MovieTweetings: a Movie Rating Dataset Collected From Twitter

@sidooms

Simon Dooms

Research ❤ datasets

- Recsys research needs datasets
- To evaluate, experiment and demonstrate

- I need datasets

Available for download:

- MovieLens 100K
- MovieLens 1M
- MovieLens 10M

Braveheart (1995)
Director: ***Mel Gibson***
Cast: **Patrick McGoohan, Angus Macfadyen, Sophie Marceau, Mel Gibson**
Genre: **History, Biography, Drama, Action, War**
Runtime: 2 hrs 57 mins

Forrest Gump (1994)
Director: ***Robert Zemeckis***
Cast: **Robin Wright, Sally Field, Gary Sinise, Tom Hanks**
Genre: **Romance, Comedy, Drama**
Runtime: 2 hrs 22 mins

Liar Liar (1997)
Director: ***Tom Shadyac***
Cast: **Cary Elwes, Jim Carrey, Justin Cooper, Maura Tierney**
Genre: **Fantasy, Family, Comedy**
Runtime: 1 hr 26 mins

Research ❤ datasets

- Recsys research needs datasets
- To evaluate, experiment and demonstrate

- I needed datasets

Available for download:

▪ MovieLens 100K	~ most recent movie: 1998
▪ MovieLens 1M	~ most recent movie: 2000
▪ MovieLens 10M	~ most recent movie: 2008

I need up-to-date movie ratings

Finding data

- Data is all around us

IMDb | Find Movies, TV shows, Celebrities and more... | All | Simon Dooms | Help
Movies | TV | News | Showtimes | Community | IMDbPro | Apps | Your Watchlist

Death Proof (2007)

Top 5000

16 | 113 min - Action | Thriller - 7 June 2007 (Netherlands)

7,1

Your rating: ★★★★★★★★★★ /10

Ratings: **7,1**/10 from 147.254 users

Reviews: 545 user | 193 critic

Two separate sets of voluptuous women are stalked at different times by a scarred stuntman who uses his "death proof" cars to execute his murderous plans.

Director: Quentin Tarantino

Writer: Quentin Tarantino

Stars: Kurt Russell, Zoe Bell, Rosario Dawson | See full cast and crew

✓ Watchlist | Share...

3 nominations. See more awards »

Photos

Quick Links

Full Cast and Crew | Plot Summary
Trivia | Parents Guide
Quotes | User Reviews
Awards | Release Dates
Message Board | Company Credits

