Article

Improving Bibliographic Coupling with Category-Based Cocitation

Rey-Long Liu ^{1,*} and Chih-Kai Hsu ²¹ Department of Medical Informatics, Tzu Chi University, Hualien 97004, Taiwan² Department of MIS, Buddhist Tzu Chi Medical Foundation, Hualien 97004, Taiwan; 107325106@gms.tcu.edu.tw

* Correspondence: rliutcu@mail.tcu.edu.tw; Tel.: +886-3-856-5301 (ext. 2370)

Received: 30 October 2019; Accepted: 26 November 2019; Published: 28 November 2019

Featured Application: The technique presented in the paper can support search engines and researchers in identifying highly related articles, which are essential for the analysis of conclusive findings on specific research issues (e.g., associations of biomedical entities) reported in scientific literature.

Abstract: Bibliographic coupling (BC) is a similarity measure for scientific articles. It works based on an expectation that two articles that cite a similar set of references may focus on related (or even the same) research issues. For analysis and mapping of scientific literature, BC is an essential measure, and it can also be integrated with different kinds of measures. Further improvement of BC is thus of both practical and technical significance. In this paper, we propose a novel measure that improves BC by tackling its main weakness: two related articles may still cite *different* references. *Category-based cocitation* (category-based CC) is proposed to estimate how these different references are related to each other, based on the assumption that two *different* references may be related if they are cited by articles in the *same categories* about specific topics. The proposed measure is thus named BCCCC (Bibliographic Coupling with Category-based Cocitation). Performance of BCCCC is evaluated by experimentation and case study. The results show that BCCCC performs significantly better than state-of-the-art variants of BC in identifying *highly related* articles, which report *conclusive* results on the same *specific* topics. An experiment also shows that BCCCC provides helpful information to further improve a biomedical search engine. BCCCC is thus an enhanced version of BC, which is a fundamental measure for retrieval and analysis of scientific literature.

Keywords: bibliographic coupling; scientific articles; article similarity measures; highly related articles; category-based cocitation

1. Introduction

Given two scientific articles a_1 and a_2 , bibliographic coupling (BC) is a measure to estimate the similarity between a_1 and a_2 by considering how a_1 and a_2 cite a similar set of references [1]. BC works based on an expectation that two articles with a similar set of references may focus on related (or even the same) research issues. The expectation is justified in practice, as BC has been an effective measure for many literature analysis and mapping tasks, such as classification [2] and clustering [3,4] of scientific articles, as well as clustering of scientific journals [5]. Moreover, BC can be applied to more scientific articles, because it works on the references cited by two articles a_1 and a_2 , and the *titles* of these references are more publicly available than other kinds of information about a_1 and a_2 , including full texts of a_1 and a_2 , as well as how a_1 and a_2 are cited by others (many articles are even not cited by any article). BC is thus an effective measure to retrieve similar legal judgments [6] and detect plagiarism [7]

as well. It can also be integrated with different similarity measures, such as those that work on main contents (titles, abstracts, and/or full texts) of articles to cluster articles and map research fields [3,8].

Therefore, further improvement of BC is of practical and technical significance. An improved version of BC can be a fundamental component for the scientometric applications noted above. Equation (1) defines BC similarity between two articles a_1 and a_2 , where R_{a1} and R_{a2} are the sets of references cited by a_1 and a_2 respectively. Therefore, the main weaknesses of BC include: (1) two related articles may still cite different references; and conversely, (2) two unrelated articles may happen to cite the same references. Many techniques have been developed to tackle the weaknesses by collecting additional information from different sources, including titles of the references cited by a_1 and a_2 [9,10] and full texts of a_1 and a_2 [11,12]); however, full texts of many articles are not publicly available.

$$BC(a_1, a_2) = \frac{|R_{a1} \cap R_{a2}|}{|R_{a1} \cup R_{a2}|} \quad (1)$$

In this paper, we aim at improving BC by tackling the first weakness noted above (i.e., two *related* articles may cite *different* references) without relying on full texts of the articles. A different kind of information is considered: category-based cocitation (category-based CC), which measures how a_1 and a_2 cite those references that are cited by articles in the *same categories* about specific topics. A category contains a set of articles that focus on the same research topic (e.g., association between specific entities or events). As articles in the same category are highly related to each other, the references cited by these articles may be related to each other as well, and hence if a_1 and a_2 cite these references, the similarity between a_1 and a_2 can be increased, even though these references have different titles. The enhanced version of BC is thus named BCCCC (Bibliographic Coupling with Category-based Cocitation). Experimental results will show that BCCCC performs significantly better than several state-of-the-art variants of BC that worked on references and their titles, which are more publicly available than full texts of articles.

The intended application of BCCCC is the identification of *highly related* articles. Two articles are highly related only if they *focus* on the *same specific* research issues. Researchers often routinely strive to analyze highly related articles on specific research issues that are in their own interest. A typical example of such applications is the identification of articles that report conclusive results on associations of specific biomedical entities. Many databases are built and maintained to include the associations already published in biomedical literature. Examples of such databases are CTD (Comparative Toxicogenomics Database, available at <http://ctdbase.org>), GHR (Genetic Home Reference, available at <https://ghr.nlm.nih.gov>), and OMIM (Online Mendelian Inheritance in Human, available at <https://www.omim.org>). However, maintenance of the databases is quite costly, as it requires a large number of domain experts to routinely collect and analyze highly related articles to curate the databases [13–15]. BCCCC is a better tool for the domain experts to prioritize those articles that report conclusive results on the same associations. The article prioritization service provided by BCCCC can also be used to improve those bibliometric techniques that analyze articles about specific topics obtained by a keyword-based search (e.g., using the topic names and their related terms as keywords to get articles about equipment maintenance [4], pavement management [16], and pollution by particulate matter [17]). With the keyword-based search, many articles that do not focus on the topics may be retrieved, and conversely, many articles that focus on the topics may not be retrieved. BCCCC can be used to identify those articles that are highly related to the specific topics.

Moreover, as BCCCC is an enhanced version of BC, it can be used to improve those techniques that relied on BC in various domains (e.g., computer science [2], biomedicine [3], legal judgments [6], smart cities [18], and mapping of multiple research fields [5]). BCCCC can also be an improved measure for article clustering, which is a fundamental task of scientometrics studies in various domains (e.g., Internet of things [19] and smart cities [20]).

2. Background

Typical measures to estimate the similarity between scientific articles include text-based measures and citation-based measures. Text-based measures work on textual contents (titles, abstracts, and full texts) of each article, while citation-based measures work on citations of each article. Hybrid measures can be built by integrating multiple measures. These measures are discussed to highlight technical contributions of BCCCC.

