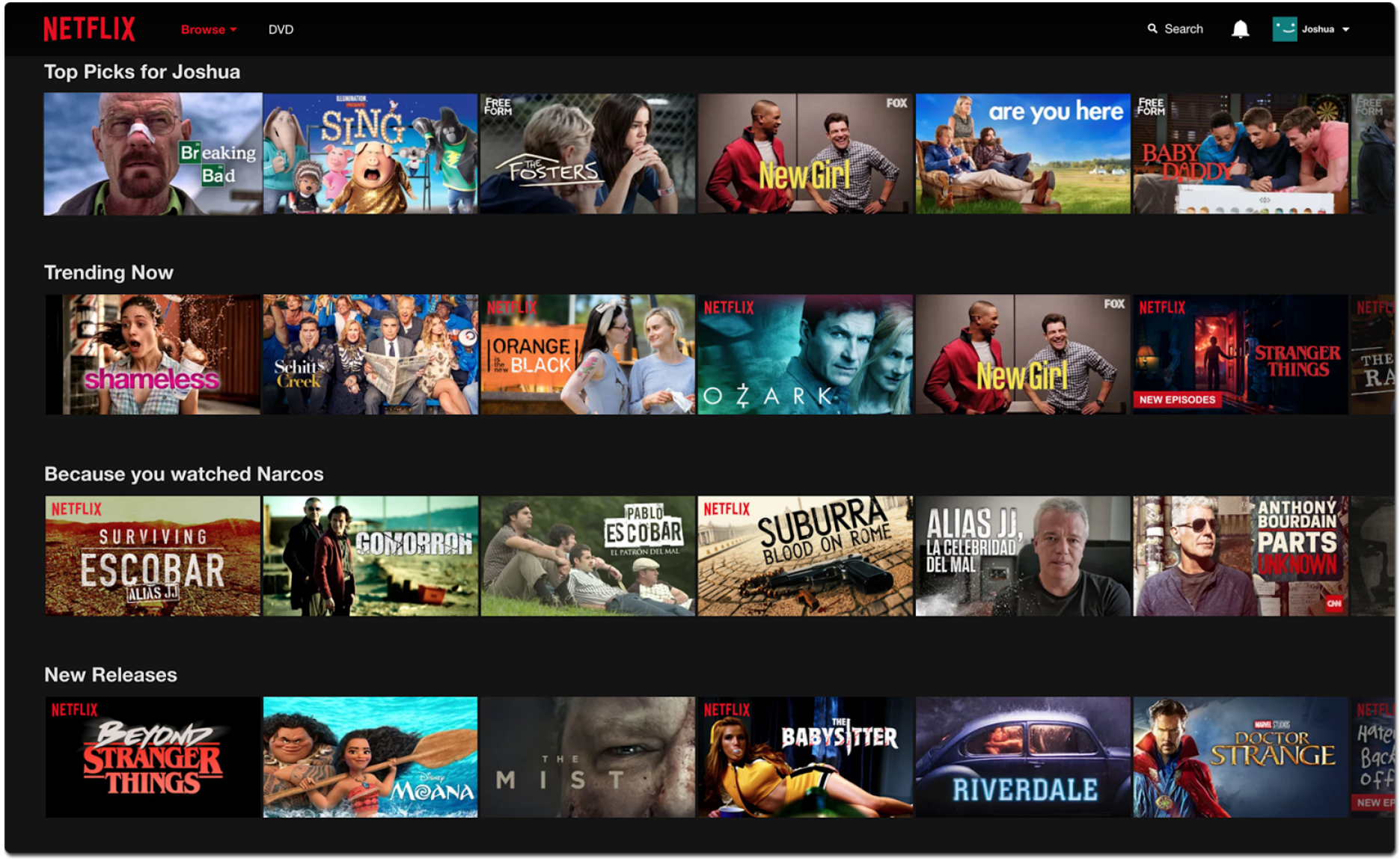
Foursquare Recommendations from Random Network Walks

Nick Dulchin

Recommendation Systems



Dataset

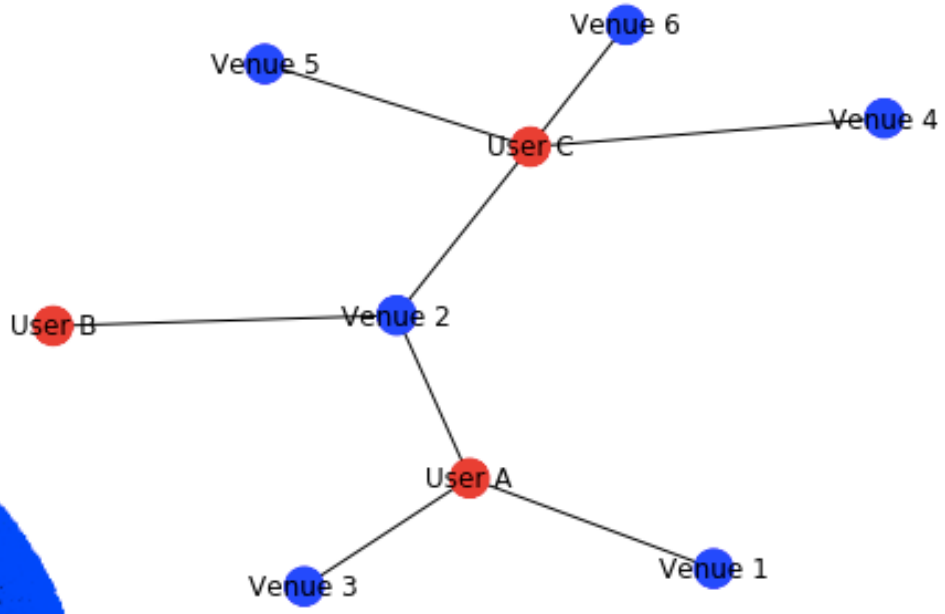
- Foursquare check-ins in New York City
  - April 3rd, 2012 to February 16th, 2013
- Unique users: 1083
- Unique venues: 38333

