

## BIBLIOMETRIC APPLICATION OF MARKOV CHAINS†

MIRANDA LEE PAO and LAURIE MCCREERY

Matthew A. Baxter School of Information and Library Science, Case Western Reserve  
University, Cleveland, Ohio 44106

(Received 4 March 1985; in revised form 8 July 1985)

**Abstract**—A rudimentary description of Markov chains is presented in order to introduce its use to describe and to predict authors' movements among subareas of a discipline. Other possible applications are suggested.

### INTRODUCTION

Bibliometric techniques have been used to describe stationary distributions of authors, citations and other observables in literature. Much progress has also been made in utilizing associative measures such as citations and coauthors to establish communication and/or subject relationship in the study of the structure of knowledge and the sociology of science. Fairthorne noted in 1969 that the study of the dynamic processes in scholarly communication had not enjoyed as much success in bibliometric research[1]. He cited Goffman's epidemic model as an exception. In 1971, Goffman made another important contribution in the study of the dynamics of literature[2]. He used the theory of Markov chains to describe and predict the movement of authors in sub-topics in symbolic logic. Zunde and Slamecka also modeled the process of science development as a Markov chain[3]. In a recent monograph, Goffman analyzes the authors in schistosomiasis and in mast cells[4]. The Markov theory in recent decades has found many applications in the physical and social sciences, and in engineering and business. It has been recognized as an important technique to characterize many non-deterministic processes. In this paper, we present the rudimentary knowledge needed to use this technique, to illustrate the method with the results of an experiment, and to suggest other useful areas of bibliometric applications using the Markov chains.

#### *Probabilistic process*

A physical *stochastic process* is any process governed by probabilistic laws. An example that comes to mind is the succession of heads or tails at repeated tossings of a coin. For example, having obtained two heads followed by two tails, the outcome of the fifth throw is completely independent of any or all of the previous throws. In mathematics, there are several types of stochastic processes. Of special interest to this paper is an important type known as the Markov process. It is named after a Russian mathematician, Andrei Andreevich Markov, who laid the foundation of the Markov theory in a series of papers starting in 1907. Its importance has been much enhanced in recent years by the many practical applications found in the sciences, engineering and commerce.

A *Markov process* is a simple form of stochastic process with the property that the conditional probability of an outcome depends only on its immediately preceding outcome, and not on any of the previous outcomes. That is, in running an experiment with the Markov property, as long as the results of the present experiment is known, the chance of the next experiment can be estimated. Data from any previous experiments can be ignored. It is also usual to think of this process as a sequence of positions

† This publication was supported in part by NIH Grants R01-LM-04177 and K04-LM-00078 from the National Library of Medicine.

Table 1. History of the movement of researchers among three institutes in the last four years.

States:	a	b	c
	$A_1$	$C_1$	$B_1$
	$C_2$	$D_1$	$D_2$
	$A_2$	$B_2$	$D_3$
	$A_3$	$B_3$	$D_4$
	$C_3$	$B_4$	$A_4^*$
	$C_4$		

\* Examples of State Transitions: Researcher  $C$  started in Institute  $b$ ; moved to Institute  $a$  in the second year; and remained there for the third and fourth years. Researcher  $A$  started in Institute  $a$ ; remained there for the second and third years; and moved to Institute  $c$  in the fourth year.

occupied by a moving particle. It is as though the process is being observed by a bystander at a fixed position. The successive times at which a particular position is visited by the moving particle gives enough information that by taking all the available positions, we can learn about the system as a whole from such fragmentary data[5].