Explore More

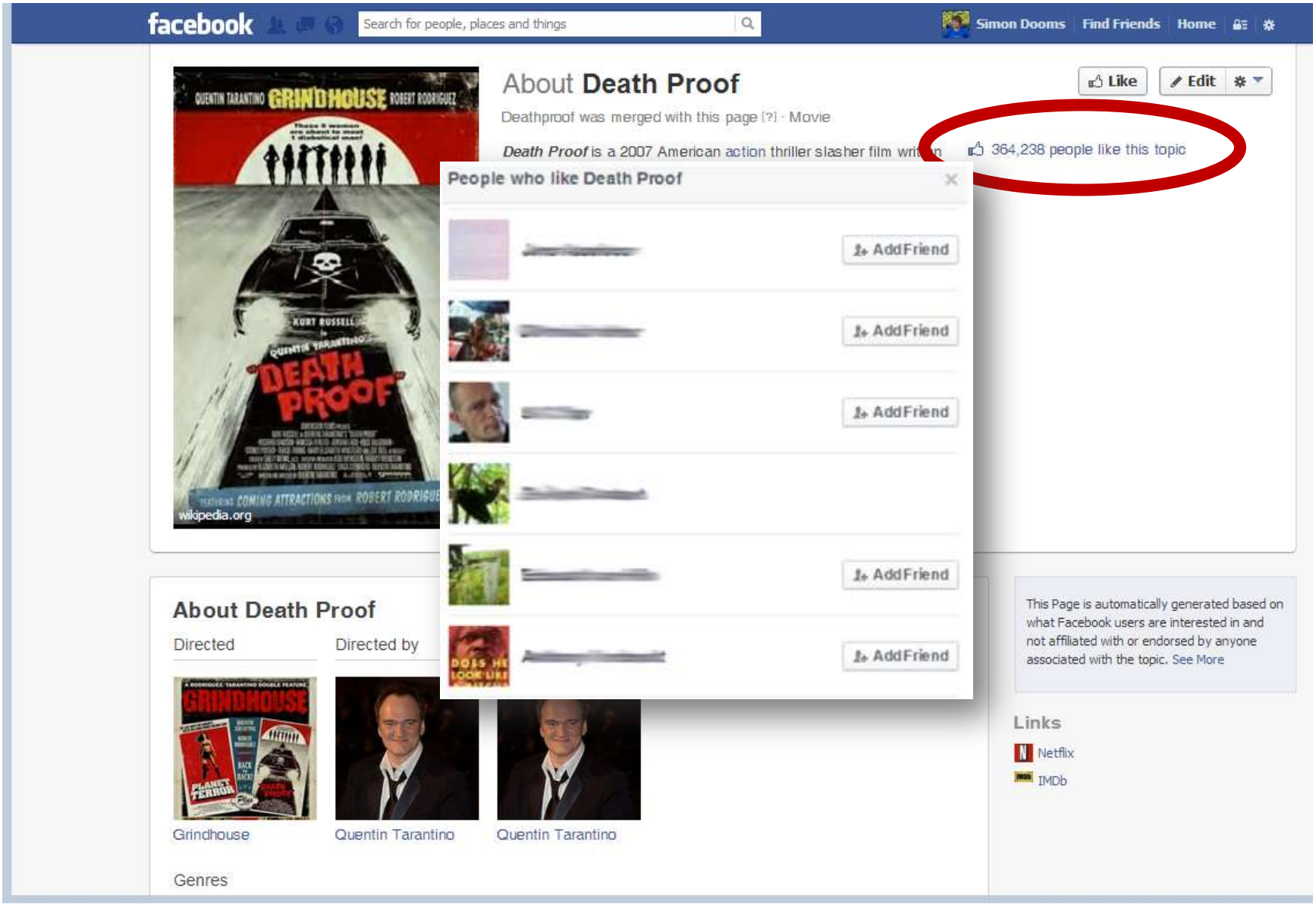
Like 7,118 people like this. Be the first of your friends.

Related News

Another Clip, TV Spot & Poster For 'Machete Kills'
27 September 2013 5:21 PM, +01:00 | The Hollywood News

The Avengers: Age Of Ultron Comic Con Teaser (Finally) Leaked
27 September 2013 7:56 AM, +01:00 | Obsessed with Film

It's official! Kurt Russell to join 'Fast and Furious 7' cast
26 September 2013 6:34 AM, +01:00 | RealBollywood.com


facebook
Search for people, places and things
Simon Dooms
Find Friends
Home
About Death Proof
Like
Edit
Deathproof was merged with this page [?] · Movie
Death Proof is a 2007 American action thriller slasher film
364,238 people like this topic
People who like Death Proof
Add Friend
About Death Proof
Directed
Directed by
Grindhouse
Quentin Tarantino
Quentin Tarantino
Genres
This Page is automatically generated based on what Facebook users are interested in and not affiliated with or endorsed by anyone associated with the topic. See More
Links
Netflix
IMDb

Home
Connect
Discover
Me
death proof
@vrunt
death proof was a good movie and a great plot point
Expand
Oldtimer @auldtimer 18h
All those clever "but why not Chauhan? Why not Parrikar?" arguments seem to have died a quiet death. Proof that Con bribed oped writers
Expand
Sean Raitt @seanraitt1 28 Sep
Right enough of this Tweeting back to my Next Movie #DeathProof with StuntMan Mike ... He's BadAss pic.twitter.com/9dglivatNR
View photo
PrinceCharles Cinema @ThePCCLondon 28 Sep
TONIGHT our Grindhouse Mini-Thon finally arrives. This 4-headed beast features Planet Terror & Death Proof,... fb.me/2I8uHQ65z
Expand
michael @michaelestelle
Death Proof is one of my all time favorite movies.
Expand

