FPRank: Learning to Predict the Future Popularity Ranking of Scientific Publications

Lequn Wang
Shanghai Jiao Tong University
lqwang1994@gmail.com

Yuning Mao
Shanghai Jiao Tong University
morningmoni@sjtu.edu.cn

Zhenyu Song
Shanghai Jiao Tong University
sunnyszy@sjtu.edu.cn

Yuting Jia
Shanghai Jiao Tong University
Jiahnxxjyt@sjtu.edu.cn

Fengyu Deng
Shanghai Jiao Tong University
dengfengyu@sjtu.edu.cn

Hongtao Lin
Shanghai Jiao Tong University
Linkin_HearT@sjtu.edu.cn

Luoyi Fu
Shanghai Jiao Tong University
yiluofu@sjtu.edu.cn

Xinbing Wang
Shanghai Jiao Tong University
xwang8@sjtu.edu.cn

ABSTRACT

Predicting the future popularity of scientific publications enable researchers to discover high quality papers and potentially promising research directions. However, the fast-growing volume and the dynamic nature of scientific publications make it extremely hard to acquire a reasonable prediction. In this work, we propose a new ranking framework FPRank to deal with two problems: (1) Future Citation Ranking which means ranking scientific publications in the whole academic network by thier predicted future popularity and (2) Zero Citation Ranking which is to rank only newly published literatures without citation information. FPRank considers the dynamic time-aware representations of papers and is able to incorporate previous graph-based ranking models and features in citation count prediction methods into a unified learning-to-rank framework. The number of future citation each paper will receive is taken as the ground truth future popularity and time is regarded as query in learning-to-rank. We show empirically that FPRank significantly improves performance over state-of-the-art future popularity prediction methods.

1. INTRODUCTION

Many researchers concern about "which papers will be popular in the near future?", for these papers can not only provide potentially promising research directions, but enables discovering academic rising stars and promote academic reward system as well. Nevertheless, nowadays it is increasingly hard to discover potentially popular papers since the fast-growing volume of scientific literatures makes it impossible for researchers to skim through all those papers. As a result, automatically predicting the future popularity of scientific publications has received considerable attention in these years. Previous popularity prediction methods can be mainly divided into graph-based ranking models and citation count prediction methods.

There have been many graph-based models proposed for scientific publication ranking. Specifically, based on the paper citation netowkr, some recent graph-based models constructed heterogeneous academic networks (an example is illustrated in Figure 1) and took into consideration publication time [1, 2], text [3], author [4, 2, 5], and venue [2] information in the construction of academic networks. They incorporated the extra information into the academic network by either assigning weights on the edges or adding new vertices. Then a random walk algorithm like *PageRank*[6] or *HITS* [7] is run to get the final score of each vertex which is used for ranking. A common assumption shared by all these models is that the popularities of papers, authors and venues are mutually reinforced.

While the aforementioned works leveraged information from different factors (e.g. author, venue), they still suffer from several obvious drawbacks. First, the models were all run in an unsupervised manner while future citation number, which was a good metric to measure future popularity, was not used. Though the assumption of mutual reinforcement seems reasonable, the graph is a constraint where the factors can only be mutually reinforced in the way defined by the graph. Moreover, they treated the graph as static and thus ignored the dynamic and evolving nature of the academic network.

Another line of works on future popularity prediction is Citation Count Prediction (CCP). Figure 1 shows the basic structure for CCP. The main idea of CCP is to first extract features from the academic network and then use regression models to predict the future citation count for each paper. Many researchers discovered different kinds of predictive features such as textual content [8, 9], network features [10] and graph evolution rules [11]. They also tried different regression models like linear regression or support vector regression for prediction.

Though these methods use massive of features and learned prediction models in a supervised way, they did not consider the dynamic time-aware representations of the papers. For example, topic popularity of a paper intuitively varies with

DOI: http://dx.doi.org/10.1145/0000000.0000000

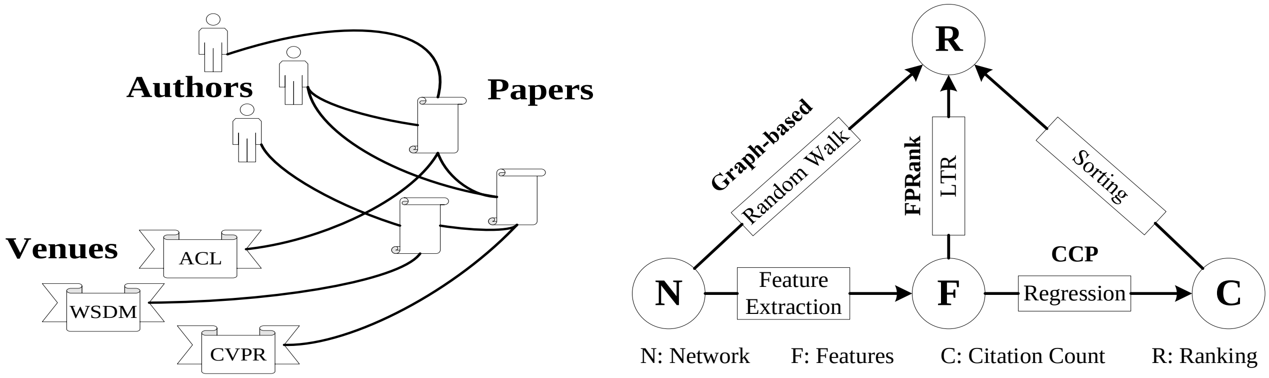

Figure 1: A sketch that shows the structure of three kinds of popularity prediction methods, namely graph-based ranking models, CCP and FPRank. Graph-based methods run a random walk algorithm on the academic network to get the ranking. CCP and FPRank both need to extract features from the network. CCP leverages conventional regression models to predict future citation count while FPRank uses Learning-to-Rank technique to get a ranking.

