Scientometrics (2017) 110:915–935
DOI 10.1007/s11192-016-2196-7

A new bibliographic coupling measure with descriptive capability

Rey-Long Liu[1]

Received: 16 August 2016 / Published online: 3 December 2016

Abstract *Bibliographic coupling* (BC) is an effective measure to estimate the similarity between two scholarly articles (i.e., inter-article similarity between the two articles). It works on out-link references of articles (i.e., those references cited by the articles), and is essential for relatedness analysis and topic clustering of scholarly articles. In this paper, we present a new BC measure DescriptiveBC, which employs the *titles* of the out-link references to improve BC in two ways: given a target article *a*, DescriptiveBC provides more accurate information about *how* (based on numerical inter-article similarity) and *why* (based on textual descriptive terms) a scholarly article is related to *a*. Visualization of the information can support the identification, clustering, mapping, and navigation of the related evidence in scientific literature. Empirical evaluation justifies the contributions of DescriptiveBC. Release of the reference titles in each article is thus helpful for the dissemination of research findings in scientific literature, and DescriptiveBC can be incorporated into search engines of scholarly articles to help prospective researchers to navigate through the space of related articles online.

Keywords Bibliographic coupling · Inter-article similarity · Descriptive capability · Reference titles

Mathematics Subject Classification 68P20 · 68U35 · 00B15

JEL Classification D83 · L86

✉ Rey-Long Liu
rlliutcu@mail.tcu.edu.tw

[1] Department of Medical Informatics, Tzu Chi University, Hualien, Taiwan

Introduction

Bibliographic coupling (BC) is a measure to estimate the similarity between two scholarly articles (i.e., inter-article similarity between the two articles). It estimates the inter-article similarity by considering how two scholarly articles co-cite the same references (Kessler 1963). Equation 1 defines the BC similarity between two articles $a1$ and $a2$ (Couto et al. 2006; Calado et al. 2003), where R_{a1} and R_{a2} are the sets of references cited by $a1$ and $a2$ respectively (i.e., sets of out-link references in $a1$ and $a2$ respectively). Similarity between two articles will be larger if they have a higher percentage of out-link references in common.

$$\mathrm{BC}(a_1, a_2) = \frac{|R_{a_1} \cap R_{a_2}|}{|R_{a_1} \cup R_{a_2}|} \tag{1}$$

BC is a practical and effective measure for relatedness analysis and topic clustering of scholarly articles, based on several reasons: (1) BC works on out-link references of articles, which are more publicly available than several kinds of information, including in-link citations of the articles (i.e., how the articles are cited by others, employed by Couto et al. (2006), Small (1973) and full texts of the articles (employed by Liu 2015; Boyack et al. 2013; Aljaber et al. 2010; Gipp and Beel 2009); (2) out-link references were found to be more effective than in-link citations in classification (Couto et al. 2006) and clustering (Boyack and Klavans 2010) of scholarly articles; and (3) BC performed well in other applications as well, including retrieval of similar legal judgments (Kumar et al. 2011) and detection of plagiarism (Gipp and Meuschke 2011).

However, BC has limitations as well, including (1) two articles that are related to each other may still cite different references; and (2) the inter-article citation matrix is often quite sparse, making the BC similarity dominated by very few citations. Several approaches were thus proposed to deal with the limitations, such as employing second-order similarities (Thijs et al. 2015) and incorporating additional information from main textual contents of the articles, including titles and abstracts of the articles (e.g., Boyack and Klavans 2010; Couto et al. 2006) and full texts of the articles (e.g., Liu 2015; Janssens et al. 2008).

It is thus of practical and theoretical significance to improve BC. In this paper, we investigate the contributions of the *titles* of the out-link references to BC. The out-link reference titles are publicly available. They have been noted as a helpful resource to index scholarly articles (Qin 2000; Garfield 1990). We further investigate their contributions to BC, based on two expectations: (1) titles of two references r_1 and r_2 may be used to measure the similarity between r_1 and r_2 (as they respectively indicate the goals of r_1 and r_2), even when r_1 and r_2 are different (and hence these titles may be used to tackle the limitations of BC noted above); and (2) titles of the similar references cited by two articles a_1 and a_2 may provide *descriptive terms* to indicate the *aspect of relatedness* between a_1 and a_2, because an article often cites proper references for each of its research aspect (issue), and titles of the similar references cited by a_1 and a_2 may indicate the common research aspect of a_1 and a_2.

We justify the two expectations by developing a new measure DescriptiveBC (a new BC measure with descriptive capability), and empirically evaluate and analyze the performance of DescriptiveBC in the identification and navigation of highly related articles that have been judged (by domain experts) to be focusing on the same research topics. The results show that, given a target article a, DescriptiveBC provides more accurate information about *how* (based on the inter-article similarity) and *why* (based on the descriptive terms) a scholarly article is related to *a*. Visualization of the information can support the identification, clustering, mapping, and navigation of the related evidence in scientific literature. Release of the reference titles in each article is thus helpful for the dissemination of research findings in scientific literature.

Related work

As DescriptiveBC aims at providing numerical inter-article similarity as well as textual descriptive terms to indicate why two articles are related, we compare DescriptiveBC with previous techniques for *inter-article similarity estimation* and *indicative term identification*.

Inter-article similarity estimation

Previous inter-article similarity measures include (1) *citation-based* measures, which work on citations among the articles; (2) *text-based* measures, which work on textual contents of the articles (i.e., titles, abstracts, and main bodies of articles), and (3) *hybrid* measures, which employ both citation-based and text-based information.

Citation-based measures employ two kinds of citation relationships: *in-link* citations and *out-link* references. In-link citations of an article *a* are those articles that cite *a*, while out-link references of *a* are those articles that *a* cites. Two articles that share many in-link citations or out-link references are expected to be related to each other. *Co-citation* is a representative technique that considers in-link citations (Small 1973). Two articles may be related if they are co-cited by other articles. *Bibliographic coupling* (BC, ref. Eq. 1) is a representative technique that considers out-link references (Kessler 1963). Two articles may be related if they co-cite other articles. As noted above, DescriptiveBC is developed based on BC, because BC works on out-link references of articles, which are more publicly available than in-link citations (many scholarly articles have very few or even no in-link citations). Some previous studies found that out-link references could provide more accurate information than in-link citations for classification (Couto et al. 2006) and clustering (Boyack and Klavans 2010) of scholarly articles.

Text-based measures work on textual contents of the articles, including titles, abstracts, and main bodies of the articles. They are developed based on the expectation that two articles may be related if they share certain similar textual contents, and hence similarity between the two articles are often estimated by measuring how they share those terms with higher weights. A typical term weighting method is *TFIDF*, which is the product of term frequency (TF, number of occurrences of a term in an article) and inverse document frequency (IDF, total number of articles/number of articles in which a term appears). Various techniques were then developed to estimate inter-article similarity by the term weights. The vector space model (VSM) and Latent Semantic Analysis (LSA) are typical techniques that employ the term weights to represent scholarly articles as vectors for similarity estimation (Glenisson et al. 2005; Landauer et al. 2004). However, they did not perform well in several cases (Whissell and Clarke 2013; Boyack et al. 2011). BM25 (Robertson et al. 1998) was found to be one of the best techniques in finding related scholarly articles (Boyack et al. 2011). Given an article a_1 as the target, BM25 employs Eq. 2 to estimate the score (similarity) of another article a_2 with respect to a_1. In Eq. 2, k_1

and b are two parameters, $|a|$ is the number of terms in article a (i.e., length of a), *avgal* is the average number of terms in an article (i.e., average length of articles).

$$\mathrm{BM25}(a_1, a_2) = \sum_{t \in a_1 \cap a_2} \frac{\mathrm{TF}(t, a_2)(k_1 + 1)}{\mathrm{TF}(t, a_2) + k_1 (1 - b + b \frac{|a_2|}{avgal})} \mathrm{Log}_2 \mathrm{IDF}(t) \tag{2}$$

DescriptiveBC works on a textual resource that is different from those employed by the text-based measures. It works on *titles* of out-link references in the articles, rather than the textual contents of the articles. Two highly related articles should share similar core textual contents, which may be expressed in different ways and scattered in the textual contents of the article, making them quite difficult to recognize and measure. We thus aim at investigating the potential contribution of the out-link reference titles. If the contribution is confirmed, DescriptiveBC can provide another kind of useful information that can be used as a complement to the text-based measures.

