Anime Recommendation System

Rohan Bera

17MCME21

Why I chose this topic

- Not many people watch anime, hard to get good
  recommendations

- People who do watch, know only a few basic titles

The problem

If a user provides his watched anime data, can we develop an algorithm to recommend which anime to watch next?

Roadmap

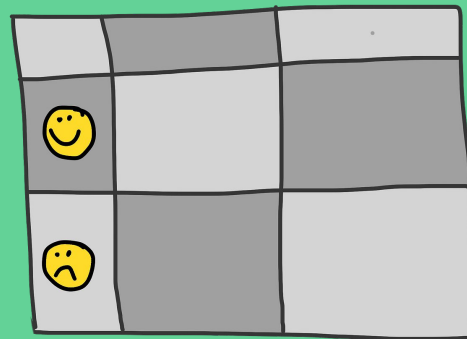Filter and get data from datasets to generate graphs

1

Generate a relationship between the main user and other users based on common anime

2

Award scores to animes. List of anime with highest scores are returned

3

Let's take a look at our dataset.

Ratings data

Filename: rating_complete.csv

This file is used to construct our graph

	user_id	anime_id	rating
0	-1	a1535	8
1	-1	a1575	10
2	-1	a1953	10
3	-1	a15197	9
4	-1	a15199	9
5	-1	a15201	9
6	-1	a20	8
7	-1	a1735	8
8	-1	a32182	8
9	-1	a14308	8

Anime data

Filename: anime.csv

MAL_ID	Name	Community rating	English
a1	Cowboy Bebop	8.78	Cowboy B
a5	Cowboy Bebop: Tengoku no Tobira	8.39	Cowboy Bebop:The M
a6	Trigun	8.24	T
a7	Witch Hunter Robin	7.27	Witch Hunter R
a8	Bouken Ou Beet	6.98	Beet the Vandel B
a15	Eyeshield 21	7.95	Unkr
a16	Hachimitsu to Clover	8.06	Honey and C
a17	Hungry Heart: Wild Striker	7.59	Unkr
a18	Initial D Fourth Stage	8.15	Unkr
a19	Monster	8.76	Mo

