



The Marvel Universe Social Network

An artificial social network of heroes

Week6 : Social Network Analysis

Data Mining – 이원상 교수님
실습 6주차

김서연 김연재 윤수진 이가람

Content

1. Intro

2. Dataset

2.1 Data Explanation

2.2 EDA & Preprocessing

3. SNA(전체)

4. Grouping

5. SNA(Group1)

6. Conclusion

7. Creative Application Area of SNA

Reference

1.Intro

미국의 박스오피스 집계 사이트에 따르면 '어벤져스: 엔드게임'은 세계에서 26 억 8871 만달러(약 3 조 2122 억원)의 흥행 기록으로 2019 년 개봉한 영화 중 가장 수익을 낸 영화에 뽑혔다. 2 위 역시 마블에서 제작한 영화인 '캡틴 마블' 11 억 2725 만달러(약 1 조 3437 억원)이었다.

마블 영화는 본국인 미국을 넘어 전세계에서 사랑받고 있다. 마블영화는 마블 코믹스 만화책을 원작으로 한다. 1939 년 설립된 타임리 코믹스가 마블 코믹스의 전신이다. 이 회사의 만화가 스탠 리는 수많은 캐릭터를 만들어냈고, 그들이 같은 세계관 속에서 활동할 수 있도록 작품을 썼다. 마블 코믹스는 스파이더맨, 아이언맨, 엑스맨, 캡틴 아메리카, 어벤져스, 판타스틱 포 등 다양한 슈퍼히어로물을 출간하며 전세계인들의 사랑을 받고 있다. 2000 년대 이후로는 만화는 물론 영화로까지 콘텐츠를 넓혀 인기를 얻고 있다.

스탠 리가 구축한 만화책 세계관을 영화로 재해석한 것이 '마블 시네마틱 유니버스(MCU·마블의 영화적 세계관)'다. 마블의 제작담당 사장 케빈 파이기는 MCU 팬 사이에서 세계관의 모든 것을 관장하는 신과 같은 존재로 여겨진다. MCU 는 방대한 세계관으로도 유명한데, 정확한 수치는 알려진 바가 없으나 약 7000 여명의 캐릭터가 존재하는 것으로 추정된다. 따라서 이러한 무수한 등장인물과 스토리가 존재하는 세계관 내에서 인물 간의 관계를 파악하기는 결코 쉽지 않다. 이에 대해 본 실습에서는 마블 코믹스의 등장인물 데이터를 이용하여 SNA 를 진행하여 각 인물관 관계를 파악할 것이고, 세밀한 세계관 이해를 진행할 것이다.

2.Dataset

2.1 Data Explanation

캐글에 공개되어 있는 The Marvel Universe Social Network dataset 의 hero-network.csv 을 활용했다.

¹

데이터는 마블 코믹스의 한 장면에서 등장하는 두 히어로의 정보를 담고 있다.

	hero1	hero2
0	LITTLE, ABNER	PRINCESS ZANDA
1	LITTLE, ABNER	BLACK PANTHER/T'CHAL
2	BLACK PANTHER/T'CHAL	PRINCESS ZANDA
3	LITTLE, ABNER	PRINCESS ZANDA
4	LITTLE, ABNER	BLACK PANTHER/T'CHAL

2.2 Preprocessing

hero-network.csv 은 두 인물이 동시에 장면에 등장했을 때마다 기록된 데이터이다. 따라서 해당 hero 조합이 동시에 몇 번 등장했는지를 weight 로 설정하기 위해 hero1,hero2 를 기준으로 groupby 하여 count 한 열을 추가하였다.

1. hero1 : 해당 장면 등장인물 1
2. hero2 : 해당 장면 등장인물 2
3. weight : 등장인물 1, 등장인물 2 가 동시에 등장한 횟수

두 인물이 동시에 등장한 횟수를 세서 edge 에 attribute 를 부여했다.

```
1 hero['weight'] = 1
2 data = hero.groupby(['hero1', 'hero2']).count().reset_index()
```

