting model as “within”, with the assumption that there are unique attributes of each hero unvarying across the phases.

EXPONENTIAL RANDOM GRAPH MODELS (ERGM)

With the purpose of understanding the nature of both networks, ERGM models were constructed to examine intrinsic network attributes.

As introduced above, the MCU contains all the Marvel heroes. Hence, the network’s nodes can be assigned with a binary attribute - whether the character was in a movie or not. With the consideration of the computational burden, four terms were selected in the ERGM model - M1:

$$M1:\ heronet \sim edges\ +\ mutual\ +\ nodematch("Movie")\ +\ nodefactor("Movie") \qquad \ldots(3)$$

Edges represent the baseline connectivity of the network; mutual represents the symmetrical ties in the network; nodematch(“Movie”) represents the homophily effect of hero appearing in the movie; nodefactor("Movie") represents the node connectivity comparison between pure comic hero and movie hero.

Parallelly, we constructed model M2 to examine the effect of local structures in the comic network: in M2, “gwesp(0.1, fixed = TRUE)” and “gwdsp(0.1, fixed = TRUE)” replaced “nodematch” and “nodefactor” terms in M1 for computational advantage, where “gwesp(0.1, fixed = TRUE)” (the geometrically-weighed edgewise shared partner) can be treated as a more robust version of “triangle” structure to represent the effect of friend of friend and “gwdsp(0.1, fixed = TRUE)” (the geometrically weighted dyadwise shared partner) represents an “elbow” structure. By statistical definitions, we expect “gwesp” to be non-negative.

$$M2:\ heronet \sim edges\ +\ mutual\ +\ gwesp(0.1, fixed\ =\ TRUE)\ +\ gwdsp(0.1, fixed\ =\ TRUE) \qquad \ldots(4)$$

For the MCU, we constructed different models due to the following reasons:

1. Network size is substantially small, which will cause singularity and converging issues.

2. Computation capability is no longer a constraint.
3. Binary attribute - “Movie” is not feasible as the MCU network is a subgraph of Comic Network based on the “Movie” attribute.

Hence, two ERGM models were created - M3 and M4:

$$M3:\ hero\_movie\_net \sim edges\ +\ mutual\ +\ nodecov("rating")\ +\ nodecov("rewards") \quad \dots(5)$$

$$M4:\ hero\_movie\_net \sim edges\ +\ mutual\ +\ gwesp(0.1, fixed\ =\ TRUE)\ +\ gwdsp(0.1, fixed\ =\ TRUE) \quad \dots(6)$$

In M3, numerical attributes, “rating” and “rewards” were assigned to each movie hero in order to examine main covariate and homophily effects of both attributes. M4 is in the same format with M2 except the dependent variable becomes MCU. In addition to model formulas, ERGM control elements were added with considerations of model accuracy and computational burden:

$$control\ =\ control.ergm(MCMLE.maxit\ =\ 20, MCMC.burnin = 10000, MCMC.interval = 200) \quad \dots(7)$$

RESULTS & INTERPRETATIONS

LINEAR REGRESSION MODEL & LINEAR PANEL MODEL SUMMARY RESULTS

Figure 5 indicates the ratings and awards are the most significant predictors. The degree centrality can improve the model fit but is not significant as a predictor. The eigen centrality measures if the heroes are in the central by evaluating its own connections and its neighbor’s connections collectively. The positive coefficient shows that if a hero is more “centered” and connects with more influential heroes, he/she is more likely to bring in more box office revenue. Yet, the closeness has a negative coefficient in this model. If we looked at the correlation between box office earnings and closeness, the relationship is positive. When interacting with other variables in the model, higher closeness, lower the mean earnings.

```
Call:
lm(formula = MN.Earning ~ MN.Rating + MN.Awards + count + deg +
    close + eigen, data = box3_movie)

Residuals:
       Min         1Q     Median         3Q        Max
-239473575  -39842094    2052230   33552078  163037299

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.154e+09  9.621e+08   1.200   0.2319
MN.Rating    1.854e+08  1.700e+07  10.907   <2e-16 ***
MN.Awards    3.242e+07  2.447e+06  13.246   <2e-16 ***
count        6.498e+06  3.421e+06   1.899   0.0593 .
deg         -5.238e+04  3.709e+04  -1.412   0.1598
close       -3.215e+14  1.314e+14  -2.447   0.0155 *
eigen        2.596e+08  1.240e+08   2.094   0.0378 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 69740000 on 163 degrees of freedom
Multiple R-squared:  0.7269,    Adjusted R-squared:  0.7168
F-statistic:  72.3 on 6 and 163 DF,  p-value: < 2.2e-16
```

