

MOVIE CLUB RECOMMENDER

TOM MADSEN
TDI PROJECT PROPOSAL

THE MOVIE CLUB PROBLEM

QMC Rankings

File Edit View Insert Format Data Tools Add-ons Help

Share

	A	B	K	L	M	N	O	P	Q	R	S	T	U	V	W	AN	AO	Std
1	Nominate	Chronology	Ge	Levine	Birnba	Eugen	Tom	Emily	Charlie	Kate	Shane	JL	Armst	Heath	Avg%Ra	Med%Ra	Std	
2	Parasite	Charlie†	1	1	3	1	2	2	1	1	2	1	9	2	94%	98%		
3	Sorry to Bother You	Charlie	2	13	23	33	39	26	28	17	32				18%	9%		
4	The Good, The Bad, and the Ugly	Birnbaum	3	25	24	42	40	44	27		25		5		14%	0%		
5	Luce	Levine	4	14	4	30	38	40	17	19	30		17	11	26%	16%		
6	Portrait of a Lady on Fire	Charlie*	5	16	1	4	1	1	19	2	1	2	13		79%	93%		
7	Howl's Moving Castle	Birnbaum	6	15	21	32	30	28	4	18	10	11			35%	26%		
8	Children of Men	Charlie	7	23	16	15	13	25	26	9	18		1		48%	46%		
9	Melancholia	Birnbaum	8	3		2	35	20	9	7	7				68%	72%		
10	Honey Boy	Charlie	9	6	5	20	6	18	22	5	5	8			66%	78%		
11	You Were Never Really Here	Birnbaum	10		15	12	21	42	25	8	8				46%	49%		
12	Palm Springs	Tom	11	10	12	25	12	19	20	12	23	6	6	5	50%	52%		
13	The Lighthouse	Tom	12	22	20	36	36	30	14	16	17				26%	17%		
14	Atlantics	Birnbaum	13		6	37	34	35	12	13	12	12			36%	27%		
15	The Last Black Man in San Francisco	Charlie	14	24		40	33	24	23	15	13	9			26%	23%		
16	Shoplifters	Birnbaum	15	7	13	6	7	4	2	4	6				81%	84%		
17	A Ghost Story	Charlie	16			14	18	36	13	6	9				58%	63%		
18	Zodiac	Levine	17	20		11	16	32	7	11	16				51%	52%		
19	The Thing	Birnbaum	18			34	29	37	18	14	26				26%	25%		
20	Borat Subsequent Moviefilm	JL	19			35	11	12	29		29	7	3	4	49%	60%		
21	Mean Streets	Charlie	20				37	43		10	31		10		24%	8%		
22	I'm Thinking of Ending Things	Tom	21	9	14	18	9	8	30		14		19		49%	58%		
23	First Reformed	Charlie	22			7	39	27	39	16		22			35%	33%		
24	Coherence	Levine	23			9	22		23	15		10			46%	49%		
25	Whiplash	Levine	24	4	8	16	4	5	6	3	15				79%	85%		

How to find *new* movies which *all* of us will enjoy?

DESIRED PRODUCT

MANY MOVIE RECOMMENDATION SYSTEMS EXIST ALREADY.

WE WANT A RECOMMENDATION SYSTEM WHICH:

1. IS GROUP-FOCUSED: MAXIMIZES THE PROBABILITY THAT *ALL* GROUP MEMBERS WILL ENJOY A SELECTION, AND
2. SUGGESTS NEW MOVIES: CONSTANTLY INCORPORATES NEW REVIEWS, RATHER THAN RELYING ON A STATIC DATABASE OF OLD REVIEWS