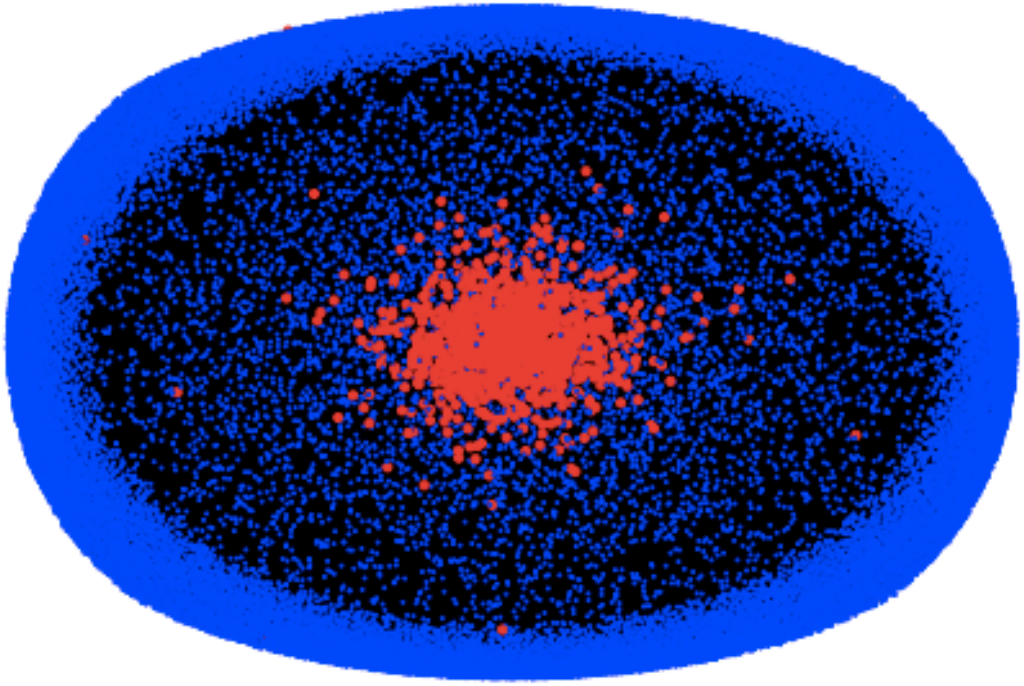
Check-ins per user-venue combination

	Quantity
count	91024.000000
mean	2.498550
std	6.577072
min	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
max	257.000000

Heterogeneous Network

- Red nodes represent users
- Blue nodes represent venues
- Edges only exist between a user and venue
- Edge weight is based on how many times a user checked in at a venue

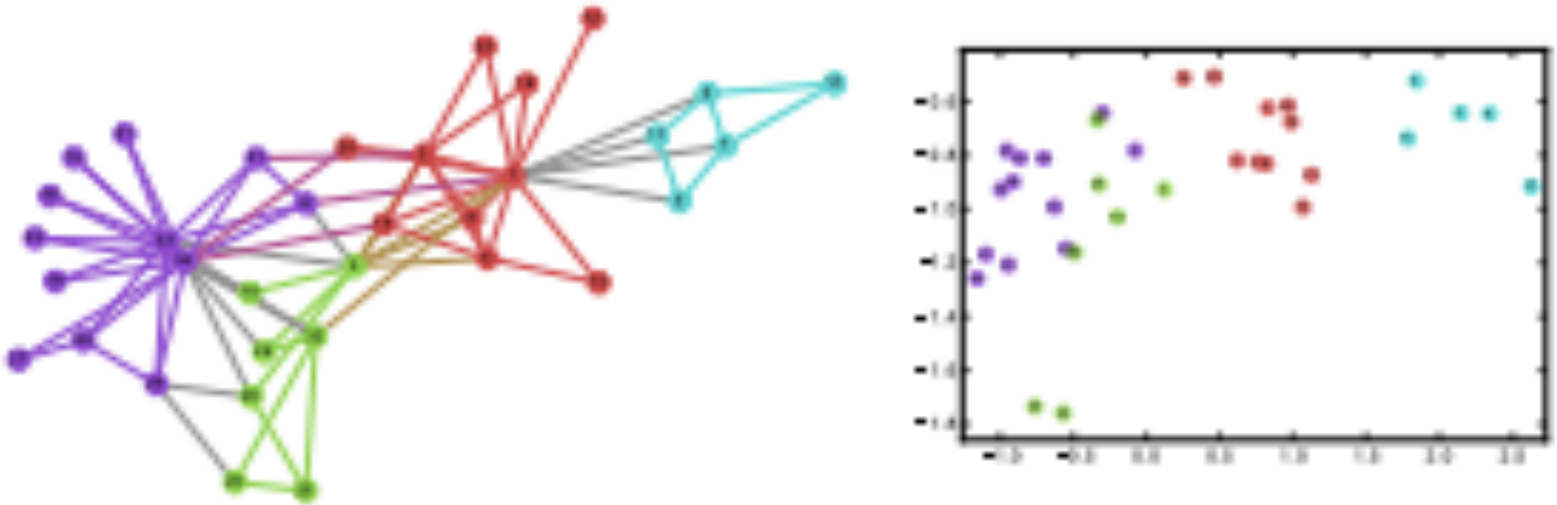




```
Type: Graph
Number of nodes: 35100
Number of edges: 75014
Average degree:   4.2743
```