Figure 5 - Linear Regression Model summary

```
Oneway (time) effect Within Model

Call:
plm(formula = MN.Earning ~ MN.Rating + MN.Awards + prank, data = box3_movie,
    effect = "time", model = "within", index = "Phase")

Unbalanced Panel: n = 3, T = 33-79, N = 170

Residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.51e+08 -2.50e+07  0.00e+00  0.00e+00  2.60e+07  1.35e+08

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
MN.Rating   204102919   22392643  9.1147 2.382e-14 ***
MN.Awards    29493360    3932386  7.5001 4.804e-11 ***
prank      8155139192 3644439794  2.2377   0.02777 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1.496e+18
Residual Sum of Squares: 4.5465e+17
R-Squared:      0.69608
Adj. R-Squared: 0.41634
F-statistic: 67.1844 on 3 and 88 DF, p-value: < 2.22e-16
```

Figure 6 - Linear Panel Model Summary

Figure 6 shows the significant predictors are mean ratings, mean number of awards, and the page rank. Page rank is a very interactive measurement that considers both a hero's own network and the page rank of the hero's linking with him/her. The positive coefficient indicates that the hero with higher link importance and link to other important heroes leads to higher earning, provided the movie is of good quality. Yet the adjusted r square and the RSE perform much worse than the linear regression model. Both the simple linear regression model and linear panel model show the centrality measures that also examine the influence power of the neighbor nodes tend to have a higher significance level. Although the linear panel model is informative, we decided to not use this model for our analysis.

The above regression results show that the centrality measurements are less influential on the box office earnings than our assumption and is very sensitive to other predictors in the regression. The centrality measurements show a higher significance level given that the measurements on the movie's quality, such as awards and ratings, are taken into consideration.

Different analytical approaches indicate different influencers to the centrality measures on the box office earnings and business outcomes. The overall analysis reveals some interesting findings:

1. Social centrality measurement does not necessarily have a positive coefficient in the model which shows heroes that are closer to the center are not necessarily a positive

impact on a movie's revenue/reputation. A possible explanation is that although the hero with higher centrality tend to be more popular in the Marvel universe and has a big fan base even if the quality of the movie is not good, the fans may feel disappointed and may not want to follow the franchise in the future.

2. The network dimensions that evaluate both the social power of the heroes and the social power of their connections have higher significance level. Eigen centrality and page rank examine the whole network collectively and interactively. In the MCU, the movie that contains several important heroes, generate more box office earnings.

ERGM SUMMARY RESULTS

From figure 7, all four terms gained high statistical significance. "egdes" had a negative coefficient of -12.1379, referring to the Marvel comic universe network is sparser than we expect; "nodematch.Movie" had a positive coefficient, which means that movie heroes are more likely to form a connections with each other than pure comic heroes. Similarly, to represent the effect of covariate, "nodefactor.Movie.1" had a positive coefficient, which means comparing to pure comic heroes, movie heroes were more likely to make connections with others in the whole network.

```
Formula:   heronet ~ edges + mutual + nodematch("Movie") + nodefactor("Movie")

Iterations:  300 out of 300

Monte Carlo MLE Results:
                   Estimate Std. Error MCMC % z value Pr(>|z|)
edges              -12.1379     0.6132      3 -19.794   <1e-04 ***
mutual               7.3074     0.2099      3  34.813   <1e-04 ***
nodematch.Movie      6.1351     0.6188      3   9.914   <1e-04 ***
nodefactor.Movie.1   7.5732     0.5610      4  13.500   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 57236005  on 41287050  degrees of freedom
 Residual Deviance:  2682611  on 41287046  degrees of freedom

AIC: 2682619    BIC: 2682681    (Smaller is better.)
```