Hybrid measures work on citation-based information and text-based information, which are different kinds of helpful information that may be integrated to improve inter-article similarity estimation. As there are two kinds of citation-based information (i.e., in-link citations and out-link references), the hybrid measures fall into two types as well, which respectively integrate text-based information with in-link citations and out-link references. The hybrid measures that integrate in-link citations often considered the *positions* and the *context passages* around the place where an article is cited in the full text of another article. They were developed based on the expectation that two articles a_1 and a_2 may be related if (1) they are cited in nearby areas in many articles that cite them (Boyack et al. 2013; Gipp and Beel 2009), or (2) they have similar context passages commented by the authors of their citing articles (Liu et al. 2013; Aljaber et al. 2010), although the citing articles may focus on different parts of a_1 and a_2 (Elkiss et al. 2008; Kumar et al. 2011) with different sentiments (Small 2011). The context passages could also be used for topic-based article retrieval (Liu et al. 2014; Ritchie et al. 2008) and disambiguation of named entities (Nakov et al. 2004). Therefore, these hybrid measures relied on in-link citations and full texts of articles, which are often not publicly available (many articles have very few or even no in-link citations). DescriptiveBC works on another kind of information: out-link-references, which are more publicly available.

Another type of hybrid measures integrate textual information with out-link references. They worked on full texts of the articles (Liu 2015; Janssens et al. 2008), which are often not publicly obtainable, and hence several hybrid measures relied on titles and abstracts of the articles only (Boyack and Klavans 2010; Couto et al. 2006). These hybrid measures did not always perform significantly better than BC (Couto et al. 2006). One of the hybrid measures performed better than BC by treating a co-reference cited by two articles as a co-word in titles and abstracts of the two articles (Boyack and Klavans 2010). DescriptiveBC can be a hybrid measure as it employs both out-link references and their titles. However, the out-link reference titles are not the main contents (i.e., titles and abstracts) of the articles. We aim at showing that the out-link reference titles can be another helpful resource for inter-article similarity estimation.

Indicative term identification

Another main contribution of DescriptiveBC is the provision of descriptive terms to explain *why* two articles are related. The descriptive terms can support the navigation of the articles that may be related to a target article in different ways, as these articles often have different aspects of relatedness to the target article. To our knowledge, no previous inter-article similarity measures provide the *relatedness-indicative* terms.

Previous studies have noted the contribution of providing certain texts for navigation of articles. However, these texts are often for describing a group of articles of interest, rather than indicating how two articles are related. For example, those terms with the best TFIDF weights in a cluster of scholarly articles were used to describe the cluster (Janssens et al. 2009). Those terms that have better discriminative capability within a domain of interest are extracted to build a term map for the readers to navigate (van Eck et al. 2010). These previous studies thus aimed at selecting representative and discriminative terms for a group of articles, rather than *relatedness-descriptive* terms between two articles.

Titles of bibliographically related articles were noted as a resource to represent articles (Salton and Zhang 1986). More specifically, titles of out-link references in a scholarly article were noted as a helpful resource from which additional keywords may be extracted to index the article. Given an article a, the KeyWords Plus[1] in the ISI Web of Knowledge database provides additional keywords extracted from titles of the out-link references in a (Garfield 1990). The additional keywords and the original keywords of a could provide different information (Qin 2000), and hence readers can employ the additional keywords to expand their search. Therefore, the additional keywords aim at indicating the *main contents* of a, while the descriptive terms extracted by DescriptiveBC aim at describing the *relatedness* between a and another article. The KeyWords Plus service and DescriptiveBC thus have different goals.

An enhanced bibliographic coupling measure

DescriptiveBC improves bibliographic coupling by employing *titles* of out-link references in scholarly articles. Given a target article a_T, DescriptiveBC estimates the similarity between a_T and a given article a_x, based on the out-link references in a_T and a_x. No other parts of the articles (e.g., titles, abstracts, and main bodies of the articles) are required. DescriptiveBC also returns a set of descriptive terms to indicate why a_x is related to a_T.

Challenges of DescriptiveBC include (1) estimation of the inter-article similarity by the titles of the out-link references, and (2) selection of the descriptive terms to indicate why two articles are related. Figure 1 illustrates the basic idea of DescriptiveBC. A scholarly article (e.g., the target article a_T) often discusses several issues, and for each issue, cites proper references. These references may have certain degrees of similarity with those references cited by another article a_x (e.g., the candidate articles a_1 to a_3 in Fig. 1). Titles of two references may be helpful for the similarity estimation, especially when the two references are *different* but *related* to each other. Based on the similarity values, similar references cited by a_T and a_x can be identified to indicate the *aspect of relatedness* between a_T and a_x, and hence terms in the titles of these references can be selected to describe why a_x is related to a_T.

More specifically, similarity between two out-link references r_1 and r_2 is defined in Eq. 3. When r_1 and r_2 are identical, the similarity is 1.0 (i.e., the largest similarity), which is the way employed by bibliographic coupling; otherwise the similarity is defined by a *Jaccard index*, which is commonly employed to estimate the similarity between two sets of objects (in this case, we treat the terms in a title as a set, see Eq. 4). Therefore, two

[1] Basic description of the "KeyWords Plus" service can be found at http://interest.science.thomsonreuters.com/content/WOKUserTips-201010-IN.

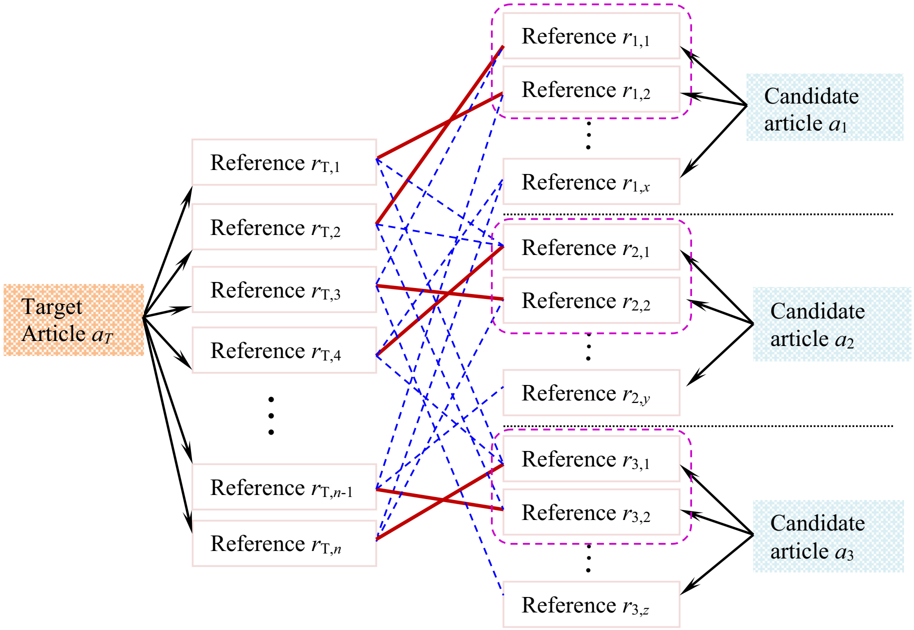


Fig. 1 Main idea of DescriptiveBC: A scholarly article (e.g., the target article a_T) often discusses several issues, and for each issue, cites proper references. These references may have strong similarity (see the *solid lines* between citations) or weak similarity (see the *dashed lines* between citations) with those references cited by other articles (e.g., the candidate articles a_1, a_2, and a_3). Based on these similarity values, terms in the titles of certain references cited by each candidate article (i.e., the citations in the *dashed boxes*) can be selected to describe *why* the candidate article is related to the target article

different references can still have a certain degree of similarity (between 0.0 and 1.0) based on their titles.