INPUT: A RANKED LIST OF MOVIES FOR 2+ GROUP MEMBERS

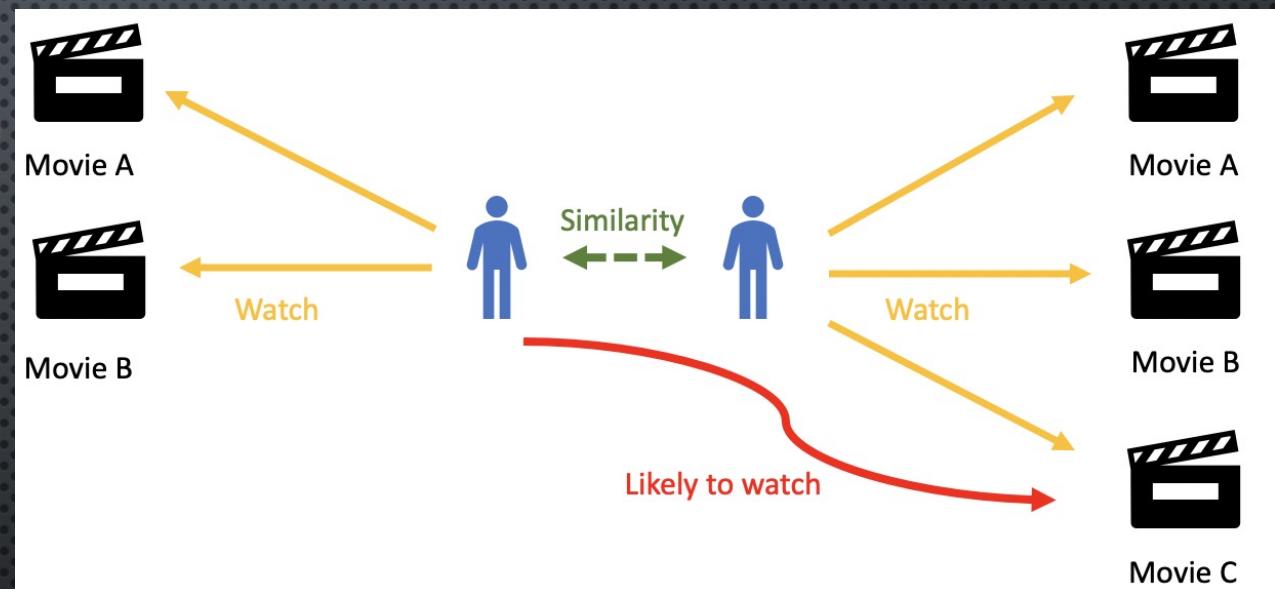
OUTPUT: A SET OF ~5 MOVIE SUGGESTIONS

THE DATA

-  (SOCIAL FILM DISCOVERY SITE)
 - 3M+ USERS (2X SINCE Q1 2020), 600M+ MOVIES WATCHED
 - EMPHASIS ON NEW MOVIES: TOP 2020 MOVIES HAVE 50K+ REVIEWS EACH
 - REVIEW DATA AVAILABLE VIA API
-  (OPEN SOURCE MOVIE REVIEW DATABASE)
 - 27M REVIEWS OF 58K MOVIES BY 280K USERS
 - EMPHASIS ON OLDER MOVIES: 1947-2018
 - 1.1M USER-SUBMITTED TAGS (USEFUL FOR CONTENT-BASED FILTERING)
- BOTH SERVICES USE THE SAME RATING SYSTEM (0-5 STARS)

APPROACH: COLLABORATIVE FILTERING

1. FOR EACH GROUP MEMBER, IDENTIFY USERS IN DATASET WITH SIMILAR TASTE
 - ('SIMILAR TASTE' = HAVING WATCHED SEVERAL OF THE SAME MOVIES, AND RATED THEM SIMILARLY) → NEAREST NEIGHBOR ALGORITHM
2. HAVING FOUND SIMILAR USERS, IDENTIFY NEW MOVIES THAT ALL (OR MANY) RATED HIGHLY