2.1. Text-Based Similarity Measures for Scientific Articles

Text-based measures often extract terms from the textual contents of articles, and the similarity between articles is estimated by several factors that are often employed by information retrieval studies. These factors are concerned with each term t , each article a , and how t appears in a . Typical factors include length of a , average length of articles, frequency of t appearing in a (i.e., term frequency, TF), as well as inverse document frequency (IDF) of t in a collection of articles, which measures how *rarely* t appears in these articles. BM25 [21] was one of the best techniques that integrate these factors to identify related scientific articles [22]. Other factors include occurrence of the stem of t (i.e., the base or root form of t), positions of t in a , and key terms specified for a . These factors, together with some factors noted above (TF, IDF, and article length), were employed by the article recommendation service provided by PubMed, which is a popular biomedical search engine [23,24]. This service was found to be one of the best to cluster scientific articles [22].

Instead of working on textual contents of articles, BCCCC is a citation-based measure that works on *out-link references* in each article. BCCCC can thus contribute a different kind of information that can be integrated with the text-based measures noted above. In the experiments, we investigate the performance of two baselines: (1) a system that applies BM25 to reference titles of articles (see Sections 4.2 and 4.4); and (2) the article recommendation service of PubMed (see Section 4.7). Experimental results show that BCCCC can contribute helpful information to the text-based measures.

2.2. Citation-Based Similarity Measures for Scientific Articles

Citation-based similarity measures mainly fall into two types: (1) those that consider *in-link* citations (how an article is cited by other articles) and (2) those that consider *out-link* citations (i.e., how an article cites references). Cocitation (CC [25]) and bibliographic coupling (BC [1]) are representative techniques of the two types, respectively. By integrating their main ideas, a hybrid citation-based measure can be built (e.g., [26]).

CC is based on the idea that two articles a_1 and a_2 may be related to each other if they are cited by the same articles. It was successfully applied to certain applications, such as support of transdisciplinary research [27], professional similarity analysis for authors of articles [28], classification of webpages [2,29], and patent analysis [30]. A typical way to improve CC is to consider *proximity* of a_1 and a_2 in the citing articles [31–34] and *context passages* for a_1 and a_2 in the citing articles [35]. However the proximity and the context passages need to be collected from full texts of the articles, which are often not publicly available. Another way to improve CC is to analyze the cocitation network to deal with the case where an article is cited by very few articles [31]. However, for those articles that are not cited by any article, it is still hard to use CC to identify similar articles. Typical examples of such articles are those new articles that are published recently, which are often the main targets for research professionals to update their knowledge.

BCCCC is an enhanced version of BC, which works on out-link references (rather than in-link citations employed by CC), making it able to deal with those articles that get very few or even no citations. As noted in Section 1, BCCCC is developed to tackle a weakness of BC: two *related* articles may still cite *different* references. To tackle the weakness, BCCCC does not rely on full texts of the articles, which are often not publicly available. When estimating the similarity between two articles a_1 and a_2 , BCCCC employs *category-based* CC, which estimates the similarity between the *references* cited

by a_1 and a_2 by considering how these references are cited by *articles in the same categories*. We will show that, with category-based CC, BCCCC performs significantly better than several state-of-the-art baselines that work on out-link references. As BCCCC is an enhanced version of BC, it can be used to improve the applications of BC. It can also be used to improve those hybrid measures that integrate BC with text-based similarity measures [2,3,8,12] and citation-based similarity measures [26].

3. Development of BCCCC

The main ideas of BCCCC are illustrated in Figure 1, which also highlights main ideas of BC and its two variants IBS (Issue-Based Similarity [9]) and DBC (DescriptiveBC [10]). BC estimates the similarity between two articles a_1 and a_2 by considering whether they cite the *same references*, i.e., Type I similarity in Figure 1. For example, in Figure 1, r_{1m} cited by a_1 is simply r_{23} cited by a_2 , and hence BC may increase the similarity between a_1 and a_2 . Both IBS and DBC estimate the similarity by considering how a_1 and a_2 cite *references with similar titles*, i.e., Type II similarity in Figure 1. For example, in Figure 1, r_{13} (cited by a_1) and r_{22} (cited by a_2) have similar titles (even though they are different references), and hence IBS and DBC may increase the similarity between a_1 and a_2 . In addition to the above two types of similarity (i.e., Type I and Type II), BCCCC also considers how a_1 and a_2 cite those references that are cited by articles in the same *categories* (i.e., Type III similarity, which is measured by category-based CC). For example, in Figure 1, r_{11} (cited by a_1) and r_{21} (cited by a_2) are cited by two articles in the same category (even though the two citing articles are different articles), and hence BCCCC may increase the similarity between a_1 and a_2 .

Figure 1. Main ideas of Bibliographic Coupling with Category-based Cocitation (BCCCC), which is a version of BC enhanced by considering three types of similarity (Type I to Type III), including the one based on *category-based CC* (i.e., Type III).

More specifically, the similarity between two references r_1 and r_2 is composed of two parts: (1) text-based similarity (for Type I similarity and Type II similarity noted above); and (2) citation-based similarity (for Type III similarity noted above). Equation (2) defines the text-based similarity, which is 1.0 if r_1 and r_2 are the same reference; otherwise it is estimated based on the percentage of the terms shared by titles of r_1 and r_2 . When computing the percentage of the shared terms, each term has a weight measured by its IDF (inverse document frequency, see Equation (3)). IDF of a term t measures how *rarely* t appears in the titles of the references of articles. If fewer articles with references mentioning t in their titles, t will get a larger IDF value. Two references will have larger text-based similarity if their titles share many terms with larger IDF values.

$$TextSim_{ref}(r1, r2) = \begin{cases} 1, & \text{if } r1 = r2; \\ \frac{\sum_{t \in Title(r1) \cap Title(r2)} IDF(t)}{\sum_{t \in Title(r1) \cup Title(r2)} IDF(t)}, & \text{otherwise.} \end{cases} \quad (2)$$

$$IDF(t) = \log_2 \frac{\text{Total number of articles} + 0.5}{\text{Number of articles whose references mention } t \text{ in their titles} + 0.5} \quad (3)$$

The citation-based similarity between two references r_1 and r_2 is defined in Equation (4), where *CategorySpaceVec* is a vector with C dimensions (C is the number of categories), and *CosineSim* is the cosine similarity between two vectors (i.e., cosine of the angle of the two vectors). Each dimension in *CategorySpaceVec*(r) corresponds to a category, and the value on the dimension is the number of articles (in the category) that cite r . Therefore, references r_1 and r_2 will have large citation-based similarity if they are cited by articles in a similar set of categories. In that case, r_1 and r_2 are actually cited by articles with a similar set of research focuses, and hence may be related to each other.

$$CitationSim_{ref}(r1, r2) = CosineSim(CategorySpaceVec(r1), CategorySpaceVec(r2)) \quad (4)$$

