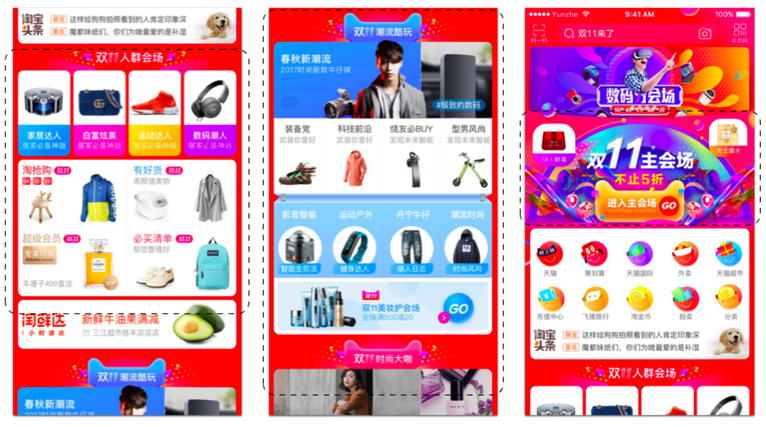
# Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba

Jizhe Wang Pipei Huang Huan Zhao
Zhibo Zhang Binqiang Zhao Dik Lun Lee

## RS IN TAOBAO

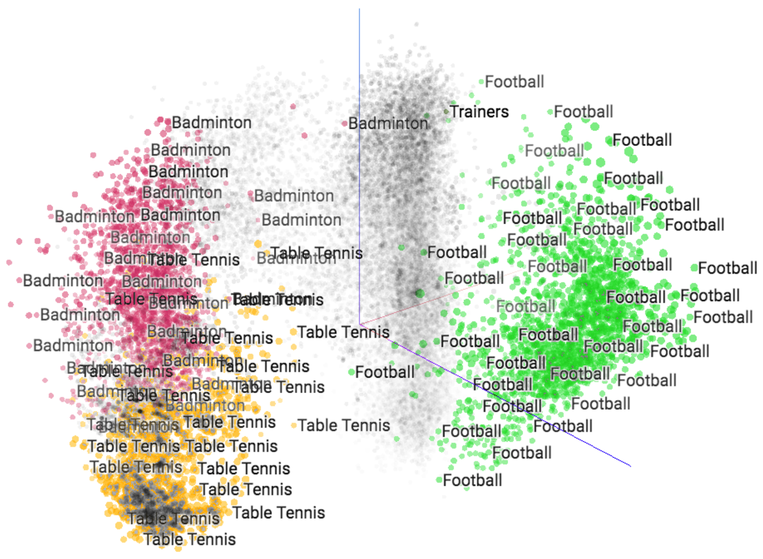
**RS on Mobile Taobao App Homepage**



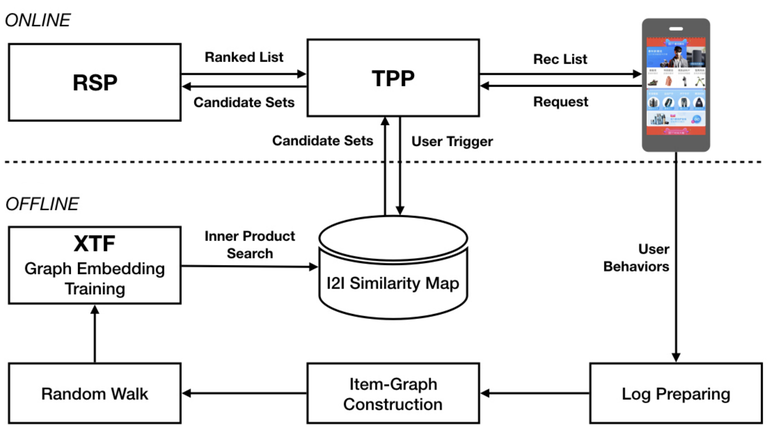
**Major challenges facing RS in Taobao**

- **Scalability**: Existing recommender system work well on smaller scale datasets, they fail on the much larger scale dataset in Taobao which has 1 billion users and 2 billion items.
- **Sparsity**: Extremely difficult to train an accurate recommending model since the interactions between users and items are sparse.
- **Cold Start**: Millions of new items are continuously uploaded each hour in Taobao. It is challenging to process these items or predict the preferences of users for these items.