time since the topic evolves quickly. Capturing the topic popularity each paper in different time periods enables the models to better learn to what extent topic popularity can possibly contribute to future popularity. Features of other factors also possess this kind of dynamic nature. In addition, it is much harder to predict the exact citation count for each paper than to give papers a ranking due to the noise in the dataset and lack of information. An intuition example is that it is easier for a researcher who has recently read 2 papers in his field to tell which one will gain more citations than to predict the exact number of citations for both papers. As we will subsequently show in section *Ranking Is Reasonable*, the predictions of CCP methods are neither robust nor accurate while those of FPRank are reasonable.

To address the limitations of both graph-based ranking models and CCP methods, we propose a new ranking framework called FPRank. FPRank tackles both Future Citation Ranking(FCR) problem, which means ranking scientific publications in the whole academic network by their predicted future popularity, and Zero Citation Ranking(ZCR) problem, which is to particularly rank newly published papers without citation information. The number of future citation count is taken as ground truth popularity.

Note that, though we can get a ranking list of ZCR directly from FCR by simply extracting newly published papers from the ranking list of FCR, we still particularly focus on ZCR since the features extracted are always biased to old papers in previous methods. These methods are all designed to predict the popularity of papers in the whole network, and therefore tend to neglect the ranking results of the latest papers without citation information, since these papers are only a small portion of the dataset and make little contribution to the performance measure function. For example, graph-based models tend to give new papers low scores since these papers do not have in degrees in the academic network and we can never access to these papers in the graph through random walk. As a result, they tend to give unreasonable predictions to these new papers. However, We believe that ranking such papers could benefit researchers in the community. Some researchers, especially researchers that are domain experts who have already mastered important publications in the their fields, care more about promising new papers than older ones which they have known. Consider the following situation: after *WSDM 2017* is held, hundreds of papers will be published in the proceedings. It is a natural question to ask, "which one will receive the largest number of citations after several years?". Answering this question provides these researchers with easier ways to discover new views and research topics.

In this work, FPRank is proposed to deal with the two problems. FPRank is a two-step framework as illustrated in Figure 1. First, time-aware features are extracted from the academic network. And then learning-to-rank (LTR) technique is leveraged to rank the papers according to their future citation count. FPRank considers dynamic, time-aware representations by extracting time-aware features from the dynamic academic network. In addition, FPRank can combine the results of previous graph-based ranking models and features in citation count prediction methods into a unified LTR framework. Compared to graph-based ranking models, FPRank can break the constraints imposed by the graph by learning weights of each factor with future citation count as ground truth. Compared to CCP methods, FPRank is more robust in terms of future popularity prediction since it deals with a relative easier problem.

We did extensive experiments on a public dataset to empirically show that FPRank is much more competent in terms of both FCR problem and ZCR problem compared with graph-based ranking models and CCP methods. We also show that the predictions of CCP methods are neither robust nor accurate while those of FPRank are more reasonable. What is more, we found that author plays the dominant role in determining whether a paper will gain a large number of citations in the future by experimenting on different set of features.

The contributions are summarized as follows:

- We consider dynamic, time-aware representations of scientific publications for ranking and propose methods for extracting time-aware features.

- We establish a new scientific publication ranking framework FPRank which leverages time-aware features and LTR technique. The framework is able to combine results of graph-based ranking models and features in citation count prediction methods.

- We analyze features of different factors in LTR and find that author plays the dominant role in determining the future popularity of a paper in the filed of natural language processing.

The rest of this paper is organized as follows, Section 2 presents literature review on scientific impact ranking, citation count prediction and LTR. Section 3 introduces the proposed FPRank framework and the ways we extract time-aware features. Section 4 describes the setup and the results of the empirical evaluation of our model. At last, we give concluding remarks in Section 5.

2. RELATED WORK

2.1 Scientific Impact Ranking

The ordering of scientific literatures has been under intensive study for long. Based on the methodology of both *PageRank* [6] and *HITS* [7], a flurry of works modeled citation network as *WWW*, where they utilized graph based methods for prestige discovering [12, 13, 14]. However, these solutions only focus on homogeneous citation network, while ignoring a lot of other useful information.

In the next period, lots of methods aimed to leverage information like authors or venues to construct a heterogeneous network. For example, Zhou *et al.* [15] proposed a method for co-ranking authors and their publications using co-author network, citation network and authorship network. Yan *et al.* [16] introduced *P-rank* to rank scholarly prestige based on the network of papers, authors and journals. Yet these methods ignore the dynamic nature of citation network, making them biased to old articles. Liang *et al.* [5] proposed *HHGBiRank* based on the multinomial heterogeneous academic hypernetwork which models the citation relationship between authors in a more natural way.