Let us illustrate with an example. Suppose there are only three cancer research institutes in existence, and there are only four researchers working in this area of endeavor. Let us further suppose that each individual can be employed by any one of the three institutions, and at the beginning of each year, there is an opportunity to change jobs among the three institutes. We denote the institutes by  $a$ ,  $b$ ,  $c$ , and the individuals by  $A$ ,  $B$ ,  $C$ ,  $D$ . The researchers are free to move from one location to either of the other two locations directly. Table 1 gives information of the movement of the researchers among these institutes. At any given year, we are able to distinguish a distribution of researchers among the three institutes. Since there is a limited way of shuffling the four individuals among these three locations, we can specify all possible combinations. For example, one possibility would be that all researchers are employed at Institute  $a$ . Another would be that  $A$  and  $B$  are at  $a$ ,  $C$  and  $D$  are at  $b$ , leaving no one at  $c$ .

Some definitions will be presented in terms of an example. We speak of the distinct values the process can assume as *states*, and the totality of states is the *state space*[6]. There are two states each for the coin toss example. There are three states that our researchers can move into. If the state space of a stochastic process is finite or countable, the process is called a *chain*. This assumes that the set of states is exhaustive. By exhaustivity, we mean that the set of states consists of a complete set of alternatives and that we have enumerated all the possible states that our researchers can move into. Moreover, at any given time, one and only one of these states must be occupied by each researcher. Strictly speaking, we are ignoring the possibility that our researchers may elect to retire from this area of employment. If the boundary of each state does not overlap with any other state, they are said to be mutually exclusive. Each particle can be in only one state at each chosen unit of time. In our example, we do not allow for the possibility that an individual can be partly employed by both  $a$  and  $b$  at the same time, since this blending of two states does not constitute a distinct state in our set of states. Each researcher can only enter one of the three institutes. Elements that

can assume one of the possible values of the state space are referred to as *particles*. In our experiment, there are a total of four particles. Thus with three institutes as all possible states, and four particles that can assume any one of these three states, there is a finite number of ways these two parameters can be combined. Each combination is known as an *outcome*, or a *space point*. A *sample space* is a set consisting of all possible outcomes of an experiment. An *event* is any subset of the sample space. Although an event may have only one outcome, it may consist of more than one. For example, the event that each of any two institutes has exactly two researchers would consist of the following six outcomes:

Outcome	Institute a	Institute b	Institute c
1	a(A, B)	b(C, D)	c(0)
2	a(A, B)	b(0)	c(C, D)
3	a(C, D)	b(A, B)	c(0)
4	a(C, D)	b(0)	c(A, B)
5	a(0)	b(A, B)	c(C, D)
6	a(0)	b(C, D)	c(A, B)

For example, in Outcome #1, researchers *A* and *B* are employed at institution *a*; researchers *C* and *D* at institution *b*. Unfortunately, some confusion exists in the common usage of the two terms “event” and “outcome”.

It is customary to say “the process is in state *j* at time *t*”. Thus we are regarding these components as a system evolving in time. If the state space is finite the process is known as a finite Markov process. One may define the unit of one step as one throw, one year, or the time between the publication of two successive papers. Thus a Markov process is a special type of stochastic process.

### Markov chains

A discrete-time or finite Markov chain is a special stochastic process with three restrictions[6]:

1. The process must be a discrete-time process, that is, the movement of the particles among the states occur at finite intervals. In our experiment, the observations may be made at a yearly interval.
2. The process must have a set of finite states.
3. The process must possess the Markov property, that is, the estimated probability of each outcome relies only on the outcome of its immediate predecessor.

In our example, although employers may consider one’s long term experience, we can make a convincing case that in a rapidly developing field such as cancer research, one’s present performance is the sole consideration for hiring. A regular Markov chain is one in which every state can be reached from every other state. For example, our researchers are free to move to any institute from any of the three institutes.

