國立臺灣大學文學院圖書資訊學系
博士論文

Department of Library and Information Science
College of Liberal Arts
National Taiwan University
Doctoral Dissertation

應用語意分析於衡量文獻引用關係之探討
A Study on Applying Semantic Analysis in Measuring Citation Relationships

蕭宗銘
Tsung-Ming Hsiao

指導教授：陳光華博士
Advisor: Kuang-Hua Chen, Ph.D.

中華民國 111 年 7 月
July 2022

doi:10.6342/NTU202201326

國立臺灣大學博士學位論文
口試委員會審定書

應用語意分析於衡量文獻引用關係之探討

A Study on Applying Semantic Analysis in Measuring Citation Relationships

本論文係蕭宗銘君（學號 F03126010）在國立臺灣大學圖書資訊學研究所完成之博士學位論文，於民國一一一年六月二日承下列考試委員審查通過及口試及格，特此證明

指導教授：　陳光華 陳光華

口試委員：

唐牧群 唐牧群

羅思嘉 羅思嘉

曾元顯 曾元顯

林雯瑤 林雯瑤

系主任、所長 林奇秀 （簽名）

doi:10.6342/NTU202201326

致謝

在整個完成學位攻讀與論文研究的過程中，得到協助甚多。若非如此，相信這個過程會更為艱難，甚至無法完成。謹感謝指導教授陳光華老師，除了在研究上給予全力支持，過程中所分享的諸多寶貴經驗與提點，讓我獲益良多。亦感謝林雯瑤老師、唐牧群老師、曾元顯老師、羅思嘉老師，於計劃書到論文審查的過程中，所悉心給予的指正與建議。本論文受惠於諸位老師的指導，得以更為完善。論文若有任何錯漏，則為本人之過。除此之外，感謝所有系上老師，在學術專業上的教導，與日常生活中的關心。其中，特別感謝楊東謀老師與藍文欽老師，總是不吝撥空瞭解我的近況並給予支持，深表感謝。

同儕與朋友的支持，是另一個讓我得以完成這個學位的重要力量。難以盡列有多少友人，陪我走過這段時間中的各個時期，謹此感謝。其中，要特別感謝好友依柔，從碩班以來持續不斷的打氣與鼓勵，排解了諸多研究生活中的壓力。感謝映涵在計劃書與論文口試中，所盡心提供的協助。以及有容、惟中，在各次聚會中的經驗交流。此外，感謝中哲、欣嵐、思嘉、思穎、弼心、舒軒、靚柔，即使已是忙碌的社會人，總會給予關心與問候。

最後，感謝家人在這段期間的包容與支持。特別感謝姊姊微霓的大力協助，特別是在首次出國時的諸多安排，讓我得以專注於會議準備之上，無須分心他顧。感謝父親蕭鋒山、母親邱春月與春花阿姨，總是持續給予關心與鼓勵，並盡心竭力給予各種支援。

謹將此論文獻與各位。

宗銘　謹識

2022 年 7 月

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

摘要

本研究分析三種被廣泛應用的引用關係，包含直接引用、書目耦合、共被引。以瞭解於不同模型衡量引用關係後，其所得分析結果之異同。研究分析的模型，除實務應用之經典模型與本研究設計之兩種語意模型外，另納入相關研究提出之頻率模型、距離模型、辭彙模型，總計六種。語意模型的部分，本研究使用基於 Wordnet 與 BERT 設計之自然語言處理開源工具，以 Awais（2011）資料集進行訓練後，判斷引用句的情感傾向與語意相似度。一方面，經由判斷引用句的情感傾向，分類直接引用關係；另一方面，則衡量引用句間的語意相似度，修正書目耦合與共被引的關係強度。頻率模型部分，則依文內引用頻率，調整直接引用、書目耦合、共被引之引用強度。辭彙模型上，則依引用句所用辭彙的相似程度，調整書目耦合、共被引之引用強度。距離模型則依文內引用的對應位置，調整共被引強度。

對於各模型衡量之結果，則比較其網路結構、分群結果、關鍵節點與強關係之情形，以確認在引用分析結果上的表現情形。本研究將各模型形成之引用網路，區別為整體網路與核心網路。對這兩類網路，比較各模型的節點數、關係數、網路密度、連結元件數（number of connected components）、傳導性（transitivity）、平均群聚係數（average clustering coefficience）上的差異。對分群結果的比較上，本研究以 Modularity 分群演算法對各模型核心網路之節點進行分群。於初步檢視分群數、孤立節點（singleton）數、群規模後，再以 Adjusted Rand Index 確認分群結果間的相似程度。接著，則以文字群聚度（textual coherence），量化衡量分群結果的表現。並以各群文獻標題中之高頻詞彙，確認各群的主題後，比較各模型間主題分析結果。最後，於節點與關係的部分，則檢視各模型中，來源文章在直接引用網路中的被引用的傾向與次數，以及書目耦合、共被引網路中的強關係書目組。經由前述方式分析不同模型在衡量各類引用關係的表現，本研究對在目前引用分析中，應用語意分析技術之優缺，加以綜整分析。

基於上述研究設計，本研究選定圖書資訊學領域之十五種期刊，以其中所刊登之 10,088 篇文章做為研究對象。由網路層級的分析結果來看，在直接引用關

doi:10.6342/NTU202201326

係上，判斷情感傾向並移除負面引用之後，對於整體網路與核心網路的結構影響並不明顯。而在書目耦合網路中，對關係強度進行調整後，核心網路的結構上有較大差異，但在整體網路結構上則無明顯變化。於共被引網路時，則不論在整體網路或核心網路上，各網路指標均指出有明顯差異。

由核心網路分群結果的相似程度來看，直接引用的部分，僅有經典模型的結果明顯不同，而頻率模型、Wordnet 模型、BERT 模型三者的分群結果則十分相似。書目耦合的部分，各模型的結果略有差距，但除了詞彙模型的較為明顯，其它模型間的差距並不明顯。在共被引的部分，各指標則指出，多數模型相互存在明顯差異。而文字群聚度、主題分析結果則顯示，語意模型應用在共被引時，文字群聚度較高，則具發掘研究領域新議題的能力。但當應用於直接引用、書目耦合時，除了沒有明顯改善文字群聚度外，主題分析的結果亦十分類似。

在節點與關係層次上，當來源文獻有被正面引用過時，其被直接引用數更可能高於未被正面引用過的文獻。此一傾向，在多個語意模型均判定此來源文獻有被正面引用或考慮進累積引用所需時間之後，會更為明顯。但在書目耦合與共被引關係的部分，則未觀察到使用語意分析的模型會提供更為優秀的表現。

綜觀而言，目前設計之語意分析模型的影響，依引用關係類型、分析層次的差異，有著不同影響。以網路層次而言，排除負面引用對於網路結果的影響甚微，這可能代表目前語意分析模型在負面引用偵測上仍力有未逮，或負面引用影響不如先前學者預期的明顯。而於書目耦合、共被引上，則對於核心網路結構均產生明顯影響。分群結果的比較，則顯示目前語意分析模型僅應用於共被引時有得到較明顯的改善。除了在文字群聚度上有較佳表現外，主題分析的結果也較能反映出領域變動情形。但應用於直接引用、書目耦合上時，則未有明顯改善。而由節點與關係層次的分析來看，應用語意分析模型區別引用句的情感傾向，有助於判斷被引用文獻的影響力。但使用語意相似度修正書目耦合與共被引時，則未觀察有進一步的改善。

關鍵詞：語意分析；引用分析；直接引用；書目耦合；共被引

doi:10.6342/NTU202201326

Abstract

The present study investigates three kinds of citation relationships, including direct citation (DC), bibliographic coupling (BC), and co-citation (CC), to understand the effects of considering semantic meanings when conducting citation analysis. Six models were included in this study. The classical model is the general way to implement citation analysis. The frequency model adjusts the strength of DC, BC, and CC by the number of citations. The lexical model revises the BC and CC strength based on the lexical similarity of citances. The distance model weights CC strength by considering the relative locations between citations. Another two models, Wordnet and BERT models, are based on the open-source tools and trained by the corpus provided by Awais (2011) to decide the citations' sentimental polarity and measure the semantic similarity between two citations. The sentimental polarity and semantic similarity were used to classify DC and weight BC/CC, respectively.

To evaluate these models, the present study compares their results at three levels: network, cluster, and node/relationship. At the network level, six indicators were used, including number of nodes, number of edges, network density, number of connected components, transitivity, and average clustering coefficient. At the cluster level, the clusters resulting from the clustering algorithm based on modularity were first examined by number of clusters, number of singletons, and cluster size. Then, Adjusted Rand Index was used to measure the similarity between the clustering results. This study further evaluated the quality of clustering results based on textual coherence and subject analysis. At node/relationship level, this research examined the correlation between a reference's sentimental types and its DC counts. Whether the citation strength will be higher if two works' topics are highly similar was also investigated.

The present study chose the 10,088 articles published in the fifteen journals of Library and Information Science (LIS) as the research subjects. The examination of network level showed that removing negative citations does not significantly affect the DC citation network. As to BC/CC citation network, weighting strength by the

doi:10.6342/NTU202201326

semantic meaning reveals different whole networks, especially the core networks.

Comparing the clustering results of DC core networks indicated that the results of the frequency, Wordnet, and BERT models were highly similar. Only that of the classical model shows a different pattern. As to the BC core networks, no noticeable differences existed between the results of these models except the lexical model. Examining the clustering results of CC core networks revealed the existence of evident divergence. Textual coherence and subject analysis supports that the clustering results of CC core network based on the Wordnet/BERT models have higher textual coherence. The subjects identified from the clustering results of the two models better reflected the development of LIS in this period.

The examination at node/relationship level revealed that the DC is probably higher if the source article has been cited positively. The tendency will be more evident when using multiple semantic models or considering the time effects. However, applying semantic models in weighting BC and CC did not improve their results.

Overall, the effect of the semantic models proposed in this study varies by the type of citation relationship and at which level researchers analyze the result. At the network level, removing negative citations affects slightly. It shows that the current semantic tools may have difficult in identifying negative citations or that the effects of negative citations are not as critical as the arguments of the previous studies. As to BC/CC, however, applying semantic models does significantly affect. The examination at the cluster level indicates that applying semantic models in CC improves its textual coherence and better reflects the evolution in the domain. Yet, no similar effect is found when using semantic models in DC and BC. Additionally, classifying citations by their sentimental polarity helps identify the influence of the cited works. At the node/relationship level, however, adjusting BC and CC based on the semantic similarity may not improve the result.

Keywords: Semantic Analysis; Citation Analysis; Direct Citation; Bibliographic Coupling; Co-Citation

doi:10.6342/NTU202201326

Contents

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

List of Figures

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

List of Tables

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

List of Abbreviations

ABCA	author bibliographic coupling analysis
ACA	author cocitation analysis
ACL	the Association for Computational Linguistics
AoT	applications of technology
ARI	Adjusted Rand Index
BC	bibliographic coupling
BCS	bibliographic coupling strength
CBOW	continuous bag-of-words model
CC	co-citation
CCS	cocitation strength
CPACA	content- and proximity-based author co-citation analysis
CS	computer science techniques
DC	direct citation
DCC	distance cocitation
FBC	frequency bibliographic coupling
FCC	frequency cocitation
FDC	frequency direct citation
HIT	health information and technology
IBIR	information behavior and information retrieval
ITC	in-text citation
JSD	Jensen–Shannon divergence
KCA	Keyword cocitation analysis
LBC	lexical bibliographic coupling
LCC	lexical cocitation
LIS	library and information science
LS	library services and management
NLP	natural language processing
NLTK	Natural Language Toolkit
NSI	Normalized Similarity Index
POS	part-of-speech tagging
RI	Rand Index

doi:10.6342/NTU202201326

SCI	Science Citation Index
SCS	scholarly communication and scientometrics
SG	skip-gram model
SIA	Sentiment Intensity Analyzer
SSCI	Social Science Citation Index
SVM	support vector machines
SVR	ε-support vector regression
WBC	Word Bibliographic Coupling
WDC	weighted direct citations
WoS	Web of Science

doi:10.6342/NTU202201326

Chapter 1

Introduction

The development of library and information science (LIS) aims at improving the efficiency of utilizing information. Measuring and deciding relationships between works, authors, and subjects lays the foundations for this task. How to gauge and organize them affects the efficiency of accessing information significantly. For improving the efficiency, people created various tools like thesauruses, classification schemata, and indexes. In the last half of the 20th century, Garfield proposed a new index tool, citation index database, as another approach to measure relationships between numerous resources and organize them. Similarly, how to measure the citation relationship accurately becomes a critical issue for the studies regarding citation index databases.

Since the creation of citation databases and the emergence of citation analysis, measuring and differentiating citation relationships has been one critical research issue in this field. Although authors cite works for numerous reasons, scholars usually treat all citations equally while applying citation analysis. The lack of discriminating

doi:10.6342/NTU202201326

different citation relationships causes doubts and debates about the effectiveness of citation analysis (Cole & Cole, 1971, 1972; Gilbert, 1977; MacRoberts & MacRoberts, 1986; 1987; 1988; 1996; 2018). Scholars had started to investigate this research issue since the beginning, but the early studies' scale was relatively small because the human experts played the main roles in analyzing these citations (Chubin & Moitra, 1975; Frost, 1979; Moravcsik & Murugesan, 1975; Murugesan & Moravcsik, 1978; Oppenheim & Renn, 1978; Speigel-Rosing, 1977). It is extremely difficult, if not impossible, to differentiate citation relationships with human experts on a large scale. Hence, as indicated by Smith (1981, p. 89), the de facto practice of general citation analysis studies is "all citations are equal."

Since the 1990s, thanks to the decreasing cost of computing power and the easier access of machine-readable full-texts, utilizing works' metadata and full-texts on a large scale has become possible. Citation analysis studies benefit from this trend in several ways. Firstly, researchers can explore various citation features by analyzing enormous data. These features include the number of references, distribution of in-text citations (ITCs), and words used in citances, which are the sentences authors cite other works (Boyack, van Eck, et al., 2018; Hsiao & Chen, 2018). These studies reveal more details about how authors cite other works. Secondly, researchers utilize additional metadata or elaborate methods to analyze citation relationships (Ahlgren et al., 2020; Bu et al., 2018; Liu, 2017; Liu & Hsu, 2018; Waltman et al., 2019). The modifications overcome some weaknesses of classical citation analysis methods and improve related applications, like information retrieval and mapping the scientific structure. Thirdly, researchers study how to differentiate citation relationships with data extracted from text body, including article structure or language parameter (Boyack, Small, et al.,

doi:10.6342/NTU202201326

2013; Elkiss et al., 2008; Eto, 2008; Hsiao & Chen, 2017; Jeong et al., 2014; Kim et al., 2016; Yaghtin et al., 2019). The aforementioned studies showed that including additional data could emphasize critical relationships and nodes. In addition, these data also help scholars explain the results of citation analysis more accurately. Overall, the advances in research methodology and data collection demonstrate more possibilities and enhance the related applications while conducting citation analysis. By discriminating different citation relationships and their usages, the researchers can further analyze authors' citation behavior and explain the results of citation analysis more appropriately.

Most previous studies rely on metadata or language parameters, e.g., location of ITCs or lexical similarity between different citances. However, authors' semantic meaning when citing works play the crucial role in deciding the meaning of citations. Given the rapid development of natural language processing (NLP) techniques in recent years, e.g., Devlin et al. (2018) and Mikolov, Chen et al. (2013), and their achievements in various NLP tasks, measuring citation relationships by applying NLP techniques is promising. Yet, the potential of applying semantic analysis in differentiating citation relationships remains unknown, especially when measuring bibliographic coupling strength (BCS) and co-citation strength (CCS). Applying these techniques in analyzing citation relationships may improve the applications of citation analysis and the development of LIS in assisting users accessing information.

Overall, an investigation of using semantic analysis may improve the development of citation analysis, especially exploring how to measure and differentiate citation relationships. It helps scholars better utilize citation analysis in different applications of citation analysis and investigate authors' citation behaviors. The present study aims to

doi:10.6342/NTU202201326

achieve the above purposes by comparing the results of several methods in gauging by three types of citation relationships, namely direct citation (DC), bibliographic coupling (BC), and co-citation (CC). The results of this study can reveal the advantages and weaknesses of conducting citation analysis with NLP semantic analysis and improve the related applications of citation analysis. The result of this study also benefits the development of LIS which aims at helping users manage and utilize information.

1.1 Citation Analysis

1.1.1 Citation relationship and citation entities

The citation analysis is generally based on three kinds of citation relationships: DC, BC, and CC. DC is "a relationship between a part or the whole of the cited document and a part or the whole of citing document" (Smith, 1981, p.83). Based on DC, Garfield (1955) proposed using the citation index as a new subject control tool. In the following years, citation index databases, such as Science Citation Index (SCI) and Social Science Citation Index (SSCI), give the momentum to utilize DC in developing various citation indicators and fulfills other needs of using citation data. Another two types, BC and CC, were proposed in 1963 and 1973, respectively (Kessler, 1963a, 1963b; Marshakove, 1973; Small, 1973). Kessler (1963a, 1963b) proposed BC as a new measurement of the relationships between documents and defined BC as the number of common references cited by two documents. Instead of measuring the relationship between two documents based on their common references, Small (1973) and Marshakove (1973) respectively proposed CC to measure the relationships between documents by counting how many times two documents are co-cited in the

doi:10.6342/NTU202201326

later publications.

The three kinds of citation relationships are widely used in building scholarly networks, and each one represents a different perspective. According to Yan and Ding (2012), DC represents real connections between nodes, whereas BC and CC are artificial connections used to measure the similarity between nodes. Furthermore, CC "is a relationship which is recognized and maintained by current researchers" and BC "is static because it depends only on citation contained in the coupled documents" (Small & Griffith, 1974, pp. 19-20). At first, they are used in measuring relationships between different publications. With the development of citation analysis, they are further applied in gauging relationships between authors (White & Griffith, 1981, 1982; Zhao & Strotmann, 2008a, 2008b) and subjects (Hsiao & Chen, 2019; Huang, Wang, et al., 2018; Moya-Anegón et al., 2004). Therefore, in addition to three kinds of citation relationships, researchers can investigate scientific structure based on three kinds of citation entities, i.e., work, author, and subject.

1.1.2 Differentiate citation relationships

Although researchers can observe the scientific structure from different perspectives, differentiating citation relationships is still a question. The question was noticed and verified in several studies during the 1960s and 1970s. Garfield (1965) and Lipetz (1965) firstly indicated that authors cite articles for numerous reasons. In the 1970s, some studies investigated this question and reported the proportion of different citation functions (Moravcsik & Murugesan, 1975), the distribution of ITCs (Voos & Dagaev, 1976), and the importance of multiple mentioned references (Herlach, 1978). In addition to studying the citation functions and analyzing the ITCs, scholars also

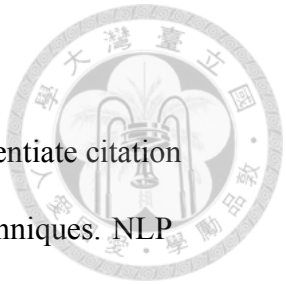
doi:10.6342/NTU202201326

investigated the citances' text. Researchers argued that the text could be the symbols representing the concepts of cited works (Small, 1978). Some studies applied similar propositions, e.g., O'Connor (1982) utilized terms extracted from citances and improved the recall ratio of information retrieval tasks. Briefly, in the early studies of citation analysis, researchers had noticed and argued that data extracted from text body could be used to differentiate citation relationships.

Despite the results of these studies mentioned above, the assumption that all citations are equal is still widely accepted in practice (Smith, 1981). As mentioned above, the primary reason is that the differentiating process largely relies on human experts and hard to expand the scale. However, with the ease of accessing full-texts and computing power, more and more studies proposed several methods in differentiating citation relationships. Prior studies utilized various indicators and citances to identify the citation functions or weight DC (Giuffrida et al., 2019; Nakov et al., 2004; Small et al., 2017). As to BC and CC, numerous methods are proposed to measure BCS or CCS. These methods can be categorized into several classes, which are distance model (Boyack et al., 2013; Callahan et al., 2010; Eto, 2007, 2008; Gipp & Beel, 2009; Liu & Chen, 2011a, 2011b), lexical model (Hsiao & Chen, 2017; Jeong et al., 2014; Liu, 2017; Liu & Hsu, 2018), and hybrid model (Kim et al., 2016). By applying these models, researchers can calculate the similarity between different citation entities more accurately, measure citation relationships in finer granularity, and provide more details about the structure of science.

doi:10.6342/NTU202201326

1.2 Semantic Analysis



Although many studies have investigated how to measure and differentiate citation relationships, most studies do not utilize NLP semantic analytics techniques. NLP concerns how to use computers to process and analyze natural language. The human language comprises a colossal number of words to represent kaleidoscopic meanings. NLP researchers propose various approaches to identify the meaning of words. These approaches can be categorized into three models: thesaurus model, count model, and predict model (Baroni et al., 2014; Saitoh, 2019).

The significant differences between these models are the approaches used to identify the words' relationships. Thesaurus model keeps the words' relationships in dictionary form. A typical example is Wordnet. The relationships, usually annotated by human experts, are used to define the meanings of words. The major limitation of thesaurus model is the expensive cost due to heavily relying on human annotation. It also restricts the efficiency of adding new words or adjusting the words' meanings (Saitoh, 2019). The count model extracts the words' meanings by analyzing the contextual representations of words, e.g., co-occurring words. According to Miller and Charles (1991), there is an inverse linear relationship between the similarity of words' meaning and the discriminability of context. They argued that the linguistic context extracted from large corpora, composed of enormous text data, can represent the words' meaning. In addition, the linguistic context, e.g., raw co-occurrence counts, could be represented as a word vector, the distributed representation of a word (Baroni et al., 2014). Although the count model reduces the difficulties in extracting meanings and measuring the relationships between words, the time complexity and space

doi:10.6342/NTU202201326

complexity still bother researchers when counting a huge corpus (Saitoh, 2019).

The prediction model also assumes that the contextual representations of words can reveal the words' meanings. Additionally, "the vector weights are directly set to optimally predict the contexts in which the corresponding words tend to appear" (Baroni et al., 2014, p.238). Firstly, Bengio et al. (2003) proposed the neural probabilistic language model, a neural network architecture used to learn the distributed representation of words. Years later, Mikolov, Chen et al. (2013) released an open-source project, word2vec, and proposed two different learning architectures: continuous bag-of-words model (CBOW) and continuous skip-gram model (SG). Compared with the model proposed by Bengio et al. (2003), CBOW and SG significantly reduce the computing complexity and outperform prior models at extracting semantic and syntactic meaning. Mikolov, Sutskever et al. (2013) reported more details about improving the efficiency of learning the distributed representation. Recent developments, e.g., Attention and BERT, could extract more context-sensitive features and show better performance in different NLP tasks (Devlin et al., 2018; Vaswani et al., 2017). In addition, the fine-tuning approach is another noteworthy trend. Researchers could benefit from the pre-trained language models and refine relatively few parameters for their interesting downstream tasks with appropriate datasets.

Overall, the early studies had shown that authors do not cite works equally. However, due to the factors like computing power limitation, counting all citations equal is a practice standard when conducting citation analysis. This de facto standard causes the continuing debate on the usefulness of citation analysis. Recently, scholars studied how to differentiate citation relationships and report their improvements in various applications. Given the findings of these studies, it is reasonable to assume that

doi:10.6342/NTU202201326

discriminating citations will improve citation analysis and enhance their applications. Additionally, with the advancement of NLP studies, the cutting-edge techniques of NLP semantic analysis have shown their ability in numerous NLP tasks. Together, whether applying NLP semantic analysis can benefit citation analysis is worthy of attention. It may provide more insights by appropriately categorizing citation entities and measuring citation relationships based on semantic meanings. Therefore, the present study investigates whether using NLP semantic analysis techniques in citation analysis can improve our ability to observe and analyze the scientific structure. The results can help us understand the effectiveness of utilizing NLP semantic analysis in citation analysis and make us one step closer to properly weighting different citation relationships instead of counting all citations equally.

1.3 Research Questions

The primary purpose of this study is to investigate the effectiveness of using NLP semantic analytics in measuring all three kinds of citation relationships and explore how this modification affects the result of citation analysis. Several models proposed by the previous studies, i.e., the classical model, the frequency model, the distance model, and the lexical model, are also considered to evaluate whether using NLP semantic analytics improves citation analysis more obviously. In addition, the semantic model is presented in this study by introducing NLP semantic analysis techniques to gauge citation relationships.

Nowadays, citation analysis is widely used to sketch the overview, research topics, and important entities of a discipline. Usually, the researchers use the citation network

doi:10.6342/NTU202201326

as a general summary for a domain, the large groups as the subfields in this domain, and the noticeable entities as the influential entities. In other words, researchers usually use the result of citation analysis from three perspectives, namely networks, clusters, and nodes. Therefore, the present study compared the result of the models mentioned above from three dimensions and explored whether the semantic model, compared with the classical model and the models proposed by other studies, can improve the outcome of citation analysis.

Three dimensions bring three research questions of this study, and these questions are detailed as follows:

1. **Do these models uncover different citation networks?**

   The network structure can be taken as the whole picture shown by a model. If the semantic model can sketch a different scientific structure landscape, its network structure should differ from those based on other models. Previous studies argued that the lexical and distance models revealed the network structures different from the classical model when measuring CC (Jeong et al., 2014; Liu & Chen, 2011a, 2011b). Hence, a reasonable conjecture is that semantic model also uncovers a different network. The present study compares several network indicators and the distributions of node degree.

2. **Does the semantic model uncover the sub-fields that differ from other models? If yes, which model is of a better ability in provides more relevant results?**

   After examining the network structures, the present study further scrutinizes the clustering results in the citation network. Generally, clusters revealed by

doi:10.6342/NTU202201326

clustering algorithms in a citation network are taken as the research subfields. The difference between the clustering results of the models represents how different the research subfields revealed by these models are. The present study conjectures that the research subfields revealed by the semantic model will differ from other models. In addition to verifying this conjecture, another focus is on which model, primarily the semantic model, can better identify research subfields.

3. **Does the semantic model identify the influential entities or relevant relationships that other models cannot?**
   The present study analyzes the difference at network and cluster levels by answering the previous two questions. The remaining questions are whether the influential entities or relevant relationships identified by these models are different and whether any model can reveal influential entities or relevant relationships better.

Investigating the questions above can answer whether the semantic model provides a different perspective to analyze the scientific structure. The present study discusses the possible reasons why the semantic approach offers a different perspective. Overall, this research studies the effectiveness of applying NLP semantic techniques in citation analysis, helps researchers find ways to refine the outcomes of citation analysis, and develops different applications in the future.

1.4 Definition of Terminologies

1. Citance

doi:10.6342/NTU202201326

Nakov et al. (2004) coined citances "to mean the sentence(s) surrounding the citation within a document" (p. 2). This study uses the citance to represent the sentence in which authors cite one or multiple citations.

2. Citation

   In this research, citations refer to the behaviors that authors quote, rephrase, or mention other works in the text body.

3. Citation Entity

   The citation entity represents the target analyzed by researchers via citation analysis. The prior studies have identified three kinds of entities: work, author, and subject.

   - The work entity includes artificial creations carriers, like journal articles, books, or multimedia resources.
   - The author entity represents who owns the authorship, including people, groups, organizations, etc.
   - The subject entity represents the concepts described, discussed, considered, or studied by a work. From the micro view to the macro view, a subject may represent a research topic, a domain, or a discipline.

4. Citation Feature

   Citation features represent objective and quantitative indicators that reflect characteristics of a citation, like the location, the distribution, and the length of ITCs.

5. Citation Function

doi:10.6342/NTU202201326

Citation functions are defined as purposes, judged by readers, that authors intend to achieve by citing. It represents the readers' viewpoint of explaining citations. Usually, readers determine the citation function by examining the surrounding sentences of a citation.

6. Citation Motivation

   Similar to citation functions, citation motivations are authors' purposes when citing works. The difference between them is who decides it. Citation functions are judged by readers, and citation motivations are determined by authors. Namely, citation motivations are authors' explanations for their citing purposes.

7. Citation Selection

   This term refers to the criteria that authors choose their references among all works.

8. Classical Model

   The present study uses the term to represent the general practice of counting DC, BC, and CC.

9. Count Model

   One way to represent the meaning of a word is to count its co-occurrence frequency with all other words. In general, a corpus is used to calculate the frequency. The method is defined as the count model in this study.

10. Distance Model

    Some methods adjust the strength of citation relationships by the distance between two entities' ITCs. These methods are based on a similar assumption that the strength between two entities is inversely related to the distance between

doi:10.6342/NTU202201326

them. In this study, the distance model is used to represent these methods.

11. Frequency Model

    The frequency model assumes that the strength of DC, BC, and CC can be weighted by the frequency of reference mention in the text body instead of counting equally.

12. Hybrid Model

    Some methods mixed multiple methods from different models. For example, the method proposed by Kim et al. (2016) used the lexical and distance models. These methods are categorized as the hybrid model.

13. In-text Citation (ITC)

    In this study, in-text citation means the exact point that authors cite other's work in the text body.

14. Language Model

    The language model represents how possible a series of tokens, like words or numbers, will happen in natural language. The language model can be trained by the count or predict models based on corpora. For some training models, e.g., the recurrent neural network, the order of tokens will be considered while training. As to CBOW and SG, only the composition of tokens will be considered.

15. Language Parameter

    Language parameters mean those features of the words in a text. They include location in an article, POS tagging of a word, spellings of words, etc. Prior studies have shown that these parameters can improve the results of citation analysis (Callahan et al., 2010; Eto, 2007, 2008, 2019; Gipp & Beel, 2009; Hsiao

doi:10.6342/NTU202201326

& Chen, 2017; Jeong et al., 2014; Kim et al., 2016).

16. Lexical Model

    Some methods, e.g., Jeong et al. (2014), determine the relationship strength between two citation entities by the similarity between the words of two citances. The higher similarity between the citances, the stronger the relationship between the entities. These methods are defined as the lexical model.

17. References

    References represent the works listed in the notes or at the end of a publication by the authors.

18. Prediction Model

    Like the count model, the prediction model represents the words' meaning in a mathematical form, namely representation. The prediction model relies on a neural network and calculates word representations by examining part of the corpus iteratively. In principle, neural networks are applied to predict the probability of the events, e.g., how possible a specific word comes after a series of tokens. After completing the training processes, neural networks' parameters are used as words' representations.

19. Research Branch

    The research branch represents a specific research direction composed of the same research subfields across different periods.

20. Research Trend

    The research trend also represents a research direction that includes similar research subfields with more generalized criteria than the research branch. Hsiao

doi:10.6342/NTU202201326

and Chen (2020) identified six research trends of LIS, and the present study uses them to examine each model's abilities in showing the scientific structure.

21. Research Subfield

    After categorizing a citation network, the present study identifies the topic of a cluster based on its high-frequency words. In this study, the topic is defined as the research subfield.

22. Semantic Analysis

    In this study, semantic analysis is defined as those approaches, algorithms, and techniques developed or applied to identify the meaning of the text in computational linguistics.

23. Semantic Model

    The present study proposes the semantic model, which measures the strength of citation relationship between citation entities based on the semantic similarity of their citances. In this research, the semantic similarity is decided by NLP semantic analysis techniques, including the thesaurus and prediction models.

24. Source Articles

    The present study uses the term, source articles, to represent the articles whose HTML full-text are successfully mapped to WoS records.

25. Thesaurus Model

    The thesaurus model is another way to represent the meanings of words by connecting the relationships between similar words, e.g., synonyms, hypernyms, and hyponyms. Unlike the count and prediction models, the thesaurus model usually relied on human experts to maintain the relationships between words. It

doi:10.6342/NTU202201326

provides accurate representations of words' meanings but costs high expense.

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

Chapter 2

Literature Review

This study investigated the effectiveness of applying NLP semantic analysis in citation analysis. Researchers have proposed three types of citation relationships and used them to measure the relationships between citation entities since several decades ago (Garfield, 1955; Hsiao & Chen, 2019; Kessler, 1963a, 1963b; Marshakove, 1973; Moya-Anegón et al. 2004; Small, 1973; White & Griffith, 1981; Zhao & Strotmann, 2008a). In practice, most citation analysis studies widely accept the assumption that all citations are equal. However, early studies have indicated that authors cite works for numerous reasons (Garfield, 1965; Lipetz, 1965). Even so, the deficiency of computing power and full-texts limits the further possibility of implementing related studies on a large scale (Voos & Dagaev, 1976).

Since the late 20th century, a series of studies have tried to utilize various features extracted from citances or ITCs when conducting citation analysis (Elkiss et al., 2008; Eto, 2007, 2008, 2019; Gipp & Beel, 2009; Hsiao & Chen, 2017; Jeong et al., 2014; Kim et al., 2016; Teufel et al., 2006a, 2006b). These studies have shown the possibility

doi:10.6342/NTU202201326

of utilizing more information, extracted from citances and ITCs, in different applications. The rapid development of NLP techniques, like word2vec (Mikolov, Chen et al., 2013) and BERT (Devlin et al., 2018), enhances the machine's ability to conduct various NLP tasks and demonstrates the possibility of differentiating citations based on their semantic meaning.

This chapter reviews the studies about citation relationships and citation entities for briefing the background knowledge at first. Then, the previous studies of citation behaviors are discussed to show how scholars interpret the results of citation analysis and why further differentiating citation relationships is necessary. Two primary citation theories are reviewed to provide different viewpoints for explaining citation analysis research. The reviewed studies of citation motivations and selections also reveal the citation behavior from authors' perspectives. As a result, the discussions show the necessity of discriminating citation relationships.

Instead of authors' perspectives, researchers discover citation functions and differentiate citation relationships by utilizing the text content. Then, the present study reviews the recent propositions of weighting citation relationships with the text content and indicates the possibility of using text content in identifying different citation relationships. The final section of this chapter presents the recent development of NLP techniques for semantic analysis and shows why applying them in conducting citation analysis is promising.

doi:10.6342/NTU202201326

2.1 Citation Relationships and Citation Entities

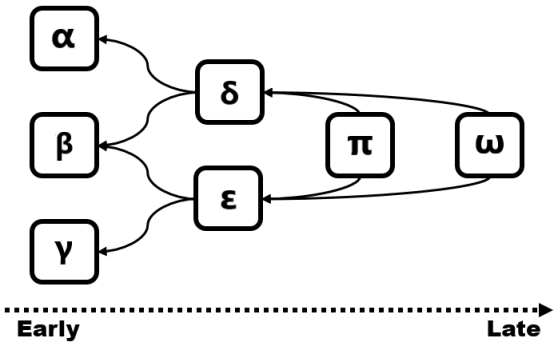
Citation relationships are the foundation of citation analysis, and scholars proposed three kinds of citation relationships, namely DC, BC, and CC. Among them, DC is the actual connection between the later work and prior work (Yan & Ding, 2012), and the other two types, BC and CC, are two works' relationships claimed by authors of the later works (Smith, 1981). In 1955, Garfield proposed a bibliographic system, namely citation index, as a tool to help scholars check bibliographic descendants of the antecedent papers (Garfield, 1955). The general subject index, limited by cost-effective concern, cannot reveal various relationships between different works. Garfield argued that the citation index could reveal unknown relationships among works and improve information communication (Garfield, 1955, 1957, 1964). Sher and Garfield (1983) proposed utilizing citation index as an evaluation tool. Several studies also suggested some indicators, e.g., impact factor, to evaluate the impact of articles and journals (Garfield, 1955; 1972). Instead of evaluating the impact of studies, Price (1965) utilized bibliographic references to investigate the scientific structure and research fronts. Other indicators, e.g., immediacy index, cited half-life, and citing half-life, can be used to analyze the life period of disciplines or publications from different perspectives (White, 2009).

During the 1960s to 1970s, scholars proposed two indirect citation relationships, i.e., BC and CC. Kessler (1963b) defined BC as the number of common references between two papers and "postulated that a number of scientific papers bear a meaningful relationship to each other (they are coupled) when they have one or more references in common" (Kessler, 1963a, p.49). The preliminary results confirmed the

doi:10.6342/NTU202201326

possibility of grouping papers into small and valid subgroups of the parent group. Additionally, both studies pointed out BC independent of words, languages, and expert reading or judgment. Ten years later, Small (1973) and Marshakove (1973) proposed CC, a new type of citation relationship, that measures the relationship between two documents based on how many times other documents cite them together. CC "is the logical opposite of the methods of bibliographic coupling" (Marshakove, 1973, p.3), and the relationship is established by citing authors (Small, 1973). Figure 2-1 shows how DC forms BC and CC. The BC of δ and ϵ is one due to their common citation β. The CC of δ and ϵ is two because they are co-cited by the later documents, π and ω. Because CC is established by the later citing authors, relationships between documents can change over time and reflect the evolution of the scientific structure.

Figure 2.1:
Direct Citation, Bibliographic Coupling, and Co-Citation



At first, citation relationships are applied to observe relationships between journal articles. Quickly, scholars use them to evaluate the authors or analyze the relationships between several authors. To the best of our knowledge, Clark (1957) first reported the correlation between an expert's journal citation counts and how often this expert is chosen as the highly visible persons in psychology field. Sher and Garfield (1983) indicated that the average citation counts of Nobel Prize winners differed significantly. Cole and Cole (1967) reported a similar finding and argued that citation counts could

doi:10.6342/NTU202201326

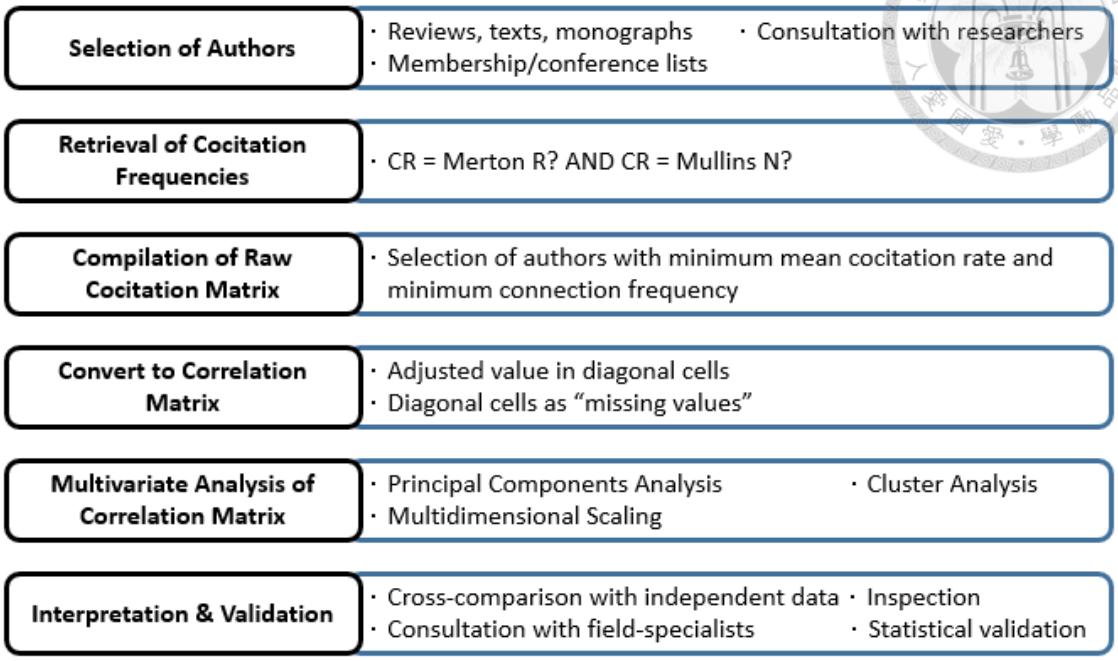
reflect an author's importance. In a series of later essays, Garfield examined several issues about DC, including the most cited primary authors, the correlation between cited times and awards, and the most cited publication of each author (Garfield, 1970, 1977a, 1977b, 1977c).

DC is the primary relationship used by the above studies when analyzing authors. The author co-citation analysis (ACA) proposed by White and Griffith (1981, 1982) applies co-citation analysis in gauging relationships between different authors. White and Griffith (1981, 1982) defined an *oeuvre*, the complete works of a person in French, for each author. Two authors are co-cited when any work from their oeuvre is co-cited. McCain (1990) concluded a set of procedures for executing ACA based on the actual experiences of the research team. Please refer to Figure 2-2. Some procedures, e.g., conversion of correlation and principal components analysis, almost become de facto ACA standards (Eom, 2003, 2009). ACA reveals the relationships between numerous authors from the perspective of the later authors. Many following studies utilized ACA to explore the knowledge structure (Ding et al., 1999; Ma et al., 2009; Tsay et al., 2003; White & McCain, 1998; Zhao & Strotmann, 2011, 2014). In a similar vein to ACA, Zhao and Strotmann (2008a, 2008b) proposed author bibliographic coupling analysis (ABCA), another method to map authors' research activities based on BC. ABCA gauges an author pair's relationship based on the number of common references between two authors' reference collections, composed of all references cited in each author's *oeuvre*.

The subject entity is much vaguer because it represents human knowledge, which is an abstract concept and might be a well-developed discipline or an emerging research topic. Some scholars define the subject entity as aggregators of different

doi:10.6342/NTU202201326

Figure 2.2:
The Procedures of ACA



Notes: Redraw based on McCain (1990)

journals, articles, or authors (Cronin & Pearson, 1990; Moya-Anegón et al., 2004; Narin et al., 1972); the others categorize similar words as the entity (Hsiao & Chen, 2019; Huang et al., 2018).

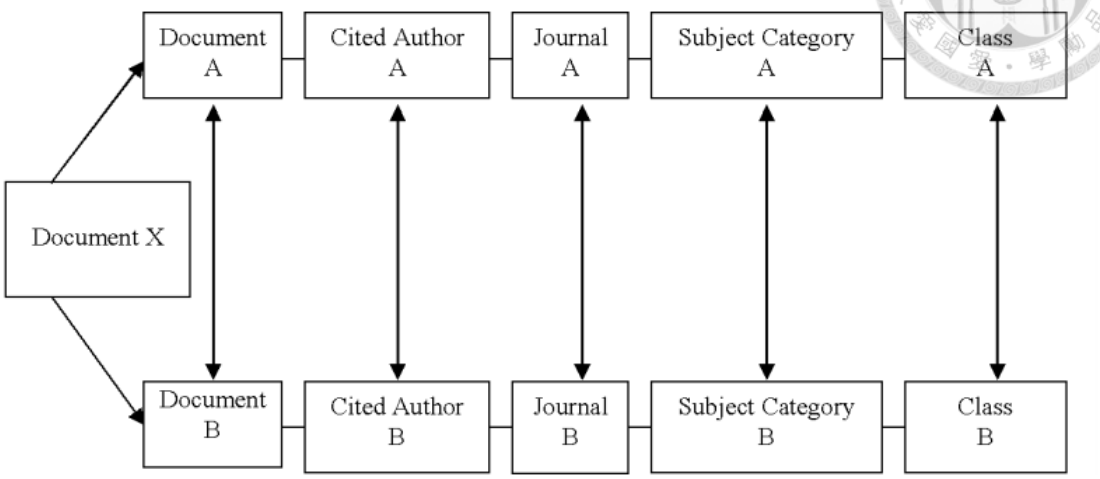
Most of the early studies defined subjects as the collections of various journals. Narin et al. (1972) investigated the interrelationships between different scientific disciplines by analyzing 275 journals about physics, chemistry, biochemistry, biology, and mathematics. Before 2000, several studies had used DC to investigate how different disciplines affect each other (Cronin & Pearson, 1990; Hargens, 1986; Urate, 1990).