Figure 7 - ERGM M1's summary results

```
Formula:   heronet ~ edges + mutual + gwesp(0.1, fixed = TRUE) + gwdsp(0.1,
    fixed = TRUE)

Iterations:  20 out of 20

Monte Carlo MLE Results:
                 Estimate Std. Error MCMC % z value Pr(>|z|)
edges           -9.671240   0.036811     87  -262.7   <1e-04 ***
mutual           5.256899   0.051854     99   101.4   <1e-04 ***
gwesp.fixed.0.1  4.811853   0.031309     82   153.7   <1e-04 ***
gwdsp.fixed.0.1 -0.005383   0.000254     20   -21.2   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 57236005  on 41287050  degrees of freedom
 Residual Deviance:  2778999  on 41287046  degrees of freedom

AIC: 2779007    BIC: 2779069    (Smaller is better.)
```

Figure 8 - ERGM M2's summary results

In figure 8, we concluded the effect of "friend of a friend" exist in the network as "gwesp" term had a positive coefficient. Hence, local structure of triangles was confirmed. There is no scientific tendency of "elbow" structure formed in the network as "gwdsp" had a negative coefficient.

```
Formula:   hero_movie_net ~ edges + mutual + nodecov("rating") + nodecov("rewards")

Iterations:  4 out of 20

Monte Carlo MLE Results:
                 Estimate Std. Error MCMC % z value Pr(>|z|)
edges           -4.487384   0.655270      2  -6.848  < 1e-04 ***
mutual           4.051830   0.105004      1  38.587  < 1e-04 ***
nodecov.rating   0.117667   0.043078      2   2.732  0.00630 **
nodecov.rewards -0.015319   0.005178      2  -2.959  0.00309 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 13724  on 9900  degrees of freedom
 Residual Deviance:  8014  on 9896  degrees of freedom

AIC: 8022    BIC: 8051    (Smaller is better.)
```

Figure 9 - ERGM M3’s summary results

```
Formula:   hero_movie_net ~ edges + mutual + gwesp(0.1, fixed = TRUE) +
    gwdsp(0.1, fixed = TRUE)

Iterations:  20 out of 20

Monte Carlo MLE Results:
                Estimate Std. Error MCMC % z value Pr(>|z|)
edges           -6.53380    0.58535      0 -11.162   <1e-04 ***
mutual           4.07279    0.61646      1   6.607   <1e-04 ***
gwesp.fixed.0.1  3.30079    0.41757      0   7.905   <1e-04 ***
gwdsp.fixed.0.1 -0.25152    0.06158      2  -4.085   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 13724  on 9900  degrees of freedom
 Residual Deviance:  8704  on 9896  degrees of freedom

AIC: 8712    BIC: 8741    (Smaller is better.)
```

Figure 10 - ERGM M4’s summary results

Referring to figure 9, for the MCU, although the network is still sparse, it’s condenser comparing to the comic network. Similar to comic universe, mutual/symmetrical ties were more likely to be formed in the MCU. With a positive coefficient, “nodecov.rating” represents the casualty between high ratings and high connectivity. However, the negative coefficient for “nodecov.rewards” indicated that movie hero with more connections do not necessarily win more awards.

Looking at figure 10, and comparing with figure 8, we can conclude the effect of “friend of a friend” also exists in the MCU as “gwesp” term has a positive coefficient. Hence, local structure of triangles is also confirmed. Similarly, there is no scientific tendency of “elbow” structure formed in the network as “gwdsp” has a negative coefficient. However, such results indicate higher transitivity nature of the MCU than the Comic Network.

Comparing the ERGM results for both networks, we are confident to conclude that the MCU inherits the spirit of the Marvel comic universe. Both networks reveal sparseness through the coefficients, which makes sense for Marvel universe having over 6000 heroes in the comic books and over 100 heroes the movie series. Ideally, MCU will bring more comic heroes to the movie productions. Also, from the M4 results, high transitivity nature of the MCU indicates the potential of bringing more heroes in the movies with substantial interactions and low cost. Such interactions will potentially improve the creativity of movie production and reflect through the general movie ratings as indicated in the M3 model. The

Avengers series received higher ratings than individual hero films in general. Hence, we shall expect a more diversified and more creative MCU in the future.