If all these conditions are met, the theory of Markov chains states that given that the initial probabilities of the states are known, and given that the conditional probabilities of transition from any state *i* to any state *j* are known, the Markov process for this experiment has been completely described. By this we mean that if we are satisfied that our researcher’s experiment met all the conditions for a Markov chain, to describe all the movements of our system, we need (1) a reasonable assumption of the percentages of researchers at each institute as we start our experiment and (2) estimate of the chance that any researcher can move from one institute to any one of the three institutes. Notice that the possibility that one remains in the same location is also included. The probability of transition from state *i* to state *i* is a distinct outcome or space point. Thus for (1) we are referring to an initial probability distribution of the three states, an example of which would be that initially, each institute probably starts off with one of the four researchers. For (2) we must have data for the conditional probabilities of moving from state *i* to state *j*. A *conditional probability* is explained as the probability of event *x* given that event *y* has occurred. In our example, the

conditional probability of  $P_{ab}$  means the probability of moving to state  $b$  given that the previous state is  $a$ . Or  $P_{ij}$  is the *conditional probability* of transition from state  $i$  to state  $j$ . Thus there are altogether nine transition probabilities for our example.

	a	b	c
a	$P_{aa}$	$P_{ab}$	$P_{ac}$
b	$P_{ba}$	$P_{bb}$	$P_{bc}$
c	$P_{ca}$	$P_{cb}$	$P_{cc}$

*Transition matrix*

It is convenient to arrange the conditional probabilities by a square matrix, that is, an array of numbers with the same number of rows and columns corresponding to the number of states for this process. This is known as the transition matrix of this chain. Each of the columns represents the state to which the transition is made. Each cell  $P_{ij}$  represents the likelihood of moving from state  $i$  to state  $j$ . Each cell thus contains a non-negative value with the sum of each row equal to unity. In other words, from a given state  $a$  the sum of the probabilities of moving to any of the possible states is one. That is, the sum of  $P_{aa}$ ,  $P_{ab}$  and  $P_{ac}$  is unity. This transition matrix corresponds to the discrete-time Markov chain of our experiment and it contains all relevant information regarding the movement of particles among the states.

The question now is how do we derive values for the matrix? In practical situations, specific conditional probability  $P_{ij}$  is difficult to ascertain. However, a good estimate can be deduced from past data. Referring back to our example, suppose we are given the data from the past four years of the movement of these four researchers among the three institutes. Table 1 shows that  $A$  started the first year in  $a$ ;  $C$  and  $D$  in  $b$ ;  $B$

Table 2. Data for the movement of researchers among three institutes.

	a	4/5	0	1/5
M =	b	1/4	2/4	1/4
	c	0	1/3	2/3

Table 2-a. Initial probability distribution:  $a, b, c = (1/4, 2/4, 1/4)$ .

	a	0.80	0	0.20
M =	b	0.25	0.50	0.25
	c	0	0.33	0.67

Table 2-b. Transition matrix.

	a	.640	.067	.293
M <sup>2</sup> =	b	.325	.333	.342
	c	.083	.389	.528

Table 2-c. Stationary probability distributions.

	a	b	c	
L = M <sup>5</sup> =	a	.3335	.2662	.3996
	b	.3332	.2664	.3997
	c	.3328	.2667	.3997

Table 2-d. Reciprocal of stationary distributions of the three states:  $a, b, c = (3, 4, 25)$ .

in  $c$ . Then  $A$  remained in  $a$  for the second and third years before moving to  $c$ , and so on. We note that initially there is only one individual at  $a$ , two at  $b$ , and one at  $c$ . Thus the initial probability distribution for  $a, b, c$  is  $(1/4, 2/4, 1/4)$  (See Table 2-a). The conditional probability of  $P_{aa}$ , the estimated chance of any one remaining at  $a$  during successive years, is computed by counting the number of times a researcher staying at Institute  $a$  from the previous year. In this case, we have  $A1$  to  $A2$ ,  $A2$  to  $A3$ ,  $C2$  to  $C3$ ,  $C3$  to  $C4$ , a total of four. What other choices does anyone located at Institute  $a$  have? One can move to Institute  $b$ . The number of researchers who actually made the move is zero. Thus  $P_{ab}$  is 0. The number of individuals moving from  $a$  to  $c$  is  $A3$  to  $A4$ , a total of one. Therefore the conditional probabilities of transition from state  $a$  to  $a, b$ , and  $c$  are  $4/5, 0, 1/5$  respectively. The sum of the values of each row equals to unity. Continuing along this line of calculation, we obtain the values for each transition in the Matrix  $M$ :  $M$  is the transition matrix for this Markov chain of cancer researchers among the three institutes. The Matrix  $M$  gives all the information on any possible change from one location to another taken in one step.

