SNA

Witcher Network 분석

데이터마이닝이론및응용 에이쁠원하조

> 2020147024 김우영 2020251009 김혜리 2020195053 노가원 2020147018 조윤영

RQ

소설<더 위쳐시리즈>를 원작으로 하여 현재 시즌 2까지 방영된 넷플릭스 오리지널 드라마 <위쳐> 시리즈는 탄탄한 구성으로 사랑받는 드라마이다. 하지만 방대한 소설 내용 때문에 드라마만을 시청한 시청자들은 인물간의 관계를 이해하기 어렵다는 평가가 있다. 우리는 소설 1~8권까지의 인물 관계 데이터를 활용하여 <위쳐>의 인물 관계를 분석하고자 한다. 하지만 단순히 1~8권 까지의 내용을 네트워크 분석하는 것이 아니라 시즌1에 해당하는 4권까지의 관계와 소설 전체에 해당하는 관계를 각각 분석하여 시즌1, 시즌2에서 변화하는 인물의 양상을 살펴보고자 한다. 또한 각 중심성 지수가 소설 내용이 진행됨에 따라 변화하는 양상을 바탕으로 인물 간에 어떠한 변화가 있는지, 내용의 흐름이 어떻게 진행되는지 확인해보고자 한다.

Data, 변수 설명

witcher_network.csv는 다음과 같은 변수를 가지고 있고 우리는 소설 1,2,3,4권에 해당하는 data_half와 소설 전권의 데이터인 data_full로 나누어서 비교하여 분석을 진행했다.

Source: <더 윗쳐시리즈>의 등장인물

Target: 등장인물 중 Source의 인물과 연결되어 관계가 있는 인물

Type: 연결의 종류인 Directed/Undirected를 구분한다.

Weight: 두 등장인물이 상호작용한 횟수로서 관계의 깊이를 나타내는 수치라고 볼 수 있다.

book : 처음으로 <더 윗쳐시리즈>에서 등장한 책의 권수.

Outlier 제거

두 데이터 각각에서 다른 노드들과 어떠한 연결도 없는 노드는 제거하였다. data_full -> 'Vreemde', 'Morteisen' 제거 data_half -> 'Voymir', 'Sigismund' 제거



```
df_f = data_full[['Source', 'Target', 'Weight']]
df_f.drop_duplicates(subset=['Source', 'Target', 'Weight'], inplace=True)

#these characters were way off to the side which made the rest of the networl
df_f = df_f.drop(df_f.index[df_f('Source'] == 'Vreemde'])
df_f = df_f.drop(df_f.index[df_f('Target'] == 'Worteisen'])
df_f = df_f.drop(df_f.index[df_f('Target'] == 'Vreemde'])
df_f = df_f.drop(df_f.index[df_f('Target'] == 'Morteisen'])
```

```
df_h = data_half[['Source', 'Target', 'Weight']]
df_h.drop_duplicates(subset=['Source', 'Target', 'Weight'], inplace=True)
#these characters were way off to the side which made the rest of the network
df_h = df_h.drop(df_h.index[df_h['Source'] == 'Voymir'])
df_h = df_h.drop(df_h.index[df_h['Source'] == 'Sigismund'])
df_h = df_h.drop(df_h.index[df_h['Target'] == 'Voymir'])
df_h = df_h.drop(df_h.index[df_h['Target'] == 'Sigismund'])
```

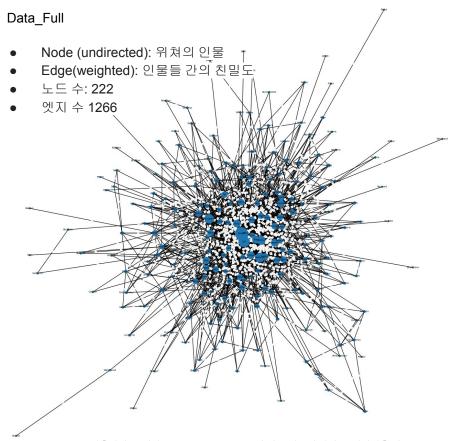
node와 edge 및 네트워크 그림

Data_Half

Node(undirected): 위쳐의 인물

Edge(weighted): 인물들 간의 친밀도 노드 수: 153 엣지 수 663

Geralt, Ciri와 같은 인물이 가장 큰 영향력을 갖는 것을 볼 수 있다. Dennis, Rennie, Ekkerhard 등의 인물은 중심에서 멀리 plot되어 있어 적은 영향력을 갖는것을 볼 수 있다. 이는 소설에서 큰 비중을 갖지 않는 인물임을 의미한다.