Finding data

- Data is all around us
  BUT extremely unstructured

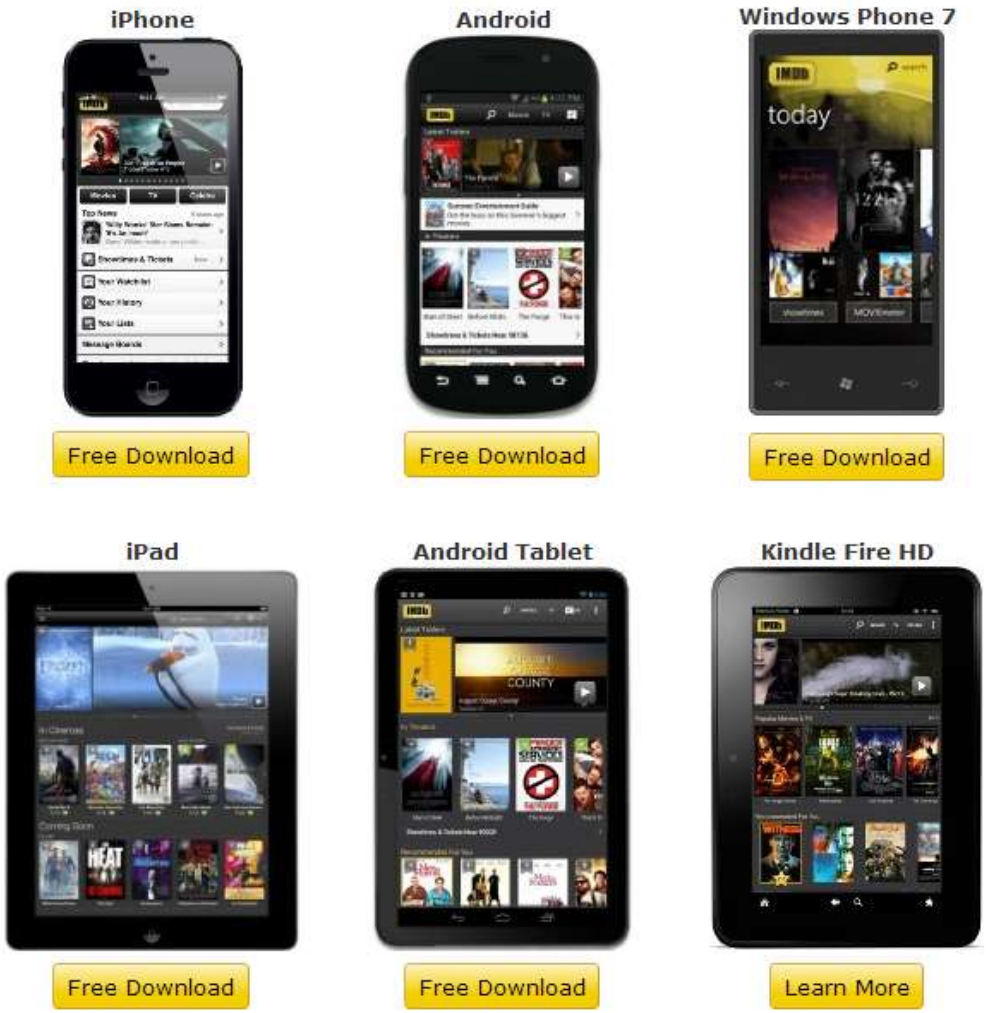
- What we want:

1::122::5::838985046
1::185::5::838983525
1::231::5::838983392
1::292::5::838983421
1::316::5::838983392

