Citation Genetic Genealogy: A New Perspective for Citation Analysis in Scientific Literature

Fengjun Sun[1]

[1] *fengjunsun@gmail.com*
Institute of Scientific and Technical Information of China (ISTIC), 15 Fuxing Road Beijing 100038 (P.R. China)

Abstract

Citation relationships are commonly described with citation index or citation graph, but in this article, the author introduced the notion of citation genetic genealogy and apply it in citation analysis. A citing document usually only uses pieces of its cited reference, so the author of this paper defined those pieces of a scientific document, which carry the information that have been used or may be used in the future by other documents as its document genes. Besides, with the definition of symbolic information of a scientific document, the conclusion that a citing document inherited the document genes from its references can be drawn. Based on these understandings, citation genetic genealogy was constructed to describe citation relationships. With citation genetic genealogy, it is easy to map the citation relationships, like bibliographic coupling and co-citation, with familiar family relationships and illustrate the inheritance relationships in scientific literatures. Also, citation genetic genealogy may provide an interface between the citation analysis of a document set and the content analysis for each individual document inside this document set.

Introduction

Citation index is the basic tool for citation analysis. It is an ordered list of cited documents, each accompanied by a list of citing documents (Egghe & Rousseau, 1990). The pioneer work of modern citation index was done by Garfield and his colleagues in ISI (Garfield, 1955, 1964; Garfield, Sher, & Torpie, 1964). Today citation index is not only used for the dissemination and retrieval of scientific information but also for research evaluation (Garfield, 2006, 2007). Meanwhile, citation relationships can also be described by the mathematical discipline of graph theory. It is also Garfield who first suggested using graph to illustrate citation index (Garfield, 1963). Later, Price recognized the linking properties of citations (Price, 1965), and Garner applied mathematical notion of graph to citation network (Garner, 1967). In the citation graph, each node represents a document and the directed links between each node indicate the citing relationship. With the analysis of the graph, researchers can reveal the communication relationships between each node or each document. And we may go farther to find the relationships between the researchers, publications, institutes, and even regions or countries.

Science is about extending the range of a theory and deepening and strengthening the theory's foundations (Brookes, 1980). And scientific documents, as the carriers of scientific ideas, are not only the channels of communication but also the channels to spread these scientific ideas to the following researchers. The fact that a document is cited indicates that some information have been inherited by its citations (Small, 1978). Quite recently, Garfield developed a software system called Histcite to generate chronological historiographs or genealogical microhistories of authors or topics, highlighting the most-cited works in the retrieved collection(Garfield, 2009; Garfield, Paris, & Stock, 2006). But there may be a better way than graph to illustrate the inheritance of scientific ideas, so in this article, we introduce citation genetic genealogy, as a new way to describe the inheritance relationships in scientific literatures.

Citation genetic genealogy

On the Shoulders of Giants

Isaac Newton once said, "If I have seen farther, it is by standing on the shoulders of giants." (Merton, 1965) From his words we may draw such a conclusion that a research work can be roughly divided into two parts: the researchers' creative jobs in this research process and the information inherited from other research works which have already been done.

Creative jobs are the core of a research process. Newton contributed his work to other scientists, but without his own creative jobs, the development of science may be postponed for many years. On the other hand, the information inherited from the previous works also plays a very important role in research. Without knowing what others have already done, researchers will have to start from the very beginning. Today, it is even unimaginable for a scientist if he or she ignores others' research achievements.

The goal of scientific research is the production of public scientific knowledge and the creation of consensus, concerning this knowledge, in the research community (Ziman & Crane, 1969). The scientific documents are very important carriers in the process of building this consensus (McCain & Turner, 1989) and they are also assumed to represent scientific activity (Ziman, 1987). So for a document, it carries both the information created by the document itself and the information provided by other research works.

Inherited information, document gene and document genomes

When we are talking about what information have been inherited, the first things come to our minds are those valuable pieces of the previous documents' content. These pieces may include data, concepts, theories, methods or techniques (Egghe & Rousseau, 1990).

But for scientific documents, we define another type of information that must be inherited by the citations, "symbolic information". Symbolic information of a document is shown by each item in the reference list. For example, in journal articles, the symbolic information usually contains the authors, the title, the journal and the published year. Even the author did not use any piece of the content of the cited document, there is no doubt that he or she inherited the symbolic information, since it had already been in the reference list.

