

A network analysis of

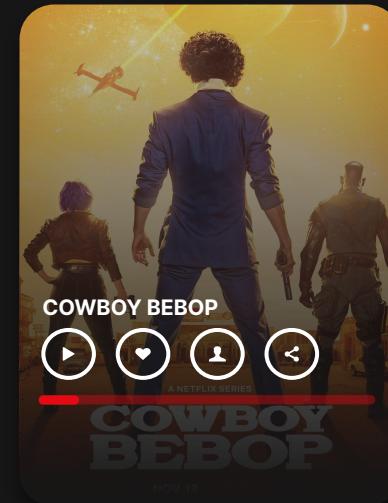
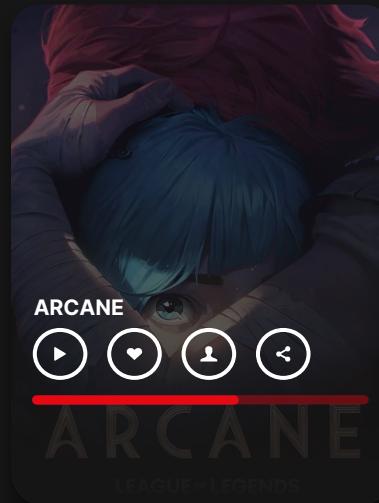
**NETFLIX**

**Actors/Directors and Movies**

Simrat Benipal, Jose Perales, Ishrat Naba  
Group # 11  
Netflix II

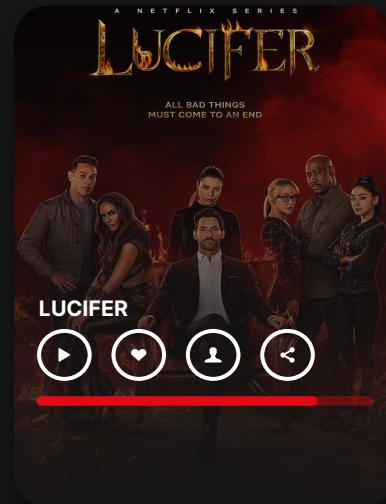
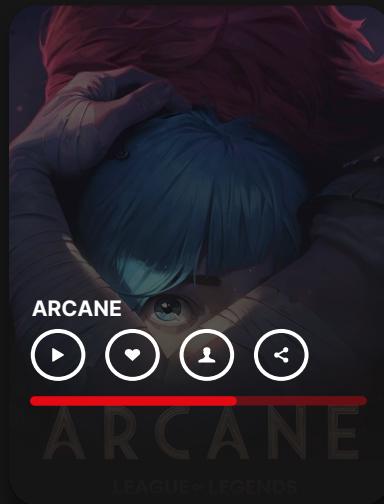
# Nodes, Links

- Nodes: People (Cast, Directors) and Movies
  - Total Number of Nodes = **35,133**
- Links: Represent the involvement of people in a Movie
  - Total Number of Links = **49,577**

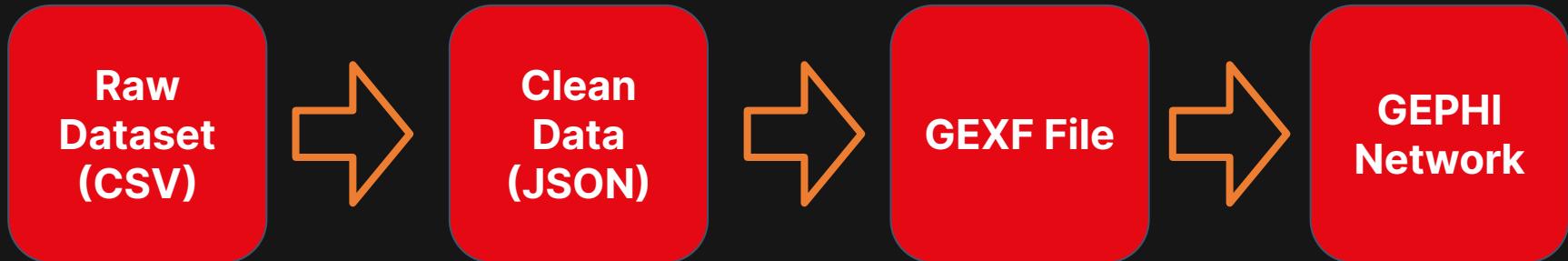


# DATASET

- Dataset link: [Netflix Movies and TV Shows](#) (last updated 2021)
- Contains data on all TV Shows and Movies.