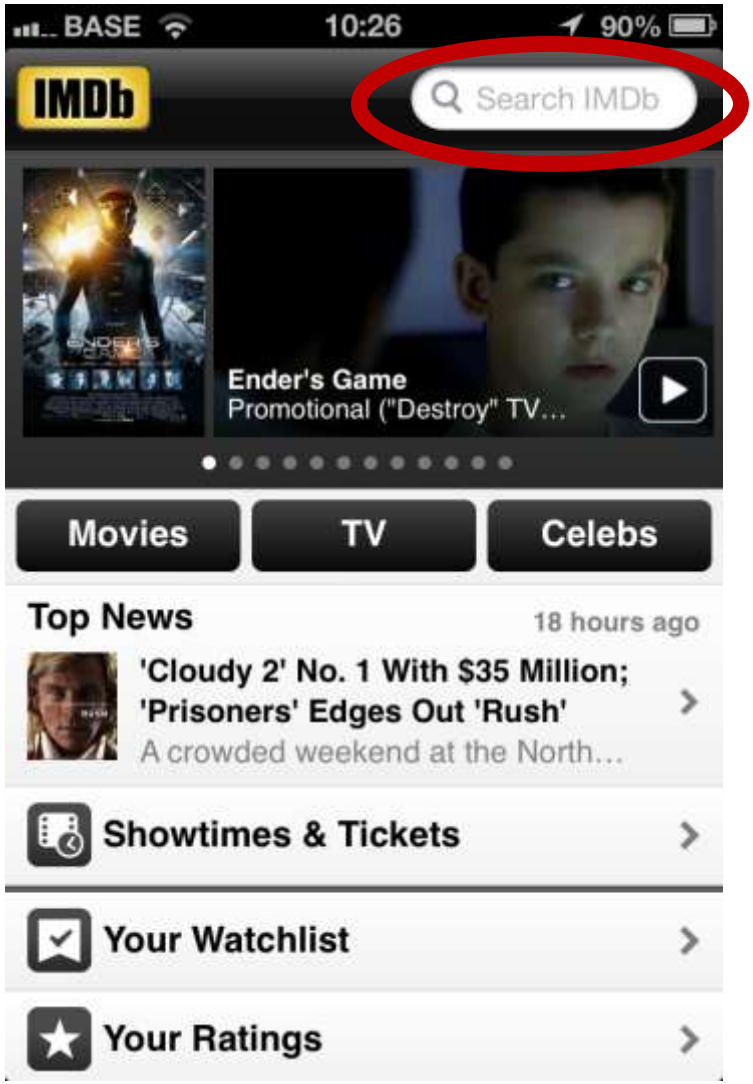
(user, item, rating, time)