$$Sim_{ref}(r_1, r_2) = \begin{cases} 1, & \text{if } r_1 = r_2; \\ \dfrac{|Title(r_1) \cap Title(r_2)|}{|Title(r_1) \cup Title(r_2)|}, & \text{otherwise.} \end{cases} \tag{3}$$

$$Title(r) = Set\ of\ terms\ in\ the\ title\ of\ r \tag{4}$$

With the similarity between two references, DescriptiveBC estimates the similarity between two articles a_1 and a_2 by Eq. 5, where R_{a1} and R_{a2} are the sets of out-link references in a_1 and a_2 respectively. The similarity falls between 0.0 and 1.0 as well. For each reference in a_1, DescriptiveBC identifies the most similar reference in a_2, and vice versa. Similarity between these references are employed to estimate the similarity between a_1 and a_2. Therefore, if a_1 and a_2 have a higher percentage of *similar* references (in terms of Eq. 3), similarity between them will be larger, even when they do *not* co-cite many references. DescriptiveBC is thus based on seamless integration of pure citation-based information (as employed by traditional bibliographic coupling) and text-based information from the titles of the out-link references. The integration can properly deal with the common case where two references are different but related to each other.

$$DescriptiveBC(a_1,a_2) = \frac{\sum_{r_1 \in R_{a_1}} Max_{r_2 \in R_{a_2}} Sim_{ref}(r_1,r_2) + \sum_{r_2 \in R_{a_2}} Max_{r_1 \in R_{a_1}} Sim_{ref}(r_1,r_2)}{|R_{a_1}| + |R_{a_2}|} \tag{5}$$

The second challenge of DescriptiveBC is to properly select *descriptive terms* (DTerms) to indicate why a_2 is related to a_1. All non-stop words in the titles of the out-link references in a_2 are candidate DTerms. Selection of DTerms are based on the *descriptive strength* defined in Eq. 6. The descriptive strength of a term t falls between 0.0 and 1.0 as well. It is estimated by considering the out-link references in which t appears. If t appears in a higher percentage of out-link references in $a2$ and these references are similar to the references in a_1, the descriptive strength t will be larger. In this case, t is more capable of indicating *why* a_2 is related to a_1.

$$Dstrength(a_1,a_2,t) = \frac{\sum_{r_2 \in R_{a_2}; t \in Title(r_2)} Max_{r_1 \in R_{a_1}} Sim_{ref}(r_1,r_2)}{|R_{a_2}|} \tag{6}$$

The final set of DTerms selected to indicate why a_2 is related to a_1 simply consists of α terms (in the titles of the references in a_2) with the highest descriptive strengths (see Eqs. 7 and 8). As the DTerms are mainly used as a guide for readers to navigate through the space of related articles, we expect that α should be set to about 10, which is close to the length of the title of a scholarly article, making the DTerms both brief and informative to indicate the relatedness between two articles.

$$DTermSet(a_1,a_2) = \{t | Dstrength(a_1,a_2,t) \text{ is at top } \alpha \text{ in } RefTitle(a_2)\} \tag{7}$$

$$RefTitle(a) = \bigcup_{r \in R_a} Title(r) \tag{8}$$

DescriptiveBC thus has three interesting features: (1) it works on out-link references, which are more commonly available than several typical kinds of information (e.g., in-link citations and main bodies of articles); (2) it enhances bibliographic coupling by seamlessly integrating the citation-based information and text-based information from the titles of the out-link references; (3) it selects certain terms from the titles of references to describe the aspect of relatedness between two articles. DescriptiveBC can thus provide the information about *how* (by the inter-article similarity) and *why* (by the descriptive terms) two articles are related. The information can support the identification, clustering, and navigation of the related evidence published in scientific literature.

Empirical evaluation

Experiments are designed to empirically evaluate DescriptiveBC in two ways: (1) performance of the inter-article similarity (estimated by Eq. 5) and (2) descriptive capability of the DTerms (extracted by Eq. 7). For the former, we compare DescriptiveBC with several baselines in identifying those articles that are highly related to each other, while for the latter we investigate whether DTerms of a candidate article a_i with respect to a target article a_T can be used to explain *why* a_i is related to a_T.

The data

We evaluate DescriptiveBC in the context of identifying of *highly related* articles, which are those articles that focus on the same research topic. Researchers often need to check multiple articles to cross-validate the evidence about specific research topics. Retrieval of the highly related articles is thus of practical significance. The retrieval is also challenging for inter-article similarity measurement techniques, because the highly related articles should share similar *core textual contents*, which are quite difficult to extract and measure. Contribution of DescriptiveBC and bibliographic coupling to this retrieval task is thus interesting, as they work on out-link references (rather than textual contents) of the articles. Therefore, instead of relying on the main bodies of articles (as done in several previous studies, e.g., Liu 2015), we focus on the out-link references in the articles.

We employ the data in DisGeNET,[2] which maintains a database of articles that focus on specific gene-disease associations. A gene-disease association is a specific research topic, and hence the articles selected for the association are highly related to each other. To facilitate the research of disease diagnosis and therapy, several databases of gene-disease associations have been developed and maintained (e.g., the databases maintained by Genetic Home Reference and Online Mendelian Inheritance in Human), however the maintenance is quite costly because a large number of experts are often recruited to carefully and frequently retrieve and check multiple articles.[3] A good inter-article similarity measure is thus essential for the maintenance of the databases, as it can recommend potential highly related articles for the experts to check.

More specifically, we select from DisGeNET those gene-disease associations that had the largest number of articles annotated by Genetic Association Database[4] (GAD) or Comparative Toxicogenomics Database[5] (CTD) for human. Both GAD and CTD recruit domain experts to select articles to annotate each gene-disease association (Wiegers et al. 2009; Becker et al. 2004). For CTD, doctoral-level experts were trained to select the articles, with a high degree of inter-expert agreement (Wiegers et al. 2009). For each gene-disease association <*g*, *d*>, we designate one article as the *target*, while the others as the *highly related candidates*. Given the target article, an inter-article similarity measure should be able to rank high these highly related candidates, among other candidates that focus on other research topics (i.e., not dedicated to the association between *g* and *d*).

Therefore, for each gene-disease association <*g*, *d*>, we also collect a large number of candidate articles that are *not* dedicated to <*g*, *d*>. These candidate articles are collected by sending two queries to a popular search engine PubMed Central[6] (PMC): "*g* NOT *d*" and "*d* NOT *g*". These articles thus share a certain amount of contents with the target article, however they are *non-highly related* articles for <*g*, *d*> because they mention *g* or *d* but not both. For each gene-disease association, at most 200 non-highly related candidate articles are collected.

We thus totally have 53 topics (gene-disease associations) for which 9928 articles are tested: 9740 articles are *non-highly related* to these topics (and hence on average one topic has 183.77 non-highly related articles), while 188 articles are *highly related* to their

[2] DisGeNET is available at http://www.disgenet.org/web/DisGeNET/menu/home.

[3] The ways of database update by Genetic Home Reference and Online Mendelian Inheritance in Human can be found at http://ghr.nlm.nih.gov/ExpertReviewers and http://www.omim.org/about, respectively.

[4] GAD is available at available at http://geneticassociationdb.nih.gov.

[5] CTD is available at available at http://ctdbase.org.

[6] PubMed Central is available at http://www.ncbi.nlm.nih.gov/pmc.

respective topics (and hence on average one topic has 3.55 highly related articles). These articles totally have 435,786 out-link references whose titles are collected as well. We conduct 188 testes so that each highly related article is designated as the target article (for its respective topic) for exactly one time.