¹ <https://www.kaggle.com/datasets/csanhueza/the-marvel-universe-social-network>

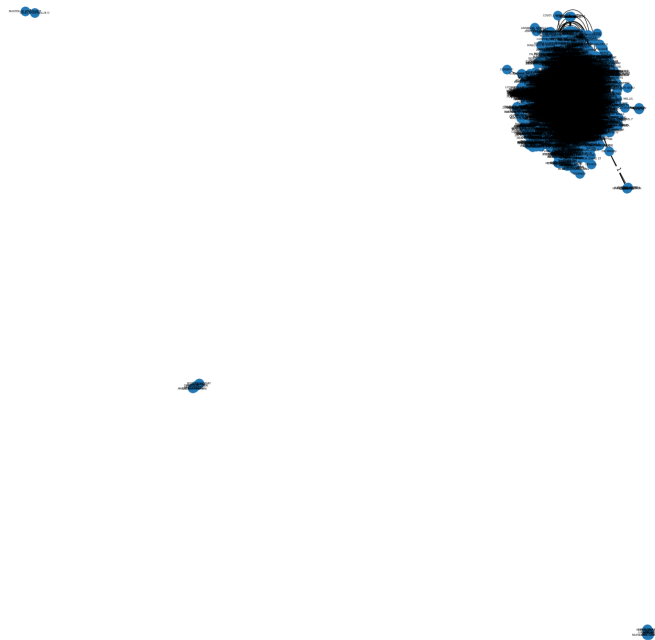
3. SNA(전체)

Python 의 NetworkX 라이브러리로 네트워크를 그려보았더니, 총 6426 개의 node 와 167219 개의 edge 로 이루어진 네트워크를 확인할 수 있었다.

```
1 nx.info(G_data)
```

```
'Graph with 6426 nodes and 167219 edges'
```

네트워크의 시각화를 진행해보았다. 가장자리에 있는 몇몇 관계를 제외하고 중심부에 매우 많은 노드가 몰려있는 것을 확인할 수 있었다.

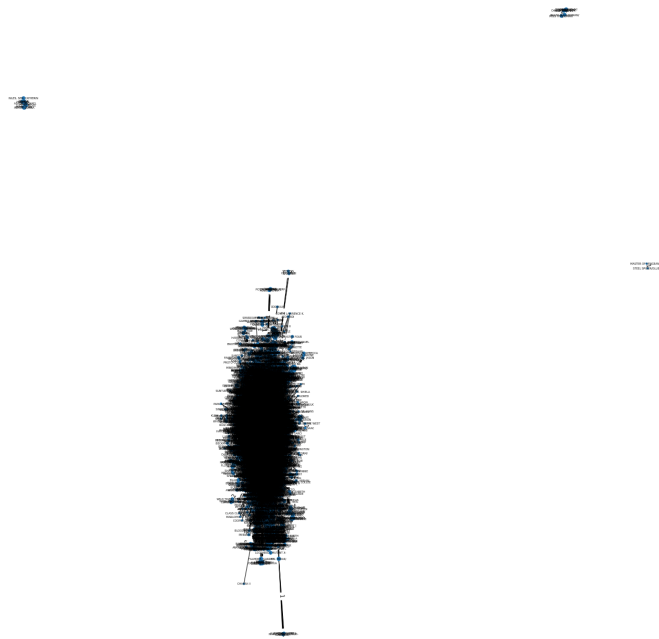


추가적으로, 노드의 크기에 degree 를 반영하여 degree 가 클수록 노드의 크기도 커지도록 네트워크를 그려보았다.

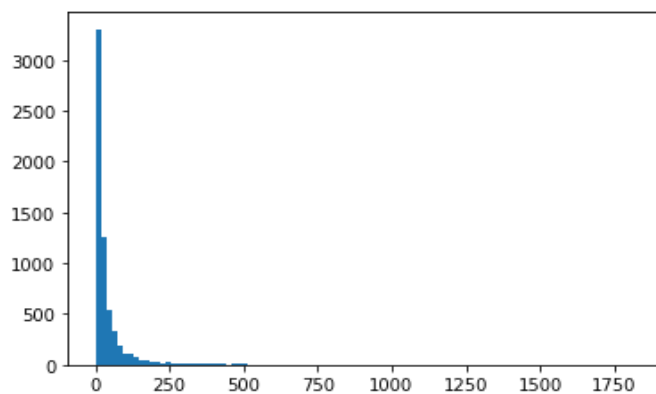
```

4 plt.figure(figsize=(20, 20))
5
6 degree = nx.degree(G_data)
7 pos=nx.spring_layout(G_data)
8 nx.draw(G_data, pos, with_labels=True, node_size=[5 + v[1]*5 for v in degree], font_size = 5)
9 edge_weight = nx.get_edge_attributes(G_data, 'weight')
10 nx.draw_networkx_edge_labels(G_data, pos, edge_labels = edge_weight)
11 plt.show()

```



이전에 비해 가장자리의 노드는 더 작아지고, 중심부는 더 밀집해있는 것을 볼 수 있는데, degree 가 노드와 연결된 모든 엣지의 수를 의미한다는 것을 고려했을 때 가장자리는 엣지의 수가 적기 때문에 노드의 크기가 작아지고, 중심부에는 많은 노드가 서로 연결되어 있기 때문에 더 밀집되었다는 것을 알 수 있다.



degree 를 확인했을 때 degree=0 인 노드가 상당히 많은 것을 알 수 있었다. 이는 연결된 edge 의 수가 0 인 node 가 상당히 많은 것을 의미한다. 현재 데이터는 마블 코믹스의 작품 안에서 다른 등장인물과 동시에 등장하는 인물, 그 횟수를 count 한 것이다. 따라서 현재 dataset 에서 degree=0 인 node 가 많은 것은 작품에 많이 등장하지 않는 인물이거나 다른 인물과 동시에 등장하는 것이 아닌, 단독 등장 횟수만 높은 등장 인물일 것이다.

