# For the Love of Board Games

Board Game Recommendation Engine

For BoardGameGeek

By Jen Wadkins

# What's the Data?

Vital Stats

- 182k users
- 22.5k most popular board games
- From the BoardGameGeek API
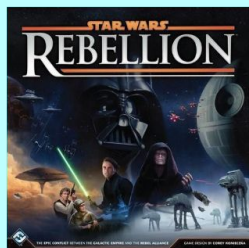- 117k users over 5 native ratings

# Content Recommender - Business Problems

- Cost of acquiring and maintaining data
- Overcome common problems with recommender systems
- Overcome BGG specific problem
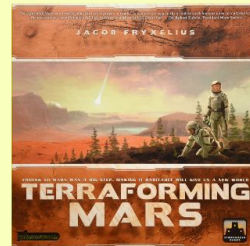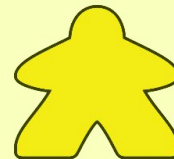- Computational cost of recommendations

Collaborative Filtering

# Collaborative Filter Challenges

## Data Acquisition /Maintenance

- Data can be refreshed/cleaned in about 30 minutes a day of computing time
- Automated retrieval and cleaning

## Computational Cost Of Recommendation

- End model is purely mathematical, and requires no ongoing model maintenance and deployment.
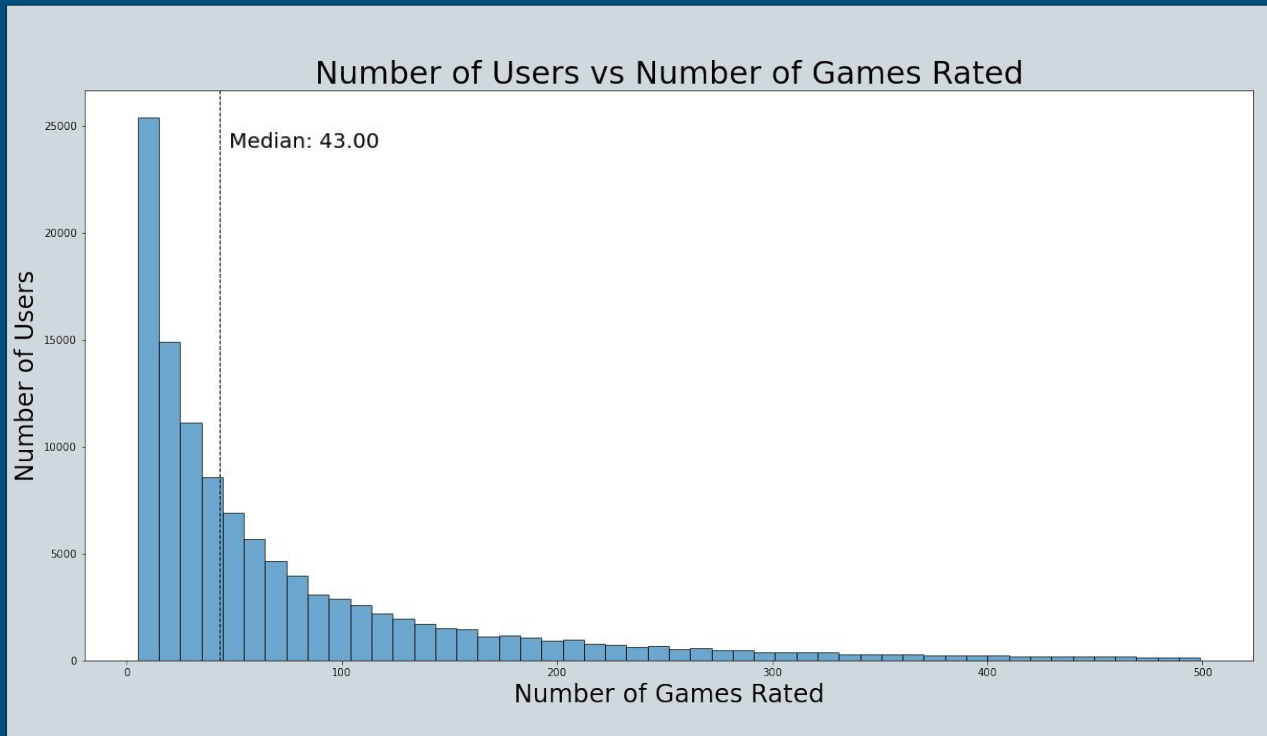
# Collaborative Filter Challenges

## Cold Start

- Address in common fashion with user questionnaire.
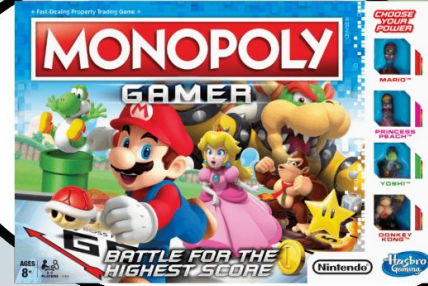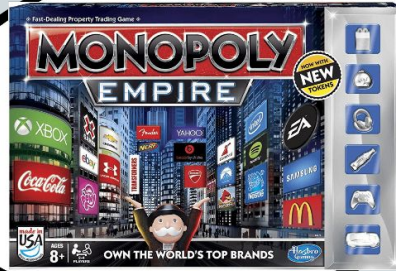- Insufficient time - allocated to Future Work
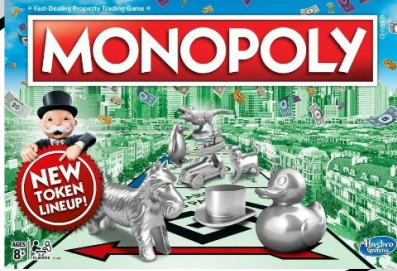
## Catalog Coverage

- Our solution results in high catalog coverage

# Collaborative Filter Challenges

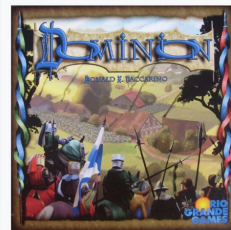Low number of user ratings - very sparse matrix

BGG Specific
Problem:
Missing
Relationships

# Leveraging Content-Based Filtering with Synthetic Ratings

- Increase total user ratings
  - Create ratings for very similar items
- Increase catalog coverage

Content Based Filtering

100 Synthetic Ratings for a 5-Rating User

# What are "RECOMMENDED" and "RELEVANT" Items?

A RECOMMENDED item is a game that the user will like more than their average.