Huge efforts have been made to address the dynamic nature of citation networks. Walker *et al.* [1] presented a network traffic model CiteRank to mimic researchers starting their research from recent publications with probability proportional to:

$$\rho_i = e^{\mathrm{age}_i/\tau_{\mathrm{dir}}},$$

where age_i is the time between publication an current year, τ_{dir} is a constant. By adding an exponential weight, it remedies the aging effect. Moreover, Sayyadi and Getoor [4] introduced their model *FutureRank*, involving a random walk on citation network and author-paper network with a time-discounted preference vector. [2] defined a ranking algorithm utilizing citations, authors, venues, and charactered the publication time as time-discounted edge. Wang *et al.*[3] presented *MRFRank* which constructed time-aware weighted graph and leveraged words and words co-occurrence to model the innovativeness of a paper.

Though they considered the chronological information, they still ignored the dynamic representations of papers. In addition, the scores of the papers were learned in an unsupervised manner while future citation count, which is a good metric for future popularity was only considered in model selection part. These issues can be handled with FPRank well.

2.2 Citation Count Prediction

Another direction for estimating the future potential of scientific publications is to predict a paper's future citation count. This direction has gained increasing attention since the citation count is a common metric for a paper's popularity and used frequently as a measurement for scholarly impact.

The early works on citation count prediction, including [17, 18, 19], extracted heuristic features and adopted simple models such as linear regression on relatively small datasets with fewer than one thousand publications.

Later on,Fu and Aliferis [8] utilized the information in the textual content by identifying important and discriminative keywords in the text. Yan *et al.* [9] studied more sophisticated features such as authors' productivity, sociality, *h-index* and venue's centrality. They found that the "author" and "venue" determined whether a paper was attractive while the content features in isolation were not predictive. Following this direction, Didegah and Thelwall [20] found that venue prestige played a central role in determining a paper's future citation count. Livne et al. [10] also tried network features and found that the network features were the most predictive in many fields. Pobiedina et al. [11] regarded citation count prediction as a link prediction problem and found that the graph evolution rules were powerful features. Some researchers made stronger assumptions on the features and tried to predict individual paper citation count over time. They proposed generative probabilistic models using a reinforced Poisson process or Point process to model the process through which individual paper gained citations so as to predict individual paper citation count over arbitrary time.[21, 22, 23]

However, these methods still did not take into account the dynamic, time-aware representations of the publications. In our work, we tackle the problem of ranking rather than the exact citation count due to noises in the dataset and lack of full information. In addition, different from [20], we found "author" played a dominant role in determining a paper's future popularity.

2.3 Learning to Rank

The idea of Learning-to-Rank (LTR) is to learn ranking functions automatically by applying machine learning techniques. In a typical setting, a machine learning algorithm is used in the training data consisting of queries, documents (represented by features), and their labels to make the specified loss function minimized. After training, the model can then predict rankings for documents in new queries.

Existing LTR algorithms can be categorized into three approaches by their input representation and loss function: the pointwise , pairwise, and listwise approaches [24]. Typical LTR algorithms include Ranking SVM [25], LambdaMART [26] and etc.

LTR algorithms are usually used to improve retrieval performance. However, there is also work leveraging LTR for academic ranking. For example, [27] aims to model authors' authority by incorporating LTR into topic model.

3. METHOD

3.1 Learning to Rank

In the setting of learning to rank, a set $\boldsymbol{X}$ of n queries $\boldsymbol{x}_i \sim P(\boldsymbol{x})$ is given and we have the relevances $r(\boldsymbol{x}, y)$ between each query $\boldsymbol{x}$ and document y. We need to design a utility function $U(\boldsymbol{x}, \boldsymbol{y})$ which measures the quality of the ranking $\boldsymbol{y}$ for query $\boldsymbol{x}$. Then the overall performance of a ranking function R that returns rankings $R(\boldsymbol{x})$ is defined by the expected utility

$$U(R) = \int U(\boldsymbol{x}, R(\boldsymbol{x}))dP(\boldsymbol{x}). \quad (1)$$

We are going to find a ranking function $R \in \boldsymbol{R}$ that maximizes $U(R)$ using $\boldsymbol{X}$ and the relevances $r(\boldsymbol{x},y)$. Since $U(R)$ cannot be computed directly, it is typically estimated via the empirical risk

$$\hat{U}(R) = \frac{1}{|\boldsymbol{X}|} \sum_{x_i \in \boldsymbol{X}} U(\boldsymbol{x}_i, R(\boldsymbol{x}_i)). \quad (2)$$

The typical estimation strategy is Empirical Risk Minimization, which corresponds to choosing the ranking function $\hat{R}$ that optimizes the empirical risk

$$\hat{R} = \text{argmax}_{R \in R} \hat{U}(R). \quad (3)$$

The time complexity of LTR is high when the number of documents in one ranking is large. To address this problem, one way is to use pooling in which only the top few documents retrieved by some simple ranking model are reordered. Another way is to sample a part of the documents each time so that the number of documents in a single ranking is affordable.

