Co-citation, bibliographic coupling and a characterization of lattice citation networks

Peer-reviewed author version

EGGHE, Leo & ROUSSEAU, Ronald (2002) Co-citation, bibliographic coupling and a characterization of lattice citation networks. In: Scientometrics, 55(3). p. 349-361.

DOI: 10.1023/A:1020458612014
Handle: http://hdl.handle.net/1942/773

Co-citation, bibliographic coupling and a characterization of lattice citation networks

LEO EGGHE
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
and UIA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
E-mail: leo.egghe@luc.ac.be

and

RONALD ROUSSEAU
KHBO, Industrial Sciences and Technology,
Zeedijk 101, B-8400 Oostende, Belgium
& UIA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
& LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
E-mail:ronald.rousseau@kh.khbo.be

Abstract

In this article we study directed, acyclic graphs. We introduce the head and tail order relations and study some of their properties. Recalling the notions of generalized bibliographic coupling and generalized co-citation, and introducing a new property, called the I - property, we come to a characterization of lattices. As document citation networks are concrete realizations of directed acyclic graphs all our results are directly applicable to citation analysis.

Keywords: bibliographic coupling, co-citation, lattice characterization, indirect citation relations, I - property, relative bibliographic coupling, relative co-citation, Jaccard similarity measure.

1. Introduction

Bibliographic coupling is a term made popular by M.M. Kessler (1963), although the idea itself is due to Fano (1956). A coupling unit between two documents is an item of reference used by these two documents. If such an item exists the two documents are said to be bibliographically coupled. Their bibliographic coupling strength is then the number of references they have in common. Similarly, two documents are said to be co-cited when they both appear in the reference list of a third document. The notion of co-citation was proposed independently by Irina Marshakova (1973) and Henry Small (1973) from ISI. Also for this notion the idea originated somewhat earlier, namely in Rosengren's co-mention approach (1968). The co-citation frequency is defined as the frequency with which two documents are cited together. Thus, while bibliographic coupling focuses on groups of papers which cite a source document, co-citation focuses on references coming frequently in pairs.

Many elementary mathematical properties of the bibliographic coupling and co-citation relation can be found in "Introduction to Informetrics" (Egghe & Rousseau, 1990). It is also clear that besides co-citation one may also study tri-citation (three articles being cited in the same document) or, more generally multi-citations (Small, 1974; Sen & Gan, 1983). A similar remark holds for the notion of bibliographic coupling. Besides document co-citations (or bibliographic couplings) one may also study author or journal co-citations (or bibliographic couplings), see e.g. (Ding et al., 2000). Yet, from a graph-theoretic point of view these forms of citation studies are totally different. Article citations are nearly always one-directional (article A cites article B, but then article B, being older, will not cite article A), while journal citations are very often bi-directional: journal A cites journal B, while, during the same period, journal B also cites journal A. A similar remark applies to author citations.

Bibliographic coupling and co-citation relations have in recent times been studied e.g. by Glänzel and Czerwon (1995), White (2000), Cawkell (2000) and Fang & Rousseau (2001). The ideas underlying these notions have been applied to search engines on the Web, in particular in Google (Dean & Henzinger, 1999; Henzinger, 2001). We also recall that the 'related records' feature in ISI's databases uses bibliographic coupling (Garfield, 1988; Atkins, 1999).

In this paper we introduce two partial order relations (called the head and the tail relation) and study some of their properties. This leads to the notions of generalized bibliographic coupling and generalized co-citation, also studied by Sen and Gan (1983). Finally, introducing another property, called the l - property (l for 'lattice') leads, together with the generalized bibliographic coupling and co-citation property to a characterization of a lattice network.

This paper belongs to a series of papers in which we try to develop algebraic tools to study networks, e.g. citation networks, hyperlink graphs (such as the Internet) and collaboration networks (Rousseau, 1987; Egghe & Rousseau, 2001). As such this work is related to earlier work by Garner (1967), Pritchard (1984), Botafogo, Rivlin and Shneiderman (1992), De Bra (2000) and Leazer & Furner (1999).

