Check for updates

The MIT Press

RESEARCH ARTICLE

# Towards establishing a research lineage via identification of significant citations

**Tirthankar Ghosal[1]**, **Piyush Tiwary[2]**, **Robert Patton[3]**, and **Christopher Stahl[3]**

[1]Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic
[2]Indian Institute of Science, India
[3]Oak Ridge National Laboratories, US
**equal contribution.

## ABSTRACT

Finding the lineage of a research topic is crucial for understanding the prior state of the art and advancing scientific displacement. The deluge of scholarly articles makes it difficult to locate the most relevant previous work. It causes researchers to spend a considerable amount of time building up their literature list. Citations play a crucial role in discovering relevant literature. However, not all citations are created equal. The majority of the citations that a paper receives provide contextual and background information to the citing papers. In those cases, the cited paper is not central to the theme of citing papers. However, some papers build upon a given paper and further the research frontier. In those cases, the concerned cited paper plays a pivotal role in the citing paper. Hence, the nature of the citation that the former receives from the latter is *significant*. In this work, we discuss our investigations towards discovering *significant citations* of a given paper. We further show how we can leverage significant citations to build a research lineage via a *significant citation graph*. We demonstrate the efficacy of our idea with two real-life case studies. Our experiments yield promising results with respect to the current state of the art in classifying *significant citations*, outperforming the earlier ones by a relative margin of 20 points in terms of precision. We hypothesize that such an automated system can facilitate relevant literature discovery and help identify knowledge flow for a particular category of papers.

## 1. INTRODUCTION

Literature searches are crucial to discover relevant publications. The knowledge discovery that ensues forms the basis of understanding a research problem, finding the previously explored frontiers and identifying research gaps, which eventually leads to the development of new ideas. However, with the exponential growth of scientific literature (including published papers and preprints) (Ghosal, Sonam et al., 2019b), it is almost impossible for a researcher to go through the entire body of the scholarly works, even in a very narrow domain. Citations play an important role here in finding the relevant articles that further topical knowledge. However, not all citations are equally effective (Zhu, Turney et al., 2015) effective in finding relevant research. A majority of papers cite a work contextually (Pride & Knoth, 2017a) for providing additional background context. Such background contextual citations help in the broader understanding; however, they are not central to the citing paper's theme. Some papers use the ideas in a given paper, build upon those ideas, and displace the body of relevant

research. Such papers are expected to acknowledge the prior work (via citing them) duly. However, the nature of citation, in this case, is different from that of contextual citations. These citations, which heavily rely on a given work or build upon that work, are *significant citations*. However, the current citation count metric puts equal weights on all the citations. Therefore, it is inadequate for identifying the papers that have significantly cited a given work and may have taken the relevant research forward. Identifying such *significant* citations is hence crucial to the literature study.

It is not uncommon that authors sometimes fail to acknowledge relevant papers' role in influencing their ideas (Rousseau, 2007; Van Noorden, 2017). As a result, researchers spend a lot of their time searching for the papers most relevant to their research topic, thereby locating the subsequent papers that carried forward a given scientific idea. It is usually desirable for a researcher to understand the story behind a prior work and trace the concept's emergence and gradual evolution through publications, thereby identifying the knowledge flow. Researchers ideally curate their literature base by identifying significant references to a given paper and then hierarchically locating meaningful prior work.

The idea of recognizing *significant citations* is also important to understand the true impact of given research or facility. To understand how pervasive particular research was in the community, it is essential to understand its influence beyond the direct citations it received. To this end, tracking the *transitive* influence of research via identifying significant citations could be one possible solution.

In this work, we develop automatic approaches to trace the lineage of given research via transitively identifying the significant citations to a given article. The overall objective of our work is twofold:

- Accelerate relevant literature discovery via establishing a research lineage.
- Find the true influence of a given work and its pervasiveness in the community beyond citation counts.

There are two aspects to the problem: *identifying the relevant prior work* and *identifying the follow-up works that stemmed from or are influenced by the current work*. The first aspect would facilitate relevant prior literature discovery for a paper. In contrast, the second aspect would facilitate discovering the knowledge flow in subsequent relevant papers. Obviously, our approach would not be a *one size fits all* approach. Still, we believe it is effective to find investigations that build upon relevant priors and facilitate relevant literature discovery, and thereby steer towards identifying the pervasiveness of a given piece of research in the community. We base our work on classifying citations as *contextual* or *significant* and trace the lineage of research in a citation graph via identifying significant edges. The major contributions of the current work are the following:

- We use a set of novel and rich features to classify citations as significant or contextual.
- A graph-based approach to tracing the lineage of a given research work leveraging on citation classification.

## 2. RESEARCH LINEAGE

The mechanism of citations in academia is not always transparent (Van Noorden & Singh Chawla, 2019; Vîiu, 2016; West, Stenius, & Kettunen, 2017). Problems such as coercive

citations (Wilhite & Fong, 2012), anomalous citations (Bai, Xia et al., 2016), citation manipulation (Bartneck & Kokkelmans, 2011), *rich get richer effects* (Ronda-Pupo & Pham, 2018), and discriminatory citation practices (Camacho-Miñano & Núñez-Nickel, 2009) have infested the academic community. However, in spite of all these known issues, citation counts and *h*-indices still remain the measures of research impact and tools for academic incentives, though long-debated by many (Cerdá, Nieto, & Campos, 2009; Laloë & Mosseri, 2009). Usually, we measure the impact of a given paper by the direct citations it receives. However, a given piece of research may have induced a transitive effect on other papers, which is not apparent with the current citation count measures. Figure 1 shows a sample citation network where A could be a paper or a research facility. We want to know how pervasive was the research or facility A in the community. At $d = 1$ are the direct citations to A. We see that article B cite A significantly, or B is inspired by A. Other citations to A are background. At citation depth $d = 2$, we see that article C and article D significantly cite B (direct citation). We see that C also cites A significantly. Finally, at citation depth $d = 3$, E significantly cites C. We intend to understand if there is a lineage of research from A to E (A → B → C → E). Although E does not cite A directly, can we identify A's influence on E? If E is a seminal work receiving hundreds of citations, can we infer that A was the prior work that indirectly inspired E? We are interested in discovering such hidden inspirations to honestly assess the contributions of a research article or a facility.