# **DATASET → NETWORK**



- GEXF (Graph Exchange XML Format), type of XML-based format for storing graphs.
- These files are used by GEPHI to render graphs and create adequate visualizations.

# Raw Data (CSV)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
1	show_id	type	title	director	cast	country	date_added	release_ye	rating	duration	listed_in	description											
2	s1	Movie	Dick Johnson	Kirsten Johnson		United States	25-Sep-21	2020	PG-13	90 min	Documentary	As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help him say his final words.											
3	s2	TV Show	Blood & Water	Ama Qama	South Africa	24-Sep-21	2021	TV-MA	2 Seasons	International	After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her sister or not.												
4	s3	TV Show	Ganglands	Julien Leclerc	Sami Bouajila, Tracy G	24-Sep-21	2021	TV-MA	1 Season	Crime TV	S To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers are pulled into a violent crime.												
5	s4	TV Show	Jailbirds	New Orleans			24-Sep-21	2021	TV-MA	1 Season	Docuseries	Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans Justice Center in New Orleans.											
6	s5	TV Show	Kota Factory	Mayur Moi	India	24-Sep-21	2021	TV-MA	2 Seasons	International	In a city of coaching centers known to train India's finest collegiate minds, an earnest but unexceptional student and his friends must prove themselves to be the best.												
7	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford	24-Sep-21	2021	TV-MA	1 Season	TV Dramas	The arrival of a charismatic young priest brings glorious miracles, ominous mysteries and renewed religious fervor to a small town.												
8	s7	Movie	My Little Pony	Robert Cullinan	Vanessa Hudgens, Kim	24-Sep-21	2021	PG	91 min	Children & Equestria	Children & Equestria's divided. But a bright-eyed hero believes Earth Ponies, Pegasi and Unicorns should be pals after a hoof to the head.												
9	s8	Movie	Sankofa	Haille Gerin	Kofi Ghana	United States	24-Sep-21	1993	TV-MA	125 min	Dramas, In-On	A photo shoot in Ghana, an American model slips back in time, becomes enslaved on a plantation and bears witness to a secret.											
10	s9	TV Show	The Great	Andy Devine	Mel Giedroyc	United Kingdom	24-Sep-21	2021	TV-14	9 Seasons	British TV	A talented batch of amateur bakers face off in a 10-week competition, whipping up their best dishes in the hopes of being crowned champion.											
11	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy	United States	24-Sep-21	2021	PG-13	104 min	Comedies,	A woman adjusting to life after a loss contends with a feisty bird that's taken over her garden and a husband who's still in love with her.											
12	s11	TV Show	Vendetta: Truth, Lies & The Mafia				24-Sep-21	2021	TV-MA	1 Season	Crime TV	S Sicily boasts a bold "Anti-Mafia" coalition. But what happens when those trying to bring down organized crime are accused of being the ones who started it?											
13	s12	TV Show	Bangkok Bi	Kongkiat K	Sukollawat Kanarot	St	23-Sep-21	2021	TV-MA	1 Season	Crime TV	Struggling to earn a living in Bangkok, a man joins an emergency rescue service and realizes he must unravel a citywide mystery.											
14	s13	Movie	Je Suis Karl	Christian Siriano	Luna Wedli	Germany	23-Sep-21	2021	TV-MA	127 min	Dramas,	After most of her family is murdered in a terrorist bombing, a young woman is unknowingly lured into joining the very group responsible.											