The baselines

As noted in Related work, previous techniques to measure inter-article similarity between two scholarly articles a_1 and a_2 can be *citation-based* (those that worked on citation relationships among articles), *text-based* (those that worked on textual contents of a_1 and a_2), or *hybrid* (those that worked on both text-based and citation-based information). As DescriptiveBC is an enhanced version of bibliographic coupling that works on out-link references of the articles, the main baseline in the experiments is bibliographic coupling (i.e., BC defined in Eq. 1), which is a good measure for the analysis of scientific literature, as well as the retrieval of similar legal judgments (Kumar et al. 2011) and detection of plagiarism (Gipp and Meuschke 2011). We aim at investigating whether DescriptiveBC may be a better bibliographic coupling measure, and if so the contribution of integrating DescriptiveBC with other text-based and hybrid measures can be expected.

Therefore, text-based and hybrid measures are *not* main baselines in the experiments. However, we do implement two state-of-the-art text-based and hybrid measures, with a goal to further compare the contribution of textual contents (employed by the text-based and hybrid measures) and out-link reference titles (employed by DescriptiveBC). The text-based measure is BM25 (Robertson et al. 1998), which was one of the best techniques in finding related scholarly articles (Boyack et al. 2011). It is defined in Eq. 2, with the two parameter k_1 and b being typically set to 2 and 0.75 respectively (Boyack et al. 2011; Liu and Huang 2011). On the other hand, the hybrid measure is HybridK50, which performed better than BC in certain cases (Boyack and Klavans 2010). Similarity between two articles a_1 and a_2 is defined based on the intersection of words and out-link references in a_1 and a_2. HybridK50 estimates the similarity by treating a reference co-cited by a_1 and a_2 as a co-word in the titles and abstracts of a_1 and a_2. Both BM25 and HybridK50 work on titles and abstracts of the articles, which are publicly available on the Internet.

The evaluation criteria

Two criteria are employed to evaluate the performance of the inter-artcile similarity measures. They are *Mean average precision* (MAP) and average P@X, which were routinely employed in text ranking studies (e.g., Liu 2015). MAP is defined in Eq. 9, where |*T*| is the number of tests (recall that we conduct 188 tests, ref. The data), and *AvgP*(*i*) is the average precision for the *i*th test. MAP is simply the average of the *AvgP* values for all the tests.

$$\text{MAP} = \frac{\sum_{i=1}^{|T|} AvgP(i)}{|T|} \tag{9}$$

$$AvgP(i) = \frac{\sum_{j=1}^{h_i} \frac{j}{Seen_i(j)}}{h_i} \tag{10}$$

$AvgP(i)$ is defined in Eq. 10, where h_i is the number of articles that are judged (by domain experts) to be *highly related* to the target article for the ith test (i.e., the ones that focus on the same research topic as the target article), and $Seen_i(j)$ is the number of articles that readers have seen when the jth highly related article for the ith test is shown (i.e., number of articles whose ranks are higher than or equal to that of the jth highly related article for the ith test). Therefore, given a target article a_T for the ith test, if those articles that are highly related to a_T are ranked higher, $AvgP(i)$ will be higher.

On the other hand, average P@X only concerns those articles that are ranked at top-X positions. It is defined in Eq. 11. It is the average of the P@X values for all the 188 tests. Equation 12 defines P@X, which is the precision when top-X articles are shown to the readers. Typically X is set to a small value, and hence P@X evaluates how highly related articles are ranked very high. In the experiments, we set X to 1, 3, 5, and 10.

$$\text{Average P@X} = \frac{\sum_{i=1}^{|\mathrm{T}|} \mathrm{P@X}(i)}{|T|} \tag{11}$$

$$\mathrm{P@X}(i) = \frac{\text{Number of top} - \text{X articles that are highly related to the target for the } i\text{th test}}{\mathrm{X}} \tag{12}$$

Moreover, to verify whether the performance differences between DescriptiveBC and each of the baselines are *statistically significant*, we conduct significance tests by two-sided and paired t-tests.

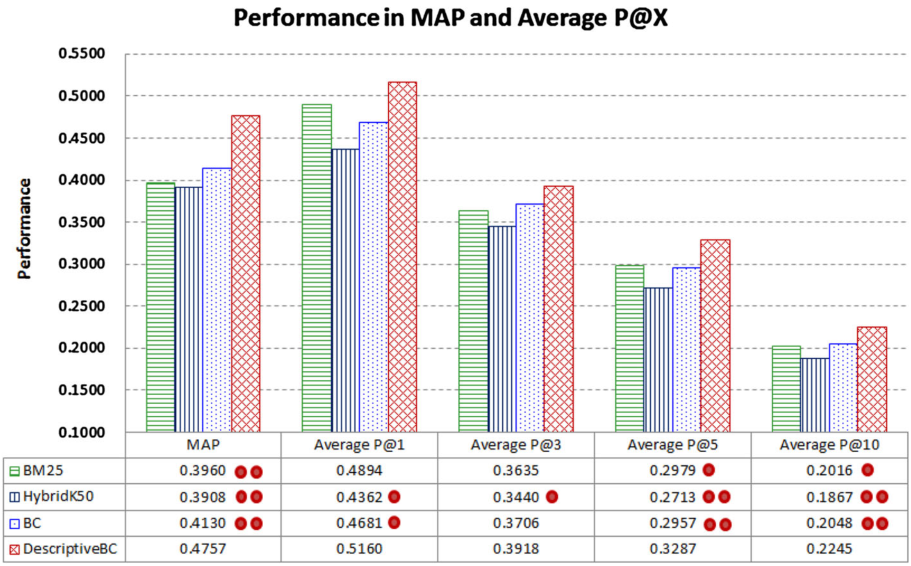
Results

We report the experimental results on (1) performance of DescriptiveBC in identifying those articles that are highly related to each other, and (2) descriptive capability of the DTerms in indicating why two articles are related.

Performance of inter-article similarity estimation

As shown in Fig. 2, DescriptiveBC performs better than BC in all evaluation criteria, with statistically significant performance improvements in MAP ($p < 0.005$), Average P@1 ($p < 0.05$), Average P@5 ($p < 0.005$), and Average P@10 ($p < 0.005$). The improvements in MAP is 15.2% (0.4757 vs. 0.4130). The results confirm the contribution of the titles of the out-link references in an article: these titles provide additional useful information that can be used to properly indicate the possible similarity between two *related* but *different* out-link references, which may share certain research targets reflected in their titles. The original version of BC does not consider the possible similarity, while DescriptiveBC is an enhanced version that employs a proper way to estimate (by Eq. 3) and utilize the similarity (by Eq. 5).

Figure 2 also shows the performance of BM25 and HybridK50, which consider the title and the abstract of each article as the resources to estimate inter-article similarity. The hybrid approach HybridK50 (which considers both the text and the out-link references) tends to have comparable performance with the text-based approach BM25 and the citation-based approach BC. The difference in MAP between HybridK50 and BM25 is *not*

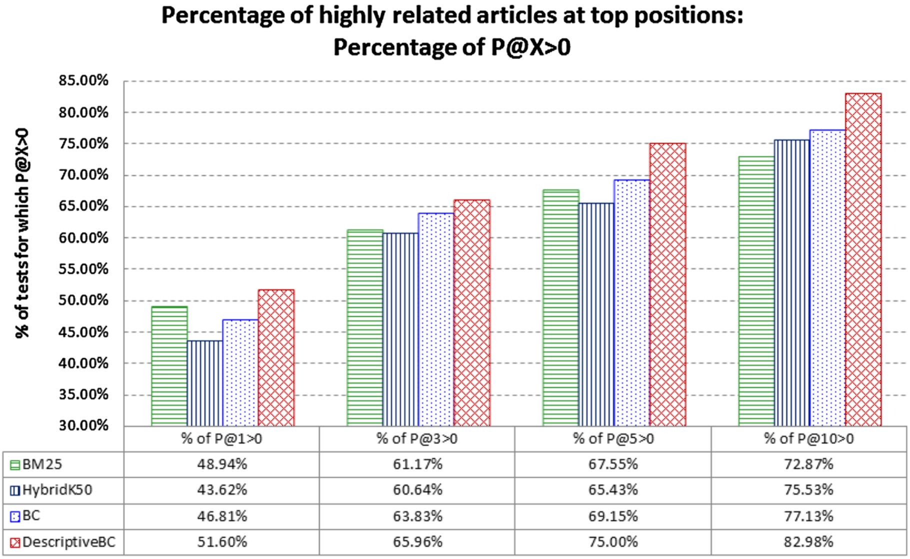


	MAP	Average P@1	Average P@3	Average P@5	Average P@10
▤ BM25	0.3960 ●●	0.4894	0.3635	0.2979 ●	0.2016 ●
▥ HybridK50	0.3908 ●●	0.4362 ●	0.3440 ●	0.2713 ●●	0.1867 ●●
□ BC	0.4130 ●●	0.4681 ●	0.3706	0.2957 ●●	0.2048 ●●
☒ DescriptiveBC	0.4757	0.5160	0.3918	0.3287	0.2245