BUSINESS IMPLICATIONS, SUGGESTIONS AND EXPECTATIONS

MCU is the biggest and most valuable cinematic franchise[8], it is important for Disney to focus on their and gain an undisputed advantage over their competitors. The MCU reached the 12 billion box office revenue mark in 2017 though the ride was not smooth. In the first phase MCU was competing with big franchises like Batman (Dark Knight series), Jason Bourne (The Bourne trilogy) and the Ocean's series. Disney Studios also faced the challenges of not having the copyrights for all the heroes in the network/universe. Disney or Marvel has since then been able to tackle these challenges strategically. The strategy included releasing movies in phases and not being phased by the low box office earnings (The Incredible Hulk earned $171 M in 2008 and Thor earned $ 205 M in 2011) of certain hero films.

From our ERGM analysis, we saw the awards not being of any importance in the model and combined with the historical strategic trends of Marvel, we can assume their focus is to capture audience share and keep up with high box office earnings. The first step towards increasing box office earning would be focusing similar marketing strategies for similar hero and movies. For example, ant man, Hawkeye (Ronnin) and the hulk play a big role in the universe but always seen as the sidekick heroes in the movies. To have a synergy effect, one marketing initiative should focus on heroes with higher degree, betweenness, closeness or page rank centralities and the other should focus on the "sidekick" heroes. For example: rather than forcing huge spending on advertisement of the sidekicks, let the audience get more involved with them over social media platforms or press junkets.

Our ERGM models provides valuable information on heroes and their interactions. It is noted that some heroes are very close in the comics but not in the movies. Marvel should focus on such interactions to gain and do something remarkable to gain audience attention. As phase four of the MCU rolls out beginning 2019, there is a scope of being able to provide the audience with some of these interactions. For example: Falcon and Captain America are considered as a strong duo but in MCU they have had not a strong relationship.

Recently, Disney was able to acquire assets of FOX movies and get copyrights to many of the heroes and movie franchises. This has eased ways for Disney to have a crossover between different franchises like the X-men and the Avengers. In our presentation, we focused on how Hank the Beast is on the border of a structure, which puts him in a brokerage position of the network. By having Beast being a mutant ambassador to the Avengers universe, there are many possibilities of increasing box office earnings. Also, at the same time there are more avenues to explore. As we see Mr. Fantastic has the highest closeness centrality and is very influential, in case of Iron Man/Tony Stark dying in the MCU, Mr. Fantastic can take his spot and carry on the role of unifying the universe (same as what Tony Stark does).

Based on our Marvel network analysis and predictive regression results, it is recommended that the stakeholders and the producers should focus on the interactions of different movie series for the MCU and bring more new and centralized characters from comic universe to the movie world. Implementing different strategies towards to central and peripheral heroes in the MCU is vital. From our regression result, having a breakdown in villains and heroes will help us see the effect in potential earnings. Based on both the transitivity characteristic of the MCU and our regression results, movies that contains several central heroes connected with several movie series will add in to the success of the movies, as they garner high audience share, and box-office earnings. This can be demonstrated by the Avengers' success since both Wolverine and Spiderman are statistically centralized characters in our network model. The Marvel stakeholders are also suggested to take advantage of the creative interactions of Marvel heroes. Thus, we recommend that Mr. Fantastic to be the next candidate connecting different movie series together since he has the highest centrality value in our analysis. Furthermore, the Marvel Comic universe should work on bringing more new centralized characters to the movie world. The stakeholders could do this by focusing on making efforts to get the copyrights for marvel comic heroes belonging to SONY (SONY has over 900 characters)[10].

