Unit-1:Measuring of Information

1.0 : Objectives

This Unit introduces the concept of measurement and highlights the problems of measurement and lack of devices of measurement of intangible entities like information. This Unit also focuses on problems of definition in regard to information.

After reading this Unit, you will be able to:

- understand the difficulties faced for developing any operational framework for tackling measurement of information and informational objects;
- grasp the feature of measurement and measuring techniques for different types of entities, phenomena, objects or situations; and
- know the notion of Informativeness and how it can be applied to user-text interaction and database evaluation.

1.1 : Introduction

Lord Kelvin (William Thompson), one of the most important physicists of the 19th century, remarked that there should be written on every laboratory door the phrase - "through measurement to knowledge". The result of a measurement is usually expressed by a numerical value. This numerical value is again a real number. The human nature is such and the socio - cultural development of man is such that every person has a better understanding or insight of something if it is expressed as a number or a set of several numbers. Thus, measurements or numerical values play a vital role in human communication and in man's knowledge systems.

The simplest and most straightforward form of measurement is counting. In counting we get values of what we count in positive whole numbers. We can count objects, which exist as discrete entities, usually, the material bodies. We can talk of 543 books, 20 lacs 92 thousand 5 hundred unemployed persons (20,92,500), 25 words per page, 4.2 Megabytes (MB) of memory, 225 fields of a bibliographic record and so on. It shows that if there are objects, which may be physical or even conceptual, which exist or are available in discrete independent forms or units then we can count them and express the result of the counting as a natural number.

But not everything, with which we deal or are concerned with, is available in discrete unitary form. Water or electricity or performance in an examination or records of our senses such as colour of an object or smell or taste of something; or the amount of information in a document or the knowledge that a person possesses is not available in forms, which allow enumeration or counting. Our aspiration is to have all of these expressed as numbers so that we can understand and utilize them better. Water is a tangible material object but cannot be found as discrete or separable units. So water cannot be counted. Therefore, man has found some properties or so-called parameters with which water can be labeled such as volume, weight, density, viscosity, transparency, etc. So we can talk of 2 liters of water (as volume) or 200 grams of water (as mass or weight) or of refractive index for white light through water as 1.2 or of viscosity of water as 0.01 poise at 20°C etc.

Water is a tangible material object. The measuring devices and the scales, with which measurement of properties of water are expressed, are physical devices and physical scales such as a measuring glass, a weighing machine, a refraction apparatus or a viscometer. But when we want to measure the performance of students through an examination, we want to measure knowledge, understanding and capacity of expression of a student on a particular topic or issue. Knowledge, understanding and capacity to express are not tangible. Yet, examiners give marks say, 40%, 81%, 11%, etc. in numerical values. Therefore, to arrive at such results or measures we need scales and devices. Devices and scales for intangible things cannot be usually physical. They are mostly conceptual set up and logical processes.

Till AD 2000 there has not been sufficiently reliable, objective and universally accepted scales and devices for measurement of information but attempts are there for a long time. To understand the problem we have to understand the problem of understanding the nature of information and of the parameters that can be attributed to various aspects of information and the devices that can be invented and developed to measure those parameters. The crucial problems lie in the matter of finding out suitable units of measurement and inventing or devising suitable parameters.

1.2 : Information Revisited

Already in MLIS *01: Information, Communication and Society,* various approaches to information have been considered. It is clear that: information does not mean the same thing to every body. Yet we accept information as a basic entity. Relations of information with knowledge, wisdom, and data were also discussed. Some of the properties were also enumerated. For our purpose we need once again to present a strategic view of information. We can then deal with the issues of measurement of informational properties, of information content of message or text or discourse or document and of information processes including information service and use of databases and texts.

1.2.1 : Status of Information

Information is everywhere. Saying information is everywhere is like saying energy or gravitation is everywhere. Yet we have a feeling that information has something subtler than other fundamental entities or concepts like energy, matter or force.

Everyone talks of information and apparently understands it, but is not clear how information should be meaningfully defined. Indeed there is no workable general definition of information, which can be applied to all types of information processes and things such as multimedia, hypermedia, books and other documents, texts, messages, symbols, art objects, etc.

One can now supply precise well-formulated definitions to a number of fundamentally important concepts such as matter, energy, number, etc. The answers might not be philosophically fully satisfactory, but they are in most cases scientifically sound. By scientifically sound, we mean that the concept has got an unambiguous, literal description (i.e., through a statement or proposition formed in natural language terms) and the concept goes into some mathematical relations as a functional or as a parameter in relation to other concepts. It is neither too general to be obvious, nor too specialized to be applied or useful only in a single context. In essence we search for reality to a concept and try to establish structural solidarity in a conceptual edifice for that reality. Such a system or program is lacking for information.

1.2.2 : Three Attitudes toward Information

One basic difficulty in analyzing information exists because of three widely different attitudes. First, information is existent or meaningful only in `human' context. Only in human cognition and in human communication `information' is relevant. In other words, there cannot be any physics of information.

Second, information is just an expression or measure of organization or orientation of elements in a set. Suppose some (not all identical) items were arranged in a certain way. It depicts a pattern to an observer. Fay alteration of items the pattern is changed. Information in this view gives just an indication of the change, but not of the pattern or the elements (Communication theorists' approach is this).

Third is the action-oriented view. Information can only be considered in terms of a response to a stimulus. Information gives the idea of action or activity of a system being stimulated by input. The input is the information (Idea of Informativeness is just an extension of this view).

1.2.3 : Information in Discourse

Deer gives the requirements of the nature and property of the concept of information in ordinary discourse for defining its nature as:

i) Information be a representation;
ii) The representation be abstract;
iii) It (representation) be meaningful;
iv) It consists of determinations, which have been made;
v) It should have been made of certain objects.

1.2.4 : Defining Information

Belkin (1978) described information as the `structure' of any text, which is capable of changing image structure of some recipient. Mackay (1969) said informatior does logical work on the organizational orientation of a system.

Following and modifying B. C. Brookes' observation we define `information' as al entity manifesting communicable knowledge, capable of reducing or removing logica uncertainty and switching changes of action and/or pattern of organizational stab of a system. The communicable knowledge or information is semantic in essence however; we use the attribute semantic in its broadest sense for want of a bette term. Communication is sharing meaning. Therefore in information exchange thi most important point is that the receiver understands the meaning (semantic import of a message or text that is comnunicated.)

We say that:

i) For information to exist, communication and so, existence of material system including human beings as generator, receiver, preserver through space and time is necessary;

ii) Any information process i.e., generation, transmission, documentation, receiving etc. is embedded to some form of energetic process, that is, change of energy takes place whenever an information process occurs. But there exists no direct linear relationship between the amount of energy and the amount of information.

1.2.5 : Informetrics and Information Measurement

Informetrics includes bibliometrics as a narrower term and is best manifested by it. It is the tool of quantitative (and to some extent qualitative) accounting of information processes underlying the socio-cultural evolution of man. In bibliometrics we attempt at understanding diffusion of knowledge, growth and decay of invisible college growth of subject fields, productivity of authors, etc. We attempt these through word frequency counts, ranking of journals by counting articles on a topic, counting bibliographic references or counting the number of theorems in mathematics, etc. In doing all these, we are guided by an implicit hypothesis:

`All such items which we try to count are somehow (expectantly linearly related to cognitive knowledge vis-a-vis information per se: '

In effect we use the bibliometric parameters or variables as `information un (IU)'. Choosing a proper information unit is the most important aspect of informetrics. Otherwise we can never be sure whether our bibliometric analysis; can have any relation with cognitive flow of information, which is the basis of sociocultural evolution.

In the recent past there had been some debates about naming the subject, which had been variously called as bibliometrics, librametrics, informetrics, scientometrics, etc. It is now generally accepted that if we use the compound terms ` or the composite term *`sciento-informetrics'* then we can cover all the aspects of the subject. Sciento-informetrics is now applied to documents on internet. This is called *cybermetrics* by some authors. Consequently, we can consider scientometrics, informetrics, bibliometrics, librametrics, cybermetrics, etc. as components of sciento-informetrics.

The essence of sciento-informetrics or so to say of any of its components i.e., librainetrics, bibliometrics, informetrics, scientometrics and cybermetrics is precise measurement. Yet measurements in this area are steeped in ambiguity. Abraham Bookstein has shown the situation prevailing in sciento-informetrics as that of a developing science where measurement is not foolproof or straightforward. He points out that we require theoretical models, which should be able to connect observable parameters to devise methods of measurement to give meaningful results. Such models depend on, at their developmental stage, on intuition and on empirical or theoretical regularities of different variables. Empirical understanding is based on measurements. This understanding helps in model building. Then we rely on the model to guide our investigations towards measurement. That again helps us to develop insight and understanding. This is nothing but a manifestation of spiral of cumulative knowledge (as propounded by Ranganathan and others) and the triad formulated by Karl Popper, the philosopher of science, in his book *Objective Knowledge* as $PP_1 \rightarrow TT_1 \rightarrow EE_1 \rightarrow PP_2 \rightarrow$..., which means the same spiral of proposition, theoretical modelling, empirical investigation and re-statement of proposition or problems and further development of theoretical model and so on.

So-called informetric laws or sciento-informetric laws of scattering show a set of well-established regularities.

In Lotka's law the number of scientists (N), publishing n papers is proportional to the inverse of
n^a: $N(n) \propto n^{-a}$ *where a* is a constant.

Similarly in Zipf's law the regularity is shown by the rank, r of occurrence of a word in descending order of frequency of use in a text as proportional to the inverse of some power of the frequency, *f* such that constant

Similar also is the case with Bradford's law, which is a relation between the rank (again in descending order) of journals with cumulative number of articles on a subject or topic produced by those journals.

All these three cases can be generalized in some types of regularities of distributions called Zipfian distribution. Unfortunately these regularities cannot be shown as statistical regularities. Bookstein also questioned the concepts that guide data collection in informetrics. What are actually the core journals on which Bradford's law is defined and tested? What constitutes a word? How to define uniqueness of a word? Are specialisation and specialization are two words or one word? Are bay and bays are two words or one word? Similarly in Lotka's law, how a person is defined as an author or as a specialist in some field? Is a person Chemist because his paper is abstracted in the Chemical Abstracts?

Yet with all these ambiguities and imprecision the regularities are observed with all sorts of bibliographic data sets. It is therefore still an open situation for the question of measurement of sciento-informetric parameters and of the relationship between a group of sources and the items they produce.