Some changes

- I added a prefix 'a' to the anime_ids (and MAL_IDS) so they don't mix up with the user_ids
- I added my personal anime ratings with user_id = -1
- I used only the top 500000 rows as the process is time consuming. (The dataset contains over 57 million rows)
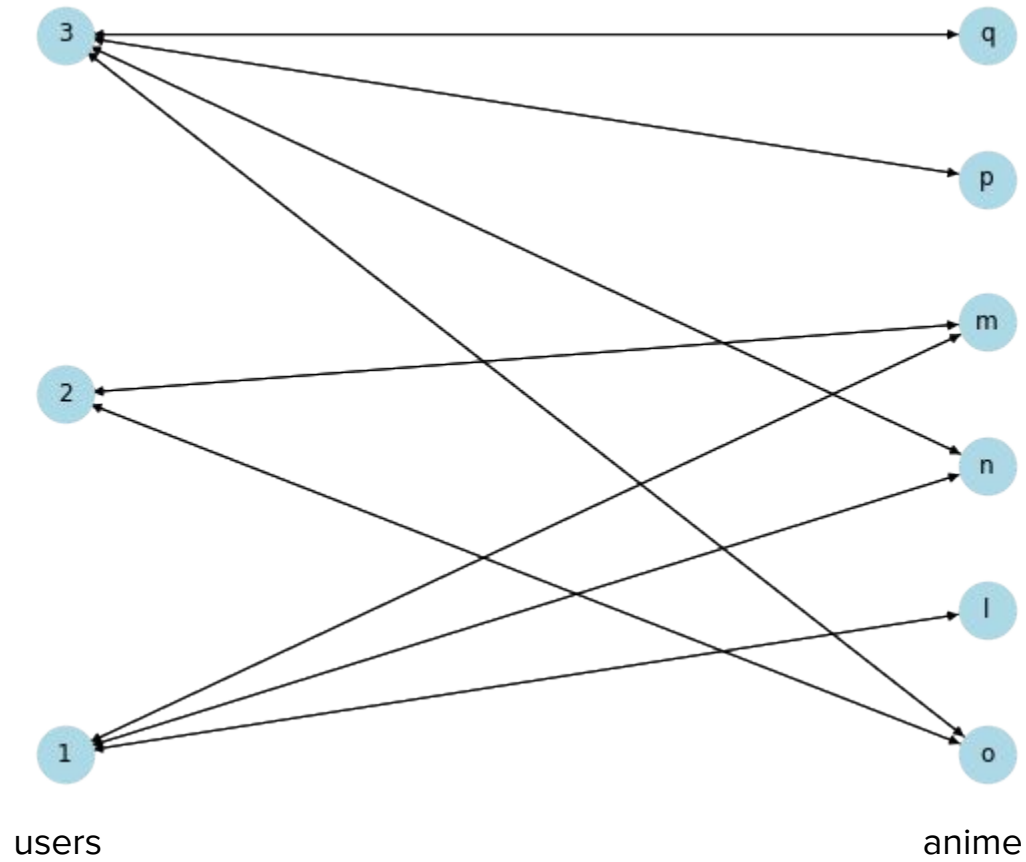
*MAL stands for My Anime List: Worlds biggest anime and manga community
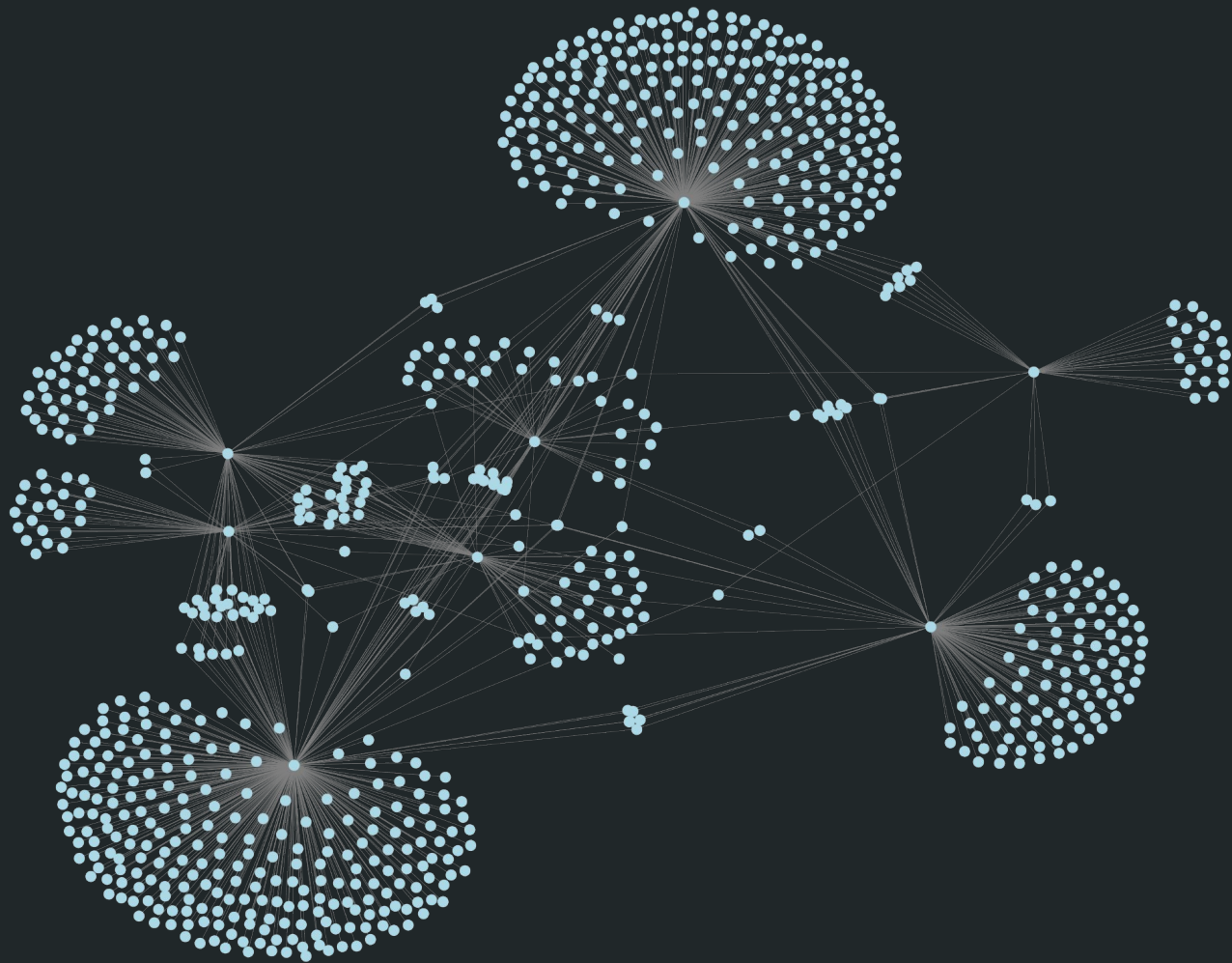
Network description