DeepWalk

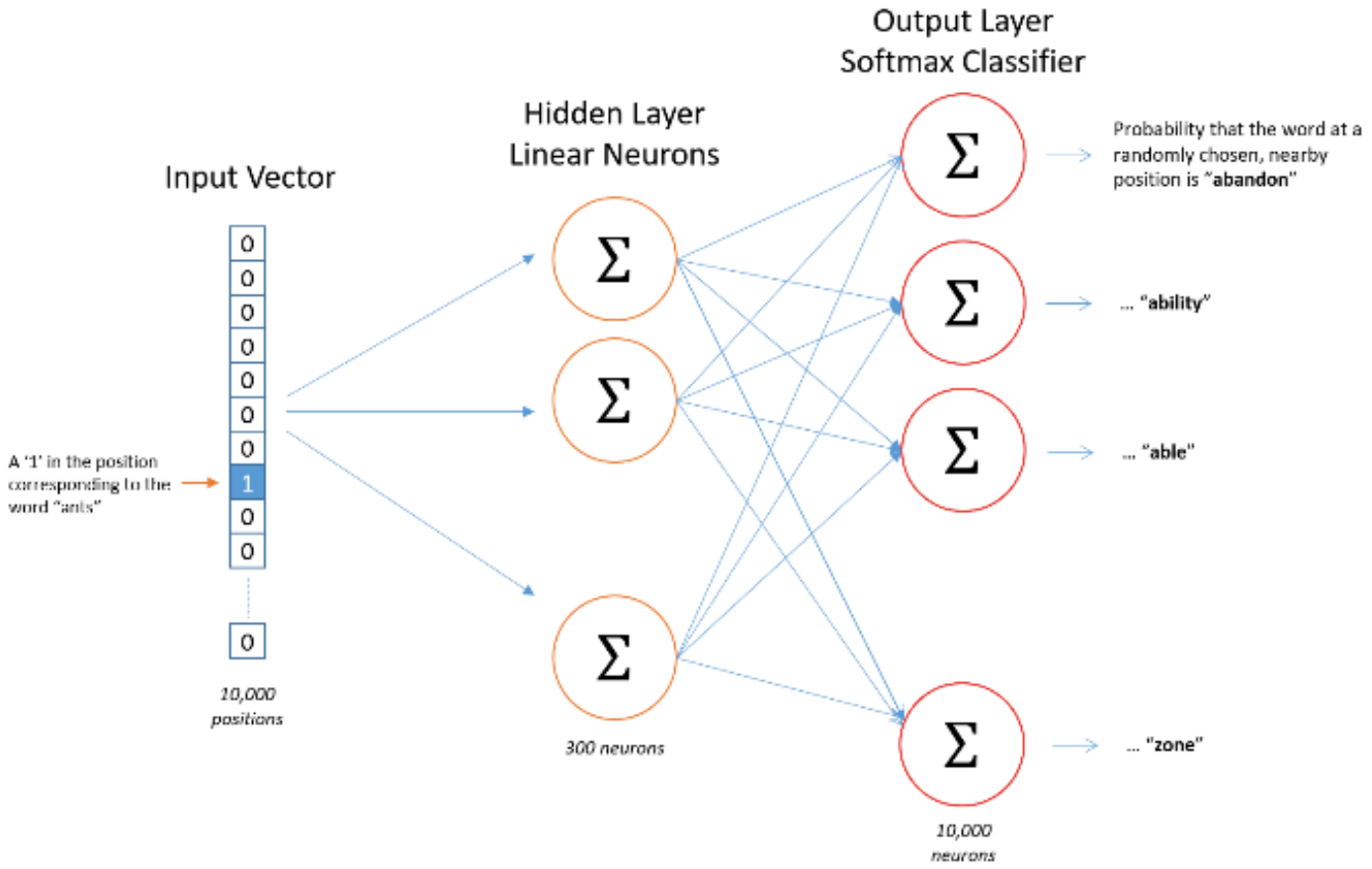
- Algorithm published in 2014
- Learns a latent representation of an adjacency matrix using random walks
- Uses neural network techniques developed for language modeling



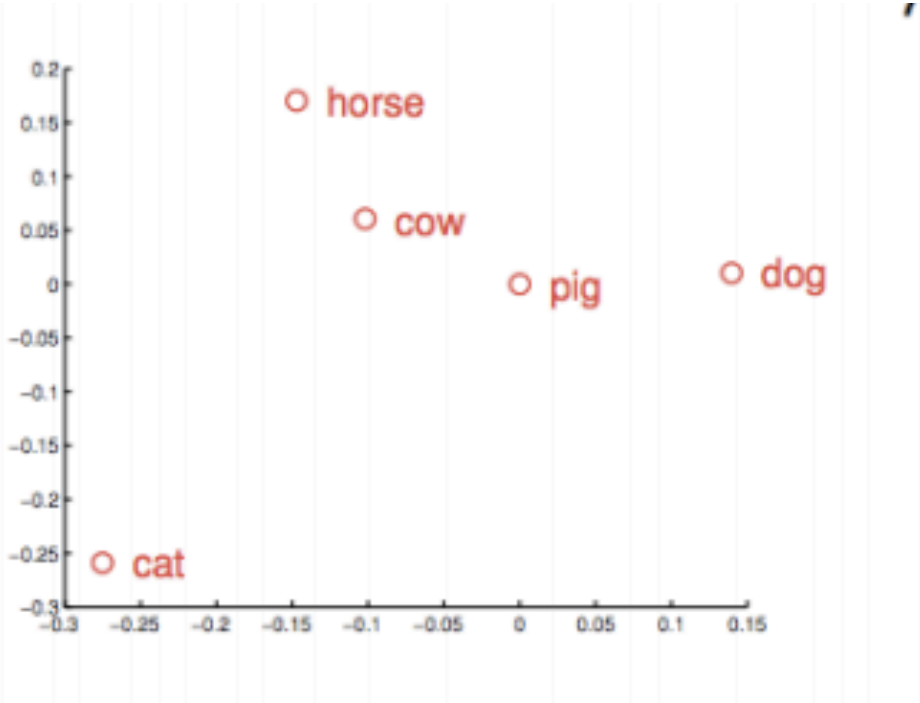
(a) Input: Karate Graph (b) Output: Representation