2. Graph theory and partially ordered sets

A directed graph G, in short: digraph, consists of a set of nodes, denoted as N(G), and a set of links, denoted as L(G). In this paper the words 'network' and 'graph' are synonymous. A path from node a to node b is a sequence of distinct links (a, n_1), (n_1,n_2), ... , (n_k,b). The length of this path is the number of links (here k+1). Note that, in general, a path from a to b does not necessarily imply a path from b to a. A cycle or loop is a path of length > 1, beginning and ending in the same node. A graph that does not contain any cycle is called an acyclic graph. In this paper we will always assume that edges are unweighted, or, equivalently, have a weight equal to one. Any directed graph has an underlying undirected graph in which links may be traversed in any direction.

A unidirectional graph is a graph in which a link between nodes a and b, implies that there is not a (direct) link from b to a. In a unidirectional graph cycles may exist, but the smallest possible length is 3.

A strongly connected component of a digraph is a set of nodes such that any two of them are joined by a path. Different strongly connected components in a network consist of disjoint sets of nodes. If a digraph consists of one strongly connected component it is said to be strongly connected. If the underlying undirected graph consists of one component, the graph is said to be weakly connected.

When applying these notions to citation networks (a network where nodes are articles and a link from article a to article b means that article a refers to article b) we always assume that these are unidirectional (because of the time difference between linked articles), although this is in reality not always the case. All citation networks considered in this paper are moreover assumed to be acyclic. Hence, a citation network is a concrete realization of an acyclic, unidirectional graph.

A binary relation on a set S is called a partial order on S (denoted as $\leq$) (Birkhoff & Bartee, 1970) if it satisfies the following three conditions:
P1. For all $x \in S : x \leq x$ (reflexivity)
P2. For all $x, y \in S : x \leq y$ and $y \leq x$ imply $x = y$ (antisymmetry)
P3. For all $x, y, z \in S : x \leq y$ and $y \leq z$ imply $x \leq z$ (transitivity).

A set S, equipped with a partial order $\leq$, is called a partially ordered set, in short a poset . Examples of posets are:
- the usual 'smaller than or equal' relation $\leq$, in the set of natural numbers;
- for any set U, the inclusion relation, denoted as $\subset$, is a partial order in the power set of U (the set of all subsets of U).

A poset satisfying the additional relation that, given x and y, either $x \leq y$, or $x \geq y$, is called a complete order. Elements can then be arranged in a chain from largest to smallest, or vice versa. The natural numbers form such a chain with the usual $\leq$ relation. An inclusion poset of a set with a least two elements is never a complete order. We further note that every subset of a poset is again a poset (perhaps even a chain).

Greatest lower and least upper bounds (Birkhoff & Bartee, 1970). Given a subset X of a poset S, we say that an element a in S is an upper bound of X if, for all $x \in X$, $x \leq a$. Similarly, we say that $b \in S$ is a lower bound of X if, for all $x \in X$, $b \leq x$. We define c to be the least upper bound of X if 1°) c is an upper bound of X, and 2°) for every other upper bound a of X, we have: $c \leq a$. This is denoted as: c = lub X. Similarly, define d to be the greatest lower bound of X if 1°) d is a lower bound of X, and 2°) for every other lower bound b of X, we have: $b \leq d$. This is denoted as: d = glb X. Subsets of a poset do not always have a least upper bound or a greatest lower bound, but if a glb or a lub exists, it is unique.

The notions of glb and lub lead us to a special, and very important class of posets, namely lattices. A *lattice* is defined as a poset such that every two elements x and y have a glb called the meet of x and y, denoted as $x \wedge y$, and a lub called the join of x and y, and denoted as $x \vee y$.

3. The tail relation

Definition: the tail and the head of a node A.

Let G be a directed, acyclic graph and let N(G) be the set of nodes in G. Let further $A \in N(G)$, then we denote by T(A) the set of all nodes C in N(G) such that there exists a finite path starting in A and ending in C. T(A) is called the tail of node A. By definition we include A in T(A), hence: $A \in T(A)$. Similarly, we denote by H(A) the set of nodes C in N(G) such that there exists a finite path starting in C and ending in A. H(A) is called the head of node A. Here too we assume that $A \in H(A)$. As G is acyclic, $H(A) \cap T(A) = \{A\}$.