Original

- Nodes: 16,872 Users, 310,059 Anime
- Edges: 57 million

Edited

- Nodes: 2,825 Users; 11,309 Anime
- Edges: 500,000
- Average degree: 70.7514

users   anime

Roadmap

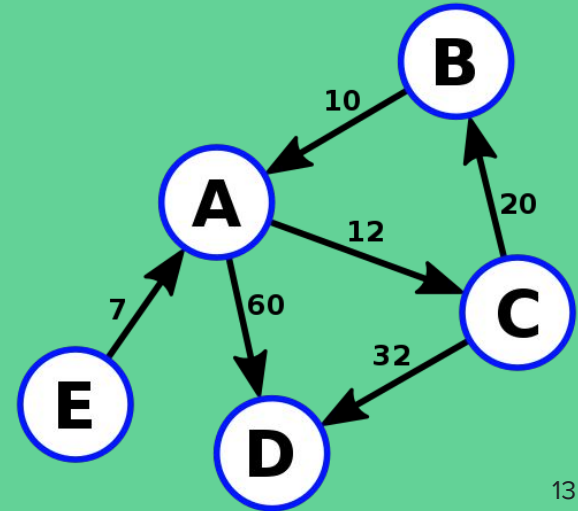Filter and get data from datasets to generate graphs

1

Generate a relationship between the main user and other users based on common anime

2

Award scores to animes. List of anime with highest scores are returned

3

Let's now generate relationships between main user and other users.

- Some users may be highly related with the main user while others might not.
- I decided not to take the "not related" users into consideration for anime suggestions.
- To understand how much the users relate to the main user, I added weighted edges between the main user and other users
- The weights are calculated by the following formula.

$$\text{weight}(uw) = \frac{1}{degree(u)} \sum_{i=1}^{n} \frac{\delta(u,w,i)}{degree(v_i)}$$

$$\text{where } \delta(u,w,i) = \begin{cases} 1 \text{ if } v_i \text{ is a common neighbour} \\ 0 \text{ otherwise} \end{cases}$$

and $v_1...v_n$ is set of all anime neighbours of u
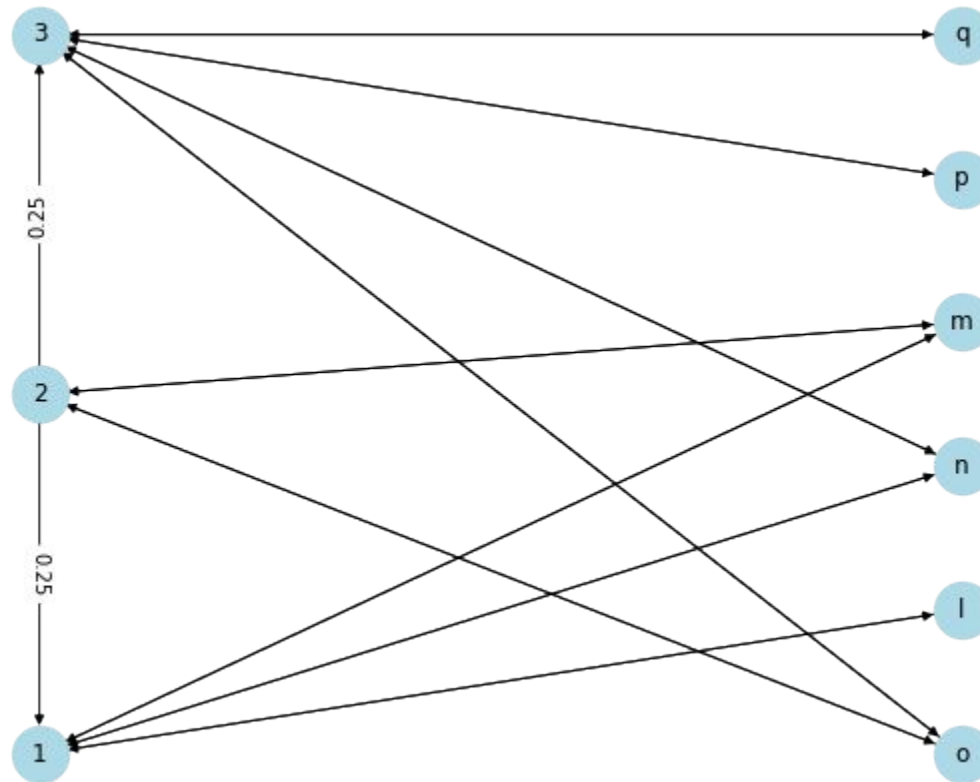
weights among users: {(2, 1): 0.25, (2, 2): 0.5, (2, 3): 0.25}