### 3. RELATED WORK

Measuring academic influence has become a research topic because publications are associated with academic prestige and incentives. Several metrics (impact factor, eigen factor, *h*-index, citation counts, altmetrics, etc.) have been devised to comprehend research impact efficiently. Still, each one is motivated by a different aspect and has found varied importance across disciplines. Zhu et al. (2015) did pioneering work on academic influence prediction leveraging citation context. Shi, Wang et al. (2019) presented a visual analysis of citation context-based article influence ranking. Xie, Sun, and Shen (2016) predicted paper influence in an academic network by taking into account the content and venue of a paper, as well as the reputation of its authors. Shen, Song et al. (2016) used topic modeling to measure academic influence in scientific literature. Manju, Kavitha, and Geetha (2017) identified influential researchers in an academic network using a rough-set based selection of time-weighted
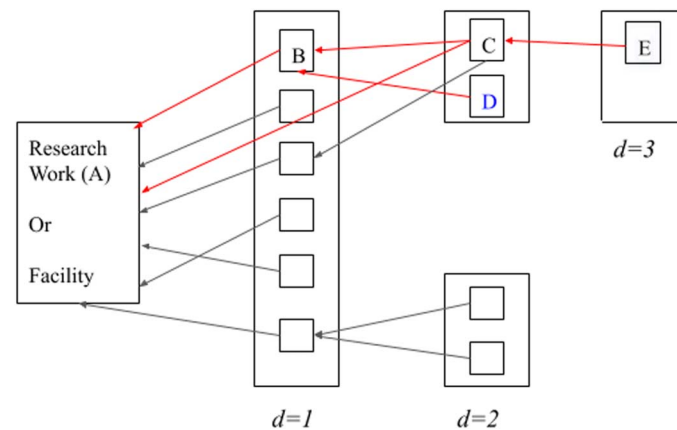


**Figure 1.** Research lineage.

academic and social network features. Pileggi (2018) did a citation network analysis to measure academic influence. Zhang and Wu (2020) used a dynamic academic network to predict the future influence of papers. Ji, Tang, and Chen (2019) analyzed the impact of academic papers based on improved PageRank. F. Wang, Jia et al. (2019) assessed the academic influence of scientific literature via altmetrics. F. Zhao, Zhang et al. (2019) measured academic influence using heterogeneous author-citation networks. Recently, many deep learning–based methods have been explored for citation classification. Perier-Camby, Bertin et al. (2019) attempt to compare deep learning-based methods with rule-based methods. They use deep learning–based feature extractors such as BCN (McCann, Bradbury et al., 2017) and ELMo (Peters, Neumann et al., 2018) to extract semantic information and feed it to various classifiers for classification. They conclude that neural networks could be a potential dimension for citation classification when a large number of samples are available. However, for a small data set such as the one we use, rule-based methods clearly hold an advantage. Apart from this, the features used in rule-based methods are more comprehensible than features extracted from deep learning methods, thus providing deeper insights into analyzing factors that make a citation significant or contextual.

The closest literature for our task is that on citation classification. Citation classification has been explored in the works of Alvarez, Soriano, and Martínez-Barco (2017), Dong and Schäfer (2011), Qayyum and Afzal (2019), and Teufel, Siddharthan, and Tidhar (2006). These works use features from the perspective of citation motivation. On the other hand, there are works that emphasize on features from a semantic perspective. Wang, Zhang et al. (2020) use syntactic and contextual information of citations for classification. Aljuaid, Iftikhar et al. (2021) and Amjad and Ihsan (2020) perform classification based on sentiment analysis of in-text citations. Athar (2011) and Ihsan, Imran et al. (2019) propose sentiment analysis of citations using linguistic studies of the citance. More recently, several open-source data sets for citation classification have been developed in the work of Cohan, Ammar et al. (2019) and Pride and Knoth (2020). Valenzuela, Ha, and Etzioni (2015) explored citation classification into *influential* and *incidental* using machine learning techniques which we adapt as *significant* and *contextual* respectively in this work.

In this work, we propose a rich set of features informed from both citation and context (semantics) perspectives, leveraging advantages of both types, thus performing better than all of the methods mentioned above. However, our problem is motivated beyond citation classification. We restrict our classification labels to *significant* and *contextual*, unlike Valenzuela et al. (2015), as these labels are enough to trace the lineage of a work. Furthermore, to the best of our knowledge, we did not find any work leveraging citation classification for finding a research lineage. Hence, we only compare our performance for the citation significance detection subtask with other approaches.

## 4. DATA SET DESCRIPTION

We experiment with the Valenzuela data set (Valenzuela et al., 2015) for our task. The data set consists of incidental/influential human judgments on 630 citing-cited paper pairs for articles drawn from the 2013 ACL anthology, the full texts of which are publicly available. Two expert human annotators determined the judgment for each citation, and each citation was assigned a label. Using the author's binary classification, 396 citation pairs were ranked as incidental citations, and 69 (14.3%) were ranked as influential (important) citations. For demonstrating our research lineage idea, we explore knowledge flow on certain papers of *Document-Level Novelty Detection* (Ghosal, Salam et al., 2018b) and the High Performance Computing (HPC)

algorithm *MENNDL* (Young, Rose et al., 2015). The actual authors of these two topics helped us with manual annotation of their paper's lineage.

## 5. METHODOLOGY

To identify significant citations, we pursue a feature-engineering approach to curate several features from cited-citing paper pairs. The objective is to classify the citations received by a given paper into SIGNIFICANT and CONTEXTUAL. The original cited citing papers in the Valenzuela data set are in PDF. We convert the PDFs to corresponding XMLs using GROBID (Lopez, 2009). We use GROBID to parse our PDFs into XMLs as well as manually correcting a few inconsistent files so that there is no discrepancy.

1. **Citation frequency inside the body of citing paper (F1)**: We measure the number of times the cited paper is referenced from within the citing paper's body. The intuition is that if a paper is cited multiple times, the cited paper may be significant to the citing paper.

2. **Are the authors of citing and cited paper the same? (Boolean) (F2)**: We check if the authors of the citing and cited papers are the same. This might be a case of self-citation or can also signal the extension of the work.

3. **Author overlap ratio (F3)**: This measures number of common authors in the citing and cited papers normalized to the total number of authors in the citing paper. The intuition is similar to F2.

4. **Is the citation occurring in a table or figure caption? (Boolean) (F4)**: The intuition is that most of the citations in tables and figures appear for comparison/significantly referencing existing work. Hence, the citing paper might be an extension of the cited article or may have compared it with earlier significant work.

5. **Is the citation occurring in groups? (Boolean) (F5)**: We check if the citation is occurring along with other citations in a group. The intuition is that such citations generally appear in related works to highlight a background detail; hence, they might not be a significant citation.

6. **Number of citations to the cited paper normalized by the total number of citations made by the citing paper (F6)**: This measures the number of citations to the cited paper by the citing paper normalized by the total number of citation instances in the citing paper. This measures how frequently the cited paper is mentioned compared to other cited papers in the citing paper.

7. **Number of citations to the cited paper normalized by the total number of bibliography items in the citing paper (F7)**: This measures the number of citations to the cited paper normalized to the total number of bibliography items in the citing paper. The intuition is similar to F6.

8. ***tf-idf* similarity between abstracts of the cited and citing paper (F8)**: We take cosine similarity between the *tf-idf* representations of the abstracts of cited and citing papers. The intuition is that if the similarity is higher, the citing paper may be inspired/extended from the cited paper.

9. ***tf-idf* similarity between titles of the cited and citing paper (F9)**: We take cosine similarity between the *tf-idf* representations of the titles of cited and citing papers.

10. **Average *tf-idf* similarity between citance and abstract of the cited paper (F10)**: We calculate the similarity of each citance with the abstract of the cited article and take the average of it. Citances are sentences containing the citations in the citing paper. Citances reveal the purpose of the cited paper in the citing paper. Abstracts contain the

contribution/purpose statements of a given paper. Hence similarity with citances may suggest that the cited paper has been used significantly in the current paper.

11. **Maximum *tf-idf* similarity between citance and abstract of the cited paper (F11)**: We take the maximum of similarity of the citances (there could be multiple citation instances of the same paper in a given paper) with the abstract of the cited paper.

