



# For the Love of Board Games



Board Game Recommendation Engine

For BoardGameGeek

By Jen Wadkins



# What's the Data?

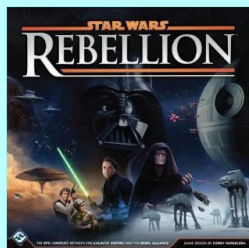
## Vital Stats

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

# Collaborative Filtering



10



9



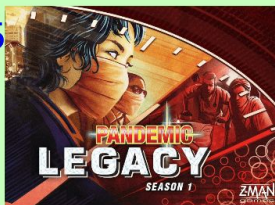
3.5

4



10

9.5



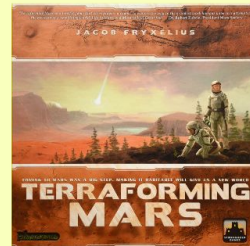
8.5

9

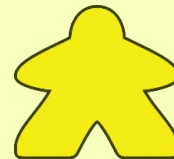


7

7



10



9

# 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
- Requires no ongoing model maintenance and deployment.

# Collaborative Filter Challenges

---

## Cold Start

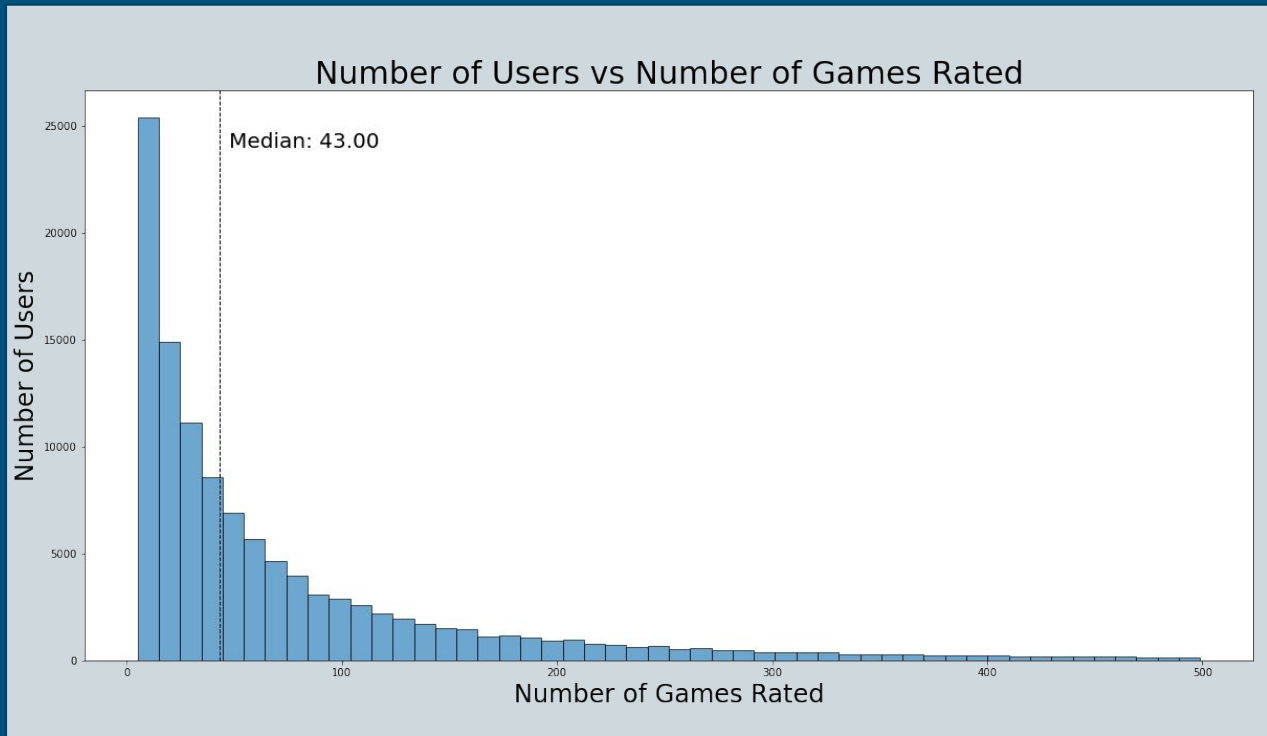
- Address in common fashion with user questionnaire.
- Insufficient time - allocated to Future Work

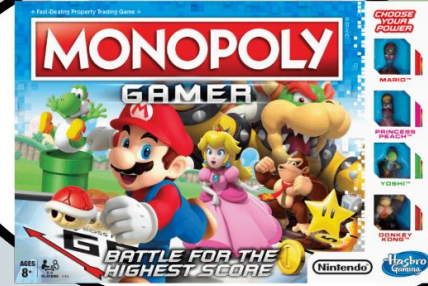
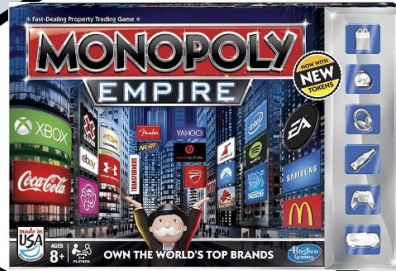
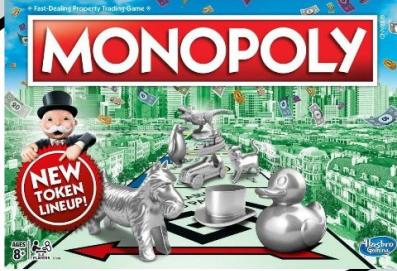
## Catalog Coverage