Instead of using DC, Moya-Anegón et al. (2004) utilized CC in measuring relationships between subjects. Their study used pre-defined disciplines, Journal Citation Report (JCR) classes, and focused on interrelationships of disciplines. They proposed that the CC between two JCR subject classes can be measured by how often

doi:10.6342/NTU202201326

the documents within each class are co-cited in later documents, shown in Figure 2-3.

Figure 2.3:
The co-citation between classes



Notes: From Moya-Anegón et al. (2004)

The above studies defined scopes of disciplines first and analyzed inter-relationships between different disciplines later. Recently, two studies proposed that researchers can use words, the signifier of subjects, to analyze the inter-relationships between various subjects. Huang et al. (2018) presented Keyword Co-citation Analysis (KCA). KCA uses references' keywords as subjects and measures relationships of two subjects by how often these references are co-cited. Instead of using CC, Hsiao and Chen (2019) gauged two words' relationship based on the intersection of their related references and coined it as Word Bibliographic Coupling (WBC). The related references of a word are composed of all references cited by the works related to this word. Their study showed that WBC network provides a different viewpoint in comparison with keyword co-occurrence network.

This section briefs the development of citation relationships and citation entities. Table 2-1 reports the first study that utilized a special citation relationship in analyzing a particular citation entity. Researchers have studied most of the combinations in the

doi:10.6342/NTU202201326

last few decades, and these methods help scholars observe or evaluate the scientific structure from diverse perspectives. However, most citation analysis studies still count all citations as equal. In other words, the current practices may ignore the differences and gauge the relationships inaccurately. The following section reviews the related studies of citation theories and associated topics, including 1) the reasons motivating authors' citing, 2) the criteria used to select references, 3) the functions of these cited works, and 4) the features of the ITCs. It shows the necessity of differentiating citation relationships and the possible way to implement them.

Table 2.1:
Citation Relationships and Citation Entities

	Citation Entity		
	Work	Author	Subject
DC	Garfield (1955) Price (1965)	Clark (1957) Price and Gursey (1976)	Narin et al. (1972)
BC	Kessler (1963b)	Zhao and Strotmann (2008a)	Hsiao and Chen (2019)
CC	Small (1973) Marshakove (1973)	White and Griffith (1981)	Moya-Anegón et al. (2004) Huang et al. (2018)

Notes: For each citation relationship, the left/right columns represent the effect compared with the networks of the classical and other modified models, respectively.

2.2 Citation Behavior

As described above, researchers have developed numerous approaches for analyzing different kinds of citation entities. With the development of these methods, the discussions and studies about why authors cite and what citations mean emerge. Seeking the answers to these questions helps scholars better understand, explain, and utilize the result of citation analysis. Researchers have proposed several theories to

doi:10.6342/NTU202201326

answer these questions. Section 2.2.1 details these theories and the debates about them.

To verify these theories, researchers investigated what motivates authors to cite works and how authors select their citations. Related studies are presented in Section 2.2.2 and Section 2.2.3. Although these studies explain the authors' citation behavior, it will be hard to enlarge the research scale and keep the reliability at the same time. Scholars also analyze citation functions, which are decided by readers, to examine citation theories. Researchers also investigate citation features on a large scale to understand better how works are cited and differentiate these citation functions more effectively. Section 2.2.4 and Section 2.2.5 report the studies about them, respectively.

2.2.1 Citation theory

The emergence of citation indexes in the 1960s gives momentum for related studies. While this domain, which applied citation analysis in different applications, develops rapidly, the explanation for citation behavior remains uncertain. The normative citation theory and the social-constructivist theory of citing are proposed to explain citation behavior (Nicolaisen, 2007; Tahamtan & Bornmann, 2018).

According to Nicolaisen (2007), Norman Kaplan is the first person who takes citation behavior as normative behavior. Kaplan (1965) proposed that the primary function of citation practices is to reaffirm the underlying general norms of scientific behavior. The sociologists of science indicated that norms are "a set of rules supposed to establish trust in, and guarantee the reliability of, the knowledge claims produced by scientists" (Nicolaisen, 2007, p. 616). In 1942, Merton posited four basic norms as the scientific ethos, including universalism, communism, disinterestedness, and organized

doi:10.6342/NTU202201326

skepticism (Merton, 1973). The primary function of citation practices, proposed by Kaplan (1965), is to reaffirm these norms. Therefore, Merton's proposition becomes the foundation of normative citation theory, the first and most prominent theory in citation analysis, and provides a theoretical basis for scientometrics (Nicolaisen, 2007; Tahamtan & Bornmann, 2018).

The normative citation theory provides the foundations of citation analyses and motivates many subsequent studies. However, as pointed out by Kaplan (1965), "little is known about the norms operating in actual practice" (p. 179). A crucial question is to what extent citation practices reflect these norms, which take citations as a tool for coping with problems of scientific property rights and priority claims. Cole and Cole (1971) reviewed several significant issues regarding using citation index to measure the quality of sociological research and discussed this issue. They concluded that "the value of using them as rough indicators of the quality of a scientist's work should not be overlooked" (Cole & Cole, 1971, p. 28). They also indicated that the citing authors rarely fail to give credit correctly from the normative theory perspective. The author's omission of some crucial works is rarely due to malice but usually because of unawareness.

The second theory, social-constructivist theory of citing, takes citations "as rhetorical devices which are not related to the theory of Merton" (Tahamtan & Bornmann, 2018, p. 204). Its advocators argue that "scientific closure is the outcome of a negotiation process in which one party convinces the other by mere persuasion" (Nicolaisen, 2007, p. 620). Given the existence of negational and perfunctory citations, Gilbert (1977) proposed that scientific papers are tools to persuade the scientific community and share authors' opinions. Scholars typically use some rhetorical devices

doi:10.6342/NTU202201326

to enhance their arguments, and citations are used to persuade others and demonstrate the validity and significance of their works (Gilbert, 1977).

Since 1985, a series of articles by MacRoberts and MacRoberts (1986; 1987; 1988; 1996; 2018) questioned using citations as the indicator of measuring influences. Based on their examination of randomly selected papers in the history of genetics published after 1950, MacRoberts and MacRoberts (1986) concluded:

> *The mere presence of a reference is not a marker of influence, nor is the absence of a reference evidence that it is uninfluential. References are simply obvious historical leads and evidence of influence only when they have been demonstrated to be so. (p. 167)*

In their following article, MacRoberts and MacRoberts (1987) re-examined the works of genetics history and noted three patterns: (1) some works were used without citations, (2) some works were mentioned with secondary sources, and (3) some works were always credited when used. Therefore, they argued that citation counts could not fully reflect works' influence. Their studies also reported numerous reasons, e.g., uncited influent works, self-citations, and biased databases, to criticize the usage of citations as an indicator reflecting the influences and the proposition of the normative citation theory.

In addition to the theories mentioned previously, some scholars also proposed alternative perspectives. Small (2004) tried to integrate two theories into a framework, and Nicolaisen (2007) proposed the handicap principle to explain citation behavior. These theories present different perspectives to explain citation behavior. Many scholars have analyzed the citations to verify these theories. The following section

doi:10.6342/NTU202201326

briefs those studies focused on investigating the citation motivation.

2.2.2 Citation motivation

The present study defines citation motivations as the purpose that authors intend to achieve by citing works. Brooks (1985; 1986) were the early studies that surveyed authors' citation motivation. Brooks (1985) surveyed 26 faculty members to gather their citation motivations and concluded that citation motivations include currency scale, negative credit, operational information, persuasiveness, positive credit, reader alert, and social consensus. The study also indicated that persuasiveness is the most significant citing reason. Brooks (1986) interviewed 20 scholars about citation motivations in their published articles. The majority of their 437 citations were attributed with multiple motivations. Brooks (1986) also categorized the seven citation motivations into three groups: "(1) persuasiveness, positive credit, currency, and social consensus, (2) negative credit, and (3) reader alert and operational information" (Brooks, 1986, p. 34). The two studies showed that authors usually cite a reference for multiple motivations, not necessarily positive credit only. Case and Higgins (2000) also reported similar results. Their study argued that promoting authority and criticizing the citing works are the main citation motivations.

Bonzi and Snyder (1991) further investigated the citation motivations of 51 authors in several natural science disciplines, mainly aiming at finding the differences between self citations and other citations. They concluded 14 motivations from the previous works and asked the participants why they cited the references. The results showed:

[The] most substantial difference in reasons for self citations as opposed to

doi:10.6342/NTU202201326

citation to others is that of establishing the writer's authority in the field ... Demonstrating knowledge of important work in the discipline is another reason whose use differs significantly in self citation as opposed to citation to others ... The third significant difference is identification of earlier research on which the reported work builds. (p. 249)

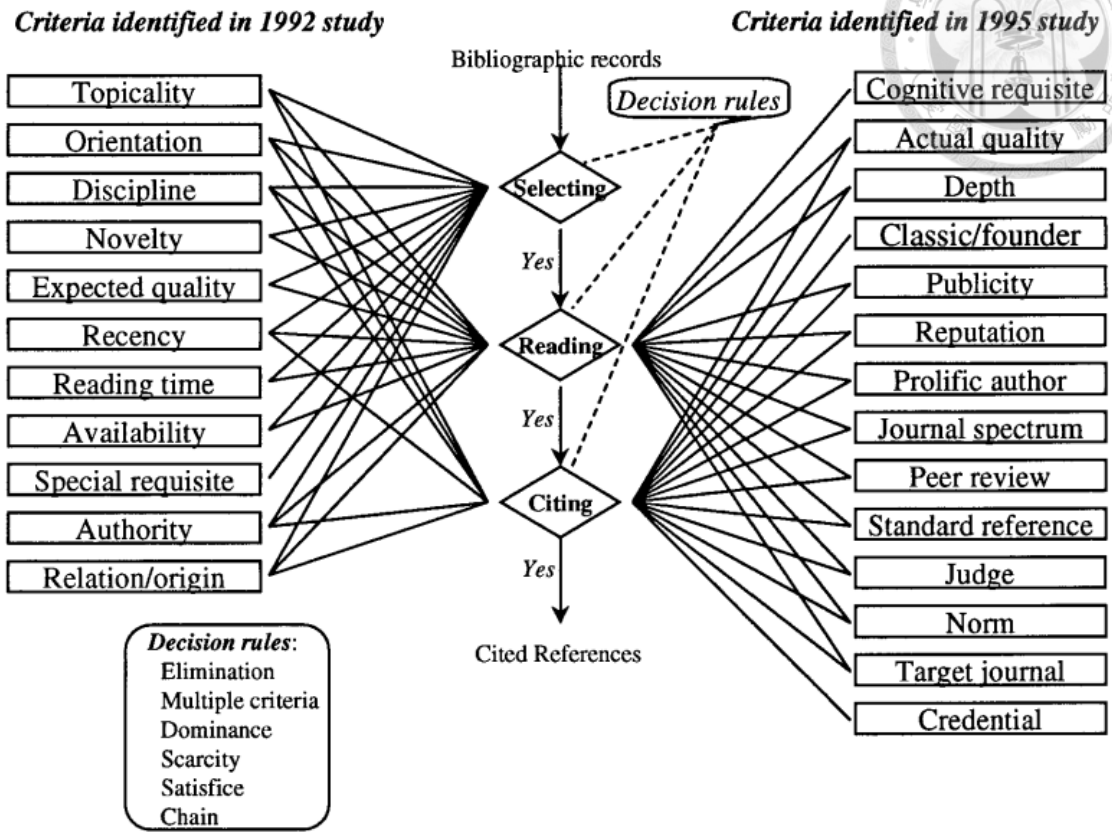
As regarding other motivations, the results reported no significant difference. In sum, these studies explored the reasons of authors' citing by questionnaires and interviews; their results showed the diversity of authors' motivations. Besides examining the authors' motivations, scholars are also interested in how authors select their citing works.

2.2.3 Citation selection

The present study defines citation selection as the authors' criteria for using a work. Both the citation motivation and selection are authors' reasons for citing. The motivation is more psychological purposes of authors' citation, and the selection is the standards which authors judge to use a work or not. Shadish et al. (1995) developed a questionnaire to explore how the colleagues chose their citations. They categorized the 28 items into six groups, including 1) negative citations, 2) personally influential citations, 3) creative citations, 4) classic citations, 5) citations for social reasons, and 6) supportive citations (Shadish et al., 1995). Another early attempt is a series of studies carried out by Wang and White in the 1990s (Wang & White, 1999; White & Wang, 1997a, 1997b). Wang and White (1999) interviewed 15 users who participated in their 1992 study about document selection. They interviewed authors to understand why they read or cite works. Their main research results, the cognitive model of document

doi:10.6342/NTU202201326

Figure 2.4:
Cognitive model of document use



Notes: From Wang and White (1999)

use, are shown in Figure 2-4.

Based on the items and constructs developed by Shadish et al. (1995), Case and Higgins (2000) used a questionnaire with 32 items to study citation selection. They also used another 11 questions to investigate the relationship between citing authors and cited authors. Their research focused on the criteria for citing highly-cited authors' works and considered the social relationships of citing authors. Recently, Thornley et al. (2015) conducted semi-structured interviews to investigate why researchers cite works. According to their interview, familiarity and knowledge with the authors are the critical reasons for citing others' works. Besides, only a few reasons for negative citations are about criticizing another researcher's works. These aforementioned studies were interested in how social relationships affect citation counts, and some

doi:10.6342/NTU202201326

recent studies, like Milard (2014) and Zingg et al. (2020), also reported their findings for this question. From a different perspective, Harwood (2008) argued several reasons which affect authors' citation selection, including publication citing policy, audience location, space restrictions, and publication speed.

To sum up, these studies reveal that numerous reasons affect citation selections. Although the results help researchers understand and verify how diverse the citation reasons can be, it is hard to further conclude a generalized theory due to lacking comparability among their research designs. The inconsistency between their classification schemes also hardens the difficulty of further generalization. Besides, although authors' opinions are the golden standard for explaining how they choose citations, the crucial difficulty is how to operationalize and enlarge the research scale. Hence, some researchers analyze the citation behavior from readers' perspectives to achieve better coders' reliability. The present study reviews those studies in the following section.

2.2.4 Citation function

The two sub-sections above review the studies investigating citation behavior from authors' perspectives. Another way to analyze citation behavior is from readers' perspectives. In the present study, citation functions are defined as readers' interpretations of authors' purposes when citing works. Many studies have analyzed the citation context to investigate citation functions.

One prominent result of early studies is the Moravcsik-Murugesan classification scheme, shown in Table 2-2. They developed the scheme based on their two studies,

doi:10.6342/NTU202201326

which examined citation functions of physics articles. According to their studies, the percentage of perfunctory citations was 41%, and 19% were non-confirmative. They argued that the portion was not negligible. Chubin and Moitra (1975) examined the usage of citations in several physics journals based on a scheme revised from Moravcsik-Murugesan's. Their study reported only around 5% of negational citations and 20% of perfunctory citations. They concluded that researchers could not abandon citation counts or simply take them at face value. Some other studies focused on citation functions in different disciplines, like studies of science (Speigel-Rosing, 1977), literary (Frost, 1979), social sciences (Peritz, 1983), multi-disciplines (Hurt, 1987), psychology (Krampen et al., 2007), information science (Tabatabaei, 2013), and humanities and social sciences (Lin, 2018). All studies reported various kinds of citation functions, including confirmation and contradiction. In other words, numerous kinds of citation functions exist and are not necessarily positive. Although many studies focus on analyzing citation functions, the inconsistency among their classification schemes still makes it hard to deduce a consolidated and generalized conclusion.

Some researchers aimed at highly cited papers and authors. Oppenheim and Renn (1978) investigated how authors used highly cited papers. They reported that "about 40% of the citations were for historical reasons, but that in the remaining 60% of the cases, the old paper is still actively used" (Oppenheim & Renn, 1978, p. 225). Using the classification scheme developed by Oppenheim and Renn (1978), Ahmed et al. (2004) randomly sampled 98 articles, which cited a work of Watson and Crick and published during 1993 and 2003. They examined 100 different citances in these articles. According to their result, "exact 75% of the citations in the articles cited

doi:10.6342/NTU202201326

Table 2.2:
Moravcsik-Murugesan classification scheme

Class	Description
Conceptual	The citing paper directly or indirectly uses the concept or theory of the cited paper to lay foundations to build on it.
Operational	The citing paper utilizes the cited papers as a tool to substantiate its claim; it borrows techniques, results, references, or conclusions from the cited papers.
Organic	The cited paper provides the necessary foundation of the citing paper; the foundation includes concepts, theories, or results.
Perfunctory	The cited paper describes alternative approaches, compares different results, or provides general background knowledge.
Evolutionary	The cited paper directly contributes to the logical development of the subject of the citing paper.
Juxtapositional	The cited paper does not contribute to the development of the citing paper and only shows the alternative approaches or provides material in parallel or divergent lines.
Confirmative	The author of the citing paper considers the cited paper as correct.
Negational	The author of the citing paper disputes the cited paper or claims that it is incorrect.
Redundant	When multiple papers are cited to show that various authors have dealt with a topic without being directly used by the citing study, these cited papers are redundant except for one of them.

Notes: From Moravcsik and Murugesan (1975) and Murugesan and Moravcsik (1978)

Watson and Crick for historical acknowledgment or background discussion of the work itself" (Ahmed et al., 2004, p. 154).

Although citation functions may not precisely correspond to citation motivations and citation selections, it is much easier to use operational procedures to identify citation functions. Therefore, researchers can reasonably enlarge their studies' scale. The citation motivation and citation selection should be the ground truth to discriminate citation relationships. However, the difficulty of the general survey makes it extremely hard to accomplish, if not impossible. Therefore, analyzing citation functions is relatively reasonable when researchers investigate how works are cited with massive

doi:10.6342/NTU202201326

data. Despite its advantage, the inconsistent schemes used in these reviewed studies restrict its further progress. Another limitation is that the studies mentioned above relied on human experts when identifying citation functions. A possible way to improve citation function analysis is to utilize features extracted from citation contexts.

2.2.5 Citation feature

Since the 1970s, researchers have investigated how authors cited references in the text body. They used several objective features like the distribution of ITCs and words in citances. Voos and Dagaev (1976) was an early study analyzing the citation features and investigating whether these features correlate with the characteristics of the cited articles. They calculated the frequency of reference mention, how many times a reference was mentioned in an article, and the distribution of ITCs in different sections.

The relationship between the cited and citing works is one focus. Herlach (1978) explored the correlation between the frequency of reference mention and how much a reference related to its citing article. Herlach concluded that "multiple mention of a relevant reference within the same research paper can be taken as an indication of a close relationship between a given cited paper and the citing papers" (Herlach, 1978, p. 310). Hou, Li, and Niu (2011) supported this conclusion after analyzing 651 papers published in 2008 under Biochemistry & Molecular and Genetics & Heredity in Web of Science (WoS). These results show that the frequency of reference mention is significantly higher when the reference is closely related to the citing article.

Scholars are also interested in the correlation between the frequency of reference mention and the importance of a reference. McCain and Turner (1989) proposed the

doi:10.6342/NTU202201326

utility index formula, which used the frequency of reference mention as one of the variables to measure the reference utility. Hooten (1991) and Tang and Safer (2008) also supported that the frequency of reference mention likely correlates with a reference's importance. Recently, Ding et al. (2013) ranked publications by the sum of how many times a publication was mentioned in the later articles and compared with the result of DC. Although the correlation between the rank order of both methods was moderate, the top articles ranked by both methods were very similar. Hence, they argued that the two methods generated different ranks for a similar set of citations. In addition, Hu et al. (2017) indicated that several types of citations are more likely to be mentioned repeatedly, including author self-citing citations, journal self-citing citations, and citations of recent works. Overall, these studies showed that the frequency of reference mention correlates with citation behaviors or reference characteristics and may help researchers differentiate important references from common references.

The location of ITCs is another crucial feature. Several studies reported that the distribution of ITCs in scientific articles follows a similar pattern (Bertin et al., 2016; Bornmann & Daniel, 2008; Boyack et al., 2018; Hsiao & Chen, 2018; Hu et al., 2013;). In general, the distribution of ITCs reaches its peak in the first 5% of text progression. Then, it drops and reaches the bottom in 20~30% of text progression. In other words, the density of ITCs in the introduction section is higher than that in other sections (Bertin et al., 2016; Hu et al., 2013). Both method and result sections have a relatively low density (Bertin et al., 2016).

Additionally, several studies indicate that the location of ITCs can be used to measure the citation's importance. Utility index formula mentioned above used the location of ITCs as one of its variables. Bornmann and Daneil (2008) compared the

doi:10.6342/NTU202201326

distributions of ITCs with high DC and low DC. According to their results, high DC articles are more frequently cited in method and result sections, and low DC articles are more likely cited in the discussion section. Tang and Safer (2008) also reported that "references cited in the Method were rated as significantly more important than references not cited in the Method section ... references cited in the Introduction section only and nowhere else were rated as significantly less important" (p. 263). Ding et al. (2013) made a similar conclusion after analyzing the frequency of citation mention. In sum, these studies argue that the location of ITCs correlates to the importance of a reference.

Recently, Zhao et al. (2017) explored to what extent the frequency and location can differentiate essential citations from nonessential ones. According to their findings, using both the frequency and location help identify nonessential citations. Besides, scholars also further investigate other citation features, like the number of words in a citance (Tang & Safer, 2008), the citation interval, the number of citations (Boyack et al., 2018, Hsiao & Chen, 2018), the distribution of part-of-speech (POS) tags, and frequently used words (Hsiao & Chen, 2018).

To sum up, the studies mentioned above investigate citation features such as frequency, location, and length of citance. Their results provide more understanding of how authors use citations and show the utility of citation features. Given the availability of machine-readable documents and the abundance of computing power, utilizing citation features extracted from full texts is possible when conducting citation analysis. It may help researchers deal with the question about the diversity of citation functions. Some studies have proposed alternative methods to measure citation relationships by using kinds of citation features. The related studies are reviewed and

doi:10.6342/NTU202201326

discussed in the following section.

2.3 Different Weighting Schemes

Although the previous studies have indicated that the importance of each citation varies, the expensive computing power and rare machine-readable documents make it extremely difficult to differentiate them by analyzing citation features on a large scale during the most time of the 20th century. No late than the 1990s, with the advance of computer technology and the popularity of machine-readable full-text documents, scholars can analyze citations on a large scale. Researchers propose various ways to differentiate citation relationships. Section 2.3.1 details those studies that suggested different weighting schemes for DC. Section 2.3.2 reports the studies which focused on two indirect citation relationships, BC and CC.

2.3.1 Weighting direct citation

Instead of counting all DCs equally, many studies have tried to weight DCs differently since the 1980s. Generally, these studies fall into two categories. Some utilize citation features, and others consider additional citation relationships.

Based on citation features

As reviewed in Section 2.2.3, several studies have used citation features to identify citation functions or measure the importance of a citation. McCain and Turner (1989) was an early study that tried to weigh DC differently. They developed a utility index to measure the relationship between a citing paper and a cited paper; the formula is shown below:

doi:10.6342/NTU202201326

$$UI = W_{sc}\left[W_i ln\left(X_i + 1\right) + W_m ln\left(X_m + 1\right) + W_d ln\left(X_d + 1\right) + W_r ln\left(X_r + 1\right)\right] \quad (2.1)$$

where W is the weighted coefficient; X is the DC number; i, m, d, and r denote Introduction section, Methods section, Discussion section, and Review section; sc means the relationship between the authors of citing paper and cited paper.

This formula considered three features: frequency of reference mention, location of ITC, and degree of self-citation. They also used the logarithm to leverage weight to each incremental occurrence (McCain & Turner, 1989).

In comparison with the formula proposed by McCain & Turner (1989), Wan and Liu (2014) added three additional citation features: (1) the time interval between a cited paper and a citing paper, (2) the average number of words in citances, and (3) average number of words between two ITCs. They labeled 820 citations from 40 papers selected in the Association for Computational Linguistics (ACL) anthology website. The ε-support vector regression (SVR) method, whose parameters were optimized by support vector machines (SVM), was used to measure the relation between the labeling results and six citation features. Their results showed that SVR could identify high-quality papers which were defined by two tutorials in two NLP domains. As pointed out by Wan and Liu (2014), "if we can determine the different importance levels of citations, we can improve almost all the citation-based bibliometrics by incorporating a citation importance value" (p. 1930). However, due to no authoritative ranking list for researchers, they noted that the author list ranked by SVR method was just for references.

Zhu et al. (2015) also studied how to differentiate citations by citation features and used them to measure author influence. In the first part of their study, they asked the citing authors to identify the key citations in their works and used machine learning to evaluate the effectiveness of the context information in identifying the key citations.

doi:10.6342/NTU202201326

According to their results, the feature based on the frequency of reference mention was the best, and they proposed an influence-primed h-index (hip-index) based on this feature. Zhu et al. (2015) argued that hip-index is a better indicator of researcher performance due to its higher precision in identifying ACL Fellows. Overall, these studies proposed different schemes to weight direct citation by utilizing citation features such as locations of ITCs and frequency of reference mention. Another approach, which weights DC by using other citation relationships, is discussed in the following sub-section.

Based on related citation relationships

Some scholars have tried to identify important relationships between two documents by using additional citation relationships. Small (1997) proposed combined linkage as a measure of document similarity. The combined linkage measures document similarity by direct citation linkage, DC, and indirect citation linkage, BC and CC. In addition, Small proposed longitudinal coupling, which "connects older and younger papers by taking either two steps forward or two steps backward" (Small, 1997, p. 277). The combined linkage between document i and j is represented in the following formula:

$$CombinedLinkage = \frac{\sum_{m \neq i,j} (C_{i,m}C_{j,m} + C_{i,m}C_{m,j} + C_{m,i}C_{m,j}) + (2 \times C_{ij})}{\sqrt{(k_i + 1) \times (k_j + 1)}} \quad (2.2)$$

where C_{ij} is DC that i cites j; $C_{i,m}C_{j,m}$ represents BC; $C_{i,m}C_{m,j}$ represents longitudinal coupling; $C_{m,i}C_{m,j}$ represents CC. k_i represents the sum of the cited times and the citing times of i.

The proposed formula weighted direct and indirect citation linkages with high and low coefficients, respectively. The sum is normalized by the cited times and the citing times.

Similarly, Persson (2010) proposed integrating DC with BC and CC into one indicator and coined it as weighted direct citations (WDC). WDC simply equals the sum of DC, BC, and CC. In formula form, WDC is defined as follows:

$$WDC = C_{ij} + \sum_{m \neq i,j} (C_{i,m}Cj, m + C_{m,i}Cm, j) \quad (2.3)$$

doi:10.6342/NTU202201326

In addition, Persson also proposed normalized weighted direct citation and said:

> *Some papers cite substantially more papers than other papers, and some papers are considerably more cited than others. Therefore, one could normalize a given shared reference by the number of citations to that particular paper ... Similarly, for a given co-citation we could take the inverse of the number of papers that the citing paper cite (Persson, 2010, p. 416).*

Generally, both Small (1999) and Persson (2010) noticed the possible effect of highly cited papers and publications with an unusually high number of references. They designed different approaches to deal with this question.

Normalized Similarity Index (NSI) is another similar method proposed by Nassiri et al. (2013). NSI is defined as:

$$NSI = \frac{\sum_{m \neq i,j} \left(C_{i,m}C_{j,m} + C_{i,m}C_{m,j} + C_{m,i}C_{m,j}\right) + C_{ij}}{k_i + k_j - \left[\sum_{m \neq i,j} \left(C_{i,m}C_{j,m} + C_{i,m}C_{m,j} + C_{m,i}C_{m,j}\right) + C_{ij}\right]} \tag{2.4}$$

The numerator of NSI is very similar to the numerator of combined linkage. The main difference between the two methods lies in their denominator's composition, namely how to normalize. Nassiri et al. (2013) indicated that the normalization step of NSI is Jaccard type measures instead of square roots used by combined linkage. In other words, the similarity between sample sets is "defined as the size of the intersection divided by the size of the union of the sample sets" (Nassiri et al., 2013, p. 93).

The three studies above modified the relationships based on the citation relationships closed to two documents. Another type of method in modifying the relationship is graph-based ranking algorithm. For example, several studies have utilized PageRank to obtain the relative importance of each node by analyzing the link relations of nodes in a directed graph recursively (Fiala et al., 2008; Wan & Liu, 2014). Besides, Valenzuela et al. (2015) proposed identifying meaningful citations based on citation features and citation relationships. They annotated the citation importance of 450 citations of the ACL anthology papers, concluded 12 citation features, and investigated the effectiveness of two clustering algorithms, including SVM and random

doi:10.6342/NTU202201326

forests. The Precision-recall curve showed that both methods outperform the baseline that labels are randomly assigned.

2.3.2 Weighting bibliographic coupling and co-citation

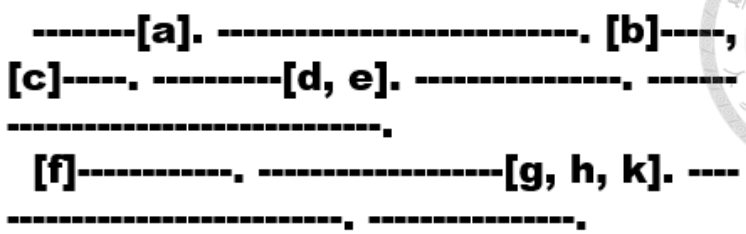
The approaches applied in weighting BC and CC can be categorized into several models: the distance model, the lexical similarity model, and the hybrid model. The basic assumption of the distance model is: the closer two ITCs are, the stronger their CCS is. As to the lexical similarity model, most proposals are based on the similarity of the words between two citances or titles. It usually assumes that the relationship between two works can be measured by the similarity between the words relating to two works. The difference between these proposals lies in how to define these words. Some scholars also proposed hybrid approaches, which combined both models or introduced other ways to measure the relationships.

Distance model

The distance model emerged in the 2000s. At first, Eto (2007) pointed out the possibility of weighting CCS by considering the locations of the ITCs. Based on the distance between two ITCs, he classified the classical CC relationships into two categories: enumerated CC, which means two papers are cited within one statement, and the others. Using the CiteSeer dataset, he reported that the two documents of enumerated CC have higher similarity in several similar indicators, including TF/IDF, BC, and CC. In the following study, Eto (2008) further divided the classical CC into four categories: CC between different paragraphs, CC in a paragraph, CC of the same sentence, and enumeration CC (showed in Figure 2-5). Based on five similarity indicators, he argued that CC between different paragraphs and enumeration CC have the lowest and highest similarity, respectively. Eto (2012) investigated "the effects of using co-citation context more deeply and more widely by comparing the search performance of six retrieval methods" (p. 651). The results supported his previous findings and showed the possible benefit of applying normalization techniques.

doi:10.6342/NTU202201326

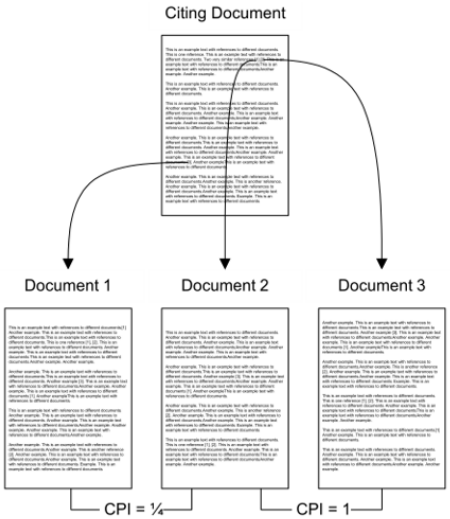
Figure 2.5:
Eto's four types of CC relationships



Notes: [a] & [g]: different paragraph. [a] & [d]: same paragraph. [b] & [c]: same sentence. [d] & [e]: enumeration. Redraw based on Eto (2008)

Gipp and Beel (2009) proposed another method and coined it as Citation Proximity Analysis, which measures CCS by considering the relative position of two ITCs. They suggested a weighting scheme, Citation Proximity Index, to measure CCS between two citations based on the distance. Figure 2-6 presents the details of Citation Proximity Analysis and Citation Proximity Index. According to the responses from 21 participants, the works recommended by Citation Proximity Analysis, compared with those recommended by classical CC, were more likely to be identified as related works.

Figure 2.6:
Citation Proximity Analysis



Occurrence	CPI value
Sentence	1
Paragraph	1/2
Chapter	1/4
Same journal / same book	1/8
Same journal but different edition	1/16

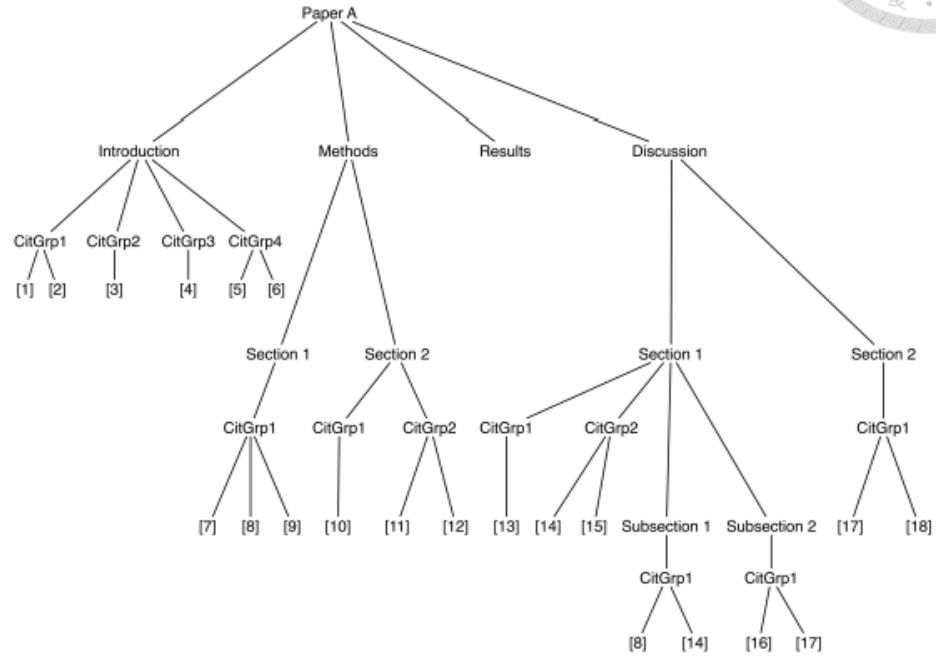
Notes: From Gipp and Beel (2009)

Callahan et al. (2010) proposed a more delicate method named as contextual cocitation. They viewed the article structure as a tree, shown in Figure 2-7, and calculated CCS based on the depth of the deepest common parent node of two ITCs. The idea of Callahan et al. (2010) is that "the relationship quantified is between cited

doi:10.6342/NTU202201326

documents and the strength will depend on the context" (p. 1134). They further discussed the potential of contextual cocitation:

Figure 2.7:
A document tree



Notes: From Callahan et al. (2010). CCS of [1] & [2] equals to 2(the depth of common parent node) plus 1.

> *Using contextual cocitation …has the potential to increase the granularity of the relationships recognized between works in a field, and also to change the unit of analysis used for studying these relationship ... it will be possible to analyze cocitation between documents at multiple levels of granularity and, thus, to more accurately map the shape of a field. (Callahan et al., 2010, p.1139)*

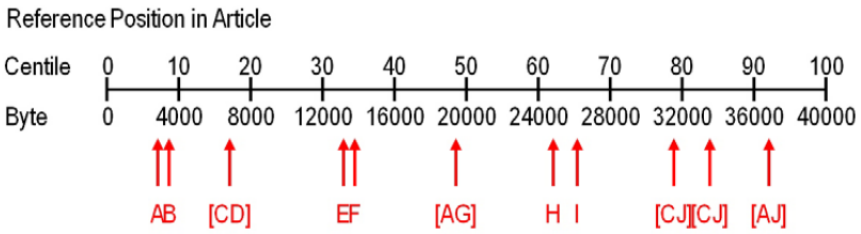
Contextual cocitation can also be applied to other kinds of citation entities, e.g., author, to analyze the relationships between a citing author and the cited authors. In sum, they argued that contextual cocitation is quite promising for informetrics and library science.

The studies reviewed above proposed several methods to improve the accuracy of CC based on the structure of an article. Boyack et al. (2013) proposed another way to gauge the CCS by considering the distance between two ITCs. Their research used two different kinds of weighting schemes: 1) normalizing the distances between ITCs to

doi:10.6342/NTU202201326

centiles; 2) the number of character offsets. For example, in Figure 2-8, the CCS between C and D is 1 when using the classical CC and 4 when using other weighting methods. The CCS between B and D is 1, 2, and 1 when calculating by the classical CC, CC based on centiles, and CC based on characters, respectively. Using 270,521 full-text documents from 2007, they compared their methods with the classical CC. The results showed that utilizing the full-text information increases the textual coherence of the clusters.

Figure 2.8:
Weighting scheme based on character offsets and centiles



	4	3	2	1	0
B)				all	none
P1)	same bracket	< 5 centiles	5-15 centiles	15-25 centiles	> 25 centiles
P2)	same bracket	< 5 centiles	5-10 centiles	10-15 centiles	> 15 centiles
O)	same bracket	< 375 bytes	375-1500 bytes	1500-6000 bytes	> 6000 bytes

Notes: From Boyack et al. (2013). In the table: B) Classical CC; P1) CCS adjusted by centile; P2) CCS adjusted by centile; O) CCS adjusted by character offsets.

These studies proposed various weighting schemes based on article structure or relative distance, and their results showed that introducing additional information improves co-citation analysis. The distance model assumes that the similarity of two ITCs is inversely related to the distance between them. This assumption is reasonable but not without question since the similarity between meanings of two sentences will not necessarily decrease as the distance increase.

Lexical model

The lexical model assumes that relationships between citation entities can be measured by gauging the similarity between their related words. If the related words can be defined reasonably, the similarity between related words can reflect the citation relationships. The major differences between the methods of this model will be 1) the

doi:10.6342/NTU202201326

scope of possible related words, 2) the ways used to extract related words, and 3) the methods used to measure the similarity between two texts.

Jeong et al. (2014) proposed content-based ACA. This method extracts the related words of an ITC by removing the stop words in the citance and stemming the remaining words. The cosine similarity between the related words of two ITCs determines the CCS value, and the CCS of two references is the maximum CCS among all the pairs of their ITCs. As shown in Figure 2-9, the CCS of Daubechies and Mallat is the value of cosine similarity between sentence X and sentence Y. They followed the classical ACA approach to gauge CCS between authors and conduct other follow-up procedures. Using 1,402 full-text articles published in the JASIST from January 2003 to June 2012, they compared the results of content-based ACA and the classical ACA. The factor analysis showed that content-based ACA outperforms the classical ACA by explaining more variance and revealing more subfields. The citances, concluded by Jeong et al. (2014), can "help discover the essential structural components of the corresponding traditional co-citation network" (p. 209).

Figure 2.9:
Examples of Content-based ACA

Sentence X: The **standard** 2D **discrete wavelet packet transform**, as **introduced earlier**, is very **efficient** and can **compute** the **wavelet packet coefficients** with a **complexity** of O(n) (where n is the number **pixels** in the given **image**) (Daubechies, 1992).

Sentence Y: **Discrete wavelet** and **wavelet packet transforms** have been **shown** to be **useful** for **image analysis** (Mallat, 1989).

	analysi	complex	discret	effici	introduc	pixel	standard	use	coeffici	comput	earlier	imag	packet	shown	transform	wavelet
X	0	1	1	1	1	1	1	1	1	2	1	0	1	1	0	2
Y	1	0	0	0	1	0	0	1	0	1	0	1	0	1	1	2

Notes: From Jeong et al. (2014). Words with bold faces are the related words of the sentences.

As to BC, Liu (2017) proposed using titles to improve BC and coined it as DescriptiveBC. The value of classical BC is dichotomous, either 0 or 1. In DescriptiveBC, the value is decided by the Jaccard index similarity of the words of the references' titles. In brief, the more similar the titles of two references are, the higher their BCS will be. The way to calculate DescriptiveBC is shown as follows:

$$Sim_{ref}\left(r_1, r_2\right) = \begin{cases} 1 & \text{, if } r_1 = r_2 \\ \frac{|Title(r_1) \cap Title(r_2)|}{|Title(r_1) \cup Title(r_2)|} & \text{, otherwise} \end{cases} \quad (2.5)$$

doi:10.6342/NTU202201326

where r_1 and r_2 are two references cited by different works. $Title(r)$ is a set of terms in the title of r.

$$DescriptiveBC\,(a_1, a_2) = \frac{\sum_{r_1 \in R_{a_1}} \max_{r_2 \in R_{a_2}} Sim_{ref}\,(r_1, r_2) + \sum_{r_2 \in R_{a_2}} \max_{r_1 \in R_{a_1}} Sim_{ref}\,(r_1, r_2)}{|R_{a_1}| + |R_{a_2}|} \quad (2.6)$$

where a_n represents the article; R_{a_n} is the set of citation in a_n; r_n is the any one citation of a_n.

Besides, he also proposed to extract the descriptive terms from the non-stop words in titles, and these terms can be used to explain why an article is related to another one. The result shows that DescriptiveBC "(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" (p. 932).

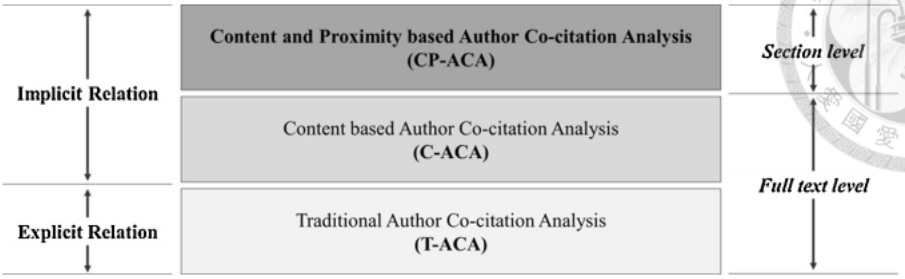
In sum, these studies utilized the lexical similarity of two citation entities as the criteria for measuring BCS or CCS. The lexical similarity approaches show a different perspective to observe the scientific structure and outperform in identifying related scholarly articles and subfields.