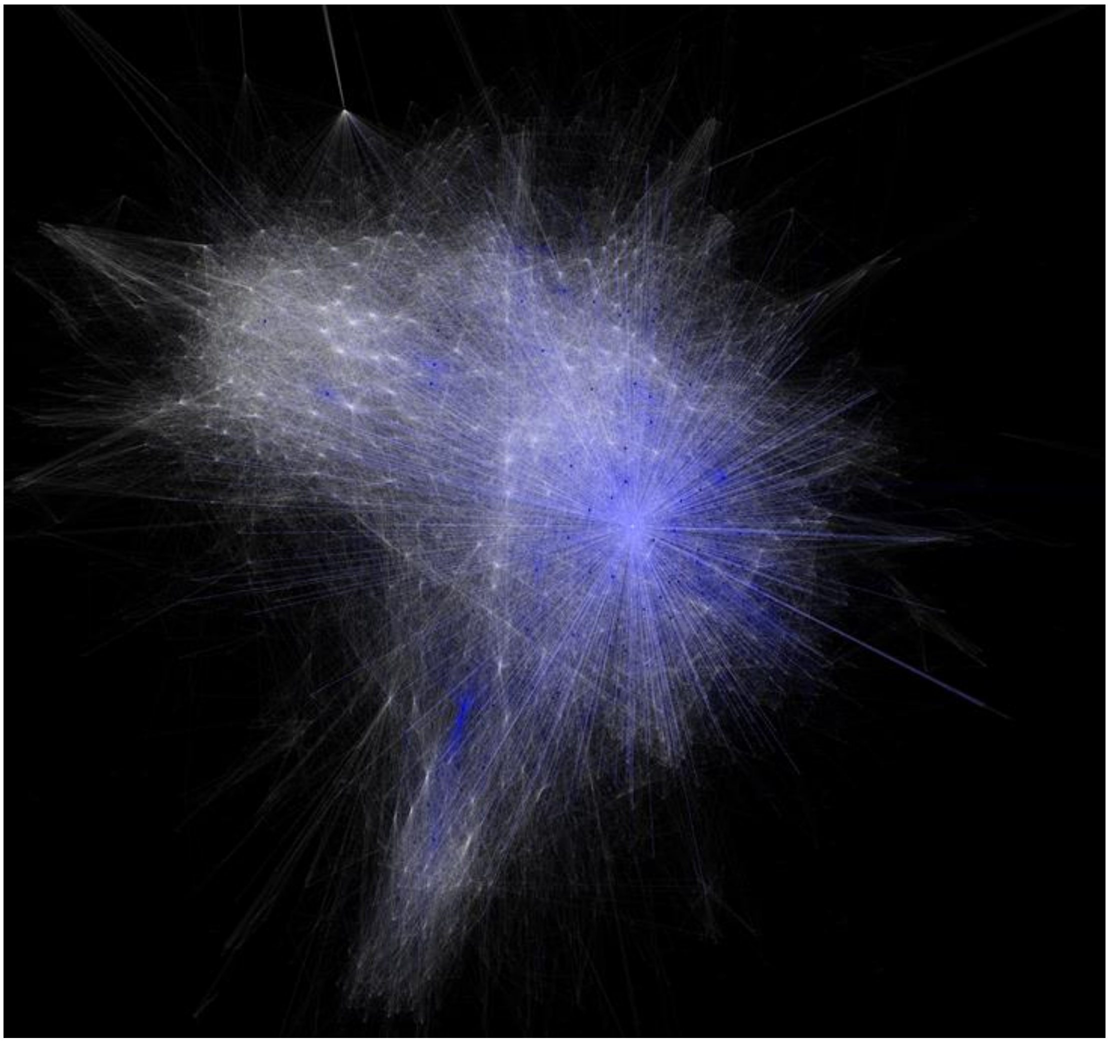
In terms of the future improvement of the network models, it is recommended to include several categorical variables regarding heroes' personal backgrounds in the analysis, for example, a variable defining the role of superstar like Robert Downey Jr. or a new rookie actor like Tom Holland. Also, taking the connections between the Marvel Movie network and actor's real-life relationships into considerations could be of importance to the Marvel network analysis. Bringing in Mel Gibson (Robert Downey Jr. close friend) in the MCU can stir up a reaction in the audience and might acquire more of the audience share and fan base.