Definition: the tail relation in G

Let A and B be nodes in G. Then we say that B -< A if and only if $B \in T(A)$. Similarly, B -< A if and only if $A \in H(B)$. The relation -< is called the tail relation (actually tail *order*, see further) in G.

In words: B is "smaller" than A in the tail relation if B belongs to the tail of A, or A belongs to the head of B, i.e. there exists a finite path beginning in A and ending in B. In the next proposition we will show that this relation is a partial order. Hence it will be called 'the tail order', and hence the term 'smaller' has been used correctly.

Proposition 1

The tail relation is a partial order, i.e. -< is reflexive, asymmetric and transitive.

Proof.

a) reflexivity

By definition $A \in T(A)$, hence A -< A

b) antisymmetry

If A -< B and B -< A then there is a finite path beginning in B and ending in A, and moreover a finite path beginning in A and ending in B. This means that there exists a loop from B to B, unless A = B. As G is acyclic, this means that A = B.

c) transitivity

If A -< B and B -< C , then there exists a finite path from B to A and a finite path from C to B, hence there exists a finite path from C to A, or A -< C. This proves the transitivity of the tail relation.

Note

If B -< A then T(B) $\subset$ T(A), i.e. T is an order-preserving operator in N(G), -< .

4. Applications in citation analysis: indirect citation relations

We let G be a citation network, i.e. document A is connected to document B (in that order) if A cites B. Then B -< A means that there exists a finite path of citation relations beginning in A and ending in B. We will express this by saying that A cites B indirectly. If the length of the path connecting A with B is n then we say that A cites B indirectly on level n. It is possible that a document A cites a document B on different levels. We assume that a document D always cites itself on level 0. A direct citation is a citation on level 1.

We can, similarly, consider the 'is cited by' and the 'is indirectly cited by' relations. They are the dual relations of the 'cites' and 'cites indirectly' relations.

As the 'cites' and 'is cited by' relations are not transitive, it are the 'cites indirectly' and 'is indirectly cited by' relations that have the better mathematical structure. This is natural from a categorical point of view. In category theory (Mac Lane, 1971; Rousseau, 1992) one requires an arrow from A to C, once there is an arrow from A to B and from B to C. Only the indirect citation relation meets this requirement.

We next generalize *Proposition 1* (Egghe & Rousseau (1990), p. 230, based on a result by Kochen (1974)) and its corollary.

Proposition

For any document C we have:

$$C \in \bigcap_{D \in T(C)} H(D)$$

Proof. This formula merely states that if we form the collection of all documents that are indirectly cited by C, this is C's tail, and if we consider any document, say D, in this tail, then C belongs to the head of D. This is trivial.

Corollary

For every document C we have:

$$H(C) = \bigcap_{D \in T(C)} H(D)$$

Proof. We denote the set on the right-hand side by R. As C $\in$ T(C), R is clearly a subset of H(C). Conversely, if D $\in$ T(C), then H(C) $\subset$ H(D), consequently, H(C) is a subset of R. This proves the equality.

Similarly, we generalize Theorem 2 (Egghe & Rousseau (1990), based on Kochen (1974)). This is the basic reference cycle (Atkins, 1999).

Theorem. If G is a non-empty, weakly connected citation network with N vertices, i.e. containing N documents then, for any C $\in$ G:

$$G = \bigcup_{j=0}^{N-1} K_j \quad \text{with}$$
$$K_j = \bigcup_{D \in K_{j-1}} \left(T(D) \cup H(D)\right)$$
$$\text{and } K_0 = \{C\}$$

Proof. Consider any document C in G. Then K_1 is the set of all documents that either cite C indirectly, or are indirectly cited by C. By the requirement of weak connectedness, K_1 contains more elements than C, unless G = {C}, in which case the theorem is proved. So, we can proceed and form K_2. In order to prove that G is indeed the union of all K_j, j = 0, ...,N -1, we assume that there exists a document D_0 that does not belong to any of the K_j. We know, however, that there exists a path, ignoring directions, necessarily finite, and hence with length at most N -1, connecting D_0 with C. Hence there must be a number j $\leq$ N –1, such that D_0 belongs to K_j. This proves the basic cycling theorem.