Structured data

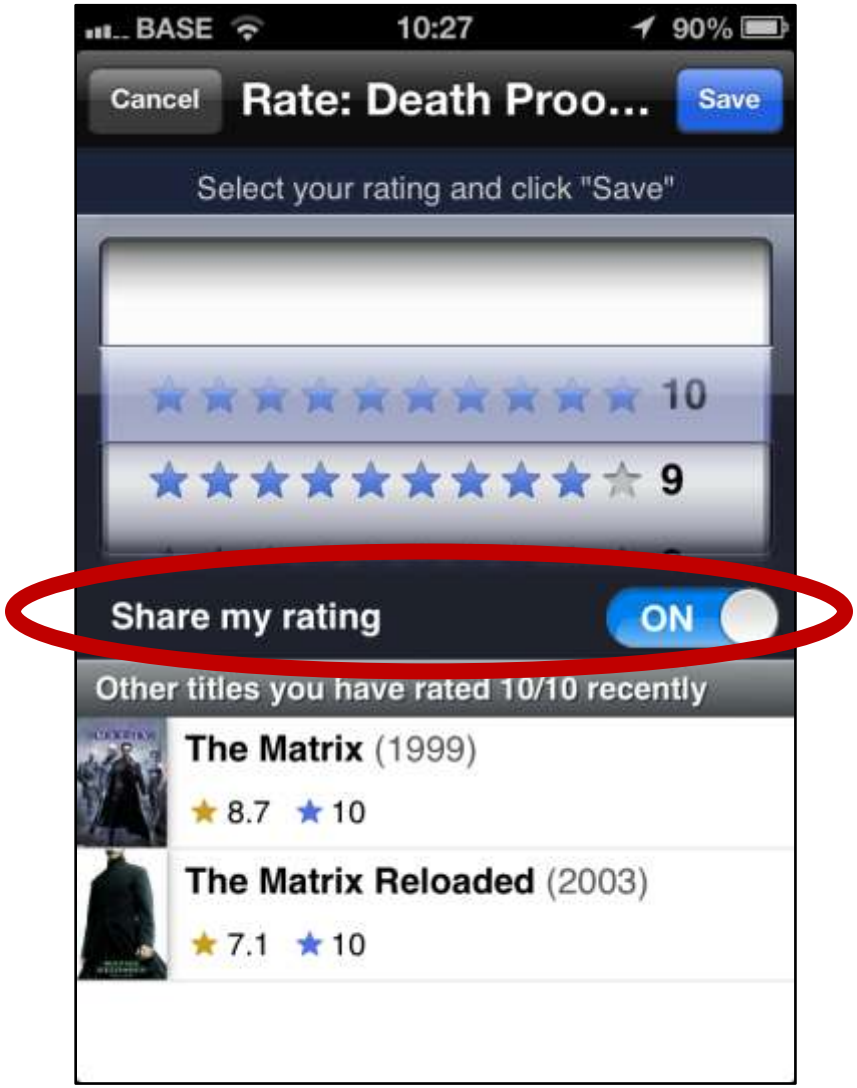
Structured data



Structured data



Structured data

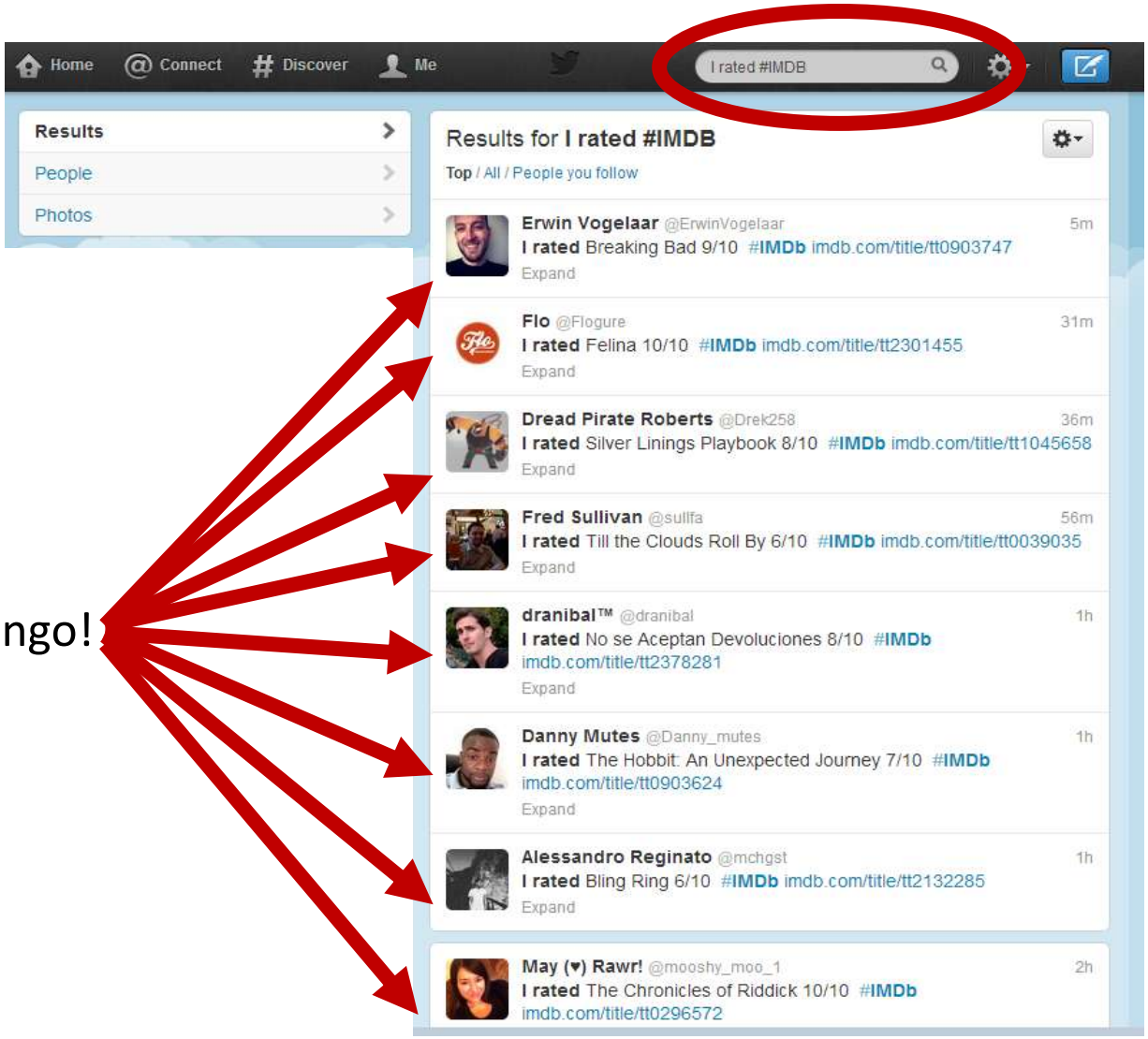


Structured data

"I rated Death Proof 10/10 #IMDb"

- User
- Item (movie)
- Rating
- Hashtag

Structured data



Bingo!