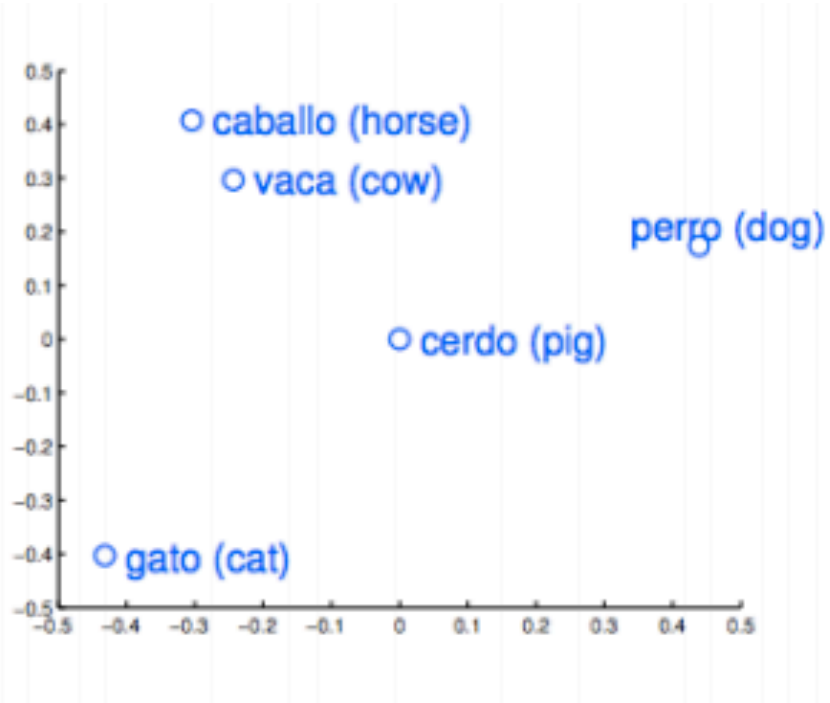
Word2vec

- First introduced by Google researchers in 2013
- Used for Natural Language Processing
- Trains a neural network with one hidden layer



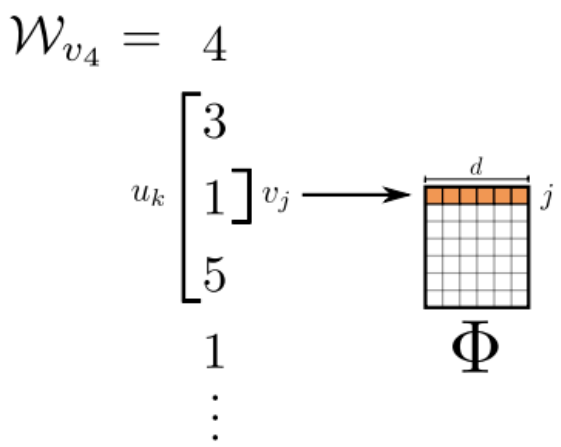
Word2vec

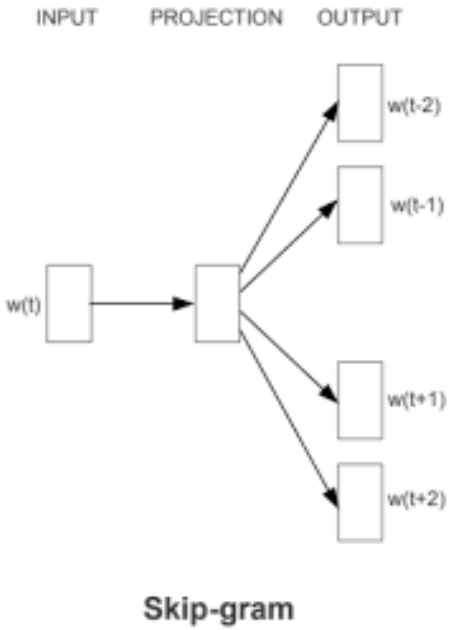

horse
cow
pig
dog
cat


caballo (horse)
vaca (cow)
perro (dog)
cerdo (pig)
gato (cat)

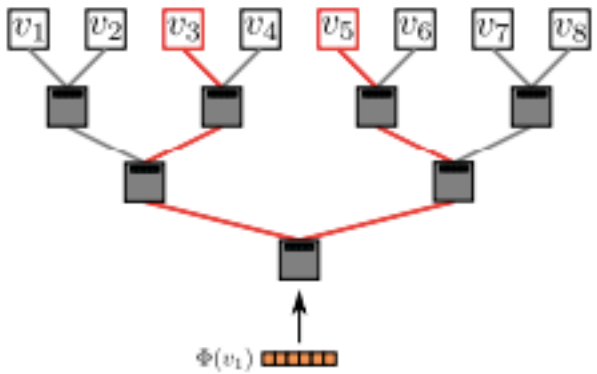
Skipgram/Hierarchical Softmax

- Skipgram takes “words” and adjusts weights to predict “words” that will be around it more accurately
- Hierarchical softmax reduces computational complexity of the output layer function from: $O(V) \rightarrow O(\log_2 V)$