By integrating the two kinds of similarity (i.e., the text-based similarity and the citation-based similarity), the similarity between two references r_1 and r_2 can be estimated. The two kinds of similarity are integrated in a linear way (see Equation (5)), with the text-based similarity having a larger weight because when the text-based similarity is quite high (e.g., it is 1.0 if r_1 and r_2 are the same reference), the similarity between r_1 and r_2 should be high as well, no matter whether the citation-based similarity is large or not. Therefore, the weight of the text-based similarity is set to 1.0, while the citation-based similarity (i.e., parameter k in Equation (5)) may be set to 0.5. We also expect that proper setting for k should not be a difficult task (the expectation will be justified by experimental results in Section 4.6).

$$Sim_{ref}(r1, r2) = Minimum\{1.0, TextSim_{ref}(r1, r2) + k \times CitationSim_{ref}(r1, r2)\} \quad (5)$$

Based on the similarity between references, the similarity between two articles a_1 and a_2 is defined in Equation (6), where R_a is the set of references cited by article a . The similarity is estimated by considering how a_1 cites those references that are similar to the ones cited by a_2 , and vice versa. Therefore, for each reference r in R_{a1} (R_{a2}), BCCCC identifies its most similar reference r_{max} in R_{a2} (R_{a1}). The similarity between each r and its r_{max} is the basis on which the similarity between a_1 and a_2 is estimated. Two articles are similar to each other if they cite references that are assessed (by Equation (5)) to be similar to each other. In that case, the two articles may be related to each other as they cite those references that are related to each other.

$$BCCCC(a1, a2) = \frac{\sum_{r1 \in R_{a1}} \max_{r2 \in R_{a2}} Sim_{ref}(r1, r2)}{|R_{a1}|} \times \frac{\sum_{r2 \in R_{a2}} \max_{r1 \in R_{a1}} Sim_{ref}(r1, r2)}{|R_{a2}|} \quad (6)$$

4. Experiments

4.1. Experimental Data

Development of BCCCC is motivated by the need of researchers who routinely collect new research results on specific associations between biomedical entities, such as genes, diseases, and chemicals. The experimental data was thus collected from CTD, which recruits biomedical experts to curate associations to support biomedical professionals to do further research on the associations already published in literature [36,37]. The associations are of three types: <chemical, gene>, <chemical, disease>, and <gene, disease>. All the associations in CTD were downloaded in August 2017. As in [9], we exclude those associations that are not supported by direct evidence (i.e., those diseases that have no ‘marker/mechanism’ or ‘therapeutic’ relations to chemicals or genes). Each association

has scientific articles that have been confirmed (by CTD experts) to be focusing on the association. Therefore, an association can thus be seen as a *category* of articles that are *highly related* to each other (i.e., they report conclusive findings on the same association, rather than a single entity). Given a target article a_T , a better system should be able to rank higher those articles that are highly related to a_T (i.e., those that are in the same category of a_T). Such a system is essential for researchers (including CTD experts and biomedical professionals) to retrieve, validate, and curate conclusive findings on specific topics reported in literature.

As we are investigating whether BCCCC is an improved version of BC, those articles whose references cannot be retrieved from PubMed Central are removed (PubMed Central is available at <https://www.ncbi.nlm.nih.gov/pmc>). Categories (associations) without multiple articles are removed, and categories with the same set of articles are treated as a category. There are 16,273 categories, within which there are 12,677 articles for experimentation. The data items are available as Supplementary Materials, see Tables S1–S3 for the three types of categories, respectively. Titles of the references in the articles are preprocessed by removing stop words with a stop word list provided by PubMed (available at <https://www.ncbi.nlm.nih.gov/books/NBK3827/table/pubmedhelp.T.stopwords>). To process synonyms properly, MetaMap (available at <https://metamap.nlm.nih.gov>) is employed to replace biomedical entities in the titles with their concept IDs.

The 12,677 articles are then randomly and evenly split into 20 parts so that we can conduct 20-fold cross validation: each fold is selected as a *test fold* exactly one time and the other folds are used to collect category-based CC information for BCCCC to rank the articles in the test fold, and the process repeats twenty times. In the experiment on a test fold f , each article a_T in f is selected as the *target article* exactly one time. BCCCC and all the baseline systems (see Section 4.2) rank the other articles in f based on how these articles are similar to a_T . To objectively evaluate BCCCC, no information should be available to indicate the categories of the articles in a test fold (i.e., these articles should be treated as “new” articles for which category information is *not* provided). Therefore, when BCCCC collects category-based CC information to rank the articles in a test fold, category information of these articles is removed.

As noted above, each article a_T in a test fold f is selected as the target article exactly one time. Among the articles in a test fold f to be ranked, those that belong to the same categories of a_T are *highly related* to a_T , and hence should be ranked higher. Performance of each system can be evaluated by measuring how the systems rank the highly related articles for each target article (those target articles without highly related articles in f are excluded in the experiment). Average performance on the target articles is then reported (see the evaluation criteria defined in Section 4.3).

4.2. Baseline Systems for Performance Comparison with BCCCC

Table 1 lists the baseline systems in the experiments. The first three baselines are BC and its variants: IBS (Issue-Based Similarity [9]) and DBC (DescriptiveBC [10]). They are the main baselines in the experiments, as BCCCC aims at being an improved version of BC. Instead of relying on full texts of articles that may not be publicly available [11,12], IBS and DBC improve BC by references in the articles. Therefore, performance comparison with BC and the two variants can identify the contribution of BCCCC to further improvement of BC, which has been a fundamental technique for literature retrieval, analysis, and mapping.

Table 1. Baseline systems for performance comparison with BCCCC.

Baseline System	Usage of the References		
	Individual		Concatenated
	Object-Based	Title-Based	
(1) BC (Bibliographic Coupling)	√		
(2) IBS (Issue-Based Similarity)		√	
(3) DBC (DescriptiveBC)		√	
(4) BM25ref (BM25 on references)			√

To estimate the similarity between two articles a_1 and a_2 , the three baselines consider the references cited by a_1 and a_2 . The key difference is that BC treats each reference in an “object-based” manner, because the similarity between a_1 and a_2 is increased only when they cocite the same objects (i.e., references, ref. Equation (1)). On the other hand, IBS and DBC were developed to improve BC by “title-based” similarity estimation, in which the similarity between a_1 and a_2 can be increased if they cite references with *similar* titles, even though these references are *different* from each other. IBS estimates the similarity between two articles based on a certain number of *most-similar* references titles in the articles, while DBC estimates the similarity based on *all* references’ titles in the articles. It was shown that, by considering the title-based similarity, IBS and DBC performed significantly better than BC in article clustering [9] and article ranking [10]. Therefore, BC and the two state-of-the-art variants (IBS and DBC) can be the main baselines to verify whether BCCCC is a further improved version of BC. For more detailed definitions for IBS and DBC, the readers are referred to [9] and [10], respectively.