- Our solution results in high catalog coverage

# Collaborative Filter Challenges

Low number of  
user ratings -  
very sparse  
matrix

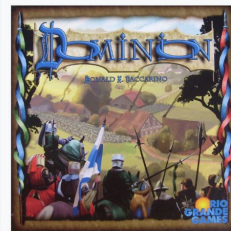




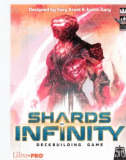
BGG Specific  
Problem:  
Missing  
Relationships

# Leveraging Content-Based Filtering with Synthetic Ratings

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



Content  
Based  
Filtering







# 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	▼ 7.66% .571 to .527	▼ 5.82% 61.75% to 58.15%	▲ 9.69% 83.93% to 92.06%
Users UNDER median	▼ 18.39% .608 to .497	▼ 5.6% 59.3% to 56%	▲ 9.59% 83.64% to 91.66%
Users OVER median	▲ 4.86% .533 to .559	▼ 5.96% 64.3% to 60.4%	▲ 49.17% 54.89% to 81.88%

# 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

---

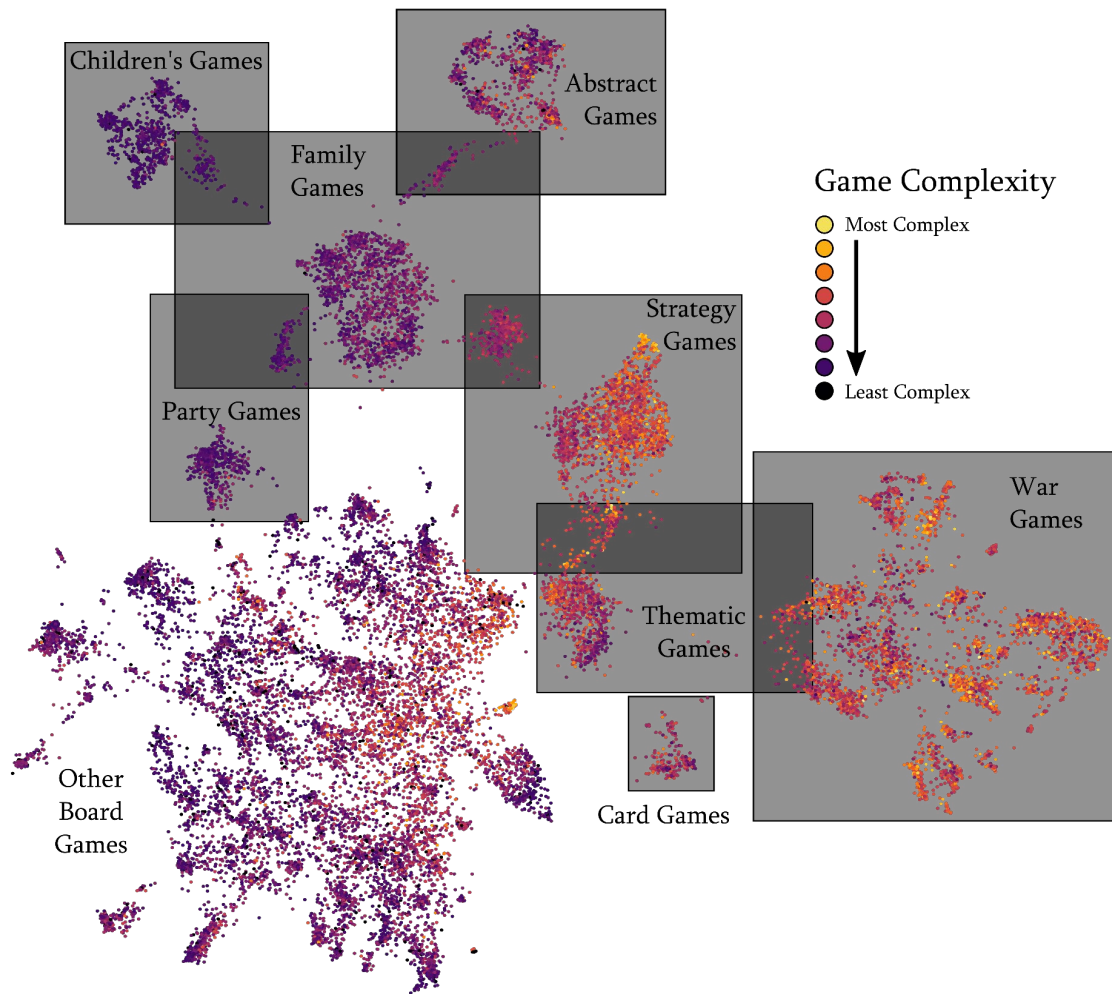
Jen Wadkins

Email: [threnjen@gmail.com](mailto:threnjen@gmail.com)

GitHub: [GitHub Repo](#)

Publications: [Jen Wadkins on Medium](#)

Kaggle: [Kaggle Profile](#)



## APPENDIX:

### Board Game Similarity Clusters

Identifying important features for  
Content-Based Filtering