12. **Average *tf-idf* similarity between citance and title of the cited paper (F12)**: We calculate the similarity of each citance with the title of the cited paper and take an average of it.

13. **Maximum *tf-idf* similarity between citance and title of the cited paper (F13)**: We take the maximum of similarity of the citances with the title of the cited paper.

14. **Average length of the citance (F14)**: Average length of the citances (in words) for multiple citances. The intuition is that if the citing paper has spent many words on the cited article, it may have significantly cited the corresponding article.

15. **Maximum length of the citance (F15)**: Maximum length of the citances (in words).

16. **No. of words between citances (F16)**: We take the average of the number of words between each pair of consecutive citances of the cited paper. This is set to 0 in the case of a single citance.

17. **In how many different sections does the citation appear in the citing paper? (F17)**: We take the number of different sections in which the citation to a cited paper occurs and normalize it with the total number of sections present in the citing paper. The intuition is that if a citation occurs in most sections, it might be a significant citation.

18. **Number of common references in citing and cited paper normalized by the total number of references in citing article (F18)**: We count the number of common bibliographic items present in the citing and cited papers and normalize it with total bibliographic items present in the citing paper.

19. **Number of common keywords between abstracts of the cited and citing paper extracted by YAKE (Campos, Mangaravite et al., 2018) (F19)**: We compare the number of common keywords between the abstracts of the citing and cited papers extracted using YAKE. Our instinct is that a greater number of common keywords would denote a greater similarity between abstracts.

20. **Number of common keywords between titles of the cited and citing paper extracted by YAKE (F20)**: We compare the number of common keywords between the titles of the citing and cited papers extracted using YAKE.

21. **Number of common keywords between the body of the cited and citing papers extracted by YAKE (F21)**: We compare the number of common keywords between the body of the citing and cited papers extracted using YAKE.

22. **Word Mover's Distance (WMD) (Huang, Guo et al., 2016) between the abstracts of the cited and citing papers (F22)**: We measure the WMD between the abstracts of the citing and cited papers. The essence of this feature is to calculate semantic distance/similarity between abstracts of the two papers.

23. **WMD between titles of the cited and citing papers (F23)**: We measure the WMD between the titles of the citing and cited papers.

24. **WMD between the bodies of the cited and citing papers (F24)**: We measure the WMD between the bodies of the citing and cited papers.

25. **Average WMD between citance and abstract of the cited and citing papers (F25)**: We take the average of WMDs between the citance and abstract of the cited paper.

26. **Maximum WMD between citance and abstract of the cited and citing papers (F26)**: We take the maximum of WMDs between the citance and abstract of the cited paper.

27. **Average VADER (Gilbert & Hutto, 2014) polarity index—Positive (F27), Negative (F28), Neutral (F29), Compound (F30)**: We measure the VADER polarity index of

all the citances of the cited paper, and take their average for each sentiment (positive, negative, neutral, and compound).

28. **Maximum VADER polarity index—Positive (F31), Negative (F32), Neutral (F33), Compound (F34) of citances**: We measure the VADER polarity index of all the citances of the cited paper, and take the maximum among them for each sentiment (positive, negative, neutral, and compound). The intuition to use sentiment information is to understand how the citing paper cites the cited paper.

29. **Number of common venues in the bibliographies of the citing and cited papers (F35)**: We count the number of common venues mentioned in the bibliographies of the citing and cited papers and normalize it with the number of unique venues in the citing paper. Higher venue overlap would signify that the papers are in the same domain (Ghosal et al., 2019b).

30. **Number of common authors in the bibliographies of the citing and cited papers (F36)**: We count the number of common authors mentioned in the bibliographies of the citing and cited papers and normalize it with the number of unique authors in the citing paper (Ghosal et al., 2019b).

As mentioned earlier, only 14.3% of total citations are labeled as significant, which poses a *class imbalance* problem. To address this issue, we use SMOTE (Chawla, Bowyer et al., 2002) along with random undersampling of the majority (contextual citation) class. We first split the data set into 60% training and 40% testing data. Then we undersample the majority class by 50%, and oversample the minority class by 40% on the training partition of the data set.

## 6. EVALUATION

Our evaluation consists of two stages: First, we evaluate our approach on the citation significance task. Next, we try to see if we can identify the research lineage via tracing significant citations across the two research topics (*Document-level Novelty* and *MENNDL*). We ask the original authors to annotate the lineage and verify it with our automatic method. We train our model on the Valenzuela data set and use that trained model to predict significant citations of *Document-Level Novelty* and *MENNDL* papers, and thereby try to visualize the research lineage across the citing papers. We curate a small citation graph to demonstrate our idea. Note that our task in concern is Citation Significance Detection, which is different from Citation Classification in Literature. Whereas Citation Classification focuses on identifying the citation's intent, Citation Significance aims to identify the value associated with the citation. Obviously, the two tasks are related to each other, but the objectives are different.

### 6.1. Citation Significance Detection

The goal of this task is to identify whether a citation was SIGNIFICANT or CONTEXTUAL. We experiment with several classifiers for the binary classification task such as kNN ($k = 3$), Support Vector Machines (kernel = RBF), Decision Trees (max depth = 10) and Random Forest (n estimators = 15, max depth = 10). We found Random Forest to be the best performing one with our feature set. Table 1 shows our current results against the earlier reported results on the Valenzuela data set. We attain promising results compared to earlier approaches with a relative improvement of 20 points in precision. As the data set is small, neither earlier works nor we attempted a deep neural approach for citation classification on this dataset. Like us, Qayyum and Afzal (2019) also used Random Forest as the classifier; however, they relied on metadata features rather than content-based features for their work. Their experiments tried to answer the following questions: To what extent can the similarities and dissimilarities

**Table 1.** Results on Citation Significance Detection on the Valenzuela data set

| Methods | Precision |
|---|---|
| Valenzuela et al. (2015) | 0.65 |
| Qayyum and Afzal (2019) | 0.72 |
| Nazir, Asif, and Ahmad (2020a) | 0.75 |
| Nazir, Asif et al. (2020b) | 0.85 |
| **Current Approach** | **0.92** |

between metadata parameters serve as useful indicators for important citation tracking? And: Which metadata parameters or their combinations are helpful in achieving good results? We specifically work with full-text content-based features; hence our approach leverages richer information because it takes into consideration the full text of the works, whereas Qayyum and Afzal (2019) is solely based on metadata which helps us to achieve better performance.

Table 2 shows the classification results of the various classifiers we experimented with. Clearly, our features are highly interdependent (Section 5), which explains the better performance of Random Forests.

Figure 2 shows the importance of the top 10 features ranked by their *information gain*. However, our experimental data set is small and our features corelated; hence it seems that some features have marginal contributions. We deem that in a bigger real-life data set, the feature significance would be more visible. Here, we can see that features such as *distance between citances, the number of concerned citations normalized by the total number of citations, similarity between cited-citing abstracts, in-text citation frequency*, and *the similarity between citance and cited abstract* play an important role in the classification. The other features in the top 10 are *distance between citance, number of citations from citing to cited normalized by the total citations made by the citing paper, the similarity between cited-citing abstracts, in-text citation frequency, the average similarity between citance and cited abstract, number of citations from citing to cited normalized by the total references made by the citing paper, number of common YAKE keywords between the body of citing and cited paper, the average similarity between citance and title of cited paper, the max similarity between citance and abstract of cited paper*, and *neutral sentiment polarity of citance*. We explain the possible reasons behind the performance of these features in the subsequent sections. The precision using only the top 10 features is 0.73. Hence, other features play a significant role as well. A complete list of features and the corresponding information gain is given in Table 3.