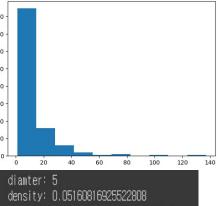


weighted를 사용하여 본 결과 Geralt, Ciri, Dandelion과 같은 인물이 가장 큰 영향력을 갖는 것을 볼 수 있다. 반면 Ramon, Ekkerhard 등의 인물은 중심 바깥에 작은 원으로 표현된 것을 보아 적은 영향력을 갖는 인물인 것을 알 수 있다. 또한 기존에 큰 영향력을 갖지 못했던 Dandelion이 비교적 큰 영향력을 갖게된것을 볼 수 있다.

네트워크의 degree distribution 확인 및 density/diameter 등 확인

1. Data full

[('Velerad', 3), ('Geralt', 137), ('Foltest', 32), ('Adda', 4), ('Ostrit', 3), ('Two', 37), ('Yennefer', 72), ('Nenneke', 16), ('Tola', 15), ('Roach', 12), ('Nivellen', 2), ('Fenne', 1), ('Asse 9), ('Vereena', 3), ('Hereward', 5), ('Falwick', 5), ('Caldemeyn', 140 ('Stregobor', 7), ('Count', 30), ('Eltibald', 1), ('Fredefalk', 3) ('Renfri', 7), ('Nohorn', 4), ('Tavik', 5), ('Civril', 6), ('Fifte 4), ('Audoen', 4), ('Libushe', 1), ('Baron', 29), ('Haxo', 5), ('Pavetta', 24), ('Calanthe', 39), ('Rainfarn', 8), ('Windhalm', 5 ('Eist', 11), ('Crach', 21), ('Draig', 7), ('Drogodar', 5), ('Vissegerd', 23), ('Ravix', 1), ('Mousesack', 11), ('Roegner', 6) 100 ('Sir', 30), ('Rumplestelt', 2), ('Akerspaark', 7), ('Nettly', 3), ('Dhun', 3), ('Lille', 7), ('Little', 23), ('Torque', 7), ('Galarr 4), ('Aen', 26), ('Vanadain', 5), ('Toruviel', 12), ('Filavandrel' 9), ('Field', 19), ('King', 43), ('Chireadan', 7), ('Errdil', 7), ('Vratimir', 4), ('Beau', 21), ('Triss', 45), ('Laurelnose', 4), ('Neville', 5), ('Eskel', 7), ('Vesemir', 14), ('Dennis', 11), ('Istredd', 4), ('Herbolth', 4), ('Seven', 16), ('Vespula', 2), ('Dandelion', 69), ('Gar', 22), ('Dudu', 2), ('Vimme', 2), ('Sword 40-12), ('Agloval', 6), ('Teleri', 3), ('Essi', 7), ('Duke', 9), ('Zelest', 4), ('Veverka', 1), ('Braenn', 9), ('Freixenet', 6), ('Eithné', 7), ('Ciri', 102), ('Venzlav', 7), ('Guy', 6), ('Yurga' 9), ('Visenna', 2), ('Sheldon', 6), ('Donimir', 3), ('Rience', 31) ('Lambert', 5), ('Vanielle', 4), ('Ori', 12), ('Forest', 7), ('Yarpen', 17), ('Xavier', 3), ('Regan', 5), ('Yannick', 4), ('Paulie', 8), ('Henselt', 18), ('Aelirenn', 2), ('Everett', 5), ('Shani', 11), ('Philippa', 45), ('Meve', 17), ('Ethain', 6), ('Embyr', 62), ('Ithlinne', 19), ('Vilgefortz', 44), ('Ekkehard', 1), ('Artaud', 11), ('Tissaia', 