### Limiting condition

Consider next the situation in which an individual moves from state  $b$  to state  $c$  via an intermediary stop at  $a$ .  $P_{bc}$  is accomplished in two steps:  $P_{ba}$  and  $P_{ac}$ . The transition matrix  $M$  does not give the conditional probability of  $P_{bc}$  in these specific two steps nor  $P_{bc}$  in any two steps directly. One must compute the conditional probability of  $P_{ac}$  given  $P_{ba}$ . That is the intersection of  $P_{ba}$  and  $P_{ac}$  or  $1/4 \times 1/5$  which is  $1/20$  or  $0.05$ . However,  $P_{bc}$ , the transition from state  $b$  to  $c$  in two steps may be accomplished in three possible paths. They are:

		total
$b$ to $a$ ( $1/4$ )	$a$ to $c$ ( $1/5$ )	$1/20 = 0.05$
$b$ to $b$ ( $2/4$ )	$b$ to $c$ ( $1/4$ )	$1/8 = 0.125$
$b$ to $c$ ( $1/4$ )	$c$ to $c$ ( $2/3$ )	$1/6 = 0.167$
		$0.342$

Thus the conditional probability of transition from state  $b$  to state  $c$  in two steps is  $0.342$  which is considerably higher than  $0.05$ . The calculation of the conditional probabilities of transition from any state  $i$  to any state  $j$  in two steps is equivalent to taking the power of the transition matrix  $M$ . In other words, multiply  $M$  by  $M$ , or  $M^2$  (Table 2-c). In matrix multiplication, the operation is row by column with each element of the row multiplied into the corresponding element of the column. Then the products are summed. The following illustrates the operation:

$$\begin{pmatrix} p & q & r \\ s & t & u \\ v & w & x \end{pmatrix} \times \begin{pmatrix} p & q & r \\ s & t & u \\ v & w & x \end{pmatrix} = \begin{pmatrix} pp + qs + rv & pq + qt + rw & pr + qu + rx \\ sp + ts + uv & sq + tt + uw & sr + tu + ux \\ vp + ws + xv & vq + wt + xw & vr + wu + xx \end{pmatrix}$$

By the same token, the transition matrix for any transition in five steps is  $M^5$  (Table 2-d). We may extend this method of finding the transition probabilities  $P_{ij}$  in  $n$  steps by the matrix  $M^n$ . The  $(i, j)$ -entry of the  $n$ th power of the transition matrix  $M$  contains the probability that the chain moves from state  $i$  to state  $j$  in  $n$  steps.

A theorem in the theory of Markov chains states that if  $M$  is a regular Markov chain, as the chain progresses in a large number of steps  $k$ , taking the powers of  $M$  will approach a probability matrix  $L$  in which the distribution of probabilistic values in each row is the same. These values are strictly positive. In other words, after a long sequence, the matrix converges. One obtains a stationary probability distribution. The values in each cell will not change with time, and they are independent of the number of steps taken. They are absolute probability values. This is variously known as invariant probability distribution, long run distribution, limiting probability distribution, or stationary probability distribution for the chain. This specific property has enormous implication. Given an experiment that can be described by a Markov chain, the results

of past outcomes may be used to predict the proportion of particles which will eventually end up in each of the available states. Since the initial distribution of the population of particles no longer affects the eventual distribution among the states and since the final distribution is no longer dependent on any of the conditional probabilities, one is able to make appropriate decisions based on the estimated percentages of particles eventually ending in each state. In various applications, this method has been shown to produce results which approximate the actual distributions.