Moreover, to evaluate BCCCC more comprehensively, BM25ref is implemented as a baseline as well. This baseline represents a way that ranks articles by text-based similarity on reference titles. As noted in Section 2, BM25 [21] was one of the best text-based techniques to identify related scientific articles [22]. We apply BM25 to estimating the similarity between *concatenated reference titles* (CRTs) of articles. For each article, a CRT is constructed by concatenating all titles of the references cited by the article. The similarity between a target article a_T and another article a_x is simply the BM25 similarity between their CRTs (denoted by CRT_T and CRT_x , respectively). BM25ref similarity is defined in Equation (7), where k_1 and b are two parameters, $|CRT|$ is the number of terms in CRT (i.e., length of CRT), $avglen$ is the average length of CRTs (following several previous studies [10,22], the two parameters k_1 and b of BM25ref are set to 2 and 0.75 respectively).

$$BM25ref(a_T, a_x) = \sum_{t \in CRT_T \cap CRT_x} \frac{TF(t, CRT_x)(k_1 + 1)}{TF(t, CRT_x) + k_1(1 - b + b \frac{|CRT_x|}{avglen})} \log_2 IDF(t) \quad (7)$$

Therefore, BM25ref is *not* a version of BC, although it relies on the references cited by articles as well. It is thus actually *not* the main baseline in the experiments. However, comparison of BCCCC and BM25ref can provide additional evidence to further validate whether BCCCC can perform better than a typical text-based approach, which works on reference titles as well. BCCCC can be an enhanced version of BC only if it performs significantly better than all the baselines.

4.3. Evaluation Criteria

As noted above, each article will be a target article exactly one time. Therefore, for each target article, we evaluate how the systems rank its highly related articles (i.e., those that are judged by CTD experts to be focusing on the same research topic as the target article). Two evaluate criteria that are commonly employed by previous studies (e.g., [10]) are employed to evaluate the systems. The first criterion is Mean Average Precision (MAP), which measures how highly related articles are ranked at higher positions. MAP is defined in Equation (8), where T is the set of target articles, and $AvgPrecision(i)$ is the average precision for the i th target article. MAP is thus the average of the $AvgPrecision$ values for all the target articles.

$$MAP = \frac{\sum_{i=1}^{|T|} \text{AvgPrecision}(i)}{|T|} \quad (8)$$

For each target article, AvgPrecision is defined in Equation (9), where H_i is the number of articles that are highly related to the i th target article, and $Rank_{i,j}$ is the rank of the j th highly related article of the i th target article. As the system being evaluated aims at ranking articles, $Rank_{i,j}$ is determined by the system, and hence $Rank_{i,j}$ is actually the number of articles that readers have read when the j th highly related article is recommended by the system. The ratio $j/Rank_{i,j}$ can thus be seen as the precision (achieved by the system) when the j th highly related article is shown. AvgPrecision(i) is simply the average of the precision values on all highly related articles of the i th target article. It is in the range [0–1], and it will be 1.0 when all the highly related articles are ranked at top- H_i positions.

$$\text{AvgPrecision}(i) = \frac{\sum_{j=1}^{H_i} \frac{j}{Rank_{i,j}}}{H_i} \quad (9)$$

Therefore, MAP is concerned with how all highly related articles are ranked at higher positions. In some practical cases, readers may only care about how highly related articles are ranked at *top* positions (e.g., readers only read a certain number of articles at top positions). Therefore, another evaluation criterion average P@X is employed as well. This criterion considers those articles that are ranked at top-X positions only. It is defined in Equation (10), where P@X(i) is the precision when top-X articles are shown to the readers for the i th target article (as defined in Equation (11)). As readers often care about a limited number of top positions only, X should be set to a small value, and hence we investigate performance of the systems when X is set to 1, 3, 5, and 10.

$$\text{Average P@X} = \frac{\sum_{i=1}^{|T|} \text{P@X}(i)}{|T|} \quad (10)$$

$$\text{P@X}(i) = \frac{\text{Number of top-X articles that are highly related to the } i^{\text{th}} \text{ target article}}{X} \quad (11)$$

By simultaneously measuring performance of the systems in both MAP and average P@X, we can comprehensively evaluate how the systems rank all highly related articles, as well as how highly related articles are ranked at top positions. A better system should be able to perform significantly better than others in both evaluation criteria.

4.4. Performance of BCCCC and the Baselines

Figure 2 shows performance of all systems. To verify whether differences of the performance of BCCCC and the baselines are *statistically significant*, a two-tailed and paired *t*-test with 99% confidence level is conducted. The results show that BCCCC performs significantly better than each baseline in all evaluation criteria MAP and Average P@X ($X = 1, 3, 5$, and 10). When compared with the best baseline DBC, BCCCC contributes 10.2% improvement in MAP (0.5708 vs. 0.5180), indicating that it is more capable of ranking highly related articles at higher positions. When only the top positions are considered, BCCCC yields a 7.1% improvement in Average P@1 (0.5383 vs. 0.5025), 9.8% improvement in Average P@3 (0.3620 vs. 0.3297), 9.3% improvement in Average P@5 (0.2850 vs. 0.2607), and 10.4% improvement in Average P@10 (0.1967 vs. 0.1782).

Figure 2. Performance of all systems in terms of the evaluation criteria mean average precision (MAP) and Average P@X ('•' on a system indicates that difference of the performance of BCCCC and the system is statistically significant with $p < 0.01$).

The results justify the contribution of category-based CC to BC. The best baselines, DBC and IBS, improve BC by considering text-based similarities between reference titles. BCCCC performs significantly better than them by considering category-based CC. BCCCC is thus a further improved version of BC, which is a critical method routinely used to retrieve, cluster, and classify scientific literature. Development of BCCCC can thus significantly advance the state of the art of literature analysis.

We further measure the percentage of the target articles that have highly related articles ranked at top- X positions ($X = 1, 3, 5$, and 10). A higher percentage indicates that the system performs more stably in identifying highly related articles for different target articles, making the system more helpful in practice. Figure 3 shows the results. BCCCC achieves the best performance again. When compared with the best baseline, DBC, it yields a 7.1% improvement when $X = 1$ (53.83% vs. 50.25%), 6.0% improvement when $X = 3$ (73.52% vs. 69.39%), 4.7% improvement when $X = 5$ (80.29% vs. 76.71%), and 4.2% improvement when $X = 10$ (87.59% vs. 84.08%). BCCCC contributes larger improvements when X is smaller, indicating that it is more capable in ranking highly related articles at top positions for more articles.

Figure 3. Percentage of highly related articles that are ranked at top positions by the systems.

4.5. A Case Study

We conduct a case study on a target article (PubMed ID: 22707478 [38]) to further analyze the contribution of BCCCC, as seen in Figure 4. Based on the curation by CTD experts, the article focuses on associations of the chemical Bisphenol A with several genes such as ESRRG and ESR1 (i.e., associations <Bisphenol A, ESRRG> and <Bisphenol A, ESR1>). Bisphenol A is a synthetic compound that exhibits estrogen-mimicking properties. ESRRG (Estrogen Related Receptor Gamma) and ESR1 are two genes that respectively encode estrogen receptor-related receptors and Estrogen Receptor α (ER α).