3.2 Future Citation Ranking

We argue that the papers should have different representations in different years. As a result, the features extracted must be time-aware. More specifically, the features of each document are query-aware in LTR, where query is publication year. The popularity of a publication is highly related to its citation count. Therefore, we assume that the future citation count of each paper is its ground truth future popularity. In Future Citation Ranking, our goal is to predict the ranking of papers in the whole academic network according to the number of future citations they will obtain p years from the current year t_{cur}.

In order to extract time-aware features and construct the training set, we shift our current time t from t_0 to $t_{\text{cur}} - p$. For each year t, we hide all the information after t to obtain a paper set PS_t. We then extract features for each paper in PS_t using the information in PS_t. By applying such method, we can obtain a training set containing $t_{\text{cur}} - t_0$ queries

$$\boldsymbol{X} = \{(\boldsymbol{x}_t^i)_{i=1}^{n_t}\}_{t=t_0}^{t_{\text{cur}}-p} \quad (4)$$

where $\mathbf{x}_t^i$ is the feature vector for paper i in PS_t, n_t is the number of papers in PS_t. In practice, we observe that t_0 is not necessarily to be set to the first year of the dataset because the query years that are too old could help little to reveal the authority at present. Therefore, in our experiments, we set $t_0 = t_{\text{cur}} - p - 10$.

As for relevance in LTR, for each query time t, we regard the logarithm of the future citation count each paper would obtain in the next p years as the relevance for the papers in PS_t,

$$\boldsymbol{r}(\boldsymbol{t}, y) = \{(\log c_t^i)_{i=1}^{n_t}\}_{t=t_0}^{t_{\text{cur}}-p} \quad (5)$$

where c_t^i is the number of citations paper i would obtain within the next p years after year t.

3.3 Zero Citation Ranking

The setting for ZCR is similar to that of FCR except that for each query time t, we only consider the papers published in year t, thus constructing a training set

$$\boldsymbol{X}_{\text{ZCR}} = \{(\boldsymbol{x}_t^i)_{i=1}^{n_{\text{Pub-In-t}}}\}_{t=t_0}^{t_{\text{cur}}-p}, \quad (6)$$

where $\boldsymbol{x}_t^i$ is the feature vector for paper i published in time t, $n_{\text{Pub-In-t}}$ is the number of papers published in year t, and the relevances are

$$\boldsymbol{r}(\boldsymbol{t}, y) = \{(\log c_t^i)_{i=1}^{n_{\text{Pub-In-t}}}\}_{t=t_0}^{t_{\text{cur}}-p} \quad (7)$$

where c_t^i is number of future citations in the next 3 years for paper i which is published in year t. For testing, unlike FCR to rank all the papers published, ZCR only considers the papers published in year t_{cur}.

3.4 Features

All the features below are processed by the method mentioned in the previous section so that they are time aware and consider the dynamic nature of the academic network.

3.4.1 Network Based

One set of features we use is network-related according to [28]. They claimed that high, low and medium impact papers position their citations differently. More specifically, they found that medium impact papers create citations in a narrow field while high and low impact papers cite more diverse papers. They also found that high impact ones differ from the low in that they can find bridges and connections between scientific communities.

Intuitively, the impact of one paper is related to its popularity and may decide its future citations. Therefore, we follow their discoveries for future citation ranking. There are features such as *Graph Density*, *Clustering coefficient*, *Connectivity*, and *Maximum betweenness*. Note that all the features of a specific paper are extracted from the subgraph consisting of the references of that paper rather than the whole citation network. Readers could refer to [28] for more details.

3.4.2 Author Based

Author is apparently a key factor in determining the future citations of one paper. We refer to existing methods and leverage multiple author based features.

The first category of author based features we leverage is simply designed features. They are author's total publication number, average publication number per year, h-index, total citation number, average citation number per year, and rank of the aforementioned features. The number of co-authors is also considered because authors with a large number of co-authors tend to be popular and their writings are likely to acquire plenty of citations. Similar features of affiliations to which the authors affiliate to are extracted as well.

Besides, we consider author's authority defined in [9]. It's based on the whole citation network and similar to PageRank. Note that the citation network is a homogeneous graph, which makes this feature different from the extracted features in methods [4] and [5], where they leverage heterogeneous graph. The linkages in the citation network $G(V, E)$ are weighted by standard cosine similarity between two corresponding papers. The transition probability between two papers is then defined by normalizing the corresponding affinity weight. Based on the transition matrix $M = M_{i,j}$ constructed by the transition probability, the authority of a paper p_i is defined as

$$\text{Authority}(p_i) = \alpha \sum_{j \neq i} \text{Authority}(j) \cdot M_{j,i} + \frac{1-\alpha}{|V|} \quad (8)$$

where $\alpha = 0.85$. The authority of an author a is defined as:

$$\text{Authority}(a) = \sum_{p \in P_a} \text{Authority}(p) \quad (9)$$

where P_a is the set of papers that author a has published.

3.4.3 Venue Based

The venue in which a paper published is also a nonnegligible aspect that may determine the paper's future popularity. For venue based features, we utilize a venue's average publication number, average citation number since they reflex the size and prestige of a venue to some extent.