1.3 : Framework for Information Exchange

For our purpose we consider that every information system (from human being to a computer to even an atom) has an inherent *K -structure.* At any space - time position, the system is in *a K-state.* Any output of information from the system doe: not necessarily change its K-state, only changes the *Energy State.* But any input of information may or may not change the K-state. K-structural activities within the system may also change the K-state. The 'K-amount' is an abstract subjective .counterpart of `information amount' of a system. K may decrease or increase. The decrease is most easily related to memory loss or forgetting and increase to discover: or gaining new experience in a human system. Example of K-state change b' information input is learning.

It is obvious that we have taken `K' as representative of `subjective knowledge or `knowledge within' in an abstract generalized manner. We know that what w know, are not always possible to express or communicate i.e., not always amenable to be transformed in `objective information'. `Objective information' is the `knowledge without' or `communicated knowledge'. `K' is an inherent property of an information system.

Communication is sharing meaning. In other words it is incorporation of information in different K-structures to give rise to change in K-state in an equivalent way. Sharing meaning means the amount of K- state change of different systems ca give rise to the same amount of information and vice versa.

Information to be communicated requires to be coded in symbols transferred through channels or preserved in some material body in the form of messages. Message are embedded in language (we again use such a term for lack of better one). The information can be shared only if the sharers can produce and receive sernantic elements in *atomic information units (AIU or aiu).* AIU's are information-carrying concepts. There may be information units (IU or iu) or semantic elements which may carry meaning of a large string of aiu's. By analogy we may say that IU's at molecular concepts. For the purpose of measurement IU s may be taken as equivalent to the units of information. A unit of information is an item (informational) c convenience. It can be a byte, a word, a document, a part of a document and anything that can have an 1U counterpart.

An iu or aiu may be transmitted through strings of message symbols in some code. I f the symbolic string breaks up or get disorganized the semantic essence of the 1U is lost. And one cannot divide an aiu and preserve the information. Information (and the K-state) is then a particular pattern or organization of information elements based upon organization and pattern (as well as state) of some symbols or representations embedded of material and energetic elements (of molecules, atoms, larger objects, acoustic waves, electro-magnetic (EM) waves, movements of electrons, etc.).

Let us now represent information by I, information units by J, energy by E and K-state by K, measure of any parameter by m, such as m (i), information on or about information by I(I), uncertainty by U, etc. We shall use other symbols in common use with their common meanings.

Information equations and relations may be considered as follows:

1) $$I = |I_1 - I_2| = *J_{pq}(*j_{qr})$$

*representing composition on p and r.

J are the AIU's or core or prominent information elements and j are the aiu's or supporting information elements. A statement like `want articles on air pollution in Indian cities' would give a composition of core information elements of pollution, air, articles with supporting elements of cities, Indian. Some of the concepts may not be atomic as is cities. Question would be which cities.

2) $m(\Delta I) = in(*J)(*j)$

$= m(a_P\, a_r)$

a_P and a_r representing values of aiu's equivalent to J and j.

The values of a's are to be determined in regard to the pattern of arrangements in some units (by proposition).

3) Output ΔI: $K \rightarrow K$, when ΔK amount is output as ΔI amount of information, such that $\Delta I \geq 0$; there is no change in K even if ΔK is output.

4) Input $\Delta I \rightarrow K \rightarrow K + \Delta K$, $\Delta K \geq 0$, $\Delta I > 0$. This means that when someamount of information (ΔI) is input to a system from outside, its K amount may change if the knowledge is new ($\Delta K > 0$) or does not change if the knowledge is already known ($\Delta K = 0$).

5) $\Delta K = f.(E)$, $\Delta I = g(E)$; $\Delta I > 0$, $\Delta E >> 0$, $\Delta K \geq O$. This means that whenever there is some change in information or knowledge, there is a corresponding change in energy (E) of the system.

6) $m(I(I)) << m(I)$. This means amount or measure of information about information is much less than the amount or measure of information itself. For this reason, indexing or cataloguing or classification is important.

If there is a decisional problem and uncertainty is say, U_1, then

7) $U_1 + \Delta I \rightarrow U2 < U1$

This illustrates that information AI reduces uncertainty from U2 to U1.

We have talked of logical uncertainty because uncertainties inherent in a *system* like that connected with uncertainty principle of quantum physics cannot be *reduced* or removed by any amount of information from or regarding that system.

By logical we have primarily meant `inductive'. Information processing or output is essentially inductive. We should remember statistical accounting is inductive. Accounting of pattern or arrangement is also inductive. Our approach is here only indicative and the research dealing with these issues is at the preliminary stage.

1.4 : Measurement

The subject of measurement is a complex but well-established domain in its own right. Measurement is essential in any science. Without measurement no science can be developed; indeed no useful knowledge is possible without measurement.

Simply defined, measurement is the assignment of numbers to observed phenomena according to some rules excepting any type of randomization rule. In fact measurement enhances information value on any system and reduces uncertainty or randomness.

Defined in a slightly different way, a measurement is a quantified observation o: some attribute or aspect of (parameter related to) an object, process, project o system (and a set of objects, processes, projects or systems). Mathematically speaking, measurement is a correspondence or a one to one mapping from some assignments that may be natural or artificial on an observable through an observation on to the real line (i.e., the real number system). It is important in measurement theory and practice that only those parameters, which are observable, are to b measured. The correspondence or mapping preserves the observed relationship and operations performed on the parameter in the same manner on the real line. This is called homorphism' or homomorphism' (in mathematical terminology).

1.4.1 : Requisite Conditions of Measurement

Measurement is used differently for different types of phenomena, parameters and contexts. The way we measure time is very different from the way we measuring the action of a drug. Both these are very different from the way we measure the performance of a student in an examination. Again it is different when we try measure (determine) the marketable value or price of a product or of a service.

Any measurement -

i) Is an energetic process; it needs effort and it involves expense of energy. The process of measurement involves a stimulus-response complex; this affects the system under measurement. It generates a fluctuation in energy distribution in the system. In other words, it enhances entropy. This reduces the precision of result of measurement and thus, makes a deviation from expected exact information. Because of involvement of exchange of energy, attempts of increasing sensitivity of the measurement mechanism or of measurement instrument results in increasing noise. The issue of exactitude or precision in measurement is the problem of reducing noise, which is the problem of minimizing the effects of energetic fluctuation, which again means minimizing informative fluctuation.

ii) Interferes with the system on which measurements are made. Therefore the measurement is always imprecise and tells about a condition, which may not permit repetition or second time measurement in any future time. The process of measurement increases entropy.

iii) Is to be made on a well defined parameter with an exact expectation about the outcome of the measurement, not in terms of the `value' or `result' of the measurement but in terms of its `meaning'.

iv) Is to be expressed as a `datum', a value that can be translated or transformed into a numerical expression. Therefore each and every measurement must follow a `code' (encoding process), a `scale', a `unit', a `point of reference' or `origin' or `zero' of the scale; in other words any measurement must have a well-developed `frame of reference'. The scale is usually a linear scale.

v) Requires a device or an `equipment' or a `tool' or a `measuring instrument' and a set of `step by step' measuring process.

However measurements follow some conditions to be performed. Then there are some general facts about measurements. All measurements are: (i) arbitrary to some extent; (ii) approximate; (iii) relative; and (iv) in some sense or other a `validation' and an `evaluation'.

1.5 : Scales of Measurement

A scale is a function or a functional f '. Mathematical concept of a scale is a homomorphism between two relational systems. An n-dimensional scale is a homomorphism in of an irreducible empirical relation system (a,, t_1'..., t_m) into an n' dimensional numerical relation system (R^n, Si, ...S_m) where R^n

is the nth Cartes product of the set of real numbers. The empirical relation system is the assignation, on the parameter, which are to be represented as numerical measurement values.

Keeping aside this abstract (mathematical) approach of an n - dimensional se let us consider a linear or one dimensional scale, that is, when n = 1. A 1 dimensional scale (the special case for n = 1) is a mapping from an empirical to the real number system R. Most of the time we use such a linear scale. Bu times we may use a set of real numbers in order of preference usually called tuple or in specific cases binary or duplet, triplet, quadruplet or 4-tuples, 5-tup etc. when we need to represent measurements (or vectors) of multi-dimensional representations.

Scales are neither unique nor the same for all purposes, they can be of differ types. The major types are: (1) Category scales which are also called nominal qualitative; (2) Sequence scales - these are usually words or numbers in ordered sequence with uniform spacing between them like as the months of a 3 -January, February, etc.; they are also called ordinal scales; (3) Quantitative sc which can be of many varieties but always give results in numbers; major sub-t} are interval, difference, ratio and absolute scales.

Nominal (or category or qualitative) Scale:

Function `f is a nominal scale if the admissible transformation • from empirica numerical is the set of all one to one mapping from R (real line) to R. Such a s consists of an ordered or unordered series of words and/or numbers, that na identify or describe observable. Each word or number defines a distinct cater that connects one or more entities say for example *good, fair, average, poor* say, *very informative, fairly*
informative, poorly informative, not informative all. Such qualitative assignments can easily be changed into numerical nom assignments through some rational weightage e.g., 10, 6, 3, O.

Ordinal Scale:

Function f is an ordinal scale if the admissible transformation is set of monotonically increasing continuous mapping from R to R. The examples are ratings or ranking or numerical rolls such as accession number in an Accession Register or scores which can be expressed first, second, third, fourth, (1st, 2nd, 3rd, 4th), in a competition, say. In many of the sequence scales the numbers are used identification and do not have quantitative significance. Time series and order occurrence scales are the most widely used sequence scales.

Interval Scale:

Function '*f*" is an interval scale if the admissible transformation is the set a positive linear transformation of '*f*" with the functional relation *off(x) = ax* examples are almost all scales of physical measurements like temperature, ti distance, etc.

Numbers are used for denoting the values of the parameter under measurers and not just for identification or tagging.

Difference Scale:

Function '*f*"is a difference scale if the admissible transformation is the se mapping differing from *f* by a constant: *G (f') = f (n) + C.* It is sometimes ca step up scale. A difference scale must have a unique unit of measure but may I an arbitrary origin.