Roadmap

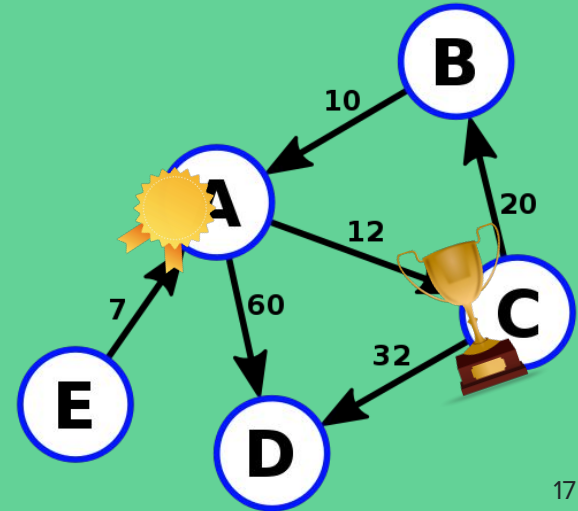Filter and get data from datasets to generate graphs

1

Generate a relationship between the main user and other users based on common anime

2

Award scores to animes. List of anime with highest scores are returned

3

Time to award scores to our anime list.
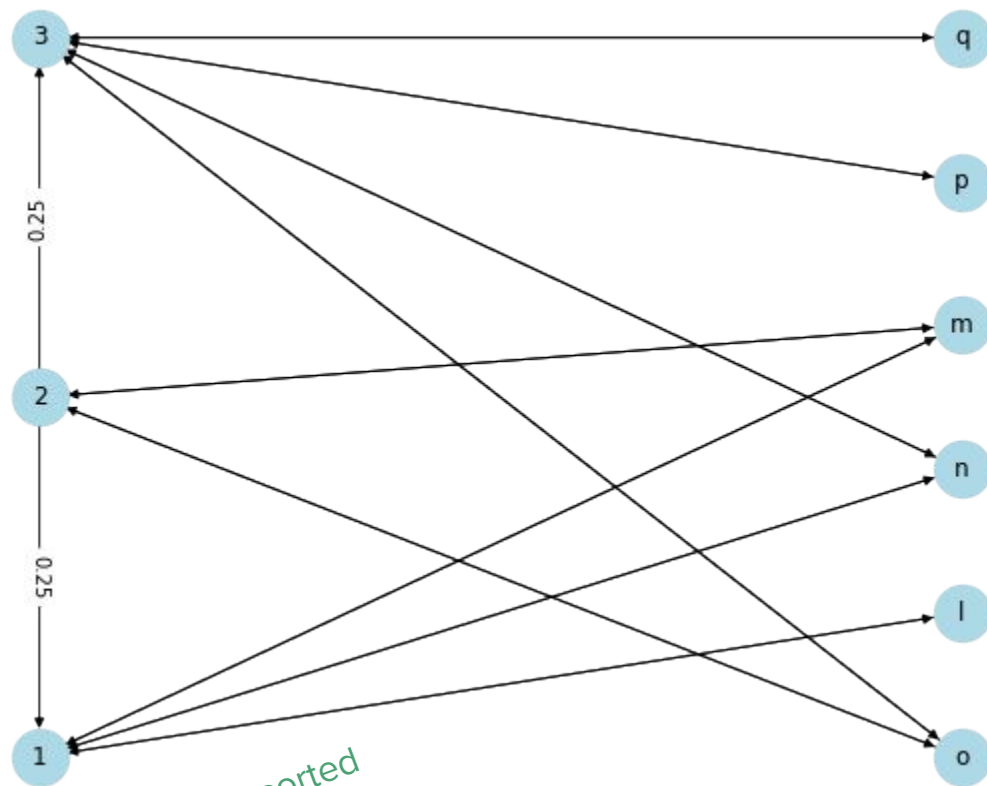
- Collect all animes reviewed by the 'user neighbours' of 'u'. These animes will be scored.
- The scores between the main user 'u' and the animes are awarded by the following formula.