In our fictitious cancer researcher experiment, we can reasonably predict the relative size of the cancer research departments in each of the three institutes after many years. There is no need to take into consideration the past history of each researcher's employment. Table 2 shows the intermediary matrices before  $M$  reaches the stationary probability distribution  $L$  after 5 steps. We may infer from these figures that after several years, 33%, 27%, 40% of researchers will be found in Institutes  $a$ ,  $b$ ,  $c$  respectively. Comparing the initial distribution of 25%, 50%, 25% and the limiting distribution for these three institutes, it appears that there is a tendency of a build-up at Institute  $c$ . Its limiting probability is a high of 40% from its initial value of 25%. Thus this Markov chain indicates a predictable pattern by its transition matrix.

Another theorem of the Markov chains is the following[5]. A regular Markov chain is irreducible and ergodic, in that every state can be reachable from every other state in one or more steps. If all the elements in the transition matrix of such a Markov chain after  $n$  number of steps are greater than zero, the theorem states that the limiting probability for any state  $i$  is the reciprocal of the expected mean recurrence time of the state  $i$ . This means that if a particle moves away from state  $i$ , the average time for this particle to return to state  $i$  again, measured in numbers of steps, is equal to the reciprocal of the limiting probability of state  $i$ . Since the unit for each step in our experiment is a single year, and the limiting probabilities for Institutes  $a$ ,  $b$ ,  $c$  are .33, .27, .40, their reciprocals are 3, 4, 2.5 number of years respectively. Therefore if any researcher should leave Institute  $c$ , the average individual would return to  $c$  again in an average of two and a half years. In Goffman's two experiments on medical investigators, he defined the unit of one step as the publication time between any two successive papers by an individual. He found that the investigators of schistosomiasis wrote an average of 1.5 papers[4]. Since the shortest recurrence time for any of his seven subtopics in schistosomiasis is three papers, there is little chance that the average investigator in this subject after leaving a subarea will ever return to it again. Similar conclusion was made for investigators in the subject of mast cell.

#### EXPERIMENT

Our objective is to illustrate the application of the theory of Markov chain to study the movement of authors in a non-biomedical subject. Comprehensive author data for the subject ethnomusicology was compiled from a ten year period, 1967–1976[7]. 2018 authors wrote a total of 3302 publications. To satisfy the three restrictions of a Markov chain, we chose the unit for each step as the time between the publication of two successive publications by the same author. Although strictly speaking, one could publish two papers simultaneously, there is usually a time difference. Such steps are discrete and finite. Second, there are a total of 9 subfields within ethnomusicology, most of which are divisions by geographic locations. These are commonly considered as all the subfields studied, and there is no overlap at least in the way each paper was classified in the abstracting journal *Repertoire Internationale de Litteratur Musicale*. The conditions of exhaustivity and mutual exclusivity applied. Thirdly, that the transition of an author from one subfield to another depends only on his or her present research interest can be defended adequately. Although availability of funds, one's educational background, availability of resources, materials and opportunity, and one's association can be influencing factors, one's present research focus is probably the major contributing force. To summarize, the process of author's movement among the 9 subfields is finite; the exhaustive list of 9 states or subfields is finite; each state transition depends only on the present state. Thus we may consider this experiment to satisfy the con-

Table 3-a. Initial probabilities for the movement of musicologists among nine subfields in ethnomusicology.

a: Discipline	=	0.0258
b: General	=	0.0258
c: Africa	=	0.0733
d: Asia	=	0.2027
e: Europe	=	0.3043
f: North America	=	0.0961
g: South America	=	0.0867
h: Australia	=	0.0173
i: Popular Music	=	0.1680

ditions for a Markov chain. This is a regular Markov chain since there are no external artifact to deter one from publishing in any of the 9 subfields allowing each of the available states to be reachable from every other state.