An author usually used a little part of his or her citing document (Egghe & Rousseau, 1990), sometimes only a word or a sentence. This little part is what Henry Small called information molecule (Small, 1999), and it leads us to think about using the biology notion gene to describe the "information molecule".

In genetics, a gene is a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions (Pearson, 2006). Similarly, we define the notion of document gene like this: a document gene is a piece of a scientific document which carries the information that have been used or may be used in the future by other documents. With this definition, we draw such a conclusion that a citing document must inherit the document genes from its references. This is because the citing document inherited those genes that carry the symbolic information even though it did not use any content part of its reference. For a scientific document, it basically has two types of genes, the genes carrying the creative information provided by this document and the genes inherited from other documents, and we simply call them creative genes and inherited genes respectively. Table 1 shows the comparison between human genes and document genes. Those data about human genes is from Robinson's book (Robinson, 2005).

Table 1. The comparison between human genes and document genes

Table	*Human genes*	*Document genes*
Number	22,258	Indeterminable
Source	His or her ancestors, and directly from parents	The previous documents or other sources
Function	Storing the genetic information	Storing the document genetic information
Obtaining pattern	Obtaining the genes passively from his or her parents	The authors deliberately choose the inherited genes and produce creative genes for the document

Before we discuss further about the topics in this article, we would like to define another notion also based on genetic knowledge, document genome. In modern molecular biology and genetics, the genome is the set of both the genes and the non-coding sequences of the DNA, and each organism has one genome (Ridley, 2000). But for a document, as we mentioned before, contains two different types of genes. Creative genes produced by the document itself and inherited genes coming from previous documents. Unlike the genome in genetics, there is more than one genome in a document if the document has more than one reference. Instead of document genome, we define the notion of document genomes. Here is the definition: document genomes are the sets of the document genes within a scientific document. Among the document genomes, one genome contains the genes created by this document itself and other genomes contain the genes inherited from other sources including scientific documents.

Citation genetic genealogy

Traditionally, genetic genealogy is the method using genetics to research genealogy. It applies the new genetic technology in the process to identify the relationships between individuals (Smolenyak & Turner, 2004). And the results of genealogy are usually illustrated by family trees or pedigree chart.

So far, we have already defined document gene and document genomes, so in some sense, we have already treated a scientific document like a living organism and built the genetics for scientific documents. Based on these understandings, we are able to build the genetic genealogy for citations, and in this article we call it citation genetic genealogy.

To illustrate citation genetic genealogy, we first use family relationships to map citation relationships. In order to achieve this goal, we need to define the notion of parent document in this way: if document A is the reference of document B, then document A is the parent document of B. From this definition, we are able to find three types of documents related with document A.

- A's child document: A's citation
- A's sibling document: sharing at least one parent document with A
- A's spouse document: having at least one child document with A

Here the words "spouse documents" mean the two documents have child documents, this is different from the real world. We know two spouses may not have children biologically, and two people having children may not be described as spouses.

Now with the notions we have already defined, all the citation relationships can be mapped with the familiar family relationships. Here is the mapping table of citation relationships to citation family relationships.

Table 2. The mapping table of citation relationships to citation family relationships.

Table	*Citation relationships*	*Citation family relationships*
Item 1	Reference	Parent document
Item 2	Citation	Child document
Item 3	Reference's references	Grandparent document
Item 4	Citation's Citation	Grandchild document
Item 5	Bibliographically coupled documents	Sibling documents
Item 6	Co-cited documents	Spouse documents

Kessler defined bibliographic coupling between two documents (Kessler, 1963), and here in our citation genetic genealogy, bibliographic coupling means the two documents share parent documents, in other words, the two documents are sibling documents. And the bibliographic coupling strength can be measured as how many parent documents shared by these two documents. Similarly, co-citation(Marshakova, 1973; Small, 1973) in our system means the two documents are spouse documents. And the strength of co-citation between the two documents is determined by how many child documents they have.

Even though we find citation family relationships very interesting to map the relationships in citation system with quite familiar human family relationships, it is necessary to point out the differences between human family relationships and citation family relationships. Here are some of the differences.