22), ('Lydia', 14), ('Jarre', 22), ('Jan' 11), ('Blue', 16), ('Aplegatt', 10), ('Demavend', 20), ('Codringher', 19), ('Sabrina', 30), ('Horm', 2), ('Yaevinn', 4), ('Molnar', 6), ('Fabio', 6), ('Anna', 17), ('Margarita', 20), ('Falka', 28), ('Rayla', 4), ('Bran', 2), ('Esterad', 16), ('Francesca', 30), ('Carduin', 8), ('Marti', 17), ('Dorregaray', 8), ('Keira', 16), ('Leticia', 2), ('Sigismund', 2), ('Voymir', 1), ('Morvran', 3), ('Fauve', 3), ('Bronibor', 9), ('Menno', 12), ('Ardal', 8), ('Vattier', 15), ('Stefan', 13), ('Xarthisius', 4), ('Ceallach', 4), ('Cahir', 32), ('Kayleigh', 10), ('Iskra', 3), ('Mistle', 14), ('Rudiger', 2), ('Milva', 33), ('Maria', 4), ('Shilard', 14), ('Assire', 28), ('Sheala', 6), ('Riordain', 2), ('Zoltan', 31), ('Percival', 11), ('Munro', 8), ('Figgis', 6), ('Emiel', 9), ('Brouver', 2), ('Aubry', 6), ('Daniel', 7), ('Fringilla', 27), ('Merlin', 4), ('Albrich', 2), ('Stella', 3), ('Ida', 17), ('Adalia' 7), ('Connor', 2), ('Dyffryn', 4), ('Zuleyka', 11), ('Sigrdrifa', 7) ('Zyvik', 2), ('Vysogota', 9), ('Kelpie', 16), ('Carthia', 2), ('Joanna', 2), ('Leo', 6), ('Windsor', 5), ('Stavro', 5), ('Golan', 1), ('Julia', 20), ('Radovid', 8), ('Fergus', 1), ('Lebioda', 4), ('Jacob', 1), ('Hjalmar', 6), ('Milo', 3), ('Arthur', 2), ('Nimue', 12), ('Reynart', 12), ('Reynard', 1), ('Palmerin', 2), ('Fulko', 1), ('Joachim', 6), ('Eredin', 7), ('Auberon', 5), ('Ihuarraquax', 3), ('Milton', 1), ('Adam', 5), ('Puttkammer', 4), ('Barclay', 5), ('Ramon', 1), ('Isengrim', 2), ('Kobus', 1), ('Elan', 1), ('Sievers' 3), ('Aillil', 1), ('Braathens', 1), ('Cyrus', 5), ('Ambassador', 2) ('Lucienne', 3), ('Eudora', 3), ('Alzur', 2), ('Lvtta', 4)]



density: 0.05160816925522808 transitivity: 0.25670526709002617 reciprocity: 0.0

dict (degree), values

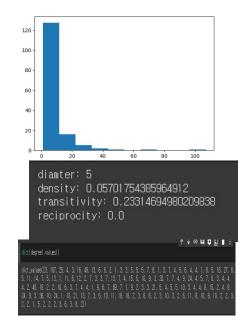
node와 연결된 모든 edge의 수인 degree를 통해 각 인물당 다른 인물들과의 연결 정도를 파악할 수 있었다. degree distribution을 확인한 결과 대부분의 node들의 edge는 0-40사이에 분포해 있는 것을 알 수 있고, 이를 초과하는 node들의 경우 시리즈 내에서 상당히 영향력이 있는 인물들 즉 주인공 혹은 주인공에 준하는 인물들임을 예상할 수 있다. edge가 가장 많은 node들의 경우 그 수가 137,72,69,62로 다른 node들보다 edge수가 특별히 많음을 알 수 있다.