Figure 4. A case study to analyze how BCCCC successfully ranks a highly related article (article 18197296) higher than a less related article (article 17850458) for a given target article (article 22707478).

As a test article, article 18197296 [39] focuses on associations of the gene ESRRG with two chemicals including Bisphenol A (i.e., association <Bisphenol A, ESRRG>). This article is thus highly related to the target article (i.e., article 22707478 noted above), with the association <Bisphenol A, ESRRG> as their common research focus. Another test article is article 17850458 [40]. It focuses on associations of the chemical Estradiol with two genes ESR1 and ESR2 (i.e., associations <Estradiol, ESR1> and <Estradiol, ESR2>). Estradiol is a female sex hormone, while ESR2 is a gene that encodes Estrogen Receptor β (ER β). Therefore, although this article and the target article (22707478) have a common focus on the gene ESR1, they are *not* highly related, as they actually focus on associations of ESR1 with *different* chemicals (article 17850458 focuses on <Estradiol, ESR1>, but article 22707478 focuses on <Bisphenol A, ESR1>).

Therefore, given article 22707478 as a target, article 18197296 is a highly related article, while article 17850458 is a less related article, as seen in Figure 4, and hence the former should be ranked higher than the latter. However, better baselines in the experiment fail to do so. They prefer article 17850458 to article 18197296 by ranking article 17850458 at the top three positions (DBC: top position; IBS: top position; BM25ref: the 3rd position; BC: the 3rd position), but article 18197296 after the 11th position. BCCCC successfully ranks the less related article at the lower position (the 7th position) and the highly related article at the top position.

We further analyze why BCCCC can rank the highly related article (i.e., article 18197296) at the top position for the target article (i.e., article 22707478). References cited by the two articles tend to have low text-based similarities in their titles (i.e., $TextSim_{ref}$ is low), while many of these references have high citation-based similarities (i.e., $CitationSim_{ref}$ is high). Only 15 pairs of the references have $TextSim_{ref}$ 0.15, but 67 pairs of the references have $CitationSim_{ref}$ 0.5. This is the reason why BCCCC can successfully rank article 18197296 high, but the baselines cannot. Figure 5 shows an example to illustrate the analysis. Article 22707478 (the target article) and article 18197296 (the highly related article) respectively cite articles 22101008 [41] and 12185669 [42] as references (see r_1 and r_2 in Table 2). The two references share no terms in their titles, and hence their text-based similarity ($TextSim_{ref}$) is 0. On the other hand, although the two references are *not* cocited by any articles, they are cited by *different* articles in the same categories (see categories c_1 to c_3 in Figure 5). Therefore, by category-based cocitation, $CitationSim_{ref}$ between the two references is high (0.647), indicating that the two references may be highly related, and hence BCCCC similarity between the target article and the highly related article can be increased.

Figure 5. An example to show how BCCCC successfully increases the similarity between the target article (ID: 22707478) and its highly related article (ID: 18197296).

Table 2. Example references (and the similarity between them) noted in the case study.

Article	A Reference Cited by the Article	Text-Based Similarity & Citation-Based Similarity
Target article 22707478 [38]: Gestational exposure to bisphenol a produces transgenerational changes in behaviors and gene expression.	r_1 : cited reference 22101008 [41]: Neonatal Bisphenol A exposure alters sexually dimorphic gene expression in the postnatal rat hypothalamus.	$TextSim_{ref}(r_1, r_2) = 0$ $CitationSim_{ref}(r_1, r_2) = 0.647$
Highly related article 18197296 [39]: Direct evidence revealing structural elements essential for the high binding ability of bisphenol A to human estrogen-related receptor-gamma.	r_2 : cited reference 12185669 [42]: To ERR in the estrogen pathway.	
Target article 22707478 [38]: Gestational exposure to bisphenol a produces transgenerational changes in behaviors and gene expression.	r_3 : cited reference 9454668 [43]: Masculine sexual behavior is disrupted in male and female mice lacking a functional estrogen receptor alpha gene.	$TextSim_{ref}(r_3, r_4) = 0.21$ $CitationSim_{ref}(r_3, r_4) = 0$
Less related article 17850458 [40]: Effects of organisational oestradiol on adult immunoreactive oestrogen receptors (alpha and beta) in the male mouse brain.	r_4 : cited reference 10536018 [44]: Survival of reproductive behaviors in estrogen receptor beta gene-deficient (betaERKO) male and female mice.	

We then analyze why BCCCC can rank the less related article (i.e., article 17850458) at a lower position (the 7th position) for the target article (i.e., article 22707478). Many references cited by the two articles have high text-based similarities, but a smaller number of them have high citation-based similarities (49 pairs of the references have $TextSim_{ref}$ 0.15, and 39 pairs of the references have $CitationSim_{ref}$ 0.5). This is the reason why BCCCC can successfully rank article 17850458 lower, but the baselines cannot. Figure 6 shows an example to illustrate the analysis. Article 22707478 (the target article) and article 17850458 (the less related article) respectively cite articles 9454668 [43] and 10536018 [44] as references, see r_3 and r_4 in Table 2. The two references share many terms in their titles (e.g., ‘behavior’, ‘estrogen receptor’, ‘gene’, ‘male’, ‘female mice’), and hence their text-based similarity ($TextSim_{ref}$) is high (0.21). However, they are not cited by any article in the same categories, as illustrated in Figure 6, and hence their citation-based similarity ($CitationSim_{ref}$) is 0.

Figure 6. An example to show how BCCCC successfully reduces the similarity between the target article (ID: 22707478) and a less related article (ID: 17850458).

Detailed analysis also justifies that the two references actually focus on different issues. As noted in their titles, see r_3 and r_4 in Table 2, they actually focus on $ER\alpha$ and $ER\beta$, respectively. Estrogen receptors modulate many different biological activities (e.g., reproductive organ development, cardiovascular systems, and metabolism), and $ER\alpha$ and $ER\beta$ are encoded by different genes and have different biological functions [45]. Therefore, term overlap in titles of references (as considered by DBC and IBS) may not be reliable in measuring the similarity between the references. BCCCC considers category-based CC to collect additional information to further improve the similarity estimation.

4.6. Effects of Different Settings for BCCCC

We further investigate the effects of different settings for BCCCC. There is a parameter k that governs the relative weight of the category-based CC component of BCCCC (see Equation (5)). In the above experiments, k is set to 0.5. It is interesting to investigate whether this parameter is difficult to set (i.e., whether performance of BCCCC changes dramatically for different settings for k).

Figure 7 shows performance of BCCCC with ten different settings for k in [0.1–1.0]. It is interesting to note that performance in each evaluation criterion (i.e., MAP, and Average P@X) does not change dramatically. BCCCC with $k = 0.5$ does not have significantly different performance than BCCCC with some of the other settings for k , especially, when k is in [0.3–0.5], all performance differences are not statistically significant. Therefore, it is not a difficult task to set the parameter k in practice. Setting k as [0.3–0.5] may be good for BCCCC.