We tracked the movement of publications by each author among the 9 subfields, which were labelled from *a* to *i*. The first paper of each author was checked with respect to its relevance to subfield by the heading used by the abstracting service. Each paper was assigned to only one of the 9 subfields. There were 52, 52, 148, 409, 614, 194, 175, 35 and 339 papers in topics *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h* and *i* respectively. Thus our initial probability distribution is listed in Table 3-a. Next the number of transitions from any

Table 3-b. Number of transitions among nine subfields.

from	to →	a	b	c	d	e	f	g	h	i	Total number of transitions
↓	a	11	10	6	11	28	0	3	1	0	70
	b	7	4	5	15	10	3	1	0	2	47
	c	7	2	82	19	5	2	2	1	9	129
	d	18	5	13	78	21	1	2	1	3	142
	e	32	14	0	11	135	44	0	1	5	202
	f	2	4	3	2	0	17	4	0	12	44
	g	3	1	1	1	1	1	24	0	2	34
	h	4	2	1	2	0	0	0	14	0	23
	i	1	1	5	1	6	14	4	1	26	59
	number of transitions	85	43	116	140	206	42	40	19	59	750

Table 3-c. Transition matrix.

	a	b	c	d	e	f	g	h	i
a	.1571	.1429	.0857	.1571	.4000	.0000	.0429	.0143	.0000
b	.1489	.0851	.1064	.3191	.2128	.0638	.0213	.0000	.0426
c	.0543	.0155	.6357	.1473	.0388	.0155	.0155	.0078	.0698
d	.1268	.0352	.0915	.5493	.1479	.0070	.0141	.0070	.0211
e	.1584	.0693	.0000	.0545	.6683	.0198	.0000	.0050	.0248
f	.0455	.0909	.0682	.0455	.0000	.3864	.0909	.0000	.2727
g	.0882	.0294	.0294	.0294	.0294	.0294	.7059	.0000	.0588
h	.1739	.0870	.0435	.0870	.0000	.0000	.0000	.6087	.0000
i	.0169	.0169	.0847	.0169	.1017	.2373	.0678	.0169	.4407

Table 3-d. Stationary probability distribution:  $E^7$ .

	a	b	c	d	e	f	g	h	i
a	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
b	.1179	.0609	.1250	.1779	.3041	.0532	.0702	.0170	.0737
c	.1179	.0609	.1250	.1779	.3041	.0532	.0702	.0170	.0737
d	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
e	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
f	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0737
g	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0738
h	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
i	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0737

subfield to any other subfields was counted. Since 1441 individuals contributed only one publication to the subject, they were eliminated because they did not make any transition from their original subfield. For example, in Table 3-b, cell  $P_{ca}$  contains transitions from subfield  $c$  to subfield  $a$  and there were a total of seven transitions made from subfield  $c$ . Table 3-c is the transition matrix  $E$  for our experiment, which contains the conditional probability of movement from subfield  $i$  to subfield  $j$ . The sum of each row equals to unity. Consider subfield  $h$ , Australian music. There was very little activity in row  $h$  and column  $h$ , indicating that there was little transition from and to this subfield. On the other hand, there was relatively more activity for subfield  $e$ , European music.

Next, we raised the matrix  $E$  to the power of  $n$ . In the 7th step, the matrix converged to a stationary probability distribution. (Table 3-d) A comparison of initial and limiting probabilities indicates that the areas of European and Australian folk music are the most stable while the Popular Music and discipline areas are the least stable. The long run probability distribution also revealed that, in a few years, although Asian and



European folk music may retain their predominance among musicologists, one can expect a shift in interest from Popular Music and possibly North American folk music to African music. The proportion of authors in African music will rise from 7% to 13% while areas of North American folk music and Popular Music will lose a few percentage points. (See Tables 3-a, 3-d).

