X-Rank: Explainable Ranking in Complex Multi-Layered Networks

Jian Kang*
Arizona State University
Tempe, Arizona
jkang51@asu.edu

Scott Freitas*
Georgia Institute of Technology
Atlanta, Georgia
safreita@gatech.edu

Haichao Yu
Arizona State University
Tempe, Arizona
haichaoy@asu.edu

Yinglong Xia
Huawei Research America
Santa Clara, California
yinglong.xia@huawei.com

Nan Cao
Tongji University
Shanghai, China
nan.cao@gmail.com

Hanghang Tong
Arizona State University
Tempe, Arizona
hanghang.tong@asu.edu

ABSTRACT

In this paper we present a web-based prototype for an explainable ranking algorithm in multi-layered networks, incorporating both network topology and knowledge information. While traditional ranking algorithms such as PageRank and HITS are important tools for exploring the underlying structure of networks, they have two fundamental limitations in their efforts to generate high accuracy rankings. First, they are primarily focused on network topology, leaving out additional sources of information (e.g. attributes, knowledge). Secondly, most algorithms do not provide explanations to the end-users on *why* the algorithm gives the specific ranking results, hindering the usability of the ranking information. We developed X-Rank, an explainable ranking tool, to address these drawbacks. Empirical results indicate that our explainable ranking method not only improves ranking accuracy, but facilitates user understanding of the ranking by exploring the top influential elements in multi-layered networks. The web-based prototype (X-Rank: http://www.x-rank.net) is currently online—we believe it will assist both researchers and practitioners looking to explore and exploit multi-layered network data.

CCS CONCEPTS

• **Information systems → Content ranking**;

KEYWORDS

Ranking, knowledge, explainability, multi-layered network

ACM Reference Format:
Jian Kang*, Scott Freitas*, Haichao Yu, Yinglong Xia, Nan Cao, and Hanghang Tong. 2018. X-Rank: Explainable Ranking in Complex Multi-Layered Networks. In *The 27th ACM International Conference on Information and Knowledge Management (CIKM '18), October 22–26, 2018, Torino, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3269206.3269224

* The first two authors contributed equally to this work.

CIKM '18, October 22–26, 2018, Torino, Italy

ACM ISBN 978-1-4503-6014-2/18/10…$15.00
https://doi.org/10.1145/3269206.3269224

1 INTRODUCTION

A fundamental challenge in understanding complex networks is the task of ranking. Ranking is a core task that arises in multiple high impact areas, including: (i) web search, (ii) recommender systems, (iii) local graph partitioning and many more. Traditional ranking methods, such as PageRank [11] and HITS [8], are primarily focused on data graphs (e.g. co-authorship, co-purchase networks)—often acting as a black-box to the end-users who are not data mining experts. Since earlier research focuses largely on the accuracy of rankings in the data graph, there is a dearth of explainable ranking platforms that integrate the use of data and knowledge.

We define the problem of explainable ranking on a multi-layered network, by incorporating both data and knowledge layers in the network model. In complex multi-layered networks, it is often difficult to intuitively understand the ranking analysis. To address this challenge, we developed the X-Rank platform, which allows users to visually explore and exploit the ranking results.

Contributions. Our primary contributions are three-fold—(1) the development of a multi-layered network model that incorporates data and knowledge; (2) explainable ranking on the multi-layered network model, encompassing a suite of carefully chosen network mining algorithms; and (3) a web platform to visualize and explain the ranking analysis in complex multi-layered networks.

Demonstration. The X-Rank web platform will run live demonstrations and allow for open interactions with the audience. This includes the ability for audience members to (i) run various queries, (ii) explore the resulting rankings and explanations and (iii) interact with the network visualizations and ranking lists. Our goal is to spark a conversation on bringing the element of *explanation* to complex multi-layered networks analysis, including its algorithmic and systemic challenges, as well as the potential applications. An explanatory video of the X-Rank platform can be found online here: https://youtu.be/EAKPaCWJQxQ.