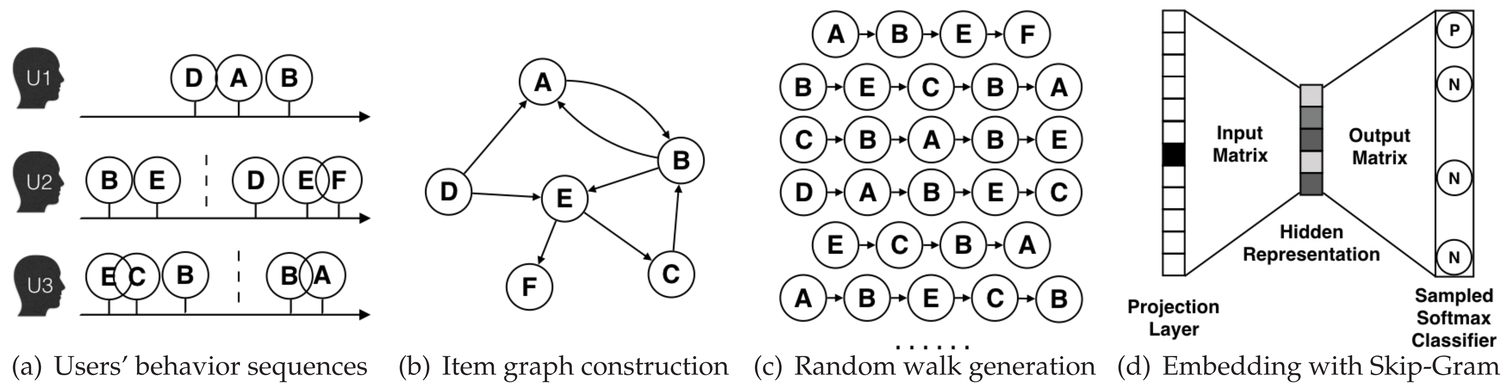
## VISUALIZATION




## SYSTEM DEPLOYMENT




## PROPOSED FRAMEWORK




(a) Users' behavior sequences (b) Item graph construction (c) Random walk generation (d) Embedding with Skip-Gram

- **Graph Construction:** We construct item graph from users' behaviors and filter out invalid data and abnormal behaviors to eliminate noise.
- **Sequence Generation:** By running random walk, we can generate a number of sequences.
- **Embedding Training:** Apply the Skip-Gram algorithm to learn the embeddings, which maximizes the co-occurrence probability of two commodities in the obtained sequences.

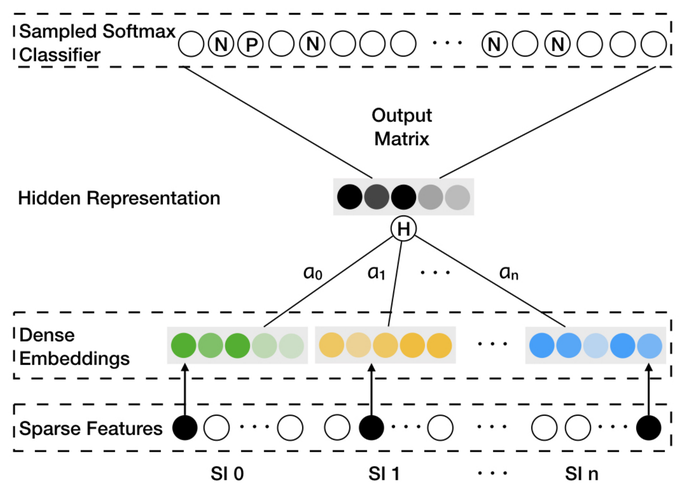
## EMBEDDING WITH SIDE INFORMATION

**Side-information**:In e-commerce, side information refers to the category, shop, purchase-level, material, etc.
**GES**: Add a layer with average-pooling operation to incorporate side information.

$$\mathbf{H}_v = \frac{1}{n+1}\sum_{s=0}^{n}\mathbf{W}_v^s$$

**EGES:** Different side information contribute differently to the co-occurrence of items in users' behaviors.

$$\mathbf{H}_v = \frac{\sum_{j=0}^{n} e^{a_v^j}\mathbf{W}_v^j}{\sum_{j=0}^{n} e^{a_v^j}}$$




## EXPERIMENTS

| Dataset | Amazon | Taobao |
|---|---|---|
| BGE | 0.9327 | 0.8797 |
| LINE(1st) | 0.9554(+2.43%) | 0.9100(+3.44%) |
| LINE(2nd) | 0.8664(-7.65%) | 0.9411(+6.98%) |
| GES | 0.9575(+2.66%) | 0.9704(+10.1%) |
| EGES | 0.9700(+4.00%) | 0.9746(+10.8%) |