The reciprocals of  $L$ , the stationary distribution, were then taken to determine the mean recurrences, or the number of papers produced before an author returns to the same subject. The values for subfields  $a$  to  $i$  were found to be: 8, 16, 8, 6, 3, 19, 14, 59, and 14 respectively. Once an author leaves North American folk music, for example, the least number of papers required before he returns to it is 3. Yet the average number of publications per author is 3302/2018, or 1.6. Since these findings held true for all the subfields, it was concluded that it would be highly unlikely for an ethnomusicologist to return to an area of research once he or she has left it to pursue another area.

#### SUMMARY AND CONCLUSIONS

To verify the validity of the predictive model, publication trends in the ethnomusicology area were examined in later volumes of RILM. Unfortunately, due to delays in publication, volumes were only available for 1977 through April of 1980. The results of comparing the number of publications in each subarea for each year's volume indicated the following. During 1977 and 1978, European music retained its dominance among musicologists, as was predicted. 1979 and the first part of 1980 reflect the beginning of a decline in the number of publications in European folk music. Similarly, Asian music retains its high numbers of publications during the same period. North American music shows a slight decline in 1979 which carries over into 1980. Summing the publications in each subarea over the 40 months, one finds a significant portion of publications concentrated in Asian and European folk music. Although the percentage of papers in Popular music has suffered a slight drop, the predicted shift into African music has yet to materialize. Consequently, based on the limited data subsequent to our original analysis, predictions made as a result of using the Markov model cannot be conclusively verified. To unqualifiably support the previous findings, a repetition of the experiment with more data at a later date would seem to be in order.

The experiment in this humanistic subject has produced results comparable to past experiments in the sciences demonstrating that properties of the Markov chains can be utilized to describe and predict the pattern of movements of authors in the research areas of a subject, assuming that reasonable operational definitions and assumptions are made on the unit of each step; that is, the states or the distinct values the process may assume. Utilization of this process has far reaching effects in many areas. With regard to library acquisitions and development, the ability to predict authors' writing practices can be used to indicate areas to be strengthened, maintained, or retired, as well as for providing a more objective basis for collection building. It can further indicate trends of current and future research interests of library patrons.

The technique may be used to study the movement of writers among a fixed number of publishers that are particularly devoted to a discipline by their change of affiliation of publishers. Similarly, the trend of publishing among a group of journals may be tracked by the shift of citations among these journals. The movement of faculty members among universities that offer a specific program may also be followed by the personnel change among these schools. Finally, in formulating funding policy, one may be interested in studying the movement of research interest within a defined scientific discipline by tracing the citations or publications shifts in these subareas. Thereby based on such information one may formulate a policy with the view to anticipate growing and declining trends in science.

Operationally, the computation of the powers of the transition matrix becomes extremely tedious for chains with many states and especially for many steps. This task can be simplified by the use of the statistical package MINITAB for chains with less than 50 states. We also offer a sample program in Appendix I which can be implemented

on any computer with a minimum of modification. Even though incompatible dialects of BASIC are offered by different manufacturers, a BASIC interpreter is routinely included in any desktop computer. One can easily alter the size of the allowable number of states by substituting a larger number than 50 in the dimension statements. One should also check the specific commands needed to access the data file.

This paper discussed only the most elementary applications of the Markov Chains. Yet the theory can accommodate more complex situations, such as the merging of the states into fewer larger sets, as in the case of the cessation of journals. The relation between states may also be quantified by the average number of steps a particle will visit state  $i$  before entering another state  $j$ . The original process before entering a given state  $i$  may also be fruitfully investigated using the model of Markov process. There appears to be other potential applications of this theory.