Related Work. To the best of our knowledge, X-Rank is the first online platform that allows for explainable ranking on complex multi-layered networks. Alternative network visualization tools exist, such as Gephi [1], MuxViz [3], PathFinder [5], Apolo [2] and Carina [4]. However, none of these tools are oriented towards web-based explainable rankings in complex multi-layered networks. In addition, the algorithmic back-end of this platform is based on work in explaining PageRank [7], network of networks model [10] and fast local subgraph identification [6].

2 SYSTEM ARCHITECTURE

The web platform has two main architectural components—(1) visualization and user controls (front-end) and (2) algorithmic processing (back-end). In Figure 2, we can see an architectural diagram illustrating a typical user interaction with the platform. The user interaction contains three primary steps: (i) the user selects a query node of interest from the dataset, (ii) a three step algorithmic process is performed on the back-end to obtain the query-sensitive explanatory rankings, and (iii) the rankings and explanations are sent to the front-end for visualization. The platform front-end, comprised of the user-interface and server, are implemented using HTML, Javascript, C# and vis.js—while the platform back-end, containing the algorithms, are written in Python.

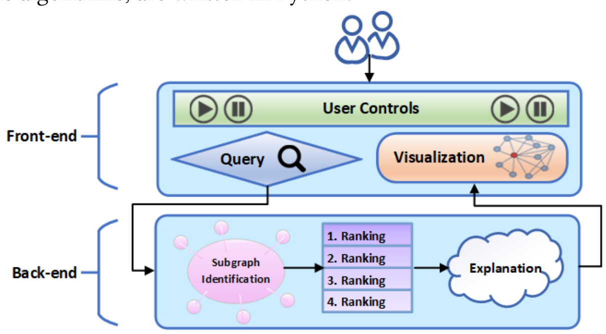

Figure 2: X-Rank platform architecture.

Front-end. In this section we discuss the implementation and development of the X-Rank web platform (corresponding to the top half of Figure 2).

In Figure 1, we can see a sample query being conducted on the X-Rank platform. The three primary components in the figure are—(1) network visualization on the left, (2) ranking tables in the upper right and (3) user controls in the lower right. (1) With respect to network visualization, the large circle represents the user query (this node can be right-clicked to explore the explanation information); small colored circles correspond to the ranking tables on the right (data layers); and rectangles correspond to the two knowledge layers. (2) The ranking tables contain the top ten products for each data layer (book, DVD, video and music) based on the ranking algorithm. In addition, the ranking tables contain additional information on data items compared to the network visualization, which contains only the rank (with the exception of the query node). (3) With respect to user controls, we allow the user to interact with two settings—(i) the query node of interest and (ii) which data/knowledge layers to utilize in the algorithmic processing.

Back-end. In this section we describe the algorithmic design behind the X-Rank platform (corresponding to the bottom half of Figure 2). Three fundamental algorithms comprise the core of the X-Rank platform, consisting of the following stages—(i) local subgraph identification around a given query node, using a variation of random walk with restart [6]; (ii) ranking of the resulting subgraph using the network of networks model and CrossQuery algorithm [10]; and (iii) explanation of the ranking using Aurora algorithms [7]. Each algorithm is run sequentially as seen in Figure 2..

3 TECHNICAL DETAILS

The proposed X-Rank algorithm encompasses three key algorithmic components—(i) subgraph identification (LocalProximity), (ii) ranking over multi-layered networks (CrossQuery) and (iii) ranking explanations (Aurora-E and Aurora-N). In this section, we briefly describe each of the three components, followed by the description of the overall X-Rank algorithm.

3.1 Subgraph Identification

Given a network and a query node in the network, the goal of this algorithm is to find a subgraph containing only the relevant nodes in the vicinity of the query node. By pruning the original network into a smaller one, we can significantly improve the computational efficiency of the ranking and explanation algorithms. Having this in mind, we utilize the LocalProximity algorithm introduced in [6] to identify a subgraph around the given query node before further computation.