APPENDIX

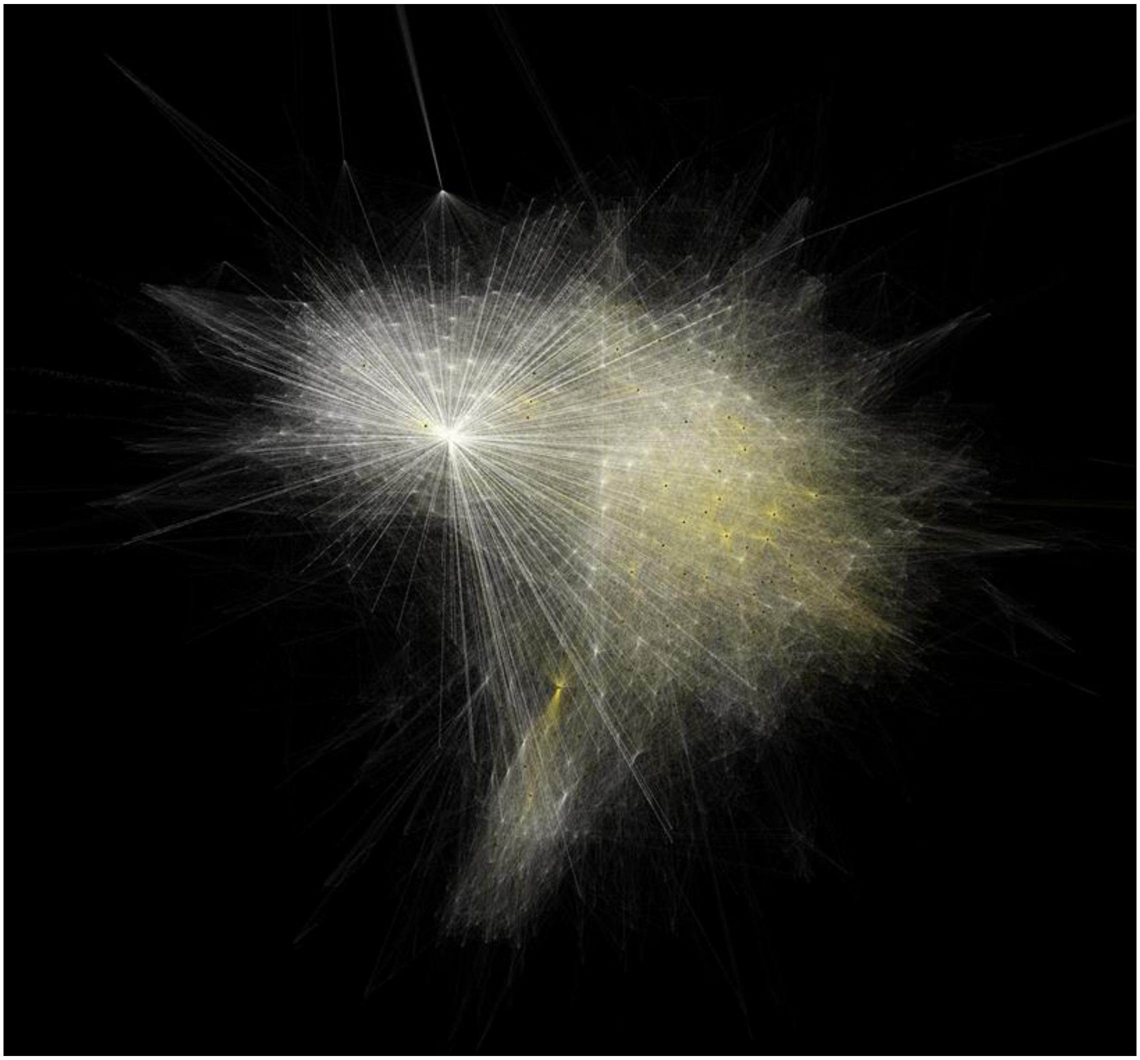
1. https://www.imdb.com/list/ls066122784/
2. https://www.boxofficemojo.com/franchises/chart/?id=avengers.htm
3. https://www.imdb.com/list/ls026696709/?sort=list_order,asc&st_dt=&mode=simple&page=1&ref_=ttls_vw_smp
4. http://marvelcinematicuniverse.wikia.com/wiki/Movie/Awards
5. https://en.wikipedia.org/wiki/Marvel_Cinematic_Universe
6. https://www.forbes.com/sites/robcain/2017/07/10/the-marvel-cinematic-universe-has-now-topped-12-billion-in-total-worldwide-gross/#3051e20a7775
7. https://www.youtube.com/watch?v=hA6hldpSTF8
8. https://www.businessinsider.com/most-successful-movie-franchises-of-all-time-at-box-office-2018-3#2-star-wars-9097-billion-22
9. Jackson, Social and Economic Networks. Princeton Univ. Press, 2008.
10. https://screenrant.com/sony-marvel-comics-characters-movie-rights/