#### REFERENCES

1. Fairthorne, R. A. Empirical Hyperbolic Distributions (Bradford-Zipf-Mandelbrot) for Bibliometric Description and Prediction. *Journal of Documentation*, 25: 319-343; 1969.
2. Goffman, W. A Mathematical Method for Analysing the Growth of Scientific Discipline. *Journal of the Association for Computing Machinery*, 18(2): 173-185; 1971.
3. Zunde, P. and Slamecka, V. Predictive Models of Science Progress. *Information Storage and Retrieval*, 7: 103-109; 1971.
4. Goffman, W. and Warren, K. S. *Scientific Information Systems and the Principle of Selectivity*. Praeger Press, New York; 1980.
5. Cinlar, E. *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, N.J., 1975.
6. Isaacson, D. L. and Madsen, R. W. *Markov Chains Theory and Applications*. Wiley, New York; 1976.
7. McCreery, L. S. Bibliometric Study of Ethnomusicology, A Humanities Subject, (Unpublished dissertation) Case Western Reserve University, Cleveland, Ohio; 1984.

#### APPENDIX 1

```

10 REM* PROGRAM FOR MATRIX MULTIPLICATION
20 REM*
30 REM*
40 REM* THIS PROGRAM READ DATA CELLS FROM A SQUARE MATRIX—M.DAT
50 REM* INTO A TWO DIMENSIONAL ARRAY A(I,J). IT THEN MULTIPLES
60 REM* THE MATRIX BY ITSELF AND STORES IT IN C(I,J). IT IS
70 REM* THE RESULTING MATRIX. FINALLY, IT ALLOWS THE USER TO
80 REM* CONTINUE TO MULTIPLE THE RESULTING MATRIX BY ANSWERING
90 REM* THE QUESTION: MULTIPLE THE RESULTING MATRIX AGAIN? (Y/N)
100 REM*
110 REM* MAXIMUM NUMBER OF ROWS ALLOWED IN THIS VERSION IS
120 REM* 50 BY 50.
130 REM* HOWEVER, THE DIMENSION MAY BE INCREASED BY CHANGING THE
140 REM* PARAMETERS IN STATEMENTS #200 AND #210.
150 REM*
160 REM* PLEASE NOTE THAT THE INPUT FILE MUST BE NAMED M.DAT.
170 REM* PLEASE NOTE THAT COMMAS MUST BE INSERTED BETWEEN NUMBERS IN M.DAT
180 REM*
190 REM*
200 DIM A(50,50)
210 DIM C(50,50)
220 INPUT "ENTER NAME OF THIS MATRIX: "; B$
230 INPUT "ENTER # OF ROWS FOR THIS MATRIX: "; K
240 REM*
250 REM* THE FOLLOWING SEGMENT READS DATA FROM DATA FILE M.DAT
260 OPEN "I", #2, "M.DAT"
270 FOR I = 1 TO K
280 FOR J = 1 TO K
290 INPUT #2, A(I,J)
300 PRINT A(I,J);
310 NEXT J
320 PRINT
330 NEXT I
340 CLOSE
350 REM*
360 REM* DATA FILE IS CLOSED
370 REM* THE FOLLOWING SEGMENT MULTIPLES THE MATRIX
380 PRINT

```

```
390 PRINT "THE RESULT OF MATRIX MULTIPLICATION IS: "  
400 LET S = 0  
410 FOR M = 1 TO K  
420   FOR J = 1 TO K  
430     FOR I = 1 TO K  
440       LET S = S + A(I,J) * A(M,I)  
450     NEXT I  
460     LET A = M  
470     LET B = J  
480     LET C(A,B) = S  
490     LET S = 0  
500   NEXT J  
510 NEXT M  
520 REM* THE FOLLOWING SEGMENT PRINTS THE MATRIX  
530 REM*  
540 FOR I = 1 TO K  
550   FOR J = 1 TO K  
560     PRINT USING "#.###:" C(I,J),  
570     LET A(I,J) = C(I,J)  
580   NEXT J  
590 PRINT  
600 NEXT I  
610 INPUT "MULTIPLE THE RESULTING MATRIX AGAIN? (Y/N)"; Z$  
620 IF Z$ = "Y" (or Z$ = 'Y') THEN GOTO 390  
630 END
```