https://miro.medium.com/max/1400/1*2_eaF6211RoGZBvZAm_sEQ.png

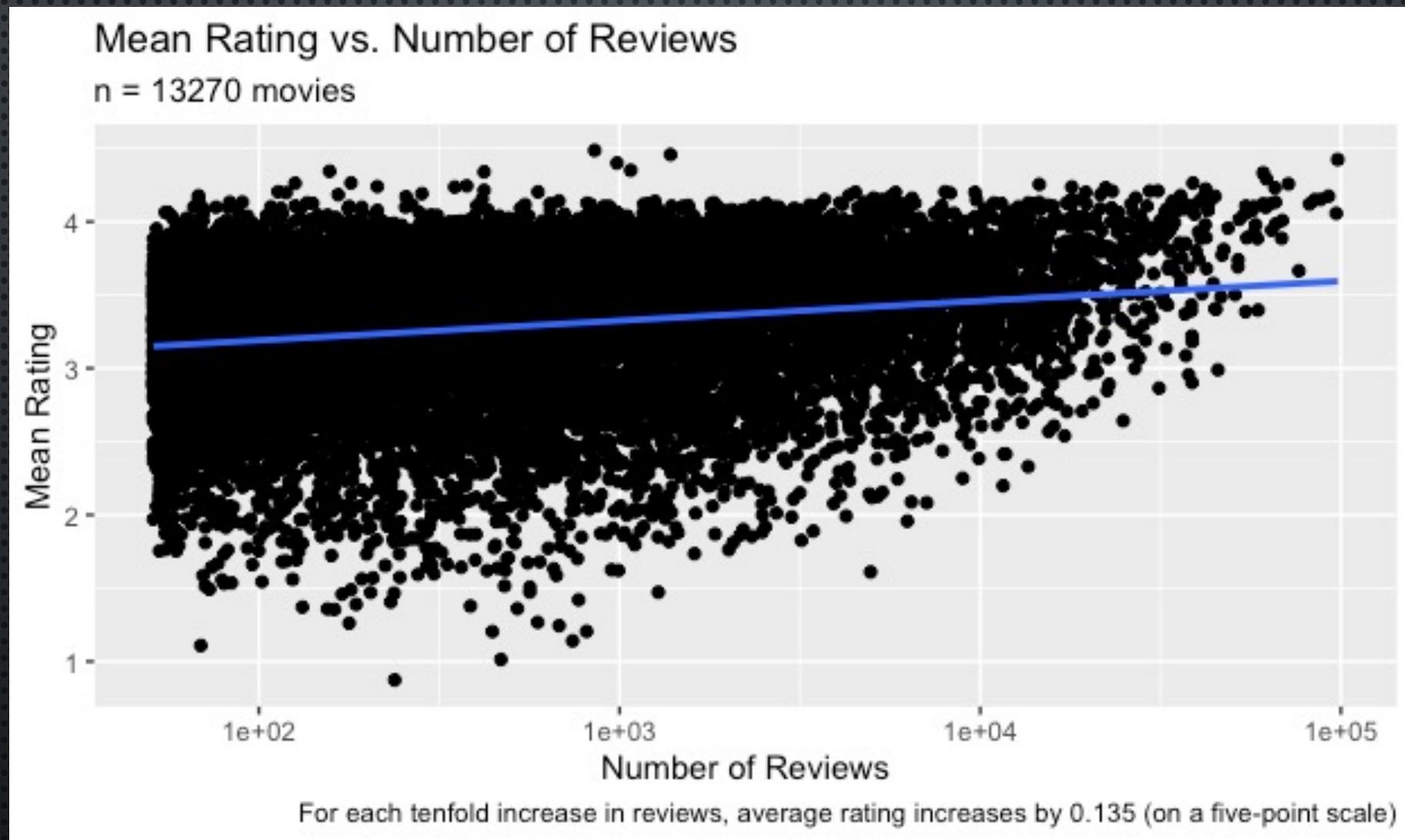
SCALABILITY WILL BE A CHALLENGE → MATRIX FACTORIZATION!

MODEL EVALUATION

CROSS-VALIDATION, FORMING ARTIFICIAL "GROUPS" FROM RANDOMLY SELECTED SETS OF USERS

- RATHER THAN MSE, SUCCESS IS DEFINED BY % OF RECOMMENDATIONS ENJOYED BY ALL USERS
- E.G. "IN SIMULATION STUDIES, 94% OF RECOMMENDATIONS WERE RATED BY ALL GROUP MEMBERS AS 4 STARS OR ABOVE."

EDA: MOVIELENS DATA



EDA: MOVIELENS DATA

