information of multiple types of nodes and links. A snippet of this network is presented in Fig. 1. The blue squares denote different advertisers, green circles denote keywords bought by corresponding advertisers, red circles denote the keyword candidates retrieved from different data sources. The green lines denote the bidding relations between advertisers and keywords, blue lines denote the relations of co-occurring in same auctions between keywords, black lines denote the co-click relations, orange lines denote the adjacent relations in search sessions.

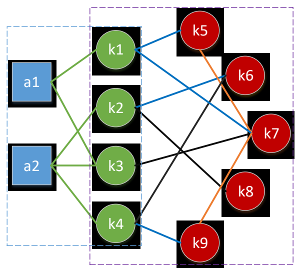


**Figure 1: A snippet of this multi-relational network with three types of nodes and four types of relationships.**

We collect and sample the data from a commercial search engine in two months to build the complete network, the statistics of this network can be found in Table 1,

**Table 1: Information of the datasets**

| | |
|---|---|
| Number of nodes | 123713 advertisers, 153291895 keywords |
| Number of links | 211478885 green links, 744963799 blue links, 985014321 black links, 428775103 orange links |

## 3. METHODOLOGY

### 3.1 Mining Informative Features

For this task, four types of features can be learned as follows,

*3.1.1 Meta-Path-Based Similarity Features.* Meta-paths [6] can be used to facilitate node proximity measurement in heterogeneous network. We adopt *normalized path count* and *path-constrained random walk* as the similarity features to represent the connectivity between advertisers and keyword candidates along different types of paths. Some possible meta-paths in Fig. 1 are: a1→k3→a2→k2, a1→k1→k5, a1→k3→k7, a1→k1→k7→k5.

*3.1.2 Network Embedding Features.* Since this multi-relational network is composed of four bipartite networks $\{V_A \cup V_B, E_k\}$ where the keyword vertices are shared across these networks. We can apply the second-order proximity [7] to learn the network embedding features as $min\{-\sum_{k=1\sim 4}\sum_{(i,j)\in E_k} w_{ij}.logp(v_i|v_j)\}$, where $p(v_i|v_j) = \exp(\vec{u}_i'^T.\vec{u}_j) / \sum_{k=1}^{|V|} exp(\vec{u}_k'^T.\vec{u}_j)$ denotes the conditional probability of vertex $v_i$ given vertex $v_j$.

*3.1.3 Semantic Features.* We use a deep learning neural network model [5] to measure the similarity of candidate keywords and advertiser's original keywords using historical clicked data. A click weighted similarity metric of different pairs of keyword and low-dimension semantic vectors are chosen as features.

*3.1.4 Statistical Features.* Some other features of advertisers and candidate keywords such as historical impressions and clicks, average cost per click, average post-click quality score, average click-through rate, biding and competition statistics.

### 3.2 Supervised Link Prediction Model

Based on the informative features mining from multi-relational network. We use the *Field-aware Factorization Machines* model [8] to model and predict if the (advertiser, keyword) pair has a relationship. The FFM model has the form as: $f(X) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(V_{i,f_j} \times V_{j,f_i}\right) x_i x_j$ and can be trained based on advertisers' historical good bidding keywords and can generate probabilistic score for each candidate keyword.

## 4. EVALUATION AND CONCLUSIONS

By splitting the dataset into two periods: first month for training and second month for testing, we deploy and verify our approach on the real sponsored search data and compared the top-20 recommendation performance with several existing methods such as random walk (*RW*) and matrix factorization (*MF*) in precision and recall metrics. The performance of different recommendation method is shown in Table 2.

**Table 2: Performance of keyword recommendation**

| Method | Precision | Recall |
|---|---|---|
| RW | 0.462 | 0.377 |
| MF | 0.535 | 0.396 |
| SLPM | **0.623** | **0.540** |

With rich representation of structure in the multi-relational network and capability of retrieving more candidates, the experimental results demonstrated the effectiveness of proposed approach in recommending suitable keywords to advertisers.

## 5. REFERENCES


[1] Fuxman A, Tsaparas P, Achan K, et al. Using the wisdom of the crowds for keyword generation. *Proceedings of the 17th international conference on World Wide Web.* ACM, 2008.

[2] Chen Y, Xue G R, Yu Y. Advertising keyword suggestion based on concept hierarchy. *Proceedings of the 2008 international conference on WSDM.* ACM, 2008.

[3] Zhang W, Wang D, Xue G R, et al. Advertising keywords recommendation for short-text web pages using Wikipedia. *ACM Trans. Intell. Syst. Techno*, 2012.

[4] Agrawal R, Gupta A, Prabhu Y, et al. Multi-label learning with millions of labels: Recommending advertiser bid phrases for web pages. *Proceedings of the 22nd international conference on World Wide Web.* ACM, 2013.

[5] Huang P S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data. *Proceedings of the 22nd ACM international conference on CIKM.* ACM, 2013.

[6] Meng C, Cheng R, Maniu S, et al. Discovering meta-paths in large heterogeneous information networks. *Proceedings of the 24th International Conference on World Wide Web.* ACM, 2015.

[7] Tang J, Qu M, Mei Q. Pte: Predictive text embedding through large-scale heterogeneous text networks. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2015.

[8] Juan Y, Zhuang Y, Chin W S, et al. Field-aware factorization machines for CTR prediction. *Proceedings of the 10th ACM Conference on Recommender Systems.* ACM, 2016.