**Table 2.** Classification result of various classifiers for Citation Significance

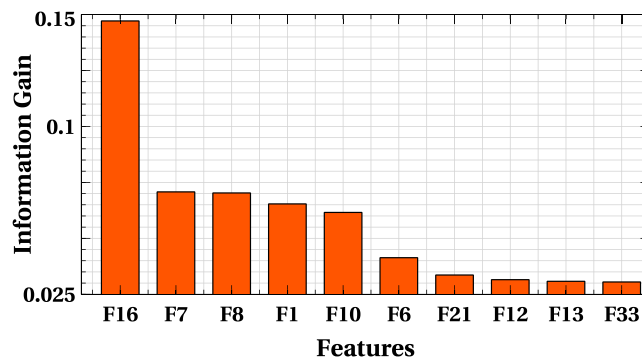| Methods | Precision | Recall | F1 score | Accuracy |
|---|---|---|---|---|
| kNN | 0.80 | 0.87 | 0.83 | 0.81 |
| SVM | 0.79 | 0.67 | 0.73 | 0.81 |
| Decision Tree | 0.80 | 0.82 | 0.81 | 0.86 |
| **Random Forest** | **0.92** | **0.82** | **0.87** | **0.90** |

**Figure 2.** Feature importance ranked via information gain.

To analyze the contribution of each feature, we evaluate our model using single feature at a time similar to Valenzuela et al. (2015). The precision after considering each feature individually is shown in Table 4. It is seen that the first 28 features in the table contribute significantly to the classification, and the overall precision after considering all the features is even better (an improvement of 14 points). F1, F4 (suggesting that significant citations do occur in tables or figures), and F21, followed by F7, F19, and F3 are the best performing features. This indicates that features obtained from a citation perspective are more useful. On the other hand, the worst performing features are F20 (perhaps due to the small size of the data set), F13, F5 (suggesting that significant citations also occur in groups), F17, F2, F22, and F33. Most of our observations are in line with Valenzuela et al. (2015).

Pride and Knoth (2017b) found *Number of Direct Citations*, *Author Overlap*, and *Abstract Similarity* to be the most important features. Our approach performs well enough to proceed with the next stage.

**Table 3.** Information gain (IG) due to each feature. Features are ranked in decreasing order of information gain

| Feature | IG | Feature | IG | Feature | IG |
|---|---|---|---|---|---|
| F16 | 0.147 | F24 | 0.024 | F30 | 0.015 |
| F7 | 0.070 | F13 | 0.022 | F27 | 0.015 |
| F8 | 0.070 | F33 | 0.021 | F31 | 0.015 |
| F1 | 0.065 | F18 | 0.020 | F32 | 0.014 |
| F10 | 0.061 | F3 | 0.020 | F15 | 0.014 |
| F6 | 0.041 | F23 | 0.019 | F9 | 0.013 |
| F21 | 0.033 | F35 | 0.019 | F17 | 0.011 |
| F12 | 0.031 | F34 | 0.017 | F36 | 0.011 |
| F13 | 0.030 | F19 | 0.017 | F4 | 0.006 |
| F33 | 0.030 | F22 | 0.016 | F5 | 0.006 |
| F35 | 0.025 | F28 | 0.016 | F2 | 0.004 |
| F26 | 0.024 | F25 | 0.016 | F20 | 0.003 |

**Table 4.** Performance of Random Forest model by using individual features at a time. The features are listed in decreasing order of the precision

| Feature | Precision | Feature | Precision | Feature | Precision |
|---|---|---|---|---|---|
| F1 | 0.78 | F15 | 0.28 | F25 | 0.15 |
| F4 | 0.76 | F8 | 0.27 | F11 | 0.14 |
| F21 | 0.71 | F10 | 0.27 | F24 | 0.13 |
| F7 | 0.68 | F9 | 0.25 | F31 | 0.11 |
| F19 | 0.61 | F27 | 0.23 | F26 | 0.10 |
| F3 | 0.50 | F23 | 0.20 | F33 | 0.08 |
| F16 | 0.47 | F36 | 0.20 | F22 | 0.07 |
| F28 | 0.43 | F35 | 0.20 | F2 | 0.04 |
| F35 | 0.37 | F12 | 0.19 | F17 | 0.04 |
| F6 | 0.33 | F34 | 0.19 | F5 | 0.03 |
| F32 | 0.33 | F30 | 0.17 | F13 | 0.03 |
| F33 | 0.29 | F18 | 0.15 | F20 | 0.01 |
| **Total** | | | | | **0.92** |

It is important to note that despite so many features, it is possible that some features might be correlated. Hence, we find the Pearson's correlation coefficient between each pair of features to see how dependent they are on each other. The heatmap of the correlation matrix is shown in Figure 3.
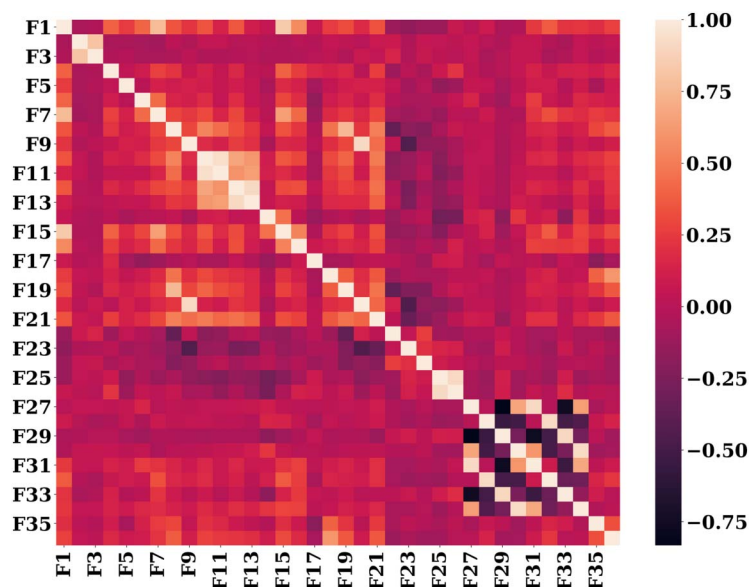


**Figure 3.** Heatmap of correlation between various pair of features.

**Table 5.**    Feature pairs with high correlation coefficient

| Feature pair | Correlation coefficient | Feature pair | Correlation coefficient |
|---|---|---|---|
| F10 & F11 | 0.937 | F25 & F26 | 0.910 |
| F30 & F34 | 0.919 | F9 & F20 | 0.907 |
| F28 & F32 | 0.917 | F29 & F33 | 0.905 |
| F27 & F31 | 0.914 | F1 & F15 | 0.835 |
| F12 & F13 | 0.910 | F27 & F29 | 0.832 |

We find that the average correlation coefficient between all the features is 0.074. However, there are few pairs of features that have high correlation coefficients. We have listed such pairs in Table 5.