Hybrid model and others

Some researchers combined both models in recent years. For example, Kim et al. (2016) proposed a method based on both models and coined it as content- and proximity-based author co-citation analysis (CPACA). According to their research results, classical ACA shows the explicit relations on the full-text level, content-based ACA reveals the implicit relations on the full-text level, and CPACA only focuses on the implicit relations on the section level (shown as Figure 2-10). Figure 2-11 shows the difference between content-based ACA and CPACA. The graph shows that CPACA only considers those ITCs pairs co-cited in the same section. Their study selected and downloaded 6,360 full-text articles, published in 15 journals of oncology in JCR, from PubMed Central and investigated the effectiveness of CPACA. They argued that CPACA "enables the identification of distinct sub-fields of authors to represent authors' subject relatedness" (p. 954).

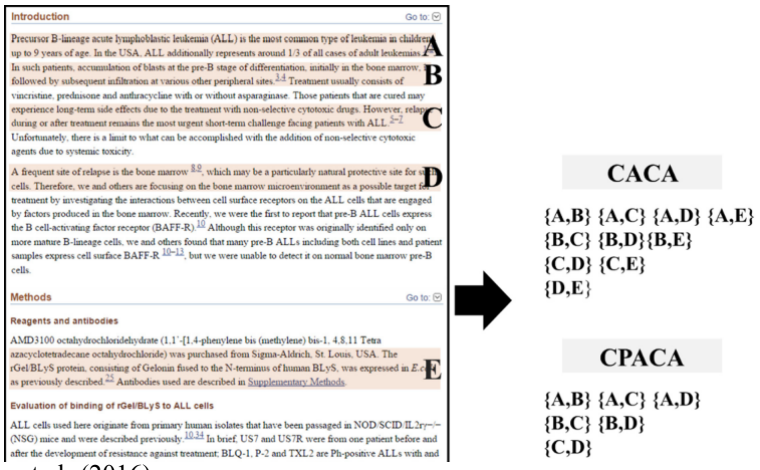
doi:10.6342/NTU202201326

Figure 2.10:
Comparison of three ACA methods



Notes: From Kim et al. (2016)

Figure 2.11:
Difference between content-based ACA and CPAC



Notes: From Kim et al. (2016)

Bu et al. (2018) proposed another approach that uses the frequency of reference mention in the full-text and the number of context words, namely words of ITCs, as the additional variables. They weighted CCS by the weighting sum of three values: 1) the raw value of classical ACA, 2) the number of citation appearance, and 3) the number of context words. The citations with a higher citation appearance or more context words will be advantageous. They compared the results of classical ACA and their proposed method by using factor analysis, network analysis, and MDS-measurement. Their method showed a better clustering result and revealed more scientific structure details.

To sum up, most studies focus on improving the results of CC by numerous models. Their results show that differentiating citation relationships more carefully can reveal more details of the scientific structure. The lexical similarity approaches provide some different insights when analyzing a discipline. However, only a few studies

doi:10.6342/NTU202201326

adopted the lexical similarity approaches, and there is a lack of studies investigating the possibility of differentiating citation relationships by semantic analytics. Due to the rapid development of NLP techniques in recent years, numerous studies aimed at implementing NLP techniques in identifying the importance and emotion of DC. Compared with the development of related studies about DC, the potential of applying semantic analytics in BC and CC is noteworthy, especially considering the recent advancement of NLP technology. These studies are reviewed in the following section.

2.4 NLP, Sentiment Analysis, and Citation Analysis

Sentiment analysis of citations

Since the mid-1990s, the development of data mining and NLP algorithms inspired new research interesting about classifying citations and measuring their sentiment polarity (Ding et al., 2014; Goodarzi et al., 2014). The new research field, i.e., sentiment analysis of citations, "aims to determine the sentiment polarity conveyed through a segment of text concerning a specific entity (opinion target)" (Xu et al., 2015, p. 1334). Depending on the granularity, the text considered may be the whole document, a set of sentences, one sentence, or a part of a sentence (Abu-Jbara et al., 2013; Xu et al., 2015). Sentiment analysis of citations can help summarize scientific articles (Athar & Teufel, 2012a), understand the evolution of a field (Athar & Teufel, 2012b), and measure the impact of a given publication by both quantitative and qualitative data (Hernández-Alvarez et al., 2017).

Sentiment analysis can benefit several applications of citation analysis, such as citation motivation classification, citation summarization, information retrieval, citation recommendation and prediction, and knowledge graph mining (Ding et al., 2014). The writing styles of scientific texts make the sentiment analysis of citation harder than other genres. Authors usually try to be neutral and hedge any personal bias (Athar, 2011; Athar, & Teufel, 2012a). Therefore, the sentiment polarity of citations is often unclear. Citances are often neutral, and both negative citations and criticisms to

doi:10.6342/NTU202201326

other works are often veiled (Athar, 2011; Ghosh et al., 2016). Authors may use different forms, like contrastive terms, to show negative polarity (Athar, 2011). Besides, another challenge is the terms of the specific domain. Technical terms and specific writing styles make the sentiment analysis of scientific text different from other genres (Athar, 2011; Goodarzi et al., 2014).

Constructing sentiment lexicon

Sentiment analysis can be further divided into three tasks: sentiment lexicon construction, context recognition, and determination of polarity and its degree (Goodarzi et al., 2014). The sentiment lexicon shows the sentimental polarity and plays a vital role in determining polarity and degree. Constructing the lexicon decides the meaning of words and the relationships between words. Most reviewed studies used the dictionary-based approach, which manually determines polarity and relationships between words to construct the lexicon (Goodarzi et al., 2014). The dictionary-based approach is the thesaurus model mentioned in Section 1.2. Its primary advantage is that manual annotation can represent the meanings and relationships explicitly, but its expensive cost restricts adding new words or adjusting the words' meanings (Saitoh, 2019). Some studies constructed their sentiment lexicon, e.g., Teufel et al. (2006a) identified 20 manually acquired verb clusters and 892 cue phrases in their research. Some open resources based on this model, e.g., Natural Language Toolkit (NLTK), a natural language toolkit of Python, also provides researchers a general way to implement sentiment analysis.

Recognizing citation context

Context recognition is to determine the authors' sentiment for an opinion target by analyzing the words. In sentiment analysis of citations, the task can be defined as follows:

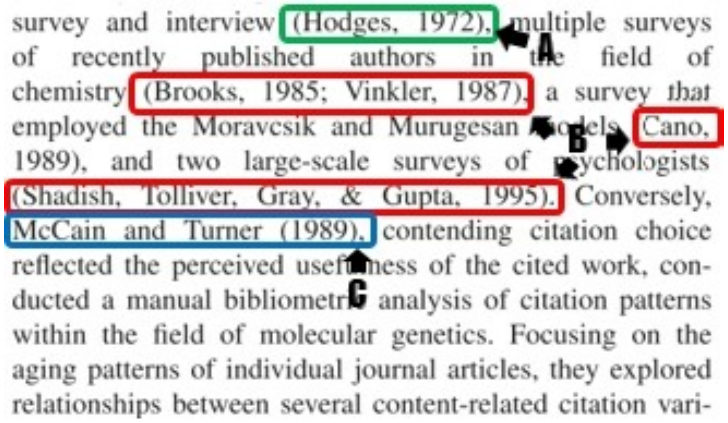
> *Given a scientific article A that cites another article B, find a set of sentences in A that talk about the work done in B such that at least one of these sentences contains an explicit reference to B. (Abu-Jbara et al., 2013, p. 599)*

doi:10.6342/NTU202201326

The range of context may be a fixed range, e.g., several words or sentences around an ITC, or a dynamic range decided by other methods (Goodarzi et al., 2014).

Context recognition in the scientific text will be challenging due to the scope of influence of citations varies widely. The possible influence scope of citations varies widely from a single clause to several paragraphs (Athar, 2011). In Figure 2-12, for example, three ITCs (red rectangle pointed by B) only influence a single clause, one ITC (green rectangle pointed by A) influences a sentence, and the other one ITC (blue rectangle pointed by C) affects two sentences. Athar and Teufel (2012a) investigated the effects of a different range of contexts on measuring the polarity of citations and concluded that "the task of jointly detecting sentiment and context is a hard problem" (p.599). After examining 300 sentences, Abu-Jbara et al. (2013) noticed that the citation context usually fell within a window of four sentences.

Figure 2.12:
The scope of influence of citations



Notes: From Ding et al. (2014)

Determining sentimental polarity and classifying citations

The polarity and purpose of citations can be identified to a degree by using sentiment lexicon and context recognition (Abu-Jbara et al., 2013; Goodarzi et al., 2014). The features used in polarity determination usually are language parameters. At the word level, they could be n-grams, cue words, cue phrases, and POS tags. Some studies used the sentence-level features, e.g., dependency structures (Athar, 2011; Athar, & Teufel, 2012a; Ghosh et al., 2016) and specific sentence structure (Dong, & Schäfer, 2011).

doi:10.6342/NTU202201326

The other parameters different from the word and sentences levels include location (Dong & Schäfer, 2011; Teufel et al., 2006a), popularity and density (Dong & Schäfer, 2011), author names and publication years (Kim & Thoma, 2015), and self-citation (Ghosh et al., 2016). Using these parameters, scholars can identify the citation polarity with numerous machine learning algorithms, e.g., Naïve Bayes, Logistic Regression, Decision Tree, k Nearest Neighbor, SVM, and Conditional Random Fields. In short, NLP techniques can help researchers extract more language parameters and use them as citation features to categorize citations. The availability of NLP tools, together with the full-text data and computing power, provides a convenient way for scholars to conduct citation analysis with NLP techniques.

Count model and predict model

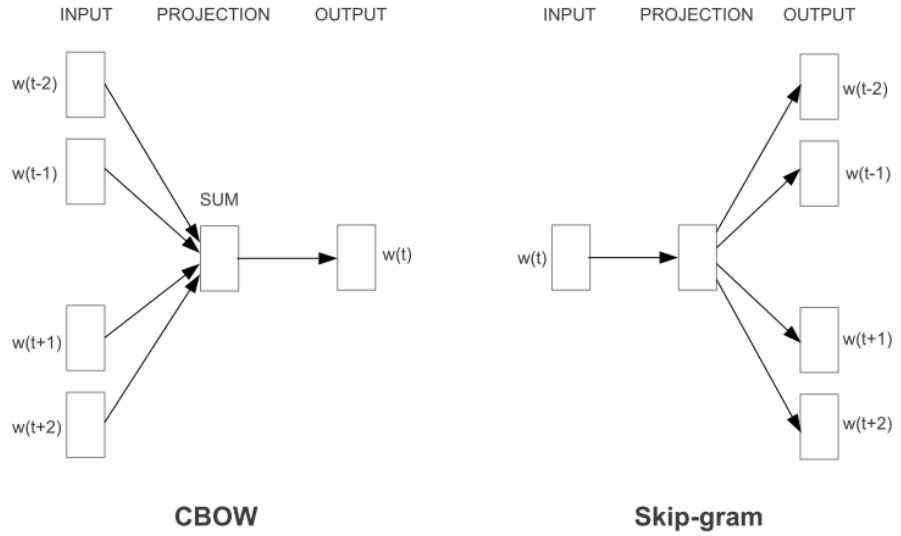
Although the thesaurus model has provided tools for extracting more citation features, constructing the semantic lexicon is time-consuming and expensive due to its dependence on manual work (Saitoh, 2019). The count model is an alternative way to analyze the meaning of words. It determines words' meanings and constructs their relationships via analyzing the co-occurrence frequency with other words. If there is a close relationship between two words, according to Miller and Charles (1991), the context around the words will be similar. In other words, the higher the co-occurrence frequency is, the closer two words' relationships will be. After analyzing corpus, researchers can transfer the raw co-occurrence counts of a word as the word vector, the distributed representation of a word, to represent its meaning (Baroni et al., 2014). The count model largely lowers the difficulties of building and updating the meanings of words. However, when determining the meaning of a word, it is necessary to count the co-occurrence frequency with all other words in a corpus. The time and space complexity will be noticeable while analyzing a huge corpus (Saitoh, 2019).

The development of NLP techniques provides another way of extracting the words' contextual representations. Mikolov, Chen et al. (2013) released an open-source project, word2vec, and proposed CBOW and SG to extract the contextual representations. The two neural network language models can compute continuous

doi:10.6342/NTU202201326

vector representations of words that measure syntactic and semantic word similarities. The idea of CBOW is to train a model to predict a word when inputting its surrounding words. After completing the training, researchers can use the model's parameters as the meaning of words. SG follows a similar idea, but its input and output are a single word and its possible surrounding words, respectively. According to Mikolov, Chen et al. (2013), "the CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word" (p. 5). Figure 2-13 presents the architectures of CBOW and SG. The model's parameters represent the meaning of words in the training corpus. The vector representation of a word, known as word embedding, has good accuracy on the Semantic-Syntactic Word Relationship test set and Microsoft Research Sentence Completion Challenge. In addition, Mikolov, Sutskever et al. (2013) further reported two different strategies, hierarchical softmax and negative sampling, to lower the complexity of computing.

Figure 2.13:
The architectures of CBOW and Skip-gram



Notes: From Mikolov, Chen et al.(2013)

Devlin et al. (2018) introduced BERT, Bidirectional Encoder Representations from Transformers, a language model that supports transfer learning and fine-tuning on NLP tasks. The pre-trained models are available at https://github.com/google-research/bert. Researchers can utilize them in their studies and research topics, e.g., Tseng (2020). When BERT was released, it outperformed other NLP techniques in four types of NLP

doi:10.6342/NTU202201326

tasks, including sentence pair classification tasks, single sentence classification tasks, question-answer tasks, and single sentence tagging tasks. After fine-tuning with proper datasets, the pre-trained models can be applied to classify citations. Based on BERT, Reimers and Gurevych (2019) proposed Sentence-BERT (SBERT) and provided its package at https://www.sbert.net/docs/installation.html. Their results show that SBERT outperforms other sentence embeddings methods on common semantic textual similarity tasks and transfer learning tasks. Because SBERT can improve the results on semantic textual similarity tasks, it may help scholars identify the similarity between citances.

To sum up, numerous studies have utilized different approaches to investigate citation behavior since the 1960s. In the early days of citation analysis, as reviewed in Section 2.2, scholars studied citation behavior by interviewing authors, examining citation contexts, and analyzing citation features. Their results confirmed the complexity of citation behavior and showed the importance of differentiating different types of citations. Those studies mentioned in Section 2.3 have proposed several possible methods to analyze citation relationships from various perspectives. With the availability of machine-readable full-text documents and computing power, recent studies can use more data extracted from full-text articles when conducting citation analysis, especially the studies related to co-citation analysis. The studies of NLP semantic analysis in citation analysis, reviewed in Section 2.4, show the possibility of applied NLP techniques in citation analysis and its possible applications.

Given the recent advance of NLP techniques based on neural networks, the investigation of applying semantic analysis techniques in citation analysis is necessary and promising. Introducing recent semantic analysis may improve the results of classifying citations by their polarity and measure the citation relationships better, especially for BC and CC. The improvement will enhance various applications of citation analysis. Therefore, this study compares the methods reviewed in Section 2.3 and those that gauge citation relationships based on semantic analysis. Comparing the results of different methods answers whether conducting citation analysis with NLP

doi:10.6342/NTU202201326

semantic analysis provides a different perspective to analyze the scientific structure. The analysis for the clustering results tells whether NLP analysis techniques can uncover the sub-fields better. The investigation for the influential entities and relevant relationships also shows the differences between these methods. Therefore, the research results answer the effectiveness of applying NLP semantic analysis techniques in citation analysis and help future studies that aim at improving the methods of differentiating citation entities and relationships.

doi:10.6342/NTU202201326

Chapter 3

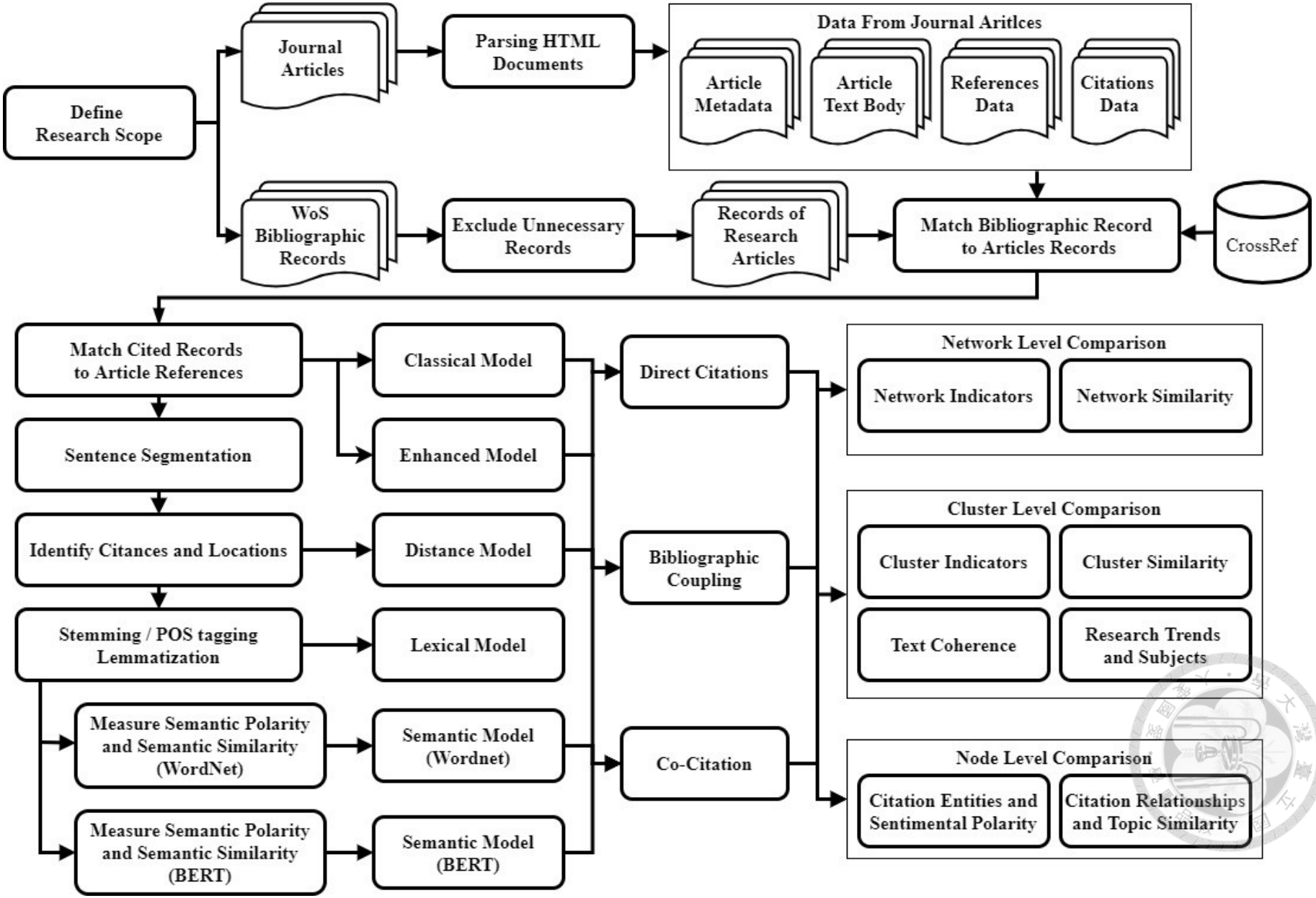
Research Design

The present study examines the results of citation analysis based on different models to investigate the effectiveness of utilizing NLP semantic analysis in citation analysis. The models are the classical model, the frequency model, the distance model, the lexical model, and the semantic model. The semantic model includes both the BERT and Wordnet models. Three models, including the classical, frequency, and semantic models, provided the results of all three types of citation relationships, namely DC, BC, and CC. The lexical model provided the results of BC and CC, and the distance model is used to measure CC only. At the network, cluster, and node levels, the results of citation analysis based on these models were examined and compared to each other. The comparisons showed the effectiveness of applying NLP semantic analytics and revealed the advantages and weaknesses.

This chapter reports the research design and implementation. Figure 3-1 shows the research design, which comprises three main stages: data preparation, citation relationship measurement, and citation network analysis. The top half of Figure 3-1 describes the procedures for gathering the research data. The first column from the left in the bottom half shows a series of steps to extract the data required for different models. The second and third columns are the models and citation relationships examined in the present study. The fourth column briefs the procedures used to

doi:10.6342/NTU202201326

Figure 3.1:
Research Design



doi:10.6342/NTU202201326

evaluate our results. The following sections detail each stage sequentially,

3.1 Data Preparation

The research data were composed of two datasets, including HTML full texts and WoS bibliographic records (WoS records). Given the research purposes of this study, the full texts were necessary to identify the semantic information of each citation. Besides, as shown in Figure 3-2, one article may be cited in various forms. Hence, the WoS records were used as the unified format of these references. In addition, because two datasets, HTML full texts and WoS records, were used in this study, mapping their reference data is required. Subsection 3.1.3 briefs the mapping algorithms. The remaining parts of this section also detail the data collecting, data preprocessing, and NLP procedures.

Figure 3.2:
Example of various forms between two references

Small, H. (1982). Citation context analysis. In *Progress in communication sciences* (pp. 287–310). Nor wood: Ablex.
Small, H. (1982). Citation context analysis. In B. Dervin, & M. J. Voigt (Vol. Eds.), *Progress in communication sciences: 3*, (pp. 287–310).

Gipp, B., & Beel. J. (2009a). Citation proximity analysis (CPA)—A new approach for identifying relate work based on co-citation analysis. In *Proceedings of the 12th international conference on sciento metrics and informetrics* (Vol. 2, pp. 571–575). Rio de Janeiro: International Society for Sciento metrics and Informetrics.
Gipp, B., & Beel, J. (2009). Citation proximity analysis (CPA) - a new approach for identifying related work based on co-citation analysis. *Proceedings of the 12th International Conference on Scientometrics and Informetrics* (pp. 571–575).

Notes: The source of two references in black fonts are from Eto (2012), and another two references, light blue fonts, are from Eto (2019). Eto cited Small's same publication in two different appearances. The references of the same work of Gipp and Beel are also listed in two forms. In other words, the forms of a cited work may be different when the same author cites the same work.

3.1.1 Defining research domain and download research data

The present study chose LIS domain as the research target because a series of the previous studies, Hsiao and Chen (2018, 2019, 2020), have focused on this domain and provided the results for examining the outcome of the present study. The cumulative expertise in this domain makes the subject analysis feasible. Journals were used to

doi:10.6342/NTU202201326

form the research dataset. To define the scope of LIS, two studies, Abrizah et al. (2015) and Huang et al. (2019), were used in the present research. Abrizah et al. (2015) survived 243 active authors/editors publishing in LIS to classify the journals of a sub-categorization, information science-library science (IS-LS), of WoS into several classes: information science, library science, information system, and undecided. Huang et al. (2019) also differentiated the IS-LS journals into MIS and LIS journals based on Abrizah et al. (2015) and several classification sources, including the Global Serials Directory in Ulrichsweb, SCImago subject classification, and the US Library of Congress catalog. These two studies have classified the LIS journals by surveying experts' opinions and examining several classification schemas. Hence, the present study used these two studies and combined their results to define the LIS journals.

In principle, only the journals classified as LIS journals by both studies were included. Of the 83 journals examined by Abrizah et al. (2015), 39 and 23 were classified as LS and IS, respectively. As to Huang et al. (2019), 63 of 88 journals were categorized into LIS-related categories, including LS, IS, and scientometrics. In total, 44 journals were categorized into LIS-related categories by both studies. Then, the journals that could not be accessed were excluded due to the requirement of machine-readable full texts. Besides, given the complexity of parsing articles, the present study removed the journals if their published papers lacked proper HTML structure. For the same reason, the journals using footnote as their primary writing style were also excluded. Additionally, only English journals were included due to the limitation of NLP tools.

Accordingly, the final list included 15 journals, as shown in Table 3-1. The articles published in these journals from 2010 to 2019 are used as the research data. Two exceptions are Journal of Librarianship and Information Science and Journal of Information Science because both journals do not provide HTML documents for the articles published before 2012. In addition to the full texts of journal articles, the bibliographic data of these articles were downloaded from WoS. To prevent the possible unexpected effects due to including multiple types of publications, WoS

doi:10.6342/NTU202201326

records were applied to filter out publications like editorial letters, book reviews, opinions, and review articles. Only the bibliographic records whose data type is "Article" were downloaded.

Table 3.1:
Journals included in the final list and the number of their HTML docs

Journal Name	Period	Issues	Docs
Aslib[1]	2010-2019	58	414
Government Information Quarterly	2010-2019	44	816
Health Information and Libraries Journal	2010-2019	40	402
Information Processing & Management	2010-2019	60	872
JASIST[2]	2010-2019	120	2,034
Journal of Documentation	2010-2019	60	661
Journal of Information Science	2012-2019	48	444
Journal of Informetrics	2010-2019	40	926
Journal of Librarianship and Information Science	2012-2019	32	361
Library and Information Science Research	2010-2019	38	481
Library Hi Tech	2010-2019	40	518
Online Information Review	2010-2019	66	822
Program: Electronic Library and Information Systems[3]	2010-2019	40	326
Scientometrics	2010-2019	120	3,210
The Electronic Library	2010-2019	60	865
Total		866	13,152

Notes: (1) *Aslib* represents both *Aslib Proceedings* and *Aslib Journal of Information Management*, which the latter is the journal's new name since 2014. (2) *JASIST* is the abbreviation of the *Journal of the Association for Information Science and Technology*, renamed from the *Journal of the American Society for Information Science and Technology* since 2014. (3) Since 2018, this journal has been renamed *Data Technologies and Applications*.

3.1.2 Extracting data

After collecting the articles and bibliographic records, the present study extracted the necessary data. First, four kinds of data were extracted from each article, including metadata, text body, references, and citations. The metadata includes basic information about this article, e.g., the title and authors. The text body contains the raw text of the

doi:10.6342/NTU202201326

publication. The references contain the text part and HTML attributes of each reference listed in a document. The citations include the location, citances, and HTML attributes, used to identify a reference's citances in the text body.

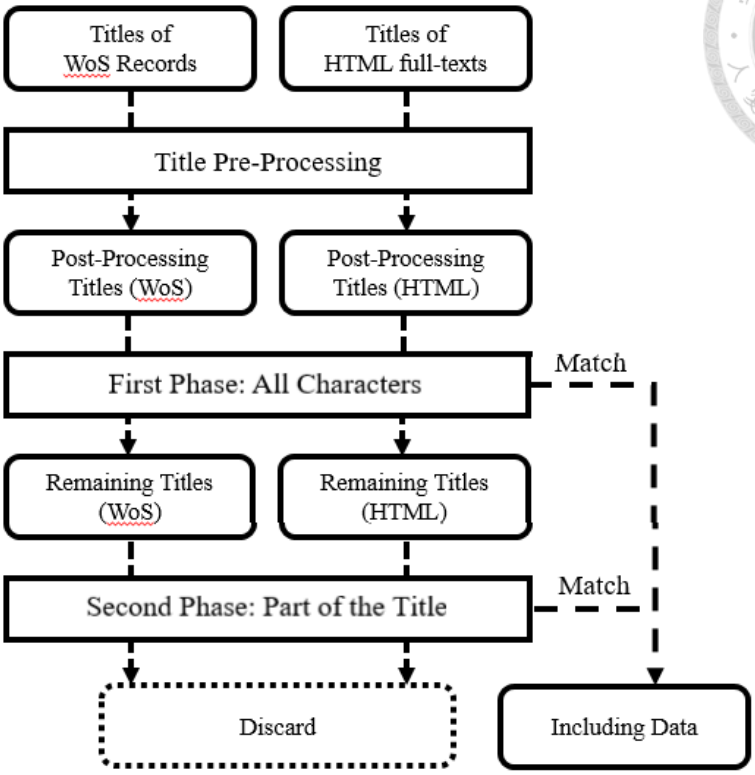
3.1.3 Mapping WoS records and HTML full-texts

As mentioned above, WoS records were used to associate the different forms of the same cited work. The present study compared the titles of two datasets to map HTML full-texts to corresponding WoS records with the following procedures. In the first phase, all titles from two datasets were collected, the characters in the titles were lowercased, and the spaces were removed. Then, if a post-processing title from WoS records could be fully matched to one and only one post-processing title from HTML full texts, the WoS record would be assigned to this HTML full text. For the unmatched records, the criterion would be lowered as matching the first and the last 50 characters of the post-processing titles in the second phase. Similarly, a record would be mapped to a full text if its title could be matched to one and only one title of a full text. The flow chart of this algorithm is shown in Figure 3-3. The WoS records and HTML full texts mapped during the two phases were used in this study, and the remaining were excluded. After the mapping procedure, 10,088 articles were included in this study. These articles were the source article of this study.

Then, the present study examined the citing records of the matched WoS records and found out their corresponding references, as shown in Figure 3-4. First, if any DOI could be identified in a WoS citing record, the DOI would be used to search its possible corresponding reference from the HTML full text. If there were no available DOI in the citing records or references, the WoS citing record would be split into tokens and examined whether any token indicates page or volume. If yes, the page token, volume token, and the first token in the citing record, usually the first author's family name, were used to find the possible corresponding reference. If some records still remain unmatched after completing the previous steps, this study measured the number of

doi:10.6342/NTU202201326

Figure 3.3:
Mapping WoS records and HTML full-texts



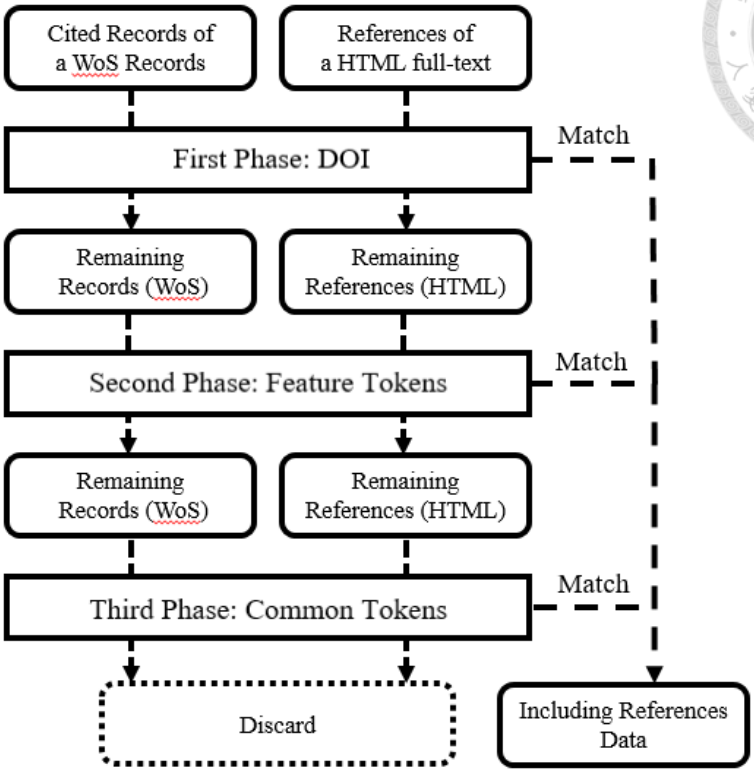
common tokens between the unmatched cited records and references. Be noted that only the tokens composed of more than three characters would be considered. In principle, a cited record would be assigned to the reference with the highest number, at least 3, of common tokens. After the mapping procedure, 447,984 references were included in this study.

3.1.4 NLP and other preparing procedures

The steps described above provided the data used to measure citation relationships by the classical model. The ITCs data, namely the frequency of reference mention, can enhance the classical model, and more data derived from analyzing HTML full texts is necessary for other models. The distance model requires the location of ITCs, and the citance content is necessary for the lexical and semantic models. Besides, NLP procedures, like POS tagging or lemmatization, are also required for lexical and semantic models.

doi:10.6342/NTU202201326

Figure 3.4:
Mapping WoS cited records and references from HTML full-texts



In this study, the location of an ITC was decided by counting the number of characters. Specifically, it is the number of characters between the starting character of an article's text body and the first character of this ITC. Sentence segmentation, which splits a full text into numerous sentences, was applied to collect the citances in an article. This study used spaCy 2.3.1, a python tool for NLP, to accomplish this task. In general, this study randomly chose two or three research articles for each journal and manually checked their sentence segmentation results. Some additional rules were used to refine the result of sentence segmentation to ensure correctness. As to POS tagging and lemmatization, another NLP toolkit, NLTK 3.5, was used for this processing. Totally, 688,879 citances and 21,152,967 tokens were identified and included in this study.

doi:10.6342/NTU202201326

3.2 Citation Relationships Measurement

In this study, the results of several different models, including the classical model, the frequency model, the distance model, the lexical model, and the semantic model, were compared to investigate the effectiveness of applying NLP semantic techniques in citation analysis. The classical model, which measures all references equally, was regarded as the baseline model. The semantic model, which measures citation relationships by considering the sentimental polarity or semantic similarity, was proposed by this study to modify the classical model. The other models included were used to see whether the results of semantic model provide a different perspective and outperform these modified models. The following sub-sections detail the algorithm of each model.

3.2.1 Classical model

The classical model represents the general ways to measure the citation relationships in practice. The formulas used to measure three kinds of citation relationships in this model are shown as follows.

$$DC_{i,k} = \begin{cases} 1 & \text{, if } i \text{ uses } k \text{ as its reference} \\ 0 & \text{, otherwise} \end{cases} \tag{3.1}$$

$$BC_{i,j} = \sum_{k=1}^{d} DC_{i,k} \times DC_{j,k} \tag{3.2}$$

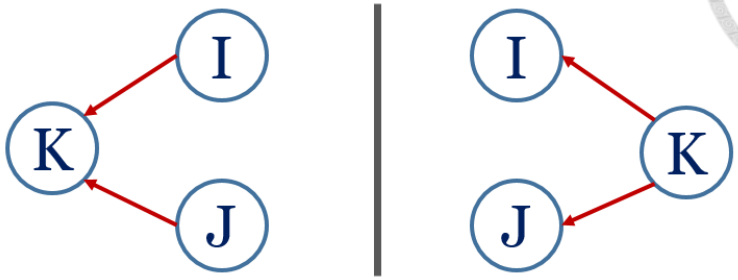
$$CC_{i,j} = \sum_{k=1}^{d'} DC_{k,i} \times DC_{k,j} \tag{3.3}$$

i, j, and k represent three different citation entities, namely three different works in this study. $DC_{i,k}$ indicate whether work i cites work k. $BC_{i,j}$ and $CC_{i,j}$, two relationships between work i and j, can be decided by DC. Figure 3-5 shows the basic concept of the classical model. d represents the number of the distinct references cited by work i or j,

doi:10.6342/NTU202201326

and d' is the number of distinct works citing work i or j.

Figure 3.5:
DC, BC, and CC (Classical Model)



Notes: In these graphs, the arrow's direction shows which works cite which. For example, in the left graph, i and j cite k. $DC_{k,i}$ in the left and right graphs are 0 and 1, respectively. The left and right graphs also show $BC_{i,j}$ and $CC_{i,j}$, respectively. In this example, $BC_{i,j}$ and $DC_{i,j}$ are 1.

3.2.2 Frequency model

The studies reviewed in Section 2.2.5 and Section 2.3 shows that the frequency of reference mention can differentiate the importance of citations. Hence, the present study defined the frequency model, which modifies the results of the classical model by weighting the citation relationships based on the frequency of reference mention instead of counting them equally. Compared with the classical DC, shown in Formula 3-1, the frequency DC (FDC) is decided by the number of a work mentioned by another work in its text body. As to frequency BC (FBC) and frequency CC (FCC), the present study gauges them by formula 3-5 and 3-6.

$$FDC_{i,k} = \begin{cases} x & \text{, if } i \text{ mentions } k \ x \text{ times.} \\ 1 & \text{, if } k \text{ is only existed in references.} \\ 0 & \text{, otherwise} \end{cases} \tag{3.4}$$

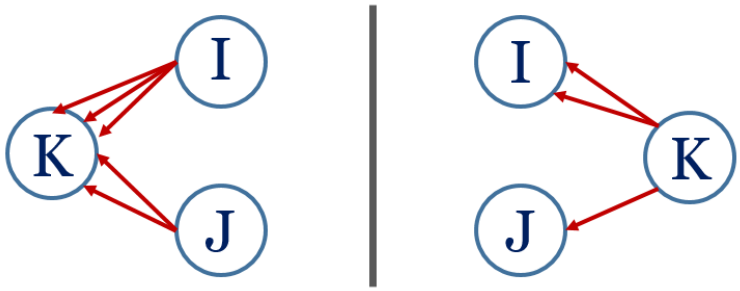
$$FBC_{i,j} = \sum_{k=1}^{d} \min\left(FDC_{i,k}, FDC_{j,k}\right) \tag{3.5}$$

$$FCC_{i,j} = \sum_{k=1}^{d'} \min\left(FDC_{k,i}, FDC_{k,j}\right) \tag{3.6}$$

doi:10.6342/NTU202201326

FBC is decided by accumulating the citation strength of all their distinct references of work i and j. A similar way also calculates FCC. Figure 3-6 represents the basic concept of the frequency model.

Figure 3.6:
FDC, FBC, and FCC (Frequency Model)



Notes: The arrow's direction shows which works cite which. $FDC_{k,i}$ in the left and right graphs are 0 and 2 separately, and the left and right graphs show $FBC_{i,j}$ and $FCC_{i,j}$ respectively. In this example, $FBC_{i,j}$ is 2, and $FDC_{i,j}$ is 1..

3.2.3 Distance model

As reviewed in Section 2.3.2, the distance model assumes that the citation strength is inversely related to the distance between two ITCs. The farther two ITCs are, the less their citation strength will be. The present study applied the method of the fifth method tested in Eto (2012) with different normalization way, generated the result of distance model, and compared it with other models. The algorithms used to measure distance CC (DCC) between work i and j are shown as bellows.

$$DCC_{i,j} = \frac{\sum_{k=1}^{d'} DistanceSim\left(C_{k,i}, C_{k,j}\right)}{\sqrt{F_i \times F_j}} \tag{3.7}$$

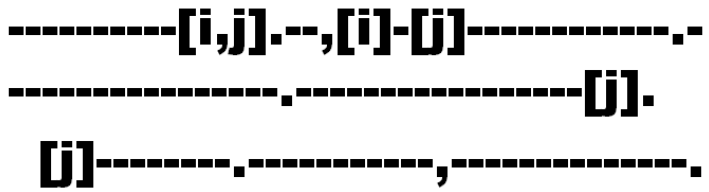
$$DistanceSim\left(C_{k,i}, C_{k,j}\right) = \sum_{x \in C_{k,i}} \sum_{y \in C_{k,j}} weight\left(x, y\right) \tag{3.8}$$

$$weight\left(x, y\right) = \begin{cases} 4 & \text{, if } x \text{ and } y \text{ is enumeration.} \\ 3 & \text{, if } x \text{ and } y \text{ is in the same sentence.} \\ 2 & \text{, if } x \text{ and } y \text{ is in the same paragraph.} \\ 1 & \text{, if } x \text{ and } y \text{ is across paragraphs.} \end{cases} \tag{3.9}$$

doi:10.6342/NTU202201326

According to Eto (2012), he defined four types of co-citations, please refer to Figure 2-5, and weighted them in different values, shown as Formula 3-9. $C_{k,i}$ is a set which contains all locations where authors cite i in k. F_i is the total number which i was mentioned in all works' text body. Figure 3-7 is an example of calculating DCC. Given that there is no reasonable way to use this method in gauging DC and BC, the distance model is applied in measuring CC in the present study.

Figure 3.7:
Example of DCC



Notes: In this example, a work cites i and j for 2 and 4 times, respectively. Hence, there are 8 combinations of (i, j), including an enumeration, one same-sentence, four same-paragraphs, and two across-paragraphs co-citations. If a dataset only includes this work, $DCC_{i,j}$ will be $(4 + 3 + 2 \times 4 + 1 \times 2) / (2 \times 4) = 2.125$.

3.2.4 Lexical model

The lexical model applied in this study is based on the proposition of Jeong et al. (2014), which measures the CCS of two authors based on cosine similarity between the citances citing their works. Figure 2-9 provides an example of this model. The present study determines the lexical CC (LCC) of two references i and j based on the cosine similarity between the citances citing them. As to the lexical BC (LBC) of two works i and j, the value is decided by the sum of the cosine similarity between the citances of their common references. The $LBC_{i,j}$ given by reference k is the maximum cosine similarity between the m and n citances of this reference in work i and j, respectively. Likewise, the $LCC_{i,j}$ decided by the work k is the maximum cosine similarity between the m and n citances regarding the references i and j in work k. The final LCC or LBC between the works i and j are the sum, shown as Formula 3-11 and Formula 3-12.

doi:10.6342/NTU202201326

$$F\text{”}(s_x, s_y) = consineSimilarity(s_x, s_y) = \frac{\sum_{w=1} n_{xw} n_{yw}}{\sqrt{\sum_{w=1} n_{xw}^2 \sum_{w=1} n_{yw}^2}} \quad (3.10)$$

$$LBC_{i,j} = \sum_{k,m,n} \max\left(F\text{”}(s_{i,k,m}, s_{j,k,n})\right) \quad (3.11)$$

$$LCC_{i,j} = \sum_{k,m,n} \max\left(F\text{”}(s_{k,i,m}, s_{k,j,n})\right) \quad (3.12)$$

Formula 3-9 is used to decide the cosine similarity of two citances x and y. n_{xw} and n_{yw} represent the frequency of a distinct word w in x and y, respectively. Note that not all words were included in this study. Before counting cosine similarity, the texts used to identify its cited references, e.g., Jeong et al. (2014), were removed first, and the stopwords were excluded then. All remaining words were stemmed by Snowball Stemmer. Besides, only BC and CC were examined in the lexical model.

3.2.5 Semantic model

As mentioned in Section 2.4, the relationships between words can be defined by three models: the thesaurus model, the count model, and the predict model. The present study measures the sentimental polarity of a citances and the semantic similarity between two citances by the thesaurus and predict models. The result of the sentimental polarity was used to classify citations into three categories: positive, negative, and neutral citations. The current study investigated how the negative citations affect DC citation network by comparing the DC network of the classical models with the DC network without the negative citations. The result of the semantic similarity was used to adjust the BCS and CCS. The current study explored the differences between BC and CC citation networks based on the classical and Wordnet/BERT models.