Figure 7. Effects of setting different weights for the citation-based similarity component of BCCCC (i.e., k in Equation (5), and ‘•’ indicates that performance difference produced by setting k to the value and 0.5 is statistically significant with $p < 0.01$).

Another different setting for BCCCC is the way to compute the citation-based similarity. In the above experiments, BCCCC employs category-based CC (see Equation (4)). Another setting for BCCCC is to replace Equation (4) with *article-based* cocitation (i.e., article-based CC), which is a traditional cocitation measure defined in Equation (12) [2,29], where I_{a1} and I_{a2} are the sets of articles that cite articles a_1 and a_2 , respectively (i.e., in-link citations of a_1 and a_2 , respectively).

$$CC_{article}(a1, a2) = \frac{|I_{a1} \cap I_{a2}|}{|I_{a1} \cup I_{a2}|} \quad (12)$$

Therefore, article-based CC can be seen as a “constrained” version of category-based CC, as cocitation is counted only if two references are cited by the same article (rather than articles in the same category). Figure 8 shows the performance of the different settings. BCCCC with category-based CC performs significantly better than BCCCC with article-based CC in all evaluation criteria. The results justify the contribution of category-based CC, which provides additional helpful information even when two references are not cocited by the same article. It is also interesting to note that, as BCCCC with article-based CC still performs better than the baselines, as seen in Figures 2 and 8, it can be a good version as well, especially when no categories of articles are provided in practice.

Figure 8. Contribution of *category*-based CC when compared with traditional CC, which employs *article*-based CC (‘•’ indicates that performance difference is statistically significant with $p < 0.01$).

4.7. Potential Application of BCCCC to Biomedical Search Engines

We further investigate the potential application of BCCCC to biomedical search engines by comparing its performance with PMS (a PubMed service), which is a service provided by PubMed to recommend related articles for a given article. As noted in Section 2, PubMed is a popular search engine for biomedical professionals, and PMS integrates several kinds of well-known indicators [23,24]. These indicators are routinely employed in information retrieval systems as well. PMS was also one of the best to cluster scientific articles as well [22]. Therefore, by comparing how BCCCC and PMS identify highly related articles, the potential contribution of BCCCC to biomedical article recommendation can be evaluated.

For each target article, related articles recommended by PubMed were collected on 18 October and 19 October 2019. We focused on those target articles for which the numbers of recommended articles were less than 200. Some of the articles recommended for a target article a_T may not be the test articles in the experiments, and hence there is no conclusive information to validate whether these articles are highly related to a_T . Therefore, to conduct objective performance comparison, these articles

are excluded so that both BCCCC and PMS can work on the same set of test articles whose relatedness to a_T has been validated by CTD experts.

More specifically, given a target article a_T , let P_T be the set of articles that are recommended by PubMed and included in the test articles judged by CTD experts. Let l be the *lowest* rank of the articles (in P_T) in the ranked list produced by BCCCC, and B_T be the set of articles that BCCCC ranks at the 1st to the l th positions. Therefore, B_T includes articles in P_T , as well as those that are ranked higher by BCCCC. Articles in B_T thus fall into two types: (1) those that are recommended by both BCCCC and PMS (i.e., the set P_T); and (2) those that are recommended by BCCCC but *not* PMS (i.e., the difference set $B_T - P_T$). With this experimental setting, it is reasonable to expect that PMS actually prefers the former (i.e., articles in P_T) to the latter (i.e., articles in $B_T - P_T$).

Performance of PMS and BCCCC can be compared by measuring their precision and recall on P_T and B_T , respectively. Precision of PMS is the percentage of highly related articles in P_T , while precision of BCCCC is the percentage of highly related articles in B_T (see Equations (13) and (15)). Computation of recall requires the number of highly related articles that should be retrieved. This number can be seen as the highly related articles in B_T (see denominators of Equations (14) and (16), and if there is no highly related articles in B_T , the target article a_T is excluded from the experiment), because BCCCC recommends all articles in B_T , while PMS may only recommend some of them (as noted above). Recall of BCCCC is thus always 1.0 (see Equation (16)), and we are investigating whether this is at the cost of recommending more articles and thus possibly reducing its precision.

$$\text{Precision}_{\text{PMS}}(a_T) = \frac{\text{Number of highly related articles in } P_T}{|P_T|} \quad (13)$$

$$\text{Recall}_{\text{PMS}}(a_T) = \frac{\text{Number of highly related articles in } P_T}{\text{Number of highly related articles in } B_T} \quad (14)$$

$$\text{Precision}_{\text{BCCCC}}(a_T) = \frac{\text{Number of highly related articles in } B_T}{|B_T|} \quad (15)$$

$$\text{Recall}_{\text{BCCCC}}(a_T) = \frac{\text{Number of highly related articles in } B_T}{\text{Number of highly related articles in } B_T} \quad (16)$$

Therefore, there often exist tradeoffs between precision and recall, and hence the F1 measure is computed. F1 is a measure commonly used in information retrieval studies to harmonically integrate precision and recall (see Equation (17)).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

The results show that the average F1 values of PMS and BCCCC are 0.7786 and 0.8420, respectively, indicating that BCCCC performs 8.1% better than PMS. A significance test (two-tailed and paired *t*-test) also shows that the performance difference is statistically significant ($p < 0.01$). PMS performs worse as it cannot recommend many highly related articles that are recommended by BCCCC. As PMS is a practical system that recommends scientific articles based on titles and abstracts of articles, it is helpful for the article recommendation services to consider the citation-based information collected by BCCCC, especially when the system aims at recommending highly related articles.

5. Conclusions and Future Work

BC is a similarity measure applicable to scientific articles that cite references, because it estimates the similarity between two articles by measuring how the two articles cite a similar set of references. BC is thus an effective and fundamental measure for retrieval, analysis, and mapping of scientific literature. However, BC has a main weakness: two related articles may still cite *different* references. The proposed new measure, BCCCC, tackles the weakness of category-based CC, which estimates how

these different references are related to each other. Development of category-based CC is based on the assumption that two different references may be related if they are cited by articles in the same categories about specific topics.

The performance of BCCCC is evaluated by experiments and validated in a case study. The results show that BCCCC is an improved version of BC, as it performs significantly better than state-of-the-art variants of BC. The contribution of category-based CC to BC is thus justified. Moreover, effects of different settings for BCCCC are investigated as well. The results show that setting a proper parameter for BCCCC is not a difficult task, and article-based CC may still be helpful, although it is less helpful than category-based CC. We also investigated the potential contribution of BCCCC to biomedical search engines. The results show that BCCCC performs significantly better than the article recommendation service provided by PubMed, which is a popular search engine routinely employed by biomedical professionals. BCCCC can thus provide a different kind of helpful information to further improve the search engine, especially in recommending articles that are highly related to each other (i.e., focusing and reporting conclusive results on the same specific topics).