15	s14	Movie	Confession	Bruno G	Klara Castanho	Lucca	22-Sep-21	2021	TV-PG	91 min	Children &	When the clever but socially-awkward Tet joins a new school, she'll do anything to fit in. But the queen bee among her peers has other ideas.											
16	s15	TV Show	Crime Stories: India Detectives				22-Sep-21	2021	TV-MA	1 Season	British TV	Cameras following Bengaluru police on the job offer a rare glimpse into the complex and challenging inner workings of law enforcement.											
17	s16	TV Show	Dear White People	Logan Browning	United States	22-Sep-21	2021	TV-MA	4 Seasons	TV Comedies	Students of color navigate the daily slights and slippery politics of life at an Ivy League college that's not nearly as "post-racial" as it claims.												
18	s17	Movie	Europe's	M Pedro de Echave	García, Pablo A	22-Sep-21	2020	TV-MA	67 min	Documentary	Declassified documents reveal the post-WWII life of Otto Skorzeny, a close Hitler ally who escaped to Spain and became a spy.												
19	s18	TV Show	Falsa Identidad	Luis Ernesto	Mexico	22-Sep-21	2020	TV-MA	2 Seasons	Crime TV	S Strangers Diego and Isabel flee their home in Mexico and pretend to be a married couple to escape his drug-dealing enemies.												
20	s19	Movie	Intrusion	Adam Salter	Freida Pinto, Logan Miller	22-Sep-21	2021	TV-14	94 min	Thrillers	After a deadly home invasion at a couple's new dream house, the traumatized wife searches for answers and leads the investigation.												
21	s20	TV Show	Jaguar	Blanca Suárez, Iván	Álvarez	22-Sep-21	2021	TV-MA	1 Season	International	In the 1960s, a Holocaust survivor joins a group of self-trained spies who seek justice against Nazis fleeing to Spain to hide.												
22	s21	TV Show	Monsters	I	Olivier Megaton	22-Sep-21	2021	TV-14	1 Season	Crime TV	S In the late 1970s, an accused serial rapist claims multiple personalities control his behavior, setting off a legal odyssey that will change his life forever.												
23	s22	TV Show	Resurrection: Ertugrul	Engin Altar	Turkey	22-Sep-21	2018	TV-14	5 Seasons	International	When a good deed unwittingly endangers his clan, a 13th-century Turkish warrior agrees to fight a sultan's enemies in exchange for his son's freedom.												
24	s23	Movie	Avvai Shan	K.S. Ravikumar	Hassan, Meena	21-Sep-21	1996	TV-PG	161 min	Comedies,	Newly divorced and denied visitation rights with his daughter, a doting father disguises himself as a gray-haired nanny in order to see her again.												
25	s24	Movie	Go! Go! Co	Alex Woo, Maisie Benson	Paul Ki	21-Sep-21	2021	TV-Y	61 min	Children &	From arcade games to sled days and hiccup cures, Cory Carson's curious little sister Chrissy speeds off on her own fo												
26	s25	Movie	Jeans	S. Shankar	Prashanth, India	21-Sep-21	1998	TV-14	166 min	Comedies,	When the father of the man she loves insists that his twin sons marry twin sisters, a woman creates an alter ego that might just work.												
27	s26	TV Show	Love on the Spectrum	Brooke Sat	Australia	21-Sep-21	2021	TV-14	2 Seasons	Docuseries	Finding love can be hard for anyone. For young adults on the autism spectrum, exploring the unpredictable world of dating is even more challenging.												
28	s27	Movie	Minsara	K Rajiv	Menc Arvind Swamy, Kajol	21-Sep-21	1997	TV-PG	147 min	Comedies,	A tangled love triangle ensues when a man falls for a woman studying to become a nun and she falls for the friend he left behind.												