Given a network $\mathbf{G}$, a query node q, number of trials n, and a relevance threshold parameter t_s—the LocalProximity algorithm will return a subgraph $\mathbf{T}$ around query node q. It works as follows:

(1) Compute the walk distribution L using random walk with restart [12] on $\mathbf{G}$ in n trials.
(2) For any node u in $\mathbf{G}$, if $L(u) > \mu(L) + \sigma(L)/t_s$, include the node into the subgraph, where $\mu(L)$ and $\sigma(L)$ are the mean and standard deviation of the walk distribution.
(3) Create a subgraph $\mathbf{T}$ based on the included nodes.

3.2 Ranking on Multi-layered Network

Given a multi-layered network $(\mathbf{A}, \mathbf{Y})$, where $\mathbf{A}$ and $\mathbf{Y}$ denote the within-layer connectivity and cross-layer dependencies respectively, and a query node q—the objective is to find the top-k ranked nodes in terms of their relevance to query q. In order to solve this problem, we utilize the network of networks model [10, 13]. In [10], the authors formulate ranking on a multi-layered network as random walk with restart with a closed-form solution: $\mathbf{r} = (\mathbf{I} - \tilde{c}\mathbf{W})^{-1}(1 - \tilde{c})\mathbf{e}$, where $\mathbf{W} = \frac{c}{c+2a}\mathbf{A} + \frac{2a}{c+2a}\mathbf{Y}$, $\tilde{c} = \frac{c+2a}{1+2a}$, and $\mathbf{e}$ is the teleportation vector with respect to query q; a, c are parameters to control the restart probability. Next, we run the CrossQuery algorithm introduced in [10] to obtain the top-k ranked nodes in a target layer d. CrossQuery performs power iterations to update the ranking vector, while shrinking the set of candidate nodes in each iteration until only k nodes remain. We summarize ranking on multi-layered networks as follows:

(1) Initialize the teleportation vector $\mathbf{e}$; parameter $\tilde{c} = \frac{c+2a}{1+2a}$, where $a = \frac{1}{4\lambda(\mathbf{Y})-2}$, $c = \frac{1+2a}{2\lambda(\mathbf{A})}$), and $\lambda()$ is the largest eigenvalue of the corresponding matrix.
(2) For each target data layer d in the multi-layered networks, run CrossQuery($\mathbf{W}$, $\mathbf{e}$, s, d, $\tilde{c}$, k), where s represents the source domain of query node and k is an integer budget.

3.3 Explaining Ranking on Multi-layered Network

A key characteristic of the X-Rank algorithm is that it provides a reasonable explanation as to *why* it gives such ranking results. This is accomplished by exploring the influence of key graph elements (e.g. edges, nodes), building upon the work in [7]. To measure the influence of a graph element, we define the influence by its gradient with respect to a loss function $f(\cdot)$ over the ranking vector $\mathbf{r}$. Therefore, the influence of edge (i, j) (Aurora-E) is defined as