In addition, we include as a feature the average citation number in recent k(=3) years to measure a venue's current quality and popularity.

Furthermore, We use a feature named *Venue Centrality* introduced by [9]. A graph of venues is first established where edges denote the citing-cited relationships between venues. The weight of each edge is estimated by the number of citations between two venues. Then a similar algorithm as in calculating author's authority is performed to estimate venue's centrality.

3.4.4 Paper Based

Another aspect we consider that may decide a paper's future popularity is from the paper itself. First, we calculate the entropy of topic distribution of a paper d. The topic distribution is estimated by running LDA [29] using the content of papers:

$$\text{entropy(d)} = \sum_{i=1}^{\text{K=100}} -p(\text{topic}_i|d) \times \text{logp}(\text{topic}_i|\text{d}) \quad (10)$$

It is based on an intuition that if the diversity of topics a paper covers is huge, it should own a vast range of audience and hence acquire more citations than the papers which narrow to a single topic.

Another feature we consider is also extracted from paper's content. We compare the content similarity between the papers published in the query year and the years before. This is based on the assumption that to some extent papers published in the current year reflex the popularity of topics. If a paper is similar to the papers published in the current year, it's likely to draw attention and acquire citations. For the measurement of content similarity, we follow the method used in [10].

Temporal information is also important for determining the popularity of scientific literatures. We hence add another feature named recency, which is the number of years since a paper published.

4. EXPERIMENTS

4.1 Dataset

We use the processed dataset *ACL Anthology Network (AAN)* provided by [5]. The 2011 release of AAN consists of 18,401 papers, 14,386 authors, and 273 venues. Compared to other datasets, AAN is preferred for several reasons. First, it contains full names of authors, which is useful for author disambiguity. Second, it has affiliation information of authors, which is not included in many other public datasets. Third, there is full text information while other datasets usually contain only abstract of papers. Last but not the least, the papers in AAN are from the same domain, which is suitable for ranking since the citation count of publications in different fields may vary heavily and the weight of each feature in LTR may also vary significantly in different fields. Note that the AAN dataset is relatively small, but FPRank is scalable and feasible for larger dataset.

4.2 Evaluation Metrics

In order to evaluate the performance of ranking algorithms, we introduce the following three common evaluation metrics.

4.2.1 Normalized Discounted Cumulative Gain

Normalized Discounted Cumulative Gain (NDCG) [30] measures the performance based on the ground truth and gives top results high weights. The NDCG score is computed as

$$\text{NDCG@k} = Z_k \sum_{i=1}^{k} \frac{2^{r_i} - 1}{log_2(i+1)} \quad (11)$$

where r_i is the relevance level of the ith paper returned by the algorithm, and we let $r_i = \log(c_i + 1)$ where c_i is the citation number of this paper. Z_k is the normalization constant to ensure that an ideal rank will get score 1.

4.2.2 Graded Average Precision

Graded Average Precision (GAP) [5] directly generalizes average precision to the multigraded relevance case and it also considers recall by estimating the area under the non-interpolated grade precision-recall curve. *GAP@k* is defined as

$$GAP@k = \frac{\sum_{d=1}^{k} \sum_{m=1}^{d} \delta(m,d)/d}{\sum_{i=1}^{c} R_i \sum_{j=1}^{i} p_j} \quad (12)$$

$$\delta(m,d) = \begin{cases} \sum_{j=1}^{min(r_d,r_m)} p_j & r_m \geq 0 \\ 0 \quad otherwise \end{cases} \quad (13)$$

where R_i is the number of relevant paper at grade i, p_j is the probability that the user sets the threshold at grade i and $r_d = i$ if c_d, the citation number of P_d, satisfies $i \leq$ log $c_d \leq i + 1$.

4.2.3 Match Ratio

We also use a straight-forward metric *Match Ratio (MR)* to measure a model's capability on predicting top-tier papers. Let P be the set of top-k returned papers and P_0 the set of top-k papers in the ground truth. *MR@k* is defined as

$$MR@k = \#(P \cap P_0)/k \quad (14)$$

The higher $MR@k$ is, the better the model is on predicting the top-k highly cited papers. In this metric, the order of papers is ignored. We only consider whether a highly cited papers in top-k is hit or not.