등장 횟수가 높은 등장인물 상위 10 개를 추출한 결과 다음과 같았다.

```
1 #상위 10개 추출
2 herolist[:10]

[('CAPTAIN AMERICA', 1908),
 ('SPIDER-MAN/PETER PAR', 1737),
 ('IRON MAN/TONY STARK ', 1522),
 ('THING/BENJAMIN J. GR', 1416),
 ('MR. FANTASTIC/REED R', 1379),
 ('WOLVERINE/LOGAN ', 1371),
 ('HUMAN TORCH/JOHNNY S', 1361),
 ('SCARLET WITCH/WANDA ', 1325),
 ('THOR/DR. DONALD BLAK', 1289),
 ('BEAST/HENRY &HANK& P', 1267)]
```

degree 가 높은 상위 10 개를 추출한 결과 다음과 같았다. 마블의 등장인물 중 'Avengers'의 멤버인 Captain America, Iron Man, Scarlet Witch, Thor 을 볼 수 있었고, 'Fantastic Four'의 멤버인 Thing, Mr. Fantastic, Human Torch 를 볼 수 있다. 그 외에도 'X-Men'의 일원인 Wolverine, Beast 와 슈퍼히어로 팀에 속하지는 않지만 역시나 마블의 인기 히어로인 스파이더맨도 많이 출연한 것을 알 수 있다.

```
1 print('density:', nx.density(G_data))
2 print('transitivity:', nx.transitivity(G_data))

density: 0.00810031232553549
transitivity: 0.19453974709267596
```

density 는 0.008, transitivity 는 0.19 로 나왔다. 해당 네트워크는 undirected graph 이기 때문에 reciprocity 는 구하지 않았다. Density 는 모든 엣지 개수 대비 실제 엣지 개수를 나타내는 지표로, 이 네트워크에서는 약 0.008 으로 나왔기 때문에, 마블 캐릭터의 수에 비해 출연이 많이 겹치는 것은 아니라고 해석할 수 있다. Transitivity 는 노드 3 개 이상이 서로 연결된 비율로, 3 명 이상의 캐릭터가 같이 출연한 지표로 이해할 수 있다. 이 네트워크에서는 약 0.19 로 높지 않은 수치이기 때문에, 3 명 이상의 캐릭터 조합이 여러 영화에 동시에 출연한 경우는 적었다고 해석할 수 있다.

4. Grouping

```
1 print('diameter:', nx.diameter(G_data))

-----
NetworkXError                                Traceback (most recent call last)
<ipython-input-19-015af52f11d1> in <module>()
----> 1 print('diameter:', nx.diameter(G_data))

-----
1 frames -----
/usr/local/lib/python3.7/dist-packages/networkx/algorithms/distance_measures.py in eccentricity(G, v, sp)
    262     else:
    263         msg = "Found infinite path length because the graph is not " " connected"
--> 264         raise nx.NetworkXError(msg)
    265
    266     e[n] = max(length.values())

NetworkXError: Found infinite path length because the graph is not connected
```

한편 diameter 를 구할 수 없었다. 이를 해결하기 위해 확인을 해보았을때, 네트워크 내의 모든 노드들이 연결되어 있는 것이 아닌 총 19 개의 그룹이 이어져 있었다. 각각의 그룹을 분리하여 네트워크 시각화를 다시 진행해보고, 다른 지표들을 다시 확인해보았다.

각 그룹의 노드 개수는 다음과 같았다.

```
1번째 그룹 : 6408개 노드
2번째 그룹 : 7개 노드
3번째 그룹 : 9개 노드
4번째 그룹 : 2개 노드
```

첫번째 그룹을 제외한 나머지 그룹은 10 개 이하의 노드를 가지고 있었기 때문에, 유의미한 분석 결과를 얻기 어려울 것이라고 판단했다. 따라서 첫번째 그룹의 노드만 남기고 다시 분석을 해보았다.

	hero1	hero2	weight	ox
0	24-HOUR MAN/EMMANUEL	FROST, CARMILLA	1	False
1	24-HOUR MAN/EMMANUEL	KILLRAVEN/JONATHAN R	1	False
2	24-HOUR MAN/EMMANUEL	M'SHULLA	1	False
3	3-D MAN/CHARLES CHAN	ANGEL/WARREN KENNETH	1	False
4	3-D MAN/CHARLES CHAN	ANT-MAN II/SCOTT HAR	1	False

1 번 그룹을 제외한 노드의 리스트를 추출하여 노드 리스트에 있지 않은 행에 대해서만 다시 네트워크를 만들었다.