Fig. 2 MAP and average P@X: DescriptiveBC performs significantly better than all the baselines ('*single round*' and '*double round*' on a system indicate that performance difference between the system and DescriptiveBC is statistically significant with $p < 0.05$ and $p < 0.005$, respectively)

statistically significant ($p = 0.7972$), and the difference between HybridK50 and BC is *not* statistically significant either ($p = 0.1858$). HybridK50 was found to have comparable performance with BC in article clustering, as it performed slightly better than BC in certain cases (Boyack and Klavans 2010). Our experimental results further show that, it is still challenging to properly employ the textual contents (i.e., titles and abstracts considered by BM25 and HybridK50) to *significantly* improve the performance in identifying those articles that are *highly related* to specific topics (in such case, there may be many non-highly related articles that share several key terms with the highly related articles). DescriptiveBC performs significantly better than both BM25 and HybridK50. Note that we are not aiming at *directly* comparing DescriptiveBC with BM25 and HybridK50, because DescriptiveBC does *not* consider the titles and the abstracts. However, the results indicate that the titles of the out-link references in an article a_x is another good resource for inter-article similarity estimation as well (when compared with the title and the abstract of a_x). The contributions are of practical significance to the identification of highly related evidence already published in literature.

Figure 3 shows the *percentage* of the topics for which P@X > 0. A higher percentage indicates that, for a larger portion of the tests, highly related articles are successfully ranked at top positions. A system that achieves a higher percentage has both good and stable performance in recommending highly related articles for different research topics. Again, DescriptiveBC achieves higher percentages than BC in all settings of X. DescriptiveBC contributes 10.2% improvement in the percentage in P@1 (51.60 vs. 46.81%), 3.3% improvement in the percentage in P@3 (65.96 vs. 63.83%), 8.5% improvement in the percentage in P@5 (75.00 vs. 69.15%), and 7.6% improvement in the percentage in P@10 (82.98 vs. 77.13%). DescriptiveBC performs better than HybridK50 and BM25 as well. The contribution is of practical significance to researchers, who often need to check a large number of scholarly articles for different research topics.



	% of P@1>0	% of P@3>0	% of P@5>0	% of P@10>0
▤ BM25	48.94%	61.17%	67.55%	72.87%
▥ HybridK50	43.62%	60.64%	65.43%	75.53%
⬚ BC	46.81%	63.83%	69.15%	77.13%
⊠ DescriptiveBC	51.60%	65.96%	75.00%	82.98%

Fig. 3 Percentage of the tests for which P@X $>$ 0: DescriptiveBC ranks the highly related articles at top-1, top-3, top-5, and top-10 for a higher percentage of tests than all the baselines

Descriptive capability of descriptive terms

Having the results on the ranking of highly related articles, we are also concerned with the descriptive capability of the DTerms identified by DescriptiveBC. Given a target article a_T, descriptive capability of the DTerms of the candidate articles with respect to a_T can be investigated by answering two questions:

(**Q1**) Do highly related articles and non-highly related articles for a_T tend to have different DTerms?

(**Q2**) Do the DTerms of highly related articles tend to be related to the main topic of a_T?

Descriptive capability of the DTerms can be verified if answers to both questions are 'yes'. Q2 is motivated by the observation: the target article a_T and its highly related articles are judged and selected (by domain experts) for a specific topic (i.e., association between a specific gene and a specific disease), and hence the highly related articles are related to a_T mainly because they all focus on the topic, making their DTerms (with respect to a_T) related to the topic. On the other hand, Q1 is motivated by the observation: non-highly related articles tend to be not related to the topic, and hence their DTerms (with respect to a_T) would not be so related to the topic. Therefore, although all the articles may have certain degrees of similarity to a_T, the highly related articles should have those DTerms that are (1) different from those of non-highly related ones *and* (2) related to the topic (i.e., properly indicating *why* the highly related articles are related to a_T).

Investigation of question Q1

We *quantitatively* investigate Q1 by measuring two factors: (1) average DTerm similarity between highly related articles and (2) average DTerm similarity between highly and non-highly related articles. The ratio between the two factors is a kind of *Dunn index*, which is often used to evaluate the results of clustering (note that in this case, we are *not* clustering the articles, but treating highly related articles and non-highly related articles as two “groups” of articles). A higher ratio indicates that, with the DTerms extracted, the candidate articles can be well separated into a group of highly related articles and a group of non-highly related articles.

More specifically, we employ Eq. 13 to measure the similarity between two sets of DTerms (D_1 and D_2). It is a kind of *Jaccard index*, which is commonly used to measure the similarity between two sets.

$$DTermSimilarity(D_1, D_2) = \frac{|D_1 \cap D_2|}{|D_1 \cup D_2|} \tag{13}$$

For each of those topics that have multiple highly related articles (in addition to a randomly selected target article), Fig. 4 shows the average DTerm similarity between highly related articles, as well as the average DTerm similarity between highly and non-highly related articles. The result shows that the former is significantly higher than the latter (in a two-tailed and paired *t-test*, $p < 0.0002$), and the average ratio between them is 2.0217. For most topics, highly related articles share several DTerms, while non-highly related articles share fewer DTerms with the highly related articles. Highly related articles and non-highly related articles thus tend to have different DTerms, and hence the answer to question Q1 should be ‘yes’. Therefore, in addition to the inter-article similarity estimated by DescriptiveBC, the DTerms extracted by DescriptiveBC can be used to further distinguish highly related articles from non-highly related articles.

Investigation of question Q2

We *qualitatively* investigate Q2 by case study and analysis. Given a target article a_T for a topic, we analyze whether DTerms of highly related articles of a_T can semantically indicate

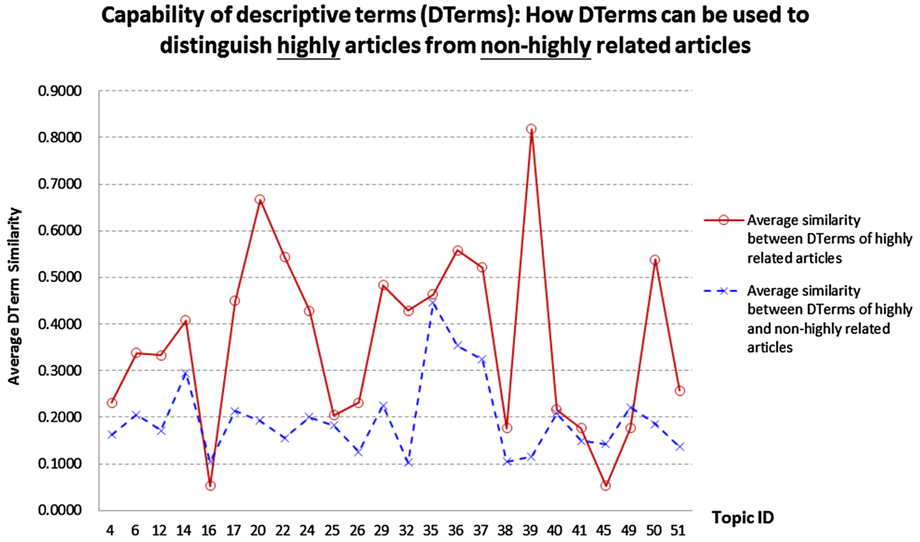


Fig. 4 DTerms similarity between highly related articles versus DTerms similarity between highly and non-highly related articles: The former is significantly higher than the latter. For most topics, highly related articles share several DTerms, while non-highly related articles share fewer DTerms with the highly related articles. DTerms extracted by DescriptiveBC can thus be used to further distinguish highly related articles from non-highly related articles

the topic (i.e., properly indicating *why* the highly related articles are related to a_T). Based on the result in Fig. 4, two representative topics are identified: Topic 39 and Topic 45, which respectively get the largest and the smallest ratios between the two factors reported in Fig. 4 (the ratios are 7.0840 and 0.3699, respectively). Highly related articles for Topic 39 (Topic 45) have the most similar (dissimilar) sets of DTerms. We thus investigate whether these DTerms are related to the two topics, respectively.