- Remove TV Shows
- Remove irrelevant fields (show\_id, country, description)
- Separate directors and cast to create individual nodes

# Data Cleaning

```
def parseData(input_csv_path: str) -> list[dict[str, str]]:  
    clean_data: list[dict[str, str]] = []  
    tv_show_counter: int = 0  
  
    with open(input_csv_path, newline='') as csvfile:  
        csvreader = csv.DictReader(csvfile)  
  
        for row in csvreader:  
            # Skip TV Shows  
            if row['type'] == 'TV Show':  
                tv_show_counter += 1  
                continue  
  
            # Delete unnecessary fields  
            for key in ['show_id', 'type', 'country', 'description']:  
                del row[key]  
  
            clean_data.append(row)  
    return clean_data  
  
def outputDataAsJSON(output_path: str, data: list) -> None:  
    with open(output_path, 'w') as jsonfile:  
        jsonfile.write('[' + '\n'  
  
        for i, entry in enumerate(data):  
            # Convert dictionary to JSON string  
            json_string: str = json.dumps(entry, indent=4)  
            jsonfile.write(json_string)  
  
            # Commas between dictionaries, don't add a comma after last one  
            jsonfile.write(',') if i < len(data) - 1 else ''  
  
    jsonfile.write('\n]' + '\n')
```

```
1  [  
2  {  
3  ... "title": "Dick Johnson Is Dead",  
4  ... "director": "Kirsten Johnson",  
5  ... "cast": "",  
6  ... "date_added": "September 25, 2021",  
7  ... "release_year": "2020",  
8  ... "rating": "PG-13",  
9  ... "duration": "90 min",  
10 ... "listed_in": "Documentaries"  
11 },  
12 {  
13 ... "title": "My Little Pony: A New Generation",  
14 ... "director": "Robert Cullen, Jos\u00e3o Luis Ucha",  
15 ... "cast": "Vanessa Hudgens, Kimiko Glenn, James Marsden, Sofia Carson, Li  
16 ... "date_added": "September 24, 2021",  
17 ... "release_year": "2021",  
18 ... "rating": "PG",  
19 ... "duration": "91 min",  
20 ... "listed_in": "Children & Family Movies"  
21 },  
22 {  
23 ... "title": "Sankofa",  
24 ... "director": "Haile Gerima",  
25 ... "cast": "Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick Medley",  
26 ... "date_added": "September 24, 2021",  
27 ... "release_year": "1993",  
28 ... "rating": "TV-MA",  
29 ... "duration": "125 min",  
30 ... "listed_in": "Dramas, Independent Movies, International Movies"  
31 },  
32 {  
33 ... "title": "The Starling",  
34 ... "director": "Theodore Melfi",  
35 ... "cast": "Melissa McCarthy, Chris O'Dowd, Kevin Kline, Timothy Olyphant",  
36 ... "date_added": "September 24, 2021",  
37 ... "release_year": "2021",  
38 ... "rating": "PG-13",  
39 ... "duration": "104 min",  
40 ... "listed_in": "Comedies, Dramas"  
41 },  
42 {  
43 ... "title": "Je Suis Karl"
```

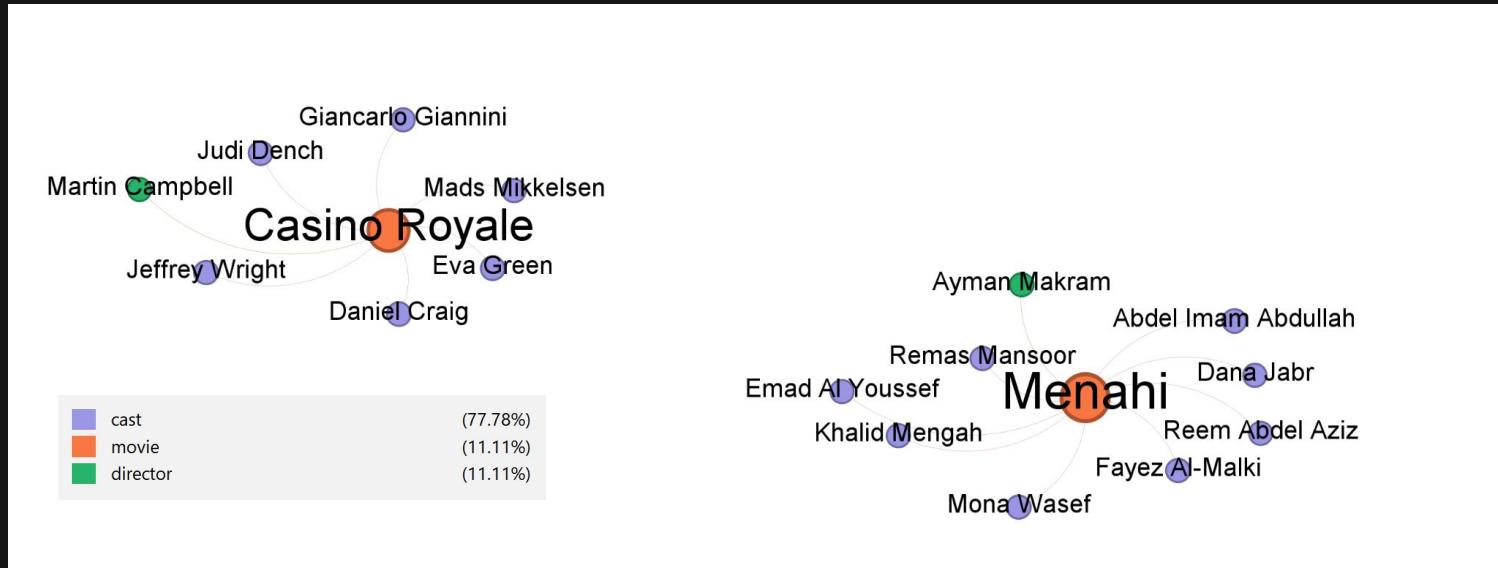
# GEXF File Creation

```
def createGraph(data: list[dict[str, str]], weights: list[list[int]]) -> list[nx.Graph]:  
    outcomes: list = ['IN', 'OUT']  
    created_graphs: list = []  
  
    for prob in weights:  
        G = nx.Graph()  
  
        for entry in data:  
            G.add_node(entry['title'], type='movie')  
  
            # Add director nodes and movie-director edges  
            for director in entry['director'].split(', '):  
                if director != '' and random.choices(outcomes, prob, k=1)[0] == 'IN':  
                    G.add_node(director, type='director')  
                    G.add_edge(entry['title'], director)  
  
            # Add cast nodes and movie-cast edges  
            for cast_member in entry['cast'].split(', '):  
                if cast_member != '' and random.choices(outcomes, prob, k=1)[0] == 'IN':  
                    G.add_node(cast_member, type='cast')  
                    G.add_edge(entry['title'], cast_member)  
  
        nx.write_gexf(G, f'output/network_{prob[0]}%.gexf')  
        created_graphs.append(G)  
  
    return created_graphs
```