An application of BCCCC is the identification of highly related articles. As noted above, BCCCC can be used to improve PubMed in recommending highly related articles, which is a service required by biomedical professionals, who often routinely analyze highly related articles on specific research issues. Identification of highly related articles is also required by domain experts that strive to maintain online databases of the associations already published in biomedical literature (e.g., CTD, GHR, and OMIM noted in Section 1). Maintenance of these databases is quite costly, as the domain experts need to routinely collect and analyze highly related articles to curate the databases. The associations already in the databases can be treated as categories for BCCCC to employ category-based CC to prioritize new articles that report conclusive results on the same associations. With the support of BCCCC, curation of new associations can be done in a more timely and comprehensive manner. The bibliographic coupling information provided by BCCCC may be used to improve other search engines in different domains as well.

Another application of BCCCC is the improvement of scientometric techniques that have been used in various domains. BCCCC can be used to improve these techniques in retrieval, clustering, and classification of scientific literature. Moreover, BCCCC is an enhanced version of BC, which is often integrated with different measures. It is thus reasonable to expect that these measures can be further improved by incorporating the idea of BCCCC.

BCCCC improves BC by category-based CC (i.e., $CitationSim_{ref}$, see Equation (4)), which is integrated with a similarity component working on the titles of the references (i.e., $TextSim_{ref}$, ref. Equation (2)). It is thus interesting to investigate how category-based CC can work with other kinds of text-based similarity components so that identification of highly related articles can be further improved. For example, text-based similarity can be measured by considering the abstracts of the references, rather than only the titles of the references. The abstract of a reference is often a commonly available part describing the goal of the reference, and hence text-based similarity based on the abstract may be helpful to further improve BCCCC. It is thus interesting to develop methods to (1) recognize the main research focus of a reference from its abstract; (2) estimate the similarity between two references based on their research focuses; and (3) integrate BCCCC with the abstract-based similarity.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-3417/9/23/5176/s1>. The datasets in the experiments are available online as three tables: Table S1: Articles in each chemical–gene association (category); Table S2: Articles in each chemical–disease association (category); and Table S3: Articles in each gene–disease association (category). Each row in the tables provides information about an article for a specific association curated by CTD experts: (1) ID of the first entity; (2) ID of the second entity; (3) ID of the article; and (4) ID of the fold in the experiment (recall that we conduct 20-fold experiment). Each article has two IDs: PubMed ID and PubMed Central ID, with which readers can access the article on PubMed or PubMed Central.

Author Contributions: Conceptualization, R.-L.L.; Data curation, C.-K.H.; Formal analysis, R.-L.L. and C.-K.H.; Funding acquisition, R.-L.L.; Investigation, R.-L.L. and C.-K.H.; Methodology, R.-L.L. and C.-K.H.; Project administration, R.-L.L.; Resources, R.-L.L.; Software, C.-K.H.; Supervision, R.-L.L.; Validation, R.-L.L. and C.-K.H.; Visualization, R.-L.L. and C.-K.H.; Writing—original draft, R.-L.L.; Writing—review & editing, R.-L.L. and C.-K.H.

Funding: The research was supported by Tzu Chi University under the grand TCRPP108007. The APC was funded by Tzu Chi University, Taiwan.

Acknowledgments: The authors are grateful to Shu-Yu Tung and Yun-Ling Lu for collecting the raw data used in the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kessler, M.M. Bibliographic coupling between scientific papers. *Am. Doc.* **1963**, *14*, 10–25. [\[CrossRef\]](#)
2. Couto, T.; Cristo, M.; Goncalves, M.A.; Calado, P.; Nivio Ziviani, N.; Moura, E.; Ribeiro-Neto, B. A Comparative Study of Citations and Links in Document Classification. In Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital libraries, Chapel Hill, NC, USA, 11–15 June 2006; pp. 75–84.
3. Boyack, K.W.; Klavans, R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2389–2404. [\[CrossRef\]](#)
4. Hoppenstedt, B.; Pryss, R.; Stelzer, B.; Meyer-Brötz, F.; Kammerer, K.; Treß, A.; Reichert, M. Techniques and Emerging Trends for State of the Art Equipment Maintenance Systems—A Bibliometric Analysis. *Appl. Sci.* **2018**, *8*, 916. [\[CrossRef\]](#)
5. Thijs, B.; Zhang, L.; Glänzel, W. Bibliographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes. *Scientometrics* **2015**, *105*, 1453–1467. [\[CrossRef\]](#)
6. Kumar, S.; Reddy, K.; Reddy, V.B.; Singh, A. Similarity Analysis of Legal Judgments. In Proceedings of the Fourth Annual ACM Bangalore Conference (COMPUTE 2011), Bangalore, Karnataka, India, 25–26 March 2011.
7. Gipp, B.; Meuschke, N. Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence. In Proceedings of the 11th ACM Symposium on Document Engineering, Mountain View, CA, USA, 19–22 September 2011.
8. Janssens, F.; Glänzel, W.; De Moor, B. A hybrid mapping of information science. *Scientometrics* **2008**, *75*, 607–631. [\[CrossRef\]](#)
9. Liu, R.-L.; Hsu, C.-K. Issue-Based Clustering of Scholarly Articles. *Appl. Sci.* **2018**, *8*, 2591. [\[CrossRef\]](#)
10. Liu, R.-L. A New Bibliographic Coupling Measure with Descriptive Capability. *Scientometrics* **2017**, *110*, 915–935. [\[CrossRef\]](#)
11. Habib, R.; Afzal, M.T. Sections-based bibliographic coupling for research paper recommendation. *Scientometrics* **2019**, *119*, 643–656. [\[CrossRef\]](#)
12. Liu, R.-L. Passage-based Bibliographic Coupling: An Inter-Article Similarity Measure for Biomedical Articles. *PLoS ONE* **2015**, *10*, e0139245. [\[CrossRef\]](#)
13. CTD. When is Data Updated? Available online: <http://ctdbase.org/help/faq/?jsessionid=92111C8A6B218E4B2513C3B0BEE7E63F?p=6422623> (accessed on 29 October 2019).
14. GHR. Expert Reviewers. Available online: <http://ghr.nlm.nih.gov/ExpertReviewers> (accessed on 29 October 2019).
15. OMIM. OMIM®—Online Mendelian Inheritance in Man. Available online: <http://www.omim.org/about> (accessed on 29 October 2019).
16. Pérez-Acebo, H.; Linares-Unamunzaga, A.; Abejón, R.; Rojí, E. Research Trends in Pavement Management during the First Years of the 21st Century: A Bibliometric Analysis during the 2000–2013 Period. *Appl. Sci.* **2018**, *8*, 1041. [\[CrossRef\]](#)
17. Błaszczak, B.; Widziewicz-Rzońca, K.; Ziola, N.; Klejnowski, K.; Juda-Rezler, K. Chemical Characteristics of Fine Particulate Matter in Poland in Relation with Data from Selected Rural and Urban Background Stations in Europe. *Appl. Sci.* **2019**, *9*, 98. [\[CrossRef\]](#)
18. Li, M. Visualizing the studies on smart cities in the past two decades: A two-dimensional perspective. *Scientometrics* **2019**, *120*, 683–705. [\[CrossRef\]](#)
19. Yan, B.-N.; Lee, T.-S.; Lee, T.-P. Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): A co-word analysis. *Scientometrics* **2015**, *105*, 1285–1300. [\[CrossRef\]](#)
20. Appio, F.P.; Lima, M.; Paroutis, S. Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 1–14. [\[CrossRef\]](#)