Topic 39 is about the association between pancreatitis and serine peptidase inhibitor. PubMed Central ID of the target article is PMC1774044.[7] Figure 5 shows the distribution of DTerms of top-30 candidate articles (based on the DescriptiveBC similarity of each article to the target article). Recall that we extract ten DTerms for each candidate article, and a candidate article has one DescriptiveBC similarity to the target. Therefore, DTerms of a candidate article have the same x-coordinate in Fig. 5. It is interesting to note that DescriptiveBC successfully ranks the two highly related articles at the top positions, and these highly related articles have almost the same sets of DTerms (denoted by 'o'). The non-highly related articles, however, have broad and different distributions of DTemrs (denoted by '×').

We further analyze whether DTerms of the highly related articles are related to Topic 39. Table 1 lists DTerms of three articles, including the two highly related articles (IDs: PMC1773194[8] and PMC1773221[9]), as well as a non-highly related article (ID: PMC2928535[10]) that gets the largest similarity to the target article. The two highly related articles share a high percentage of DTerms (nine out of ten DTerms are the same), including the name of the target disease (pancreatitis) and gene-related terms (mutation and gene), indicating that the two highly related are related to the target article in discussing genetic events of the disease. Moreover, the shared DTerms also include several terms that are related to the target gene (serine peptidase inhibitor). These terms are cationic, trypsinogen, and inhibitor, and trypsin. This is because cationic trypsinogen is actually an enzyme. It is a serine peptidase produced in the pancreas for the digestion of food.[11] Therefore, both highly related articles share many DTerms that are closely related to Topic 39. With the DTerms, readers can easily know *in what way* the two articles are related to the target article. On the other hand, the non-highly related article has several DTerms different from those of the highly related articles. The DTerms include the names of another gene (the CFTR gene) and another disease (cystic fibrosis), which is different but related to the target disease (pancreatitis).[12] Therefore, the non-highly related article is related to the target article in a way different from Topic 39 (<pancreatitis, serine peptidase inhibitor>). Different articles tend to be related to the target article in *different* ways, and the DTerms have indicated *why* they are related.

Topic 45 is another representative topic for case study. It is about the association between anemia and erythropoietin. Anemia is a condition in which the amount of red blood cells is not enough, while the erythropoietin gene controls the production of the red

[7] The title of PMC1774044 is "Absence of PRSS1 mutations and association of SPINK1 trypsin inhibitor mutations in hereditary and non-hereditary chronic pancreatitis".

[8] The title of PMC1773194 is "The N34S mutation of SPINK1 (PSTI) is associated with a familial pattern of idiopathic chronic pancreatitis but does not cause the disease".

[9] The title of PMC1773221 is "Mutations in serine protease inhibitor Kazal type 1 are strongly associated with chronic pancreatitis".

[10] The title of PMC2928535 is "Inhibition of acinar apoptosis occurs during acute pancreatitis in the human homologue ΔF508 cystic fibrosis mouse".

[11] A basic description for cationic trypsinogen and serine peptidase can be found at Genetic Home Reference: https://ghr.nlm.nih.gov/gene/PRSS1.

[12] A basic description for the CFTR gene and cystic fibrosis can be found at Genetic Home Reference: https://ghr.nlm.nih.gov/condition/cystic-fibrosis#genes.

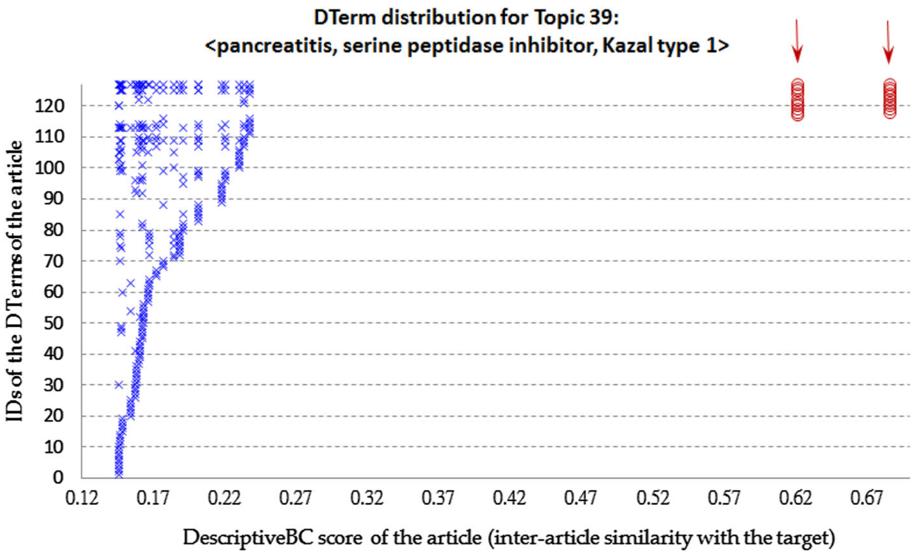


Fig. 5 A case study (Topic 39) for which highly related articles share many DTerms (denoted by *circle* and pointed by *down arrow*): The shared DTerms (see Table 1) are related to the topic of the case study (association between pancreatitis and serine peptidase inhibitor), and so can indicate *why* the articles are highly related to the target article, which focuses on the topic as well. All non-highly related articles get much lower similarity to the target article. They focus on other topics with many DTerms (denoted by *times*) not related to the association between pancreatitis and serine peptidase inhibitor, indicating that they are similar to the target in *different* ways

blood cells.[13] ID of the target article is PMC3441831.[14] It presents the correlation between erythropoietin and anemia associated with melanoma, which is a type of cancer. The target article was thus judged (by domain experts) to be highly related to Topic 45. In addition to the target article, there are two highly related articles. The first highly related article (ID: PMC1386105[15]) presents the operation of erythropoietin in treating anemia, while the second highly related article (ID: PMC2754516[16]) presents the correlation between erythropoietin and anemia associated with a digestive disease. Therefore, both articles were judged to be highly related to Topic 45 (<anemia, erythropoietin>) as well.

Topic 45 deserves case study and analysis, due to two reasons: (1) DescriptiveBC fails to rank the two highly related articles at top positions (pointed by '↓' in Fig. 6), and hence many non-highly related articles are miss-ranked at top positions; and (2) the two highly related articles have only one DTerm in common, while many non-highly related articles share several DTemrs with the first highly related article (pointed by '↓' at the right part of Fig. 6). Therefore, it is interesting to investigate whether the DTerms are closely related to Topic 45.

[13] A basic description for the erythropoietin (EPO) gene can be found at Genetic Home Reference: https://ghr.nlm.nih.gov/gene/EPO#.

[14] The title of PMC3441831 is "Erythropoietin Receptor Contributes to Melanoma Cell Survival in vivo".

[15] The title of PMC1386105 is "Signals for stress erythropoiesis are integrated via an erythropoietin receptor–phosphotyrosine-343–Stat5 axis".

[16] The title of PMC2754516 is "Use of agents stimulating erythropoiesis in digestive diseases".