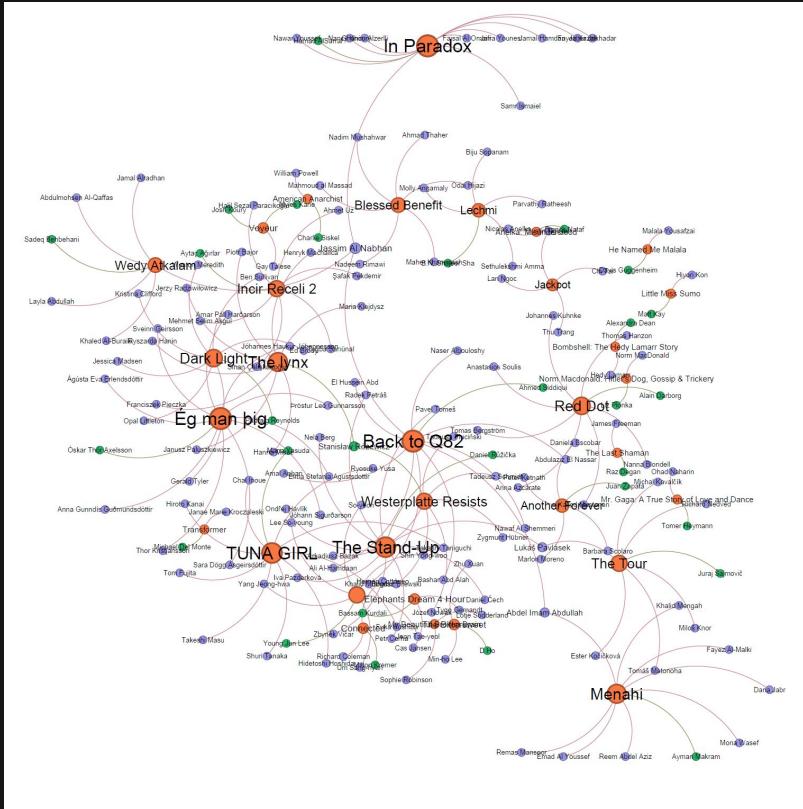
-  network\_100%.gexf
-  network\_50%.gexf
-  network\_60%.gexf
-  network\_70%.gexf
-  network\_80%.gexf