Thesaurus Model

NLTK package was used to implement the thesaurus model. As mentioned before, NLTK provides researchers with a general way to utilize Wordnet. Therefore, in the

doi:10.6342/NTU202201326

following, the present study names it the Wordnet model. Sentiment Intensity Analyzer (SIA) and Path Similarity, two NLTK built-in functions based on Wordnet, were used to identify polarity and gauge similarity.

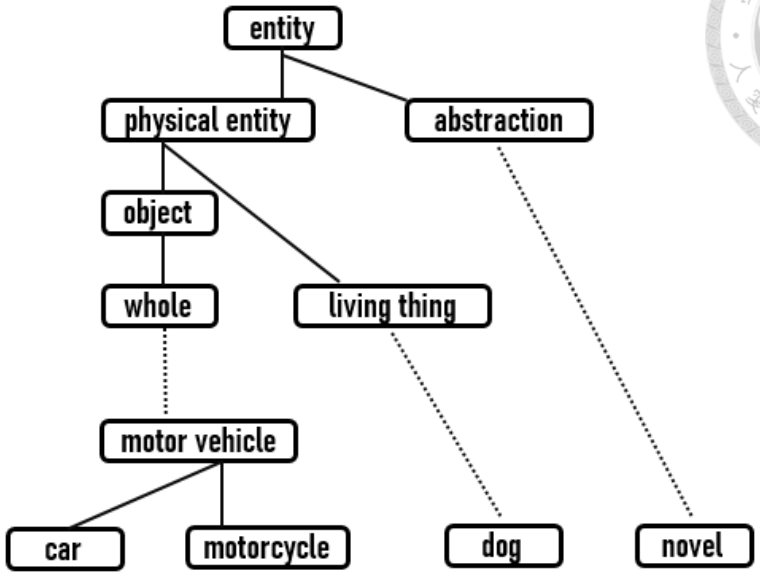
SIA provided by NLTK package is a function of Valence Aware Dictionary for Sentiment Reasoning, a model for text sentiment analysis. It bases on a dictionary that records lexical features and sentiment scores. SIA receives a string and returns fours scores, including negative, neutral, positive, and compound scores. In this study, the compound score, which is in the range -1 (very negative) to 1 (very positive), was used to classify a citance as positive, negative, or neutral. Athar (2011) has provided a corpus consisting of 8,736 citances with sentimental polarity labels annotated manually. The thresholds for the sentimental polarity were determined by comparing these manual annotations and their compound scores. After examining several compositions, the present study used the following criteria. When the score of a citance was above 0.628, it was classified as positive. A citance was negative if the score was lower than -0.278. Based on the criteria, the Micro-F1 of the results about classifying the corpus of Athar (2011) by SIA was 0.7719.

Wordnet records several word relationships, including hypernyms, hyponyms, and synonyms. Another function, Path Similarity, returns "a score denoting how similar two word senses are, based on the shortest path that connects the senses in the is-a (hypernym/hyponym) taxonomy" (Wordnet Interface, n.d.). Figure 3-7 is an example that shows how to gauge two words ' path similarity. The path similarity is between 0 (dissimilarity) and 1 (identity).

Based on the path similarity, the present study measured BCS and CCS by the following formulas. Formula 3-12 and Formula 3-13 were modified from the content-based ACA proposed by Jeong et al. (2014) and measured BCS/CCS based on the maximum path similarity between two citances. BCS between works i and j is the sum of the strength of all common references. BCS given by a common reference k is the maximum citance similarity between its citances in works i and j. The CCS is

doi:10.6342/NTU202201326

Figure 3.8:
Example of Path Similarity



Notes: Redraw based on Saitoh (2019). The common path of 'car' and 'motorcycle' is longer than that of 'car' and 'dog'. Hence, the similarity between 'car' and 'motorcycle' is higher.

determined in a similar vein. The only difference is that CCS is decided by the work k citing both references i and j.

$$WNBC_{i,j} = \sum_{k,m,n} \max\left(CitSim\left(s_{i,k,m}, s_{j,k,n}\right)\right) \tag{3.13}$$

$$WNCC_{i,j} = \sum_{k,m,n} \max\left(CitSim\left(s_{k,i,m}, s_{k,j,n}\right)\right) \tag{3.14}$$

$$CitSim_{s_x,s_y} = \sum_{u=1} TokenSim\left(t_{xu}, s_y\right) + \sum_{u=1} TokenSim\left(t_{yu}, s_x\right) \tag{3.15}$$

$$TokenSim\left(t, s\right) = \max\left(PathSimilarity\left(t, t_s\right), \forall t \in s\right) \tag{3.16}$$

where s represents the citance. $S_{i,k,m}$ indicates the m^{th} citance which i cites work k. s_x represents the collections of tokens in sentence x. t_{xu} is the u^{th} token of s_x .

The citance similarity between two citances is the normalized sum of token similarity between the tokens of two citances, shown in Formula 3-14. Formula 3-15 shows how to calculate token similarity. Only nouns, except stopwords, are used to calculate token similarity in the present study because nouns usually represent the research topics (Hsiao & Chen, 2018). For a token in a citance, path similarity is used to decide the score between it and all tokens in another citance, and the maximum score

doi:10.6342/NTU202201326

is the token similarity.

Predict Model

The present study implemented the predict model using two open-source projects based on BERT. Therefore, the BERT model is used to call it in the following. BERT project released in GitHub was used to classify a citance as neutral, positive, or negative (Devlin et al., 2018). The released project provides an available pre-trained model and python codes of classification task. The pre-trained model used in this study was BERT-Base, Uncased. The pre-trained model was fine-tuned with the corpus of Athar (2011). After testing different combinations of learning rate and epochs, the present study set learning rate and epochs as $1e^{-5}$ and 3, respectively. The accuracy of the fine-tuned classifier was 89.35%. The Micro-F1 was 0.8483. The classifier received a citance and reported the possibility of being positive, neutral, and negative. All citances were classified according to which category has the highest possibility.

As to the meaning of sentences, the present study utilized Sentence-Transformers, which provides dense vector representation for sentences based on transformer networks like BERT (Reimers & Gurevych, 2019), to get the vector representation for the citances. The cosine similarity was applied to gauge the semantic similarity between two citances. Based on the results of cosine similarity, the BC and CC based on sentence embedding were decided by Formula 3-16 and Formula 3-17, respectively.

$$BertBC_{i,j} = \sum_{k,m,n} \max\left(CosineSimilarity\left(v_{i,k,m}, v_{j,k,n}\right)\right) \tag{3.17}$$

$$BertCC_{i,j} = \sum_{k,m,n} \max\left(CosineSimilarity\left(v_{k,i,m}, v_{k,j,n}\right)\right) \tag{3.18}$$

where $v_{i,k,m}$ represents the sentence vector of the m^{th} citance citing k in i.

In sum, the present study included six models. Some limitations prevented the implementation of the lexical and distance models when measuring DC and BC. Table 3-2 reports the types of citation relationships implemented by the models.

doi:10.6342/NTU202201326

Table 3.2:
The types of citation relationships, the models, and their implementations

Relationship Type	DC	BC	CC	Implementation
Classical Model	V	V	V	General way to measure citation relationships
Frequency Model	V	V	V	Citation strength modified by the frequency of citations
Distance Model			V	Citation strength modified by the distance between two citations
Lexical Model		V	V	Citation strength modified by the lexical similarity between citations' citances
BERT Model	V	V	V	When applying to DC, identifying citations' sentimental polarity and removing the negative citations. When applying to BC and CC, weighting BCS and CCS by the semantic similarity.
Wordnet Model	V	V	V	When applying to DC, identifying citations' sentimental polarity and removing the negative citations. When applying to BC and CC, weighting BCS and CCS by the semantic similarity

3.3 Citation Network Analysis

After measuring citation relationships by these models, the results are used to build the citation networks of six sliding periods: P1 (2010~2014), P2 (2011~2015), P3 (2012~2016), P4 (2013~2017), P5 (2014~2018), and P6 (2015~2019). The design of a 5-year sliding window can reveal how LIS evolved in these ten years and prevent the possible misleading due to only treating the ten years as one or splitting it into two periods. For each period, the present study compares the results of different models at three levels. Firstly, the comparison at the network level investigated whether these models show different citation networks. Secondly, the examinations at the cluster level explored whether the subfields identified in these networks differ and which model provides a better clustering result. Finally, at the node level, the present study examined the critical entities or relationships identified in these networks and answered whether the semantic model efficiently reveals influential entities and relationships. According to these results, this study discusses and concludes the effectiveness of applying the semantic model in analyzing citation relationships.

doi:10.6342/NTU202201326

3.3.1 Network level

Building whole network and core network

The analysis at the network level is composed of two parts: the whole network and the core network. The whole network is the network of the six periods defined above. In most citation analysis studies, researchers usually remove minor citation entities, which may not relate with enough other citation entities, or weak citation relationships, whose citation strength is not high enough. The whole network provides a general aero-view for the inter-relationships between all works included, and the core network delineates how the influential works interact with each other. Both of them play a crucial role when conducting citation analysis. Therefore, the present study examines both the whole and core networks and defines the core networks of three kinds of citation relationships with the following procedures.

Common Procedures Firstly, the minor citation entities were identified and removed. After removing the minor entities, the citation relationship whose vertex connects to one of the minor entities would be removed accordingly. Then, the weak citation relationships would be identified and removed. After these procedures, the remaining citation entities and citation relationships formed the core networks.

Direct citation The citation entities whose in-degree centrality is higher than 95% of the nodes in the whole network were included as the nodes in the core network first. Then, the nodes pointing to these high in-degree nodes were also included as part of the core network. The other nodes were categorized as minor entities and excluded. Except for DC networks based on the classical model, the weak relationships were the citation relationships whose weighted strength is one.

Bibliographic coupling The citation entities whose degree centrality was higher than 80% of the nodes in the whole network were included as the nodes in the core network, and the others were minor nodes. The minor nodes and the edges connected to them would be removed first. Then, the citation relationships between these included nodes were normalized to Pearson correlation coefficients. If r is less than 0.6, a citation relationship was defined as weak and removed.

doi:10.6342/NTU202201326

Cocitation The citation entities whose degree centrality was higher than 95% of nodes in the networks were included as the nodes in the core network, and the others were minor nodes. The other procedures were the same as those of bibliographic coupling.

Calculating network indicators

All networks were viewed as undirected networks, and six network indicators provided a quick brief of these networks, including:

1. The number of nodes.
2. The number of edges.
3. Network density.

   The density for undirected graphs is defined as

   $$d = \frac{2m}{n\,(n-1)} \tag{3.19}$$

   where m and n represent the number of edges and nodes, respectively.
4. The number of connected components.

   A connected component is a subgraph in which each pair of nodes is connected via a path. The number of connected components equals the separate parts in a network.
5. Transitivity.

   The transitivity is the ratio of the number of triangles to the number of triads, two edges with a shared vertex. The formula is

   $$T = 3\frac{\#trangles}{\#triads} \tag{3.20}$$

6. Average clustering coefficient.

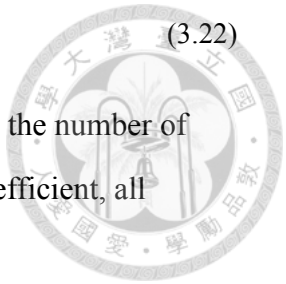
   The average clustering coefficient C is the average of all nodes' clustering coefficients, the ratio of the number of triangles to the number of all possible triangles through a node.

   $$c_v = \frac{2T\,(v)}{deg\,(v)-1} \tag{3.21}$$

doi:10.6342/NTU202201326

$$C = \frac{1}{n} \sum_{v \in G} c_v \tag{3.22}$$

$T(v)$ is the number of triangles through node v, and $deg(v)$ is the number of edges connected to v. When calculating average clustering coefficient, all networks are transferred as un-directed networks.



3.3.2 Cluster level

After examining the network indicators and comparing their differences, the present study investigates the clusters in the core networks to answer research question two. The clustering results were decided by the algorithm based on modularity, implemented by NetworkX, a free Python library. Modularity is a clustering method proposed by Newman (2004). The present study used this algorithm because it can effectively deal with large and sparse networks, which are the features of the citation networks. This algorithm also has been used by other studies which investigated the performance of different citation relationships in various applications, like Shibata et al. (2009).

Indicators of clusters

Several indicators were used to analyze the clustering results:

1. The number of the clusters
2. The number of the clusters containing only one node
3. The number of nodes of the top-n clusters
4. Adjusted Rand Index (ARI)

Indicators 1~3 provided the sketch of a clustering result. ARI is an index for measuring the similarity between two clustering results. It is modified from Rand index, which considers all pairs of samples and counts the number of pairs assigned in the same clusters in two clustering results. ARI is the corrected-for-chance version of RI by considering the expected RI. The version implemented by scikit-learn 1.0.1 was used to get ARI in this study. Then, the present study further evaluated the clustering results of different models by two methods: textual coherence and subject analysis.

doi:10.6342/NTU202201326

Textual coherence

In general, the clusters are taken as subfields. Therefore, the citation entities of the same cluster should relate to similar topics, and their title words should be more concentrated. Boyack et al. (2011) and Chandrasekharan et al. (2020) applied textual coherence to measure the degree of concentration. Textual coherence is based on JSD "calculated for each document from the word probability vector for that document, and from the word probability vector for the cluster in which the document resides" (Boyack et al. 2011, p.5).

According to Boyack et al. (2011), the JSD value is proportional to how the sets of words in documents of the same cluster differ from each other. Namely, more divergence of the set of words in documents of the same cluster results in a higher JSD value. In this study, the cluster's word probability vector was the word frequency of the title words in this cluster, and the word probability of each work was decided by its title word frequency. The title words were lemmatized, and the stop words were removed. Formula 3-22 and Formula 3-23 detail how to calculate textual coherence with JSD value.

$$Coh_i = JSD\left(random\right)_i - JSD\left(actual\right)_i \tag{3.23}$$

$$Coh = \frac{\sum\left(n_i Coh_i\right)}{\sum n_i} \tag{3.24}$$

Coh_i is the difference between $JSD(random)_i$ and $JSD(actual)_i$. Both $JSD(random)_i$ and $JSD(actual)_i$ are the average JSD of a cluster's word probability vector and its works' word probability vectors. $JSD(actual)_i$ was the average JSD of the actual clusters with i records whose title is available in WoS records or CrossRef data. Because the title of each work in a cluster was not necessarily available, i might not be equal to the size of the actual cluster. For each $JSD(actual)_i$, the present study randomly chose i articles to form a random cluster and calculated the JSD of a random cluster. The procedures were iterated 50 times, and the average JSD was $JSD(random)$.

As mentioned above, more divergence produces a higher JSD value. Hence,

doi:10.6342/NTU202201326

$JSD(random)_i$ is usually higher than $JSD(actual)_i$, and the Coh_i is proportional to how the sets of words in works of an actual cluster are in common. To reduce the possible bias due to small x, only the clusters in which x was no less than 20 were used to calculate Coh_i. In addition, clusters of small size have an advantage in this indicator. Hence, this study compared Coh_i based on all clusters whose cluster sizes were no less than 20 and examined Coh_i based on the clusters with three other ranges: 20~40, 20~60, and 20~100. This design aimed to ease the possible bias that some models may generate more small clusters in their core networks. The weighted average of Coh_i was the textual coherence of a model, namely Coh. Given that the comparisons are between the clustering results of the same citation relationship, period, and clustering algorithm, the divergences between the Coh of different models should be due to the models used to measure the citation relationship strength.

Subject analysis

Additionally, this study analyzed the subfields of the several largest clusters in each model. Instead of using keywords, the present study chose titles because titles are necessary for scholarly publications, but not every article has keywords. Title usually shows the main topic of this work and highly related to this topic. The following procedures were used to reduce the term variation between different titles. For each article, the words extracted from their titles, after lemmatizing and removing stop words, formed their topic words. The topic words were used to identify the subfields of the clusters. In general, the top 20 topic words in a cluster were used to decide its subfield.

Given that the purposes of this study are evaluating the effectiveness of the BERT and Wordnet models, the subfields identified from the core networks which belong in the same period and type of citation relationship were compared with each other. The research trends concluded from these subfields were compared with the six trends reported in Hsiao and Chen (2020) to check each model's ability to identify the research trends. The six trends include scholarly communication and scientometrics (SCS), information behavior and information retrieval (IBIR), applications of

doi:10.6342/NTU202201326

technology (AoT), library services and management (LS), health information and technology (HIT), and computer science techniques (CS).

In addition, the emergence and evolution of these largest subjects were used to analyze each model's ability to identify subfields. In the DC/BC/CC core networks, the present study examined 5/10/15 largest clusters, respectively. The number was decided by the number of clusters in the core networks and the size of these clusters. In addition to exploring the subfields revealed by these clusters to evaluate each model's clustering results, this study further compared the subfields of the twin cluster, two clusters in which the articles of one cluster in a former period make up a large proportion of those of another cluster in the following period, and investigated whether the subfields were consistent. The examination of the twin clusters was also used to evaluate these models. This study assumes that the two subfields of the twin clusters should be highly similar due to their common articles. If two subfields are dissimilar, the including articles may be too diverse to concentrate on one close related topic.

3.3.3 Node Level

The analysis at the node level aims to answer whether applying semantic analytics helps identify influential citation entities or relevant citation relationships.

Direct Citation

When examining the DC results, the focus is whether detecting sentimental polarity helps identify influential citation entities. In this study, both BERT and Wordnet models were used to decide the sentimental polarity of the citations. Based on the results of both models, a cited work would be classified as a strongly positive class, weakly positive class, and neutral class. When both BERT and Wordnet models classified one of the citances of a cited work as positive, this cited work was strongly positive. When only one model classified one of its citances as positive, it was weakly positive. Otherwise, it was neutral.

To explore the relation between a work's influence and its sentimental class, the

doi:10.6342/NTU202201326

present study examined the sentimental types of source articles and their citation counts, namely DC. Instead of scrutinizing all references, this study focused on the source article to reduce the effects of including multiple kinds of works and publications. The source articles were divided into groups based on their publication year and journal-title. In each group, the present study compared the average citation counts of the source articles of three sentimental classes. Additionally, the average ITCs of the source articles were also examined. The average citation counts and average ITCs were used to decide the influence. The present study suggests that the average citation counts and average ITCs of the strongly positive class are the highest, and those of neutral class are the lowest.

Bibliographic Coupling/Co-citation

When examining the results of BC/CC, the focus is whether the modified models help emphasize the relevant citation relationships. The SciVal Topics provided by Scopus were used to decide the topic similarity of the two works. The topic similarity of two works is equal to the Jaccard similarity between their SciVal Topics. Given that BC can "group papers into small sub-groups such that each subgroup forms a valid subdivision of the parent group" (Kessler, 1963a, p. 49), the topic similarity of two works of a high BCS pair should be high. In addition, "the significance of strong co-citation links must rely both on the notion of subject similarity and on the association or co-occurrence of ideas" (Small, 1973, p. 268). Therefore, the topic similarity of two works of a high CCS pair should be high. By comparing the topic similarity of each model's top 100 pairs, this study answers whether BERT and Wordnet models can outperform other models in identifying relevant citation relationships.

doi:10.6342/NTU202201326

Chapter 4

Results and Discussions

The present study investigates the advantages and weaknesses of utilizing NLP semantic analysis techniques in citation analysis. In this chapter, Section 4.1 briefs the preliminary statistic related to the research data firstly. Section 4.2 presents the analysis result on the network level and answers the first research question. After examining the large clusters identified by modularity for each core network, whether these models reveal different subfields is reported. Additionally, the comparison between the clustering results tells the characteristics of different models and answers which model has the better ability to provide more relevant results. Section 4.3 details this comparison and discusses the analysis result on the cluster level. Section 4.4 reports and analyzes the important entities or relationships uncovered by each model. Finally, the present study discusses the analysis results of the three levels and concludes the advantages and weaknesses when applying semantic analysis in measuring citation relationships in the last part of this chapter.

doi:10.6342/NTU202201326

4.1 Brief Statistics

4.1.1 Articles, references, and in-text citations

According to the procedures described in Section 3.1, the present study collected 10,088 articles published in 15 journals between 2010 and 2019. Table 4-1 details the number of articles in each year by journals. Almost half of the articles, 4,358 exactly, are published by two monthly journals, JASIST and Scientometrics. Six bimonthly journals publish 3,067 articles, and the remaining 2,663 articles are published in quarterly journals. The number of articles published per year increases yearly. Till 2019, the number of published articles is 20% more than that of 2012.

Based on the procedures shown in Figures 3-3 and 3-4, the present study identified 447,984 references and 688,879 ITCs. The blue round dots in Figure 4-1 shows the average number of references (AVGR) per article; the orange triangle in the same figure indicates the average number of citancesper article. Generally, a journal's AVGR of 2019 was higher than that of 2010/2012. The amount of increase varied in different journals. Several journals' AVGR increased dramatically, e.g., *Library Hi Tech* and *Online Information Review*. However, the AVGR of some journals, such as *JASIST* and *Health Information and Libraries Journal*, had no significant increase. The trends of AVGR in most journals are similar to the results reported by Hsiao and Chen (2018).

From 688,879 ITCs, the present study identified 21,152,967 tokens. From 2010 to 2019, the average citance length, which is the average number of all citances' tokens in a journal, remained stable in most journals, between 26 and 31. Compared with the increase of AVGR in most journals, the average citance length did not show a significant upward or downward tendency during this period.

doi:10.6342/NTU202201326

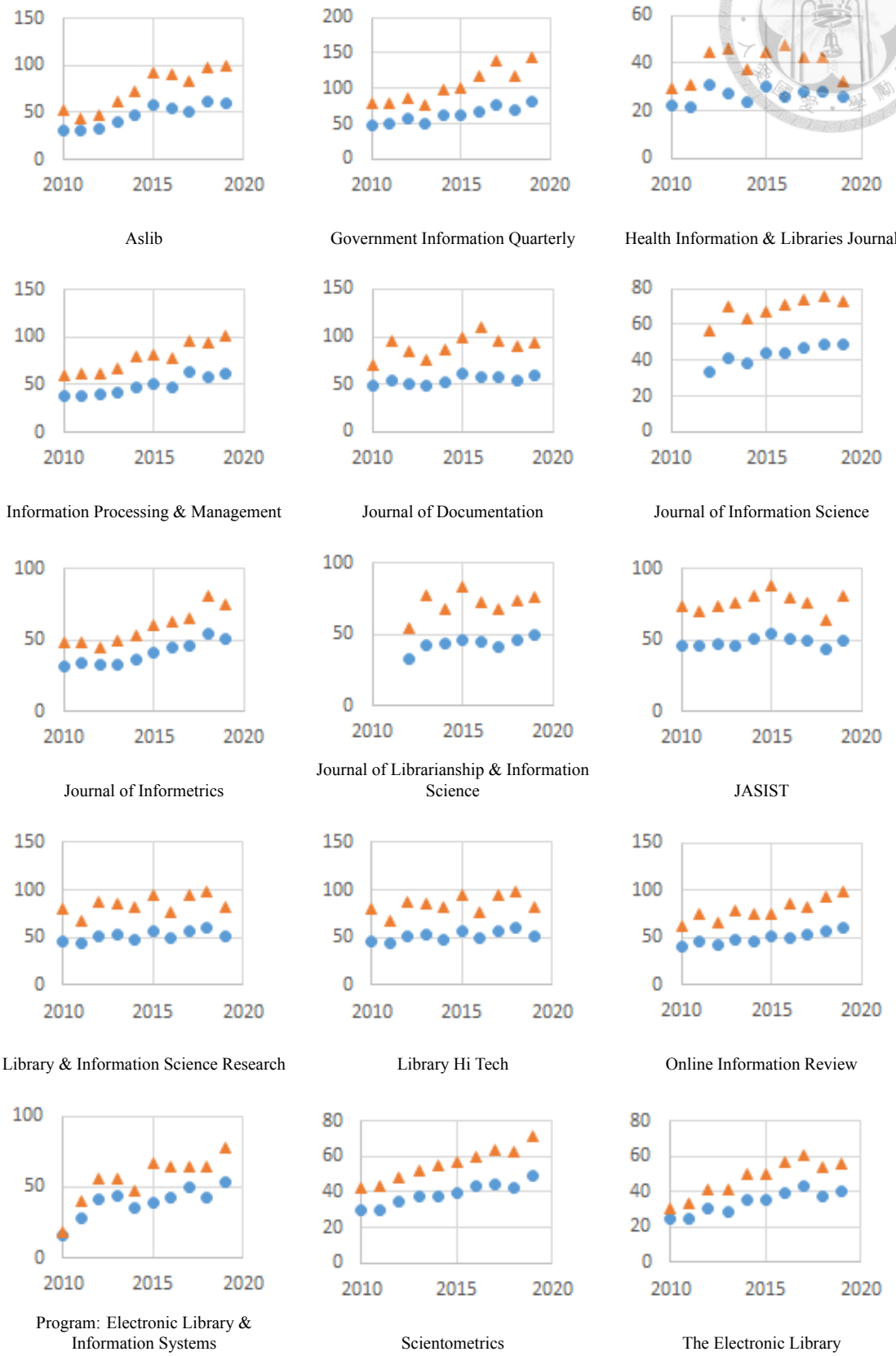
Table 4.1:
Including articles in each year by journals

Journal Title	Year										Total
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	
Aslib[2]	23	23	36	32	34	34	37	42	35	42	338
Government Information Quarterly	39	45	61	56	66	41	60	50	47	62	527
Health Information and Libraries Journal	35	28	23	22	19	21	19	17	24	18	226
Information Processing & Management[2]	55	61	76	87	44	60	70	67	70	137	727
JASIST[2]	163	174	165	175	179	181	206	188	112	100	1,643
Journal of Documentation[2]	35	39	38	36	52	54	55	61	65	73	508
Journal of Information Science[2]			38	55	59	55	40	51	50	48	396
Journal of Informetrics	55	57	66	91	82	78	76	71	74	63	713
Journal of Librarianship and Information Science			21	22	23	25	26	31	34	80	262
Library and Information Science Research	25	36	33	32	11	37	33	32	16	22	277
Library Hi Tech	41	50	41	41	42	40	38	39	37	53	422
Online Information Review[2]	48	46	49	46	51	48	58	57	67	62	532
Program: Electronic Library and Information Systems[2]	26	27	19	20	18	25	28	22	26	25	236
Scientometrics[1]	199	211	213	241	292	331	281	345	338	264	2,715
The Electronic Library[2]	56	50	48	48	51	65	56	68	64	60	566
Total	800	847	927	1,004	1,023	1,095	1,083	1,141	1,059	1,109	10,088

Notes: (1) Monthly journal. (2) Bimonthly journal.

doi:10.6342/NTU202201326

Figure 4.1:
The average number of references and ITCs at different journals



Notes: The x-axis is the publication year, and the y-axis is the average number of references and ITCs. The blue round dots and orange triangles represent the average number of reference and the average number of ITCs per article in each year.

doi:10.6342/NTU202201326

4.1.2 Citation relationships

The comparison of DC between different models

Figure 4-2 briefs the results of DC gauged by four models. The upper panel (A) reports the number of distinct references at different levels of citation counts. Each model had about 240,000 distinct references. The results show similar patterns and agree with a general notion: most references are only cited a few times, and the frequently cited references are rare. Almost 200,000 references were cited only once, and only around 50 articles were cited more than 100 times.

The lower panel (B) shows the number of distinct references at different levels of the weighted DC sum. Because the negative citations were not yet excluded in this stage, the distributions of the BERT and Wordnet models were almost identical. The weighted DC sum distribution of the frequency model was similar to that of the BERT and Wordnet models. However, the number of distinct references of the frequency model was higher because the corresponding citances of about 9,500 references were not found in the text body. Generally, the subgraphs in panel (B) show similar patterns, and the noticeable difference lies in the scale of the x-axis in the subgraph of the classical model because the frequency of ITCs is not considered in this model.

The comparison of BC between different models

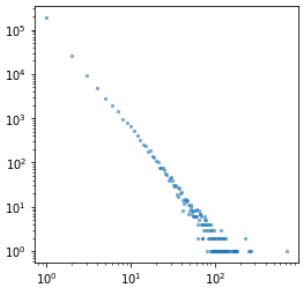
Figure 4-3 reports the BC results of five different models. The upper panel (A) shows the number of articles at different levels of BC number, which is how many articles are related to an article by BC, and the lower panel (B) represents the number of articles at different levels of the sum of BCS. The subgraphs in panel (A) present a similar pattern: the number of articles decreases as the BC number increases. Some slight distinctions existed between the results of the lexical model and the others. In the subgraph of the lexical model, the article numbers of the low BC number are higher, and its maximum sum of BCS is less than other models. It means that the lexical model measures BC more critically than other models.

doi:10.6342/NTU202201326

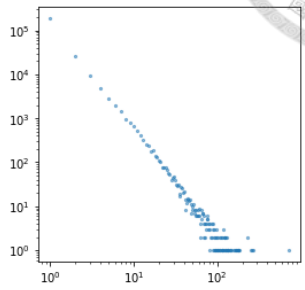
Figure 4.2:

Weighting scheme based on character offsets and centiles

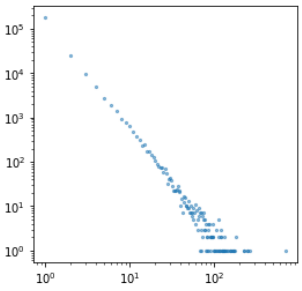
(A) The number of references cited by how many other articles



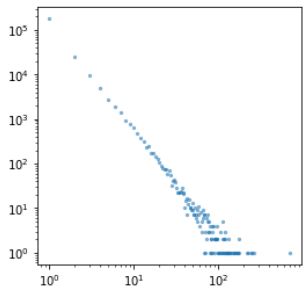
Classical Model



Frequency Model

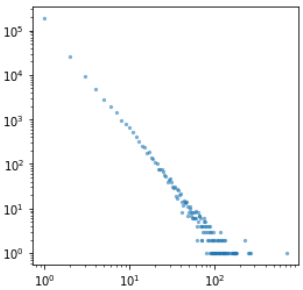


BERT Model

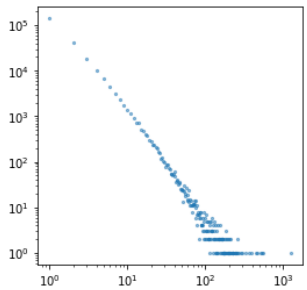


Wordnet Model

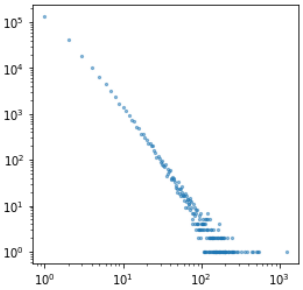
(B) The number of references at the different levels of the weighted DC sum



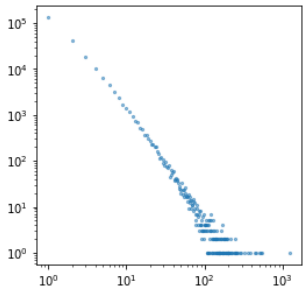
Classical Model



Frequency Model



BERT Model



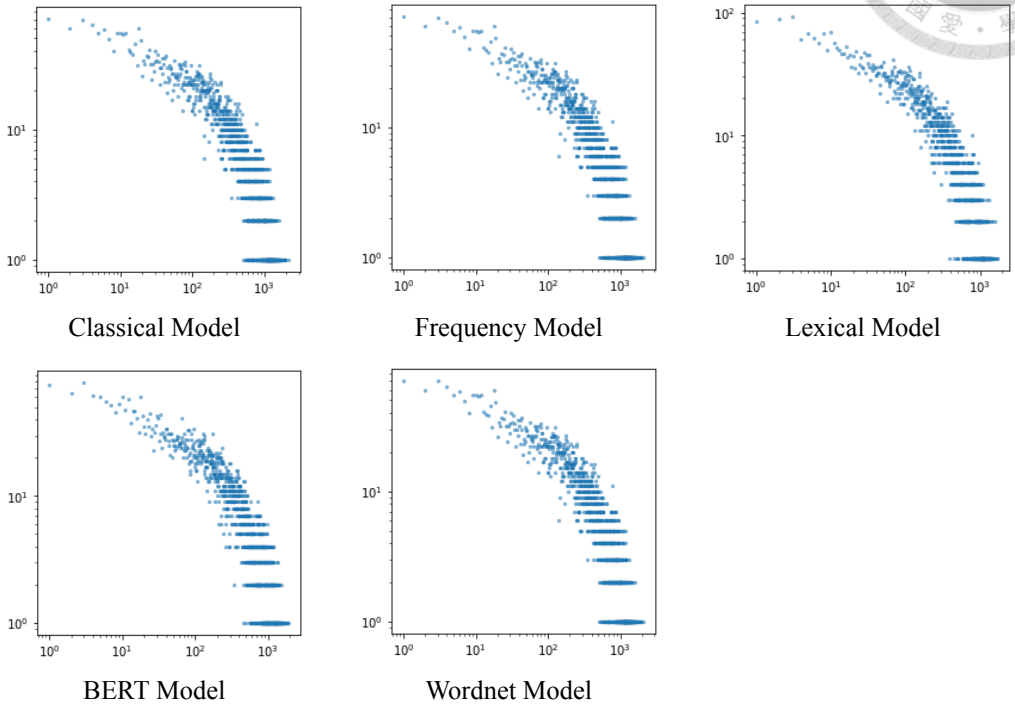
Wordnet Model

Notes: The y-axis is the number of nodes. The x-axis represents how many times a reference is cited by other articles in (A) and the sum of weighted direct citations in (B). Be noted that both axes are in log scale.

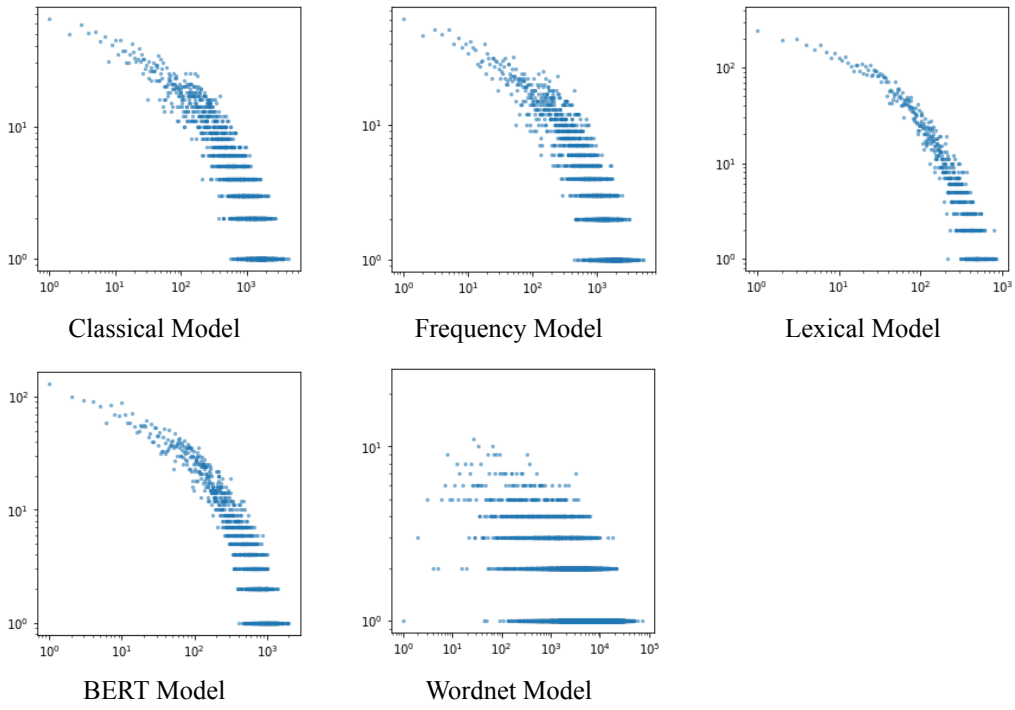
doi:10.6342/NTU202201326

Figure 4.3:
The distributions of references (BC)

(A) The number of articles at different levels of BC numbers



(B) The number of references at the different levels of the weighted DC sum



Notes: The y-axis is the number of nodes. The x-axis of (A) and (B) represent the level of BC numbers and the sum of BCS, respectively. Be noted that both axes are in log scale.

doi:10.6342/NTU202201326

In Panel (B), the decimal number of the sum of BCS was omitted to make the subgraphs concise. Some evident divergences existed in panel (B), especially the subgraph of the Wordnet model. Generally, the number of articles decreases as the sum of BCS increases. In the subgraph of the Wordnet model, however, there is only a tiny fraction, less than 1%, of articles whose sum of BCS is less than 100, and the BCS sum of almost 75% of articles is higher than 1,000. The reason is that the BCS of the Wordnet model is proportional to the length of citance. As mentioned in the previous section, the average numbers of the citances' tokens in different journals were about 26~31. It made the BSC of the Wordnet model larger than other models. The same reason also resulted in the rarity of articles with low BCS sum and the prevalence of articles with high BCS sum.

Besides, in the subgraph of the lexical model in panel (B), there were more than 200 articles whose sum of BCS was one, and the maximum sum of BCS was only close to 1,000. In other models, the numbers of articles whose sum of BCS equaled one were less than 150, and all of these models had the higher maximum sum of BCS, more than 2,000. The distribution shows that the lexical model has stricter criteria for identifying BC and giving low BCS. Besides, the distribution of the BERT model also reveals that this model is also less likely to give high BCS. Although the frequency model usually gives a higher BCS than the classical model due to considering the frequency of ITCs, there are more similarities between the distributions of the frequency and classical models compared to other models.

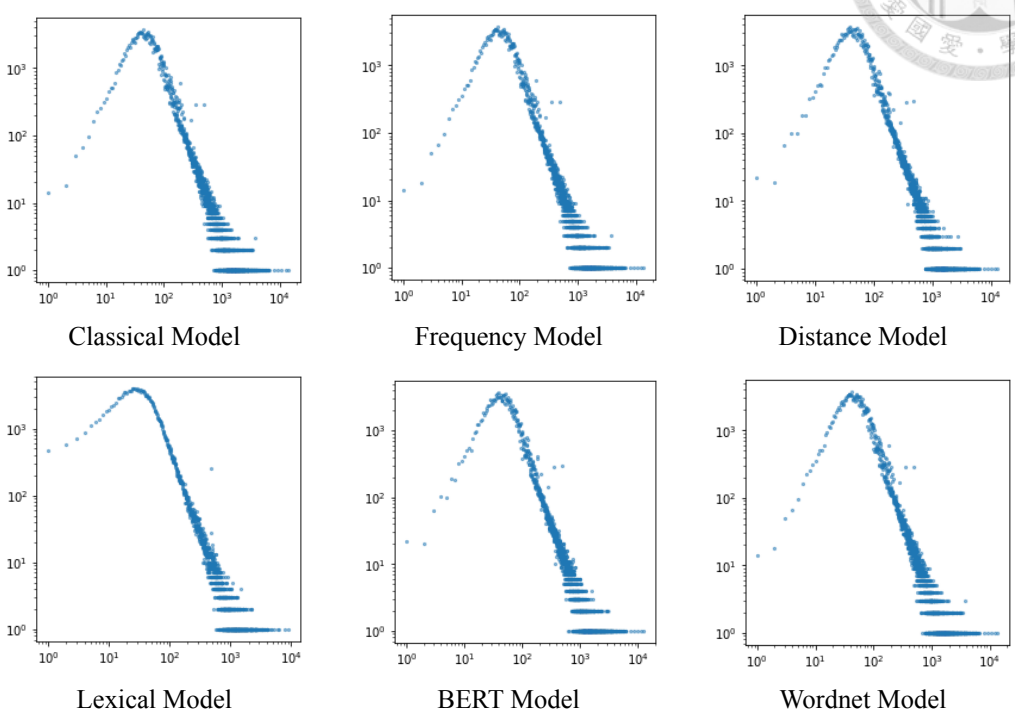
The comparison of CC between Different Models.

Figure 4-4 reports the CC results of six different models. The upper panel (A) shows how many references are at different levels of the number of co-cited references, and the lower panel (B) reports the number of references at different levels of the sum of CCS. Several outliers exist in the subgraphs of both panels. After examining the research data, the present study finds that three articles, published in Information Processing & Management in 2017 and 2019, cited more than 300 references. Hence, the works cited by the three articles are co-cited with hundreds of other works.

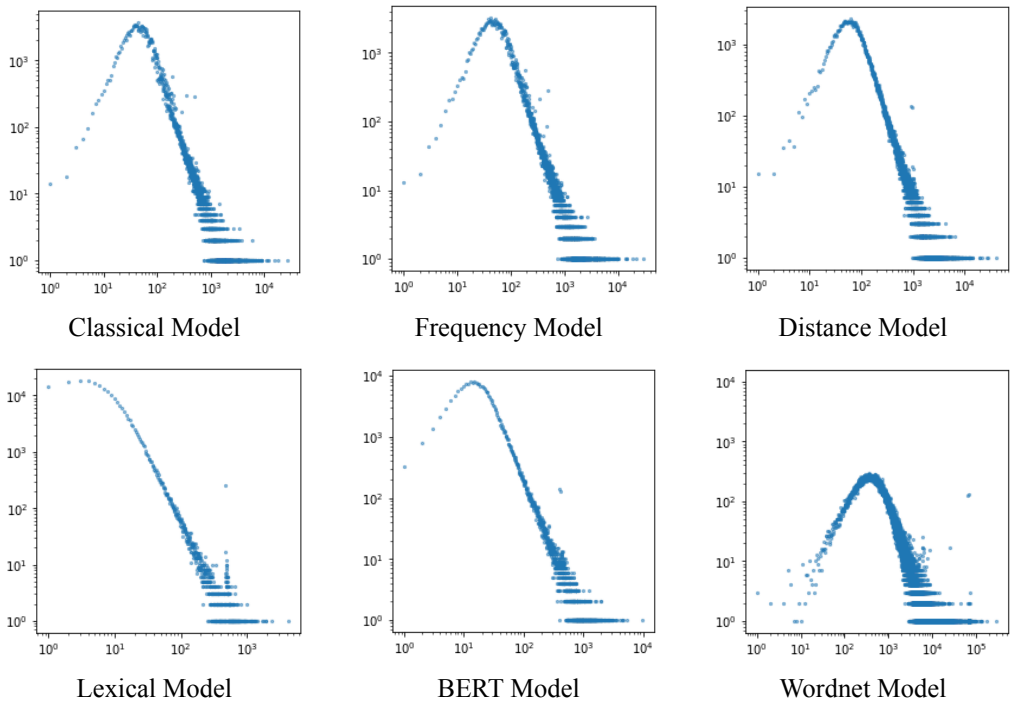
doi:10.6342/NTU202201326

Figure 4.4:
The distributions of references (CC)

(A) The number of references at different levels of CC numbers



(B) The number of references at different levels of the sum of CCS



Notes: The y-axis is the number of references. The x-axis of (A) shows how many references are at different levels of the number of co-cited references. The x-axis of (B) reports the number of references at different levels of the sum of CCS.

doi:10.6342/NTU202201326

Because their document types, categorized by WOS, are “Article”, they are still included in the following analysis.