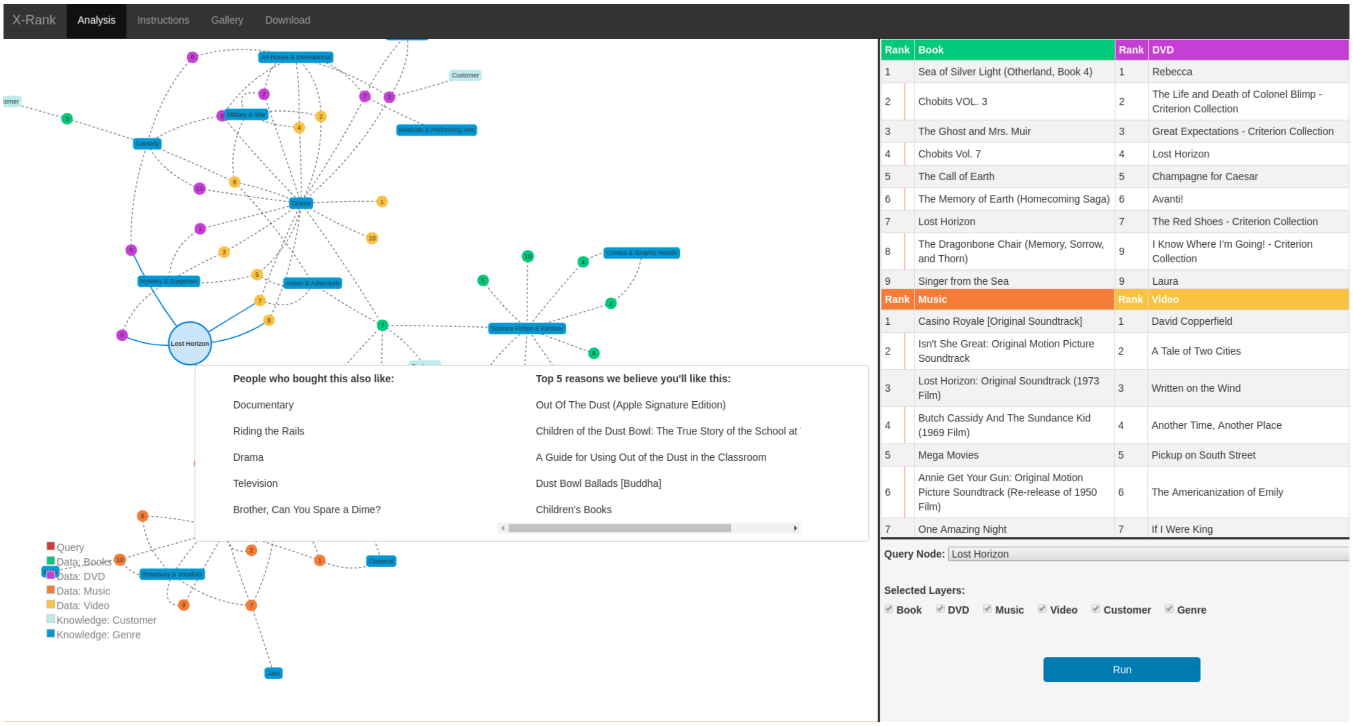

Figure 1: User interface of the X-Rank web-based prototype. Left side: network visualization, top-right: ranking lists, bottom-right: user controls. Query node can be right-clicked for explanations; node colors and shapes used to distinguish data and knowledge graphs.

$\mathbb{I}(i,j) = \frac{df(\mathbf{r})}{d\mathbf{W}(i,j)}$, where $f(\mathbf{r}) = ||\mathbf{r}||_F^2$. Similarly, we define the influence of a node (Aurora-N) as the aggregation of all inbound and outbound edges that connect to the node, $\mathbb{I}(i) = \sum_{j=1}^{n} [\mathbb{I}(i,j) + \mathbb{I}(j,i)]$, where n is the number of nodes in the network. In summary, explaining the ranking on multi-layered network is as follows:

(1) Given a multi-layered network $(\mathbf{A}, \mathbf{Y})$ and an integer budget k, run Aurora-E to obtain top-k influential edges.
(2) Given a multi-layered network $(\mathbf{A}, \mathbf{Y})$ and an integer budget k, run Aurora-N to obtain top-k influential nodes.

3.4 Overall Algorithm Description

Combining the algorithmic components from the above three subsections (as shown in Fig. 2), we present the proposed explainable ranking algorithm X-Rank.

After running the algorithm, $\mathcal{R}$ will be used to generate the ranking lists shown in our system (right side in Fig. 1). $\mathcal{E}$ and $\mathcal{V}$ will be used to explain the ranking results $\mathcal{R}$, which correspond to "People who bought this also like" and "Top 5 reasons we believe you'll like this" in Fig. 1.