$$\text{score}(uv_i) =$$

$$\sum \frac{weight(uw).(1 - \frac{1}{rating(w,v_i)})}{degree(w)}$$

where $w \epsilon$ neighbours(v_i) and $v_i \epsilon$ animes reviewed by user neighbours of (u)

Recommended anime ids and scores : [('m', 0.266), ('o', 0.25), ('n', 0.1166), ('l', 0.066), ('p', 0.05), ('q', 0.05)]

in sorted order

MAL_ID	Name	English name	Genres	Community rating
a9253	Steins;Gate	Steins;Gate	Thriller, Sci-Fi	9.11
a28171	Shokugeki no Souma	Food Wars! Shokugeki no Soma	Ecchi, School, Shounen	8.29
a11061	Hunter x Hunter (2011)	Hunter x Hunter	Action, Adventure, Fantasy, Shounen, Super Power	9.1
a34599	Made in Abyss	Made in Abyss	Sci-Fi, Adventure, Mystery, Drama, Fantasy	8.74
a37450	Seishun Buta Yarou wa Bunny Girl Senpai no Yum...	Rascal Does Not Dream of Bunny Girl Senpai	Comedy, Supernatural, Drama, Romance, School	8.38
a32281	Kimi no Na wa.	Your Name.	Romance, Supernatural, School, Drama	8.96
a23273	Shigatsu wa Kimi no Uso	Your Lie in April	Drama, Music, Romance, School, Shounen	8.74
a4224	Toradora!	Toradora!	Slice of Life, Comedy, Romance, School	8.24
a31043	Boku dake ga Inai Machi	ERASED	Mystery, Psychological, Supernatural, Seinen	8.37
a37497	Irozuku Sekai no Ashita kara	Iroduku:The World in Colors	Drama, Magic, Romance, School	7.55
a31240	Re:Zero kara Hajimeru Isekai Seikatsu	Re:ZERO -Starting Life in Another World-	Psychological, Drama, Thriller, Fantasy	8.29
a35788	Shokugeki no Souma: San no Sara	Food Wars! The Third Plate	Ecchi, School, Shounen	8.14
a14813	Yahari Ore no Seishun Love Comedy wa Machigatt...	My Teen Romantic Comedy SNAFU	Slice of Life, Comedy, Drama, Romance, School	8.07
a23847	Yahari Ore no Seishun Love Comedy wa Machigatt...	My Teen Romantic Comedy SNAFU TOO!	Slice of Life, Comedy, Drama, Romance, School	8.27
a19815	No Game No Life	No Game, No Life	Game, Adventure, Comedy, Supernatural, Ecchi, ...	8.2
a5081	Bakemonogatari	Bakemonogatari	Romance, Supernatural, Mystery, Vampire	8.36
a32282	Shokugeki no Souma: Ni no Sara	Food Wars! The Second Plate	Ecchi, School, Shounen	8.21
a9919	Ao no Exorcist	Blue Exorcist	Action, Demons, Fantasy, Shounen, Supernatural	7.59
a32188	Steins;Gate: Kyoukaimenjou no Missing Link - D...	Unknown	Sci-Fi, Thriller	8.26
a199	Sen to Chihiro no Kamikakushi	Spirited Away	Adventure, Supernatural, Drama	8.83

Challenges.
Questions.
Conclusion.

Challenges faced

- Reading data (57 million rows)
- Drawing graph for network
- Manually adding user data was tedious
- Networkx does not support mixed graph (bi-directed edges between users and anime isn't actually necessary)
- Community discovery among users was a very difficult task (weighted and directed)
- The algorithm I suggested works when the user provides his past anime data

Questions

Q. How good is network analysis compared to ML?

Ans. Pros.

- It is better explained, so one can trace how one item has the score it has.
- It gave nice recommendations even when data set was small (500 rows).

   Cons.

- It failed to give any recommendations when the main user had no 'user neighbours'.
- Factors like 'genres' were not taken into consideration.

By adding a few more functionalities, these shortcomings can be met. There may be few sophisticated model based recommendation systems that can beat our system. All in all, either of the approaches is largely an academic exercise as real world systems only use the top few results.

Questions

Q. Can this algorithm apply work in other domains?

Ans. Yes.

    For example:

- Product recommendation (dataset: users, products, ratings)
- Personalised advertisements (dataset: customers, services, frequency of visits to particular topic)

Aha!
Conclusions

1. The recommendation system suggested some good animes for me to binge on.

2. This system can further be improved in multiple ways (taking genres into consideration, updating network when a new node is added)

Thank you!

___