From Table 5 we can see that feature pairs such as F10 & F11, F30 & F34, F28 & F32, F27 & F31, F12 & F13, F25 & F26, and F29 & F33 have high correlation, which is understandable, as these pairs are nothing but the maximum and average of the same quantity measured throughout the corresponding literature. Hence, to reduce the complexity of the classifier, one may use just one of the features from each pair. The results after combining these features are shown in Table 6. It can be seen that even after combining these features, there is no significant degradation in performance of our model.

### 6.2.  The 3C Data Set

As we mention earlier, the data set used is small, due to which the significance of each feature might not be visible explicitly. Hence, we also test our method on the 3C data set, which is larger. The 3C Citation Context Classification[1] Shared Task organized as part of the Second Workshop on Scholarly Document Processing @ NAACL 2021 is a classification challenge where each citation context is categorized based on its purpose and influence. It consists of two subtasks:

- **Task A**: Multiclass classification of citation contexts based on purpose with categories BACKGROUND, USES, COMPARES CONTRASTS, MOTIVATION, EXTENSION, and FUTURE.
- **Task B**: Binary classification of citations into INCIDENTAL or INFLUENTIAL classes (i.e., a task for identifying the importance of a citation).

The training and test data sets used for Task A and Task B are the same. The training data and test data consist of 3,000 and 1,000 instances, respectively. We use the data for Task B in our experiments. However, the 3C data set does not provide us with full text, so we are only able to test only 19 of our features. We achieved an F1 score of 0.5358 with these 19 features on the privately held 3C test set. Our relevant features in use here are F1, F2, F9, F10, F11, F12, F13, F14, F15, F20, F23, F27, F28, F29, F30, F31, F32, F33, and F34. We provide the results on the validation set using a Random Forest Classifier in Table 7. The best performing system in 3C achieved an F1 score of 0.60 while the baseline F1 score was 0.30.

---

[1] https://sdproc.org/2021/sharedtasks.html#3c

**Table 6.** Citation significance results after combining features on the Valenzuela data set

| Precision | Recall | F1 score | Accuracy |
|---|---|---|---|
| 0.91 | 0.80 | 0.85 | 0.89 |

### 6.3. Research Lineage: Case Studies

Our end goal is not just citation classification but to make use of a highly accurate citation significance detection approach to trace significant citations and, thereafter, to try to establish a lineage of the given research. As explained in Section 2, by research lineage we aim to identify the idea propagation via tracking the significant citations. To achieve this, we create a *Significant Citation Graph* (SCG). This is a graph-like structure in which where each node represents a research paper. There is a directed edge between each cited-citing pair, whose direction is from cited paper node to citing paper node, indicating the flow of knowledge from cited paper to citing paper. In the usual case, all citations have equal weights in a citation graph. However, in our case, each edge is labeled as either *significant* or *contextual*, using the approach we discussed in the previous section. Our idea is similar to that of existing scholarly graph databases; however, we go one step further and depict how a particular concept or *knowledge* has propagated with consecutive citations.

Algorithm 1 shows the method to create the adjacency list for the SCG. The Citation Significance Detection ML model is trained on a given data set (Valenzuela in our case). To demonstrate the effectiveness of our method, we present an SCG for a set of papers on *Document-Level Novelty Detection* and *MENNDL*. Being the authors of the papers on these topics, we have identified the significant citations of each paper and used it to test the effectiveness of our proposed method to create an SCG.

**Algorithm 1.** Algorithm to Create Significance Citation Graph

| | |
|---|---|
| | **Input**: Trained Model & concerned research document, P |
| | **Output**: Adjacency List for Citation Graph |
| **1** | Initialize adjacency list, *A* |
| **2** | Initialize an empty queue, *Q* |
| **3** | *Q*.add(*P*) |
| **4** | **while** *Q is not empty* **do** |
| **5** |     **for** *Each citation, C in Q[0]* **do** |
| **6** |         Extract features (F1-F36) for *C* |
| **7** |         **if** *C is Significant **and** C is not in Q* **then** |
| **8** |             *Q*.add(*C*) |
| **9** |             *A*[*Q*[0]].add(*C*) |
| **10** |     *Q*.pop() |
| **11** | **return** *A* |

**Table 7.** Citation Influence Classification results on 3C Validation Set using a Random Forest Classifier

| Precision | Recall | F1 score | Accuracy |
|---|---|---|---|
| 0.569 | 0.575 | 0.572 | 0.606 |

### 6.3.1. Case study I: Document-level novelty detection

Figure 4 denotes an excerpt of a SCG from our *Document-Level Novelty Detection* papers. The red edges denote *significant* citations, whereas black edges denote *contextual* citations. Our approach determined if a citation edge is *significant* or *contextual*. In the citation graph, we are interested in the lineage among four textual novelty detection papers (P1, P2, P3, P4), which the original authors annotate. We annotated that P1 is the *pivot paper* that introduced their document-level novelty detection data set, and their other papers P2, P3, and P4 are based on P1. While P2 and P4 address novelty classification, P3 aims to quantify textual novelty. Our approach conforms to the annotation by the original authors. With P1 as the pivot we can see that there are significant edges from P1 to each of P2, P3, and P4. There is also a significant edge between P2 and P4. However, there is no edge between P2 and P3, as they were contemporaneous submissions and their objective was different (P2 was about *novelty classification* and P3 was about *novelty scoring*). P1 → P2 → P4 forms a research lineage as P2 extends on P1 and P4 extends on P2. Furthermore, we see that P12, P25, P24, and P22 (transitively) are some influential papers for P1. We verified from the authors that P25 was the paper to introduce the first *document-level novelty detection* data set but from an information retrieval perspective. P25 inspired the authors to create the data set in P1 for ML experiments. We infer that P12, P22, and P24 had significant influence on their investigations with P1. Hence, our approach (trained on a different set of papers in the Valenzuela data set) proved successful to identify the significant citations and thereby also identify the corresponding lineage.
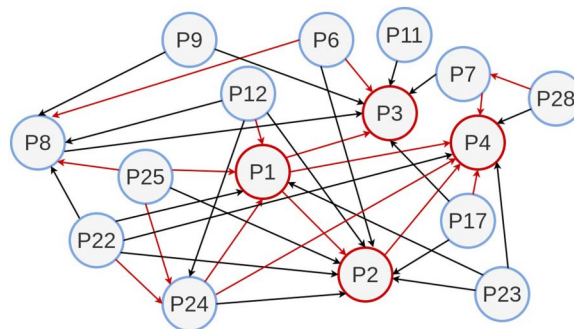