Function '*f*" is a ratio scale if the admissible transformation of the set of mapping differing from *f* by a positive multiple which is called similarity: *f(x)* = ax. Examples may be sizes or weights or such expression as `three times more informative', etc. Ratio scales have a unique origin but an arbitrary unit of measure. Zero volume or zero weight is unique as origin but a glass of water may be x milliliters or y drums or w grams or v ounces. Similarly, zero informativeness or zero relevance (not relevant at all) may be true in a unique sense in context of information retrieval or informational assessment; but on the contrary a document may be considered as of relevance weight zero point five (0.5) or zero point seven (0.7) or zero point three (0.3) depending on the nature of ratio scales used for assessment.

Absolute Scale:

Function f' is an absolute scale if the admissible transformation of the set of mapping is the identity transformation *f(x)* = x. This is the case of all types of counting. There is arbitrariness in case of counting. In the context of information measurement in most cases, the first step is counting such as, number of words, number of sentences, number of articles, number of books, etc

1.6 : Measurement Techniques

Techniques of measurement differ widely for different parameters in different subject areas. In physical science usually a physical scale and a physical device is used. Results of measurements are usually observed optically. In most cases the records are manifested as changes in length or in angle through a graduated instrument. The temperature is recorded from thermometer represented by the length or height of a mercury column. The time is recorded on a circular scale of a watch or clock by recording the angle covered by the hands of the clock. In case of measurement of volume we use sometimes graduated cylinders or pots, the graduation .represents a length. Now-a-days, however, numerical values of measurements or records can be shown directly as numbers on a display device, such as a liquid crystal device (LCD) which are called digital scales or digital devices.

Records of measurement are also obtained from comparison or equivalence or equality by standardizing a unit of the observable parameter. Volume or weight of liquids is measured in this way. In the first case a pot or vessel with a standard measure is used to compare volumes by pouring liquid into that vessel. In a common balance weights are compared to a standard weight say one kilogram or one gram.

1.6.1 : Utility of Measurement

The most important use of measurement is for Identification; but is also important for Comparison, Value judgment, Classification and knowing better. All these are very important for measurement of information because it depends upon the information state and structure of the system and upon the user's capacity to decode the message embedded in the informational item. In sciento-informetrics the utility of measurement can be best understood by consideration of exploiting o bibliographic data.

Bibliographic data represent socio-cultural output. Folk information, which is unrecorded and communicated orally, becomes bibliographic data if recorded or documented Manipulation of bibliographic data can therefore lead to in understanding of sock cultural activities. So far bibliographic data have been applied to the scientific and technological activities through bibliometrics. We have a vast field opening before us for collecting and processing other types of bibliographic data for understanding of different types of socio-cultural activities.

1.6.2 : Counting and Measurement

Counting is also called enumeration. Counting is only possible if items under counting or the things to be counted are available as a set of discrete objects. One may count human beings or a set of grains or stones but can not count a glass of water. For measurement there must be a device of measurement and a scale explicitly defining the unit of measurement. Counting is the most primitive way of measurement. In sciento informetrics all types of bibliographic data are mostly counted.

1.6.3 : Classification and Counting

Unless something is available as a unitary object with a certain property it cannot be counted. Counting is possible only when the observable property is uniquely defined over a set of countable items. When a sub-class of species is formed from a generic class it is formed on the basis of some common property or properties defining the species. In that case by counting the items we get a measure of the speciating parameter. Thus, we can count cows because we can form a sub-class of cows with common, property of cow-ness using it as a speciator from, the class of animals. In a similar manner we can count words or sentences -or documents. Thus, counting of an intangible qualification or entity can be possible if it ha a manifestation through countable observable. Therefore, many intangible entities in be counted because they manifest through countable symbols or they can be artificially represented through observable and countable manifestation.

1.7 : Standardization of Measurement

The usability of any measure or measurement depends upon a uniform practice, which can be established and ensured through standardization at international. and national levels. The standardization covers the point of origin or the zero in the scale, unit and sub-units. The scale of the dimension if it is needed (i.e., if the measurable parameter is not a pure number), the equivalence or exchange values with other scales and other units (if required) and techniques and devices of measurements are to be specified. Thus, for the temperature there is a standard absolute zero, the scales of measurement such as Fahrenheit and Centigrade: have established transformation rule from one to the other. A thermometer is to be tested for its effective working and keeping the standard. In case of information measurement we do not have any standard practice as yet.

1.7.1 : Accuracy and Stability

The first condition of any measurement is that it must be accurate but we have already noted that all measurements are in some way or other approximate. This means accuracy or precision in measurement needs to be considered in the background of purpose, content and the nature of measurable parameter. When one talks of the distance from one place to another, say from railway station `A' to railway station one expresses the distance in kilometers knowing fully well that the more accurate or more exact figure may be few meters more or less, but when one has to talk of the distance between two mirrors in a precision optical measurement the distance may be needed to be expressed in the order of accuracy of $1/10^9$ meters. Even then there will be an inaccuracy beyond 10th place of the decimal in the measurement. We, therefore, talk of the limits of error or so-called deviations almost usually expressed by standard deviation or standard error. So, we express the speed of light in vacuum as 299792 4 kilometers. This means limits of accuracy is 4 kilometers more or 4 kilometers less than the stated value which again means the exact value may lie between 299788 and 299796 kilometers.

Coming to the cases of dealing with social or intangible parameters the situation is much worse and imprecise. Because of this we usually talk of `true' score and `observed' score. It is a fact that nobody can have a direct guess about the value of the true score. From different measurements of the same `state' one can get an idea of a true score around which the observed or measured scores may be available or obtained. Take the case of evaluation of answer scripts. A student S_1 answering a set of questions Q gets 52 in a 101-point scale from an examiner $E_{1.}$ The same student writing answers to the same set Q at another point of time may get something different from 52 from the same examiner E,. Another examiner E_2 may award still another marks to S_1, against answers of Q. Only with a good number of trials we can arrive at an average score which may be considered as the best approximation to the true score. With such trials we shall also be able to, estimate the limits of error or deviation by calculating the standard deviation of all the scores, which is usually called the standard error.

Another way of approaching a precise or accurate value is to device a method of arriving at results of measurements in an exponential manner either in the descending or ascending order or both. These are done mostly through a technique called iteration.

Repeatability of observation is another pre-condition to precise measurement. This requires a stable environmental condition and a stable response. This is most important when we want to observe or measure something on the basis of human activity or human response. Answers to the same questions put to the same person in different manner or at different times but within the framework of same context should bring forth the same answer. This is the condition of stability. If the answers differ then confusion may arise. The two answers may be independent or interdependent depending on the person's memory, understanding and outlook. This point is important when we want to assess or measure the effectiveness of information services or when we try to assess the information content of a document through responses from different persons.

1.7.2 : Reliability and Validity

Reliability is the condition that the technique of measurement should give the same value, which is dependable in different situation and at different points of time. In many cases we can measure reliability through the judgment of error between 0 and 1. In all cases of statistical estimates reliability can be expressed through the ratio of true score variance to the observed variance i.e. It = st^2/sx^2 for measurement of a variable (or parameter) x.

A measure is valid if it can be assessed in relation to an observed measure. It is intuitively clear that no measure can be valid without also being reliable but a reliable measure is not necessarily a valid one. Measurement validation is a loop consisting of the functions or activities of `define (or specify) - collect - present - use -' and their feedback upon each other.

1.8 : Information Objects and Content

M. Buckland observed that information can exist or rather information can be approached in three ways information as thing; information as process and information as knowledge. When information is recorded, the record or its components exists as information as thing. This may be called informational objects. In most cases such objects can be counted. Typically enough information objects range from a whole document like as a book or CD-ROM to an alphanumeric symbol or bit. In many of the information exchange content amount of message which is considered to represent information is measured through the number of such objects. We talk of 32 megabytes of computer memory, or 3000 words for an article or 800 papers in informetrics, or 125 books issued from a library desk in a day. Such measures are considered related to the amount of information in one context or other.

1.8.1 : Information Content

It is easily understood that the number of informational objects by counting can not provide a true picture of the information content in any such object or record or message. We therefore consider something, which we call semantic value, is the amount of meaning that is carried by an informational object. For a record R, a document Ranganathan used the word `thought content'. Here, lies the difficulty. Till now we don't have any unit for measuring semantic value or thought content. In natural languages the word is usually the lowest unit carrying semantic value but the same concept can be expressed by different words in different languages and even in the same language. Thus, the problems of synonymous or equivalence terms create problem in using words as units of information. More importantly, when words are combined to form sentence or sentences are combined to form texts or message there is no uniform rule of combination of the meaning or the semantic values. Most importantly in such combination semantic values or information content are never ad S_1ditive. Indeed they do not follow any uniform rule so that any method of measurement scale can be applied taking the word or the sentence or any component of the text as unit. Classificationists attempted a way out by taking the universe of knowledge as a whole and assigning subject components in a hierarchical order to represent thought contents of documents. But the classification number can only represent the nature of the thought content of a document but not the total amount of its information content and the total amount of semantic value contained in a document

1.8.2 : Range of Information Content

There are many debates about what should be considered as information content in case of transmission or exchange of information. Should we take into consideration each and every component of a message or a text including all the details, redundancies and exemplification or should we consider an indicative usable package from which the recipient of the information would be able to work out the details.

Let us take for example the subject of plane (two-dimensional) classical Euclidean geometry. A sender may send statement of all the axioms and theorems working them out with solved problems and examples. On the other hand the sender may send to the recipient only the axioms and statements of some of the theorems and the methods or rules of solving the theorems and developing the subject. Are these two messages informationally equivalent? If so, there must be some way of showing this equivalence. Unfortunately, we don't have any means so far. There is also a similar problem when we use conceptual contraction. In the development of the subjects it happens often that a new term is coined to represent some idea or concept which required a number of terms to represent the same concept or idea earlier. In such cases the single term and the whole chain of description have equivalent information content. To overcome these a number of attempts have been made in terms of response to a stimulus, reduction of uncertainty, change of informational status or state, assessment of informativeness, etc. some of which will be discussed in latter sections of this Unit.