Table 3. The differences between human family relationships and citation family relationships.

Table	*Real human family relationships*	*Citation family relationships*
Number of parents	Two parents biologically	Multiple parent documents, equal to the number of references
Number of children	Indeterminable but limited	Indeterminable but unlimited, equal to the number of citations
What can be chosen	Children and spouse	Parent documents and elder sibling documents
What can't be chosen	Parents and siblings	Child documents, spouse documents and younger sibling documents

Here "elder sibling documents" for a document means those sibling documents published earlier, and "younger sibling document" is the opposite.

In fact, the authors of a document choose the document's references deliberately, so in our citation family relationships, the parent documents and elder sibling documents can be chosen and others cannot. This is just the opposite of the real human family relationships. If we look at this problem from the traditional citation view, we will find that those can be chosen are related to the references and those cannot be chosen are related to the citations. One document can choose its references, but cannot choose its citations.

So far, we have already introduced the genetic background of citation genetic genealogy and mapped the traditional citation relationships with citation family relationships. Next we are going to use two examples to discuss the details of citation genetic genealogy with document genomes and document gene. Unlike traditional genealogy usually displayed in family charts, we develop genome sequence as the expression of citation genetic genealogy. The first example is artificial one focusing on the process of constructing the genome sequence of each

of the selected documents. The second one is a real example which gets deep into the document genes.

Methods

An artificial example: constructing the genome sequence of a document

In scientific literature, researchers have already used citation index or citation graph to demonstrate the relationships of citations. Citation index and citation graph can be mapped to each other, and if we have the details about one of the two, we will get the other.

Since each document may have multiple citations and multiple references, we cannot use tree structure, which is the general expression of genealogy, to described citation relationships. Instead we choose to use genome sequence which can also be mapped to citation index and citation graph.

As we mentioned above, document genomes are the sets of the document genes which a document contains. And among the document genomes, one genome contains the genes created by this document and other genomes contain the genes inherited from other scientific documents. Now we just use the following example to illustrate the process of constructing genome sequence. Figure 1 is the citation graph of the example.

The citation graph in Figure 1 is very alike the citation graph in Egghe and Rousseau's book (Egghe & Rousseau, 1990), with a little change, adding document number 3 and document number 17. Next, we write the algorithm to give each document its genome sequence.

```
Algorithm of genome sequence (given document set D)
   Sort D based on published date in ascending order
      Give each document a sequential number as its identity
      Put each document's identity number as the first number in its genome sequence
      For each document d in D
            If d have reference set R in D
               Sort R based on publish date in ascending order
               Put each reference's genome sequence into d's genome sequence
Return
```