A RELEVANT item is a game that the user actually rated more than their average.

Successful Recommender:
- Correctly identifying the user's relevant items as recommended
- Getting reasonably close to the user's actual ratings
- Improves catalog coverage by recommending less popular items

# How does synthetic data change recommender quality?

| | Average Rating Error | Recall | Coverage |
|---|---|---|---|
| ALL Users | ▼ 7.03% | ▼ 6.22% | ▲ 11.16% |
| Users UNDER median | ▼ 18.07% | ▼ 5.85% | ▲ 10.9% |
| Users OVER median | ▲ 5.66% | ▼ 6.56% | ▲ 69.5% |

# Future Work

- Retune content recommender for improved metrics
- Find the sweet spot of synthetic ratings for recommendation time - try 50, 75, etc
- Implement the cold start questionnaire
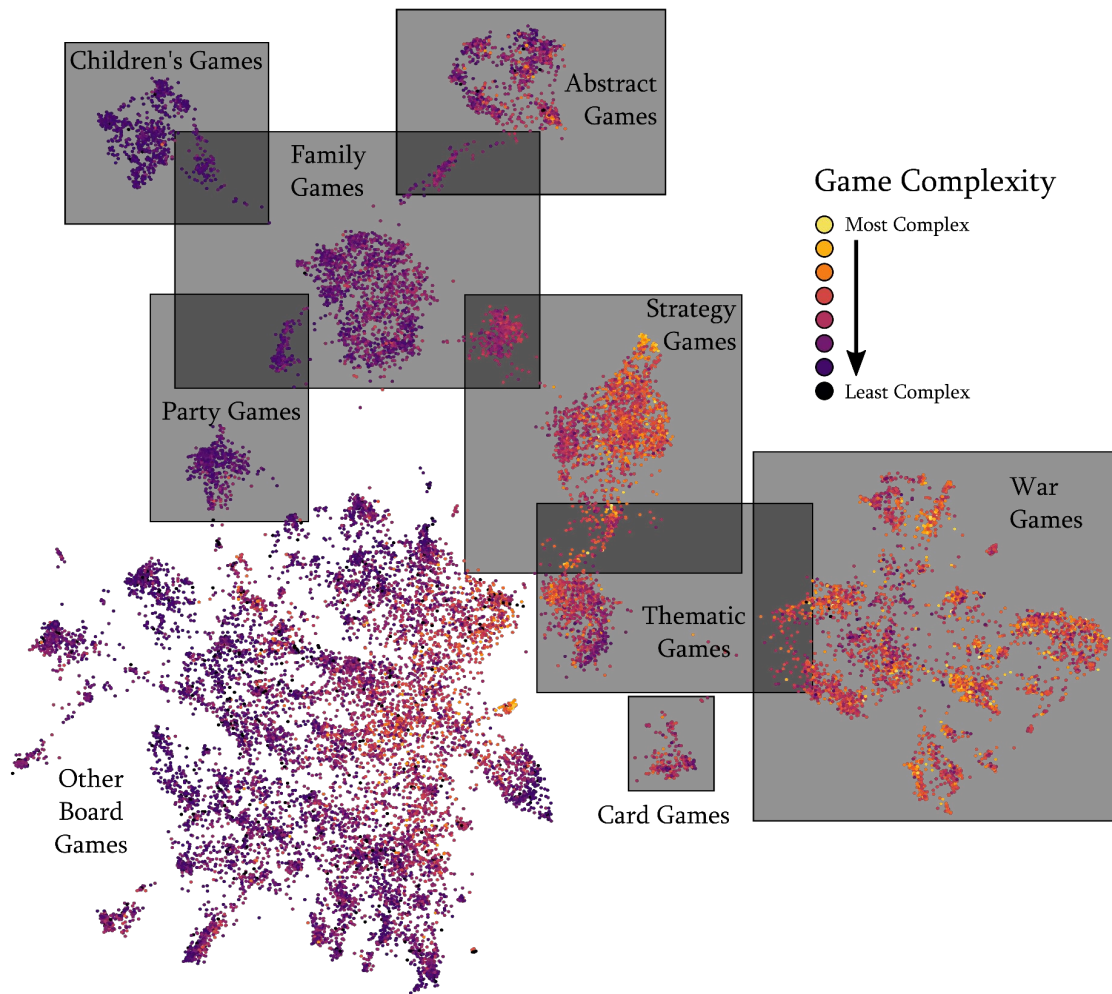- Deploy GUI

# Thank you!

# Contact Me

Jen Wadkins

Email:           threnjen@gmail.com

GitHub:          GitHub Repo

Publications:    Jen Wadkins on Medium

Kaggle:          Kaggle Profile

Game Complexity
Most Complex
Least Complex

Children's Games

Family Games

Abstract Games

Party Games

Strategy Games

War Games

Thematic Games

Card Games

Other Board Games

APPENDIX:

Board Game Similarity Clusters

Identifying important features for

Content-Based Filtering