Table 1 DTerms of three articles for Topic 39 (<pancreatitis, serine peptidase inhibitor>): Highly related and non-highly related articles tend to be related to the target article in *different* ways, and the DTerms indicate why they are related

DTerm	1st highly-related article	2nd highly related article	Top non-highly related article	Related to topic 39?
[127] pancreatitis	v	v	v	Yes
[126] mutation	v	v	v	Yes
[125] chronic	v	v	v	?
[124] gene	v	v	v	Yes
[123] hereditary	v			Yes
[122] trypsinogen	v	v		Yes
[121] cationic	v	v		Yes
[120] associated	v	v		?
[119] inhibitor	v	v		Yes
[118] trypsin	v	v		Yes
[117] idiopathic		v		?
[116] fibrosis			v	No
[115] cystic			v	No
[114] cell			v	?
[113] acute			v	?
[112] cftr			v	No
[111] transmembrane			v	?

Table 2 lists DTerms of three articles, including the two highly related articles, as well as a non-highly related article (ID: PMC1890992[17]) that gets the largest similarity to the target article (i.e., it is ranked at top-1). The non-highly related article explores the correlation between erythropoietin and neuroblastoma, which is a type of cancer but not the target disease anemia, and hence it is not closely related to Topic 45. We find that DTerms of the three articles tend to indicate why the articles are related to the target article, based on two reasons: (1) the two highly related articles have several DTerms related to the target disease or the target gene of Topic 45 (including erythropoietin, receptor, erythroid, anemia, and epoetin alfa[18]), without any DTerms related to other diseases or genes; and (2) the non-highly related article has several DTerms related to development of cancer (including angiogenesis, cancer, and tumor) rather than the target disease. The non-highly related article is ranked at top-1 simply because it has a relationship with the target article: they all have a focus on cancer (recall that they respectively focus on neuroblastoma and melanoma, which are two types of cancer), and some DTerms (cancer and tumor) of the non-highly related article have indicated the relationship.

Therefore, even the second highly related article is miss-ranked at a lower position, its DTerms can indicate why it is related to the target article (i.e., correlation between erythropoietin and anemia, which are targets of Topic 45). Similarly, even the non-highly related article is miss-ranked at the top, its DTerms have indicated why it is related to the

[17] The title of PMC1890992 is “Erythropoietin/erythropoietin receptor system is involved in angiogenesis in human neuroblastoma”.

[18] Epoetin alfa is human erythropoietin produced in cell culture.

Table 2 DTerms of three articles for Topic 45 (<anemia, erythropoietin>): The two highly related articles have several DTerms related to the target disease or the target gene of Topic 45, without any DTerms related to other diseases or genes. The DTerms can thus indicate why the highly related articles are related to the target article of Topic 45

DTerm	1st highly-related article	2nd highly related article	Top non-highly related article	Related to topic 45?
[120] erythropoietin	v	v	v	Yes
[119] receptor	v		v	Yes
[118] activation	v			?
[117] cell	v		v	?
[116] signal	v			?
[115] tyrosine	v			?
[114] stem-cell-factor	v			?
[113] interaction	v			?
[112] stat5	v			?
[111] erythroid	v			Yes
[110] patient		v		?
[109] anemia		v		Yes
[108] human		v	v	?
[107] recombinant		v		?
[106] disease		v		?
[105] chronic		v		?
[104] treatment		v		?
[103] epoetin		v		Yes
[102] alfa		v		Yes
[101] expression			v	?
[100] angiogenesis			v	No
[99] breast			v	?
[98] cancer			v	No
[97] functional			v	?
[96] tumor			v	No

target article as well (i.e., correlation between erythropoietin and cancer, which is not the target disease of Topic 45).

The answers to question Q1 and question Q2 should thus be 'yes'. Highly related articles and non-highly related articles of a target article a_T tend to have different DTerms, and DTerms of the highly related articles tend to be related to the main topic of a_T. As a_T and its highly related articles are specifically selected for the topic by domain experts, the positive answers to Q1 and Q2 indicate that DTerms of the highly related articles can be used to describe why they are related to a_T. The results thus confirm our expectations (recall Fig. 1): (1) a scholarly article a_x may discuss several issues, and for each issue, cites appropriate references; and (2) given another article a_y, terms in the titles of the references cited by a_x and a_y may be used to estimate *how* and describe *why* a_y is related to a_x. The results are also consistent with a previous finding: titles of out-link references in a scholarly article a can be used to index the main contents of a (Qin 2000; Garfield 1990). As the

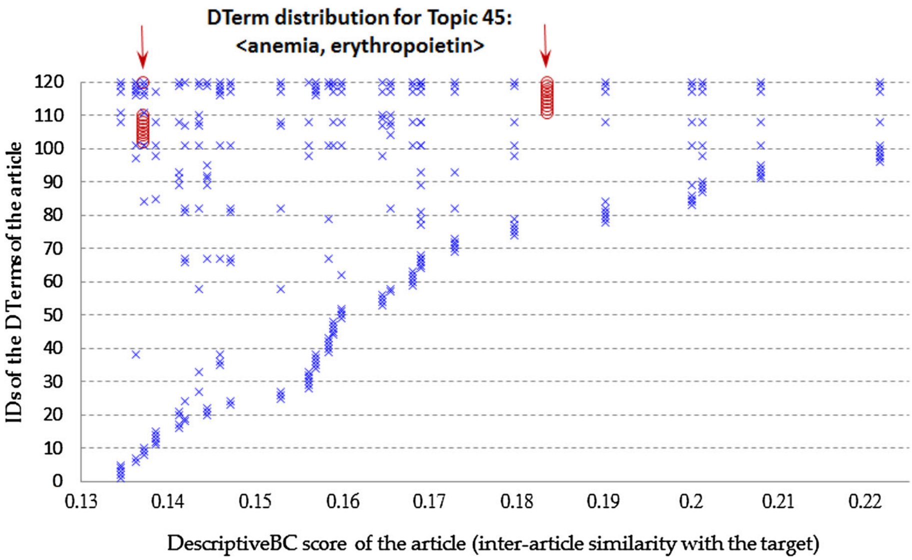


Fig. 6 A case study for which highly related articles share only one DTerm (denoted by *circle* and pointed by *down arrow*): Although only one DTerm is shared, the DTerms (see Table 2) of the highly related articles are related to the topic of the case study (association between anemia and erythropoietin), and so can indicate *why* the articles are highly related to the target article, which focuses on the topic as well. Many non-highly related articles get high similarity with the target article, however they focus on other topics with many DTerms (denoted by *times*) not related to the association between anemia and erythropoietin, indicating that they are similar to the target article in *different* ways

main contents often include several parts related to different articles, DescriptiveBC provides a novel way to further identify the aspect of relatedness between *a* and the articles.

Discussion

Application and suggestion

We have shown that DescriptiveBC is a new bibliographic coupling measure that (1) performs significantly better than the original BC in identifying related scholarly articles, and (2) provides descriptive terms to indicate why a scholarly article is related to another article. Given a scholarly article, DescriptiveBC can be invoked to support the retrieval and clustering of the related evidence already published in scientific literature. As DescriptiveBC estimates inter-article similarity and selects descriptive terms, it would be a good way to employ a two-dimensional map to visualize both *how* (based on the numerical similarity) and *why* (based on the descriptive terms) a scholarly article is related to a given target article (e.g., the maps in Figs. 5, 6). With the two-dimensional map, researchers can explore and navigate through the space of articles that are related to the target article in *different* ways.

Previous studies have found that titles of out-link references in scholarly articles can be used to index the main contents of the articles (Qin 2000; Garfield 1990). The performance of DescriptiveBC further shows that the out-link reference titles are helpful for the visualization and navigation of the relatedness between articles. Therefore, to facilitate dissemination of research findings in scientific literature, we suggest that publishers of scholarly articles should provide the titles of the references cited by the articles. Release of the reference titles can help prospective researchers to find the articles, promoting both the visibility and the impact of the articles published by the publishers.