GRAPHS

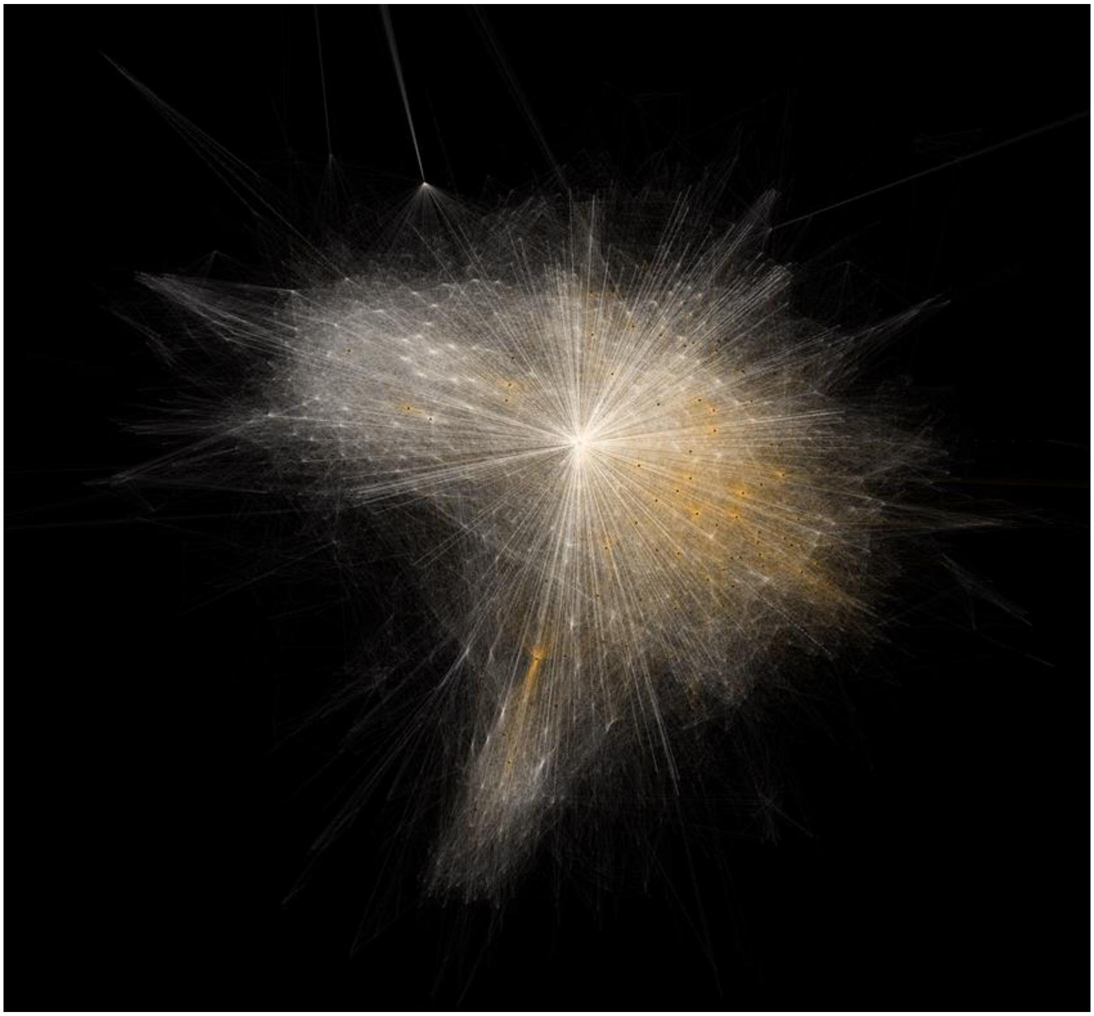
11. Degree centrality - Captain America



12. Betweenness Centrality – Wolverine



13. Closeness centrality -- Mr. Fantastic



14. PageRank centrality – SpiderMan