Algorithm 1: X-Rank

Input: a single-layered network $\mathbf{G}$, a query node q, two integer budgets k_1, k_2, and parameters a, c.
Result: a set $\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_g)$, where $\mathcal{R}_i$ represents the set of top-k_1 ranked nodes in i^{th} layer and g is the number of data layers, a set of k_2 influential edges $\mathcal{E}$, a set of k_2 influential nodes $\mathcal{V}$.

```
Initialize the number of trials n, the relevance threshold t_s;
T = LocalProximity(G, q, n, t_s);
Transform T into a multi-layered network (A, Y);
Initialize W = c/(c+2a) A + 2a/(c+2a) Y, c̃ = (c+2a)/(1+2a), s := source domain
    of q, e := teleportation vector with respect to q, R = ∅;
foreach layer d ∈ data layers (1,2, ..., g) do
    R_d = CrossQuery(W, e, s, d, c̃, k_1);
    R = R ∪ {R_d};
end
E = Aurora-E(A, Y, a, c, k_2);
V = Aurora-N(A, Y, a, c, k_2);
return R, E, V;
```

4 EMPIRICAL EVALUATIONS

To determine the effectiveness of the X-Rank algorithm we conducted a user study comparing it to multiple baseline techniques. The goal of the study was to acquire feedback on both the ranking relevance and explainability of X-Rank.

Experimental setup. *Experiment I.* We compare X-Rank to three baseline techniques: (i) random walk with restart (RWR) [12], (ii) HITS [8], and (iii) CrossQuery [10] without knowledge layers. Each user is asked to run a set of 6 queries on each algorithm and compare the relevance of the top-*10* ranking results. *Experiment II.* We ask users to rate the usefulness of explanations generated by X-Rank and baseline techniques in understanding the top-k ranking results. We note that the baseline techniques were combined with Aurora-E and Aurora-N to generate the explanations.

All results are scored on a scale of 1-5, with 1 being the least relevant/helpful and 5 being the most. Results were gathered from 11 expert judges (users that have general background knowledge in a data mining related field). We choose to utilize expert judges

Query	Type	X-Rank		RWR		HITS		CrossQuery	
		Mean	STD	Mean	STD	Mean	STD	Mean	STD
Ghost World	Book	**3.43**	1.30	3.23	1.08	3.07	1.10	3.02	0.92
The Making of Pride and Prejudice	Book	**4.02**	0.87	3.80	0.94	3.66	0.90	3.61	0.96
Fear and Loathing in Las Vegas	Book	**3.77**	0.97	3.63	0.91	3.43	1.12	3.16	0.93
Lost Horizon	DVD	**4.10**	0.95	4.05	0.88	3.22	0.76	3.0	0.95
Tai Chi Music - Dr. Paul Lam	Music	**4.36**	0.68	4.11	0.91	3.57	1.05	3.81	0.68
The American Experience	Video	3.50	1.03	**3.64**	0.96	3.00	0.93	3.11	1.09

Table 1: Results of user study. The term 'STD' in the table stands for standard deviation.

as they are more representative of the demo paper's primary user group. In addition, all experiments were performed in Windows 10 with a 3.4GHz i7-6700 CPU and 32GB memory.

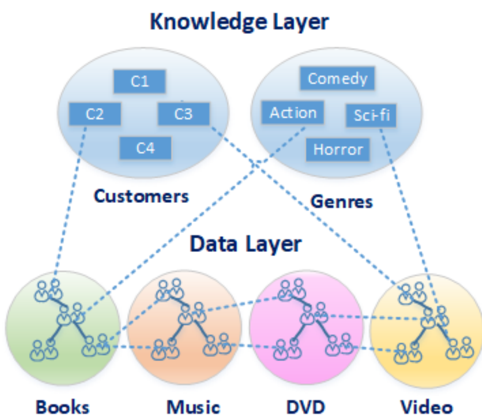

Figure 3: Example of the multi-layered network in Amazon co-purchase dataset.

Dataset. We utilize the Amazon co-purchase dataset (548,552 nodes and 1,788,725 edges) from [9] to conduct the user study. The co-purchase network contains four categories (book, DVD, music and video), each of which is used to construct a data graph as shown in Figure 3. Cross network connections (dotted lines) represent a co-purchase between different categories. In addition, we utilize the genre and customer review metadata contained in the dataset to construct the knowledge layer. Cross layer connections between knowledge and data layers represent the corresponding genre(s) and customer(s) of each product.