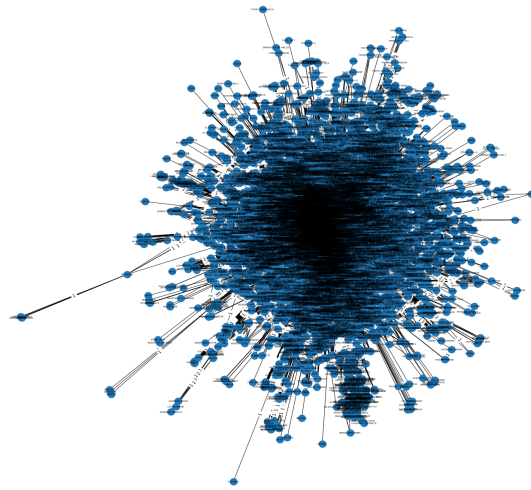
5. SNA(Group1)

1 번 그룹에 대해 다시 네트워크를 그렸을 때 6408 개의 노드와 167163 개의 엣지가 존재하는 것을 확인했다.

```
1 print(nx.info(G_data))

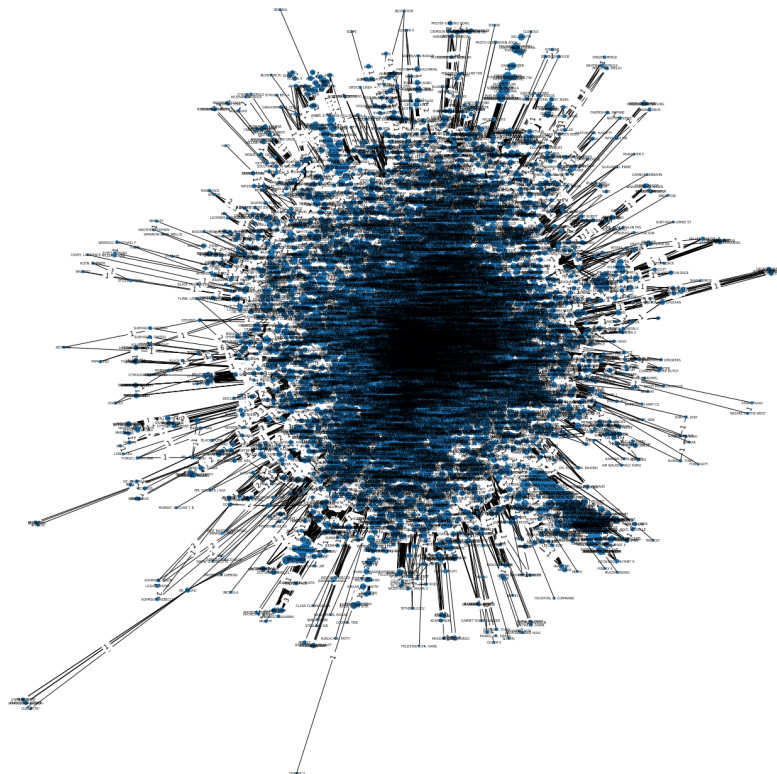
Graph with 6408 nodes and 167163 edges
```

시각화한 결과는 다음과 같다.

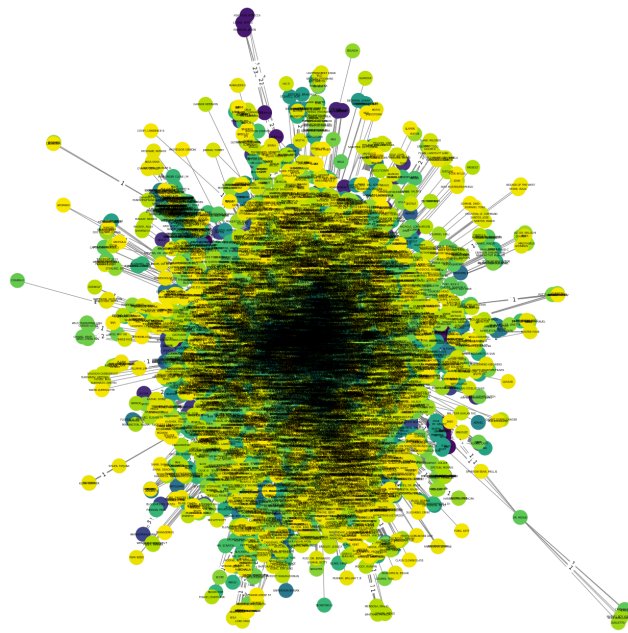


전체를 대상으로 네트워크를 그렸을 때보다 노드가 중심을 기준으로 고르게 분포했고, 주변부에 노드가 덜 분산되어 있는 것을 확인할 수 있다.