1.9 : Informativeness

Tague-Sutcliffe states that information is an, intangible aspect of interaction between a text and a reader. The way a reader benefits from a text can be used for comparing the usefulness, hence information content of a text for a reader or for different readers. One may say that a certain text P is more informative than another text *Q* for a reader R,. Similarly, the text *P* may prove to be less informative to a reader R2. The view of informativeness as a measure of information is dynamic and is similar to the concept or view of relevance presented by L. Schamber, M. B. Eisenberg and M. S. Nellan. According to them relevance is a dynamic concept and depends on the judgments of users on the quality of the relationship between information provided by the information item and information need felt by the user at that particular point of time. Degree of relevance can be used in quantitative terms if it is approved from the user's point of view both conceptually and operationally. Thus relevance can be related to the amount of ^*information wanted'* and ^*informativeness'*. A text, which is more relevant for a particular user, should be more informative to that user. If measurableness of information depends on such a concept of informativeness then it is actually an approach for measurement of information services. An information service and hence measurement of informativeness is also somewhat contextual. Ultimate user interaction with text depends on how the collection in the database has been developed, how an item of collection and the collection as a whole have been organized and described (indexed), how the text or information items are retrieved and how they are packaged and repackaged and supplied.

From the foregoing discussion it is understood that a text is informative to a user only when it adds information to the knowledge store of the user ($K + \bullet\ I \rightarrow K > K$ in the framework presented at section 1.3). Therefore, informativeness can never be negative but it can be zero. When it is zero, the user considers the text as non-relevant, useless or uninformative.

So we can say -•

Property 1 - Informativeness is a non-negative number associated with interaction of records with user. Informativeness of a record or a text may vary from user to user.

Property 2 - Informativeness can not be measured directly. However, user's reference for ranking of records relative to the amount of information. preserves any ordering of informativeness values. So a user may be asked to help prepare a rank order list of informativeness for a number of texts, documents or records.

Property 3 - Informativeness is not necessarily commutative or additive under concatenation. Concatenation means different records or texts are read or used in sequence (i.e., one after another). The order in which records are concatenated (sequence for interaction or use) may affect their informativeness ranking.

Assumptions :

Let us consider that

1) there are two texts T1 and T_2 which are related through content or topic;

2) there are two clients C_1 and C_2 with nearly equal background information state in respect of the topics of texts $T_1$and T_2. Let this information background state be j,;

3) C_1 is given the text T_1, and T_2 in this sequence i.e., T_2 is given after T_1 is read. But C_2 is given the text in alternative sequence i.e., T_1 after T_2 is read.

If C_1 and C_2,are asked to give judgments about informativeness I of T_1 and T_2. Will their judgments be same or similar ? The answer is no.

There is no cornutativity in concatenation. If the texts are of different subjects, then pure additivity or additiveness may be valid in concatenation.

$I(T_1)$, $1(T_2) \sim I(T_2)$, $I(T_1)$ [if the two texts are on similar topic].

There is no cumulativeness in concatenation. If the texts are of different subjects, then pure additiveness may be valid in concatenation. $I(T_1) * I(T_2) = I(T_2) * I(T_1) = 1(T_1) + 1(T_2) = I(T_2) + I(T_1)$, if the two texts are of different topics and both topics are unknown to C_1 or C_2.

Property 4 - The total informativeness of a record or a document is invariant under granularity or partitioning in different ways. Partition means dividing into disjointed segments. This property asserts that sum of the informativeness of the partitions is equal to the informativeness of the whole record (or document). But the sequence of the partitions should be maintained. Any breaking or alteration of the sequence of partitions does not guarantee this property. This condition is called *subsequence.* The partitioning or granularization can be taken down to the level of sentences.

Property 5 - When records are ordered according to non-increasing user reference i.e., the more informative record is given lesser rank than the less informative one; in case of ties the ranks may be arbitrarily chosen, the informativeness of a subsequence is approximately proportional to the logarithm of the number of records in the subsequence.

This indicates that effective informativeness of documents or records goes on decreasing with the use of documents i.e., acquiring of knowledge: There is a fallacy in this proposition. Unless the records or documents have related topical content/thought content, decreasing nature of informativeness can not hold good. However, for most specialist users probability of acquiring new information decreases or in other sense probability of encountering already acquired or known information increases more or less exponentially. If informativeness decreases exponentially (or equivalently in a power law sequence) then logarithm straightens out the index (power) and gives rise to direct proportionality or linearity. That is when $I(T_a)$, ¥ k^n (or, a.exp bn); then log $I(T)_a = c.n$ where $I(T)$, is the informativeness of nth text and k, a, b, c are arbitrary constants.

Property 6 - The informativeness of the response of an information service to a user query is related to the completeness and ordering of the records perused in the retrieval process with respect to those in an ideal chain of records where this chain represents the records that may completely satisfy user information need.

"Queries can be categorized in terms of the number of records in the ideal chain, the number of alternative ideal chains that could be provided, and the extent in which the records and ordering of the ideal chain will vary from person to person with the same query". Queries requiring short actual answers (fact-finding or short-range questions) will have short ideal chains and will vary very little from one user to another. On the other hand ideal chain for a major research study or a new knowledge area may be very different *for* different users.

Property 7 - The informativeness of an information service organization for a user community (actual or potential) is a measure of the extent to which the service satisfies the information needs of the community. It is a function of the degree to which the community needs are served by the organization and the timeliness and informativeness of each service.

Note : The informativeness measure must be capable of aggregation (accumulation and cumulation) over the records and users if it is to serve *as* a basis for evaluation of information service organization.

1.9.1 : Use of Informativeness Measure

Tague-Sutcliffe mentioned and illustrated uses of informativeness measure in lot areas of :

1) *Collection development* and assessment of information content of a collection or database. This can be done by comparing user assessment of informative values of different databases or for the same database at different stages of development. One may also assess the amount of information that a single document may provide to the user community over a period of time and the information, the document may be expected to provide for a certain length of time in future.

2) *Document description :* The measure may be used for assessment of effectiveness of descriptions of documents for different groans of users. The description includes all varieties of indexing, abstracting, classifying, cataloguing, extracting, etc.

3) *Retrieval processes to* estimate the information of the sequence of documents or records retrieved using a particular search strategy for a particular user for a particular need at a particular point of time. Simultaneously one can use the measure to assess the amount of useful or potentially pertinent information, which remain non-retrievable or missed. Retrieval efficiencies of different search strategies may be assessed.

4) *Repackaging and consolidation* of information can be evaluated against usefulness and against the original items of information. This may be an important application in the new development of knowledge management or content analysis and synthesis in cyberspace

1.9.2 : Observations

1) Informativeness of a text is a relative measure (not a universal or an absolute measure).

2) Informativeness values of texts depend upon:
   i) topical content or thought content of the texts;

   ii) informational state or knowledge state or K-state of the clients interacting the texts;

   iii) concatenation or sequence of interaction of the texts by the clients;

   iv) the amount I(T) (which can be hypothesized but cannot be measured or evaluated mechanically or directly) that is information content or thought content or semantic content of the texts;

   v) the commonality or disjointedness of I(T)'s of texts;

   vi) the informativeness I of texts cannot be additive unless I(T)'s are disjointed.

C_1: $K_{.1}$-- $I(T_1) \leftarrow K_1 + [\Delta K]$

ΔK represents the absolute information amount of Text T_i.

1.10 : Appendix

Algorithm to Determine the Informative Chains (This algorithm is taken from Tague-Sutcliffe, 1995).

**Input N and array PREFER

*N is the number of records in the informative set.

* The input array PREFER is defined as follows :

PREFER (I, J) = 1 if the Ith record is more "
informative than the Jth record,

= 0 if the Ith and Jth records are equally informative,
= -1 if the Jth record is more
informative than the I th record,
= 9 if the Ith and Jth records are imcomparable.

INPUT N
INPUT ARRAY PREFER

**Determine array NUM

*NUM (I) is the number of records I such that the Ith record T(I) is more informative than T.

*Q1 is an index that determines if NUM has to be recalculated because of use] revisions to preferences

```
Ql=I
REPEAT TO LABEL (2) UNTIL Q1 = 0 Q1 =0
ARRAY NUM = ZERO
FOR I = 1 TO N
        FOR J = I TO N
        FOR J=I T O N
IF PREFER (I, J) = 1 THEN NUM (I) = NUM (I) + 1
        NEXT J
NEXT I
```

**Determine pairs for which NUM is not an ordinal function
*NON contains the non-transitive triples if NUM is not ordinal
*M is the number of such non-transitive triples

```
M=0
FOR J= l TO N
        FOR K=1 TO N
                IF {PREFER (J,K) = 1 AND N(J) > N(K)}
                OR {PREFER (J,K) = -1 AND N(K) > N(J)}
                OR {PREFER (J,K) = 0 AND N(J) = N(K)}
                OR {PREFER (J,K) = 9
                THEN
                        GO TO LABEL (1)
                ELSE
                        FOR H= 1 TO N
                        IF {PREFER (J, K) = 11 AND J.< > H AND K < > H
                        AND {PREFER (J, H) = -1 OR PREFER (J, H) = 0}
                        AND {PREFER (H, K) = -1 OR PREFER (H, K) = 0} THEN
                        M = M + 1
                        NON (M, 1) = J : NON (M, 2) = K : NON (M, 3) = H
                        END IF
                        NEXT H
                END IF
LABEL (1) NEXT K
NEXT J
**Revise preferences to satisfy negative transitivity
IF M > 0
THEN
        Q 1 = 1
        PRINT "PREFERENCES FOR THE FOLLOWING RECORDS ARE NOT
        CONSISTENT. PLEASE REVISE"
        FOR J = I TO M
                FOR K = 1 TO 3
```

```
            A = NON (J, K)
            S = K M O D ( 3 ) +  1
            B NON (J, S)
            PRINT A, 13, PREFER (A, B)
        NEXT K
            FOR    K= I TO 3
                A = NON (J, K)
                S = K MOD (3) + 1
                B = NON (J, S)
                PRINT ‘ ‘ PREFERENCE FOR : ` ` ; A; ` ` VS. ‘ ‘; B
                INPUT PREFER (A, B)
                IF PREFER (A, B) = 1 THEN PREFER (B, A) = -1
                IF PREFER (A, B) -1 THEN PREFER (B, A) = 1
                IF PREFER (A, B) = 0 THEN PREFER (B, A) = 0
                IF PREFER (A, B) 9 THEN PREFER (B, A) = 9
            NEXT K
        NEXT J
END IF
LABEL (2) CONTINUE
**p Calculate the array OUTT
*The array OUTT (I, J) is calculated as follows
```