**Figure 4.** Significant Citation Graph for a set of papers on *Document-Level Novelty Detection*. Please refer to the bibliography for the paper details. P1 → Ghosal et al. (2018b), P2 → Ghosal, Edithal et al. (2018a), P3 → Ghosal, Shukla et al. (2019a), P4 → Ghosal, Edithal et al. (2020), P6 → Soboroff and Harman (2003), P7 → Zhao and Lee (2016), P8 → Colomo-Palacios, Tsai, and Chan (2010), P9 → Tang, Tsai, and Chen (2010), P11 → Kusner, Sun et al. (2015), P12 → Li and Croft (2005), P17 → Schiffman and McKeown (2005), P22 → Allan, Wade, and Bolivar (2003), P23 → Soboroff and Harman (2005), P24 → Karkali, Rousseau et al. (2013), P25 → Zhang, Callan, and Minka (2002), P28 → Zhang, Zheng et al. (2015).
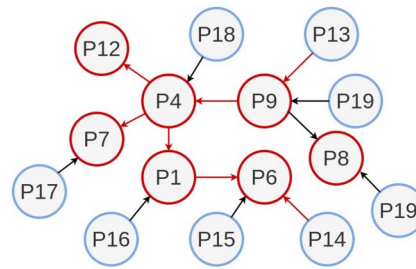
**Figure 5.** Significant Citation Graph for a set of papers on the *MENNDL HPC algorithm*. P1 → Patton, Johnston et al. (2018), P4 → Young, Rose et al. (2017), P6 → Patton, Johnston et al. (2019), P7 → Johnston, Young et al. (2019), P8 → Johnston, Young et al. (2017), P9 → Young et al. (2015), P12 → Chae, Schuman et al. (2019), P13 → Jia, Shelhamer et al. (2014), P14 → Saltz, Gupta et al. (2018), P15 → Thorsson, Gibbs et al. (2018), P16 → Bottou (2010), P17 → Noh, Hong, and Han (2015), P18 → Lucchi, Márquez-Neila et al. (2014), P19 → Baldi, Sadowski, and Whiteson (2014), P20 → Ciregan, Meier, and Schmidhuber (2012), P25 → Y. Zhang et al. (2002), P28 → Zhang et al. (2015).

### 6.3.2. Case study II: MENNDL HPC algorithm

We tested our approach's efficacy to predict the *lineage* of the high-performance computing algorithm MENNDL. We show the research lineage of MENNDL (Young et al., 2015) in Figure 5. We asked the original authors to annotate their research progression with MENNDL. According to the authors, the first paper to describe the MENNDL algorithm was published in 2015, which is deemed the *pivot* (P9). The follow-up paper that carried forward the work in P9 was P4 in 2017. Then P1 came in 2018, building upon P4. P7 and P12 came as extensions of P4. Next, P6 came in 2019 that took forward the work from P1. With P9 as the source, our approach correctly predicted the lineage as P9 → P4 → P1 → P6. Also, the lineage P9 → P4 → P12 and P9 → P4 → P7 via tracing significant citations could be visible in the SCG at Figure 4. We annotate P8 as an application of P9; hence no significant link exists between P9 and P8.

From the above experiments and case studies, it is clear that our proposed method works reasonably well when a paper cites the influencing paper meaningfully. However, there are cases where some papers do not cite the papers by which they are inspired. In such cases, our method would not work.

## 7. CONCLUSION AND FUTURE WORK

In this work, we present our novel idea towards finding a research lineage to accelerate literature review. We achieve *state-of-the-art* performance on citation significance detection, which is a crucial component to form the lineage. We leverage that and show the efficacy of our approach on two completely different research topics. Our approach is simple and could be easily implemented on a large-scale citation graph (given the papers' full text). The training data set is built from NLP papers. However, we demonstrate our approach's efficacy by testing on two topics: one from NLP and the other from HPC, hence establishing that our approach is domain agnostic. Identifying significant citations to form a research lineage would also help the community to understand the real impact of a research beyond simple citation counts. We would look forward to experimenting with deep neural architectures to identify meaningful features for the current task automatically. Our next foray would be to identify the *missing citations* for papers that may have played an instrumental role in certain papers but unfortunately are not cited. We release all the codes related to our experiment at https://figshare.com/s/2388c54ba01d2df25f38.

## AUTHOR CONTRIBUTIONS

Tirthankar Ghosal: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing—Original draft. Piyush Tiwary: Formal analysis, Implementation, Writing—Original draft. Robert Patton: Funding acquisition, Supervision. Christopher Stahl: Conceptualization, Data curation, Project administration, Writing—Review & editing.

## ACKNOWLEDGMENTS

## FUNDING INFORMATION

## DATA AVAILABILITY

We release all the codes related to our experiment at https://figshare.com/s/2388c54ba01d2df25f38.

Q2

## REFERENCES

Aljuaid, H., Iftikhar, R., Ahmad, S., Asif, M., & Afzal, M. T. (2021). Important citation identification using sentiment analysis of in-text citations. *Telematics and Informatics*, *56*, 101492. https://doi.org/10.1016/j.tele.2020.101492

Allan, J., Wade, C., & Bolivar, A. (2003). Retrieval and novelty detection at the sentence level. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 314–321). https://doi.org/10.1145/860435.860493

Alvarez, M. H., Soriano, J. M. G., & Martínez-Barco, P. (2017). Citation function, polarity and influence classification. *Natural Language Engineering*, *23*(4), 561–588. https://doi.org/10.1017/S1351324916000346

Amjad, Z., & Ihsan, I. (2020). VerbNet based citation sentiment class assignment using machine learning. *International Journal of Advanced Computer Science and Applications*, *11*(9), 621–627. https://doi.org/10.14569/IJACSA.2020.0110973

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. *Proceedings of the ACL 2011 Student Session* (pp. 81–87).

Bai, X., Xia, F., Lee, I., Zhang, J., & Ning, Z. (2016). Identifying anomalous citations for objective evaluation of scholarly article impact. *PLOS ONE*, *11*(9), e0162364. https://doi.org/10.1371/journal.pone.0162364, PubMed: 27606817

Baldi, P., Sadowski, P., & Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, *5*(1), 1–9. https://doi.org/10.1038/ncomms5308, PubMed: 24986233

Bartneck, C., & Kokkelmans, S. (2011). Detecting *h*-index manipulation through selfcitation analysis. *Scientometrics*, *87*(1), 85–98. https://doi.org/10.1007/s11192-010-0306-5, PubMed: 21472020

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *Proceedings of Compstat 2010* (pp. 177–186). Springer. https://doi.org/10.1007/978-3-7908-2604-3_16

Camacho-Miñano, M., & Núñez-Nickel, M. (2009). The multilayered nature of reference selection. *Journal of the American Society for Information Science and Technology*, *60*(4), 754–777. https://doi.org/10.1002/asi.21018

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., & Jatowt, A. (2018). YAKE! Collection-independent automatic

keyword extractor. *European Conference on Information Retrieval* (pp. 806–810). https://doi.org/10.1007/978-3-319 -76941-7_80

Cerdá, J. H. C., Nieto, E. M., & Campos, M. L. (2009). What's wrong with citation counts? *D-Lib Magazine*, *15*(3/4). https://doi.org/10 .1045/march2009-canos

Chae, J., Schuman, C. D., Young, S. R., Johnston, J. T., Rose, D. C., ... Potok, T. E. (2019). Visualization system for evolutionary neural networks for deep learning. *2019 IEEE International Conference on Big Data* (pp. 4498–4502). https://doi.org/10.1109 /BigData47090.2019.9006470

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https:// doi.org/10.1613/jair.953

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3642–3649). https://doi.org/10.1109/CVPR.2012.6248110

Cohan, A., Ammar, W., van Zuylen, M., & Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3586–3596). Association for Computational Linguistics. https://doi.org/10.18653/v1/n19-1361

Colomo-Palacios, R., Tsai, F. S., & Chan, K. L. (2010). *Redundancy and novelty mining in the business blogosphere*. The Learning Organization.

Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. *Fifth International Joint Conference on Natural Language Processing* (pp. 623–631). The Association for Computer Linguistics. https://www.aclweb.org/anthology/I11 -1070/.

Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Chivukula, S. S. S. K., & Tsatsaronis, G. (2020). Is your document novel? Let attention guide you. An attention-based model for document-level novelty detection. *Natural Language Engineering*, 1–28. https://doi.org/10.1017/S1351324920000194

Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Tsatsaronis, G., & Chivukula, S. S. S. K. (2018a). Novelty goes deep. A deep neural solution to document level novelty detection. *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2802–2813).

Ghosal, T., Salam, A., Tiwari, S., Ekbal, A., & Bhattacharyya, P. (2018b). TAP-DLND 1.0: A corpus for document level novelty detection. *arXiv*, arXiv:1802.06950.

Ghosal, T., Shukla, A., Ekbal, A., & Bhattacharyya, P. (2019a). To comprehend the new: On measuring the freshness of a document. *2019 International Joint Conference on Neural Networks* (pp. 1–8). https://doi.org/10.1109/IJCNN.2019.8851857

Ghosal, T., Sonam, R., Ekbal, A., Saha, S., & Bhattacharyya, P. (2019b). Is the paper within scope? Are you fishing in the right pond? In M. Bonn, D. Wu, J. S. Downie, & A. Martaus (Eds.), *19th ACM/IEEE Joint Conference on Digital Libraries* (pp. 237–240). IEEE. https://doi.org/10.1109/JCDL.2019.00040

Gilbert, C., & Hutto, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *8th International Conference on Weblogs and Social Media* (Vol. 81, p. 82). https://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf.

Huang, G., Guo, C., Kusner, M. J., Sun, Y., Sha, F., & Weinberger, K. Q. (2016). Supervised word mover's distance. *Advances in Neural Information Processing Systems* (pp. 4862–4870).

Ihsan, I., Imran, S., Ahmed, O., & Qadir, M. A. (2019). A corpus-based study of reporting verbs in citation texts using natural language processing: Study of reporting verbs in citation texts using natural language processing. *CORPORUM: Journal of Corpus Linguistics*, *2*(1), 25–36.

Ji, C., Tang, Y., & Chen, G. (2019). Analyzing the influence of academic papers based on improved PageRank. In E. Popescu, T. Hao, T. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Emerging Technologies for Education – 4th International Symposium* (Vol. 11984, pp. 214–225). Springer. https://doi.org/10.1007/978 -3-030-38778-5_24

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., ... Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *Proceedings of the 22nd ACM International Conference on Multimedia* (pp. 675–678). https://doi.org/10.1145 /2647868.2654889

Johnston, J. T., Young, S. R., Schuman, C. D., Chae, J., March, D. D., ... Potok, T. E. (2019). Fine-grained exploitation of mixed precision for faster CNN training. *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments* (pp. 9–18). https://doi.org/10.1109/MLHPC49564.2019.00007

Johnston, T., Young, S. R., Hughes, D., Patton, R. M., & White, D. (2017). Optimizing convolutional neural networks for cloud detection. *Proceedings of Machine Learning on HPC Environments* (pp. 1–9). https://doi.org/10.1145/3146347.3146352

Karkali, M., Rousseau, F., Ntoulas, A., & Vazirgiannis, M. (2013). Efficient online novelty detection in news streams. *International Conference on Web Information Systems Engineering* (pp. 57–71). https://doi.org/10.1007/978-3-642-41230-1_5

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015). From word embeddings to document distances. *International Conference on Machine Learning* (pp. 957–966).

Laloë, F., & Mosseri, R. (2009). Bibliometric evaluation of individual researchers: Not even right... not even wrong! *Europhysics News*, *40*(5), 26–29. https://doi.org/10.1051/epn/2009704

Li, X., & Croft, W. B. (2005). Novelty detection based on sentence level patterns. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 744–751). https://doi.org/10.1145/1099554.1099734

Lopez, P. (2009). GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. *International Conference on Theory and Practice of Digital Libraries* (pp. 473–474). https://doi.org/10.1007/978-3-642-04346-8_62

Lucchi, A., Márquez-Neila, P., Becker, C., Li, Y., Smith, K., ... Fua, P. (2014). Learning structured models for segmentation of 2-D and 3-D imagery. *IEEE Transactions on Medical Imaging*, *34*(5), 1096–1110. https://doi.org/10.1109/TMI.2014.2376274, PubMed: 25438309

Manju, G., Kavitha, V., & Geetha, T. V. (2017). Influential researcher identification in academic network using rough set based selection of time-weighted academic and social network features. *International Journal of Intelligent Information Technologies*, *13*(1), 1–25. https://doi.org/10.4018/IJIIT.2017010101

McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in translation: Contextualized word vectors. *arXiv*, arXiv:1708 .00107.

Nazir, S., Asif, M., & Ahmad, S. (2020a). Important citation identification by exploiting the optimal in-text citation frequency. *2020 International Conference on Engineering and Emerging Technologies* (pp. 1–6). https://doi.org/10.1109/ICEET48479 .2020.9048224

Nazir, S., Asif, M., Ahmad, S., Bukhari, F., Afzal, M. T., & Aljuaid, H. (2020b). Important citation identification by exploiting

content and section-wise in-text citation count. *PLOS ONE*, *15*(3), e0228885. https://doi.org/10.1371/journal.pone.0228885, PubMed: 32134940

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1520–1528). https://doi.org/10.1109/ICCV.2015.178

Patton, R. M., Johnston, J. T., Young, S. R., Schuman, C. D., March, D. D., ... Kalinin, S. V. (2018). 167-PFlops deep learning for electron microscopy: From learning physics to atomic manipulation. *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 638–648). https://doi.org/10.1109/SC.2018.00053

Patton, R. M., Johnston, J. T., Young, S. R., Schuman, C. D., Potok, T. E., ... Saltz, J. (2019). Exascale deep learning to accelerate cancer research. *2019 IEEE International Conference on Big Data* (pp. 1488–1496). https://doi.org/10.1109/BigData47090.2019.9006467

Perier-Camby, J., Bertin, M., Atanassova, I., & Armetta, F. (2019). A preliminary study to compare deep learning with rule-based approaches for citation classification. *8th International Workshop on Bibliometric-Enhanced Information Retrieval (BIR) co-located with the 41st European Conference on Information Retrieval* (pp. 125–131).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., ... Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv*, arXiv:1802.05365. https://doi.org/10.18653/v1/N18-1202

Pileggi, S. F. (2018). Looking deeper into academic citations through network analysis: popularity, influence and impact. *Universal Access in the Information Society*, *17*(3), 541–548. https://doi.org/10.1007/s10209-017-0565-5

Pride, D., & Knoth, P. (2017a). Incidental or influential? A decade of using text-mining for citation function classification. In J. Qiu, R. Rousseau, C. R. Sugimoto, & F. Xin (Eds.), *Proceedings of the 16th International Conference on Scientometrics and Informetrics* (pp. 1357–1367). ISSI Society.