The distributions’ peaks in the subgraphs of panel (A) are between 20 to 50, obviously different from DC and BC. As reported in Figure 4-1, the AVGR of each journal was higher than 20. According to the classical CC definition, the AVGR equals the average CCS of references cited by articles published by a journal. Except for the result of the lexical model, the distributions of other models show similar patterns, especially between the classical and frequency models.

In the subgraph of panel (B), the distribution of the sum of CCS in Wordnet model becomes flatter and broader because of the same reason mentioned in the paragraph regarding that of BCS in Wordnet model. Another model which shows a noticeable difference is lexical model. Its maximum value on the x-axis is much less, and that of the y-axis is much higher. It supposes that lexical model has a more critical standard in measuring citation relationships. Among the remaining four models, BERT model shows a little different pattern, and the others have akin distributions.

This section discusses the results, measured by various models, of three types of citation relationships. The present study analyzes their distributions and investigates the distinctions between them. Overall, the distributions meet the expectations. In addition, some models, like the lexical and Wordnet models, have already shown evident divergences from others. In the following section, several network indicators are used to examine the citation networks, including whole networks and core networks, of these models in different periods and answer the first research question: whether these models uncover different citation networks.

4.2 The Results of Network Analysis

The present study examined several network indicators on the network level, including the number of nodes, the number of edges, density, components, transitivity, and

doi:10.6342/NTU202201326

average clustering coefficient. Besides, JSD was used to compare the distributions of node degree between different modes. The scrutiny included the whole network and the core network, the sub-network of the whole network and composed of the nodes with high degree, and the results answer the first research question. The following text reports direct citation, bibliographic coupling, and co-citation finally.

4.2.1 Direct citation

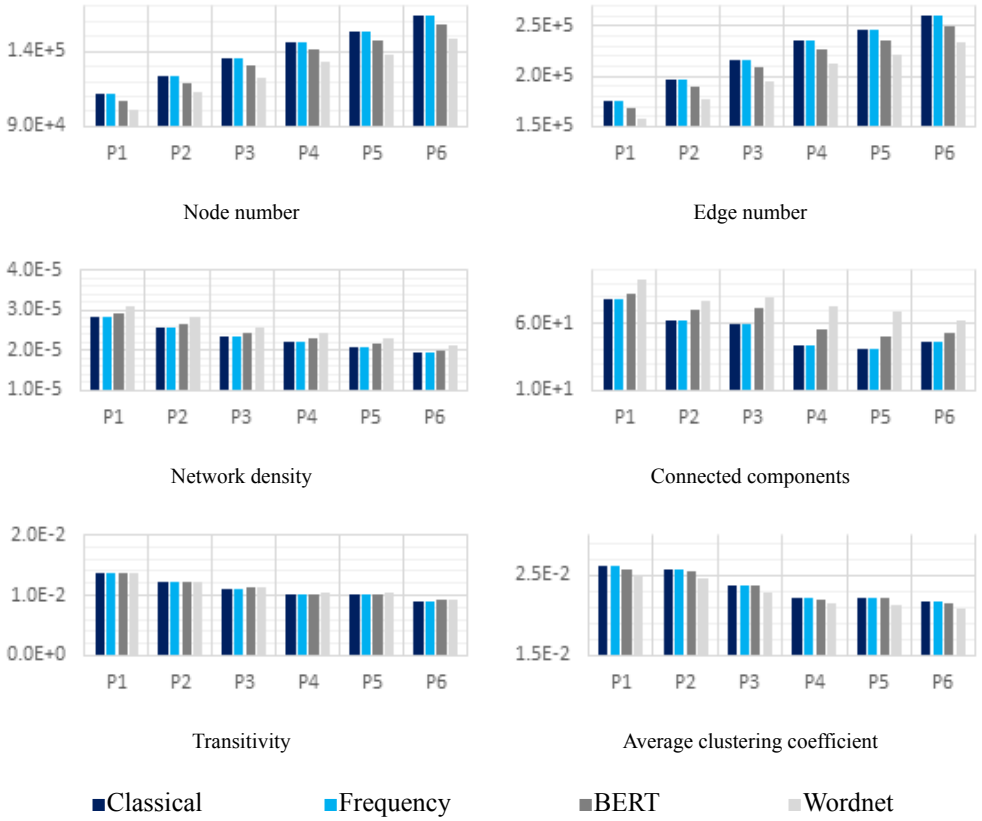
Figure 4-5 shows the statistics of the DC whole networks of four models in six periods. Accordingly, the divergences between these networks were slight. The classical and frequency models showed identical results in all indicators. In different periods, their node numbers were between 110,000~165,000, and edge numbers were between 170,000~270,000. The node numbers and edge number of the BERT and Wordnet models were slightly lower than that of the frequency model. Two reasons result in this phenomenon. One of the reasons is that not each reference has the corresponding ITCs in the text-body. In other words, authors may list some works in references without citing them in the text body. Another reason is the exclusion of negative citations when forming the whole network. As mentioned in Section 3.2.5, the present study identified the negative citation by using the sentiment score provided by SIA of NLTK and the possibility of sentimental polarity gauged by BERT. From P1 to P6, BERT model identified 141, 160, 191, 210, 215, and 248 negative citations, and the figures of Wordnet model were 10,775, 12,262, 13,429, 14,698, 15,323, and 16,403. These numbers also indicated the differences of node and edge numbers between the Wordnet and BERT models as well as explained its reason.

The divergences in networks density were so small that they could be ignored. The number of connected components showed that the networks revealed by the Wordnet model had more discrete parts, 14~29 and 9~18 more than the classical/frequency models and the BERT model, respectively. The numbers of transitivity of all models in the six periods were almost equal, and the number showed that only a tiny fraction,

doi:10.6342/NTU202201326

around 1%, of triads form the triangles. Similar patterns existed in the subgraph reporting average clustering coefficients. Overall, the differences among the whole networks were not noticeable and only occurred in the numbers of nodes, edges, and connected components.

Figure 4.5:
The statistics of the DC whole networks

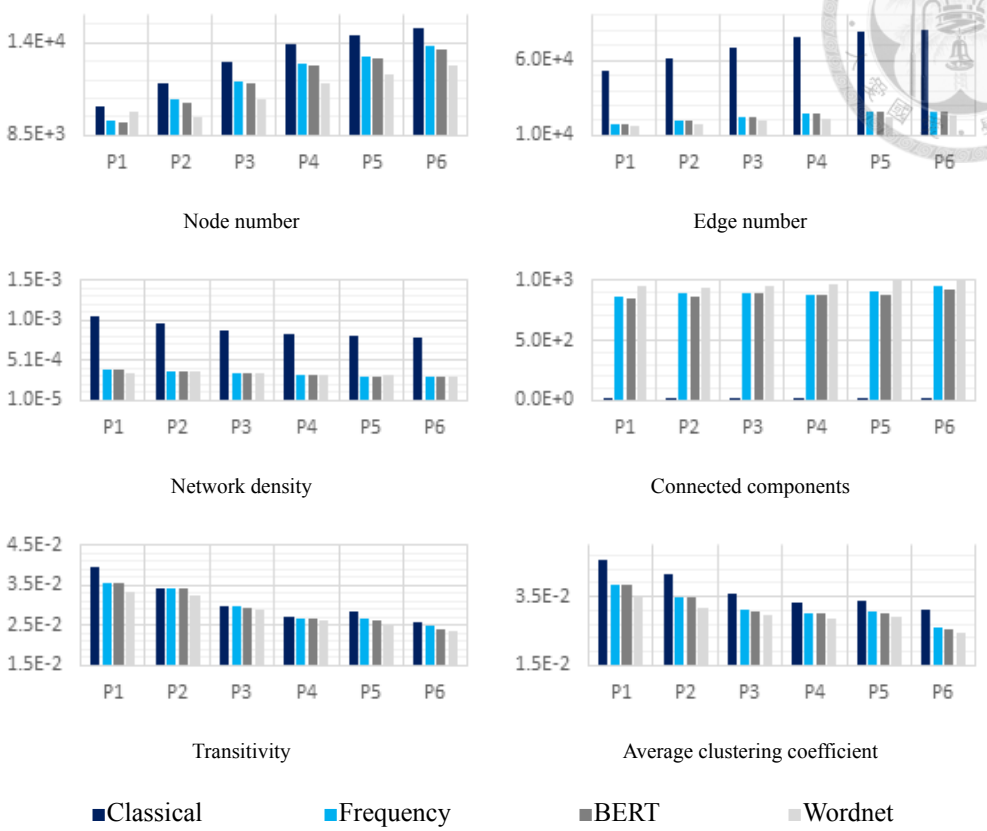


Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

Figure 4-6 shows the results of the DC core networks. The classical model did not discriminate citation relationships finely and included more nodes when the node degree was in a tie. Therefore, the indicators of the classical model showed that its core networks were different from other models. In the six periods, the node numbers of the classical model were 300~2,000 higher than other models. Besides, due to the inability to identify and exclude weak citation relationships, the core networks of the classical model had 35,000~55,000 edges more than other models.

doi:10.6342/NTU202201326

Figure 4.6:
The statistics of the DC core networks



Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

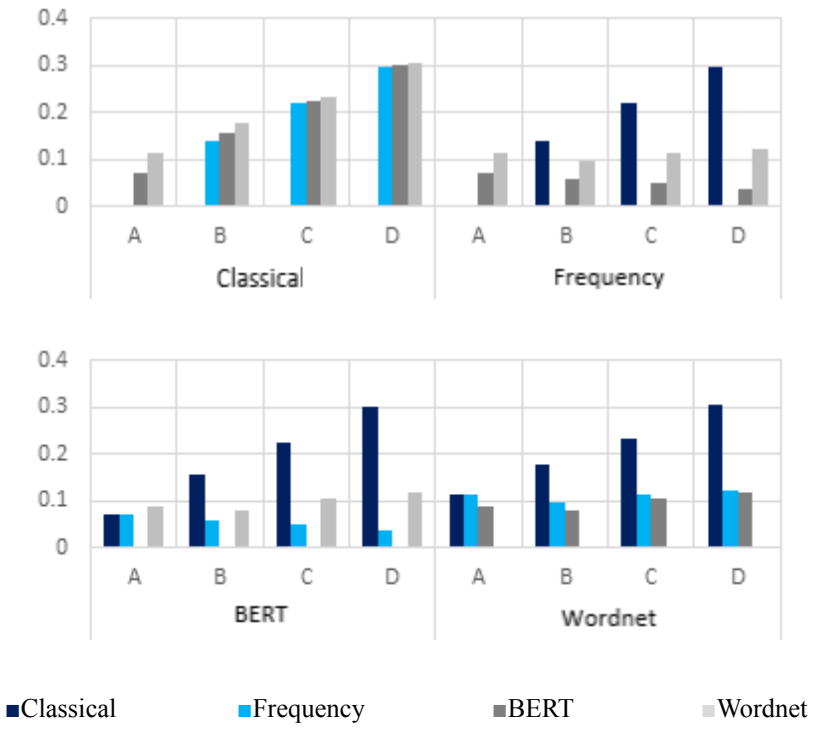
The divergences between the core networks of another three models were more minor. Because the numbers of negative citations identified by the BERT model were much less than the Wordnet model, the differences between the core networks of the frequency and BERT models were not evident. The node numbers and edge numbers of the Wordnet model were usually the lowest and ranged from 9,500 to 13,000 and 16,000 to 24,000, respectively. Also, the numbers of connected components in the core networks of the Wordnet model were the highest, between 940 and 1,040. Additionally, in all core networks of each model, the transitivity and average clustering coefficient of core networks were higher than whole networks. Such a result represented that the nodes of core networks are more likely to form triangles.

The four subgraphs of Figure 4-7 report the results of JSD, a method of measuring

doi:10.6342/NTU202201326

the similarity between two probability distributions. The present study transferred the distribution of node in-degree and weighted in-degree to the probability distributions and measured their similarity by JSD. The top-left graph reports the similarity between the distribution of the classical model and another three models. The graph shows that the core networks of the classical model were usually quite different from others, especially when comparing the core networks. JSD also indicated that the divergences between the distributions of the other three models were usually much minor, and the distributions of the Wordnet model were relatively different from the frequency and BERT models.

Figure 4.7:
The JSD between different models (DC)



Notes: Each bin represents the JSD result of a combination, including the model shown in the bottom of the sub-graph and the model indicated by the bin color. The labels on the x-axis mean the kinds of node degree distribution used to measure the JSD: (A) the node degree of the whole network; (B) the node weighted degree of the whole network; (C) the node degree of the core network; (D) the node weighted degree of the core network. The number on the y-axis is the JSD result.

Overall, the present study concludes that considering the ITCs will increase the divergence between the classical model and those considering ITCs. Although the tendency is unclear when examining the whole networks, the indicators and JSD regarding the core networks showed this pattern clearly. In other words, the core

doi:10.6342/NTU202201326

networks of the models considering ITCs differ from that of the classical model. However, the divergence between the frequency and BERT models was not clear. The main reason is that BERT model categorized rare citations as negative. The difference between the frequency and BERT models was affected mainly by the number of references without citances instead of sentiment analysis. The same reason also made the differences between the frequency and Wordnet models obscure.

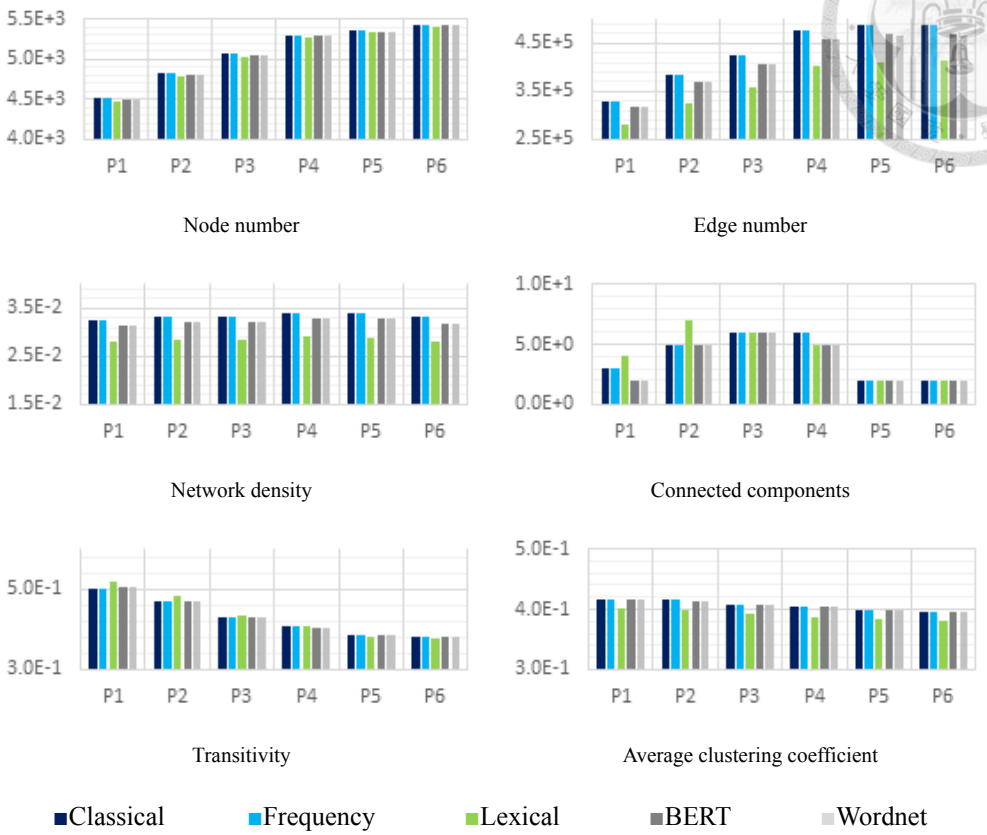
4.2.2 Bibliographic coupling

Figure 4-8 reports the indicators of the whole networks built by different models in each period. Based on these indicators, the five models can be divided into three groups: the classical/frequency models, the BERT/Wordnet models, and the lexical model. The indicators show that the whole networks of the models within the same group were similar. The differences in node numbers between these groups were minor, no more than 30. As to the number of edges, the classical/frequency models identified 12,000~20,000 edges more than the BERT/Wordnet models, and the BERT/Wordnet models found 37,000~55,000 edges more than the lexical model. The whole networks of the lexical networks had the lowest network density. The numbers of connected components in the networks were the same in three of six periods. In another three periods, their differences were less than 3. Except for the lexical model, the average clustering coefficients and transitivity of the whole networks of these models were very close. Overall, the divergences between the BC whole networks were obscure except those based on the lexical model.

As shown in Figure 4-9, the divergences between the BC core networks of each model were clear. The node numbers of these models were still similar to each other. The gap between node numbers of the core networks in the same period was less than 20. The edge numbers of these models were inconsistent, and the orders of these models by edge numbers were not the same during the six periods. For example, in P1, the edge number of the core network of the lexical model was the lowest. In P6,

doi:10.6342/NTU202201326

Figure 4.8:
The statistics of the BC whole networks



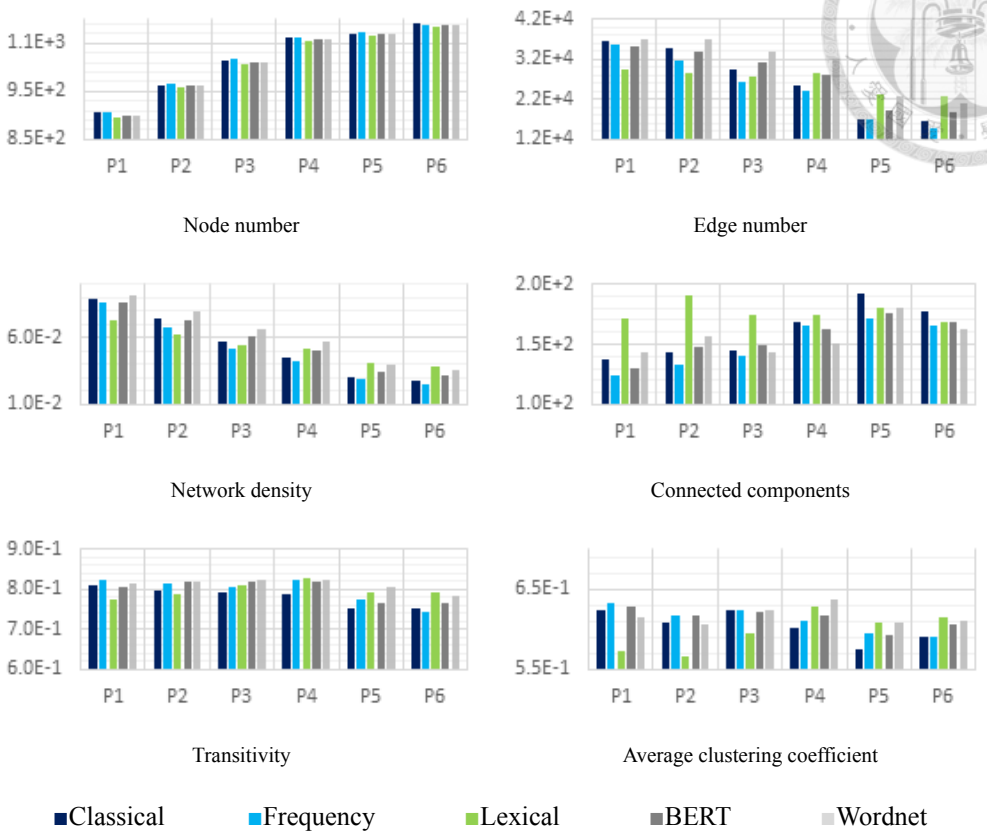
Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

however, it had the highest number of edges. The result supports that notable divergences exist between BC core networks based on different models. Similarly, the order of the models by their numbers of connected components in the core networks varied largely in the six periods. During P1 to P4, the core networks built by the lexical model had the largest number of connected components. In P5 and P6, the leader became the classical model. The inconsistency also occurred in the numbers of transitivity and average clustering coefficients.

As to JSD results, shown in Figure 4-10, JSD between each pair of BC whole networks were usually much lower than that of the same combination of BC core networks. When comparing the distribution of node degree in each BC whole network, JSD were usually below 0.05. If comparing the distribution of node degree in each BC

doi:10.6342/NTU202201326

Figure 4.9:
The statistics of the BC core networks



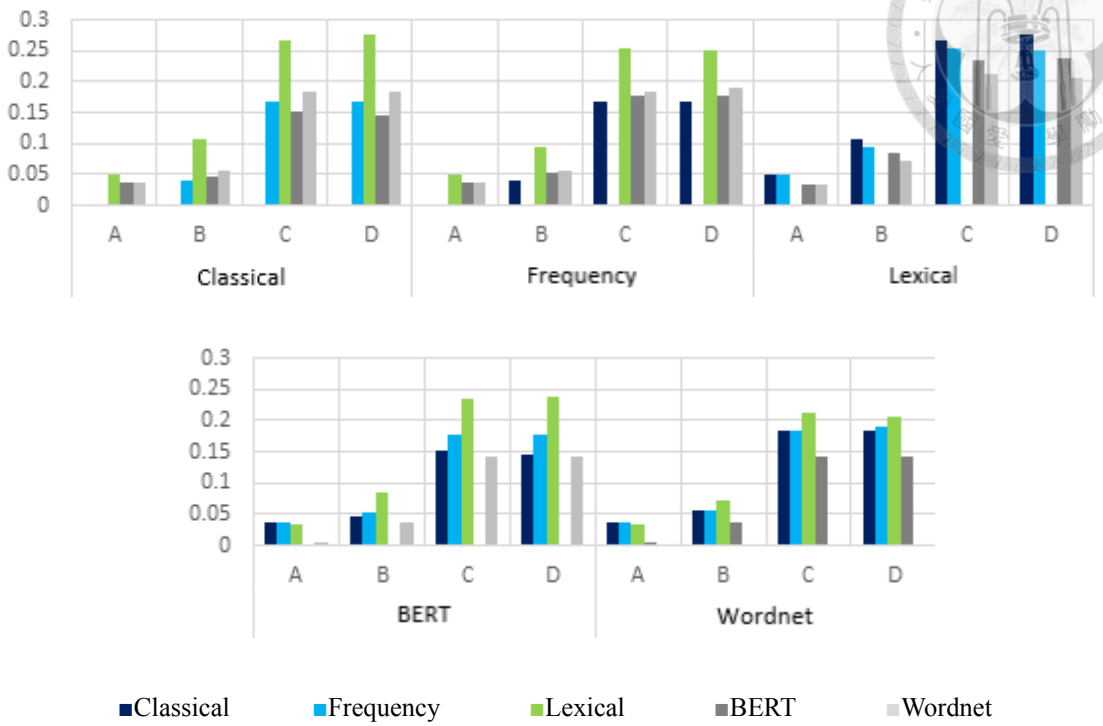
Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

whole network, JSD was 0.1 in maximum but usually below it. Such a result means that the differences between the whole networks were not apparent. The JSD between core networks were usually above 0.14. If the pair includes the core network based on the lexical model, the JSD would be more than 0.2. The increase of JSD shows the difference between the core networks based on different models.

Overall, the results show a conclusion similar to DC: the whole networks based on different models are similar, but the divergences between the core networks are obvious. In general, the network indicators and JSD show that the differences between the whole networks based on these models are obscure. The only exception is the lexical model, whose whole networks have had higher JSD when compared with the networks based on other models. However, when focusing on the core networks, the

doi:10.6342/NTU202201326

Figure 4.10:
The JSD between different models (BC)



Notes: Each bin represents the JSD result of a combination, including the model shown in the bottom of the sub-graph and the model indicated by the bin color. The labels on the x-axis mean the kinds of node degree distribution used to measure the JSD: (A) the node degree of the whole network; (B) the node weighted degree of the whole network; (C) the node degree of the core network; (D) the node weighted degree of the core network. The number on the y-axis is the JSD result.

differences between these models are obvious. Therefore, the present study concludes that these models show different BC core networks.

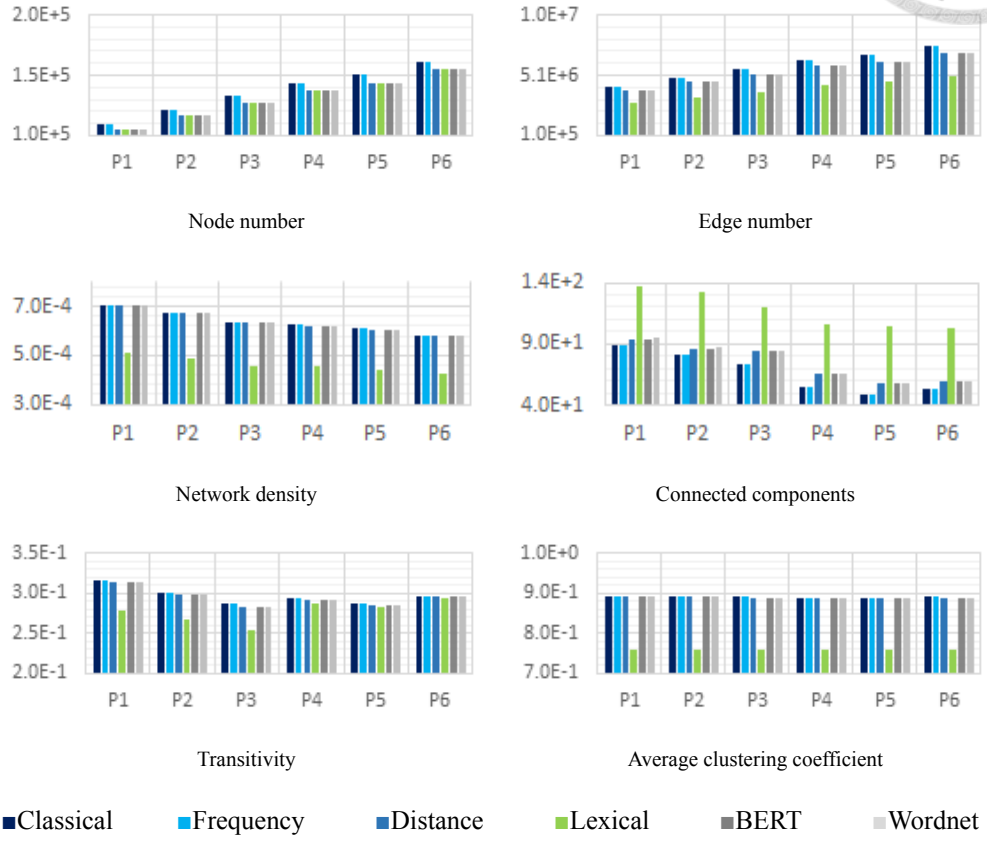
4.2.3 Co-citation

Figure 4-11 reports the indicators of CC whole networks built by six models. The differences in the node numbers between the classical/frequency models and the other models were about 4,000~6,000. In all periods, the edge numbers of the lexical model were lower than those of other models, about one million in P1 and two million in P6. The networks of other models had relatively similar edge numbers. The lexical model usually built a network with more connected components and a lower average clustering coefficient. Namely, the whole networks of the lexical model differed from those of other models. Moreover, according to these indicators, a high similarity

doi:10.6342/NTU202201326

existed between the classical and frequency models, and the remaining three models, the distance, BERT, and Wordnet models, were akin to each other.

Figure 4.11:
The statistics of the CC whole networks



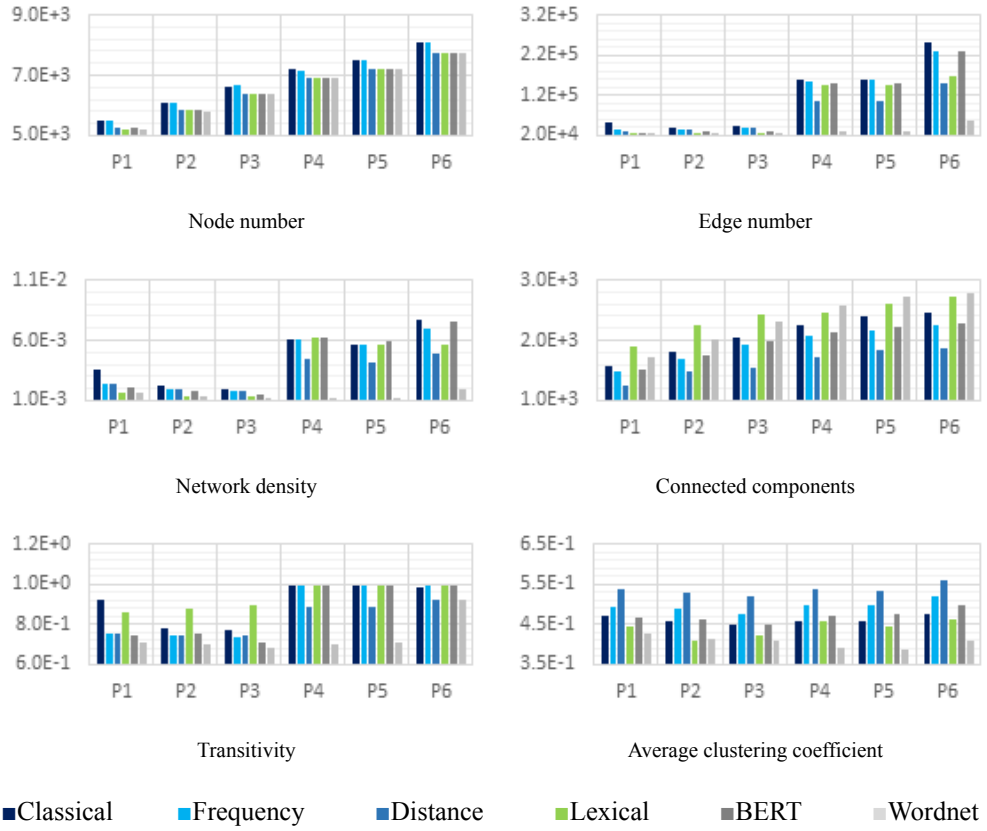
Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

The divergences of most indicators, except for the node numbers, increased when comparing the core networks. Figure 4-12 shows the details. The edge number of the networks of the Wordnet model was much less than other models. Compared with the classical and frequency models, the core networks of the lexical and distance models usually had fewer edge numbers. Besides, the number of connected components supports that the core networks of each model were quite different. The core networks of the distance model had the minimum number of connected components, and the maximum number of connected components in each period was 1.5 times more than the minimum. Transitivity also indicated that the triads of some networks were more

doi:10.6342/NTU202201326

likely to form triangles. For example, the networks of the lexical model during P2 and P3 had much high transitivity. The networks of the Wordnet model in P4 and P5 had much lower transitivity. Another indicator, average clustering coefficients, also suggested that the structures of the core networks were inconsistent. Overall, the CC core networks based on these models differed from each other.

Figure 4.12:
The statistics of the CC core networks

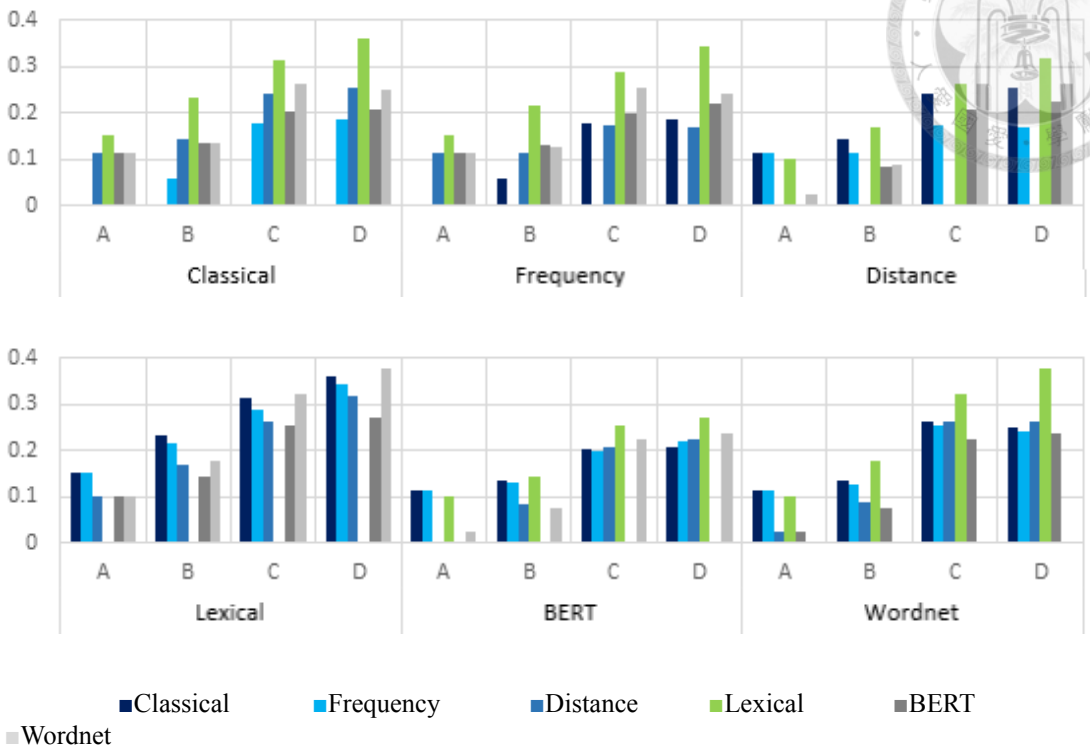


Notes: The x-axis shows six periods, including **P1**: 2010~2014, **P2**: 2011~2015, **P3**: 2012~2016, **P4**: 2013~2017, **P5**: 2014~2018, and **P6**: 2015~2019. The y-axis indicates the number of the indicator reported in the sub-graph. Be noted that the scales of the y-axis are adjusted for the concise representations.

Figure 4-13 reports the JSD between different models. The JSD showed that the node degree distributions of the lexical model were quite different from others. When comparing the whole networks and whole weighted networks, the JSD were usually above 0.1 and 0.14, respectively. As to the core networks and the core weighted networks, the JSD were 0.25 at least. The numbers reveal how different the results of the lexical model were when compared with others.

doi:10.6342/NTU202201326

Figure 4.13:
The JSD between different models (CC)



Notes: Each bin represents the JSD of a combination. The bottom is one model, and the bin color represents another model. The labels on the x-axis mean the kinds of node degree distribution used to measure the JSD: (A) the node degree of the whole network; (B) the node weighted degree of the whole network; (C) the node degree of the core network; (D) the node weighted degree of the core network.

The JSD were much smaller, 0~0.02, between the whole networks of three models: the distance, BERT, and Wordnet models. As to the whole weighted networks, the JSD raised to 0.08~0.09. JSD between the core networks further increased to 0.15 at least and usually higher than 0.2. The invisible differences between the whole networks of the three models became noticeable when comparing their core networks. Additionally, when comparing the whole networks and core networks of the BERT and Wordnet models with the classical and frequency models, the JSD were above 0.1 and 0.2, respectively. Moreover, although the differences between the whole networks of the classical and frequency models were not noticeable, the JSD still showed that the differences between their core networks could not be ignored. Overall, these indicators' results show the inconsistency of the structures between the networks built by the six models, especially the core networks.

doi:10.6342/NTU202201326

In this section, the present study analyzed the citation networks based on different citation relationships and models by seven indicators, including node number, edge number, network density, number of connected components, transitivity, average cluster coefficients, and JSD. The examinations aim at answering whether these models uncover different citation networks, namely research question one.

According to the results regarding the DC networks, the networks of the classical model were quite different from those of other models, including the frequency, BERT, and Wordnet models. However, the divergences between the later three models were slight, and it means that removing negative citations does not affect the network structures due to only a tiny percentage of citations labeled as negative. In other words, the divergences between these networks are due to considering ITCs, not removing negative citations.

The investigation about the BC and CC networks shows that the models used to weight citation relationship strength affect the divergences between these networks, especially the core networks. The networks structures of the classical and frequency models are much different from those of other models. These models, weighting citation relationships by the distance, lexical, and semantic similarity, reveal the BC or CC networks with different structures, especially the core networks.

Overall, the citation networks will differ from the classical model if networks are constructed by different methods, including considering ITCs frequency, the distance between ITCs, or semantic similarity. However, identifying sentimental polarity and removing negative citations may not result in the evident divergence. Although the differences usually existed when comparing these citation networks, the further questions are whether scholars can identify different research subfields from these networks. The following section reports the results of cluster analysis, which compares the clustering results between different core networks, and discusses whether the semantic models can identify the distinct clusters.

doi:10.6342/NTU202201326

4.3 The Results of Clusters Analysis

This section examines the clustering result for each core network, identified by Clauset-Newman-Moore greedy modularity maximization implemented by Networkx, and answers the second research question. Firstly, the number of clusters and their size in each core network are reported. Then, the similarity, decided by ARI, between the clustering results of each core network are presented.

After examining the similarity, textual coherence was used to evaluate the clustering results. As mentioned in Section 3.3.2, the higher textual coherence means that the works in the same cluster are more likely to concentrate on similar topics. In other words, higher textual coherence means that a model is better at constructing the core network providing better clustering results and discriminating the works. Additionally, examining the top n largest clusters' subjects, determined by their high-frequency topic words, was also used to judge the clustering results. The evolutions of the research subfields, research branches, and research trends were also used to determine the pros and cons of each model.

4.3.1 The number of clusters and their size

Table 4-2 reports the number of clusters in each core network and the number of clusters containing only a single node by period. Among DC core networks, the number of clusters in the core network based on the classical model was the lowest, between 20 and 40. Besides, no cluster contained only one node. The patterns entirely differed from the results of other models, namely the frequency, BERT, and Wordnet models. In the core networks built by the three models, 850 to 1100 clusters were identified, and 650~900 cluster contained only one node. The number of clusters in the core network based on the BERT model was the least among the core networks based on the three models, and the highest number of clusters was of the Wordnet model. Appendix A details the differences between the sizes of the top 5 largest clusters of

doi:10.6342/NTU202201326

different DC core networks.

Table 4.2:
Clustering results of each model and citation relationship

R/M		Year											
		2010~2014		2011~2015		2012~2016		2013~2017		2014~2018		2015~2019	
		#C	#SC	#C	#SC	#C	#SC	#C	#SC	#C	#SC	#C	#SC
DC	C	39	0	19	0	24	0	26	0	30	0	20	0
	F	905	709	944	734	935	753	936	751	953	755	998	790
	B	871	688	906	719	921	742	920	740	939	733	965	767
	W	979	758	966	793	1,000	812	1,007	820	1,037	845	1,075	881
BC	C	227	199	206	177	161	138	180	155	204	179	187	168
	F	220	192	182	158	155	130	176	157	185	158	180	155
	L	236	195	244	210	214	182	194	161	193	163	180	149
	B	213	186	211	182	187	158	172	147	189	161	180	157
	W	248	220	240	211	195	169	174	153	191	166	171	147
CC	C	1,601	1,439	1,850	1,660	2,097	1,889	2,289	2,075	2,423	2,216	2,479	2,269
	F	1,517	1,325	1,727	1,540	1,964	1,766	2,098	1,900	2,200	1,994	2,279	2,043
	D	1,302	1,121	1,515	1,319	1,607	1,393	1,752	1,569	1,875	1,677	1,911	1,692
	L	1,904	1,607	2,278	1,967	2,424	2,074	2,481	2,169	2,624	2,292	2,717	2,395
	B	1,552	1,383	1,785	1,602	2,023	1,806	2,155	1,971	2,238	2,035	2,299	2,082
	W	1,737	1,528	2,064	1,845	2,356	2,115	2,610	2,361	2,750	2,478	2,817	2,566

Notes: **#C**: the number of clusters. **#SC**: the number of single-node clusters. **C**: classical model. **F**: frequency model. **D**: distance model. **L**: lexical model. **B**: BERT model. **W**: Wordnet model.

Also shown in Table 4-2, the divergences of cluster numbers decreased when comparing the BC core networks based on five models. The number of clusters in each core network was between 150 and 250, and about 90% of clusters contained only one node. The differences between the maximum and the minimum number of clusters were 35, 60, and 60 from P1 to P3. It decreased to about 20 in the later three periods. Appendix B details the differences between the sizes of the top 10 largest clusters of different BC core networks. Compared with DC core networks, the divergences of BC core networks are reduced and obscure.

Regarding the CC core networks, the lowest number of clusters was of the distance

doi:10.6342/NTU202201326

model, at least 200 clusters less than other models. During the six periods, the lexical and Wordnet models had the largest number of clusters in P1~P3 and P4~P6, respectively. The differences between the maximum and the minimum number of clusters were between 110 and 160. In most core networks, less than 15% of their clusters contained multiple nodes. Appendix C details the differences between the sizes of the top 15 clusters of different CC core networks.

Overall, the results reported above show the differences between the core networks based on different models. The CC core networks of all models differ in the number of clusters more obvious. The differences, however, are obscure in the BC core networks of all models and the DC core networks of the frequency, BERT, and Wordnet models. In the following section, the present study reports the results of investigating the similarity between clustering results.

4.3.2 The similarity between the clustering results

In this study, the similarity between the clustering results of each core network is gauged by ARI. ARI measures how similar the two clustering results are. If ARI is one, the two clustering results are parallel in determining whether a pair of nodes are in the same cluster or not. The lower ARI is, the different the two clustering results are. The present study calculated ARI between the clustering results based on different models. Only the nodes which existed in both models' core networks were included because ARI compares the clustering results composed of the same compositions. Figure 4-14 shows the average ARI of each combination of the models.

Panel (A) shows the average ARI of different combinations, which included two clustering results based on different models applied in DC. The combinations including the classical model were very different from the others. It agrees with the previous observation that the core networks of the classical model differ from those of the others. ARI also demonstrated the high similarity between the clustering results of the frequency and Wordnet models, except for one outlier. A higher similarity existed

doi:10.6342/NTU202201326

Figure 4.14:
The similarity between clustering results of different models

(A) Direct citation

(B) Bibliographic coupling

(C) Co-citation

Notes: The y-axis represents ARI. The x-axis shows the combinations of the comparing models: **C** as the classical model, **F** as the frequency model, **D** as the distance model, **L** as the lexical model, **B** as the BERT model, and **W** as the Wordnet model. The mark x and bar point the mean and median, respectively.

between the clustering results of the frequency and BERT models, but the lower fence also hinted that the variance was large. The comparison between the clustering results of the BERT and Wordnet models led to a similar conclusion. Hence, the clustering results of the BERT model might be similar to those of the frequency and Wordnet models.