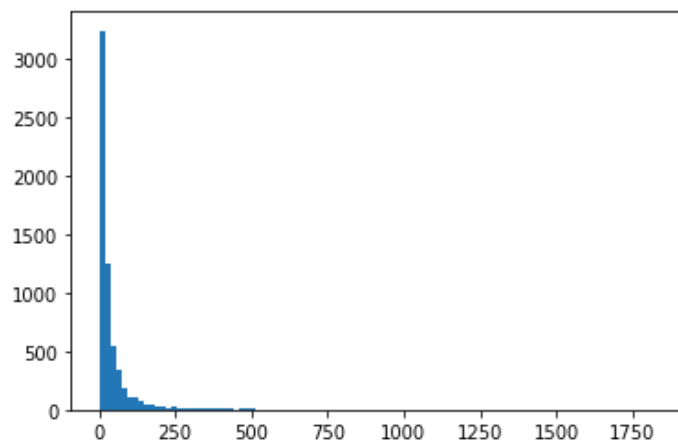
degree 를 반영하여 시각화한 결과는 다음과 같다.



끝으로 갈수록 노드 간 연결이 적어져 노드의 크기가 작아진 것을 알 수 있다.



원의 degree 를 색깔로 시각화한 결과 위와 같았다. 밝은 색일수록 degree 가 높다는 것을 의미한다. 전반적으로 네트워크의 중심에 degree 가 높은 노란색 노드가 많이 분포한 것을 확인할 수 있다.



1 번 그룹의 degree 를 확인했을 때 역시 전체 데이터의 degree 를 확인했을 때와 동일하게 degree 가 0 인 node 가 상당히 많은 것을 확인할 수 있었다. 1 번 그룹의 경우 마블 코믹스 작품에서 등장 횟수가 적은 인물이 많다는 것을 알 수 있었다.

```
1 #상위 10개 추출
2 herolist[:10]
```

```
[('CAPTAIN AMERICA', 1908),
 ('SPIDER-MAN/PETER PAR', 1737),
 ('IRON MAN/TONY STARK ', 1522),
 ('THING/BENJAMIN J. GR', 1416),
 ('MR. FANTASTIC/REED R', 1379),
 ('WOLVERINE/LOGAN ', 1371),
 ('HUMAN TORCH/JOHNNY S', 1361),
 ('SCARLET WITCH/WANDA ', 1325),
 ('THOR/DR. DONALD BLAK', 1289),
 ('BEAST/HENRY &HANK& P', 1267)]
```

degree 가 높은 상위 10 개를 추출한 결과 다음과 같다. 전체 데이터를 대상으로 한 네트워크의 리스트와 큰 차이가 없는 것으로 보아 어벤져스, 엑스맨, 판타스틱 포가 마블 세계관의 주축임을 다시 한 번 확인할 수 있었다.

```
1 print('diameter:', nx.diameter(G_data))
2 print('density:', nx.density(G_data))
3 print('transitivity:', nx.transitivity(G_data))

diameter: 5
density: 0.008143159196782077
transitivity: 0.19453467386750134
```

위에서 전체 데이터를 대상으로 했을 때와 크게 달라지지 않았다. 이를 통해 그룹 1 의 비중이 전체 데이터에서 매우 컸음을 확인할 수 있었다. diameter, density, transitivity 모두 비교적 작은 수치를 보이는데, 이에 대해 마블 캐릭터가 수에 비해 동시 출연이 많은 편이 아니며, 같은 인물 조합이 여러 편에 출연한 경우는 적은 것으로 해석할 수 있다.

각 centrality 별 가장 값이 높은 상위 10 개의 노드를 추출했다.