OUTT (I, J) = 1 if the I-th and J-th records are comparable and have yet to be assigned to a comparable set, 0 otherwise.

```
FORI= 1 TON
        FOR    J = 1 TO N
            IF PREFER (I, J) = I THEN OUTT (I, J) = 1 : NUM (I) = NUM(I)+1 IF
            PREFER (I, J) = 0 THEN OUTT (I, J) = 1
            IF PREFER (I, J) = -1 THEN OUTT (I, J) = 1
            IF PREFER (I, J) = 9 THEN OUTT (I, J) = 0
        NEXT J
NEXT I
**Sort NUM in descending order.
ID(I) is the original number of the Ith sorted record
FORK = 1 TO N
        ID(K) = K
NEXT K
FORK = N TO 2 STEP -1
        F O R J = K - 1  TO 1 STEP -1
        IF NUM (1+1) > NUM (J)
        THEN
        TEMP 1 = NUM (J) : TEMP 2 = ID(J)
        NUM (I) = NUM (J+1) : ID (J) = ID (J + 1)
         NUM (J+1) = TEMP I ID (J+1) = TEMP 2

            AND IF
        NEXT J
NEXT K
**Determine comparable sets
*The matrix COMPARE will be calculated as follows :
```

```
        COMPARE (I, J) = 1 if the Jth record is in the Ith comparable set,
                       = 0 if the Jth record is not in the Ith comparable set.
*P is the number of the comparable set currently being created.
*Determine the first comparable set, containing the record ID(1).
P = 1
K1 = ID(l)
COMPARE (P, K) = 1
OUTT (K1, K1) = 0
F O R J = 2 T O N
                K2 = ID(J)
                IF PREFER (K1, K2) = 9
                THEN
                        TEST = 0
                ELSE
                        TEST = 1
                        FOR K = 1 TO J - 1
                        K3 = ID(K)
                        IF COMPARE (P, K3) = 1 AND PREFER (K3, K2) =9 THEN
                        TEST = 0
                NEXT K
        END IF
IF TEST = 1
THEN
                COMPARE (P, K2) = 1 FORK= 1 T O J
                K3 = ID(K)
                IF COMPARE (P, K3) = 1 THEN OUTT (K3, K2) = 0 :
                OUTT (K2, K3) 0
        NEXT K
ELSE
        COMPARE (P, K2) = 0
END IF
NEXT J
**Iteratively determine if all pairs have been assigned to a comparable set. If not, create
another comparable set.

F O R I = 1 T O N
        Kl = ID(I)
        FOR J= 1 TO I
                K2 = ID(J)
        IF OUTT (K1, K2) = 1
        THEN
                P = P + 1
        COMPARE (P, ICI) 1 : COMPARE (P, K2) = 1
        OUTT (K1, K2) = 0 : OUTT (K2, K1) = 0
        FOR H = 1 TO N
                K3 = ID(H)
                IF PREFER (K1, K3) = 9 OR PREFER (K2, K3) = 9 THEN
                        TEST = 0
                ELSE
        TEST = I
        F O R K = I T O H - 1
```

```
        IF COMPARE (P, K) = 1 AND PREFER (K3, H) = 9
        THEN TEST = 0
        NEXT K
END IF
IF TEST = 1
THEN
        COMPARE (P, K3) = 1
        F O R K = 1 T O J
                K4 = ID (K)
                IF COMPARE (P, K4) = 1 THEN OUTT (K3, K4) =0
OUTT (K4, K3) = 0
                        NEXT K
                ELSE
                                COMPARE (P, K3) = 0
                        END IF
                NEXT H
                END IF
        NEXT J
NEXT I
**Determine chains
*The output matrix EDGE is defined as follows:

EDGE (I, J) = K, K>0, if the Jth record is in the Kth edge of the Ith chain,
            = 0 if the Jth record is not in the Ith chain.
FOR I = I T O P
*Redetermine array NUM for the Ith chain
        F O R J =  1 TON
        NUM(J) = 0
        IF COMPARE (I, J) = 0
        THEN
                NUM (J) = -1
ELSE
        FORK= 1 TON
                IF COMPARE (I, K) = 1 AND PREFER (*J, K)=1
                THEN
                        NUM(J) = NUM(J) + 1
                END IF
        NEXT K
        END IF
        NEXT J
*Sort array NUM in descending order
*ID(I) is the original number of the Ith sorted record
        FOR K = N TO 2 STEP -1
        FOR J = K -1 STEP -1
        IF NUM (J+1) > NUM(J)
        THEN
                TEMP I = NUM(J) : TEMP2 = ID (J)
                NUM(J) = NUM(J+1) : ID(J) = ID(J+1)
                NUM(J+l) = TEMPI : ID(J+l) = TEMP2
END IF
NEXT J
```

```
NEXT K
*Assign records to edges in Ith chain based on value of NUM
NUM(N+1) = -1
R = 0
FOR J = N TO 1 STEP -1
        IF NUM(J) NUM(J+l) THEN R = R+l
        EDGE {I, ID(J)} = R
NEXT J
NEXT I
END
```

1.11 : Summary

Measurement of information is not yet directly possible. This is because there is no suitable working definition of information and there is no effective device for measurement. Informetrics is the subject of studying quantitative aspects of informational processes. Therefore, it is essential to have clear understanding of the measurement aspect of information and informational processes for studying informetrics.

Keeping this in view we have in this Unit re-introduced the idea of information from a number of points of views. We have shown that every information system can be considered to have a K-structure (abstract knowledge or information structure - which may be considered as a sort of hardware and software capable to receive, understand, output information in an information exchange process). Based on the status of information residing on K-structure, the system is in a K-state at a particular time (analogous to active `memory'). K-structure is fixed; K-state may go on changing.

If the information is `input' from outside to the K-structure, K-state. may or may not change depending on the `informativeness' of the `input' and the K-state. In case of output from K-structure, K-state remains the same (i.e., no loss of memory or knowledge). Through this model we can consider measure of information in a message through atomic information units which are semantic i.e., have unique elementary quantum of meaning (they may be key concepts in a discourse).

We have also made it explicit that informetrics, bibliometrics, scientometrics and® librametrics may be considered as different components (they are not independent of each other) of a compound term `scientometrics and informetrics' or a composite term used by some authors as `sciento-informetrics'. We have also noted that the most crucial part in sciento-informetric studies is the determination of `information units'. Bookstein has shown that precise definitions of concepts relating to such units are lacking as yet. The nature of informetrics is that of an early stage of a developing science.

We have then presented definitions of `measurement' in the context of general theory of measurement. Any measurement is assigning a `number' or `quantified' tag to a parameter based on a unit of measurement. Any measurement is an energetic process, it interferes (through the measuring device) with the system on which measurement is done, and is to be expressed as a `datum'. It is made with appropriate expectation of the outcome in terms of its `meaning' and just not `value'. A scale for measurement is essential. There are various types of scales. They are nominal scales, ordinal scales, interval scales, difference scales, ratio scales, and absolute scales. There are measurement techniques. Much of measurement depends on the technique used. Classification and counting are two simple universal means of measurement. Measurements need to be standardized in terms of accuracy, stability, reliability and validity.

In the last section we have discussed nature of information objects and their informational contents, and then defined the idea of `informativeness'. Informativeness is very useful for measuring effectiveness of information services. Following Jean Tague-Sutcliffe we have enumerated seven properties of informativeness. In short they constitute the following: Informativeness is a non-negative number that is not usually additive and cannot be measured directly. As informativeness relates to an information object or document or record the total informativeness remains the same even if informativeness of parts are taken separately. Informativeness can be service satisfies the information needs of its user community (clientele). Informativeness measures can be used for collection development, document description (evaluative indexing), retrieval process and repackaging or consolidation of information (indeed for content management or knowledge management).

Unit-4:Sociology of Science and Scientometrics

4.0 : Objectives

After reading the Unit, you will be able to explain:

- what is sociology of science and the domain it covers;
- different approaches to sociology of science;
- the norms that controls science as a social activity;
- major contributors to the development of sociology of science;
- sociology of science within sciences;
- how the growth of knowledge occurs and different approaches to study this growth;
- modelling of the growth of knowledge and the criteria for selecting a model;
- formation of social circles in S&T research;
- different indicators of social organization in research;
- interaction between cognitive growth and social organization ;
- scientometrics approach to science;
- formation of scientometric maps and their different types;
- citation practices in scientific publications; and
- networks in research.

4.1 : Introduction

This Unit discusses sociology of science and scientometrics. This is a new field in

the area of science earlier it was dominated by Philosophy of Science and the History of Science. This new emerging area has four approaches namely; sociology of knowledge, normative approach, identification of characteristics of scientific knowledge and the fourth approach deals with the way scientist use the scientific literature. Further it also explains the role of sociology of scientometrics within the discipline of science. For studying the growth of knowledge, this Unit discusses some of the approaches and models, etc. This Unit further explains the role of social organisations in research areas.

In the area of scientometrics the technique developed by some eminent scholars such as Solla Price, Eugene Garfiled and others have been discussed. After reading this Unit you will have a clear idea of new area of sociology of science and its role within the discipline of science.

4.2 : Sociology of Science

The sociology of science is defined as the part of sociology that studies relationships between different idea systems and a variety of institutional and personality factors. It is a relatively new field in the sphere of sciences, which has traditionally been dominated by the philosophy of science and the history of science. It deals with the social relativism of scientific research and its knowledge products. It informs that scientific research is a process, which involves social interests and the major social forces that prevail upon the working of a scientist.

4.2.1 : Domain of Sociology of Science

1. The sociology of science constructs a set of highly generalized, systematic and relatively exhaustive concepts and propositions of relationships. In this venture, it uses data from all historical periods and all cultures.

2. The sociology of science encompasses the fundamental scientific ideas and their application to technology. In its study of technology, again, the sociology of science uses historical as well as contemporary data, drawn from different cultures.

4.2.2 : Approaches to the Sociology of Science

The sociological study of science is a relatively newer development. The first studies appeared in 1940s but the subject got a firm footing in the late 1950s, with the appearance of a large number of papers dealing with different aspects of sociology of science. Several approaches towards the field have emerged as described below:

1. The first approach is known as "sociology of knowledge" approach. It studies the relation between thought and society. It depicts the social or existential conditions of knowledge. This approach has been primarily concerned with the way in which ideas are influenced by the position of intellectuals in the social structure of the society. This approach has provided an orientation for the sociology of science, but its theoretical insights have not (with the exception of Weber's work) been able to exert a strong influence.