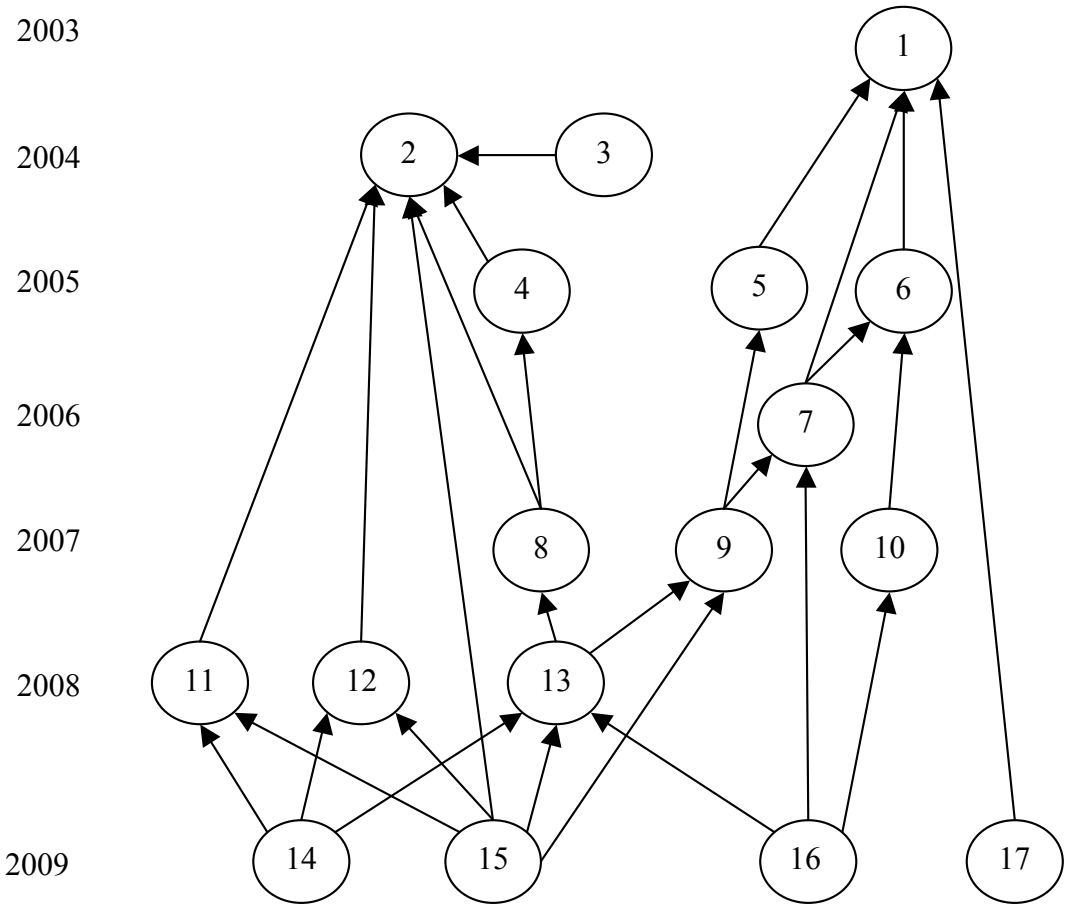

Figure 1. Citation graph of the artificial example.

Based on this algorithm, we give the genome sequence for each document. Table 4 shows the parent documents and genome sequence of each document in our example.

Table 4. The parent documents and genome sequence of each document.

Number	*Publish Year*	*Parent documents*	*Genome sequence*
1	2003	none	1
2	2004	none	2
3	2004	2	3;2
4	2005	2	4;2
5	2005	1	5;1
6	2005	1	6;1
7	2006	1,6	7;1;6;1
8	2007	2,4	8;2;4;2
9	2007	5,7	9;5;1;7;1;6;1
10	2007	6	10;6;1
11	2008	2	11;2
12	2008	2	12;2
13	2008	8,9	13;8;2;4;2;9;5;1;7;1;6;1
14	2009	11,12,13	14;11;2;12;2;13;8;2;4;2;9;5;1;7;1;6;1
15	2009	9,11,12,13	15;9;5;1;7;1;6;1;11;2;12;2;13;8;2;4;2;9;5;1;7;1;6;1
16	2009	7,10,13	16;7;1;6;1;10;6;1;13;8;2;4;2;9;5;1;7;1;6;1
17	2009	1	17;1

We give each document its genome sequence with the citation index or citation graph, and meanwhile, we can also build the citation index or graph with knowing the genome sequence of each document in reverse. Let's pick document number 16 as an example. We know the first number is the identity number, and find the second number to be 7. This means document number 7 is the reference which published earliest. And then we cut those sequence parts from 16 which are the same in 7, and find the next number to be 10, which means document number 10 is another reference. We repeat the process until we find all the references or parent documents for document 16.
Looking back to the definition of document genome, and now it is clear that the identity number represents the genome which contains all the creative genes, while the other numbers in the genome sequence represent genomes containing the genes inherited from other scientific documents. Even though there are same numbers in a genome sequence, but these genomes with the same number are not the same, and this will be discussed later.

A real example: discovering document genes through citation relationships

Now we have seen the process of constructing the genome sequence for each of the selected documents with an artificial example, it is time to use a real one to look deep into the document genes through citation relationships.
We choose the journal article references of Garfield's paper (Garfield, 2006) as the selected document set. Among the seven documents in our document set (Garfield, 1955, 1964, 1970, 1972, 1998; Garfield & Sher, 1967; Garfield, et al., 1964), two articles are written by Garfield and his colleagues and the others are completed by Garfield himself. One of the reasons that we choose these documents is that these journal articles are the references of a commentary, which mainly talks about the history of citation indexing, and those references we choose are all written by Garfield. We believe that these documents are closely related and also some of the most important documents in the history of citation indexing. Besides, thanks to Garfield's online library, we are able to get the copy of these documents. This is very important since we are going to look into the content. Then we construct the genome sequence for each document in the set. The result is shown in Table 5.