Centrality	Node
degree centrality	1. 'CAPTAIN AMERICA', 0.29779928203527395 2. 'SPIDER-MAN/PETER PAR', 0.27110972373965975 3. 'IRON MAN/TONY STARK ', 0.237552676759794 4. 'THING/BENJAMIN J. GR', 0.22100827220227878

	<ol style="list-style-type: none"> 5. 'MR. FANTASTIC/REED R', 0.2152333385359763 6. 'WOLVERINE/LOGAN ', 0.21398470422974872 7. 'HUMAN TORCH/JOHNNY S', 0.21242391134696428 8. 'SCARLET WITCH/WANDA ', 0.20680505696894025 9. 'THOR/DR. DONALD BLAK', 0.2011862025909162 10. 'BEAST/HENRY &HANK& P', 0.1977524582487904
betweenness centrality	<ol style="list-style-type: none"> 1. 'SPIDER-MAN/PETER PAR', 0.07395319167317937 2. 'CAPTAIN AMERICA', 0.05728953380407401 3. 'IRON MAN/TONY STARK ', 0.03744203870158992 4. 'WOLVERINE/LOGAN ', 0.03590348013660662 5. 'HAVOK/ALEX SUMMERS ', 0.035860776743666944 6. 'DR. STRANGE/STEPHEN ', 0.029314080324519803 7. 'THING/BENJAMIN J. GR', 0.025588667690899774 8. 'HAWK', 0.024973058476740452 9. 'HULK/DR. ROBERT BRUC', 0.02406788571775531 10. 'MR. FANTASTIC/REED R', 0.023901532279008198
closeness centrality	<ol style="list-style-type: none"> 1. 'CAPTAIN AMERICA', 0.5853279736890188 2. 'SPIDER-MAN/PETER PAR', 0.5757031179800521 3. 'IRON MAN/TONY STARK ', 0.5630052724077329 4. 'THING/BENJAMIN J. GR', 0.5593190746398953 5. 'MR. FANTASTIC/REED R', 0.557615317667537 6. 'WOLVERINE/LOGAN ', 0.5564530137224248 7. 'HUMAN TORCH/JOHNNY S', 0.5564046895353886 8. 'SCARLET WITCH/WANDA ', 0.5538554633471646 9. 'THOR/DR. DONALD BLAK', 0.5524228315226763 10. 'BEAST/HENRY &HANK& P', 0.5506660936828535
eigenvector centrality	<ol style="list-style-type: none"> 1. 'CAPTAIN AMERICA', 0.11677469223625928 2. 'IRON MAN/TONY STARK ', 0.10254077322184338 3. 'SCARLET WITCH/WANDA ', 0.10082113046617178 4. 'THING/BENJAMIN J. GR', 0.10078249100850406 5. 'SPIDER-MAN/PETER PAR', 0.10023214958686026 6. 'MR. FANTASTIC/REED R', 0.0997451725132082 7. 'VISION ', 0.09853383668565327 8. 'HUMAN TORCH/JOHNNY S', 0.0985175127988693 9. 'WOLVERINE/LOGAN ', 0.09836435239994289 10. 'BEAST/HENRY &HANK& P', 0.09549890616372651

연결 중심성, 매개 중심성, 근접 중심성, 아이겐벡터 중심성이 높은 상위 10 개의 노드를 추출한 결과 각 중심성 지표에서 상위 값을 갖는 노드들이 비슷하게 나타났다. 'CAPTAIN AMERICA'가 모든 중심성 지표에서 가장 높은 노드로 나왔고 'CAPTAIN AMERICA', 'IRON MAN/TONY STARK ', 'SPIDER-MAN/PETER PAR', 'THING/BENJAMIN J. GR', 'WOLVERINE/LOGAN '는 4 가지 지표에서 모두 상위 10 개의 노드 안에 속했다.

따라서 마블 유니버스 전체에서 캡틴 아메리카가 가장 중심적인 역할을 하며, 아이언맨, 스파이더맨, 싱, 울버린도 중심적인 역할을 한다는 것을 알 수 있다. 또한 마블 코믹스의 작품 중 장편 시리즈로 제작된 '어벤저스', '판타스틱 포', '엑스맨'의 등장인물들의 중심성이 높은 것을 알 수 있었다. 더불어 '판타스틱 포'와 '어벤저스', '엑스맨'이 동맹관계를 갖는다는 점에서 동시 등장 횟수가 증가했을 것이고 이에 따라 각 등장인물 들의 중심성이 더욱 높아졌을 것이다.

연결 중심성, 근접 중심성, 아이겐벡터 중심성은 상위 10 개의 노드가 대부분 비슷한 양상을 보였지만 매개 중심성의 경우 상위 10 개의 노드를 추출했을 때 다른 중심성 지표와 달리 눈에 띄는 다른 노드를 확인할 수 있었다. 'HAVOK/ALEX SUMMERS ', 'DR. STRANGE/STEPHEN', 'HAWK'를 찾아볼 수 있었다. 매개 중심성은 네트워크 내에서 한 노드가 다른 노드들 사이에 위치하는 정도를 나타내는 지표이다. 즉 네트워크 내에서 어디에 위치하는지를 파악함으로써 해당 노드의 영향력을 파악할 수 있다. 네트워크 내의 매개 중심성이 높은 노드는 다른 핵심 노드들 사이에서의 broker 역할을 수행하게 되며 핵심 노드로서 기능할 가능성이 높다. 즉 다른 중심성 지표에서는 최상위 10 개 노드에 속하지 못했지만 매개 중심성이 높은 'HAVOK/ALEX SUMMERS ', 'DR. STRANGE/STEPHEN', 'HAWK'는 다른 핵심 노드 사이에서 중재자 역할을 수행하게 되는 노드 들이다.