2. The second approach is known as "normative approach". It is related with the study of science as a social system. Science is regarded not as a body of knowledge or as a set of methods and techniques but as "the organized social activity of men and women who are concerned with extending man's body of empirical knowledge through the application of these techniques. The relationships among these people, guided by a set of shared norms, constitute the social characteristics of science"(Storer, 1966). It is also concerned with investigating the inter-relationships between science and other social institutions. It studies the influence of other institutions on the origins and development of science and *vice versa.* (Please note

that`institution' here is used in a broader sense and encompasses social and cultural concepts, a cluster of scientific ideas, a single or a group of scientists, etc.)

This approach is also concerned with the understanding of the values and norm of science. It views the scientific community as a single entity. It postulates very general norms without considering the variations that may occur in different specialties and sub-specialties and in different time periods.

R. K. Merton made the maximum contribution to the development of sociology o science, after its emergence in 1940s. He may be regarded as the "Father o Sociology of Science", since he contributed the major ideas that relate communication to social processes. He was the one who stated how a scientist scores in the game of science and the social rules that make this highly competitive, intricate activity possible. He has also explored the response of science as a social system subject to various historical stresses. He has suggested the ways in which the game evolves. According to him, the most important social processes and structures are those that are designed to protect awards, cognitive achievement and originality.

Merton presented his ideas in a series of papers which are listed in Table 1, along with other major contributions. In his first paper on normative structure of science, Merton was stimulated by the extreme nationalism at the beginnings of the Word War II, and he stated that the basic rules of science as a social activity are controlled by following four norms: "Universalism", "Communism", "Disinterestedness", and "Organized skepticism". The `conununism' asks scientists to share their findings, and the institution promises "returns" only on "property" that is given away. The `universalise:' enjoin scientists to evaluate knowledge claims using "pre-established impersonal criteria" so that the contributor's race, gender, nationality, social class, should not affect the allocation of rewards and resources. The norm of `disinterestedness' channels the motivations of a scientist away from his self-interests that could conflict with the institutional goal of science (extending certified knowledge). The `organized skepticism' prescribes dogmatic acceptance of claims. It urges suspension of judgment unti l sufficient evidences and arguments are available.

Merton has confirmed the theory of cumulative nature of science, and has suggested that it is the recognition through citations received which is one of the principal rewards in science.

Table 1: Major Contributors to the Development of Sociology of Science in Different Periods

Period	Researcher	Year	Title of Communication
Before 1955	Merton, R.K.	1942	Normative Structure of Science
1955-1965	Merton, R.K.	1957	Priorities in Scientific Discoveries
	Merton, R.K.	1961	Singletons and Multiples in Science
	Harber, R.	1961	Resistance to Discovery
	Hagstrom, W.O.	1965	The Scientific Community
After 1965	Merton, R.K.	1968	Mathew Effect in Science
	Crane, D	1972	Invisible Colleges
	Small, H & Griffnth,B.C	1973	The Structure of Scientific Literature - I
	Griffith, B.et al	1974	The Structure of Scientific Literature-II

These norms presented an idealized conception of science. However, the element of "motivation" was largely ignored, and the level of deviance was considered to be very low.

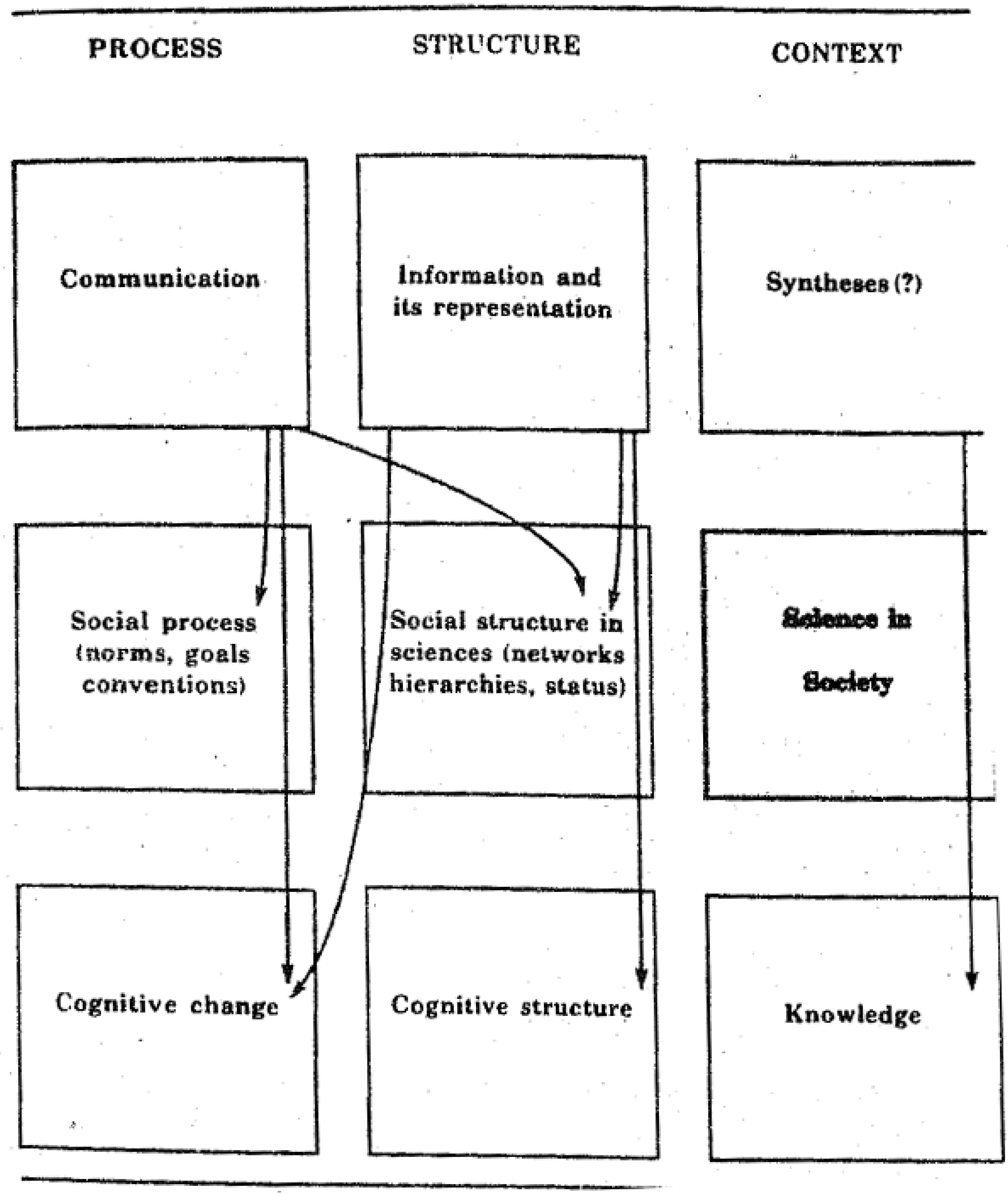
The existence of deviance from scientific norms was recognized by Hagstrom (1965). In his view, scientific behaviour is an exchange of information for recognition. Information is a gift that a scientist gives to other scientists, and in return he expects recognition. In making this gift, the scientist conforms to scientific standards concerning the production of scientific work. However, if a scientist wants to receive recognition for his work, he must produce answers to the questions that are of interest to other scientists. This sociological interpretation of the behavior of individual scientists led Hagstrom to consider the nature of scientific disciplines so as to understand how scientists select problems for their study. He has argued that the major source of deviant behaviour in science is competition for recognition, and that it is possible to differentiate between disciplines in terms of likelihood that competition will result.

1. The third approach in sociology of science is concerned with the identification of the characteristics of scientific knowledge that bring about change and development in scientific ideas, rather than identifying external influences upon this process. This approach is identified with a few historians of science whose work had considerable influence upon the sociological thinking about science (Holton, 1962; Kuhn, 1962 and Toulmin, 1963). For example, Kuhn (1962) has proposed a model of scientific change that has stirred considerable controversy. Briefly, he argues that growth and development of scientific knowledge occur as a result of the developments of paradigm or model of scientific achievement that set guidelines for research. After a period of normal "science", during which the implications of the paradigm are explored, the facts that the paradigms cannot explain become inescapable. The field then goes through a period of "crisis" during which a new paradigm is proposed and eventually accepted.

2. The last approach examines the way scientists use the scientific literature, particularly the extent to which they cite recent literature compared to older sources, and literature in their own fields compared to other fields. An analysis of the structure of literature shows that each new piece of scientific literature is related to several previous pieces, which, in turn are related to still earlier works (Price, 1965a). The scientific literature grows very rapidly. But, the "life" of an individual paper is very short, with the exceptions of few classics. The papers generally become "old" or obsolete in a period of five years. Papers published more than 15 year ago, become almost useless in scientific fields (Burton and Kebler, 1960); since the research "front" moves too far ahead.

4.2.3 : Sociology of Science within Sciences

The role of sociology of science within sciences has been depicted. by Grit th, shown in Figure 1. It contains nine boxes; the top three indicate behaviours nd artifacts used to analyze social and cognitive events in the bottom six boxes.



The first box shows that `Communication' is the only general scientific behaviour; other behaviours are specific or technical. Then, `Information and its representation', in the second box are its principal artifacts. The `scholarly synthesis', in the third box, is its final integration and mechanism for conservation of a scientific achievement. These three are the principal data sources for understanding the social and cognitive processes in science.

4.3 : Growth of Scientific Knowledge

The study of the effect of scientific communities upon the growth of knowledge shows that social interaction plays an important role in the scientific growth. If this were so, one would like to find which models of intellectual change are compatible with these findings? And then, whether these also fit in the pattern of growth of publications and new authors in the research areas? In this section, different approaches for studying the growth of knowledge discussed and models generally used for this study are briefly described. Then, the criteria for selecting a model are outlined and lastly, the life cycle of a research specialty is described.

4.3.1 : Different Approaches for Studying the Growth of Kn

For studying the growth of knowledge, a largo number of approaches are available, such as sociological approach, epidemiological approach, self-organizing systems approach, biological evolution approach, but we shall restrict our discussion only to the sociological approach which is reflected in the growth of knowledge through the publications and the authors in a scientific field.