Table 5. The parent documents and genome sequence of each document.

Number	*Publish Year*	*Parent documents*	*Genome sequence*
1	1955	none	1
2	1964	1	2;1
3	1964	1,2	3;1;2;1
4	1967	1,2	4;1;2;1
5	1970	3	5;3;1;2;1
6	1972	1,2,4	6;1;2;1;4;1;2;1
7	1998	1,2,3,4,6	7;1;2;1;3;1;2;1;4;1;2;1;6;1;2;1;4;1;2;1

We note that document number 3 is written by Garfield, Sher and Torpie, and number 2 is written by Garfield himself.
Next we form a matrix, in which each entry represents the direct inherited content genes. Since one document may inherit different document genes directly from another one, each entry may have multiple items. For example, the items 1a, 1b, 1c and 1d in entry a_{21} indicate that they are the genes inherited directly from document number 1 to document number 2. With the definition of document genes, the number of items in an entry of the matrix simply equals to how many times the same footnote number appears in the body. By looking into the

document, we find that each footnote number's positions in the text and count how many times the same number appears. The matrix for our document set is shown in Table 6.

Table 6. The direct inherited content document genes of each document.

Number	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1							
2	1a,1b,1c,1d						
3	1e	2a					
4	1f	2b					
5			3a				
6	1g	2c		4a			
7	1h	2d	3b	4b		6a	

The summary of the citation context for each document gene are listed in Table 7.

Table 7. The summary of the citation context for each gene.

Number	*The summary*
1a	
1b	
1c	
1d	
1e	a suggestion for citation indexing's use in historical research
1f	a retrieval system called "citation indexing"
1g	the Science Citation Index
1h	a paper proposing the creation of citation indexes
2a	citation indexing used for the purposes of disseminating and retrieving information
2b	the Science Citation Index
2c	the Science Citation Index
2d	the utility of the Science Citation Index as a retrieval and dissemination device
3a	the history of the genetic code using citation analysis
3b	careful citation mapping leads to the uncovering of small but important historical links overlooked by even the most diligent scholars
4a	selective dissemination of information (SDI)
4b	ASCA (Automatic Subject Citation Alert)
6a	journal impact factors

We note that even though document 1 as a reference appeared in four different positions in document 2, but in our opinion, it was just used as an example and can be replaced with another one. In other words, document 2 only inherited those genes which carry the symbolic information, genes 1a-1d are not content genes. And for document gene 1g, in fact the notion "the Science Citation Index" is not in document 1. This is because document 1 and document 2 are cited at the same position in document 6, where we obtain the gene 1g and 2c. So in fact, it is gene 2c which carries the concept of "the Science Citation Index".

Results

Through citation relationships in our document set, we obtain the key points of each gene, and with combining those we think are the same, we finally get the direct inherited content document genes of each document. See Table 8.

Table 8. The direct inherited content document genes of each document.

Number	*Document genes*
1	citation indexing
	citation indexing as a tool for information retrieval
	citation indexing as a tool for historical research
2	the Science Citation Index (SCI)
	citation indexing as a tool for disseminating and retrieving information
	SCI as a tool for disseminating and retrieving information
3	the history of the genetic code using citation analysis
	citation mapping as a tool for uncovering of small but important historical links
4	selective dissemination of information (SDI)
	Automatic Subject Citation Alert (ASCA)
5	
6	journal impact factors
7	

From the eleven document genes we discovered, six are concepts (citation indexing, SCI, citation mapping, SDI, ASCA, journal impact factors). Among the rest five, four are the usages of the six conceptual document genes and only one is a content piece (the history of the genetic code using citation analysis). Just in our example, we find that those document genes about concepts are more likely to be inherited.

We have to mention that these document genes are derived from the citation relationships in the document set. There may be more genes in each document, but not inherited by the others inside the set. In other words, with citations, we only can discover those genes are inherited. If we want those genes which are not inherited, we have to find other ways. For example, if we want to know what document genes are in document 5, we may find document 5's citations or just read the paper.