실제로 'Havok'은 엑스맨의 'Cyclops'의 동생이자 Vulcan 의 형이며, 연인인 '폴라리스'도 있다. 'Havok'이 다른 등장인물들에 비해 가족/연인 관계가 많은 것을 알 수 있다. 'Dr. strange'의 경우에는 '가디언즈 오브 갤럭시'의 고대의 존재, '마블 좀비 세계'의 좀비로 변신할 마지막 영웅, 'Thor : Viking'의 영웅으로 등장하는 등 마블 코믹스 내의 다양한 작품에 출연하여 세계관을 연결하는 역할을 한다. 'Hawk'는 연결 중심성, 근접 중심성, 아이겐벡터 중심성이 높은 어벤저스 및 토르의 등장인물들과 관계를 지니고 있다.

또한 Young Avengers 중 하나인 Kate Bishop 의 스승으로의 역할도 겸하기 때문에, 새로운 슈퍼히어로 팀과도 긴밀한 관계를 가져 매개 중심성이 높게 나왔다고 할 수 있다.

아이겐벡터 중심성에서도 눈에 띄는 'VISION' 노드를 찾아볼 수 있었다. 아이겐벡터 중심성은 네트워크 내에 한 노드와 연결된 다른 액터의 중심성을 가중치로 하여 계산하는 지표이다. 즉 자신과 연결된 다른 노드들이 네트워크 내에서 얼마나 중요한지 파악할 수 있다. 아이겐벡터 중심성이 높다면 연결정도가 적다 하더라도 높은 영향력을 가지며, 반대로 연결정도는 높지만 아이겐벡터 중심성이 낮다면 네트워크 내의 실질적인 영향력은 미미할 것으로 예측할 수 있다. 따라서 'VISION'의 연결정도 중심성이 상위 10 개 노드에 속하지 못하지만 아이겐 벡터 중심성이 높은 것에 대해 연결정도가 적지만 네트워크 내에서 'VISION'과 연결된 다른 노드들이 중요하기 때문에 실질적인 영향력이 크다고 해석할 수 있다.

6. Conclusion

마블 코믹스에 등장하는 인물의 관계를 SNA 를 활용하여 분석하였다. 전체 데이터에 대해서는 6426 개의 node 와 167219 개의 edge 를 발견하였고, 이에 대해 시각화와 structure 정보를 도출했다. structure 정보 중 diameter 이 확인되지 않는 이유에 대해 네트워크의 모든 노드가 이어지지 않았다는 문제점을 발견하고, 가장 많은 노드가 속해 있는 그룹만을 추출하여 다시 분석을 진행했다.

그 결과 모든 노드가 이어진 그래프를 그릴 수 있었고, degree 가 높은 상위 10 명의 히어로 리스트를 출력했을 때 어벤저스, 엑스맨, 판타스틱 포의 구성원이 많은 것으로 보아 이들이 마블 세계관의 주축임을 확인할 수 있었다. 새로 분석한 네트워크의 구조 정보는 diameter: 8, density: 0.008, transitivity: 0.19 로 모든 캐릭터 간 거리가 길지 않은 것으로 보아 멀리 연결되어 있지는 않으며, 인물 수에 비해 동시 출연이 많지 않다는 분석을 할 수 있었다.

마지막으로 각 centrality 별 상위 10 개 노드를 추출했는데, 모든 부문에서 Captain America 가 1 위인 것으로 보아 마블 세계관 전체에서 매우 중요한 역할을 하는 인물인 것을 확인했다. 이외에도 어벤저스와 엑스맨, 그리고 판타스틱 포의 구성원 역시 큰 영향력을 끼치는 것을 관찰했다. 눈에 띄는

노드는 매개 중심성의 Havock, Dr. Strange, Hawk, 그리고 아이겐벡터 중심성의 Vision 이었는데, 이들의 알려진 인물 정보와 인간관계를 바탕으로 보았을 때 유의미한 분석이라고 결론을 내릴 수 있었다.