Price (1961, 1963) studied the growth of knowledge in few scientific specialties, as reflected through publications. He reported that the growth of publications as well as authors in a scientific specialty passes through the following stages:

a) A preliminary period of growth in which absolute increments are small, and the rate of increase although is large, but steadily decreases;

b) A period of exponential growth, when the number of publications in a field doubles at regular intervals as a result of constant rate of growth that produces increasing absolute growth;

c) A period during which the rate of growth declines, but the annual increments remain almost constant; and finally

d) A period during which both the rates of increase and absolute increase decline and eventually approach the zero level.

This means that growth of publications and new authors follow the logistic growth curve. But the Price could not offer any reasonable explanation for this trend. Later Crane (1972) suggested that the logistic growth pattern in the growth of scientific knowledge was perhaps the result of the exploitation of intellectual innovations by a particular type of scientific community. The question according to Crane to be examined then was, as to how the scientific community affects the growth of scientific knowledge? According to her, social interaction among the members o scientific community plays a major role in the growth of scientific knowledge an(publications. When scientific literature in scientific specialties exhibits a pattern o; logistic growth, some kind of interaction between social and cognitive factors take: place, which plays a significant role in the growth of knowledge. This growth, it fact, is a kind of diffusion process in which ideas are transmitted from one scholar to another. Thus, logistic growth of knowledge can be interpreted as a "contagion" process, in which early adopters (new authors) influence later adopters (authors), which, in turn, create a logistic growth in the cumulative number of publications and authors entering a research area.

4.3.2 : Models of Growth of Knowledge

The fact that the growth of scientific knowledge and publications in scientific specialties follows a logistic growth curve suggests that there are changes in the social relationships among scientists during the process of its growth. This raises the question of the relationship between the social changes and the process of intellectual development in scientific specialties. Are they independent? The various phases of growth of scientific knowledge in scientific specialties have been conceptualized in the following three growth models:

i. **First Model:** In this model, the growth of knowledge is viewed as a cumulative progression of new ideas, developing from antecedent ideas in a logical consequence. A hypothesis derived from theory is tested against empirical evidences and it is either accepted or rejected;

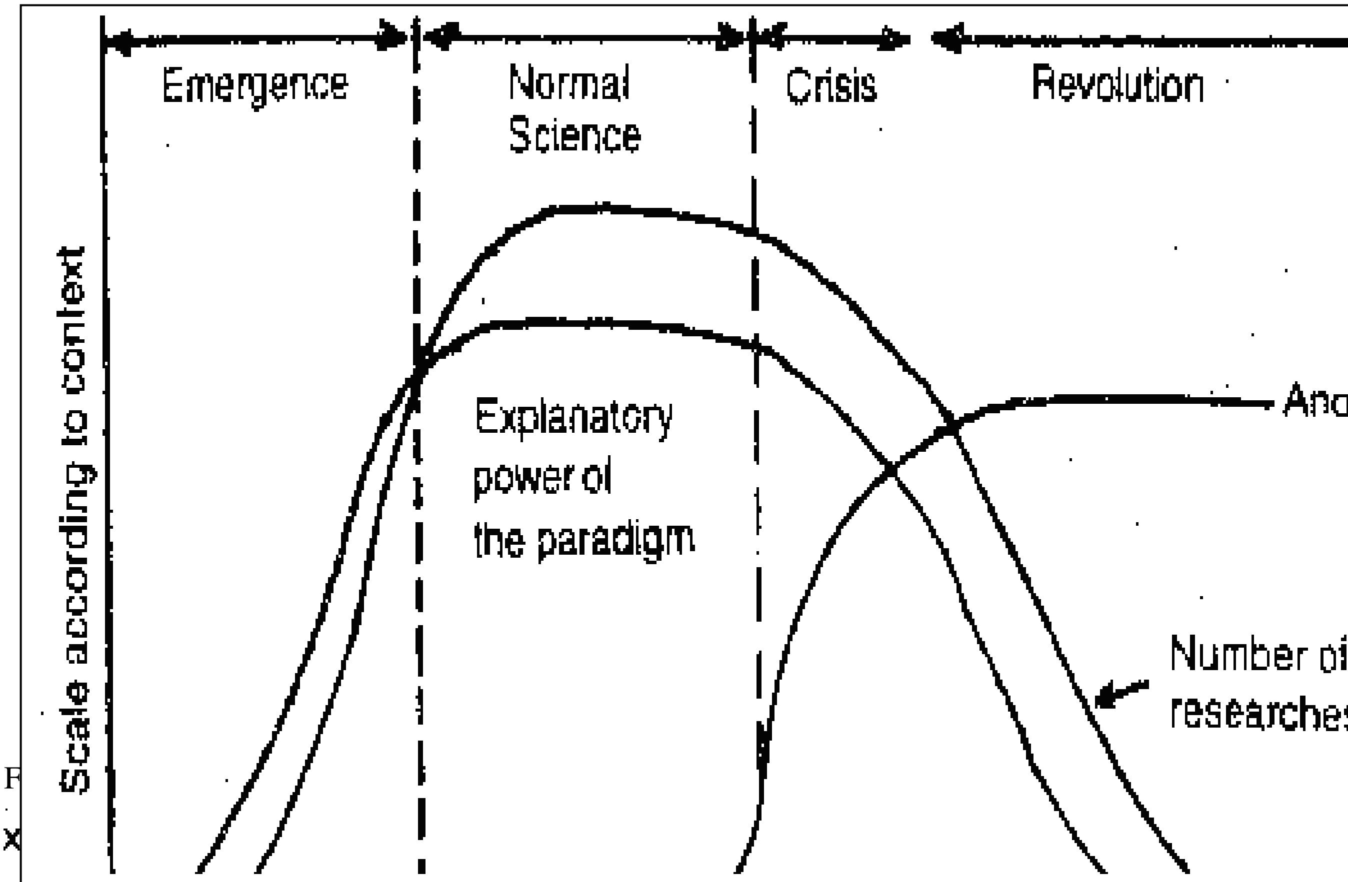
ii. **Second Model:** In this model, it is assumed that in the growth of knowledge new ideas originate not from the most recent developments, but from any previous development in the history of scientific field. In this model, there is a kind of random selection across the entire history of the cultural area; and

iii. **Third Model:** In this model, the growth of knowledge is viewed as a cumulated growth, intercepted with periods of discontinuity.

There have been many debates on the third model. Different scholars have interpreted this model in different ways, as described below:

One version of the third model can be seen in the Kuhn's analysis of scientific change (1962) that combines periods of cumulative growth, which he calls normal science, with the periods of crisis or revolutions. During the period of normal science, scholars obtain results within the framework of a given paradigm. The basic thrust of the scientific enquiry is essentially on the nature of "puzzle-solving". The inability of a given paradigm to solve certain problems and chance discoveries lead to creation of anomalies. As the explanatory power of the paradigm gets weakened, anomalies accumulate, leading to the situation of crisis. Efforts are then made to. look for solutions outside the framework of the paradigm, resulting in the multiplicity of hypothesis. Eventually, a new paradigm is discovered, which has a greater explanatory potential and the crisis is diffused. If the new paradigm provides greater explanatory power, it increasingly draws scholars from other research schools and networks associated with particular problem areas. The growth rate gradually declines when the explanatory power of the paradigm reaches a saturation level. A new chapter in the history of scientific revolution begins once again to repeat the phases of normal science, crisis and revolution. The four phases of the evolution of scientific knowledge are shown in Figures 2 and 3.

The second version of the third model as suggested by Kroeber (1957), indicates that the concentration of interest on a set of problems among scientists leads to a kind of crescendo of activities. The importance of results obtained steadily increases and then gradually tapers off, as the potentialities of the methods being used to solve the problems are exhausted or as the problems get solved. The process is repeated when the new orientation towards these problems is discovered or when an entirely new set of problems. is defined. The information generated by previous research is reinterpreted, rather than being completely discarded.



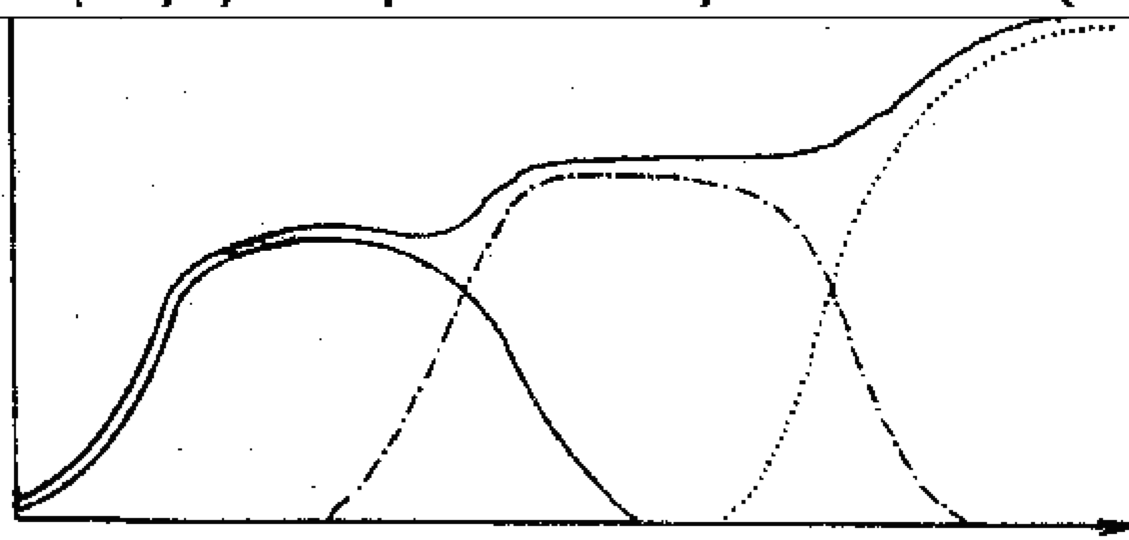


Fig.3: Dynamic model of growth of scientific knowledge: Emergence of successive paradigms
(luring the evolution of scientific knowledge

The third version of this model was provided by Toulmin (1963), who proposed a process of continuous change or "micro-revolution" that results in the "selected perpetuation of preferred intellectual variants". Toulmin shares Kuhn's views that there are periods in the development of science when scientific knowledge does not cumulate. He calls them "recurrent periods of self-doubts", during which scientists tend to question whether science can explain any thing. These crises are resolved by changes in the underlying assumptions of the scientific specialty, in contrast to Kuhn's views of discarding the old model. Toulmin argues that an old theory may survive in a new scientific specialty at a later date, as new facts are discovered.