Discussions

Loss of Document genes

In both of the examples above, we noticed the fact that same genome number appeared more than once in a genome sequence, but the genomes though with the same number, are completely different. Table 5 chooses 2 documents from the real example.

Table 5. Two documents chosen from the artificial example.

Number	*Publish Year*	*Parent documents*	*Genome sequence*
6	1972	1,2,4	6;1;2;1;4;1;2;1
7	1998	1,2,3,4,6	7;1;2;1;3;1;2;1;4;1;2;1; 6;1;2;1;4;1;2;1

For document number 6 and number 7, both of them inherited the genes from number 4, but they may not use the same ones. In fact, document 6 inherited the gene about the concept "ASCA" and document 7 inherited the gene about the concept "SDI".

This phenomenon is described as loss of genes. The main reason is the authors of a document only choose those parts of a reference they need, so some genes of the original genome are lost during the spread process. It seems that with more indirect inheritance, there is more possibility that the genes from original genome may disappear. But this may be wrong for those genes in the creative genome of a classical scientific document.

When the information in a classical document becomes "common knowledge", or the "obliteration" process (Merton & Storer, 1979), we do not have to cite the original document, which means there are no documents directly inheriting the genes in the classical one. Even after so many generations of indirect inheritance, we may find that these genes are still in a descendant document. This may provide us a new indicator for classical document by the number of descendent documents in which its creative genes exist, no matter whether the genes are inherited directly or indirectly. In the real example, the document gene about "citation indexing" is not inherited directly by document 5, 6 and 7. But from our background knowledge, we know that the gene "citation indexing" is in their genomes, in other words, this gene inherited by these documents indirectly.

Connecting citation analysis and content analysis

Both bibliographic coupling and co-citation are criticised, since they do not guarantee to refer to the same piece of information(Martyn, 1964). In order to qualify citations, researchers take content analysis into account (Elkiss, Shen, Fader, & Erkan, 2008; Nakov, Schwartz, & Hearst, 2004; Small, 1978), but these researches limited the content analysis within one document. Citation genetic genealogy just provides us a tool to connect the citations and the contents, and may facilitate needed solutions for both citation analysis and content analysis.

For more accurate content analysis, it is necessary to have a large amount of data (Och & Ney, 2004). The citation genetic genealogy just provides a large dataset. Except those words in the document's creative genome, almost all the words can be found in its ancestor documents' document genes. Besides, citation genetic genealogy might be a good tool to solve the problem of the ambiguity of language (Leydesdorff & Hellsten, 2006). When we encounter a word in a document hard to understand, perhaps it is the time to go to its "close relatives" in the genealogy to find the answer.

On the other hand, we might use content analysis to reconstruct the citation relationships. This means, we ignore those citations only inherited the symbolic information, and rebuild the citation relationships based on the content document genes. As we mentioned, except citations, we have to discover the content document genes from the title, abstract, keywords or other parts of the document with traditional content analysis method. With the combination of these methods, we are able to build the document gene database for each scientific document. And then, it will be easy to reconstruct the citation relationships, not only to make the citations analysis more accurate and precise, but also to tell the users what document genes have been inherited between each document.

A disease in citation genetic genealogy: bidirectional citations

Bidirectional citations mean two documents cite each other. Here is an example, in the two earliest documents written by the same authors(Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1999) about the algorithm page-rank, each document appeared on the reference list of the other's. Since there may be a long process before the authors' article is published, some authors choose to cite their own articles even not finished (manuscript in progress), then bidirectional citations appear. Perhaps, this is one of the last things we would like to see in citation analysis, and the problem may be even worse in citation genetic genealogy. In traditional citation network, this is just contrary to the normal flow, but in citation genetic genealogy, this means X document can be the parent document and the child document of Y at the same time. Some fundamental works in our system, like genome sequences, may get into chaos. In practice, perhaps we just ignore this kind of references, but in order to make citation genetic genealogy better, we need to find right medicine to cure this disease in citation genetic genealogy.