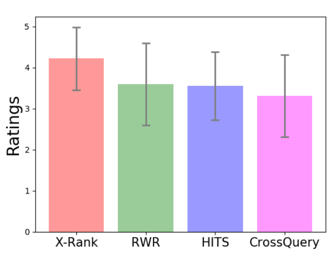

Figure 4: Explainability ratings of all compared methods. Higher is better.

Results. Findings from the user study are reported in Table 1. From this table we offer the following observations: (1) in 5 out of 6 queries, X-Rank performs the best among all compared methods; (2) comparing X-Rank and CrossQuery, the significant improvement indicates the effectiveness of adding knowledge layers. Furthermore, when measuring the usefulness of explanations from X-Rank, the users gave the explanations an average rating of 4.22 (out of 5). This is significantly higher than the ratings for RWR, HITS and CrossQuery—which are 3.60, 3.55, 3.31, respectively. Results are shown in Figure 4. This demonstrates the potential of the proposed X-Rank algorithm to provide useful and intuitive explanations. In addition, the X-Rank algorithm is capable of scaling to large networks due to its linear complexity. To see this, we note that LocalProximity, CrossQuery, Aurora-E and Aurora-N all have linear complexities, which renders a linear time complexity of the overall X-Rank algorithm.

5 CONCLUSION

The goal of this work is to develop a web-based prototype (X-Rank) for researchers and practitioners to visually explore and interact with the proposed explainable ranking algorithm. We believe the platform and algorithm will be of particular interest to both researchers and practitioners in the fields of information retrieval and data mining. In addition, an operational prototype of the X-Rank platform is currently online (http://www.x-rank.net), along with a demonstration video (https://youtu.be/EAKPaCWJQxQ). Source code will be made publicly available by the conference date.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation under Grant No. IIS-1651203, IIS-1715385, CNS-1629888 and IIS-1743040, DTRA under the grant number HDTRA1-16-0017, Army Research Office under the contract number W911NF-16-1-0168, Department of Homeland Security under Grant Award Number 2017-ST-061-QA0001, National Natural Science Foundation of China under the grant number 61602306, and gifts from Huawei and Baidu.

REFERENCES

[1] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. 2009. Gephi: an open source software for exploring and manipulating networks. (2009).

[2] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. 2011. Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning. In *CHI*. ACM, New York, NY, USA, 167–176.

[3] Manlio De Domenico, Mason A Porter, and Alex Arenas. 2015. MuxViz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks* 3, 2 (2015), 159–176.

[4] Dezhi Fang, Matthew Keezer, Jacob Williams, Kshitij Kulkarni, Robert Pienta, and Duen Horng Chau. 2017. Carina: Interactive Million-Node Graph Visualization using Web Browser Technologies. *CoRR* abs/1702.07099 (2017). arXiv:1702.07099

[5] Scott Freitas, Hanghang Tong, Nan Cao, and Yinglong Xia. 2017. Rapid Analysis of Network Connectivity. In *CIKM*. ACM, New York, NY, USA, 2463–2466.

[6] Scott Freitas, Hanghang Tong, Nan Cao, and Yinglong Xia. 2018. Local Partition in Rich Graphs. *CoRR* abs/1803.05084 (2018). arXiv:1803.05084

[7] Jian Kang, Hanghang Tong, Yinglong Xia, and Wei Fan. 2018. AURORA: Auditing PageRank on Large Graphs. *CoRR* abs/1803.05068 (2018). arXiv:1803.05068

[8] Jon M. Kleinberg. 1999. Authoritative Sources in a Hyperlinked Environment. *JACM* 46, 5 (Sept. 1999), 604–632.

[9] Jure Leskovec and Rok Sosič. 2016. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *TIST* 8, 1 (2016), 1.

[10] Jingchao Ni, Hanghang Tong, Wei Fan, and Xiang Zhang. 2014. Inside the atoms: ranking on a network of networks. In *KDD*. ACM, 1356–1365.

[11] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[12] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. 2008. Random walk with restart: fast solutions and applications. *Knowledge and Information Systems* 14, 3 (2008), 327–346.

[13] Si Zhang and Hanghang Tong. 2016. Final: Fast attributed network alignment. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1345–1354.