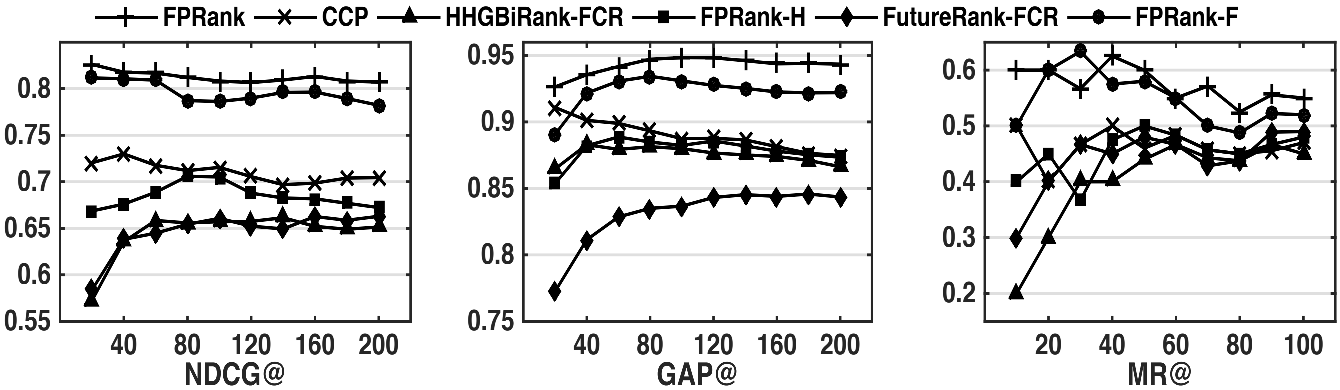

Figure 2: Evaluation of six settings on FCR.

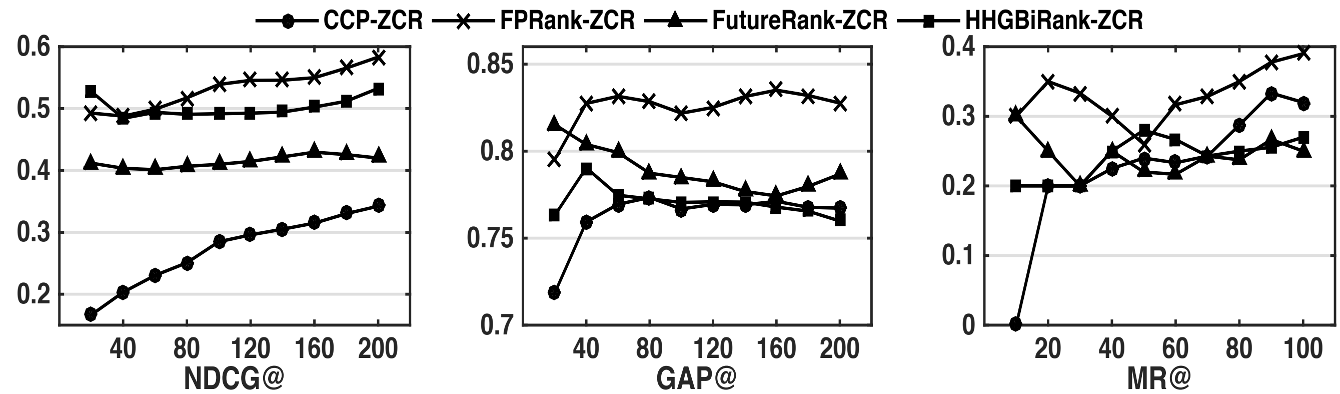

Figure 3: Evaluation of four settings on ZCR.

4.3 Experiment Setup

We use the AAN dataset described above. The goal is to rank papers according to their future citation counts during the next p years, in which we set $p = 3$. Analysis shows that the AAN dataset is complete before 2011. Therefore, we set the the test year t_{cur} as 2007, and the first query year $t_0 = 1995$. We also take year 2004 as validation year to select parameters. We compare with state-of-the-art graph-based ranking models FutureRank [4], HHGBiRank [5], and regression-based method CCP [9]. For FutureRank and HHGBiRank, we grid search the best settings of parameters using validation set. The best parameters for FutureRank are $\alpha = 0.4$, $\beta = 0.1$, $\gamma = 0.4$, and $\alpha_{11} = \alpha_{22} = \alpha_{31} = \alpha_{33} = 0.1$, $\alpha_{21} = \alpha_{32} = 0.8$, $\alpha_{12} = 0.4$, $\alpha_{13} = 0.5$ for HHGBiRank. For CCP, we extract delicately designed features as mentioned above and use classification and regression trees (CART) as the regression model. We experiment on 10 different settings:

- **FPRank**: All the designed features described above and the features extracted from FutureRank and HHGBiRank are used. LTR is performed.
- **CCP**: All the designed features described above are used to do regression. The results are then sorted to acquire a ranking.
- **FutureRank-FCR**: FutureRank is used to deal with the FCR problem.
- **FPRank-F**: Only the features obtained from FutureRank is used in FPRank.
- **HHGBiRank-FCR**: HHGBiRank is used to deal with the FCR problem.
- **FPRank-H**: Only the features obtained from HHGBiRank is used in FPRank.
- **FPRank-ZCR**: FPRank is used to solve the ZCR problem.
- **CCP-ZCR**: CCP is used to solve the ZCR problem.
- **FutureRank-ZCR**: FutureRank is used to tackle ZCR problem.
- **HHGBiRank-ZCR**: HHGBiRank is used to tackle ZCR problem.

4.4 Ranking is Reasonable

In this subsection, we show the predictions of CCP methods are neither robust nor accurate while those of FPRank are reasonable.

Table 1 shows the results in the setting of FPRank and CCP. We select top 10 papers that obtain the highest number of citations within three years after 2007 in the AAN dataset. The first column shows the true citation count of each paper; The second column shows the predicted citation count by CCP; The Third and Fourth columns present the rank of each paper which is predicted by CCP and FPRank respectively.