We also suggest that the idea of DescriptiveBC should be incorporated to search engines of scholarly articles (e.g., PubMed, and Google Scholar). The search engines have been essential portals for researchers to find scholarly articles. With the idea of DescriptiveBC, the search engines can enhance their article ranking capability and provide textual descriptions about the relatedness between articles. The search engines routinely collect and preprocess a huge amount of scholarly articles for subsequent retrieval. The inter-article similarity measure of DescriptiveBC can be incorporated into the similarity measure of the search engines so that the measure can be enhanced with bibliography information. The DTerms suggested by DescriptiveBC can be cached so that researchers can be guided online with brief textual description for each inter-article relationship.

Future work

It is interesting to further improve the strategy for visualizing the inter-article similarity and the descriptive terms returned by DescriptiveBC. The two-dimensional maps shown in Figs. 5 and 6 can be a preliminary design, however two issues deserve investigation. The first issue is the interactive visualization of the DTerms. In addition to properly showing the DTerms on the map, researchers may also have a query (composed of certain DTerms) to re-rank the articles so that those articles having certain DTerms of interest can be identified in an interactive way. The second interesting issue is the clustering of the articles based on the DTerms. A cluster of the articles may correspond to a *way* the articles are related to the target article, and hence may be helpful for the researchers to easily identify what they really want. The enhanced visualization strategy can facilitate the exploration of research findings online.

Another interesting future work is to investigate the possible improvement of DescriptiveBC by employing domain-dependent thesauri of synonyms and related terms. DescriptiveBC currently works on domain-independent bibliographic information, without considering the relatedness among terms. We expect that the inter-article similarity measure of DescriptiveBC may be improved by transforming and incorporating the term relatedness.

It is also interesting to integrate DescriptiveBC with other inter-article similarity measures. DescriptiveBC is an improved version of BC that works on out-link references, which are more publicly available than other kinds of information such as full text and in-link citations of articles. Although in the experiment DescriptiveBC performs significantly better than a citation-based measure (i.e., BC), a text-based measure (i.e., BM25), and a hybrid measure (i.e., HybridK50), we are still interested in the *fusion* of DescriptiveBC and other measures that work on textual information from titles and abstracts of articles, which are often publicly available as well. The fusion is motivated by the expectation that titles and abstracts of articles can provide other kinds of information that is different from the bibliographic information employed by DescriptiveBC. An effective fusion strategy is an interesting target for future research.

Conclusion

Bibliographic coupling is an effective inter-article similarity measure that works on out-link references, which are more commonly available than full text and in-link citations employed by other citation-based measures. We present a novel measure DescriptiveBC that is an enhanced version of the bibliographic coupling measure. DescriptiveBC is developed based on the expectations that titles of out-link references may be used to refine inter-article similarity estimation, as well as provide descriptive terms to indicate the aspect of relatedness between two articles. The expectations are justified in the identification and navigation of highly related articles that have been judged (by domain experts) to be focusing on the same research topics.

Given a target article a_T, DescriptiveBC can thus provide more accurate information about *how* (based on numerical inter-article similarity) and *why* (based on textual descriptive terms) a scholarly article is related to a_T. Visualization of the information can support the identification, clustering, mapping, and navigation of the related evidence already published in scientific literature. Release of the reference titles in each article is thus helpful for the dissemination of research findings in scientific literature, and DescriptiveBC can be incorporated into search engines of scholarly articles to help prospective researchers to navigate through the space of related articles online.

Acknowledgements This research was supported by Ministry of Science and Technology, Taiwan (Grant ID: MOST 104-2221-E-320-005).

References

Aljaber, B., Stokes, N., Bailey, J., & Pei, J. (2010). Document clustering of scientific texts using citation contexts. *Information Retrieval, 13*(2), 101–131.

Becker, K. G., Barnes, K. C., Bright, T. J., & Wang, S. A. (2004). The genetic association database. *Nature Genetics, 36*(5), 431–432.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389–2404.

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., et al. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE, 6*(3), e18029.

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology, 64*(9), 1759–1767.

Calado, P., Cristo, M., Moura, E., Ziviani, N., Ribeiro-Neto, B., & Goncalves, M. A. (2003). 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, Louisiana, USA.

Couto, T., Cristo, M., Gonçalves, M. A., Calado, P., Nivio Ziviani, N., Moura, E., et al. (2006). A comparative study of citations and links in document classification. In *Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries* (pp. 75–84).

Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology, 59*(1), 51–62.

Garfield, E. (1990). KeyWords Plus: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on current contents on diskette. *Current Contents, 32*, 3–7.

Gipp, B., & Beel, J. (2009). 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* (pp. 571–575), Brazil.

Gipp, B., & Meuschke, N. (2011). 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.

Glenisson, P., Glanzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing and Management, 41*, 1548–1572.

Janssens, F., Glänzel, W., & De Moor, B. (2008). A hybrid mapping of information science. *Scientometrics, 75*(3), 607–631.

Janssens, F., Zhang, L., De Moor, B., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing and Management, 45*, 683–702.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10–25.

Kumar, S., Reddy, P. K., Reddy, V. B., & Singh, A. (2011). Similarity analysis of legal judgments. In *Proceedings of the fourth annual ACM Bangalore conference* (*COMPUTE 2011*), Bangalore, Karnataka, India.

Landauer, T. K., Laham, D., & Derr, M. (2004). From paragraph to graph: Latent semantic analysis for information visualization. *Proceedings of the National Academy of Sciences of the USA, 101*(Suppl 1), 5214–5219.

Liu, R.-L. (2015). Passage-based bibliographic coupling: An inter-article similarity measure for biomedical articles. *PLoS ONE, 10*(10), e0139245.

Liu, S., Chen, C., Ding, K., Wang, B., Xu, K., & Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics, 101*(2), 1293–1307.

Liu, R.-L., & Huang, Y.-C. (2011). Ranker enhancement for proximity-based ranking of biomedical texts. *Journal of the American Society for Information Science and Technology, 62*(12), 2479–2495.

Liu, X., Zhang, J., & Guo, C. (2013). Full-text citation analysis: A new method to enhance scholarly networks. *Journal of the American Society for Information Science and Technology, 64*(9), 1852–1863.

Nakov, P. I., Schwartz, A. S., & Hearst, M. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR'04 workshop on search and discovery in bioinformatics* (pp. 81–88).

Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: an investigation in the antibiotic resistance literature. *Journal of the American Society for Information Science., 51*(3), 166–180.

Ritchie, A., Teufel, S., & Robertson, S. (2008). Using terms from citations for IR: Some first results. In C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, & R. White (Eds.), *Advances in information retrieval* (Vol. 4956, pp. 211–221). Berlin: Springer.

Robertson, S. E., Walker, S., & Beaulieu, M. (1998). Okapi at TREC-7: automatic ad hoc, filtering, VLC and interactive. In *Proceedings of the 7th text retrieval conference* (TREC 7) (pp. 253–264). Gaithersburg, USA.

Salton, G., & Zhang, Y. (1986). Enhancement of text representations using related document titles. *Information Processing and Management, 22*(5), 385–394.

Small, H. G. (1973). Co-citation in the scientific literature: A new measure of relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Small, H. (2011). Interpreting maps of science using citation context sentiments: A preliminary investigation. *Scientometrics, 87*(2), 373–388.

Thijs, B., Zhang, L., & Glänzel, W. (2015). Bibliographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes. *Scientometrics, 105*(3), 1453–1467.

van Eck, N. J., Waltman, L., Noyons, E. C., & Buter, R. K. (2010). Automatic term identification for bibliometric mapping. *Scientometrics, 82*(3), 581–596.

Whissell, J. S., & Clarke, C. L. A. (2013). effective measures for inter-document similarity. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (CIKM'13) (pp. 1361–1370).

Wiegers, T. C., Davis, A. P., Cohen, K. B., Hirschman, L., & Mattingly, C. J. (2009). Text mining and manual curation of chemical-gene-disease networks for the comparative toxicogenomics database (CTD). *BMC Bioinformatics, 10*, 326.