21. Robertson, S.E.; Walker, S.; Beaulieu, M. Okapi at TREC-7: Automatic ad hoc, filtering, VLC and interactive. In Proceedings of the 7th Text. REtrieval Conference (TREC 7), Gaithersburg, MD, USA, 9–11 November 1998; pp. 253–264.
22. Boyack, K.W.; Newman, D.; Duhon, R.J.; Klavans, R.; Patek, M.; Biberstine, J.R.; Schijvenaars, B.; Skupin, A.; Ma, N.; Börner, K. Clustering More than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches. *PLoS ONE* **2011**, *6*, 18029. [CrossRef]
23. PubMed. Computation of Similar Articles. Available online: https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Computation_of_Similar_Articl (accessed on 29 October 2019).
24. Lin, J.; Wilbur, W.J. PubMed related articles: A probabilistic topic-based model for content similarity. *BMC Bioinform.* **2007**, *8*, 423. [CrossRef]
25. Small, H.G. Co-citation in the scientific literature: A new measure of relationship between two documents. *J. Am. Soc. Inf. Sci.* **1973**, *24*, 265–269. [CrossRef]
26. Zhao, P.; Han, J.; Sun, Y. P-Rank: A Comprehensive Structural Similarity Measure over Information Networks. In Proceedings of the International Conference on Information and Knowledge Management, Hongkong, China, 2–6 November 2009; pp. 553–562.
27. Trujillo, C.M.; Tammy, M.; Long, T.M. Document co-citation analysis to enhance transdisciplinary research. *Sci. Adv.* **2018**, *4*, e1701130. [CrossRef]
28. Jeonga, Y.K.; Songa, M.; Ding, Y. Content-based author co-citation analysis. *J. Informetr.* **2014**, *8*, 197–211. [CrossRef]
29. Calado, P.; Cristo, M.; Moura, E.; Ziviani, N.; Ribeiro-Neto, B.; Goncalves, M.A. Combining Link-Based and Content-Based Methods for Web Document Classification. In Proceedings of the 2003 ACM CIKM International Conference on Information and Knowledge Management (CIKM'03), New Orleans, LA, USA, 3–8 November 2003.
30. Wang, X.; Zhao, Y.; Liu, R.; Zhang, J. Knowledge-transfer analysis based on co-citation clustering. *Scientometrics* **2013**, *97*, 859–869. [CrossRef]
31. Eto, M. Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information. *Inf. Process. Manag.* **2019**, *56*, 102046. [CrossRef]
32. Boyack, K.W.; Small, H.; Klavans, R. Improving the accuracy of co-citation clustering using full text. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64*, 1759–1767. [CrossRef]
33. Liu, S.; Chen, C. The proximity of co-citation. *Scientometrics* **2012**, *91*, 495. [CrossRef]
34. Gipp, B.; Beel, J. Citation Proximity Analysis (CPA)—A new approach for identifying related work based on Co-Citation Analysis. In Proceedings of the 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, 14–17 July 2009; pp. 571–575.
35. Liu, X.; Zhang, J.; Guo, C. Full-text citation analysis: A new method to enhance scholarly networks. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64*, 1852–1863. [CrossRef]
36. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; King, B.L.; McMorran, R.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2017. *Nucleic Acids Res.* **2017**, *45*, D972–D978. [CrossRef] [PubMed]
37. Wiegers, T.C.; Davis, A.P.; Cohen, K.B.; Hirschman, L.; Mattingly, C.J. Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD). *BMC Bioinform.* **2009**, *10*, 326. [CrossRef]
38. Wolstenholme, J.T.; Edwards, M.; Shetty, S.R.; Gatewood, J.D.; Taylor, J.A.; Rissman, E.F.; Connelly, J.J. Gestational exposure to bisphenol a produces transgenerational changes in behaviors and gene expression. *Endocrinology* **2012**, *153*, 3828–3838. [CrossRef]
39. Okada, H.; Tokunaga, T.; Liu, X.; Takayanagi, S.; Matsushima, A.; Shimohigashi, Y. Direct evidence revealing structural elements essential for the high binding ability of bisphenol A to human estrogen-related receptor-gamma. *Environ. Health Perspect.* **2008**, *116*, 32–38. [CrossRef]
40. Kudwa, A.E.; Harada, N.; Honda, S.I.; Rissman, E.F. Effects of organisational oestradiol on adult immunoreactive oestrogen receptors (alpha and beta) in the male mouse brain. *J. Neuroendocrinol.* **2007**, *19*, 767–772. [CrossRef]
41. Cao, J.; Mickens, J.A.; McCaffrey, K.A.; Leyrer, S.M.; Patisaul, H.B. Neonatal Bisphenol A exposure alters sexually dimorphic gene expression in the postnatal rat hypothalamus. *Neurotoxicology* **2012**, *33*, 23–36. [CrossRef]

- 42. Giguère, V. To ERR in the estrogen pathway. *Trends Endocrinol. Metab.* **2002**, *13*, 220–225. [[CrossRef](#)]
- 43. Wersinger, S.R.; Sannen, K.; Villalba, C.; Lubahn, D.B.; Rissman, E.F.; De Vries, G.J. Masculine sexual behavior is disrupted in male and female mice lacking a functional estrogen receptor alpha gene. *Horm. Behav.* **1997**, *32*, 176–183. [[CrossRef](#)] [[PubMed](#)]
- 44. Ogawa, S.; Chan, J.; Chester, A.E.; Gustafsson, J.A.; Korach, K.S.; Pfaff, D.W. Survival of reproductive behaviors in estrogen receptor beta gene-deficient (betaERKO) male and female mice. *PNAS* **1999**, *96*, 12887–12892. [[CrossRef](#)] [[PubMed](#)]
- 45. Lee, H.R.; Kim, T.H.; Choi, K.C. Functions and physiological roles of two types of estrogen receptors, ER α and ER β , identified by estrogen receptor knockout mouse. *Lab. Anim. Res.* **2012**, *28*, 71–76. [[CrossRef](#)] [[PubMed](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).