Search Twitter for "I rated #IMDb"

Collecting data

- We query the Twitter API for “I rated #IMDb”
- Extract relevant information
- Cross-reference with IMDb for extra genre data

The data

Ratings.dat

Rating scale from 1 to 10

```
1::1074638::7::1365029107
1::1853728::8::1366576639
2::0113277::10::1379466669
```

IMDb ID - http://www.imdb.com/title/tt0113277

Movies.dat

```
1028528::Death Proof (2007)::Action|Thriller
0133093::The Matrix (1999)::Action|Adventure|Sci-Fi
1670345::Now You See Me (2013)::Thriller
```

Users.dat

```
1::18405182
2::995885060
3::31260677
```

Twitter ID (NOT @handle)

Your data

- MovieTweetings dataset available on GitHub
  (https://github.com/sidooms/MovieTweetings)
- Find it on the **RecSys Wiki** (category datasets)

Latest

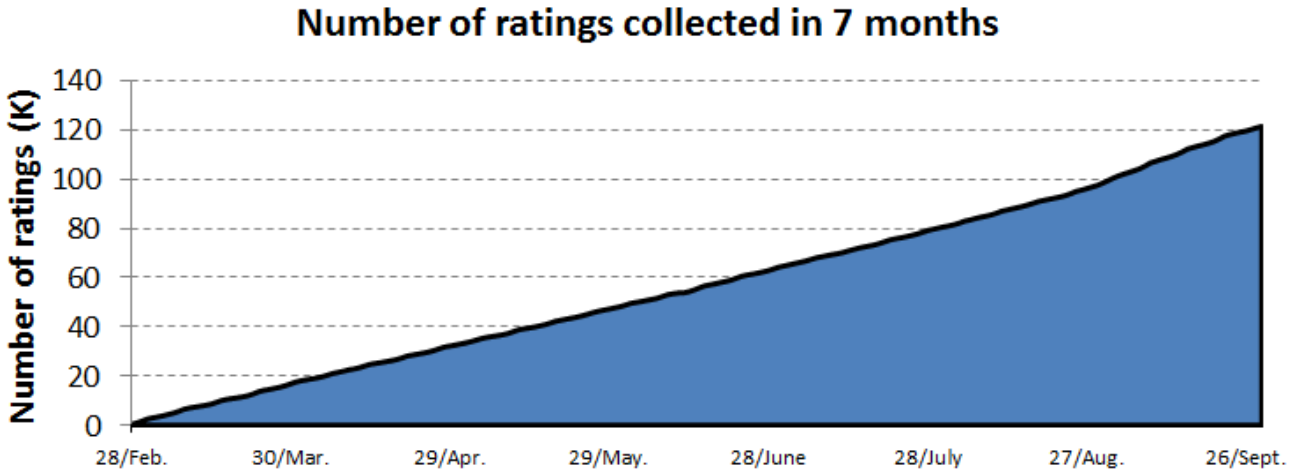
- All ratings
- Automagically updated daily

Snapshots

- Fixed portion of dataset
- Added manually when appropriate
- 10K, 20K, 30K, 40K, 50K, 100K

DISCLAIMER: Depending on Twitter API, IMDb apps and me!

Some numbers



$567_{/day}$

	MovieTweetings	MovieLens 100K	MovieLens 1M	MovieLens 10M
Ratings	121,404	100,000	1,000,209	**10,000,054**
Users	19,464	943	6,040	**71,567**
Items	**11,655**	1682	3,900	10,681

(Results on September 30, 2013)

Some fun

Top 3 most rated movies

1. Iron Man 3 (2013)
2. Man of Steel (2013)
3. World War Z (2013)

Top 3 AVG rated movies (min 20 ratings)

1. The Shawshank Redemption (1994)
2. LOTR: The Return of the King (2003)
3. The Dark Knight (2008)

Bottom 3 worst AVG rated movies (min 20 ratings)

3. Scary MoVie (2013)
2. Piranha 3DD (2012)
1. Cosmopolis (2012)

Some conclusions

- Outdated public datasets
- Social media = Unstructured data available
- Structured rating data through Twitter – IMDb
- MovieTweetings: our Movie Rating Dataset
  - Always up-to-date
  - Includes most recent and most relevant movies
  - Unfiltered rating data
  - Publicly available
- Death Proof (2007) really is an awesome movie

MovieTweetings: a Movie Rating Dataset Collected From Twitter

@sidooms

Simon Dooms