From the table, we can find that only one paper's citation count is well estimated(error within 10) in CCP and for most papers the predicted citation counts are far from the ground truth. We hence argue that this kind of estimation is unreliable in a sense that we can hardly believe in the citation counts that the existing models predicted and only the order of papers is informative. However, in terms of the relative order between papers, namely ranking, FPRank is much better and more robust since it focuses directly on ranking rather than citation count prediction.

Table 1: Predictions of CCP and FPRank on the top 10 papers in terms of future citation count in the next 3 years

True Citation	Predicted Citation(CCP)	Predicted Rank(CCP)	Predicted Rank(FPRank)
268	70	25	7
248	212	2	4
235	6	788	295
232	212	1	2
224	198	4	3
196	198	3	1
164	16	218	5
135	184	5	30
111	4	1243	19
107	52	39	14

4.5 FPRank Outperforms Existing Methods in both FCR and ZCR Problems

We first ran LTR using the features extracted from graph-based methods and compare with their original performance, namely FutureRank-FCR vs FPRank-F, and HHGBiRank-FCR vs FPRank-H. We then tested CCP and FPRank respectively using all the features we've extracted, namely FPRank vs CCP. Besides, we did experiment on ZCR problem, in which FutureRank-ZCR, HHGBiRank-ZCR, FPRank-ZCR, and CCP-ZCR were used.

Figure 2 shows the experiment results of these settings under the three performance measures described above. One can notice that FPRank-F significantly improves the performance of FutureRank-FCR by 0.142 in NDCG, 0.090 in GAP, and 0.109 in MR on average. FPRank-H on average improves the performance of HHGBiRank-FCR. by 0.033, 0.008, and 0.006 in NDCG, MR, and GAP respectively. Since we only use the features extracted from the two graph-based ranking models respectively, the improvements show that there are constraints underlying the heterogeneous network and FPRank can help break these constraints. The differences in the quantity of improvements may lie in the different factors they considered – FutureRank considers author and time information while HHGBiRank only considers author information. The more factors a model consider, the more constraints it will contain.

We can also find that FPRank outperforms all the other methods. For CCP, FPRank is superior to it by 0.103, 0.055, and 0.135 under three metrics.This suggests again that FPRank is much more competent than the CCP methods in terms of ranking and it is expected because of their different goals. For FPRank-F, FPRank further improves its performance by 0.019, 0.018, and 0.024. For FPRank-H, the results are and 0.128, 0.062, and 0.152. The performance gains over FPRank-F and FPRank-H show that the information incorporated into the two graph-based methods is not fully utilized and FPRank can be used to get better results through considering more features.

For ZCR, FPRank-ZCR also shows on the top of these three performance measures and beats both graph-based ranking models (FutureRank-ZCR, HHGBiRank-ZCR) and CCP method (CCP-ZCR). More specifically, FPRank on average improves FutureRank-ZCR by 0.122, 0.041, and 0.141, HHGBiRank-ZCR by 0.034, 0.058, and 0.107, and CCP by 0.251, 0.064, 0.086, respectively. Though the ranking of these algorithms may vary depending on the performance measure, the robustness of FPRank is proved. The performance degradation for ZCR from FCR results from the less information we have for newly published papers.

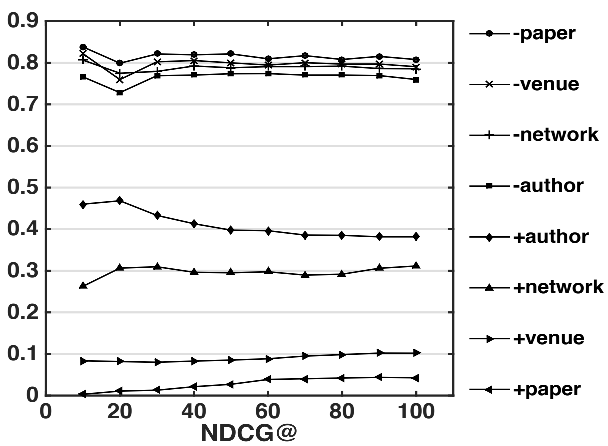

Figure 4: NDCG result for FPRank with the feature used in isolation ("+") and dropped from the model ("-")

4.6 What Features are The Most Predictive

In this experiment, we analyze which category of features is the most predictive in the field of natural language processing. The results of evaluation metric *NDCG* from 10 to 100 are shown in Figure 4. Similar results can be found when using *GAP* and *MR*. In general, author plays a dominant role. The impact of network based features introduced by [28] is also significant, which proves their discoveries about how high and low impact papers cite differently. In contrast, venue on which a paper is published and the content of a paper contribute relatively less to its future citation.

4.7 Demo System

Acemap[1] [31] is a novel academic system that aims to provide a convenient method for researchers to discover and explore connections in the research community. Acemap contains homepages for each venue, affiliation, author, and paper. FPRank was incorporated into the venue page (e.g. WSDM[2]) to recommend potentially popular papers.