# Network Visualizations (GEPHI)

- **Color** is used to differentiate between nodes
  - Movies → Orange
  - Actors → Violet
  - Directors (green)
- **Size** is used to represent the degree of the node.
  - Higher Degree ⇒ Higher Collaboration ⇒ Bigger Node

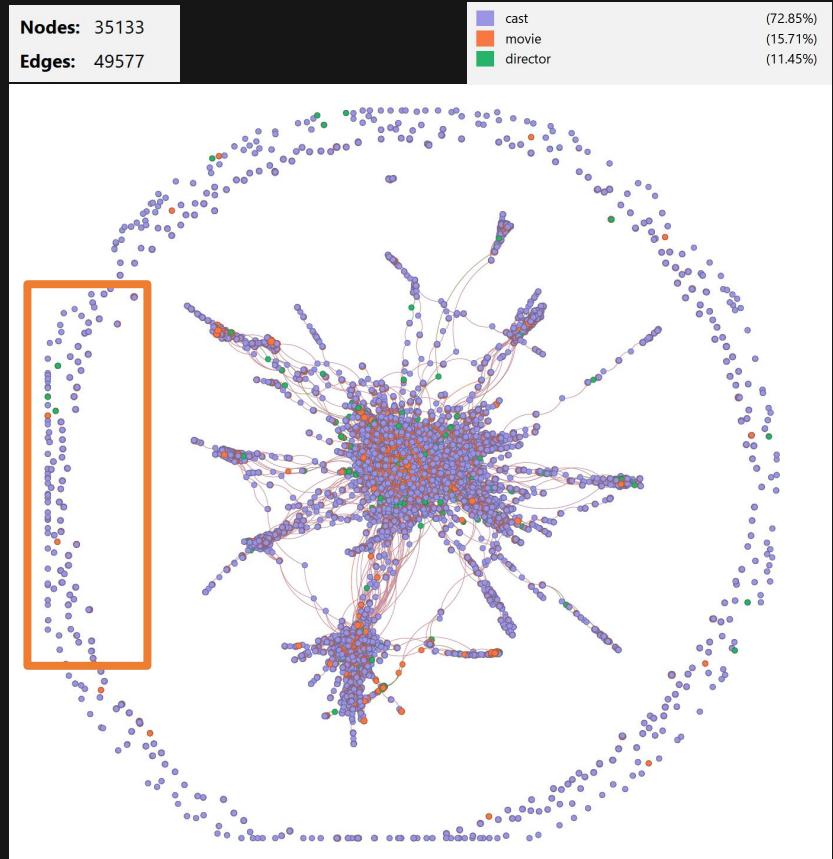


# Network Visualizations



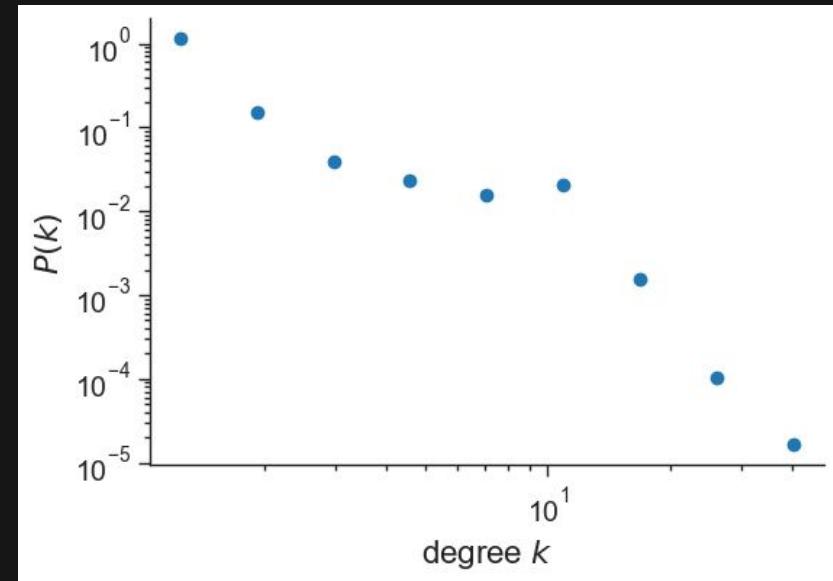
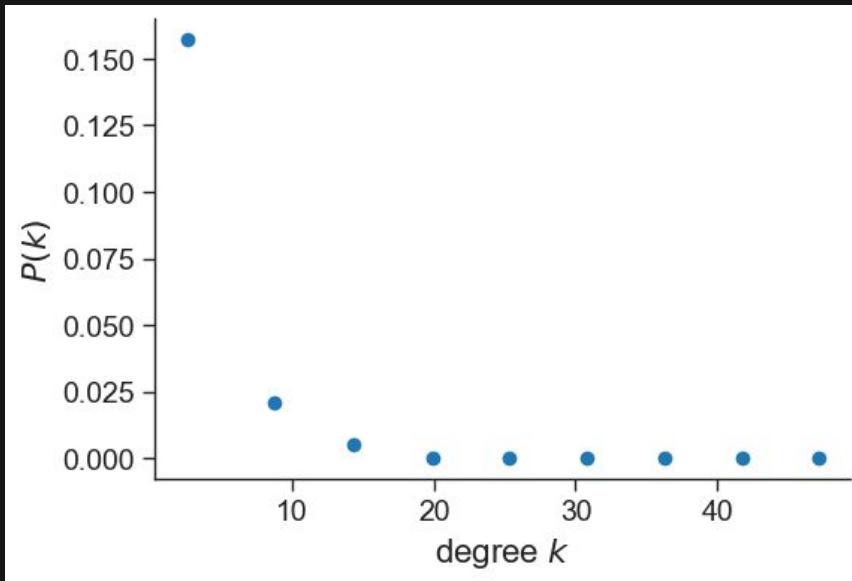
**Nodes:** 35133