Panel (B) shows the outcomes of ARI between the clustering results of different models applied in BC. Usually, the mean and median of each combination's ARI were between 0.5 and 0.7. The combinations which included the lexical model were more likely to result in relatively lower ARI when compared with other combinations. The combinations of a higher ARI were (1) the classical and frequency models, (2) the

doi:10.6342/NTU202201326

classical and BERT models, and (3) the BERT and Wordnet models. Overall, the divergences between the clustering results of the BC core networks were usually lower than the combinations of DC core networks.

Panel (C) shows the outcomes of ARI between the clustering results of different models applied in CC. Three of the fifteen combinations had a lower mean and median, below 0.4, including (1) the classical and lexical models, (2) the distance and lexical models, and (3) the frequency and lexical models. All of these combinations included the lexical model. As to another two combinations including the lexical model, their mean and median of ARI were also below 0.5. It hints that the clustering results of the CC core network based on the lexical model differ from the other models. The highest similarity existed when comparing the clustering results of the BERT and Wordnet models. The other combinations' means and medians were usually between 0.45 and 0.65.

Overall, the similarity between two clustering results relies on the type of citation relationship and the models used to measure the citation relationships. Among the clustering results of the DC core networks, ARI shows that the clustering results of the classical model are much different from the others. However, the clustering results of the DC core networks based on the other models have less divergence between each other.

The lexical model provides the most different clustering results in the BC and CC core networks. Besides, the clustering results of the Wordnet model also differ from the others in the CC core networks. As to the comparisons between the clustering results of the other models, the divergences between them are slight in the BC core networks but evident in the CC core networks. According to these results, the present study concludes that divergences exist between the clustering results of most CC core networks and part of the clustering results of BC and DC core networks.

doi:10.6342/NTU202201326

4.3.3 The textual coherence

The previous examination shows that the structures of the core networks formed by these models were usually different. In addition, depending on the type of citation relationship, the clustering results of the core networks based on different models might differ from each other. The following question is which model might be a better solution when analyzing scientific structure. Textual coherence is the quantitative indicator to evaluate the clustering results in the present study. As mentioned in Chapter 3, textual coherence is determined by calculating the difference between $JSD(actual)$ and $JSD(random)$, which is composed of x works chosen randomly. The number x equals the number of the works whose title is available in WoS records or CrossRef data in an actual cluster. The higher textual coherence represents that the clusters identified from the network differ more from those formed randomly.

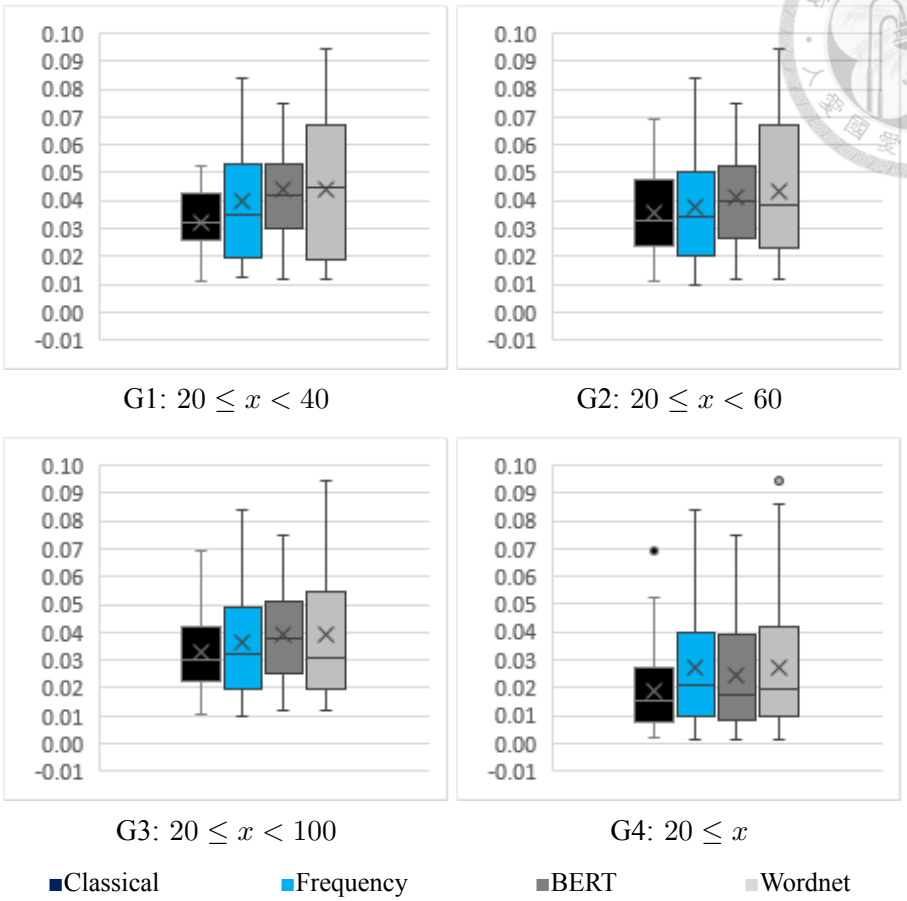
Figure 4-15 shows the textual coherence of the clusters in the DC core networks. According to x, the present study examines the textual coherence by four different groups: G1: $20 \leq x < 40$; G2: $20 \leq x < 60$; G3: $20 \leq x < 100$; G4: $20 \leq x$. The mean of the textual coherence of the classical model was consistently lower than that of other models. The largest mean difference was 1.19%, which existed in G3 between the classical and BERT models.

As to the other models, no one consistently outperforms in all groups. The average textual coherence shows that both the BERT and Wordnet models had better textual coherence than the frequency model in G1 and G2. In G4, however, the frequency model exceeded the BERT and Wordnet models. Besides, the differences between the means of these models were less than 0.55%. Hence, it is hard to argue which model can provide a better core network in which the works in the same clusters may be closer in their related topics.

Figure 4-16 represents the textual coherence of clusters in BC core networks based on five models. The largest difference between the averages was 1.03% in G1, between the classical and frequency models. In G4, the difference was reduced to 0.41%. The

doi:10.6342/NTU202201326

Figure 4.15:
The textual coherence of clusters in DC core networks



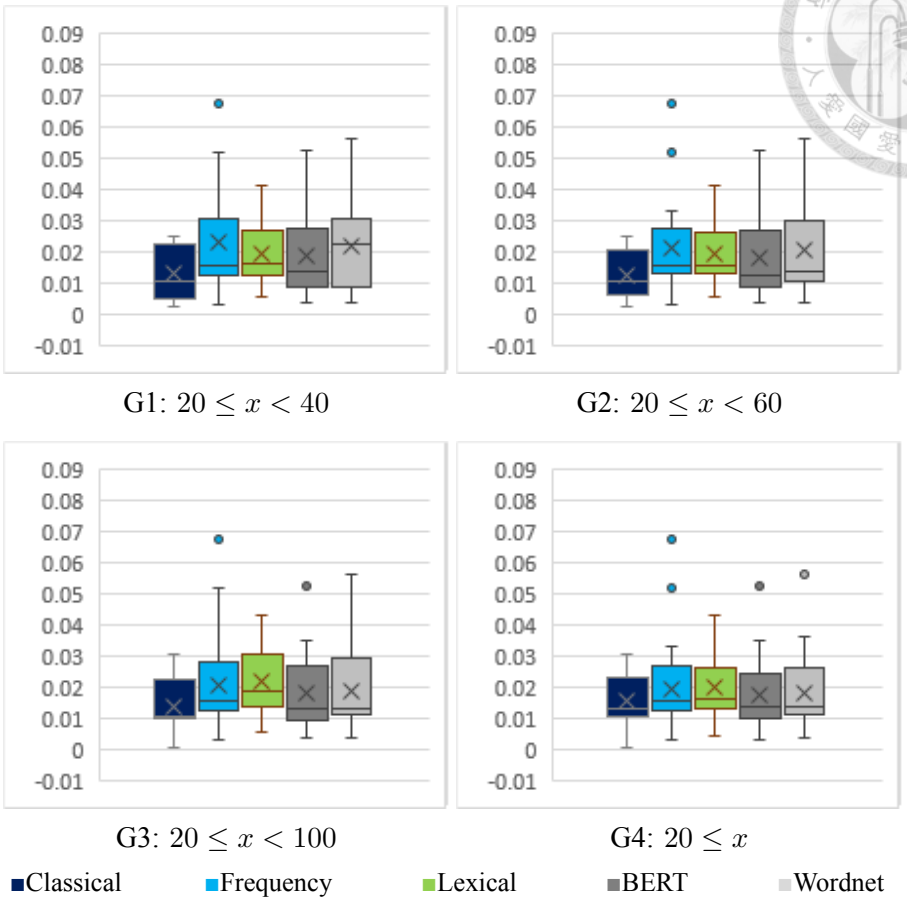
Notes: The y-axis indicates the number of the textual coherence. Mark x and bar represent the mean and median.

textual coherence of the classical model was lower than that of other models in all groups, and the textual coherence of the BERT model was only higher than that of the classical model. As to the other three models, it was also hard to argue which one's textual coherence is higher. Besides, in G4, the averages of the five models were quite close. The largest difference between the mean of textual coherence was only 0.41%. Such minor differences mean that not a single model provides a clustering result in which articles in the same clusters are much in common.

Figure 4-17 reports the textual coherence of clusters in CC core networks based on the six models and indicates several points. Firstly, compared with DC and BC, the differences between the textual coherence of different models usually increased. The largest difference was 1.49% in G4 between the average of the classical and lexical

doi:10.6342/NTU202201326

Figure 4.16:
The textual coherence of clusters in BC core networks



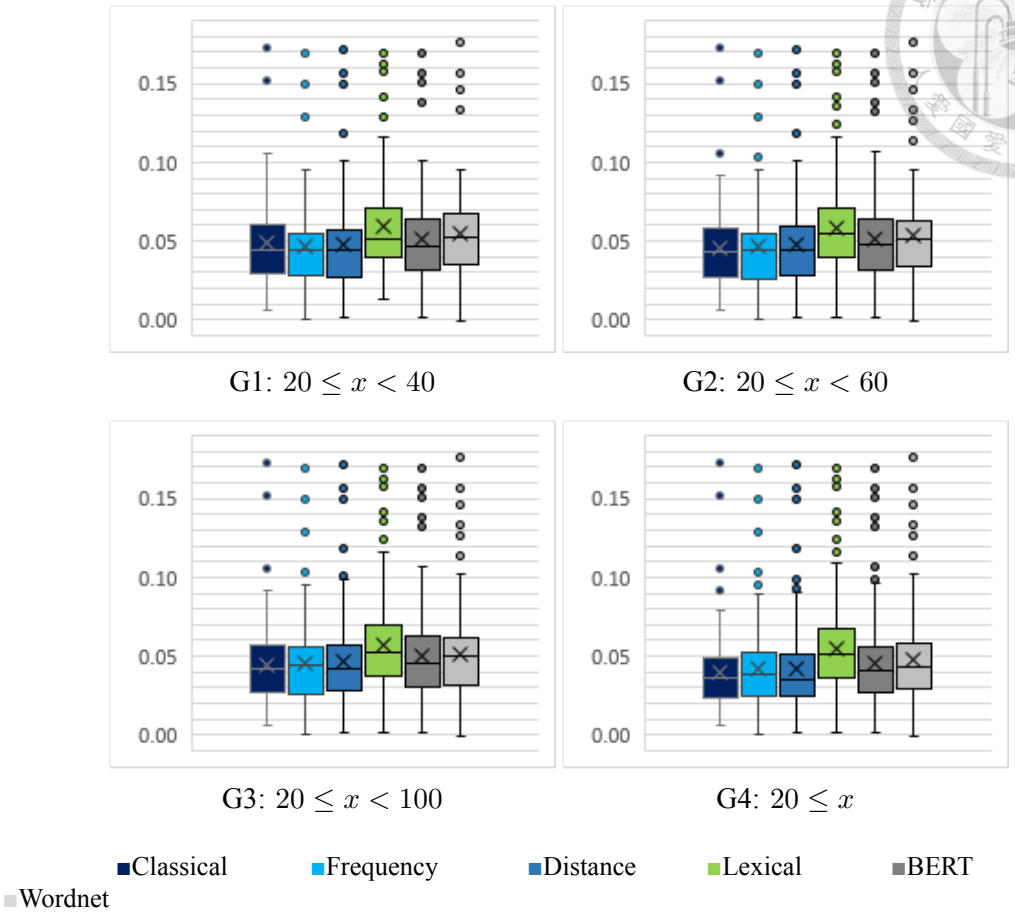
Notes: The y-axis indicates the number of the textual coherence. Mark x and bar represent the mean and median.

models. In G1, G2, and G4, the differences still exceeded 1%. Secondly, the textual coherence of the classical model was usually the lowest among all models. The only exception is G1. Thirdly, the textual coherence of the lexical model was the highest among all groups, and its average was above 5%. The followings were two models based on NLP semantic analytics, namely the BERT and Wordnet models. The average of the Wordnet model was larger than that of the BERT model.

Overall, textual coherence is a quantitative way to measure the clustering results in this study. The clustering results of core networks based on the classical model are usually worse than those based on other models. Additionally, considering additional variables usually improves the textual coherence of the clustering result, but the improvement is not apparent when modifying BCS. Furthermore, the improvement of

doi:10.6342/NTU202201326

Figure 4.17:
The textual coherence of clusters in CC core networks



Notes: The y-axis indicates the number of the textual coherence. Mark x and bar represent the mean and median.

the distance model, only applied to adjust CCS, is not obvious. The improvement of the frequency model depends on measuring which type of citation relationship. The frequency model does improve the textual coherence of the core networks built by DC but not necessarily by BC and CC.

Compared with the other models, the lexical model has a better ability to build a core network in which the works in the clusters are more similar to each other. As to the BERT and Wordnet models, the improvement also depends on the type of citation relationship. Both models may not apparently improve the textual coherence when used to build the DC and BC networks. However, in the CC networks, the results show that both models usually improve the textual coherence larger than other models, except the lexical model. In sum, the effects of using NLP semantic analytics in

doi:10.6342/NTU202201326

measuring citation relationships vary by the types of citation relationships. The following section reports the subjects identified by different models and discusses the pros and cons of each model.

4.3.4 The investigations of the largest clusters

Section 4.3.3 reports the textual coherence of the clustering result of the core network based on different models and evaluates the results. In this section, the present study first examines how large the proportion of the nodes in the core network is included in the several largest clusters, named the proportion as concentration tendency in the following discussion. Then, each cluster's research subfield was identified based on their high-frequency topic words. The identified subfields were used to find out the research trends revealed by the core networks during 2010 2019. Moreover, the composition of one or several research subfields, within a single period or across several periods, whose high-frequency topic words are lots in common, is defined as a research branch. Research subfields, research branches, and research trends were used to evaluate the results. Additionally, examining twin clusters, in which the articles of one cluster in a former period make up a large proportion of another cluster in the following period, was also used for the evaluation.

Concentration tendency and research trends

Direct citation. Unfortunately, the clustering result of the DC core networks is disappointing. The concentration tendency was very high in the DC core networks built by all four models. In different periods, the numbers of nodes varied between 10,000 and 15,000, and the numbers of nodes in the top two large clusters were usually higher than 3,000. The largest two clusters contained at least 63.57 percent of nodes, and the highest percentage rose to 91.19. The details are shown in Table 4-3. Such a high concentration tendency made naming the research subfield tough because of the lack of a specific term for representing a cluster. In the clusters which contained more

doi:10.6342/NTU202201326

than 1,500 nodes, no single topic word existed in over 10 percent of nodes. The appropriate terms used to name the subfields were usually too general to identify the research subfield specifically.

Table 4.3:
Concentration tendency of DC clustering results

Model	Year					
	P1	P2	P3	P4	P5	P6
C	65.46%	63.5%	63.80%	64.62%	64.62%	70.04%
F	81.42%	82.27%	69.03%	91.19%	80.81%	82.85%
B	78.45%	76.22%	73.70%	82.07%	85.27%	80.50%
W	75.17%	74.77%	76.34%	83.01%	87.44%	75.44%

Notes: **C**: classical model. **F**: frequency model. **B**: BERT model. **W**: Wordnet model.

The research subfields of the largest two clusters in each DC core network were usually related to SCS, IBIR, or AoT. In some of these clusters, their high-frequency topic words related to several research trends, e.g., IBIR and AoT. The inconsistency between the two research subfields of twin clusters was popular. The research subfields of some twin clusters were about IBIR in one period and SCS in another. This phenomenon indicates that some large clusters identified from the DC core networks were not specific enough. As to the remaining clusters, although they could concentrate on more specific subfields, their cluster sizes were so small that the granularity of these clusters might be unequal.

The issues mentioned above existed in the DC core networks built by all models. However, these issues did not commonly exist in the BC and CC core networks. One possible explanation is that the ways used to build the DC core networks and categorize the nodes are not appropriate enough. Another explanation is that considering ITCs or sentimental polarity cannot improve the clustering results in the DC core networks based on the classical model. Although the differences in the network structure and the clustering result exist between the DC core networks, the clustering results reported above do not support that the modified models used in the present study can improve the clustering result and reveal different research subfields and trends.

doi:10.6342/NTU202201326

Bibliographic coupling. The concentration tendency in the clustering results of the BC core networks was much less than that of DC core networks. Usually, the nodes of the largest four clusters in each core network included about 40 to 60 percent of the nodes in the BC core networks. Table 4-4 reports the details. Compared with the clusters in the DC core networks, the research subfields in each BC core network were more specific due to the smaller cluster sizes. Most research trends could be found within the identified research subfields of the largest ten clusters in each BC core network. These subfields were usually related to SCS, AoT, and IBIR. Although the clusters about CS or LS were not very popular, they still existed. In general, no noticeable difference was found when comparing the research trends revealed by the BC core networks based on different models. The research trends revealed by the modified models, which considered the ITCs, lexical similarity, or semantic similarity, were almost the same as those identified by the classical model.

Table 4.4:
Concentration tendency of BC clustering results

Model	Year					
	P1	P2	P3	P4	P5	P6
C	46.69%	48.76%	61.54%	64.88%	60.60%	61.72%
F	44.26%	48.45%	52.95%	71.50%	57.32%	55.19%
L	41.81%	43.72%	47.57%	47.01%	53.42%	53.74%
B	42.78%	44.80%	46.73%	58.85%	58.00%	59.67%
W	46.22%	45.32%	48.22%	53.93%	58.00%	59.21%

Notes: **C**: classical model. **F**: frequency model. **L**: lexical model. **B**: BERT model. **W**: Wordnet model.

The inconsistency between the research subfields of the twin clusters existed. Specifically, in some twin clusters, the subjects were about IBIR in one period and SCS in another. The phenomenon could be found in all BC core networks except those based on the frequency model. Besides, the research subfields identified from the BC core networks based on different models were similar. Hence, the approaches used in the present study may not improve the clustering results of the BC core networks compared with the result of the classical model.

doi:10.6342/NTU202201326

Co-citation. The concentration tendency of the clustering results of the CC core networks was much less than those of the BC/DC core networks and usually lower than 40 percent in most CC core networks. When the core networks were built on the lexical model, the concentration tendency of the clustering results was the lowest, usually no more than 20 percent.

Table 4.5:
Concentration tendency of CC clustering results

Model	Year					
	P1	P2	P3	P4	P5	P6
C	37.16%	28.31%	31.16%	42.96%	35.72%	38.04%
F	33.36%	27.75%	26.39%	36.37%	31.29%	38.70%
D	27.21%	25.84%	24.90%	29.64%	31.89%	30.69%
L	10.48%	8.95%	10.88%	20.43%	17.70%	19.70%
B	28.43%	22.42%	20.68%	34.65%	33.60%	36.54%
W	23.25%	19.99%	16.28%	18.89%	19.84%	23.20%

Notes: **C**: classical model. **F**: frequency model. **D**: distance model. **L**: lexical model. **B**: BERT model. **W**: Wordnet model.

The research trends identified from the largest 15 clusters in each core network included all trends reported in Hsiao and Chen (2020). It means that each models' capability in identifying research trends is close to each other. The research subfields of all twin clusters were close. Although the research subfields identified from the networks of these models were generally similar, the differences existed when comparing them in detail. Overall, using semantic analytics to measure CCS may identify the clusters with different subjects. The present study further compares and investigates the research branches and subfields identified from the CC core networks of different models.

Research branches and subfields

After examining the research trends identified from the subfields of the core networks, the present study compared the subfields identified by these models and their

doi:10.6342/NTU202201326

evolutions during the six periods. In the DC/BC core networks based on different models, the subfields of the top 5/10 largest clusters were almost identical, but several divergences existed between the subfields shown by the top 15 largest clusters in the CC core networks based on different models. Hence, the following analysis focus on the results of the CC core networks. Specifically, the subfields of three subjects, including *sentimental analysis*, *altmetrics*, and *unknown cluster*, are further discussed in the following.

Sentiment analysis. Since P3, 2012~2016, the subfields about sentiment analysis emerged in the top 15 largest clusters of the CC core networks based on the lexical model. The same subfields were also identified from the CC core networks based on the other models since P4. These subfields could be divided into two branches. The first branch, which only existed in P6, included studies applying sentiment techniques in analyzing Arabic data. The second branch, which existed in the CC core networks of all models since P4, included studies about similar issues without only focusing on Arabic. The high-frequency topic words of this branch included *sentiment analysis*, *sentiment classification*, *emotion*, *opinion mining*, and *social medium*.

In each period during P4 and P6, one of the largest 15 clusters in the CC core networks of the lexical and BERT models related to the second branch. In both models, the three relating clusters contained identical nodes. Although similar patterns also existed in the core networks of the classical, frequency, and Wordnet models, their node compositions were not identical but only similar. The scrutinization shows that the lexical and BERT models may gather citation entities about this subfield better because of the ability to categorize a cluster with the same entities consistently.

Besides, in the CC core networks based on the distance model, this branch contained two of the top 15 largest clusters in each period during P4 and P6. One was about the general topics of sentiment analysis, and another seemed to focus on the studies about analyzing the Chinese data. Namely, the distance model divided the citation entities, usually categorized as one cluster by other models, into two distinct

doi:10.6342/NTU202201326

clusters. A possible explanation is that the distance model decides citation relationship strength based on how authors cite works when writing sentences and paragraphs. They may organize cited references according to their writing purposes, not only by topic similarity, and the distance model can reflect the differences resulting from authors' opinions.

Altmetrics. Since Priem et al. (2015) coined *altmetrics* and proposed this idea, 358 related studies have been published in the journals included by WoS between 2010 and 2019. More than two-thirds of the studies were published between 2016 and 2019. The first period that the topic words of the top 15 clusters included this term was in P3, 2012~2016. The present study used the publication tendency of these articles to evaluate the clustering results of different models. After examining the clustering results of different models, the present study found some differences.

Firstly, *altmetrics* is not visible in the results of the classical and frequency models. Its relating articles were usually categorized into the cluster about general issues regarding informetrics. In the results of the frequency model, *altmetrics* was the high-frequency topic word in the clusters in P3 and P5 only. The pattern does not correspond with the publication tendency about this topic. Besides, in these clusters, only 3 to 5 percent of articles included *altmetrics* as the topic words. As to the results of the classical model, *altmetrics* was the high-frequency topic word in the clusters in P3, P5, and P6. Compared with the frequency model, this pattern is more reasonable. However, the percentage of articles containing this topic word decreased to 2%. The percentage was too low to claim that this topic was emphasized by both models.

As to the BERT model, *altmetrics* were included in a cluster of informetrics in each period from P3 to P6. The pattern corresponds with the increasing tendency of the relating publications. Besides, the percent of articles whose topic words included *altmetrics* in these clusters raised to 3~7%, higher than those of the classical and frequency models. Similar patterns are observed when examining the results of the distance model, and the percentage further increases to 3~10%. Overall, such a result

doi:10.6342/NTU202201326

shows that the BERT and distance models have a better ability than the classical and frequency models to reveal the publication tendency and emphasize this topic.

The results of the lexical and Wordnet models further stressed this topic. During P3 and P6, both models included *altmetrics* in one of their top 15 largest clusters in each period. In the results of the Wordnet model, 7 12% of the articles in the clusters contained *altmetrics* as their topic words. Additionally, *altmetrics* was the popular topic word in some of these clusters. In the lexical model, the percentage of the articles whose topic words contained *altmetrics* in the related clusters was about 10 11%, and all of these clusters focused on *altmetrics*. Although the lexical model has the best ability to reveal the cluster regarding *altmetrics* according to the results reported above, the result only supports that the lexical model has the advantage in revealing the cluster which focuses on the specific topics. However, the lexical model also suffers drawbacks due to this feature.

Unknown cluster. Among the top 15 clusters of the lexical model during P2 and P4, a research branch without the dominant topic words was noteworthy. This branch included one cluster in each period. None of any topic word was contained in more than 10% of the articles in a cluster. The only topic word contained by at least 5% of the articles in these periods was *hiv/aids*. Hence, the present study names them as *unknown cluster* and conjectures that this branch may relate to the studies about HIT or IBIR. Among the research branches of all models, such a branch only existed in the result of the lexical model. Namely, only the lexical model reveals the clusters without an apparent topic.

After retrieving the relating articles by searching *hiv/aids* in WoS, only 66 articles were published by LIS journals between 2010 and 2019. Compared with more than 350 articles about *altmetrics*, it is hard to justify that *hiv/aids* was the topic word with the highest article coverage rate. The existence of the cluster without a specific topic also indicates that the clustering results of the lexical model may be inappropriate due to the difficulty of finding out a common topic. Not to mention that the unknown

doi:10.6342/NTU202201326

cluster was the fourth largest cluster in P2.

4.4 The Results of Nodes/Relationships Analysis

In 4.2 and 4.3, the present study presents the results of citation analysis based on different models at the network and cluster levels. This section reports the investigation at the node/relationship level and discusses whether the BERT and Wordnet models help identify the critical citation entities and relevant citation relationships. These models detected the sentimental polarity of the citances, and the source articles were categorized into three types accordingly. This section discussed whether the source articles' DC and number of ITCs correlate to the type of the sentimental polarity of its citations first. Then, the present study discusses whether the BC/CC relationships, emphasized by the BERT and Wordnet models, are composed of the citation entities more similar to each other. The similarity between the citation entities was gauged by the number of common terms in their SciVal Topic, provided by Scopus. This study examined and compared the results of the top 100 BC/CC citation relationships of different models.

4.4.1 Citation counts and sentimental polarity

According to the sentimental polarity identified by the BERT and Wordnet models, the source articles were categorized into strongly positive class, weakly positive class, and neutral class. Figure 4-18 reports each journals' average DC of different sentimental classes by years. The dark blue (solid line) and light blue lines (dash line) indicate the trends of the strongly positive and weakly positive classes, respectively. The brown line (dot line) shows the trend of the neutral class.

Figure 4-18 shows the trends of the average DC of three classes. Among the 15 journals, there were seven journals whose average DC of the strongly positive class is higher than the average DC of the other classes in at least 7 of 10 years. These journals

doi:10.6342/NTU202201326

were *Government Information Quarterly*, *Information Processing & Management*, *Journal of Informetrics*, *JASIST*, *Library Hi Tech*, *Online Information Review*, and *Scientometrics*. Besides, in all journals, except for *Health Information* and *Libraries Journal*, the average DC of the neutral class was the lowest most of the time.

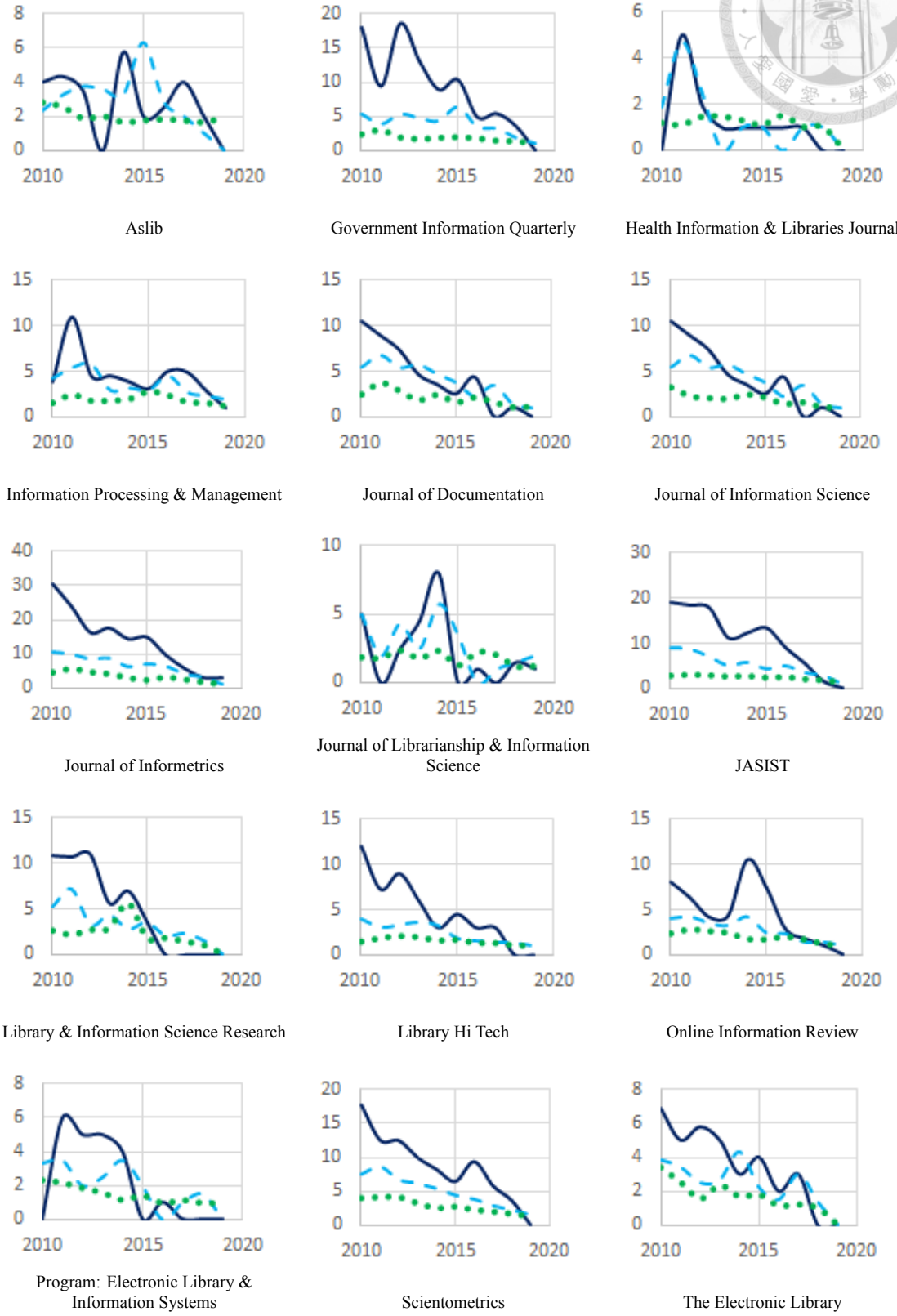
Due to considering the time lag between a work being published and cited, this study examined the average DC of the source articles published during 2010~2015. During 2010~2015, there were five journals whose average DC of strongly positive class was always higher than the average DC for another two sentimental classes. In most journals, the average DC of the neutral class was the lowest in this period. The only exception is Health Information and Libraries Journal.

Figure 4-19 reports each journal's average ITCs of different sentimental classes by year. There were more than half the journals whose average ITCs of the strongly positive class was higher than the average ITCs for another two sentimental classes in at least 7 of 10 years. The exceptions were *Aslib*, *Health Information* and *Libraries Journal*, *Journal of Librarianship and Information Science*, *Library and Information Science Research*, *Program: Electronic Library and Information Systems*, and *The Electronic Library*. Similarly, the present study examined the average ITCs of the source articles published during 2010~2015. Almost all journals' average ITCs of the neutral class was the lowest in this period. The only exception was *Health Information and Libraries Journal*.

Overall, if one of a cited work's citances is identified as positive, the works' DC and number of ITCs are usually higher than other works whose citances are not identified as positive. The tendency would be more apparent if the identification is based on the result of multiple models, e.g., the strongly positive references in this study. Hence, identifying sentimental polarity helps researchers find the possible works with high influence. Be noted that the present study does not claim a causal relationship between positive citances and high influence due to the difficulty of detecting the exact time that authors make a positive citation is hard to know by analyzing citation content

doi:10.6342/NTU202201326

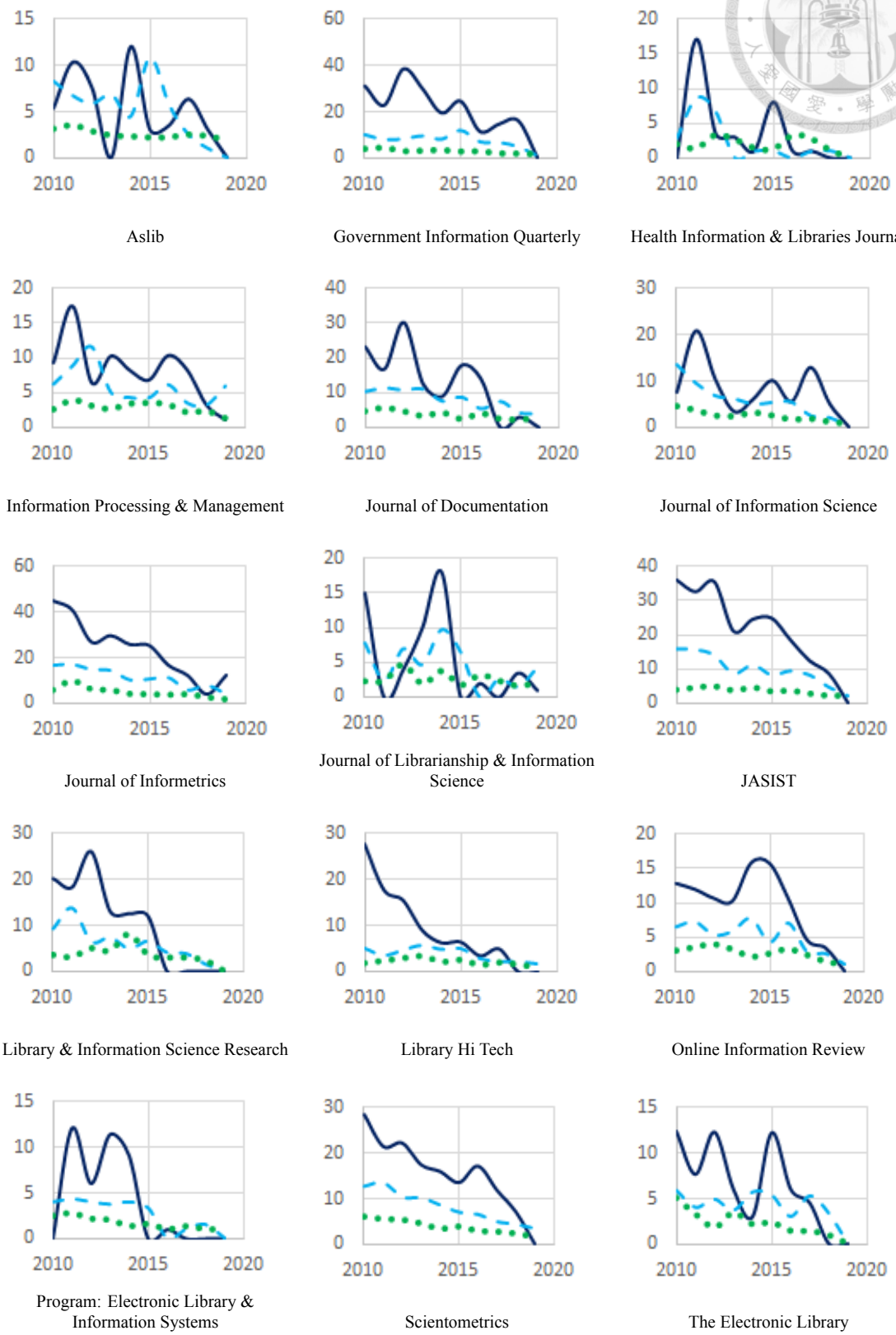
Figure 4.18:
The average DC of different sentimental classes



Notes: Dark blue solid line: strongly positive class. Light blue dash line: weak positive class. Green dot line: neutral class.

doi:10.6342/NTU202201326

Figure 4.19:
The average ITCs of different sentimental classes



Notes: Dark blue solid line: strongly positive class. Light blue dash line: weak positive class. Green dot line: neutral class.

doi:10.6342/NTU202201326

only. For testing this causal relationship, further studies will be necessary.

4.4.2 Topic similarity of the BCS pairs

The present study calculated the topic similarity of the top 100 BCS pairs ranked by each model. The topic similarity between two works of a BCS pair is the Jaccard similarity between their topic terms, SciVal Topics provided by Scopus. Examining the sum of the topic similarity of the top BCS pairs shows each model's ability to give high BCS to the pairs with highly similar or identical topics. The present study examined the top 10, 25, 50, 75, and 100 pairs to answer whether the sum of topic similarity of the top BCS pairs ranked by the BERT and Wordnet models would be higher than that of other models. Table 4-6 reports the details.

Table 4.6:
The sum of topic similarity of the top n BC pairs

n	Type	C	F	L	B	W
10	*Sim*	8	7	9	9	**10**
	NA	0	0	0	0	0
25	*Sim*	22	16	**23**	20	22
	NA	0	0	1	0	0
50	*Sim*	**41**	35.2	39.2	**41**	40.2
	NA	0	0	3	2	1
75	*Sim*	59	55.2	55.2	**59.2**	58.2
	NA	3	0	3	3	6
100	*Sim*	**81.2**	75.2	76.2	78.4	80.2
	NA	3	0	3	4	6

Notes: **C**: classical model. **F**: frequency model. **L**: lexical model. **B**: BERT model. **W**: Wordnet model. *Sim* means the sum of the topic similarity of all top n pairs. *NA* represents the number of pairs without available topic similarity due to the lack of SciVal Topics. The bold and italic number is the highest sum of the topic similarity.

Table 4-6 shows that the classical model usually performed better than the others and the frequency model was the worst. The sums of the BERT and Wordnet models were close to that of the classical model and surpassed occasionally. The results do not

doi:10.6342/NTU202201326

support that these modified models may have a better ability in identifying more relevant citation relationships than the classical model.

4.4.3 Topic similarity of the CCS pairs

The present study also calculated the topic similarity of the top 100 CCS pairs ranked by each model and examined the sum of topic similarity of the top 10, 25, 50, 75, and 100 pairs. Table 4-7 reports the details. The number of CC pairs without topic similarity was much higher than the top BCS pairs. One reason is that the top CCS pairs usually included the works published before 2010. SciVal Topics provided by Scopus are not wildly available in those old publications like the recent ones. Besides, the works included in the top CCS pairs whose citation entities may not be the journal articles indexed by Scopus. Both reasons increase the possibility of a CCS pair without topic similarity.

Table 4.7:
The sum of topic similarity of the top n CC pairs

n	Type	C	F	D	L	B	W
10	*Sim*	6	5	6	4	5	7
	NA	4	5	4	6	5	3
25	*Sim*	13	13	12	13	14	15
	NA	12	12	13	11	11	10
50	*Sim*	27	28	29	27	30	29
	NA	23	22	21	22	19	21
75	*Sim*	44	47	44	42	43	45
	NA	31	28	30	32	31	28
100	*Sim*	58	56	58	52	57	61
	NA	41	44	41	45	41	37

Notes: **C**: classical model. **F**: frequency model. **L**: lexical model. **L**: lexical model. **B**: BERT model. **W**: Wordnet model. *Sim* means the sum of the topic similarity of all top n pairs. *NA* represents the number of pairs without available topic similarity due to the lack of SciVal Topics. The bold and italic number is the highest sum of the topic similarity.

The sum of topic similarity of the Wordnet model was consistently higher than that

doi:10.6342/NTU202201326

of the classical model, and the sum of the lexical model was consistently lower than that of the classical model. In other modified models, the sum of topic similarity was usually close to that of the classical model. Although the Wordnet model, compared with the classical model, may better identify more relevant citation relationships, further study will be necessary for such a conclusion due to the tiny differences and many pairs without topic similarity.

4.5 Discussion

4.5.1 Applying semantic analysis in DC

In the present study, the results of three modified models, including the frequency, BERT, and Wordnet models, were compared with the results of the classical model when measuring DC. Several network indicators and JSD show that no notable differences existed between the whole networks built by each model. The same indicators also point out that significant divergences occurred between the core networks of the modified models and the classical model. However, the divergences may be only due to considering ITCs in the modified models because the differences between the networks based on the modified models, both the whole and core networks, are so slight that they can be ignored. Namely, the effects on the network structures by removing the negative citations are trivial.

The analysis at the cluster level points out the same conclusion. After comparing several indicators for clusters and ARI, the present study concludes that the clustering results of the core networks based on the classical model differ from those of the modified models. Considering the ITCs makes the structures of the core networks based on the classical model different from the others, and it also affects the clustering results of the classical model. However, the clustering results of the core networks based on the modified models are similar to each other. Namely, removing the negative citation from the core networks does not significantly affect the clustering results.

doi:10.6342/NTU202201326

Examining the textual coherence further confirms that the clustering results of the BERT and Wordnet models were only slightly different after removing the negative citations. The textual coherence of the classical model was lower than those of the modified models, and the differences between the textual coherence of the modified models were obscure, less than 0.002. The subject analysis also shows that research subfields revealed by the four models were almost identical. Hence, identifying and removing negative citations may not increase the similarity between works belonging to the same cluster and does not reveal differences among research subfields in comparison to the classical model.

The results show that removing negative citations does not significantly change the network structures and clustering results based on the frequency models. The phenomenon might be due to the low number of identified negative citations. The possible reasons for finding few negative citations include the researchers' tendency to represent the negative comments indirectly (Athar, 2011; Ghosh et al., 2016; Goodarzi et al., 2014) and the less usage of negative citations (Chubin & Moitra, 1975; Tabatabaei, 2013).