네트워크에서 가장 멀리 연결된 pair의 거리인 diameter 는 5로 edge수가 많은 것에 비해 diameter는 상대적으로 작음을 알 수 있다. 따라서 해당 node는 시리즈중 인물들과 큰 관계가 없는 주변 인물로 파악된다. density의 경우 0.0516으로 낮음을 알 수 있다. 이는 node들간 연결이 활발하지 않음을 의미한다. 따라서 edge수가 가장 많은 인물들을 제외한다면 다른 인물들간 연결이 크게 활성화 되지 않음을 의미한다. transitivity의 경우 0.257로 크게 높지 않음을 알 수 있다. 이는 네크워크에서 3이상의 관계에서 연결된 빈도가 크게 높지 않음을 알 수 있다. 이는 인물들이 군집을 형성하는 빈도가 낮다고 해석할 수 있겠다.

해당 데이터는 undirected data로 reciprocity 는 0이다

2.Data half

[('Velerad', 3), ('Geralt', 107), ('Foltest', 25), ('Adda', 4), ('Ostrit', 3), ('Two', 16), ('Yennefer', 48), ('Nenneke', 13), ('Iola', 6), ('Roach', 6), ('Nivellen', 2), ('Fenne', 1), ('Asse', 3), ('Vereena', 3), ('Hereward', 5), ('Falwick', 5), ('Caldemeyn', 5), ('Stregobor', 7), ('Count', 8), ('Eltibald', 1), ('Fredefalk', 3), ('Renfri', 7), ('Nohorn', 4), ('Tavik', 5), ('Civril', 6), ('Fifteen', 4), ('Audoen', 4), ('Libushe', 1), ('Baron', 8), ('Haxo', 5), ('Pavetta', 18), ('Calanthe', 27), ('Rainfarn', 8), ('Windhalm', 5), ('Eist', 11), ('Crach', 14), ('Draig', 7), ('Drogodar', 5), ('Vissegerd', 13), ('Ravix', 1), ('Mousesack', 11), ('Roegner', 6), ('Sir', 12), ('Rumplestelt', 2), ('Akerspaark', 7), ('Nettly', 3), ('Dhun', 3), ('Lille', 7), ('Little', 13), ('Torque', 7), ('Galarr', 4), ('Aen', 15), ('Vanadain', 5), ('Toruviel', 10), ('Filavandrel', 9), ('Field', 3), ('King', 20), ('Chireadan', 7), ('Errdil', 7), ('Vratimir', 4), ('Beau', 9), ('Triss', 24), ('Laurelnose', 4), ('Neville', 5), ('Eskel', 7), ('Vesemir', 6), ('Dennis', 3), ('Istredd', 4), ('Herbolth', 4), ('Seven', 4), ('Vespula', 2), ('Dandelion', 40), ('Gar', 18), ('Dudu', 2), ('Vimme', 2), ('Sword', 10), ('Agloval', 6), ('Teleri', 3), ('Essi', 7), ('Duke', 4), ('Zelest', 4), ('Veverka', 1), ('Braenn', 9), ('Freixenet', 6), ('Eithné', 7), ('Ciri', 69), ('Venzlav', 7), ('Guy', 1), ('Yurga', 9), ('Visenna', 2), ('Sheldon', 3), ('Donimir', 3), ('Rience', 21), ('Lambert', 5), ('Vanielle', 4), ('Ori', 5), ('Forest', 5), ('Yarpen', 10), ('Xavier', 3), ('Regan', 4), ('Yannick', 4), ('Paulie', 8), ('Henselt', 15), ('Aelirenn', 2), ('Everett', 4), ('Shani', 8), ('Philippa', 24), ('Meve', 9), ('Ethain', 3), ('Emhyr', 38), ('Ithlinne', 10), ('Vilgefortz', 24), ('Ekkehard', 1), ('Artaud', 10), ('Tissaia', 21), ('Lydia', ('Jarre', 7), ('Jan', 3), ('Blue', 6), ('Aplegatt', 10), ('Demavend', 11), ('Codringher', 18), ('Sabrina', 18), ('Horm', 2), ('Yaevinn', 3), ('Molnar', 6), ('Fabio', 6), ('Anna', 2), ('Margarita', 3), ('Falka', 10), ('Rayla', 3), ('Bran', 2), ('Esterad', 6), ('Francesca', 11), ('Carduin', 8), ('Marti', 10), ('Dorregaray', 8), ('Keira', 10), ('Leticia', 2), ('Morvran', 2), ('Fauve', 3), ('Bronibor', 2), ('Menno', 2), ('Ardal', 1), ('Vattier', 5), ('Stefan', 2), ('Xarthisius', 2), ('Ceallach', 2), ('Cahir', 3), ('Kayleigh', 6), ('Iskra', 3), ('Mistle', 8), ('Rudiger', 2)]