**Edges:** 49577



# Degree Distribution

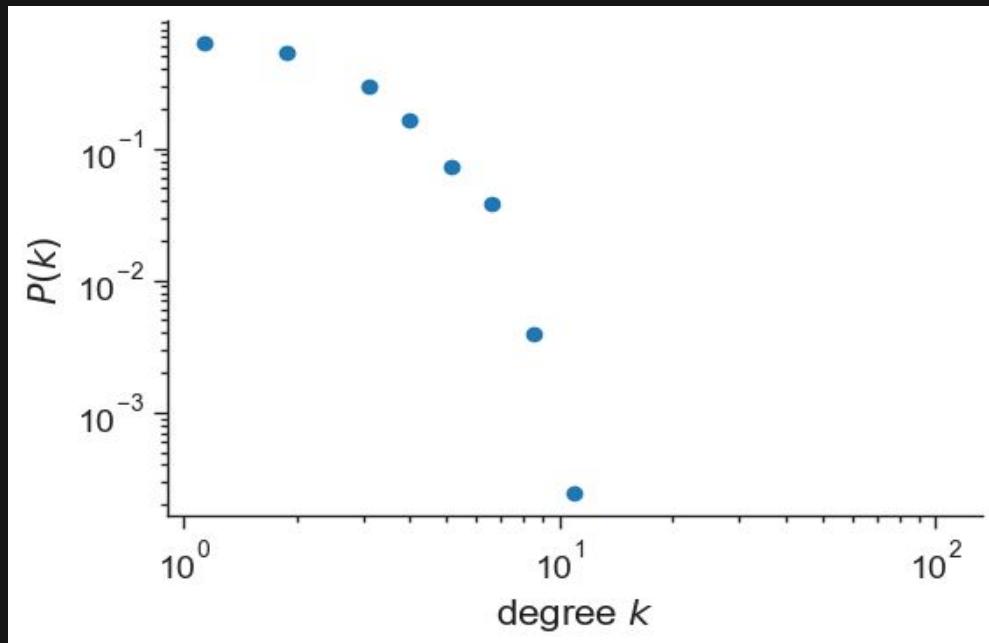
- Minimum degree  $\rightarrow 1$ , Maximum degree: 50
- Average degree  $\rightarrow 2.82$
- Nodes with degree less than 10 are very common.
- Nodes with higher degree, like 30-50, are rare.