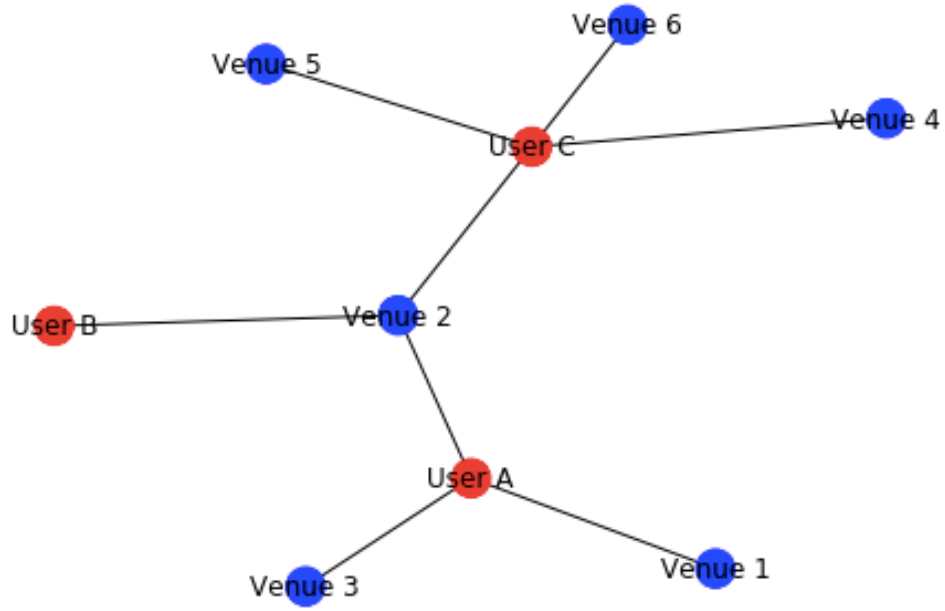


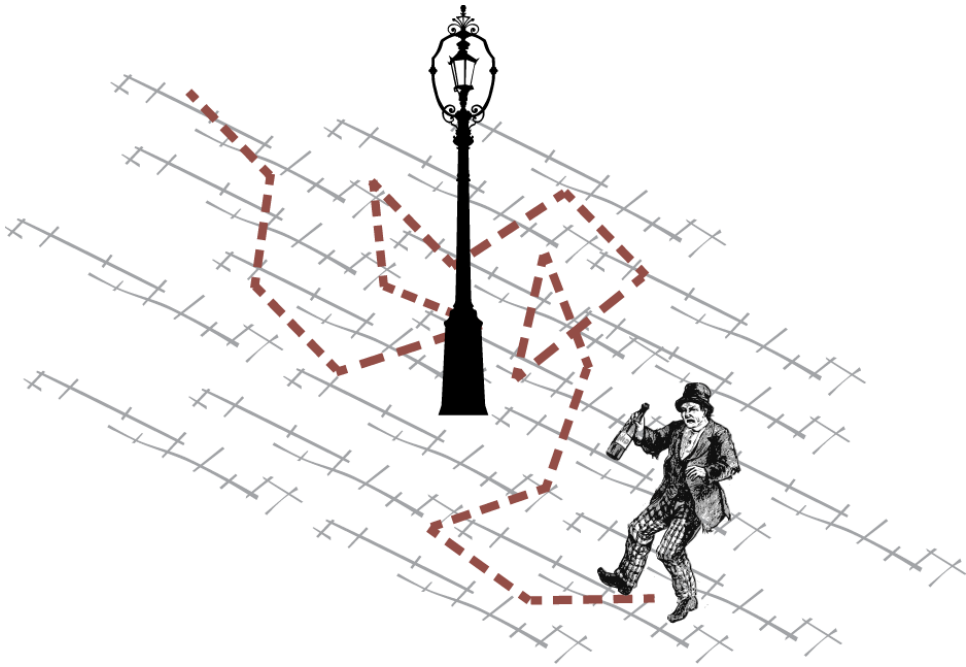
Skip-gram



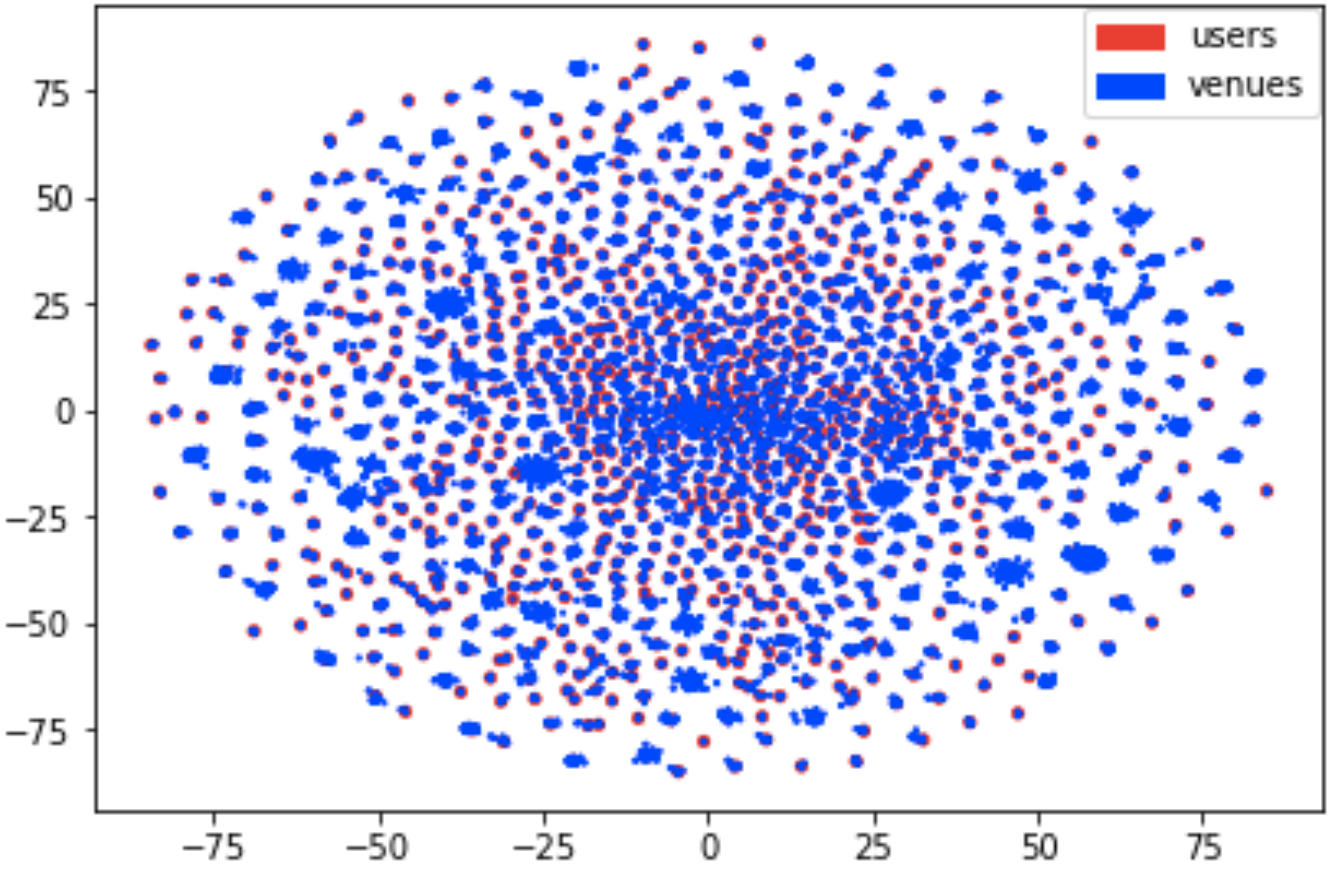
DeepWalk

- Random walks are “sentences”