Besides, the present study found that the present semantic tools have weaknesses in identifying negative citations when fine-tuning the BERT classifier and deciding the thresholds for negative citations based on the SIA compound score. Given that both BERT and NLTK are not specifically designed to analyze the scientific writing styles, the tendency to give negative comments indirectly does limit both tools' performances. Although the BERT classifier can be fine-tuned with a corpus of specific writing styles, the proportion of the negative citations in the corpus provided by Athar (2011) is still not large enough to train this model well. Therefore, whether the BERT classifier is well trained for negative citations will be a question, especially those without strong and direct representations. According to the related studies (Chubin & Moitra, 1975; Lin, 2018; Tabatabaei, 2013), the number of negative citations identified by the BERT model in the present study might be too low. Another question is whether the number of actual negative citations is large enough to affect the result of citation analysis. In

doi:10.6342/NTU202201326

the present study, the Wordnet model had categorized about 5 percent of citations as negative, but its network structures and clustering results were still highly similar to the frequency models'. It raises the question of whether negative citations affect the result of citation analysis.

At the nodes/relationships level, the positive citation will be the clue for identifying possible references with high citation counts. When any citance of an article is categorized as positive, its number of citation counts will be likely higher than the average citation counts of the articles, published in the same journal and year, without any positive citance. Additionally, if multiple NLP techniques, e.g., BERT and SIA in this study, identify at least one of an article's citances as positive, the possibility of this article cited more times than the average increases. The tendency will be more noticeable when considering the time required for being cited since its publication. Although the present study does not investigate the causal relationship between positive citations and high citation counts due to the research limitation, the results suggest that using NLP techniques may help researchers detect the possible articles with high citation counts, especially when applying multiple methods to identify sentimental polarity.

Overall, the present study argues that applying semantic analysis in DC with detecting the sentimental polarity of references helps researchers identify references with high citation counts. However, it cannot provide better clustering results and scientific networks different from that considering ITCs.

4.5.2 Applying semantic analysis in BC

In the present study, the results of four modified models, including the frequency, lexical, BERT, and Wordnet models, were compared with the results of the classical model when measuring BC. Similarly, the network indicators and JSD showed that the differences between the whole networks based on each model were minor, usually less than the differences of the DC whole networks. However, obvious divergences existed

doi:10.6342/NTU202201326

between the core networks of the modified models and the classical model. Besides, within the modified models, their core networks also differed from each other. Namely, the methods used to weight BCS in this study affect the structure of core networks and make the core networks differ from each other.

Although the core networks built by each model differed, the analysis at the cluster level showed that their clustering results were akin to each other. The indicators for clusters and ARI showed that no obvious divergences, in most cases, existed among the clustering results between the classical and modified models. Only the clustering results of the lexical model differed from those of the other models. Hence, the present study argues that measuring BCS with additional data does not create notable differences of clustering results between the modified models and the classical model.

The observations based on the textual coherence and the subject analysis also supported that utilizing additional data does not significantly change the clustering results. As mentioned above, the increase in textual coherence means that the quality of clustering increases. Although the clustering results of the modified models had higher average textual coherence than those of the classical model, the amount of the increase was slight. Besides, this result also pointed out that the modified models based on non-semantic analysis, compared with those based on semantic analysis, had a tiny advantage in the quality of clustering. Moreover, the subject analysis showed that the research subfields identified from the core networks of each model were close. In other words, the modified models did not reveal the research subfields different from the classical model.

The above discussions suggest that the effects of measuring BCS with additional data, extracted from the article content, are not significant. Although modifying the ways to weight BCS might result in different BC core networks, the clustering results of these core networks seem similar. Furthermore, using additional data only improves clustering quality on a small scale, as shown by the investigation based on textual coherence. The modification fails to categorize citation entities into clusters in which

doi:10.6342/NTU202201326

the included entities are similar to each other. The models based on semantic analysis may not be significantly more appropriate than those based on non-semantic analysis for improving the result of BC.

At the nodes/relationships level, the variation between the pairs of citation entities with the high BCS identified by different models was minor. In general, the high BCS pairs ranked by the modified models would more likely have a high topic similarity. However, it seems those with the high BCS pairs ranked by the classical model demonstrate high topic similarity. The modified models would rank some pairs with high BCS different from the classical model, but the replacement of the high BCS pairs usually decreased the sum of the topic similarity of these pairs.

Overall, weighting BCS between citation entities by the sematic similarity, decided by the semantic analysis, of their related citances may not significantly improve applications of citation analysis based on BC. Although measuring BCS with semantic analysis uncovers different structures of networks, the identified research subfields are highly similar to the subfields uncovered by the classical model. Besides, according to the textual coherence, applying semantic analysis only slightly improves the quality of categorization compared with the results of the classical model. The improvement is lower than the modified models based on the frequency of reference mention and the lexical similarity. Moreover, the top BCS pairs ranked by the semantic analysis are not more likely to have a higher topic similarity. The advantage of applying semantic analysis in measuring BCS and its benefit is not notable. The improvement provided by the modification based on the frequency of reference mention or lexical similarity may be more noticeable.

4.5.3 Applying semantic analysis in CC

In the present study, the results of five revised models, including the frequency, distance, lexical, BERT, and Wordnet models, were compared with the results of the classical model when measuring CC. The six models could be split into three groups by

doi:10.6342/NTU202201326

the network indicators. The differences between the whole networks of the models within the same group were minor, but the whole networks based on the lexical model differed much from those of the others. As to the core networks, the structures of the models were different from each other in most network indicators. JSD also showed that the divergences were minor between the whole networks, except for that based on the lexical model, and were manifest between the core networks.

The inspection of the clustering results showed that the modification of the CC weighting scheme changed the results. The ARI also indicated that the clustering results of these models differed from each other, especially the lexical and Wordnet models. Additionally, the lexical, WordNet, and BERT models usually had clustering results with higher textual coherence than those of the classical model. The subject analysis also indicated that parts of the research subfields uncovered by these models differed. Compared with the classical model, the modified models emphasized some different research subfields and were capable of identifying new subfields earlier.

Hence, the examinations of the network indicators, JSD, textual coherence, and subject analysis showed that modifying the CC weighting schema affected the network structure and research subfields identified. Additionally, the increase in the textual coherence meant that the modification usually improved the clustering results. These results have indicated the possible advantage of applying semantic analysis in measuring CC. Although the lexical model had a higher textual coherence than the Wordnet and BERT models, the results of subject analysis implied that the lexical model might categorize the articles into too small clusters to reflect the scientific structure properly. Namely, applying semantic analysis in measuring CC sketches the scientific structure from a different perspective, provides the clustering results in which the articles might be more similar to each other, and demonstrates different research subfields.

The result of the analysis at the nodes/relationships level is similar to that of BC. The sums of the topic similarity of the top n CCS pairs in different models differed

doi:10.6342/NTU202201326

slightly. Such a result hardly supports that any model has advantages in ranking the pairs of works in order of the topic similarity.

Overall, the performance of applying semantic analysis in measuring CC is promising in comparison to that in measuring BC. Firstly, weighting CCS by semantic analysis makes the network structures and clustering results differ from the classical model and other modified models. Meanwhile, according to textual coherence, the text similarity between works in the same cluster is higher when measuring CCS by semantic analysis. In addition to the higher similarity, research subfields identified by semantic analysis may uncover some emerging research topics earlier. Thus, the present study argues that applying semantic analysis in measuring CCS improves clustering results and early identification of research topics. However, at the nodes/relationships level, applying semantic analysis in measuring CCS seems to make no significant improvement.

4.5.4 Further discussion of the advantages and weaknesses

The present study improves DC by detecting the sentimental polarity of references' citances and measures BCS/CCS by considering the semantic similarity between their related citances. The discussion above shows that these modifications can classify citation entities into several groups and measure citation relationships based on the related citances. The foci of the following discussion are the pros and cons of applying semantic analysis in measuring different kinds of citation relationships and the possible reasons for these advantages and disadvantages.

Direct citation relationship

The direct citation is that a citation entity claims a citation relationship to another entity. The present study investigated the effects of classifying direct citations by their sentimental polarity. By utilizing the corpus provided by Athar (2011), the references were classified into three groups based on the sentimental polarity of their citations.

When utilizing semantic analysis, researchers can classify direct citation

doi:10.6342/NTU202201326

relationships. By using the NLP techniques of the predict model, e.g., the BERT model in the present study, researchers can fine-turn language models to fit their classification missions if the appropriate training sets are available. After the fine-turn process, researchers can classify citations and references into corresponding classes and explore how the class helps researchers improve the studies of citation analysis. When trying to use the techniques of the dictionary model, e.g., the Wordnet model in the present study, how to design and train the method to classify citations and references properly may be more difficult. Even so, the dictionary model still helps researchers accomplish the classification tasks if the methods are well designed.

Classifying citations or references into different classes by their sentimental polarity allows scholars to study the features of different kinds of citations. Scholars usually use DC to evaluate citation entities, e.g., articles or authors, by various citation indicators, like citation counts or h-index. Checking the features of different classes makes scholars understand citation behavior more and improves future applications of citation analysis, like the correlation investigated by the present study between the positive references and total citation counts. Some studies, e.g., Catalini et al. (2015) and Xu et al.(2022), have applied machine learning techniques in classifying citations and investigating their features and roles in scholarly communication. Additionally, combining the results of multiple classifiers can further classify citation entities or citation relationships in the finer granularity and provide more possibilities for further studies and applications.

Nonetheless, the experience of identifying negative citations also indicates that the semantic analysis used in the present study may not identify some classes of citations well. The possible reason is that these tools are for processing general natural language, not for scholarly publications. Scholars usually use indirect and obscure forms to represent negative opinions like criticism or disagreement in their works. Such writing style hardens the difficulty of identifying negative citations by using semantic analysis designed for processing general natural language. Despite this weakness, the experience of investigating DC relationships in this study supports that studying DC

doi:10.6342/NTU202201326

relationships may benefit from classifying DC relationships by sematic analysis.

Besides, removing negative citations identified by the semantic analysis does not significantly affect the structure of citation networks and the clustering results. Given that the current models cannot identify negative citations well, improving classifiers or enhancing training sets may reach different conclusions. Overall, using semantic analysis does help researchers in classifying citations, investigating the roles played by different kinds of citations in the citation network, improving the applications of citation analysis, and developing the citation theory.

Compared with classifying citations, utilizing the semantic analysis to adjust the strength of DC as a continuous number, instead of a dichotomous number, might be difficult. Although the studies reviewed above have reported that part of references cited by a work may affect it more than other references, more studies are required to answer how to reasonably decide the quantitative number by weighting the semantic meaning extracted from the text. To some degree, the multiclass task can provide an approximate solution for the weighting problem, but whether appropriate training sets are available for this task is another major issue. Future studies may investigate this approach's feasibility and evaluate its pros and cons.

In sum, the present study concludes that applying semantic analysis in analyzing DC can classify citation entities and citation relationships and investigate the differences between the classes. Semantic analysis can help researchers discriminate citations and references to enhance future applications of citation analysis and improve the understanding of citation behaviors. However, the effectiveness of semantic analysis, especially designed for general natural language, is not without question without the proper training set. Besides, it seems that applying semantic analysis at the nodes/relationship level will be more appropriate for DC, especially for identifying influential citation entities and relationships. Although the present study suggests that removing negative citations may not seriously affect the citation network and clustering result, investigating how other kinds of citations and references affect the network

doi:10.6342/NTU202201326

structure and the clustering result may be an interesting future research question.

Bibliographic coupling

The bibliographic coupling is a citation relationship between two citation entities based on their common related references. The present study investigated the effects of adjusting BCS with semantic analysis and compared the results with the classical model. The preliminary results show that the methods used in the present study exhibit only minor effects on improving the result. Although the network structure and clustering result changed, the textual coherence and subject analysis indicated that semantic analysis does not uncover new research subfields or enhance the text similarity between the articles within the same cluster.

Nonetheless, not only the models applying semantic analysis but also those modified by lexical similarity and ITCs fail to improve the BC result based on the classical model. Therefore, the present study considers that refining BC with stricter criteria when calculating BCS may be inappropriate. The process of forming the BC network can be decomposed into the following two steps:

1. Pick a citation entity that does not be added to this network.
2. Measure its BCS between this new entity and each node that already existed in this network.

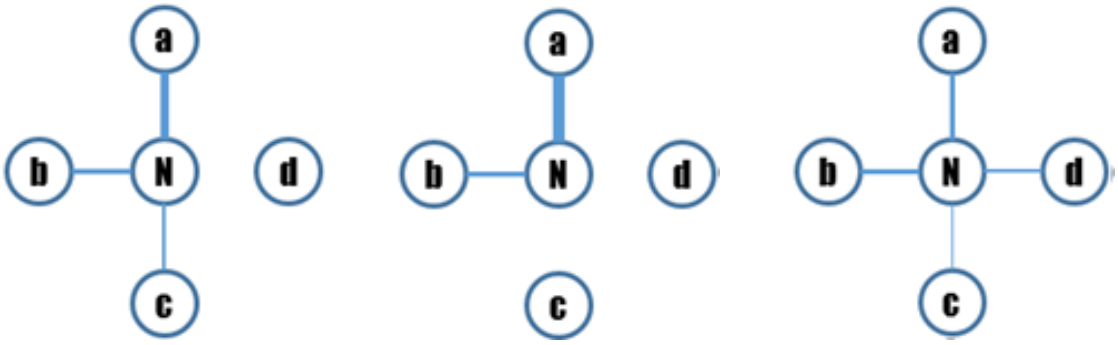
For the relationships in the BC network before adding a new entity, the addition does not affect them. Modifying BCS will not work like modifying CCS, which may affect a specific citation relationship multiple times. This feature of BC limits the possible benefit resulting from the modification. Besides, if the modified models work as expected, they will strengthen the BCS for the pair of highly related entities and vice versa. However, according to the argument of Kessler (1963a, 1963b) and the experience of the subsequent studies about BC, the classical model may have given a high BCS for the pair of high related entities. The modified models applied in the present study may only strengthen the original structure revealed by the classical model. The present study considers that the reason may explain why the modified models which use data extracted from the text body, including semantic analysis, to

doi:10.6342/NTU202201326

adjust BCS cannot significantly affect the network structure and uncover different clusters.

Another direction to revise BC is to gauge BCS based on the number of common references and the sum of the similarity between references if no references in common. In the present study, the BCS increases only when two entities cite the same reference. Some studies, like Liu (2017), proposed that BCS between two entities might increase to a degree if two different references, cited by each entity, relate to each other. The amount of increase will be decided by the similarity between references. Figure 4-20 shows the possible effects when using different stands in modifying BC. Subgraph (1) represents adding a new node N to an existing BC network containing four nodes a, b, c, and d. The width of an edge shows the BCS, measured by the classical model, for each pair of nodes. The approach used in the present study may strength strong relationships and remove weak relationships to solidify the core structure built by the classical model as shown in the subgraph (2). The subgraph (3) shows another stand that may add new relationships and may change the network structure. The present study advises that future studies investigate how to apply semantic analysis in this approach.

Figure 4.20:
The effects when measuring BCS in different approaches



Notes: N represents the new node adding to an existing BC network, a~d are the original nodes in the BC networks. (1) represents the effect of adding new node *N*. (2) is the way used in the present study to revise BCS. (3) shows the possible effects when revising by another direction.

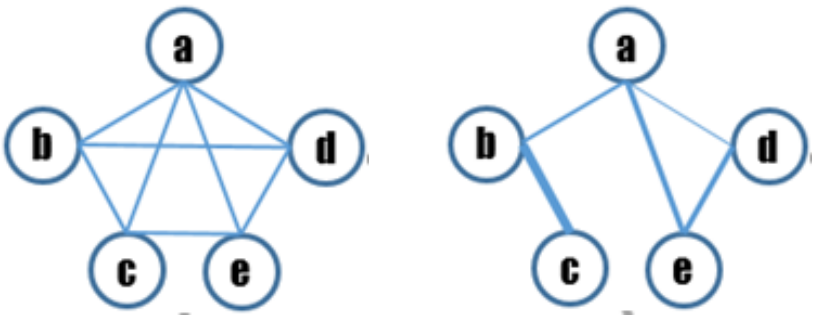
doi:10.6342/NTU202201326

Co-citation

Co-citation is a citation relationship between two citation entities based on how the later publications co-cite their related references. Although the modified models may not improve the result at the nodes/relationships level, these models usually result in clusters with a higher textual coherence. Accordingly, applying semantic analysis improves the textual coherence and helps researchers detect emerging research subfields.

A noteworthy divergence exists between the results of the revised models when analyzing BCS and CCS. The difference may be due to how the citation network of the classical model is affected when adding a new article. For each article included in the research based on CC, their citing references can form a clique network by CC, in which all references are related to each other equally. The whole CC network is composed of these cliques networks. However, the experience of using references and the findings of the studies reviewed above disagree with weighting CCS equally. The modified models can discriminate relationships and may decompose a clique network into a network that reveals more information, as shown in Figure 4-21. Hence, the modified models have the advantage of constructing a CC network because they can adjust the CCS based on how authors use two references in their articles.

Figure 4.21:
The effects when further discriminating CCS



(1) (2)

Notes: (1) represents the CC network, a clique network, of an article citing a~e based on the classical model. In the network of the classical model, all references a~e are co-cited with each other, and their CCS are the same. (2) shows the possible network built by the modified models. The modified model may remove or enhance some citation relationships. Hence, the differences between different nodes may be stressed.

doi:10.6342/NTU202201326

Accordingly, the frequency model may be inappropriate to modify the CC network based on the classical model. Although references with high ITCs in an article tend to be more influential for this article, these references may not necessarily relate to each other more. Instead, these references may highly relate to the references with low ITCs if their topics are similar. The textual coherence of the clustering result of the frequency model, lower than that of other revised models, supports this suggestion.

Compared with the classical and frequency models, the distance model can further discriminate CC relationships because authors usually use references based on subjects of references. In general, one paragraph discusses one topic; references used in the same paragraph relate to this topic hence. Within the same sentence or clause, authors shall use references that focus on a narrow topic. Therefore, the distance between two citations may reflect whether the two references' subjects are similar according to the writing structure. As a result, the distance model can differ the tight relationships from the weak relationships. The textual coherence also supports this conclusion. Additionally, the distance model can reveal some research subfields which cannot be identified by other models, e.g., the NLP studies about analyzing Chinese data.

The lexical model is the only modified model with higher textual coherence than the BERT and Wordnet models. In other words, the cluster identified by the lexical model may contain articles which more concentrate on similar topics. Such ability relies on that only the identical terms, excluding the heteronym, will be considered, and only the strong relationships will be kept in the lexical model. Namely, when modifying the CC network by the lexical model, only the references whose tokens in their citances are almost identical, usually the references cited in the same sentence, will get strong CCS. Although the textual coherence shows the advantage of the lexical model, the subject analysis also indicates this model's weaknesses that the identified clusters may be too small to reveal subjects in a discipline in a proper granularity.

When dealing with CC, the BERT and Wordnet models, two models considering semantic perspective, performed well in the present study. The two models adjust the

doi:10.6342/NTU202201326

relationship strength based on the semantic similarity and avoid the possible drawbacks of the frequency and distance models. Compared with the lexical model, both models do not limit to identical terms and can consider the complex relationships, e.g., hypernym, hyponym, and synonym, between words. Therefore, applying semantic analysis in measuring CCS has an advantage in modifying the CC network built by adding an article, as shown in Figure 4-21. Textual coherence and subject analysis also support that both models performed well when the semantic analysis is applied in CC.

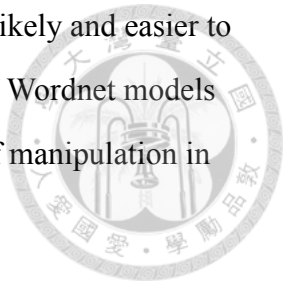
The common weakness of both models is detecting the meaning implied in the writing structure by authors. Among the modified models used in the present study, only the distance model is capable of analyzing such meaning. According to the result of the present study, the distance model may further divide a research field into several small and meaningful subfields. Considering the writing structure and integrating these models are noteworthy research questions for future studies.

Compared with the BERT model, a specific weakness of the Wordnet model is how to decide the correct meaning of a word. When measuring the semantic similarity between two words, the present study simply uses their maximum similarity despite that the two words may refer to totally unrelated meanings. For example, according to the online Cambridge Dictionary, bulb can mean (1) a round root of some plants from which the plant grows and (2) a light bulb. Another word, lamp, is relevant to the second meaning but irrelevant to the first meaning. When measuring the similarity between bulb and lamp based on the Wordnet model, the present study did not find the exact meaning of bulb used by authors but used their maximum similarity instead. Although further analyzing the surrounding text may find out a possible meaning of a word, it will require more computing power and increase the complexity of the model.

Since the BERT model has considered the surrounding text while constructing a language model and providing the sentence embedding, it is unnecessary to figure out the real meaning of each word in a sentence. Namely, when detecting the meaning of a sentence or measuring the semantic similarity between two sentences, the semantic

doi:10.6342/NTU202201326

analysis based on the predict model, e.g., the BERT model, is more likely and easier to consider the word's various meanings. Although both the BERT and Wordnet models are based on the open source Python package tool, the complexity of manipulation in the BERT model is much less than in the Wordnet model.



doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

Chapter 5

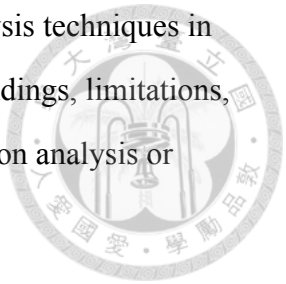
Conclusion

Given the issues studied by the previous research in terms of citation analysis and the advancement of NLP semantic analysis reviewed in Chapter 2, further discriminating citations by analyzing the content of citances instead of counting them equal is possible. Moreover, it is promising for improving applications of citation analysis. The present study aims to improve the methods of citation analysis and their applications, and investigates the effectiveness of applying semantic analysis techniques when conducting citation analysis. The advantages and weaknesses of applying these techniques appear by examining the results of this study. Two different semantic analysis techniques, Wordnet of the dictionary model and BERT of the predict model, were used to modify the methods of measuring three types of citation relationships, namely DC, BC, and CC. The two models based on BERT and Wordnet were compared with the classical model and several methods proposed by the previous studies.

The purpose of this research was further divided into three research questions. The present study proposed several ways of applying semantic analysis in improving the measurement of citation relationships and compared their results with those of the classical model and other modified models based on the reviewed studies. The comparisons were based on three levels: the network level, the cluster level, and the node/relationship level. The answers of these research questions help scholars

doi:10.6342/NTU202201326

understand the possible improvement by utilizing the semantic analysis techniques in applications of citation analysis. This chapter details the research findings, limitations, and suggestions for future studies which would like to conduct citation analysis or improve citation analysis with semantic analysis.



5.1 Research Finding

After examining and comparing the results of citation analysis with the classical and modified models, the present study reports the following findings about the effectiveness of applying NLP techniques in citation analysis.

5.1.1 The conclusion of network analysis

- *When applying NLP techniques in excluding negative citations or weighting citation relationships, its effect on the structure of a citation network depends on the type of citation relationships and networks examined.*

  The debate about the effectiveness of citation analysis and investigation of citation functions raises the question of whether utilizing semantic analysis techniques to remove negative citations may significantly change the outcome of citation analysis. Besides, the previous studies on how to measure BCS/CCS suggests that weighting BCS/CCS in different ways may obviously affect the structures of citation networks. Whether applying semantic analysis techniques in measuring BCS/CCS results in the same outcome remains unknown. Hence, the present study used several network indicators to compare the citation networks, including the whole and core networks, and answer the above questions.

  Although the DC networks of the modified models significantly differed from that of the classical model, considering ITCs is the cause of these differences. Compared with the networks of the frequency model, no noticeable

doi:10.6342/NTU202201326

effects on the DC whole network and core network existed after removing the negative citations identified by the BERT or Wordnet models. Hence, excluding negative citations failed to change the network structure further.

Similarly, measuring BCS based on semantic similarity did not significantly change the structure of the whole network. Nevertheless, the effects on the structures of the core networks exist. The network indicators and JSD indicated the differences between the core networks built by the frequency and the BERT/Wordnet models. The most evident effects existed in the CC citation networks, especially core networks. Reweighting CCS made the structure largely differ from the classical model, and the dissimilarity between the networks based on these modified models was also notable. Therefore, the effect of measuring BCS/CCS with semantic similarity on the network structure varies with the type of citation network. Table 5-1 shows the effect of the network structure when applying semantic analysis to each type of citation relationship.

Table 5.1:
The Effect of Applying Semantic Analysis

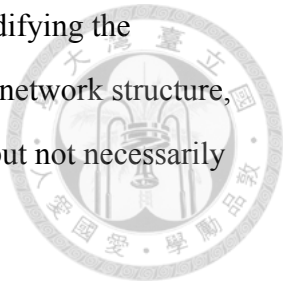
	Type of citation relation ship					
	DC		BC		CC	
Core Network	Obvious	Slight	Obvious	Obvious	Obvious	Obvious
Whole Network	Slight	No	No	No	Slight	Slight

Notes: For each citation relationship, the left/right columns represent the effect compared with the networks of the classical and other modified models, respectively.

Overall, excluding negative citations does not largely change the network structure when compared with the DC network that has been modified by ITCs. This conclusion also elicits two possibilities. One is that excluding them may not help researchers refine the DC citation network to reveal the structure of science from a different perspective before the further breakthrough in identifying negative citations. Another is that removing negative citations may not affect the DC citation network even with high accuracy when identifying them. Examining BC/CC core networks supports that weighting BCS/CCS based on semantic

doi:10.6342/NTU202201326

similarity reveals the different citation networks. Namely, modifying the weighting scheme by semantic analysis techniques affects the network structure, especially the core network, and provides a new perspective, but not necessarily a better one, to observe the structure of science.



5.1.2 The conclusion of cluster analysis

- *Adjusting CCS by sematic similarity improves the clustering results, but modifying DCS and BCS by the methods proposed by the present study do not result in the same effect.*

  According to the reviewed studies in Chapter 2, considering additional variables when measuring BCS/CCS usually results better than the classical model. It may be promising when measuring BCS/CCS by applying semantic analysis techniques. The studies about citation behavior also raise the question of whether removing negative citations affects the result of citation analysis and its related applications significantly. Hence, the present study examined and compared the results at the cluster and node/relationship levels.

  However, according to the result reported in Chapter 4, applying NLP techniques in analyzing DC may not further discriminate the clustering results of considering the ITCs. Although the ARI indicated that the clustering result of the DC core networks based on the classical model significantly differed from those of the modified models, no noticeable differences existed between the results of the modified models. In other words, excluding negative citations does not further affect the clustering results after considering ITCs in the frequency model.

  As to the clustering results of the BC core networks, the difference was more evident than that of DC core networks but still slight. The ARI showed that the diversity between the clustering results from different models is usually minor. The most obvious effects of applying semantic analysis happen in measuring

doi:10.6342/NTU202201326

CCS. ARI indicated that the similarity between the clustering results from different models is usually high. Namely, the semantic analysis largely affected the clustering results and produced evident differences compared with the classical and other modified models.

The present study further evaluated the clustering results with textual coherence. When examining the textual coherence of DC and BC clustering results, neither the evident difference nor a consistent pattern existed. This result agrees with the opinions reported in the previous paragraph: modifying DC and BC with the methods proposed by the present study does not improve the clustering results obviously. When inspecting the clustering results of CC, the textual coherence showed that the three models, i.e., the lexical, BERT, and Wordnet models, improved the clustering results of CC core networks.

Then, the present study identified the subjects of the top clusters in the core networks and examined them. For all DC core networks, the subfields identified were too general to be used to further applications, and removing negative citations does not reveal different subfields either. The subfields and the drawbacks identified from the BC core networks based on different models were similar. It means that modifying BCS by analyzing citances with semantic analysis techniques may not reveal the subfields which differ from those identified by the classical model. In terms of CC, the subfields identified from the core network of the lexical model showed its latent drawback, which might divide the citation entities into too small clusters, although this model had the largest increase in the textual coherence. On the contrary, the clustering results of the BERT and Wordnet models had high textual coherence, and the identified subfields were also appropriate for describing the LIS development between 2010 and 2019. Hence, the present study concludes that adjusting CCS by sematic similarity improves the clustering results.

- *The distance model reveals the relationship which can not be stressed by other models, including the semantic model.*

doi:10.6342/NTU202201326

Although the present study concludes that the clustering result of the CC core networks built on BERT and Wordnet models can better discriminate the entities and identify new emerging issues, the result of the distance model reveals some different clusters which are meaningful. When writing articles, authors use and organize their citing references for various purposes. Texts reflect how authors describe or comment on the citing references but not exactly how they organize or categorize these references. Some criteria for organizing citing references are hard to be identified by analyzing the semantic meanings of citances. Hence, not a single modified model can fully replace all other models in the current study. When using modified models, researchers may choose by their intentions or the perspectives they want to focus on. To consider how to integrate the information revealed by models may be a noteworthy issue for future studies.

5.1.3 The conclusion of node/relationship analysis

- *Positive citations identified by the sentimental polarity may be able to find the references with strong influence.*

The previous studies indicated that the influence of citations with different features varies. Whether discriminating citations by their sentimental polarity helps researchers differentiate the possible references with high influence from others is noteworthy. After categorizing references by their citations' sentimental polarity and examining their cited times, the result showed that the average citation counts of the journal articles, which have been cited positively at least once, are usually higher than the average citation counts of those without being cited positively. Hence, identifying positive citances by analyzing their text content with semantic analysis can find the articles that may be cited more times than the average citation counts of the articles published in the same journal and year.

Two factors make this tendency clearer. Firstly, combining the results of

doi:10.6342/NTU202201326

multiple classifiers can further classify the references into several categories, and the result is more likely to help researchers identify the references with higher citation counts. The result suggests that combining multiple classifiers as a voting system may be an approach to measure the sentiment represented by authors. Researchers who intend to identify positive citations and influential publications can consider using multiple NLP tools to judge the sentimental polarity.

Secondly, the accumulation of citation counts of a publication usually takes several years, depending on the discipline. When focusing on the works which have been cited positively and published for at least five years, the gap in average citation counts between them and those without positive citations becomes clearer.

However, because testing the causal relationship between the positive citations and the high citation counts is not included in this study, the current conclusion only indicates the possibility of using this method to identify the possible works with influence instead of claiming that researchers can predict highly cited articles by this method. Overall, the NLP tools can help researchers identify positive citations and investigate their features, and positive citations may help researchers detect influential references.

- *Discriminating negative citations is difficult for the available NLP tools, especially for those of predict model.*

The present study classified citations by their sentimental polarity into positive, neutral, and negative citations. Identifying positive citations helps scholars differentiate the possible citations with strong influence, but identifying negative citations is much more challenging.

Although the previous studies do not provide a consistent percentage for the negative citations, the numbers reported by them are always higher than 1%. Hence, the proportion of negative citations in the BERT model, less than 0.1%, is

doi:10.6342/NTU202201326

too low to be reasonable. It seems that this model obviously favored classifying citations into the neutral and positive classes in the present study. Although the Macro-F1 of the Wordnet model was much lower than that of the BERT model, the problem of identifying negative citations is relatively minor. Overall, the ability to identify negative citations is the weakness of the available NLP tools.

As discussed in Section 4.5.4, the primary reason should be the differences between how authors represent their critical opinions in the academic writing style and how people comment on their bad experiences or feelings. The NLP tools used by the present study are designed for processing general texts, not a specific genre of texts. For researchers trying to studying issues of negative citations on a large scale, further steps like combining other techniques to identify negative citations may be necessary to increase its accuracy. This approach may be also used for the NLP tools of predict model. Additionally, fine-tuning the language model with an appropriate training set or building a language model for this specific purpose may be another possible solution. To sum up, there is a large room for enhancing the ability of NLP tools in identifying negative citations.

- *Modifying BCS/CCS by semantic similarity of citances may not improve the ability of identifying more relevant relationships.*

Recently, NLP tools have shown good performance in the tasks of processing general language. Such a result shows that the state-of-the-art NLP tools may have be much better in investigating semantic meanings. Given that the purpose of BC and CC is to measure how two citation entities relate to each other, modifying BC and CC by semantic similarity may increase the chance of improving the ranking that the pairs of citation entities have highly topic similarity.

However, according to this research, the topic similarity sum of the pairs with strong BCS/CCS measured by the models based on semantic similarity is

doi:10.6342/NTU202201326

not significantly higher than those of the classical and modified models. The present study calculates and compares the topic similarity sum of the top 100 BCS/CCS ranked by different models. Although the sum of the top 100 BCS of the BERT/Wordnet models was the highest occasionally, the gaps in the sums between different models were small. When examining the results of CC, the present study also found similar phenomena. Hence, no significant improvement in the ranks of pairs with strong BCS/CCS was found after modifying BC/CC by semantic analysis. Namely, weighting BCS/CCS by semantic similarity between citances may not give advantage to the pairs of works which the topics of two works have lots in common.

5.2 Research Limitations

The findings reported above are limited by several factors. A major limitation is that the current findings are only appropriate for English documents. When applying NLP techniques to analyzing the works written in other languages, further studies are necessary due to the diversity of human languages and the functionality of NLP tools for different languages. Although the academic writing style should be similar in different languages, the two reasons make the difficulty of analyzing texts in different languages vary.

Because of the diversity of human languages, the procedures used to process the texts of different languages vary. Hence, the variation in procedures also changes the difficulty of analyzing texts of different languages. For example, how to segment a series of tokens into words is a serious question when processing Chinese, Korean, and Japanese. Similarly, for German, how to split a compound word, *kompositum*, into meaningful words is also a challenge.

Besides, another question is the power of the available NLP tools for a language. A well-developed tool can significantly reduce the difficulty of processing texts and

doi:10.6342/NTU202201326

increase the accuracy of analyzing works. The comparison between the effectiveness of the NLP tools for different languages is beyond the purposes of this study. Further investigation is necessary for future studies which try to classify or weight citation relationships of texts written in other languages.

The difference in citation behaviors is another major limitation. The proportion of positive and negative citations varies in different disciplines. Although the present study concludes that removing negative citations does not significantly affect the structure of DC core networks, this conclusion is limited to the result of citation analysis regarding LIS. In this study, the percentage of the negative citations identified by the BERT and Wordnet models were 0.05% and 4%, respectively. However, according to Tabatabaei (2013), the proportion of negative citations reported by some previous studies is more than 10%. Besides, Lin (2018) also reports that the proportion of negational citations varied in different disciplines and the maximum is 9.1% in History. Hence, the possibility of affecting the structure of citation networks after removing negative citations can not be excluded. Further studies are necessary for making sure that the conclusion is appropriate for another discipline.

5.3 Suggestions for Future Research

1. *To investigate the effectiveness of adjusting BCS by semantic similarity between two works based on their metadata.*

   According to the research findings, using the information extracted from the text body to measure BCS fails to improve the result of BC and its related applications. Although measuring BCS by analyzing citation content does not enhance the applications of BC, future studies can explore how to measure BCS of two citation entities by the semantic similarity between their titles, abstracts, or descriptions. The approach used in this study only increases the BCS of a pair of works when any of their references are identical and how authors use the identical references is similar. Instead of this approach, future studies may

doi:10.6342/NTU202201326

investigate how to decide the BCS which a reference without an identical reference cited by another work provides for a pair of works. Liu (2017) has verified the efficiency of measuring BCS by the lexical similarity of titles. The techniques for semantic analysis used in the present study may improve the method of weighting BCS based on the similarity between titles and the related applications.

2. *To verify the effectiveness of new NLP techniques.*

   With the advance in NLP techniques, more language models or techniques for semantic analysis will be proposed in the future. This study aims to investigate the effectiveness of using NLP techniques for semantic analytics in citation analysis, and the emergence of new techniques may overcome the drawbacks indentified by this study. Investigating the possibility and capability of these techniques are noteworthy.

3. *To explore the efficiency of combining different NLP tools when conducting citation analysis.*

   Another direction is exploring how to combine multiple NLP tools. The present study classified the references into three classes based on the results of two NLP tools, and the further classification provides more information about DC. Hence, combining multiple NLP tools may be helpful in further analyzing citations and improving citation analysis. An option is in line with the present study that researchers may classify citations based on the classification results of multiple tools. The purpose is to build a voting system appropriate for classifying citations. Another one is to explore how to combine sentimental polarity and semantic similarity. By combining both of them, the result may identify a closer relationship or similar nodes.

4. *To examine whether the conclusions are appropriate for analyzing the works in languages different from English.*

   The current study only examines the language models for English. For other

doi:10.6342/NTU202201326

languages, further examinations may be necessary to answer whether any unexpected difficulty exists when analyzing the works in other languages. As mentioned in Section 5.2, the development of NLP tools in different languages is different. For each language, their specific characteristics affect how well researchers can analyze their semantic meanings. Future studies can aim at investigating the ability of NLP tools to analyze academic works in different languages or multi-language.

5. *To study how to apply the semantic analysis in measuring different kinds of citation entities.*

   In this study, only the citation relationships between different publications are examined. Three kinds of citation entities are concluded from the previous studies, including works, authors, and subjects. Researcher can further study how to apply NLP tools when analyzing citation relationships between different authors and subjects.

6. *To inspect whether the conclusion of this study can be suitable for other disciplines.*

   Examining the conclusions of this study with datasets from other disciplines is also worth trying. Although collecting the machine-readable full texts with structure good enough for analysis might be difficult on a large scale, the examination and its result help researchers understand whether these conclusions can be applied to other disciplines.

7. *To investigate the features of citations of different functions.*

   The suggestions mentioned above are about further improving the application of semantic analysis in studies of citation analysis. The following advice is about future studies applying semantic analysis in improving the studies about citation analysis.

   As reviewed in Chapter 2, the previous studies have proposed various citation functions and classified citations according to their scheme. The current

doi:10.6342/NTU202201326

study has shown the possibility of classifying citations by utilizing the fine-tuned classifier based on BERT. Future studies about investigating citation functions can utilize this approach. The result of manually labeling the selected citations can be used as the training set to fine-tune the classifier for identifying citations of different functions on a large scale. If the classifier is capable of classifying citations correctly, researchers can study the roles and effects of the citations with different functions from a macro perspective as well as provide more empirical data for developing the citation theory.

8. *To investigate how to integrate the information revealed by different models.*

The previous related studies focused on evaluating the appropriate applications for different kinds of citation relationships or exploring whether the models proposed by these studies were better than the original model. In the conclusions, however, the present study reports that the information extracted by different models may vary. Hence, instead of figuring out a better solution, researchers may try to integrate the information revealed by different approaches and provide a comprehensive description for the scientific structure.

doi:10.6342/NTU202201326

doi:10.6342/NTU202201326

References

Abrizah, A., Noorhidawati, A., & Zainab, A. N. (2015). LIS journals categorization in the journal citation report: A stated preference study. *Scientometrics*, *102*(2), 1083-1099. doi: 10.1007/s11192-014-1492-3

Abu-Jbara, A., Ezra, J., & Radev, D. (2013). Purpose and polarity of citation: Towards NLP-based bibliometrics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (p. 596-606). Atlanta, Georgia: Association for Computational Linguistics.

Ahlgren, P., Chen, Y., Colliander, C., & van Eck, N. J. (2020). Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. *Quantitative Science Studies*, 1-20. doi: 10.1162/qss_a_00027

Ahmed, T., Johnson, B., Oppenheim, C., & Peck, C. (2004). Highly cited old papers and the reasons why they continue to be cited. Part II., The 1953 Watson and Crick article on the structure of DNA. *Scientometrics*, *61*(2), 147-156. doi: 10.1023/B:SCIE.0000041645.60907.57

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 Student Session* (p. 81-87). Portland, OR, USA: Association for Computational Linguistics.

Athar, A., & Teufel, S. (2012a). Context-enhanced citation sentiment detection. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*

doi:10.6342/NTU202201326

(p. 597-601). Montréal, Canada: Association for Computational Linguistics.

Athar, A., & Teufel, S. (2012b). Detection of implicit citations for sentiment detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (p. 18-26). Jeju Island, Korea: Association for Computational Linguistics.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (p. 238-247). Baltimore, Maryland: Association for Computational Linguistics. doi: 10.3115/v1/P14-1023

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*, 1137-1155.

Bertin, M., Atanassova, I., Gingras, Y., & Larivière, V. (2016). The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology*, *67*(1), 164-177. doi: 10.1002/asi.23367

Bonzi, S., & Snyder, H. W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics*, *21*(2), 245-254. doi: 10.1007/BF02017571

Bornmann, L., & Daniel, H.-D. (2007). Functional use of frequently and infrequently cited articles in citing publications: A content analysis of citations to articles with low and high citation counts. *European Science Editing*, *34*(2), 35-38.

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., … Börner, K. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE*, *6*(3), e18029. doi: 10.1371/journal.pone.0018029

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. doi: 10.1002/asi.22896

Boyack, K. W., van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of*

doi:10.6342/NTU202201326

Informetrics, *12*(1), 59-73. doi: 10.1016/j.joi.2017.11.005

Brooks, T. A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, *36*(4), 223-229. doi: 10.1002/asi.4630360402

Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, *37*(1), 34-36. doi: 10.1002/(SICI)1097-4571(198601)37:1<34::AID-ASI5>3.0.CO;2-0

Bu, Y., Wang, B., Huang, W.-B., Che, S., & Huang, Y. (2018). Using the appearance of citations in full text on author co-citation analysis. *Scientometrics*, *116*(1), 275-289. doi: 10.1007/s11192-018-2757-z

Callahan, A., Hockema, S., & Eysenbach, G. (2010). Contextual cocitation: Augmenting cocitation analysis and its applications. *Journal of the American Society for Information Science and Technology*, *61*(6), 1130-1143. doi: 10.1002/asi.21313

Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, *51*(7), 635-645. doi: 10.1002/(SICI)1097-4571(2000)51:7<635::AID-ASI6>3.0.CO;2-H

Catalini, C., Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences*, *112*(45), 13823-13826. doi: 10.1073/pnas.1502280112

Chandrasekharan, S., Zaka, M., Gallo, S., Zhao, W., Korobskiy, D., Warnow, T., & Chacko, G. (2020). Finding scientific communities in citation graphs: Articles and authors. *Quantitative Science Studies*, 1-20. doi: 10.1162/qss_a_00095

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, *5*(4), 423-441.

Clark, K. E. (1957). Indices of eminence. In *America's psychologists: A survey of a growing profession* (p. 26-61). Washington, D. C.: American Psychological Association.

doi:10.6342/NTU202201326

Cole, J., & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the ”Science Citation Index”. *The American Sociologist*, *6*(1), 23-29.

Cole, J. R., & Cole, S. (1972). The Ortega hypothesis: Citation analysis suggests that only a few scientists contribute to scientific progress. *Science*, *178*(4059), 368-375. doi: 10.1126/science.178.4059.368

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, *32*(3), 377-390. doi: 10.2307/2091085

Cronin, B., & Pearson, S. (1990). The export of ideas from information science. *Journal of Information Science*, *16*(6), 381-391. doi: 10.1177/016555159001600606

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805 [cs]*.