5. Indirect co-citation and indirect bibliographic coupling

We will use the terminology of a citation network, but note that the ideas proposed here can be applied in any acyclic digraph.

Definition: Two documents A and B are said to be indirectly co-cited if there exists a document C such that C cites A indirectly and C cites B indirectly. Using the tail order, this means:

A and B are indirectly co-cited

$\Leftrightarrow$

$\exists$ C: A -< C and B -< C

$\Leftrightarrow$

$\exists$ C: A $\in$ T(C) and B $\in$ T(C)

$\Leftrightarrow$

$$\exists\ C: T(A) \cup T(B) \subset T(C)$$

Definition: Two documents A and B are said to be indirectly bibliographically coupled if there exists a document C such that A cites C indirectly and B cites C indirectly. This gives:

$$\begin{gathered}\text{A and B are indirectly bibliographically coupled}\\ \Leftrightarrow\\ \exists\ C: C \text{ -< } A \text{ and } C \text{ -< } B\\ \Leftrightarrow\\ \exists\ C: C \in T(A) \cap T(B)\\ \Leftrightarrow\\ \exists\ C: T(C) \subset T(A) \cap T(B)\end{gathered}$$

The notions of indirect co-citation and indirect bibliographical coupling go back at least to Sen & Gan (1983).

Notes:
a) If A and B are co-cited they are indirectly co-cited.
b) If A and B are bibliographically coupled they are indirectly bibliographically coupled.
c) Any document A is always indirectly co-cited and bibliographically coupled with itself.
d) If A cites B they are indirectly co-cited (by A) and indirectly bibliographically coupled (by B).
e) If a network has a largest element in the tail order, i.e. $\exists L \in N(G)$ such that $\forall\ A \in N(G)$: A -< L, then any two nodes are indirectly co-cited.
f) If a network has a smallest element in the tail order, i.e. $\exists S \in N(G)$ such that $\forall A \in N(G)$: S -< A, then any two nodes are indirectly bibliographically coupled.
g) The algorithm studied by Rousseau (1987) and known as the Gozinto theorem describes a method to take indirect citations into account when determining the citation influence of one article on another.

6. A characterization of a lattice

Lattices are not only important mathematical structures, it is shown (Fang & Rousseau, 2001) that they do occur as substructures of citation networks. Hence the characterization given here has not only value for the theoretical development of network structures, but may also have some practical applications.

Proposition

If G, -< is a lattice, then $T(A) \cap T(B) = T(A \wedge B)$

Proof. As $A \wedge B$ -< A, we have $T(A \wedge B) \subset T(A)$, as T preserves the tail order. Similarly $T(A \wedge B) \subset T(B)$, which shows that $T(A \wedge B) \subset T(A) \cap T(B)$. If now $C \in T(A) \cap T(B)$ then C -< A and C -< B, consequently, C -< $A \wedge B$. Hence $C \in T(A \wedge B)$, showing that $T(A) \cap T(B) \subset T(A \wedge B)$.

Proposition

If G, -< is a lattice then $T(A) \cup T(B) \subset T(A \vee B)$.
Proof. A, B -< $A \vee B$, hence $T(A) \subset T(A \vee B)$. Similarly $T(B) \subset T(A \vee B)$, hence $T(A) \cup T(B) \subset T(A \vee B)$.

It is easy to see that the other inclusion is not valid, unless A -< B or B -< A. We formulate this as a characterisation of comparability in the lattice G, -<.

Proposition

Assume that G, -< is a lattice. Then two elements A and B in N(G) are comparable, i.e. A -< B or B -< A, if and only if $T(A) \cup T(B) = T(A \vee B)$.

Proof. $\Leftarrow$ The element $A \vee B$ exists because G is a lattice. Now, it is given that $A \vee B$ belongs to T(A) or T(B). Hence $A \vee B$ -< A or $A \vee B$ -< B. But we always have that A,B -< $A \vee B$. Hence $A \vee B = A$ or $A \vee B = B$, which means that B -< A, or A -< B, respectively. The other implication is trivial.