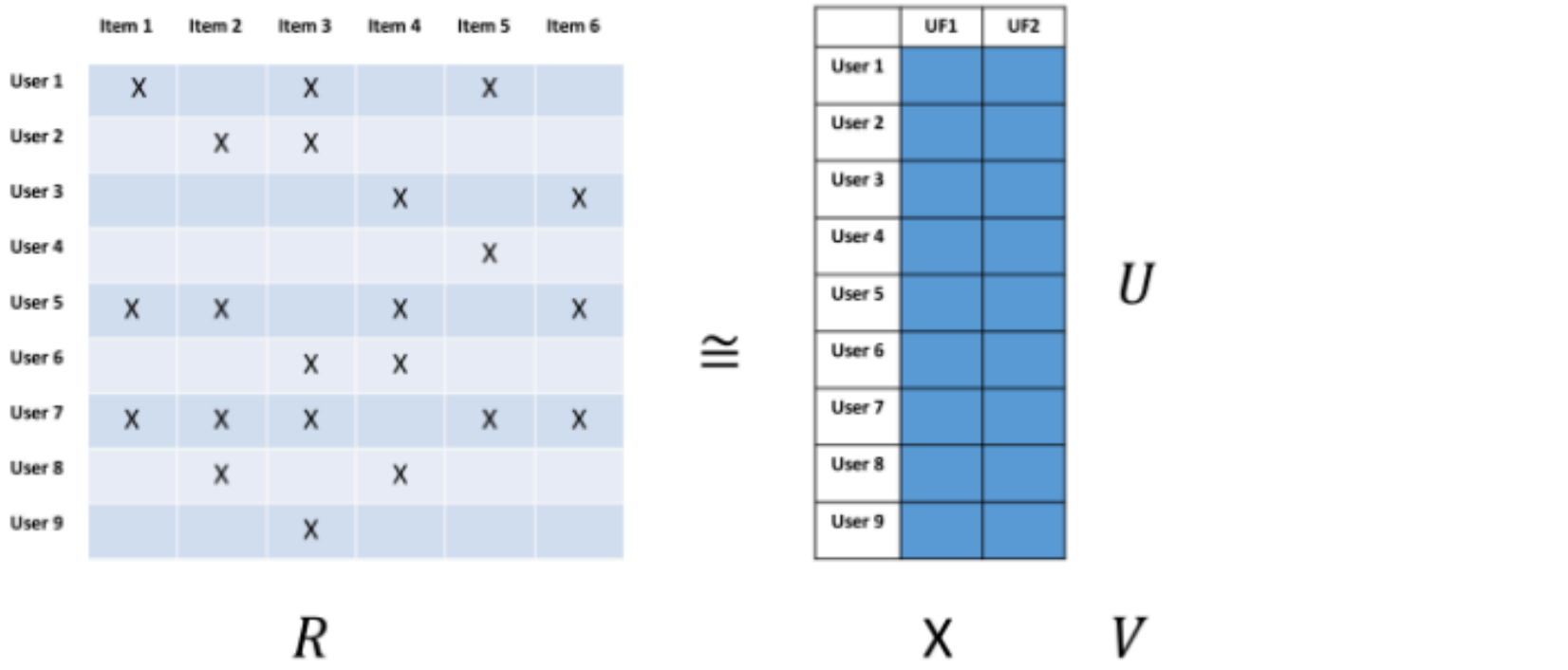




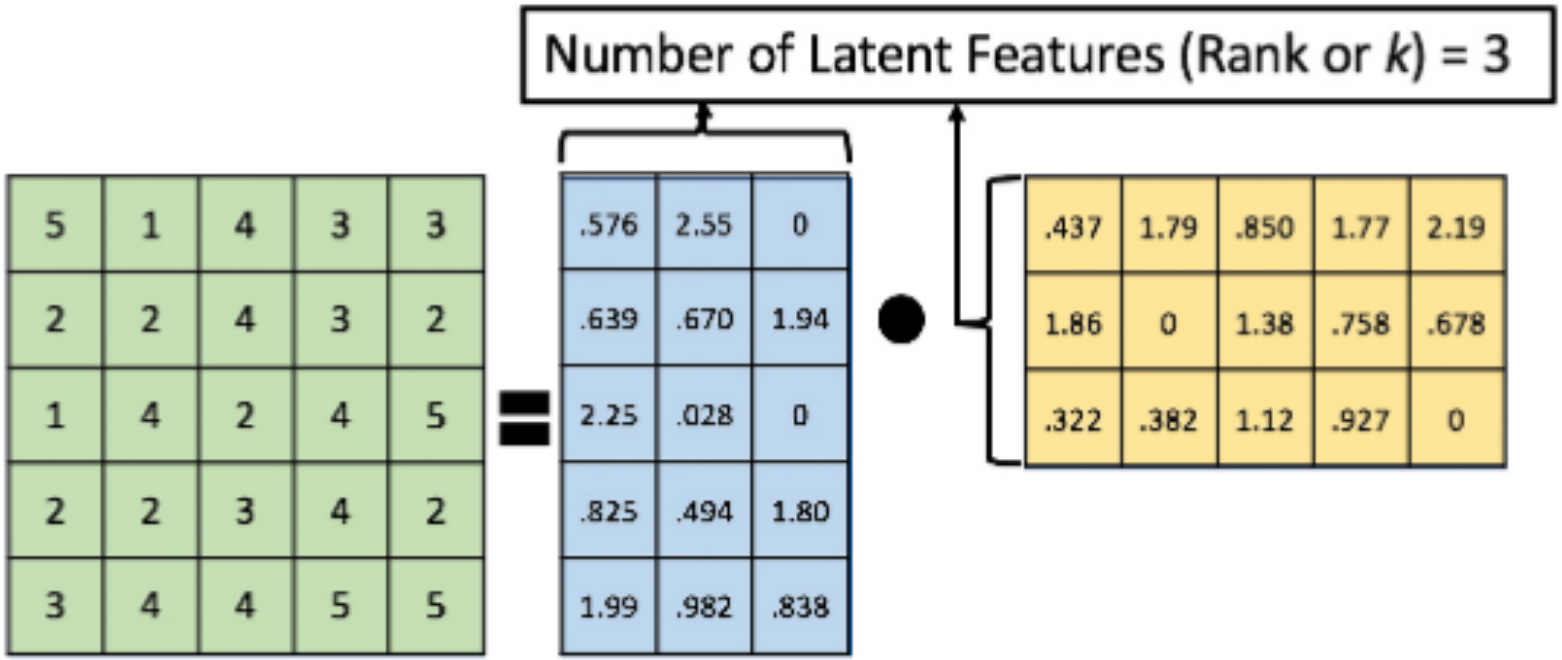
TSNE Visualization



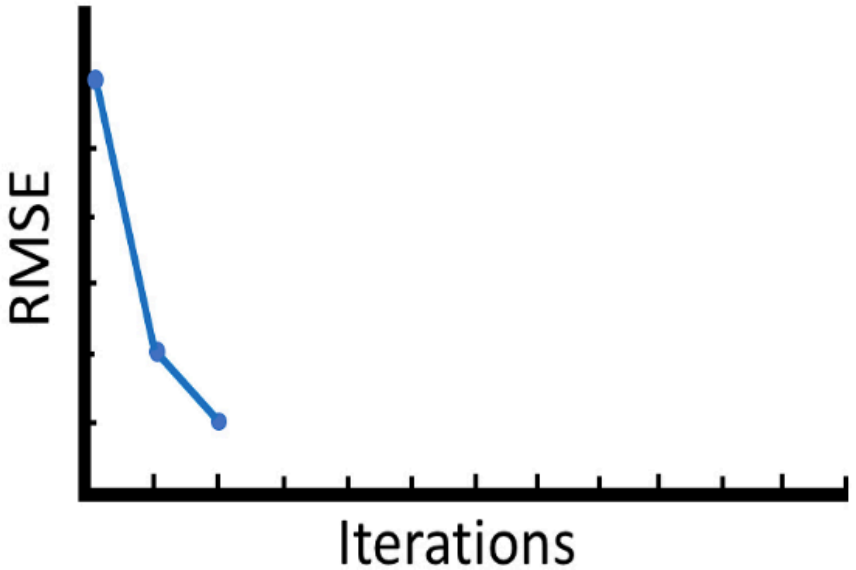
Matrix Factorization



Matrix Factorization

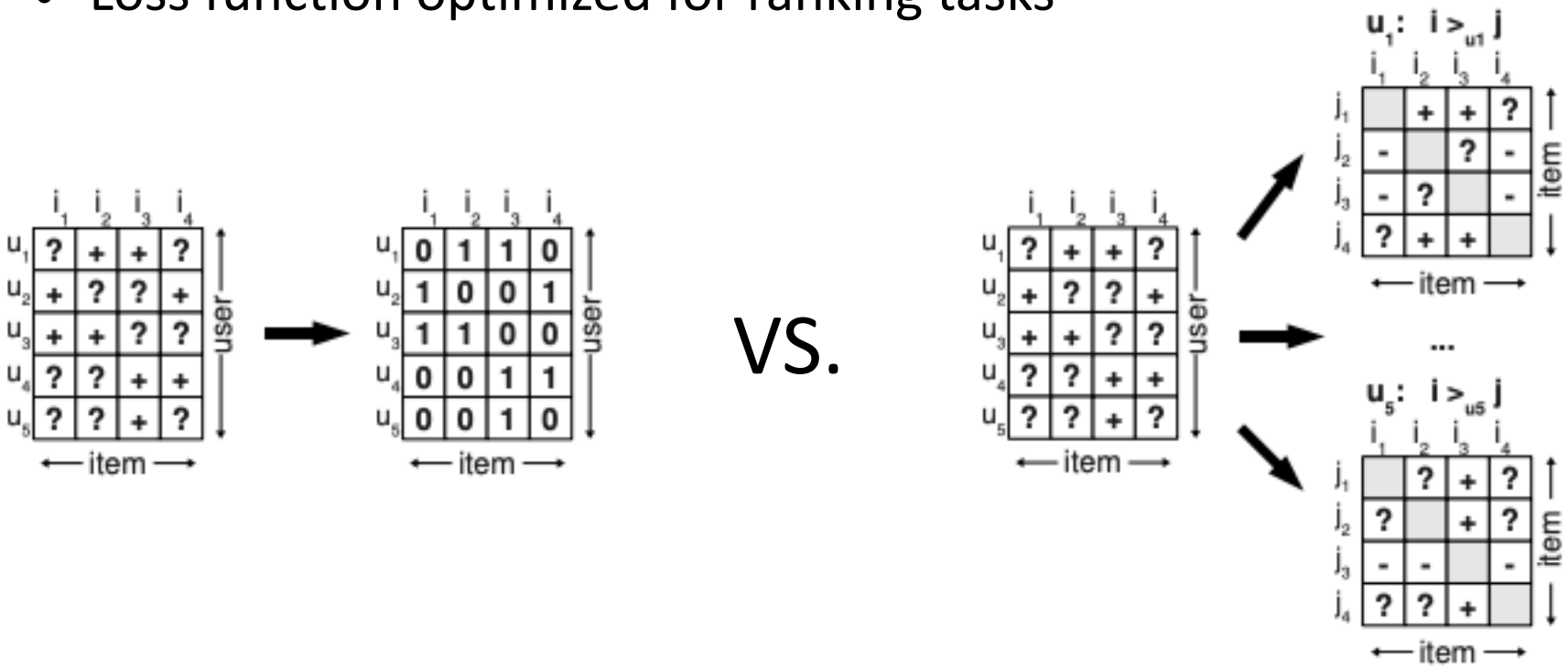


Alternating-Least-Squares



Bayesian Personalized Ranking

- Loss function optimized for ranking tasks



Ranking Criteria

- Deepwalk: Cosine similarity in embedding space
- Random: Rankings picked totally randomly
- ALS: Highest predicted values
- BPR: Highest Predicted values

Recommendation Examples

```
training_venues_print(85)
```

	User	Venue Name	Quantity
63968	85	La Bagel Delight	15
63978	85	Perelandra Natural Foods	14
63958	85	Sahadi's	9
63965	85	Borough Hall Greenmarket	5
63974	85	The Ensemble Studio Theatre	4
63992	85	Pronto Pizza	4
63976	85	My Little Pizzeria	3
63983	85	Zap Liquors & Spirits	3
63972	85	Heights Chateau	3
63970	85	Staubitz Market	2
63997	85	817 Broadway 10th Floor	2
63975	85	United Artists Court Street 12 & RPX	2
63990	85	Eataly	1
63985	85	Ensemble Studio Theater	1
63986	85	Alpine Scout Camp	1
63987	85	Delacorte Clock	1
63988	85	Ruben Liquor	1