Ding, Y., Chowdhury, G., & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997. *Journal of Information Science*, *25*(1), 67-78. doi: 10.1177/016555159902500107

Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics*, *7*(3), 583-592. doi: 10.1016/j.joi.2013.03.003

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, *65*(9), 1820-1833. doi: 10.1002/asi.23256

Dong, C., & Schäfer, U. (2011, Nov). Ensemble-style self-training on citation classification. In *Proceedings of 5th international joint conference on natural language processing* (p. 623-631). Chiang Mai, Thailand: Asian Federation of Natural Language Processing.

doi:10.6342/NTU202201326

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. doi: 10.1002/asi.20707

Eom, S. B. (2003). *Author co-citation analysis using custom bibliographic databases: An introduction to the SAS approach*. Lewiston, N.Y: Edwin Mellen Press.

Eom, S. B. (2009). *Author cocitation analysis: Quantitative methods for mapping the intellectual structure of an academic discipline*. Hershey, PA: Information Science Reference.

Eto, M. (2007). Multivalued co-citation measure based on semantic distance between co-cited papers in a citing paper: A case study focused on enumeration of citations. *Library and Information Science*(58), 49-67.

Eto, M. (2008). A new co-citation measure based on structures of citing papers. 情報処理学会論文誌データベース（*TOD*）, *49*(7), 1-15.

Eto, M. (2012). Evaluations of context-based co-citation searching. *Scientometrics*, *94*(2), 651-673. doi: 10.1007/s11192-012-0756-z

Eto, M. (2019). Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information. *Information Processing & Management*, *56*(6), 102046. doi: 10.1016/j.ipm.2019.05.007

Fiala, D., Rousselot, F., & Ježek, K. (2008). PageRank for bibliographic networks. *Scientometrics*, *76*(1), 135-158. doi: 10.1007/s11192-007-1908-4

Frost, C. O. (1979). The use of citations in literary research: A preliminary classification of citation functions. *The Library Quarterly: Information, Community, Policy*, *49*(4), 399-414.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108-111. doi: 10.1126/science.122.3159.108

Garfield, E. (1957). Breaking the subject index barrier - a citation index for chemical patents. *Journal of the Patent Office Society*, *39*(8), 583-595.

doi:10.6342/NTU202201326

Garfield, E. (1964). "Science Citation Index"–A new dimension in indexing. *Science*, *144*(3619), 649-654. doi: 10.1126/science.144.3619.649

Garfield, E. (1965). Can citation indexing be automated? In *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings* (p. 189-192). Washington: National Bureau of Standards Miscellaneous Publication.

Garfield, E. (1970). Citation indexing for studying science. *Nature*, *227*(5259), 669-671. doi: 10.1038/227669a0

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(4060), 471-479.

Garfield, E. (1977a). 250 most cited primary authors, 1961-1975 .1. how names were selected. *Current Contents*(49), 5-15.

Garfield, E. (1977b). 250 most-cited primary authors, 1961-1975 .2. correlation between citedness, Nobel-Prizes, and academy memberships. *Current Contents*(50), 5-15.

Garfield, E. (1977c). 250 most-cited primary authors, 1961-1975 .3. each authors most-cited publication. *Current Contents*(51), 5-20.

Ghosh, S., Das, D., & Chakraborty, T. (2018). Determining sentiment in citation text and analyzing its impact on the proposed ranking index. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (p. 292-306). Cham: Springer International Publishing.

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, *7*(1), 113-122. doi: 10.1177/030631277700700112

Gipp, B., & Beel, J. (2009). Citation proximity analysis (CPA) –A new approach for identifying related work based on co-citation analysis. In *Proceedings of ISSI 2009*.

Giuffrida, C., Abramo, G., & D'Angelo, C. A. (2019). Are all citations worth the same? Valuing citations by the value of the citing items. *Journal of Informetrics*, *13*(2), 500-514. doi: 10.1016/j.joi.2019.02.008

Goodarzi, M., Mahmoudi, M. T., & Zamani, R. (2014). A framework for sentiment

doi:10.6342/NTU202201326

analysis on schema-based research content via lexica analysis. In *7'th International Symposium on Telecommunications (IST'2014)* (p. 405-411).

Hargens, L. L. (1986). Migration patterns of U. S. Ph. D. s among disciplines and specialties. *Scientometrics*, *9*(3), 145-164. doi: 10.1007/BF02017238

Harwood, N. (2008). Citers' use of citees' names: Findings from a qualitative interview-based study. *Journal of the American Society for Information Science and Technology*, *59*(6), 1007-1011. doi: https://doi.org/10.1002/asi.20789

Herlach, G. (1978). Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science*, *29*(6), 308-310. doi: 10.1002/asi.4630290608

Hernández-Álvarez, M., Gomezsoriano, J., & Martínez-Barco, P. (2017). Citation function, polarity and influence classification. *Natural Language Engineering*, *23*(4), 561-588.

Hooten, P. A. (1991). Frequency and functional use of cited documents in information science. *Journal of the American Society for Information Science*, *42*(6), 397-404. doi: 10.1002/(SICI)1097-4571(199107)42:6<397::AID-ASI2>3.0.CO;2-N

Hou, W.-R., Li, M., & Niu, D.-K. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays*, *33*(10), 724-727. doi: https://doi.org/10.1002/bies.201100067

Hsiao, T.-M., & Chen, K.-H. (2017). Yet another method for author co-citation analysis: A new approach based on paragraph similarity. *Proceedings of the Association for Information Science and Technology*, *54*(1), 170-178. doi: 10.1002/pra2.2017.14505401019

Hsiao, T.-M., & Chen, K.-H. (2018). How authors cite references? A study of characteristics of in-text citations. *Proceedings of the Association for Information Science and Technology*, *55*(1), 179-187. doi: 10.1002/pra2.2018.14505501020

doi:10.6342/NTU202201326

Hsiao, T.-M., & Chen, K.-H. (2019). Word bibliographic coupling: Another way to map science field and identify core references. *Proceedings of the Association for Information Science and Technology*, *56*(1), 107-116. doi: 10.1002/pra2.10

Hsiao, T.-M., & Chen, K.-H. (2020). The dynamics of research subfields for library and information science: An investigation based on word bibliographic coupling. *Scientometrics*, *125*(1), 717-737. doi: 10.1007/s11192-020-03645-9

Hu, Z., Lin, G., Sun, T., & Hou, H. (2017). Understanding multiply mentioned references. *Journal of Informetrics*, *11*(4), 948-958. doi: https://doi.org/10.1016/j.joi.2017.08.004

Huang, M., Shaw, W.-C., & Lin, C.-S. (2015). One category, two communities: Subfield differences in "information science and library science" in journal citation reports. *Scientometrics*, *119*(2), 1059-1079. doi: 10.1007/s11192-019-03074-3

Huang, W., Wang, B., Bu, Y., & Min, C. (2018). A study on scientometrics of co-citation analysis of keywords. *Information and Documentation Services*(2), 37-42.

Hurt, C. (1987). Conceptual citation differences in science, technology, and social sciences literature. *Information Processing & Management*, *23*(1), 1-6. doi: 10.1016/0306-4573(87)90033-1

Jeong, Y. K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics*, *8*(1), 197-211. doi: 10.1016/j.joi.2013.12.001

Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, *16*(3), 179-184. doi: 10.1002/asi.5090160305

Kessler, M. (1963a). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10-25. doi: 10.1002/asi.5090140103

Kessler, M. (1963b). An experimental study of bibliographic coupling between technical papers. *IEEE Transactions on Information Theory*, *9*(1), 49-51. doi: 10.1109/TIT.1963.1057800

Kim, H. J., Jeong, Y. K., & Song, M. (2016). Content- and proximity-based author

doi:10.6342/NTU202201326

co-citation analysis using citation sentences. *Journal of Informetrics*, *10*(4), 954-966. doi: https://doi.org/10.1016/j.joi.2016.07.007

Kim, I. C., & Thoma, G. R. (2015). Automated classification of author's sentiments in citation using machine learning techniques: A preliminary study. In *2015 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)* (p. 1-7). doi: 10.1109/CIBCB.2015.7300319

Krampen, G., Becker, R., Wahner, U., & Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics*, *71*(2), 191-202. doi: 10.1007/s11192-007-1659-2

Lin, C.-S. (2018). An analysis of citation functions in the humanities and social sciences research from the perspective of problematic citation analysis assumptions. *Scientometrics*, *116*(2), 797-813.

Lipetz, B.-A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *American Documentation*, *16*(2), 81-90. doi: 10.1002/asi.5090160207

Liu, R.-L. (2017). A new bibliographic coupling measure with descriptive capability. *Scientometrics*, *110*(2), 915-935. doi: 10.1007/s11192-016-2196-7

Liu, R.-L., & Hsu, C.-K. (2018). Issue-based clustering of scholarly articles. *Applied Sciences*, *8*(12), 2591. doi: 10.3390/app8122591

Liu, S., & Chen, C. (2011a). The effects of co-citation proximity on co-citation analysis. *Proceedings of ISSI 2011 - The 13th International Conferenceon Scientometrics and Informetrics*, 474-484.

Liu, S., & Chen, C. (2011b). The proximity of co-citation. *Scientometrics*, *91*(2), 495-511. doi: 10.1007/s11192-011-0575-7

Ma, R., Dai, Q., Ni, C., & Li, X. (2009). An author co-citation analysis of information science in China with Chinese Google Scholar search engine, 2004–2006. *Scientometrics*, *81*(1), 33-46. doi: 10.1007/s11192-009-2063-x

MacRoberts, M. H., & MacRoberts, B. R. (1986). Quantitative measures of

doi:10.6342/NTU202201326

communication in science: A study of the formal level. *Social Studies of Science*, *16*(1), 151-172. doi: 10.1177/030631286016001008

MacRoberts, M. H., & MacRoberts, B. R. (1987). Another test of the normative theory of citing. *Journal of the American Society for Information Science*, *38*(4), 305-306. doi: 10.1002/(SICI)1097-4571(198707)38:4<305::AID-ASI11>3.0.CO;2-I

MacRoberts, M. H., & MacRoberts, B. R. (1988). Author motivation for not citing influences: A methodological note. *Journal of the American Society for Information Science*, *39*(6), 432-433. doi: 10.1002/(SICI)1097-4571(198811)39:6<432::AID-ASI8>3.0.CO;2-2

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, *36*(3), 435-444. doi: 10.1007/BF02129604

MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*, *69*(3), 474-482. doi: 10.1002/asi.23970

Marshakove, I. (1973). System of document connections based on references. *Nauchno-Tekhnicheskaya Informatsiya*, *2*(6), 3-8.

McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, *41*(6), 433-443. doi: 10.1002/(SICI)1097-4571(199009)41:6<433::AID-ASI11>3.0.CO;2-Q

McCain, K. W., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, *17*(1-2), 127-163. doi: 10.1007/BF02017729

Merton, R. K. (1973). The normative structure of science. In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (p. 267-278). Chicago: University of Chicago Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781 [cs]*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed

doi:10.6342/NTU202201326

representations of words and phrases and their compositionality. *arXiv:1310.4546 [cs, stat]*.

Milard, B. (2014). The social circles behind scientific references: Relationships between citing and cited authors in chemistry publications. *Journal of the Association for Information Science and Technology*, *65*(12), 2459-2468. doi: 10.1002/asi.23149

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1-28. doi: 10.1080/01690969108406936

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, *5*(1), 86-92.

Moya-Anegón, F., Vargas-Quesada, B., Victor, Herrero-Solana, Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, *61*(1), 129-145. doi: 10.1023/B:SCIE.0000037368.31217.34

Murugesan, P., & Moravcsik, M. J. (1978). Variation of the nature of citation measures with journals and scientific specialties. *Journal of the American Society for Information Science*, *29*(3), 141-147. doi: https://doi.org/10.1002/asi.4630290307

Nakov, P. I., Schwartz, A. S., & Hearst, M. A. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR'04 workshop on Search and Discovery in Bioinformatics.*

Narin, F., Carpenter, M., & Berlt, N. C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science*, *23*(5), 323-331. doi: 10.1002/asi.4630230508

Nassiri, I., Masoudi-Nejad, A., Jalili, M., & Moeini, A. (2013). Normalized Similarity Index: An adjusted index to prioritize article citations. *Journal of Informetrics*, *7*(1), 91-98. doi: 10.1016/j.joi.2012.08.006

doi:10.6342/NTU202201326

Newman, M. E. J. (2004, Jun). Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, *69*, 066133. doi: 10.1103/PhysRevE.69.066133

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609-641. doi: 10.1002/aris.2007.1440410120

O'Connor, J. (1982). Citing statements: Computer recognition and use to improve retrieval. *Information Processing & Management*, *18*(3), 125-131. doi: 10.1016/0306-4573(82)90036-X

Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, *29*(5), 225-231. doi: 10.1002/asi.4630290504

Peritz, B. C. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, *5*(5), 303-312. doi: 10.1007/BF02147226

Persson, O. (2010a). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, *4*(3), 415-422. doi: 10.1016/j.joi.2010.03.006

Persson, O. (2010b). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, *4*(3), 415-422. doi: https://doi.org/10.1016/j.joi.2010.03.006

Price, D., & Gursey, S. (1976). Studies in scientometrics .2. relation between source author and cited author populations. *International Forum on Information and Documentation*, *1*(3), 19-22.

Price, D. J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510-515. doi: 10.1126/science.149.3683.510

Priem, J., Taraboerlli, D., Groth, P., & Neylon, C. (2015). *Altmetrics: A manifesto – altmetrics.org.* Retrieved from `http://altmetrics.org/manifesto/`

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embedding using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing.*

Saitoh, K. (2019). *Deep learning from scratch 2.* Taipei: GOTOP Information Inc.

Shadish, W. R., Tolliver, D., Gray, M., & Sen Gupta, S. K. (1995). Author judgements

doi:10.6342/NTU202201326

about works they cite: Three studies from psychology journals. *Social Studies of Science*, *25*(3), 477-498. doi: 10.1177/030631295025003003

Sher, I. H., & Garfield, E. (1983). New tools for improving and evaluating the effectiveness of research. *Essays of an information Scientist*, *6*, 503-513.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, *60*(3), 571-580. doi: 10.1002/asi.20994

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269. doi: 10.1002/asi.4630240406

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, *38*(2), 275-293. doi: https://doi.org/10.1007/BF02457414

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799-813. doi: https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics*, *60*(1), 71-79. doi: 10.1023/B:SCIE.0000027310.68393.bc

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies*, *4*(1), 17-40. doi: 10.1177/030631277400400102

Small, H., Tseng, H., & Patek, M. (2017). Discovering discoveries: Identifying biomedical discoveries using citation contexts. *Journal of Informetrics*, *11*(1), 46-62. doi: 10.1016/j.joi.2016.11.001

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*(3), 327-340. doi: 10.1177/030631277800800305

Smith, L. C. (1981). Citation analysis. *Library Trends*, *30*(1), 83-106.

doi:10.6342/NTU202201326

Spiegel-Rösing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science*. doi: 10.1177/030631277700700111

Tabatabaei, N. (2013). *Contribution of information science to other disciplines as reflected in citation contexts of highly cited JASIST papers* (Unpublished doctoral dissertation). McGill University, Montreal.

Tahamtan, I., & Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, *12*(1), 203-216. doi: 10.1016/j.joi.2018.01.002

Tang, R., & Safer, M. A. (2008). Author-rated importance of cited references in biology and psychology publications. *Journal of Documentation*, *64*(2), 246-272. doi: 10.1108/00220410810858047

Teufel, S., Siddharthan, A., & Tidhar, D. (2006a). An annotation scheme for citation function. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue* (p. 80-87). Sydney, Australia: Association for Computational Linguistics.

Teufel, S., Siddharthan, A., & Tidhar, D. (2006b). Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (p. 103-110). Stroudsburg, PA, USA: Association for Computational Linguistics.

Thornley, C., Watkinson, A., Nicholas, D., Volentine, R., Mahmuei, H., Herman, E., … Tenopir, C. (2015). The role of trust and authority in the citation behaviour of researchers. *Information Research*, *20*(3), 1-21.

Tsay, M.-Y., Xu, H., & Wu, C.-W. (2003). Author co-citation analysis of semiconductor literature. *Scientometrics*, *58*(3), 529-545. doi: 10.1023/B:SCIE.0000006878.83104.61

Tseng, Y.-H. (2020). The feasibility of automated topic analysis: An empirical evaluation of deep learning techniques applied to skew-distributed chinese text classification. *Journal of Educational Media and Library Sciences*, *57*(1), 121-144.

doi:10.6342/NTU202201326

Urata, H. (1990). Information flows among academic disciplines in Japan. *Scientometrics*, *18*(3), 309-319. doi: 10.1007/BF02017767

Valenzuela, M., Ha, V., & Etzioni, O. (2015). Identifying meaningful citations. In C. Caragea et al. (Eds.), *Scholarly big data: AI perspectives, challenges, and ideas, papers from the 2015 AAAI workshop* (pp. 21–26). Menlo Park, CA: AAAI Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *arXiv:1706.03762 [cs]*.

Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or, did we op. cit. your idem?. *The Journal of Academic Librarianship*, *1*(6), 19-21.

Waltman, L., Boyack, K. W., Colavizza, G., & van Eck, N. J. (2019). A principled methodology for comparing relatedness measures for clustering publications. *arXiv:1901.06815 [cs]*.

Wan, X., & Liu, F. (2014). Are all literature citations equally important? Automatic citation strength estimation and its applications. *Journal of the Association for Information Science and Technology*, *65*(9), 1929-1938. doi: 10.1002/asi.23083

Wang, P., & White, M. D. (1999). A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *Journal of the American Society for Information Science*, *50*(2), 98-114. doi: 10.1002/(SICI)1097-4571(1999)50:2<98::AID-ASI2>3.0.CO;2-L

White, H. D. (2009). Citation analysis. In *Encyclopedia of Library and Information Sciences, Third Edition* (p. 1012-1026). Taylor & Francis.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, *32*(3), 163-171. doi: 10.1002/asi.4630320302

White, H. D., & Griffith, B. C. (1982). Authors as markers of intellectual space: co-citation in studies of science, technology and society. *Journal of Documentation*, *38*(4), 255-272. doi: 10.1108/eb026731

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation

doi:10.6342/NTU202201326

analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, *49*(4), 327-355. doi: 10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-4

White, M., & Wang, P. (1997a). *Document selection and relevance assessments during a research project* (CLIS Technical Report No. 97-02). College Park, MD: University of Maryland.

White, M., & Wang, P. (1997b). A qualitative study of citing behavior: Contributions, criteria, and metalevel documentation concerns. *Library Quarterly*(67), 122-154.

Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., & Xu, H. (2015). Citation sentiment analysis in clinical trial papers. *AMIA Annual Symposium Proceedings*, *2015*, 1334-1341.

Xu, L., Ding, K., & Lin, Y. (2022). Do negative citations reduce the impact of cited papers? *Scientometrics*, *127*(2), 1161-1186. doi: 10.1007/s11192-021-04214-4

Yaghtin, M., Sotudeh, H., Mirzabeigi, M., Fakhrahmad, S. M., & Mohammadi, M. (2019). In quest of new document relations: evaluating co-opinion relations between co-citations and its impact on Information retrieval effectiveness. *Scientometrics*, *119*(2), 987-1008. doi: 10.1007/s11192-019-03058-3

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313-1326. doi: 10.1002/asi.22680

Zhao, D., Cappello, A., & Johnston, L. (2017). Functions of uni- and multi-citations: Implications for weighted citation analysis. *Journal of Data and Information Science*, *2*(1), 51-69. doi: doi:10.1515/jdis-2017-0003

Zhao, D., & Strotmann, A. (2008a). Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1-10. doi:

doi:10.6342/NTU202201326

10.1002/meet.2008.1450450292

Zhao, D., & Strotmann, A. (2008b). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, *59*(13), 2070-2086. doi: 10.1002/asi.20910

Zhao, D., & Strotmann, A. (2011). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, *87*(1), 115-131. doi: 10.1007/s11192-010-0317-2

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology*, *65*(5), 995-1006. doi: 10.1002/asi.23027

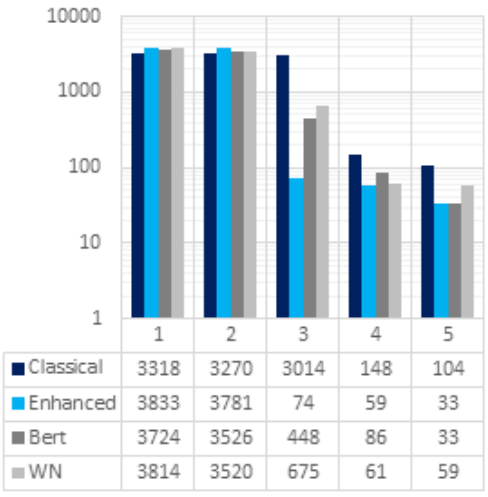
Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, *66*(2), 408-427. doi: https://doi.org/10.1002/asi.23179

Zingg, C., Nanumyan, V., & Schweitzer, F. (2020). Citations driven by social connections? A multi-layer representation of coauthorship networks. *Quantitative Science Studies*, 1-17. doi: 10.1162/qss_a_00092

doi:10.6342/NTU202201326

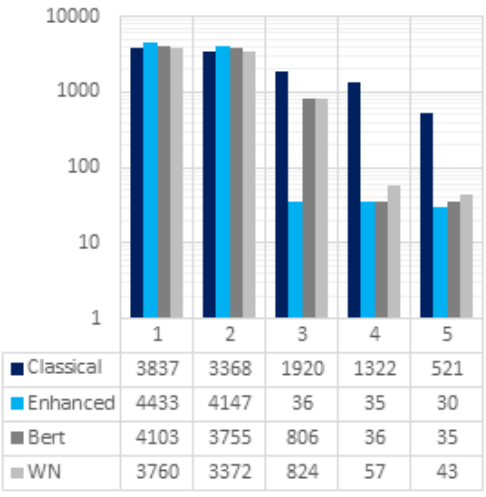
doi:10.6342/NTU202201326

Appendix A: The Size of Top 5 Clusters in the DC Core Networks



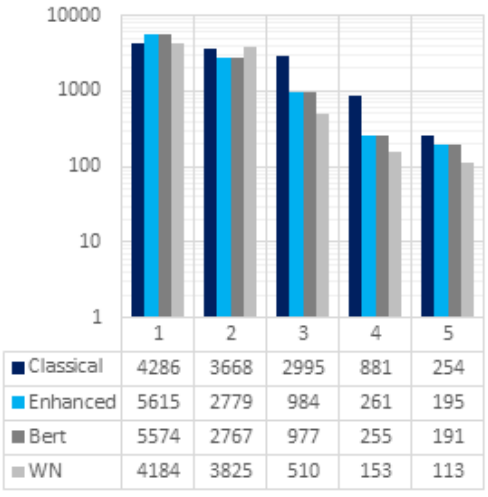
	1	2	3	4	5
Classical	3318	3270	3014	148	104
Enhanced	3833	3781	74	59	33
Bert	3724	3526	448	86	33
WN	3814	3520	675	61	59

2010~2014



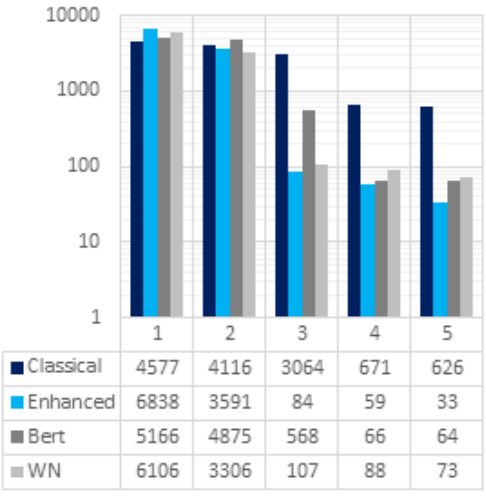
	1	2	3	4	5
Classical	3837	3368	1920	1322	521
Enhanced	4433	4147	36	35	30
Bert	4103	3755	806	36	35
WN	3760	3372	824	57	43

2011~2015



	1	2	3	4	5
Classical	4286	3668	2995	881	254
Enhanced	5615	2779	984	261	195
Bert	5574	2767	977	255	191
WN	4184	3825	510	153	113

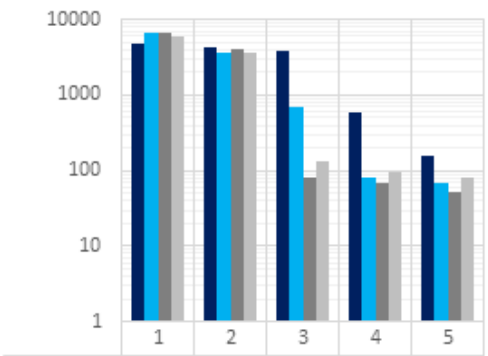
2012~2016



	1	2	3	4	5
Classical	4577	4116	3064	671	626
Enhanced	6838	3591	84	59	33
Bert	5166	4875	568	66	64
WN	6106	3306	107	88	73

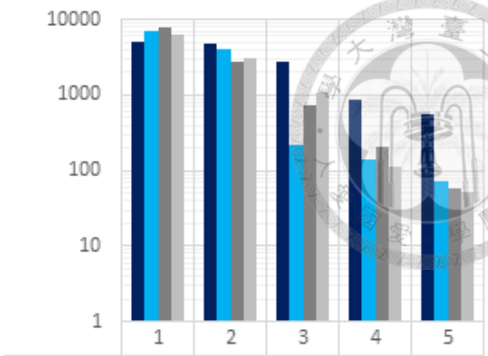
2013~2017

doi:10.6342/NTU202201326



	1	2	3	4	5
Classical	4721	4256	3778	601	159
Enhanced	6636	3684	678	79	68
Bert	6629	4149	79	68	52
WN	6047	3647	136	98	81

2014~2018

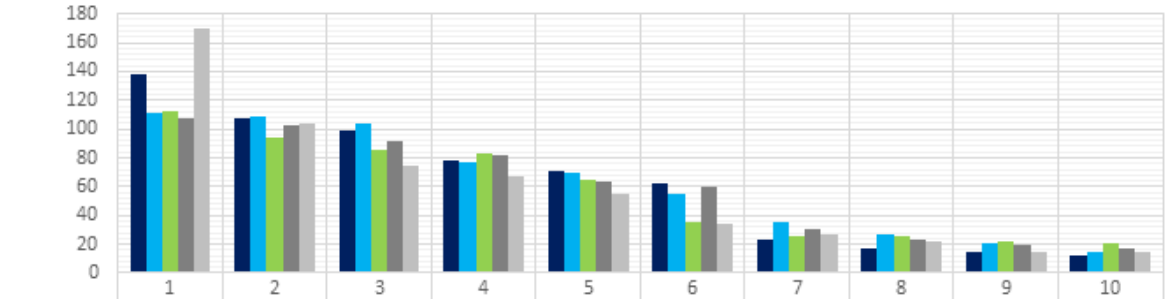


	1	2	3	4	5
Classical	5149	4872	2792	844	564
Enhanced	7013	4028	220	140	73
Bert	7850	2716	738	205	59
WN	6184	3066	1087	114	52

2015~2019

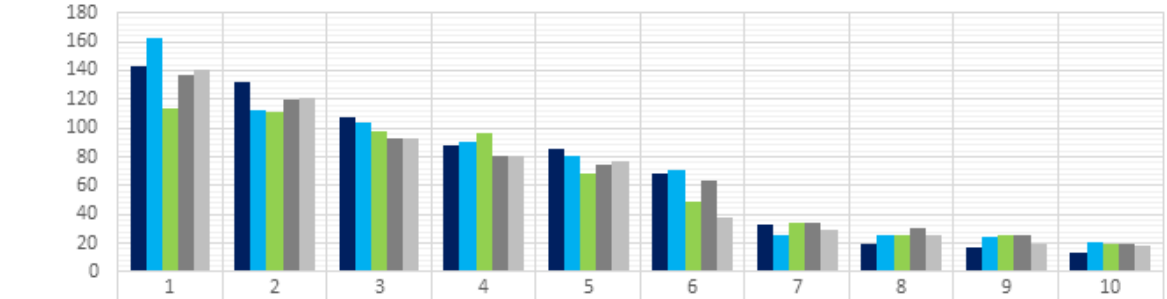
doi:10.6342/NTU202201326

Appendix B: The Size of Top 10 Clusters in the BC Core Networks



	1	2	3	4	5	6	7	8	9	10
Classical	138	108	99	78	71	62	23	17	15	12
Enhanced	111	109	104	77	70	55	35	27	21	15
Lexical	112	94	86	83	65	35	26	26	22	21
Bert	108	103	92	82	64	60	30	23	19	17
WN	170	104	75	67	55	34	27	22	14	14

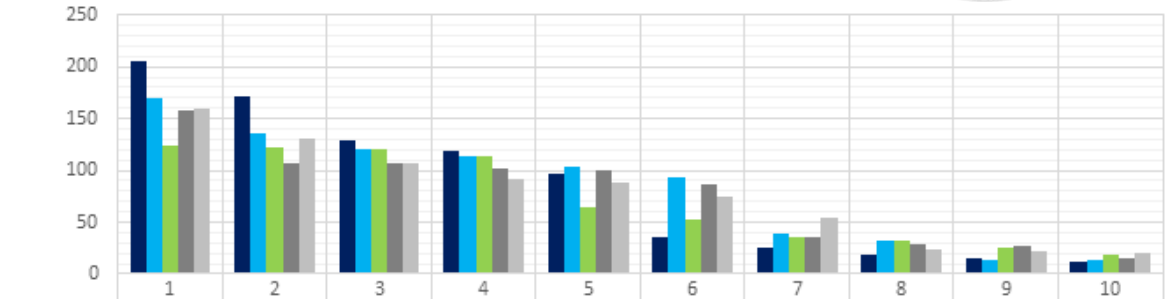
2010~2014



	1	2	3	4	5	6	7	8	9	10
Classical	143	132	107	88	86	68	33	19	17	13
Enhanced	162	112	104	90	81	71	26	25	24	21
Lexical	113	111	98	96	68	49	34	26	25	20
Bert	137	120	93	81	74	63	34	31	26	19
WN	141	121	93	81	77	38	29	25	19	18

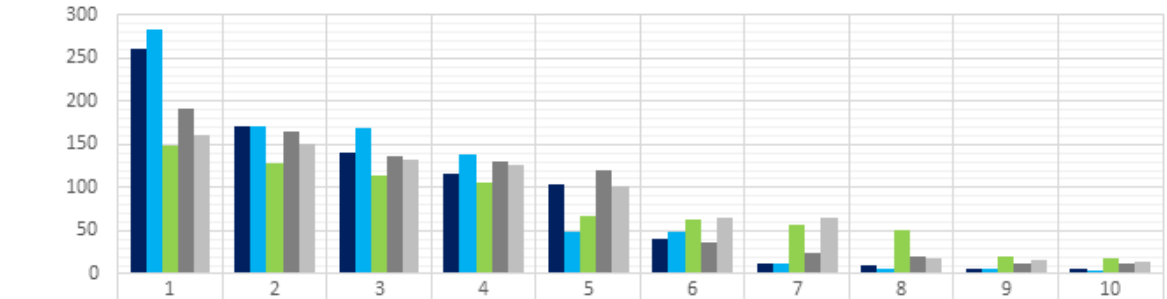
2011~2015

doi:10.6342/NTU202201326



	1	2	3	4	5	6	7	8	9	10
Classical	206	171	129	118	96	35	26	18	16	11
Enhanced	169	135	121	114	103	93	39	33	13	13
Lexical	123	122	120	114	65	53	35	33	26	18
Bert	157	107	106	102	100	87	35	28	27	15
WN	159	130	107	91	88	75	54	24	22	20

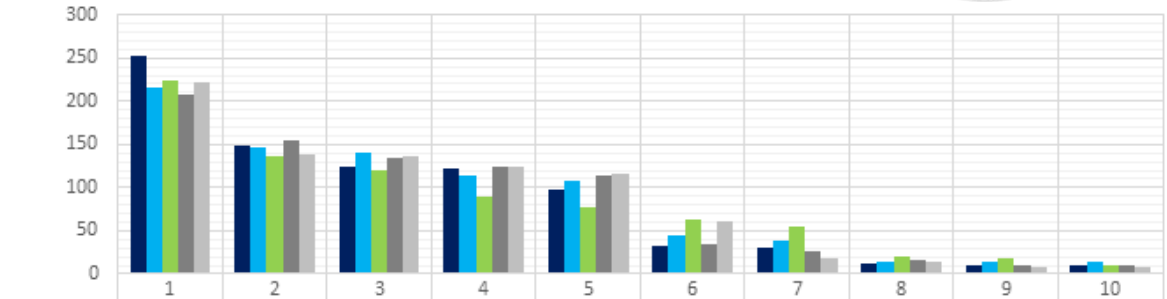
2012~2016



	1	2	3	4	5	6	7	8	9	10
Classical	260	172	140	117	104	40	13	10	6	6
Enhanced	283	171	168	138	49	48	12	6	5	4
Lexical	149	128	113	105	67	64	57	50	20	18
Bert	191	164	136	131	120	37	25	20	13	13
WN	160	151	132	127	101	66	65	18	16	14

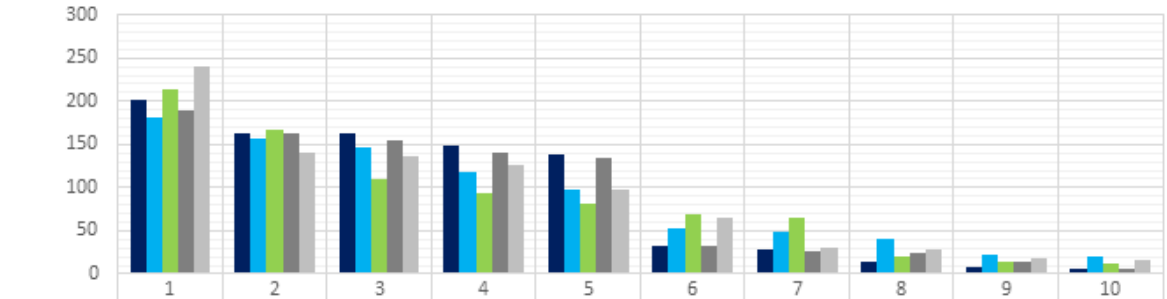
2013~2017

doi:10.6342/NTU202201326



	1	2	3	4	5	6	7	8	9	10
Classical	252	149	125	123	97	32	30	12	10	10
Enhanced	215	146	141	113	108	44	39	14	14	14
Lexical	223	137	121	89	78	64	55	21	19	11
Bert	207	154	134	125	113	34	26	16	11	11
WN	222	138	136	124	117	61	18	14	9	9

2011~2015



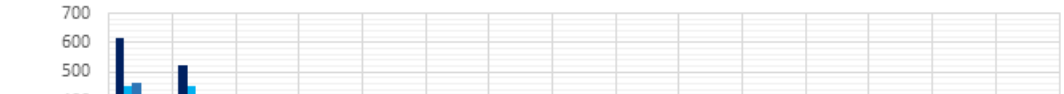
	1	2	3	4	5	6	7	8	9	10
Classical	201	163	162	148	138	33	28	14	8	7
Enhanced	181	156	146	118	98	52	49	41	22	20
Lexical	213	166	110	93	82	70	65	21	15	12
Bert	190	162	155	141	135	32	27	24	14	7
WN	240	140	136	127	97	65	30	29	19	17

2012~2016

doi:10.6342/NTU202201326

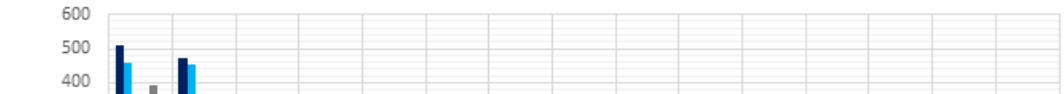
doi:10.6342/NTU202201326

Appendix C: The Size of Top 15 Clusters in the CC Core Network



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	618	523	270	226	209	189	154	130	112	99	79	73	72	52	51
Enhanced	453	451	308	303	182	132	105	95	90	68	63	58	56	51	50
Distance	465	283	192	178	168	140	139	113	112	102	91	89	79	68	68
Lexical	156	92	88	82	65	65	61	61	59	59	58	50	49	45	45
Bert	401	272	266	205	175	171	157	116	73	72	65	64	64	64	58
WN	349	234	215	143	139	135	128	116	99	87	81	78	76	67	58

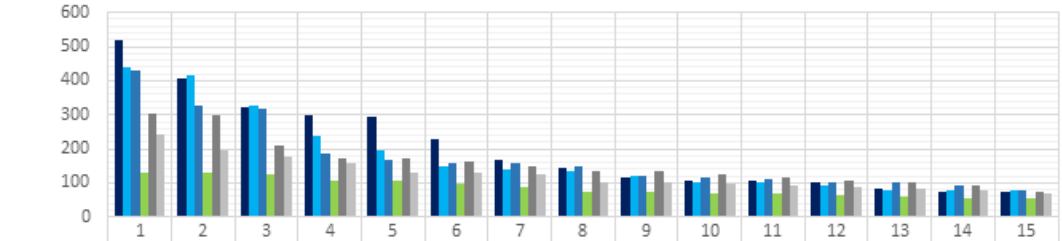
2010~2014



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	509	471	263	211	141	126	123	107	98	96	96	92	91	89	81
Enhanced	459	453	326	164	156	132	107	101	97	95	95	82	82	74	74
Distance	325	306	304	229	182	162	146	145	115	104	103	88	84	77	73
Lexical	147	86	80	72	68	68	66	64	64	61	58	55	55	50	49
Bert	393	248	217	183	141	126	117	116	105	104	90	84	83	82	67
WN	362	212	207	154	117	111	109	94	91	87	80	71	71	68	63

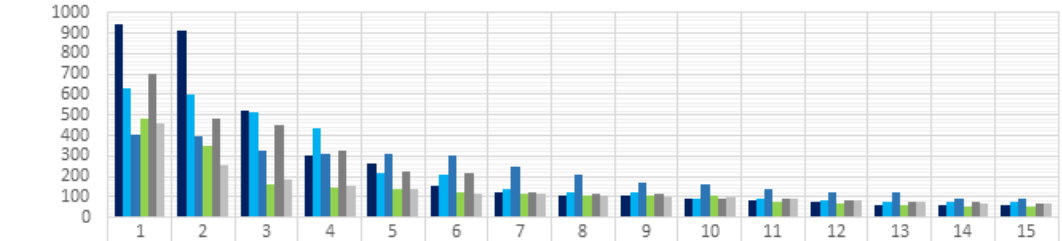
2011~2015

doi:10.6342/NTU202201326



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	518	407	324	299	294	228	169	144	117	108	106	100	83	76	76
Enhanced	437	416	327	238	195	148	141	137	120	104	100	94	81	80	79
Distance	431	328	319	188	167	158	156	147	120	116	111	101	101	94	77
Lexical	129	128	125	108	106	98	89	76	74	71	68	64	58	55	54
Bert	304	301	211	172	171	162	147	136	133	126	118	105	101	94	73
WN	244	198	177	159	132	128	126	103	102	97	95	86	83	79	68

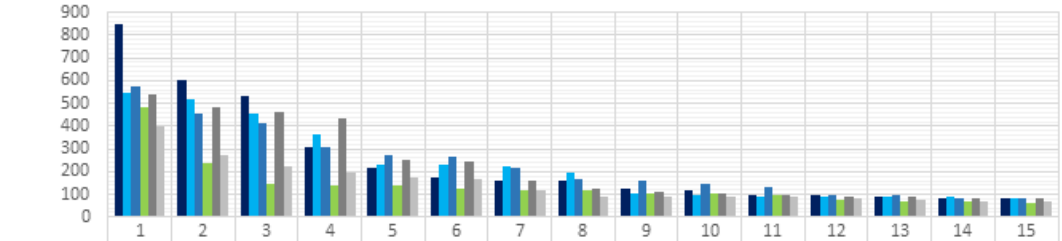
2012~2016



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	939	914	519	304	264	158	121	108	105	91	81	80	60	58	57
Enhanced	634	598	516	432	218	212	141	125	120	95	92	82	80	80	76
Distance	403	396	327	311	309	301	245	213	172	163	141	124	120	94	93
Lexical	483	352	166	145	140	123	117	109	109	107	77	72	59	56	54
Bert	698	483	453	324	221	214	124	119	112	95	92	86	77	73	72
WN	455	255	182	154	137	119	112	111	100	97	89	86	73	72	70

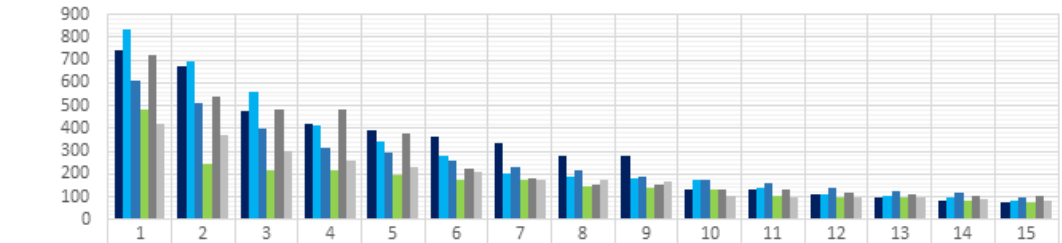
2013~2017

doi:10.6342/NTU202201326



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	845	606	533	305	219	176	159	159	123	119	100	96	91	83	82
Enhanced	547	520	457	363	232	231	226	194	104	95	92	89	88	87	83
Distance	578	453	413	311	275	268	214	168	161	145	131	95	94	84	81
Lexical	483	236	147	142	139	127	120	119	102	101	97	79	69	69	62
Bert	539	483	465	435	253	246	159	127	110	101	96	93	92	85	84
WN	396	273	223	192	172	170	118	89	88	87	87	81	74	66	66

2011~2015



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Classical	744	674	476	421	392	367	333	282	280	132	131	114	100	85	76
Enhanced	834	691	562	413	346	283	203	187	179	177	142	112	105	97	84
Distance	607	508	401	316	291	257	227	214	185	176	163	137	125	116	98
Lexical	483	244	216	216	194	173	172	144	141	134	107	98	96	83	75
Bert	721	542	483	482	380	226	180	152	151	134	131	121	110	103	102
WN	422	374	303	257	228	211	176	171	168	105	99	95	94	87	86

2012~2016

doi:10.6342/NTU202201326