data_half의 경우 data_full와 마찬가지로 degree distribution을 확인한 결과 대부분의 node들의 edge는 0-40 사이에 분포해 있는 것을 알 수 있고 , 이를 초과하는 node들의 경우 data_full과 마찬가지로 시리즈 내에서 상당히 영향력이 있는 인물들로 파악된다. 다만 크기가 큰 edge들의 경우 107,69,48인 것을 보아 시즌2에서 주인공들이 더 많은 인물들을 만나며 시즌1에 비해 영향력이 큰 node가 한개 적은 것을 보아 시즌2에서 새로운 영향력이 큰 인물의 등장 또는 기존 인물중 한 인물의 비중이 커진 것임을 예상해볼 수 있다.

diameter는 5로 data_full과 마찬가지로 해당 node는 시리즈중 인물들과 큰 관계가 없는 주변 인물로 파악된다. density의 경우도 역시 data_full과 마찬가지로 0.057로 낮고 이는 node들간 연결이 활발하지 않음을 의미하며 영향력이 큰 인물들을 제외하면 다른 인물들 간 연결이 크게 활성화 되지 않음을 알 수 있다. transitivity의 경우 0.233으로 마찬가지로 크게 높지 않으며 네트워크에서 3이상의 관계에서 연결된 빈도가 크게 높지 않음을 알 수 있고 이는 인물들이 군집을 형성하는 빈도가 낮다고 해석할 수 있다.

네트워크 중심성 분석

시즌1, 소설 1,2,3,4권에 해당하는 네트워크 분석 시즌1, 시즌2, 소설 전권에 해당하는 네트워크 분석

```
degree centrality:
  [('Geralt', 0.7039473684210525), ('Ciri', 0.4539473684210526), ('Yennefer', 0.3157894736842105)]
betweenness centrality:
  [('Geralt', 0.5177182781865106), ('Ciri', 0.18284084725903282), ('Emhyr', 0.1060502338533477)]
closeness centrality:
  [('Geralt', 0.7676767676767676), ('Ciri', 0.63333333333333), ('Yennefer', 0.5846153846153846)]
eigenvector centrality:
  [('Geralt', 0.3833791345535262), ('Ciri', 0.3063770518374445), ('Yennefer', 0.2497406642222411)]
```

```
degree centrality:
  [('Geralt', 0.6199095022624435), ('Ciri', 0.46153846153846156), ('Yennefer', 0.3257918552036199)]
betweenness centrality:
  [('Geralt', 0.3917440422628955), ('Ciri', 0.14917506159328225), ('Dandelion', 0.07982894426906378)]
closeness centrality:
  [('Geralt', 0.7196697068403909), ('Ciri', 0.638728323699422), ('Dandelion', 0.5800524934383202)]
eigenvector centrality:
  [('Geralt', 0.302577502003062), ('Ciri', 0.28564285851311527), ('Yennefer', 0.22734753044606196)]
```

degree centrality 분석을 보면 'Geralt'가 시즌1, 2 모두에서 가장 높은 값을 보였고 이는 모든 시즌을 통틀어 'Geralt'가 중심 인물로서 모든 관계의 중심에 있다고 분석된다.

betweenness centrality 분석을 보면 여전히 'Geralt'가 가장 높은 값을 가지지만 시즌 2에서 그 값이 낮아진 것을 확인할 수 있다. 이는 시즌 1보다 시즌 2에서 'Geralt'가 중계자 노드로서의 영향이 감소했다고 분석 되는데 이것은 새롭게 등장한 'Dandelion'이 그 영향력을 가져간 것으로 분석해볼 수 있다. 혹은 시즌 2로 이야기가 진행됨에 따라 새로운 인물들이 등장하면서 네트워크의 양상이 바뀌었다고도 분석해볼 수 있다.