Recommendation examples

```
get_deepwalk_recs(85)
```

Name	Similarity
Perelandra Natural Foods	0.979556
My Little Pizzeria	0.973256
817 Broadway 10th Floor	0.971372
Heights Chateau	0.970663
Poets House	0.967446
La Bagel Delight	0.967086
Court Street Bagels	0.965206
Alpine Scout Camp	0.963482
Sun Yat Sen Middle School MS 131	0.961378
Huge Meetup	0.959499

```
get_random_recs(85)
```

Name	Similarity
Consulate Of Bolivia	0.020615
Earl of Sandwich	0.082142
26 Court St	0.126363
Happy Nails & Spa	0.000608
Ralph Lauren	0.144673
LIRR - Jamaica Station	0.088661
Take-Two Interactive Software, Inc.	0.132039
Brooklyn Heights	0.129065
Original Pizza	0.177119
Wendy's	0.114440

```
get_bpr_recs(85)
```

Name	Prediction
Whole Foods Market	0.858127
The Bell House	0.842832
Landmark's Sunshine Cinema	0.838088
Dekalb Market	0.798090
Music Hall of Williamsburg	0.772296
NYU Skirball Center for Performing Arts	0.739677
d.b.a. Brooklyn	0.726816
United Artists Court Street 12 & RPX	0.719948
Peaches HotHouse	0.709125
La Colombe Coffee Roasters	0.708232

```
get_als_recs(85)
```

Name	Prediction