Conclusions

In this article, we used some basic notions in genetics and introduced citation genetic genealogy to describe citation relationships. Citation relationships in citation genetic genealogy are described as the channels for scientific document genetics. Inherited document genes, or those pieces in a reference used by its citations, spread to the reference's descendent documents through these channels. In the future, we may build the document gene database though content analysis and citation analysis, and facilitate better solutions for both content analysis and citation analysis.

Acknowledgments

I would like to thank Professor Ronald Rousseau of KHBO (Association K.U.Leuven), Senior Research Fellow Yishan Wu, Associate Research Fellow Junpeng Yuan and Associate Research Fellow Lijun Zhu of Institute of Scientific and Technical Information of China (ISTIC), for discussions and valuable suggestions.

References

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*, 107-117.

Brookes, B. (1980). The foundations of information science. *Journal of Information Science, 2*(3-4), 125.

Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Amsterdam: Elsevier

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

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

Garfield, E. (1963). Citation indexes in sociological and historical research. *American Documentation, 14*(4), 289-291.

Garfield, E. (1964). Science citation index: A new dimension in indexing. *Science, 144*(3619), 649-654.

Garfield, E. (1970). Citation indexing for studying science. *Nature, 227*(5259), 669.

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

Garfield, E. (1998). From Citation indexes to informetrics: Is the tail now wagging the dog? *Libri, 48*(2), 67-80.

Garfield, E. (2006). Commentary: Fifty years of citation indexing. *International Journal of Epidemiology, 35*(5), 1127-1128.

Garfield, E. (2007). The evolution of the science citation index. *International microbiology, 10*(1), 65.

Garfield, E. (2009). From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics, 3*(3), 173-179.

Garfield, E., Paris, S., & Stock, W. G. (2006). HistCiteTM: A Software Tool for Informetric Analysis of Citation Linkage. *Information Wissenschaft und Praxis, 57*(8), 391.

Garfield, E., & Sher, I. (1967). ISI's experiences with ASCA-a selective dissemination system. *Journal of Chemical Documentation, 7*(3), 147-153.

Garfield, E., Sher, I., & Torpie, R. (1964). The use of citation data in writing the history of science. Philadelphia: Institute for Scientific Information.

Garner, R. (1967). Computer-oriented graph theoretic analysis of citation index structures. Philadelphia: Drexel University Press.

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

Leydesdorff, L., & Hellsten, I. (2006). Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells'. *Scientometrics, 67*(2), 231-258.

Marshakova, I. V. (1973). A system of document connections based on references. *Scientific and Technical Information Serial of VINITI, 6*, 3-8.
Martyn, J. (1964). Bibliographic coupling. *Journal of Documentation, 20*(4), 236.
McCain, K., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics, 17*(1), 127-163.
Merton, R. (1965). *On the Shoulders of Giants: A Shandean Postscript*. New York: Free Press.
Merton, R., & Storer, N. (1979). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University Of Chicago Press.
Nakov, P., Schwartz, A., & Hearst, M. (2004). *Citances: Citation Sentences for Semantic Analysis of Bioscience Text.* Paper presented at the Proceedings of the SIGIR'04
Och, F., & Ney, H. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics, 30*(4), 417-449.
Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*.
Pearson, H. (2006). Genetics: what is a gene? *Nature, 441*(7092), 398.
Price, D. J. D. (1965). Networks of scientific papers. *Science, 149*(3683), 510-515.
Ridley, M. (2000). *Genome: The Autobiography of a Species in 23 Chapters*. New York: HarperCollins.
Robinson, T. (2005). *Genetics for Dummies*. Hoboken: Wiley.
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.
Small, H. (1978). Cited documents as concept symbols. *Social Studies of Science, 8*(3), 327.
Small, H. (1999). ASIS award of merit: On the shoulders of giants. *Bulletin-American Society For Information Science, 25*, 23-25.
Smolenyak, M., & Turner, A. (2004). *Trace Your Roots with DNA: Using Genetic Tests to Explore Your Family Tree*. Emmaus: Rodale Books.
Ziman, J. (1987). *An Introduction to Science Studies: The Philosophical and Social Aspects of Science and Technology*. Cambridge: Cambridge University Press.
Ziman, J., & Crane, D. (1969). Public knowledge: An essay concerning the social dimension of science. *Physics Today, 22*, 87.