closeness centrality 분석을 보면 다른 인물들은 시즌1, 2간의 차이가 미약하지만 'Vennefer' 대신 'Dandelion'이 높은 closeness centrality 값을 가지는 것을 볼 수 있다. degree centrality는 'Vennefer'가 더 높은 값을 가지는 것으로 보아 내용의 핵심 전개를 담당하기 보다는 인물간의 연결에 초점을 두고 만들어진 친화력 좋은 인물일 가능성이 있다.

eigenvector centrality 분석은 대체적으로 비슷한 값을 가진 것을 볼 수 있다. 이것은 'Geralt', 'Cirl', 'Vennefer'가 서로 비슷한 노드들과 연결되어 있어서 평균적으로 값이 비슷해졌을 가능성도 있지만 소설이 하나의 주인공만을 강조하여 진행되는 것이 아니라 3명의 인물이 모두 주인공으로서 각각 큰 비중을 갖는 형태로 진행된다고도 분석해볼 수 있다. 또한 새로운 인물 'Dandelion'이 높은 값을 가지지 않는 것으로 보아 소설의 핵심 인물보다는 보조적인 친화력 좋은 인물일 가능성이 더 높을 것으로 분석된다.

한계점 및 개선안

한계점

- 본 연구는 드라마 <위쳐> 시청자들의 이해를 돕기 위한 목적을 가지고 진행되었다. 반면, 분석에 활용한 데이터는 원작 소설의 내용을 바탕으로 추출한 데이터이기에 소설과 드라마 사이의 세부적인 차이점은 반영하지 못한다는 한계가 있다.
- 소설 텍스트 데이터에서 추출한 데이터이기에, 표면상으로 드러나지 않고 전반적인 내용 속에 함의되어있는 관계를 표현하지 못했을 가능성이 있다.
- SNA는 일반적으로 해석 과정이 중요하기에, 소설 <더 위쳐시리즈> 또는 드라마 <위쳐>에 보다 많은 지식과 이해를 가진 상태가 아니었기에 심층적인 분석에 한계가 있었다.
- 중심성 분석의 특성상, 중심인물 몇 명 위주로 분석이 진행되었다는 한계가 있다.

개선안

- 드라마 <위쳐>의 대본 데이터를 이용해 분석을 진행해본다면 소설과 드라마의 차이점을 확인해볼 수 있을 것이다.
- 소설 <더 위쳐시리즈> 또는 드라마 <위쳐>에 대한 높은 이해도를 가진 이들의 의견을 구할 수 있다면 보다 심층적인 분석이 가능할 것이다.
- 중심 인물들을 제거하고 네트워크 분석을 새롭게 진행해보는 과정을 통해, 차순위로 중요한 인물들의 관계에 대해서도 파악해볼 수 있을 것이다.

creative application areas

- 소설의 사회 연결망 분석 결과를 추후 텍스트 분석 인공지능이 소설의 내용을 정확히 파악할 수 있도록 적용하여 더욱 효과적인 학습 효과를 기대해 볼 수 있다.
- 관련하여 확장한다면, 특정 소설 텍스트를 입력하면 해당 소설 등장인물들의 네트워크 그림을 보여주고 분석해주는 인공지능 프로그램도 개발할수 있으리라 기대한다. 이러한 프로그램은 장편소설이나 여러 개의 시즌을 가지는 드라마 등을 효과적으로 이해하는 데에 도움을 줄 수 있을 것이다.

Reference

Sadasivan, Ava. "Witcher Network." Kaggle, 14 Jan. 2022, https://www.kaggle.com/datasets/avasadasivan/witcher-network.

Avasadasivan. "Witcher Network Analysis." Kaggle, Kaggle, 15 Jan. 2022, https://www.kaggle.com/code/avasadasivan/witcher-network-analysis.