Pride, D., & Knoth, P. (2017b). Incidental or influential? Challenges in automatically detecting citation importance using publication full texts. In J. Kamps, G. Tsakonas, Y. Manolopoulos, L. S. Iliadis, & I. Karydis (Eds.), *Research and Advanced Technology for Digital Libraries—21st International Conference on Theory and Practice of Digital Libraries* (pp. 572–578). Springer. https://doi.org/10.1007/978-3-319-67008-9_48

Pride, D., & Knoth, P. (2020). An authoritative approach to citation classification. In R. Huang, D. Wu, G. Marchionini, D. He, S. J. Cunningham, & P. Hansen (Eds.), *JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (pp. 337–340). ACM. https://doi.org/10.1145/3383583.3398617

Qayyum, F., & Afzal, M. T. (2019). Identification of important citations by exploiting research articles' metadata and cue-terms from content. *Scientometrics*, *118*(1), 21–43. https://doi.org/10.1007/s11192-018-2961-x

Ronda-Pupo, G. A., & Pham, T. (2018). The evolutions of the rich get richer and the fit get richer phenomena in scholarly networks: The case of the strategic management journal. *Scientometrics*, *116*(1), 363–383. https://doi.org/10.1007/s11192-018-2761-3

Rousseau, R. (2007). The influence of missing publications on the Hirsch index. *Journal of Informetrics*, *1*(1), 2–7. https://doi.org/10.1016/j.joi.2006.05.001

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., ... Thorsson, V. (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Reports*, *23*(1), 181–193. https://doi.org/10.1016/j.celrep.2018.03.086, PubMed: 29617659

Schiffman, B., & McKeown, K. (2005). Context and learning in novelty detection. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 716–723). https://doi.org/10.3115/1220575.1220665

Shen, J., Song, Z., Li, S., Tan, Z., Mao, Y., ... Wang, X. (2016). Modeling topic-level academic influence in scientific literatures. In M. Khabsa, C. L. Giles, & A. D. Wade (Eds.), *Scholarly Big Data: AI Perspectives, Challenges, and Ideas—2016 AAAI Workshop*. AAAI Press. https://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/view/12598.

Shi, C., Wang, H., Chen, B., Liu, Y., & Zhou, Z. (2019). Visual analysis of citation context-based article influence ranking. *IEEE Access*, *7*, 113853–113866. https://doi.org/10.1109/ACCESS.2019.2932051

Soboroff, I., & Harman, D. (2003). Overview of the TREC 2003 novelty track. In *TREC* (pp. 38–53).

Soboroff, I., & Harman, D. (2005). Novelty detection: The TREC experience. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 105–112). https://doi.org/10.3115/1220575.1220589

Tang, W., Tsai, F. S., & Chen, L. (2010). Blended metrics for novel sentence mining. *Expert Systems with Applications*, *37*(7), 5172–5177. https://doi.org/10.1016/j.eswa.2009.12.075

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In D. Jurafsky & É. Gaussier (Eds.), *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 103–110). ACL. https://www.aclweb.org/anthology/W06-1613/. https://doi.org/10.3115/1610075.1610091

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., ... Shmulevich, I. (2018). The immune landscape of cancer. *Immunity*, *48*(4), 812–830. https://doi.org/10.1016/j.immuni.2018.03.023, PubMed: 29628290

Valenzuela, M., Ha, V., & Etzioni, O. (2015). Identifying meaningful citations. In C. Caragea et al. (Eds.), *Scholarly Big Data: AI Perspectives, Challenges, and Ideas*. AAAI Press. https://aaai.org/ocs/index.php/WS/AAAIW15/paper/view/10185.

Van Noorden, R. (2017). The science that's never been cited. *Nature*, *552*. https://doi.org/10.1038/d41586-017-08404-0

Van Noorden, R., & Singh Chawla, D. (2019). Hundreds of extreme self-citing scientists revealed in new database. *Nature*, *572*(7771), 578–579. https://doi.org/10.1038/d41586-019-02479-7, PubMed: 31455906

Vîiu, G.-A. (2016). A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation. *Journal of Informetrics*, *10*(2), 552–566. https://doi.org/10.1016/j.joi.2016.04.010

Wang, F., Jia, C., Liu, J., & Liu, J. (2019). Dynamic assessment of the academic influence of scientific literature from the perspective of altmetrics. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *Proceedings of the 17th International Conference on Scientometrics and Informetrics* (pp. 2528–2529). ISSI.

Wang, M., Zhang, J., Jiao, S., Zhang, X., Zhu, N., & Chen, G. (2020). Important citation identification by exploiting the syntactic and contextual information of citations. *Scientometrics*, *125*(3), 2109–2129. https://doi.org/10.1007/s11192-020-03677-1

West, R., Stenius, K., & Kettunen, T. (2017). Use and abuse of citations. *Addiction science: A guide for the perplexed*, 191. https://doi.org/10.5334/bbd.j

Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, *335*(6068), 542–543. https://doi.org/10.1126/science.1212540, PubMed: 22301307

Xie, Y., Sun, Y., & Shen, L. (2016). Predicating paper influence in academic network. *20th IEEE International Conference on Computer Supported Cooperative Work in Design* (pp. 539–544). IEEE. https://doi.org/10.1109/CSCWD.2016.7566047

Young, S. R., Rose, D. C., Johnston, T., Heller, W. T., Karnowski, T. P., ... Miller, J. (2017). Evolving deep networks using HPC. *Proceedings of the Machine Learning on HPC Environments* (pp. 1–7). https://doi.org/10.1145/3146347.3146355

Young, S. R., Rose, D. C., Karnowski, T. P., Lim, S.-H., & Patton, R. M. (2015). Optimizing deep learning hyper-parameters through an evolutionary algorithm. *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments* (pp. 1–5). https://doi.org/10.1145/2834892.2834896

Zhang, F., & Wu, S. (2020). Predicting future influence of papers, researchers, and venues in a dynamic academic network. *Journal of Informetrics*, *14*(2), 101035. https://doi.org/10.1016/j.joi.2020.101035

Zhang, F., Zheng, K., Yuan, N. J., Xie, X., Chen, E., & Zhou, X. (2015). A novelty-seeking based dining recommender system. *Proceedings of the 24th International Conference on World Wide Web* (pp. 1362–1372). https://doi.org/10.1145/2736277.2741095

Zhang, Y., Callan, J., & Minka, T. (2002). Novelty and redundancy detection in adaptive filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 81–88). https://doi.org/10.1145/564376.564393

Zhao, F., Zhang, Y., Lu, J., & Shai, O. (2019). Measuring academic influence using heterogeneous author-citation networks. *Scientometrics*, *118*(3), 1119–1140. https://doi.org/10.1007/s11192-019-03010-5

Zhao, P., & Lee, D. L. (2016). How much novelty is relevant? It depends on your curiosity. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 315–324). https://doi.org/10.1145/2911451.2911488

Zhu, X., Turney, P. D., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, *66*(2), 408–427. https://doi.org/10.1002/asi.23179