# Null Model

Compared to our graph, this null model:

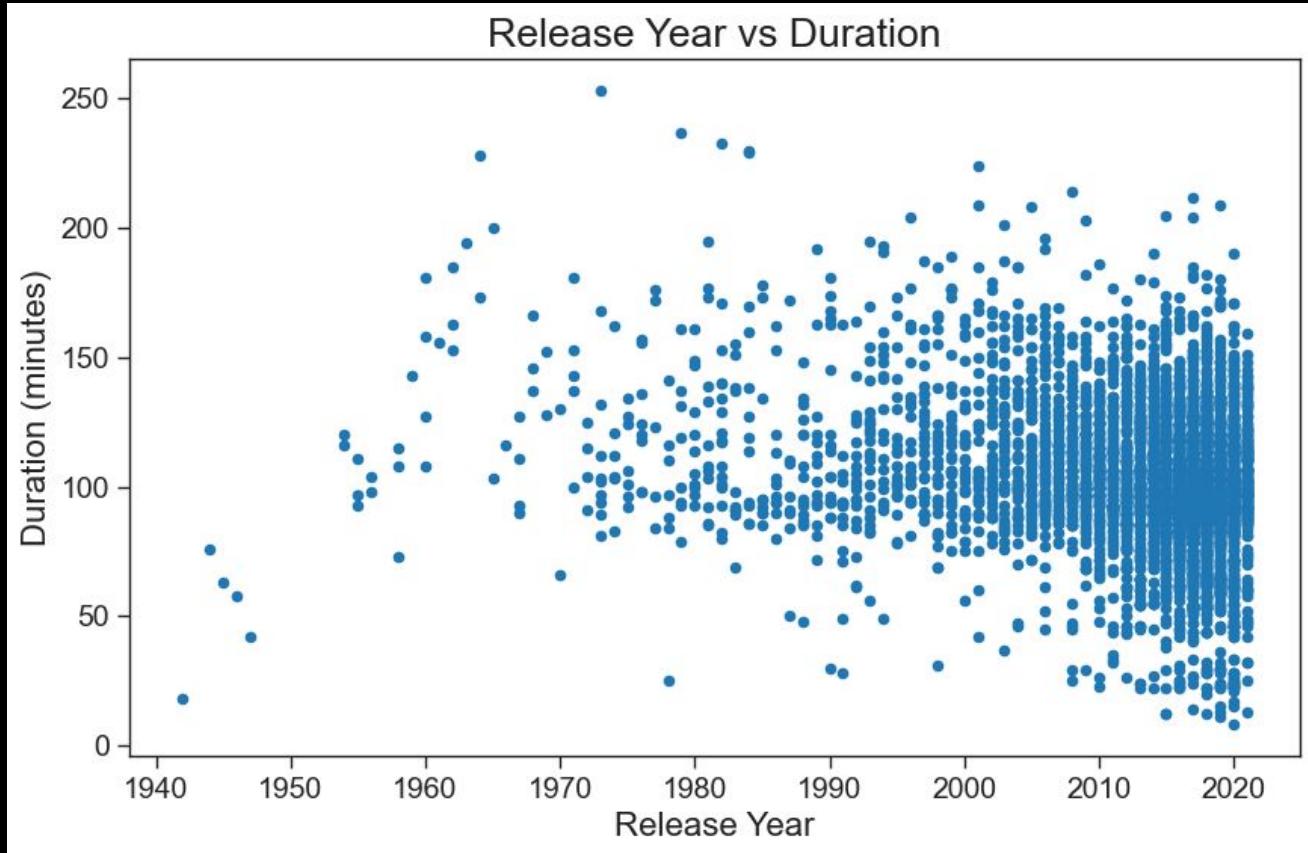
- Has less clusters/preferential attachment, likely due to the randomness of the Erdos-renyi model (tendency to have less attachment)



# Research Questions

- Which TV Ratings, categories, and movie lengths are most prevalent in Netflix?
- Is there a correlation between the duration of a movie, its TV Rating, and its release year?
- Does Netflix have a preference for newer movies as compared to older ones?
- Number of Movies per category (i.e. drama, action)?
- Are there key people acting as a bridge between different clusters of actors/directors?
- Do some directors have a preference for longer movies or for a certain TV Rating?





# THANKS

# REFERENCES

Dataset: [https://www.kaggle.com/datasets/  
shivamb/netflix-shows/data](https://www.kaggle.com/datasets/shivamb/netflix-shows/data)