Proposition

If G is a citation network, and G, -< is a lattice then any two documents A and B are indirectly co-cited and indirectly bibliographically coupled.

Proof.
In order to show that A and B are co-cited it suffice to take $C = A \vee B$. For proving bibliographical coupling we take $C = A \wedge B$.

Definition: the I - property

A graph is said to have the I - property if A -< C, B -< C, A -< C', B -< C' implies that there exists C" such that A -< C", B -< C", C" -< C and C" -< C'. This property is illustrated by Fig.1. Intuitively, the I - property assures the existence of an element C" that lies 'in-between' the (A,B)-level and the (C,C')-level. Note that this is not really correct as C" may coincide with one or more of the given elements A,B,C or C'.

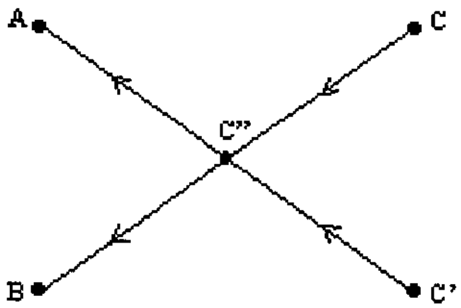

Fig. 1 A graph having the l - property

Proposition
If G, -< is a lattice then it has the l - property.

Proof
C and C' are both larger than A and B, hence taking C" = A $\vee$ B shows that N(G), -< has the l - property.

Theorem
If G is a finite citation network, then the following are equivalent

a) G, -< is a lattice
b) G, -< has the l - property and every two elements are co-cited and bibliographically coupled

Proof. We know already that if G, -< is a lattice then it has the l - property, and any two documents are co-cited and bibliographically coupled.

Assume now that G, -< has property b.

Let A and B be two nodes (documents) in G. By assumption b, we know that they are indirectly co-cited. Hence, there exists a node C, such that A and B belong to T(C). This means that C is an upper bound of A and B. Let ***C*** be the set of all upper bounds of A and B. By the previous observation, this set is non-empty. Now, by the l - property we know that for any C_1 and C_2 in ***C*** there exists a C_3 (not necessarily different) in ***C*** such that A, B -< C_3 and C_3 -< C_1,C_2. Applying this procedure a finite number of times, we find an element C' in ***C*** such that A,B -< C' -< C*, for every C* in ***C***. This proves that A and B have a lub: C' = A $\vee$ B.

Similarly, we conclude, using the fact that any A and B are indirectly bibliographically coupled, that A and B have a glb.

Remarks

1. Knowing that a poset has the l - property and that any two elements are indirectly co-cited does not suffice to be a lattice.

2. Having the *l* - property and any two elements are indirectly bibliographically coupled does not suffice to be a lattice;

3. Having the property that every two elements are co-cited and bibliographically coupled does not suffice to be a lattice.

Examples are given in figs. 2, 3 and 4.

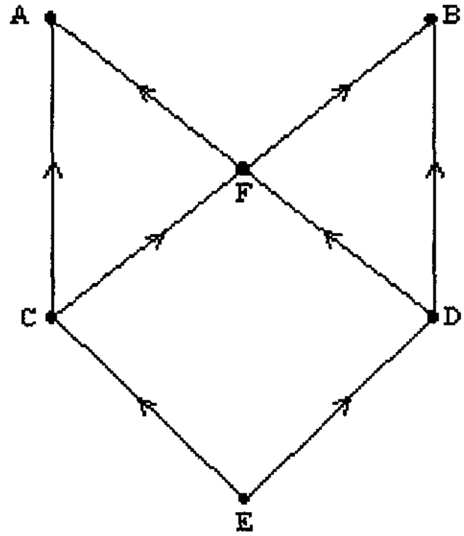

Fig. 2 An example of a poset having the *l* - property, where any two elements are indirectly co-cited but which is not a lattice, as A and B have no lub.

