Information Processing & Management Vol. 14, pp. 93–95
© Pergamon Press Ltd., 1978. Printed in Great Britain

0306-4573/78/0401-0093/$02.00/0

ON THE CITATION INFLUENCE METHODOLOGY OF PINSKI AND NARIN

NANCY L. GELLER

Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

(*Received* 2 *August* 1977; *received for publication* 3 *November* 1977)

Abstract—Using Markov Chain theory we give further insight into the citation influence methodology for scientific publications which was initially described by Pinski and Narin.

1. INTRODUCTION

In recent work Pinski and Narin[1] described a methodology to determine "influence measures" based on citations for groups of scientific journals, authors, or scientific subfields. This method assigns a value or "influence weight" to each publishing entity within the set of entities under consideration. These influence weights are normed to have average value one, and the higher the influence weight, the more "influence" the particular entity has within the collection. The methodology was applied to journals in physics in [1], to journals in biomedicine in [2], to journals in chemistry in [3], and to various collections of journals in many disciplines in [4]. In each discipline an influence hierarchy of journals was thus formed.

It is the purpose of this paper to show a relationship between the influence measure methodology and Markov Chain theory. This alternative approach yields additional results and lends insight into the methodology.

2. THE INFLUENCE METHODOLOGY

Pinski and Narin consider a collection of publishing entities such as journals, fields of research, or individuals, called "units" and put C_{ij} as the number of citations unit j receives from unit i. With n units under consideration they form the $(n \times n)$ citation matrix $\mathbf{C} = (C_{ij})$. Next they put $S_i = \sum_{j=1}^{n} C_{ij}$, which is the number of references the ith unit gives out. They form the $(n \times n)$ matrix γ whose ijth element, γ_{ij}, equals C_{ij}/S_j. With W_i as the influence weight of the ith unit and $\mathbf{W} = (W_1, W_2, \ldots, W_n)$ as the vector of influence weights, they determine that $\mathbf{W}$ satisfies the equation

$$\gamma^T \mathbf{W} = \mathbf{W}, \tag{1}$$

where γ^T is the transpose of the matrix γ.

Since (1) determines only a direction for $\mathbf{W}$ Pinski and Narin normalize by

$$\sum_{i=1}^{n} S_i W_i / \sum_{i=1}^{n} S_i = 1 \tag{2}$$

to assure that the average influence weight is one. Thus, they conclude, the influence weight vector they seek is the eigenvector $\mathbf{W}$ normalized by (2) which corresponds to the largest eigenvalue of γ (which is unity).

3. THE MARKOV CHAIN APPROACH AND ITS IMPLICATIONS

We wish to consider another matrix γ^*, with elements $\gamma^*_{ij} = C_{ij}/S_i$, which is a transition matrix for a Markov Chain (see [5] Chap. 5). We interpret the elements of γ^* in a natural way: among the references the ith unit gives out, γ^*_{ij} is the proportion given to the jth unit.

The matrix γ^* is relevant here because γ^* and γ have the same eigenvalues. This may be proved by observing

$$\gamma = \Lambda\gamma^*\Lambda^{-1}, \tag{3}$$

where Λ is an $(n \times n)$ diagonal matrix with ith diagonal element S_i. It follows from eqn (3) that γ and γ^* have the same characteristic equation, hence the same eigenvalues.

For γ^*, the equation analogous to (1) is

$$\gamma^{*T}\mathbf{W}^* = \mathbf{W}^*. \tag{4}$$

Along with

$$\sum_{k=1}^{n} W_k^* = 1 \tag{5}$$

(in place of an analogy to eqn (2)) eqn (4) gives the limiting stationary distribution of an irreducible aperiodic Markov Chain with transition matrix γ^*. Thus W_k^* is the probability of ending up in state k, independent of the initial state. In the citation context W_k^* may be interpreted as the long-run proportion of time, or probability, that the kth unit will be cited if citation patterns continue as they are at the time we form the matrix **C**.

We now determine the relationship between **W** of eqns (1) and (2) and $\mathbf{W}^*$ of eqns (4) and (5).

Substituting eqn (3) into eqn (1) and simplifying gives

$$\gamma^{*T}\Lambda\mathbf{W} = \Lambda\mathbf{W}. \tag{6}$$

Thus a scalar multiple of $\Lambda\mathbf{W}$ is a solution to eqn (4).