7. Creative Application Area of SNA

SNA 는 복잡한 관계로 얽힌 데이터를 설명하는 데에 효과적인 분석 방법이다. 본 실습에서는 마블 코믹스의 등장인물 관계도를 그리기 위해 SNA 를 사용했지만, 이외에도 다양한 분야에 접목시킬 수 있다.

SNS 상의 인물정보와 활동을 SNA 로 분석하여 위험한 인물의 활동을 파악하고 그들의 불법행위를 발견한 연구가 이루어진 바 있다. 스웨덴 스톡홀름 대학의 Berzinji, A., Abdullah, F. S., Kakei, A.H (2013)[1]에 의하면, 테러리스트 집단은 사람들이 많이 모이는 SNS 에서 그들의 생각을 전파하고 있다고 한다. 또한 전세계 테러 조직인 IS 도 SNS 를 통해 조직원들을 모집하고 테러를 모의했다고 한다. 실제로 2014 년 1 월 IS 의 SNS 를 통한 조직원 모집에 대한민국 18 세 고등학생 김모군이 현혹되어 출국한 후 현재까지 행방불명된 사건도 있다.

이에 대해 SNS 상에서 활동하고 있는 테러리스트 집단의 관계망을 분석하고 어떤 정보를 어떻게 퍼뜨리는지, 그리고 그들이 퍼뜨리는 정보에 대한 효과는 어느 정도인지 분석하여 그들의 만행을 방지할 수 있다.

SNA 를 범죄 네트워크 분석에 다음과 같이 적용할 수 있다. node 를 범죄 집단의 단위 주체(사람, 활동 계정), edge 를 테러리스트의 범죄 행위로 설정할 수 있다. 그리고 각 노드의 중심점을 통해 전체 범죄 네트워크에서 중심이 되는 범죄자들을 찾을 수 있을 것이다. degree 는 어떤 노드가 얼마나 많이 다른 노드와 연결되었는가를 측정하는 지표이므로 어떤 범죄 조직원이 얼마나 많이 다른 조직원 또는 계정에 연결되어 있는지 알 수 있다. betweenness 는 얼마나 많은 최단 경로가 특정 노드를 지나는지를 측정하는 지표이므로 정보에 흐름에 얼마나 범죄 조직원이 중요한 역할을 하는지 보여준다. closeness 는 얼마나 빨리 네트워크 안에서 노드가 정보에 접근 할 수있는지를 보여준다. Eigenvector centrality 는 누가

중요한 노드들과 연결이 많이 되어있는지를 측정하는 지표이기 때문에 어떤 조직원이 핵심인지 알 수 있을 것이다. 그런데 노드는 그들의 중심점에 대해 각각 등수가 정해져 있기 때문에 가장 상위의 노드가 중요한 역할을 할 것이다.

SNA 를 의학 분야에서도 활용할 수 있다. 우리가 네트워크하면 생각하는 사람들 간의 관계 외에도 의생명과학 영역에서도 네트워크를 쉽게 찾을 수 있다. 작게는 유전자, 단백질, 세포부터 크게는 질병 수준에 이르기까지 다양한 영역에 네트워크가 존재한다. 의학 분야에서의 네트워크 분석이 잘 알려지지 않다가 2000년대 초반, 헝가리의 과학자 바라바시가 사회연결망, 인터넷 연결망 및 생체 단백질 연결망 등 구조적 특징을 파악하게 되면서 그 중요성이 알려지게 되었고 이는 하나의 학문인 '네트워크 의학'으로 자리잡게 되었다. 이러한 네트워크 의학은 DNA, RNA 그리고 단백질의 상호작용에 관한 메커니즘을 연구하는데 활용된다. 또한 데이터 기반의 신약을 개발하는 곳에도 다각도로 활용되고 있다. 특히 기존 약물을 갖고 새로운 적응증을 찾아내는데 많이 활용되고 있다. 또한 질병의 상호 연결성을 통해 과거 앓았던 질병으로 미래에 어떤 질병이 걸릴 지, 질병이 한 사람으로부터 시작되어 어떻게 전파되는지 예측하는 데에도 사용될 수 있다. 이처럼 의학의 다양한 분야에서 네트워크 분석을 적용하면서 질병 상호관계와 약물과 유전체의 상호관계를 통해 환자 맞춤 치료를 하는데 기여할 것으로 보인다. 궁극적으로 정밀 의료의 실현에 있어 결정적인 역할을 할 것으로 보인다.

Reference

[1] Berzinji, A., Abdullah, F.S. & Kakei, A.H., "Intelligence and Security Informatics Conference (EISIC)", 2013 European, 2013.

"네트워크 의학, 환자 맞춤치료에 기여...적극 뛰어들어야"

<https://www.docdocdoc.co.kr/news/articleView.html?idxno=1064720>