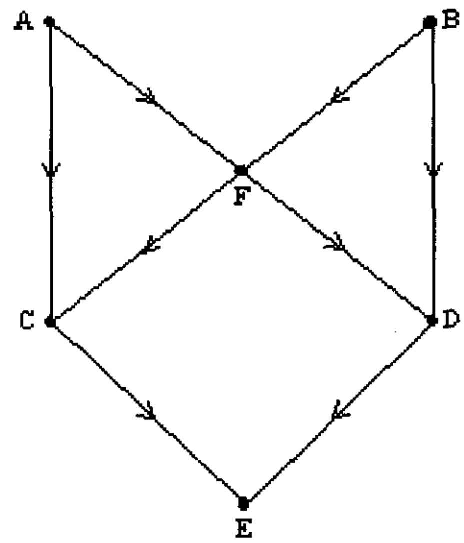

Fig.3 An example of a poset having the *l* - property, where any two elements are indirectly bibliographically coupled but which is not a lattice as A and B have no glb

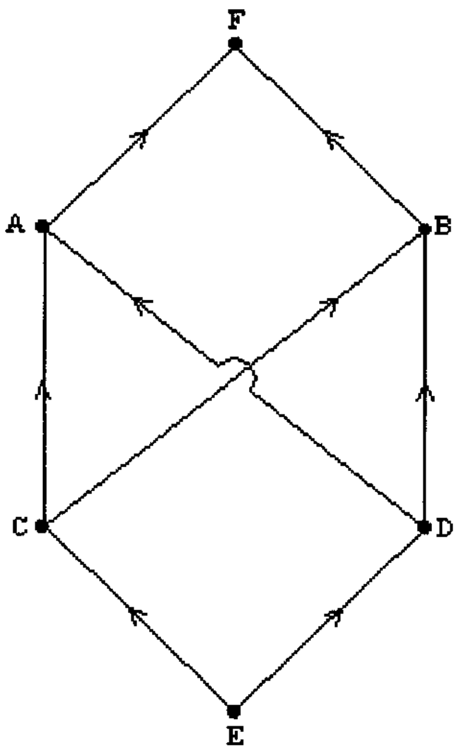

Fig.4 A poset where any two elements are indirectly co-cited and bibliographically coupled but which is not a lattice, because, e.g. A and B have no lub.

7. Relative co-citation and relative bibliographic coupling

In the same article as where he introduced the notion of co-citation, Small also introduced the notion of relative co-citation (Small, 1973). The relative co-citation of documents A and B is defined as:

$$\frac{\#\big(C(A) \cap C(B)\big)}{\#\big(C(A) \cup C(B)\big)}$$

where C(A) denotes the set of all articles that cite A, C(B) denotes the set of all articles that cite B, and #(X) denotes the number of elements in the set X. The notion of relative bibliographic coupling can similarly be defined (Sen & Gan, 1983; Egghe & Rousseau, 1990). In particular, relative co-citation trends (evolution of relative co-citation values over time) of some citation classics have been studied in Egghe & Rousseau (1990). We note that the relative co-citation of A and B is nothing but the Jaccard index of the sets C(A) and C(B) (Egghe & Rousseau, 1990).

The notions of relative co-citation and relative bibliographic coupling can easily be generalized to the 'indirect' case using the head and tail relations. In this context we consider

$$\frac{\#(T(A) \cap T(B))}{\#(T(A) \cup T(B))} \text{ and } \frac{\#(H(A) \cap H(B))}{\#(H(A) \cup H(B))}$$

Consequently, these formulae can be used as general similarity measures, generalizing the well-known Jaccard index.

8. Conclusion

In this article we have studied directed, acyclic graphs as concrete realizations of document citation networks. We introduced the head and tail order relations and studied some of their properties. Recalling the notions of generalized bibliographic coupling and generalized co-citation, and introducing a new property, called the I - property, we came to a characterization of lattices. As lattices do occur in real citation networks this result may be considered as one more building block for a mathematical, graph-based theory of networks in the information sciences. Finally, we introduced a generalization of the Jaccard similarity measure.

References

H. ATKINS (1999). The ISI Web of Science - links and electronic journals. D-Lib Magazine, 5(9). Available at: http://www.dlib.org/dlib/september99/atkins/09atkins.html

G. BIRKHOFF, T.C. BARTEE (1970), *Modern applied algebra*. McGraw-Hill, New York.

R.A. BOTAFOGO, E. RIVLIN and B. SHNEIDERMAN (1992). Structural analysis of hypertexts: identifying hierarchies and useful metrics. ACM Transactions on Information Systems, 10(2), 142-180.