It follows using eqns (2) and (5) that

$$\left(\sum_{i=1}^{n} S_i\right) W_k^*/S_k = W_k. \tag{7}$$

Notice that $\sum_{i=1}^{n} S_i$ is the total number of references given out by the n units under discussion. We therefore can interpret eqn (7) as a relationship between the influence weight of the kth unit and the long-run probability the kth unit will be cited. The influence weight W_k is the expected number of citations (from the units under consideration) which the kth unit will receive per reference the kth unit gives out. This view lends motivation to the concept of "influence weight".

Since time homogeneity of citation patterns is a reasonable assumption, it follows that with aperiodicity and irreducibility, influence weights as defined by Pinski and Narin are a measure of the expected long-run behavior of unit-to-unit citations. Among the group of units under consideration, the higher the "influence weight", the higher the expected number of citations to the unit in the long-run, per reference given out by the unit. Furthermore, Pinski and Narin's "total influence", W_kS_k, is $W_k^* \sum_{i=1}^{n} S_i$, that is, the long-run expected number of citations among the units under consideration to the kth unit. Thus a ranking of journals by their W_k^* is equivalent to a ranking by total influence.

It is also of interest that the algorithm suggested by Pinski and Narin for finding the influence weights W_k is related to the matrix γ^m. In Markov Chain theory the solution $\mathbf{W}^*$ to eqns (4) and (5) is also any row of the matrix $\lim_{m\to\infty} \gamma^{*m}$, if the Markov Chain is aperiodic and irreducible. It follows from eqn (3) that **W** is also any row of the matrix $\lim_{m\to\infty} \gamma^m$.

Several other aspects of the influence methodology work are clarified using the Markov Chain theory. Concerning aperiodic and irreducible Markov Chains, it is well-known (see [5] p. 301, 309) that the largest eigenvalue of the transition matrix γ^* is unity and unity is a unique eigenvalue. Thus γ must also have that property, as mentioned by Pinski and Narin, but without published proof.

In their paper[1], Pinski and Narin observe that the influence weights of a two unit system would not be defined if $C_{12} = C_{21} = 0$, that is, if there is no cross-citation. Our approach leads to a general sufficient condition under which influence weights will be well-defined: that the Markov Chain with transition matrix γ^* is irreducible and aperiodic. A simple sufficient condition for this is that there is some positive integer m such that the matrix γ^{*m} (or γ^m) has all non-zero elements.

It follows that for some collections of journals, authors, etc., influence weights cannot be determined. For example, if the units under consideration fall naturally into two separate clusters, unity would not be a unique eigenvalue of γ^* (or γ) and the influence weights for the entire group would be undefined.

We also observe that the Markov Chain context explains why caution is necessary in comparing influence weights which arise from separate aggregates of journals (etc). Comparing one physics journal that has influence weight four among physics journals to another physics journal which has influence weight three among the same physics journals is not the same as comparing a physics journal which has influence weight four among physics journals with a chemistry journal which has influence weight three among chemistry journals. One explanation of this is that an influence weight is a function of the long-run probability of being cited within the aggregate itself. Thus a journal's influence weight depends on which journals it is aggregated with, and different aggregations will ordinarily lead to different influence weights for the same journal.

4. CONCLUSION

Using Markov Chain techniques, the influence methodology of Pinski and Narin is seen to be related to the long-run behavior of unit-to-unit citations if citation patterns over time remain as they now are. The Markov Chain theory also gives sufficient conditions for the existence of unit-to-unit influence weights.

Acknowledgement—The author is grateful to Dr. Gabriel Pinski for helpful suggestions.

REFERENCES

[1] G. Pinski and F. Narin, Citation influence for journal aggregates of scientific publications: theory, with application to the literature of physics. *Inform. Proc. Mangmt.* 1976, **12**, 297.

[2] F. Narin, G. Pinski and H. H. Gee, Structure of the biomedical literature. *J. Am. Soc. Inf. Sci.* 1976, **27**, 25.

[3] G. Pinski, Influence and interrelationship of chemical journals. *J. Chem. Inf. Comput. Sci.* 1977, **17**, 67.

[4] F. Narin, *Evaluative Bibliometrics; the Use of Publication and Citation Analysis in the Evaluation of Scientific Activity.* Computer Horizons, Inc., Project No. 704R, Cherry Hill, New Jersey (1976).

[5] F. S. Roberts, *Discrete Mathematical Models with Applications to Social, Biological and Environmental Problems.* Prentice Hall, Inc., Englewood Cliffs, NJ (1976).