Park Slope Armory YMCA	0.349517
Trader Joe's	0.322353
Moment	0.294439
La Bagel Delight	0.292450
St. Francis College	0.269500
Perelandra Natural Foods	0.241178
MTA Subway - Jay St/MetroTech (A/C/F/R)	0.235247
Home :]	0.227509
MTA Subway - Beach 105th St/Seaside (A/S)	0.221946
St. Francis College Mailroom	0.219235

Evaluation Metrics

- Metrics are based on recall not precision
- Expected Percentile Ranking: $\overline{rank} = \frac{\sum_{u,i} rank_{ui}}{total\ \#\ of\ venues}$
- $\overline{rank} > 0.5$ is worse than random
- Mean Reciprocal Ranking: $\overline{reciprocal\ rank} = \frac{1}{total\ \#\ of\ venues} \sum_{u,i} \frac{1}{rank_{u,i}}$
- Evaluate using 20% of data held out for testing

Results

Model	EPR (all)	Rank	EPR (no repeats)	Rank	MRR (all)	Rank	MRR (no repeats)	Rank
ALS	0.096444	1	0.262232	1	0.122834	2	0.015795	1
Deepwalk	0.111273	2	0.329225	3	0.124465	1	0.001072	3
BPR	0.128094	3	0.295064	2	0.027435	3	0.010447	2
Random	0.516386	4	0.517186	4	0.000231	4	0.000173	4

Next Steps

- Hyperparameter Tuning to make models perform better
  - size of output embedding
  - # of walks/node
  - length of walks