T. CAWKELL (2000). Visualizing citation connections. In: The web of knowledge: A festschrift in honor of Eugene Garfield (B. Cronin and H.B. Atkins, eds.). Medford (NJ): Information Today, pp. 177-194.

P. DE BRA (2000). Using hypertext metrics to measure research output levels. Scientometrics, 47, 227-236.

J. DEAN and M.R. HENZINGER (1999).Finding related pages in the World Wide Web. In: Proceedings of the Eighth International World Wide Web Conference (WWW8) pp. 389-401. Elsevier Science B.V.

Y. DING, G. CHOWDHURY and S. FOO (2000). Journal as markers of intellectual space: journal co-citation analysis of information retrieval area, 1987-1997. Scientometrics, 47, 55-73.

L. EGGHE and R. ROUSSEAU (1990). Introduction to informetrics. Quantitative methods in library, documentation and information science. Amsterdam: Elsevier.

L. EGGHE and R. ROUSSEAU (2001). BRS-compactness in networks: theoretical considerations related to cohesion in citation graphs, collaboration networks and the Internet. Preprint.

Y. FANG and R. ROUSSEAU (2001). Lattices in citation networks: an investigation into the structure of citation graphs. Scientometrics, 50, 273-287.

R.M. FANO (1956). Information theory and the retrieval of recorded information. In: Documentation in action, J.H. Shera, A. Kent and J.W. Perry (eds.). New York: Reinhold Publ. Co., 238-244.

E. GARFIELD (1988). Announcing the SCI Compact Disc Edition: CD-ROM gigabyte storage technology, novel software, and bibliographic coupling make desktop research and discovery a reality. Current Contents, 22, May 30, 1988.

R. GARNER (1967). A computer oriented graph theoretic analysis of citation index structures. In: Three Drexel information science studies (B. Flood, ed.). Drexel Institute of Technology, 1-46.

W. GLANZEL and H.J. CZERWON (1995). A new methodological approach to bibliographic coupling and its application to research-front and other core documents. In: Proceedings of the fifth biennial conference of the international society for scientometrics and informetrics (M. Koenig and A. Bookstein,eds.). Medford (NJ): Learned Information inc.

M.R. HENZINGER (2001). Hyperlink analysis for the Web. IEEE Internet Computing, 5(1), 45-50.

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

M. KOCHEN (1974). Principles of information retrieval. Los Angeles: Melville.

G.H. LEAZER & J. FURNER (1999).Topological indices of textual identity networks. In: Woods, L. (Ed.), Knowledge: creation, organization and use: Proceedings of the 62nd Annual Meeting of the American Society for Information Science, Medford (NJ): Information Today, 345-358.

S. MAC LANE (1971). Categories for the working mathematician. Springer-Verlag: New York.

I. V. MARSHAKOVA (1973). System of document connections based on references (in Russian). Nauchno-Tekhnicheskaya Informatsiya, ser.2, 6, 3-8.

A. PRITCHARD (1984). On the structure of information transfer networks. M. Phil. thesis, School of Librarianship, Polytechnic of North London (UK).

K.E. ROSENGREN (1968). Sociological aspects of the literary system. Stockholm: Natur och Kultur.

R. ROUSSEAU (1987). The Gozinto theorem: using citations to determine influences on a scientific publication. Scientometrics, 11, 217-229.

R. ROUSSEAU (1992). Category theory and informetrics: information production processes. Scientometrics, 25(1), 77-87.

S.K. SEN and S.K. GAN (1983). A mathematical extension of the idea of bibliographic coupling and its applications. Annals of Library Science and Documentation, 30, 78-82.

H. SMALL (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, 265-269.

H. SMALL (1974). Multiple citation patterns in scientific literature: the circle and hill models. Information Storage and Retrieval, 10, 393-402.

H.D. WHITE (2000). Toward ego-centered citation analysis. In: The web of knowledge: A festschrift in honor of Eugene Garfield (B. Cronin and H.B. Atkins, eds.). Medford (NJ): Information Today, pp. 475-496.