[1]http://acemap.sjtu.edu.cn/

[2]http://acemap.sjtu.edu.cn/conference?id=42C7B402

5. CONCLUSION

In this work, a new ranking framework FPRank is proposed to tackle two problems: (1) Future Citation Ranking problem which aims to predict the ranking of scientific publications by their predicted future popularity and (2) Zero Citation Ranking problem which is to rank only newly published literatures without citation information. FPRank considers the dynamic time-aware representations of papers and takes the future citation count as the ground truth future popularity. LTR technique is leveraged and time is taken as query in FPRank to train ranking models. Extensive experiments on features of graph-based ranking models and citation count prediction methods show that FPRank outperforms the state-of-the-art scientific ranking and citation prediction methods in dealing with both FCR and ZCR problems. In addition, feature analysis is conducted and shows that author plays the dominant role in determining the future popularity of a paper in the field of natural language processing.

6. REFERENCES

[1] Dylan Walker, Huafeng Xie, Koon-Kiu Yan, and Sergei Maslov. Ranking scientific publications using a model of network traffic. *CoRR*, abs/physics/0612122, 2006.

[2] Yujing Wang, Yunhai Tong, and Ming Zeng. Ranking scientific articles by exploiting citations, authors, journals, and time information. In *Proc. AAAI*, 2013.

[3] Senzhang Wang, Sihong Xie, Xiaoming Zhang, Zhoujun Li, S Yu Philip, and Xinyu Shu. Future influence ranking of scientific literature. In *Proc. SDM*. SIAM, 2014.

[4] Hassan Sayyadi and Lise Getoor. Futurerank: Ranking scientific articles by predicting their future pagerank. In *Proc. SDM*. SIAM, 2009.

[5] Ronghua Liang and Xiaorui Jiang. Scientific ranking over heterogeneous academic hypernetwork. In *Proc. AAAI*, 2016.

[6] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[7] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 1999.

[8] Lawrence D Fu and Constantin Aliferis. Models for predicting and explaining citation count of biomedical articles. In *Annual Symposium Proceedings*, 2008.

[9] Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. Citation count prediction: learning to estimate future citations for literature. In *Proc. CIKM*. ACM, 2011.

[10] Avishay Livne, Eytan Adar, Jaime Teevan, and Susan Dumais. Predicting citation counts using text and graph mining. In *the iConference Workshop on Computational Scientometrics: Theory and Applications*, 2013.

[11] Nataliia Pobiedina and Ryutaro Ichise. Citation count prediction as a link prediction problem. *Applied Intelligence*, 2016.

[12] Johan Bollen, Marko A Rodriquez, and Herbert Van de Sompel. Journal status. *Scientometrics*, 2006.

[13] Peng Chen, Huafeng Xie, Sergei Maslov, and Sidney Redner. Finding scientific gems with google pagerank algorithm. *Journal of Informetrics*, 2007.

[14] Nan Ma, Jiancheng Guan, and Yi Zhao. Bringing pagerank to the citation analysis. *Information Processing and Management*, 2008.

[15] Ding Zhou, Sergey A Orshanskiy, Hongyuan Zha, and C Lee Giles. Co-ranking authors and documents in a heterogeneous network. In *Proc. ICDM*. IEEE, 2007.

[16] Erjia Yan, Ying Ding, and Cassidy R Sugimoto. P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, 62(3):467–477, 2011.

[17] Michael Callaham, Robert L Wears, and Ellen Weber. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Jama*, 2002.

[18] Carlos Castillo, Debora Donato, and Aristides Gionis. Estimating number of citations using author reputation. In *String processing and information retrieval*. Springer, 2007.

[19] Abhaya V Kulkarni, Jason W Busse, and Iffat Shams. Characteristics associated with citation rate of the medical literature. *PloS one*, 2007.

[20] Fereshteh Didegah and Mike Thelwall. Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, 2013.

[21] Dashun Wang, Chaoming Song, and Albert-László Barabási. Quantifying long-term scientific impact. *Science*, 2013.

[22] Hua-Wei Shen, Dashun Wang, Chaoming Song, and Albert-László Barabási. Modeling and predicting popularity dynamics via reinforced poisson processes. In *Proc. AAAI*, 2014.

[23] Shuai Xiao, Junchi Yan, Changsheng Li, Bo Jin, Xiangfeng Wang, Xiaokang Yang, Stephen M. Chu, and Hongyuan Zha. On modeling and predicting individual paper citation count over time. In *Proc. IJCAI*, 2016.

[24] Tie-Yan Liu. Learning to rank for information retrieval. 2011.

[25] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proc. SIGKDD*. ACM, 2002.

[26] Christopher JC Burges. From ranknet to lambdarank to lambdamart: An overview. *Learning*, 2010.

[27] Zaihan Yang, Liangjie Hong, Dawei Yin, and Brian D Davison. Ranking authors with learning-to-rank topic modeling. *International Journal of Innovative Computing, Information and Control*, 2015.

[28] Xiaolin Shi, Jure Leskovec, and Daniel A McFarland. Citing for high impact. In *Proc. JCDL*. ACM, 2010.

[29] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 2003.

[30] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 2002.

[31] Zhaowei Tan, Changfeng Liu, Yuning Mao, Yunqi Guo, Jiaming Shen, and Xinbing Wang. Acemap: A novel approach towards displaying relationship among academic literatures. In *Proc. WWW, Companion Volume*, 2016.