The fourth version of this model can be found in Holton's perception of scientific change (1962), which combines both Krober's and Toulmin's interpretations. According to Holton, science as a whole grows because individual scientific specialties grow. Each is eventually exhausted, but new ones are continuously emerging due to the discovery of linkages or connections among the old scientific specialties. Like Toulmin, he sees the birth and death of scientific specialties in terms of biological evolutionary analogy, of a struggle for survival among ideas. Certain central concepts, utilized in a variety of fields, provide continuity, but the multiplicity of efforts coupled with essential democracy of scientific institutions. This kind of growth model includes the notion of a special kind of cognitive event that plays an important role in bringing about scientific change.

Many scholars such as Toulmin, Holton, and Mulkay have given the critique of Kuhn's concept of paradigm, which suggests that there are in fact two facets of cognitive events that stimulate the growth of scientific knowledge: (a) the paradigm as a way of seeing a perspective or pattern, and (b) the paradigm as a special kind of tool or problem solving devise.

4.3.3 : Selection of Models for Growth of Knowledge

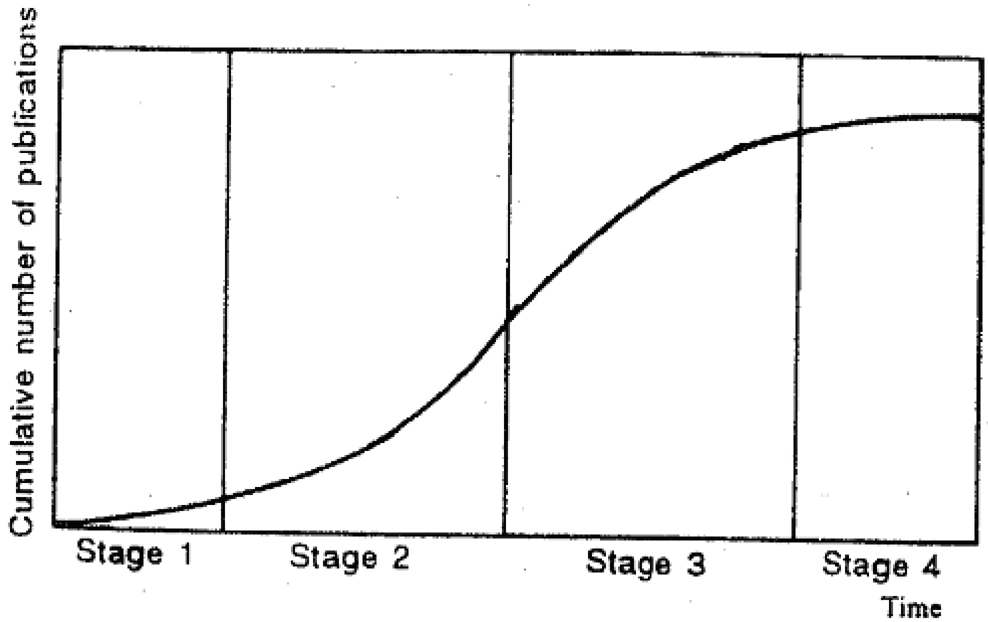
Many studies on the cumulative growth of publications and authors in scientific specialties have depicted that their growth pattern fits in some of the growth models leading to an S-shaped curve. The different parts of S-shaped curve represent different phases of the development of a scientific specialty. The first model of the growth of scientific knowledge does not imply that scientific specialties pass through different phases of growth. The second model suggests that the new developments in scientific specialties emerge from previous developments. The growth implied in the first model does not fit the pattern of the growth of publications and authors. The model also assumes an irregular growth in all stages and at all times and implies very little social organization in scientific specialties. The third model, which describes the periods of growth, followed by the periods of discontinuities or decline, fits the observed data on the cumulative growth of publications and authors of many studies. This model indicates the presence of social interaction among participating scientists, but does not imply a high level of solidarity among them, since conflict and controversy are at the minimum. It is during the period of controversy that solidarity among opposing groups of scholars is likely to appear, or alternately, during a period when new specialty is developed.

4.3.4 : Life Cycle of Scientific Specialties

Several attempts have been made by scholars to trace paradigms in terms of development model of scientific specialties. Some of them distinguish three or four stages of growth within the paradigm life cycles of scientific specialties:

a) Goffman has divided the life cycle into three stages: Stage 1: Insufficient and unordered information; Stage 2: Insufficient, but ordered information; and Stage 3: Sufficient and ordered information.

b) Crane has considered the different stages of knowledge development as coupling of four stages: Stage 1: When the paradigm appears, there is no developed social organization; Stage 2: When normal science flourishes, invisible colleges appear; Stage 3: With the solving of major problems and turning of anomalies, social splitting appears; and Stage 4: With the exhaustion of paradigm, the number of participants decreases (See Figure 4).



		1	stages	3	4
Characteristics of	Knowledge		Normal Science	Solution of major prob-lems	Exhausion
	Knowledge			Anomalies appear	Crisis
	Scientific Communities	Little or no social orga-niztion	Groups of collaborators and an invi-sible college	Increasing specialization	Decline in membership
	Scientific Communities			Increasing controversy	Decline in membership

Fig.4: Characteristics of scientific communities and knowledge development in four stages

a) Mulkay *et* al have characterized the development of scientific specialty in the following sequence of stages: Stage 1: Exploration; Stage 2: Unification; and Stage 3: Decline or displacement.

b) Mullins has provided four stages-model, (i) normal stage, (ii) network stage, (iii) cluster stage, and (iv) specialty stage.

However, all these models attempt to cover the whole life cycle of scientific specialties, i.e. from its birth to death. Figure 5 provides an augmented table of stage characteristics superimposed upon Crane segmentation of the logistic growth curve for scientific specialties. This figure combines features collected from various stage models.

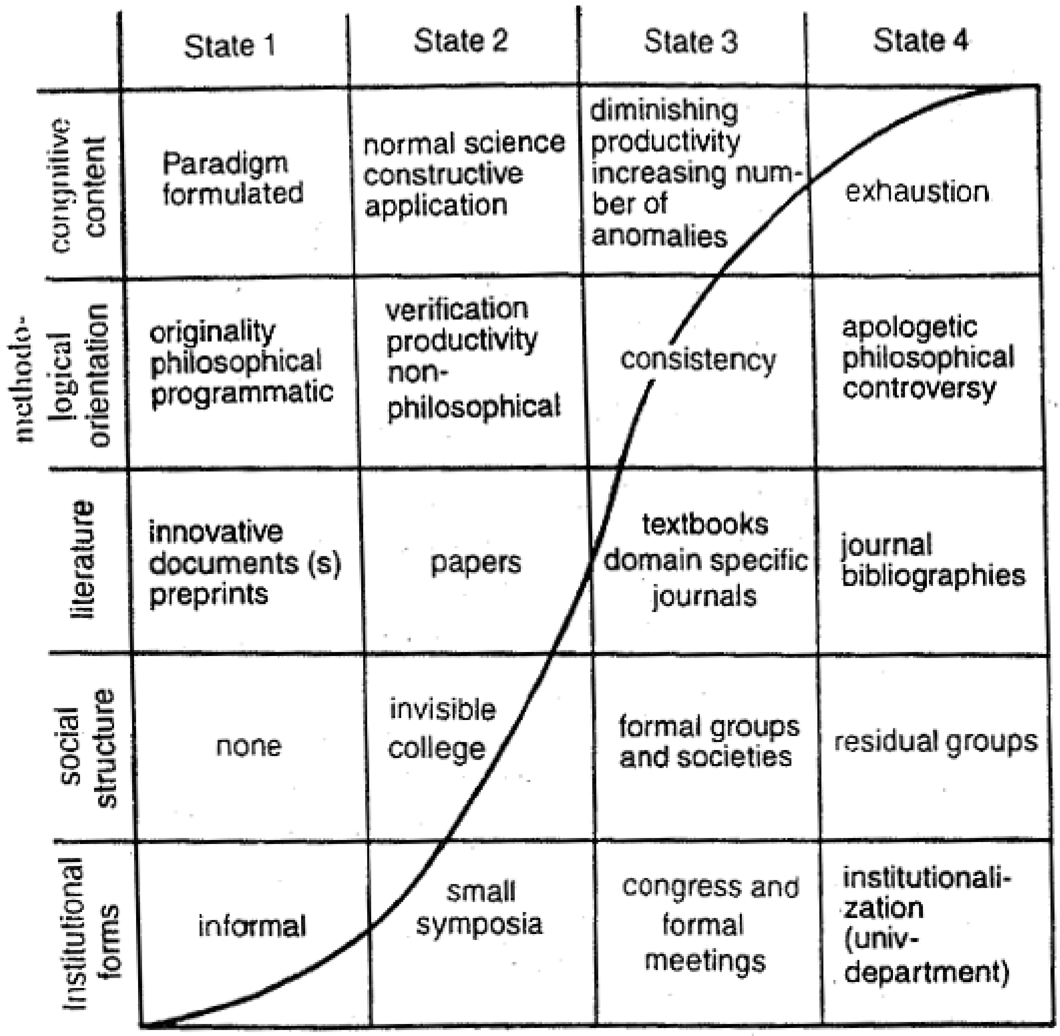


Fig.5: Characteristics of life cycle of scientific specialties in relation to the various stages superimposed upon a Logistic growth curve

4.4 : Social Organization in Research Areas

If a social organization does exist in a research area, then most of the members should be related to others in at least one way. Indirect ties are as important as direct ties, for a scientist who communicates with another scientist and in this way may obtain information that is transmitted to the second scientist. Allen (1969) has argued that it is unlikely that indirect ties between scientists reflect the actual flow of information. He has pointed out that the number of intermediaries separating the two individuals is inversely related to the probability of information being exchanged by them in this manner. It is important to realize, however, that a research area is not entirely a Nee-to-face group. Its members are widely dispersed geographically. Many exchange ideas through correspondence or publications. Third parties play important role in transmitting information between persons who may never actually meet. Examination of citations appearing in the publications of one of the areas made it possible to validate the existence of influence transmitted through intermediaries.

4.4.1 : Indicators of Social Organization in Research

The scientists in research areas can have different types of social relationships with one another. To measure the social organization, a number of indicators are used, some of which arc:

i. Teacher-Student Relationship (Thesis Directors) - This is the most common relationship between a teacher and his research student. Both collaborate in doing research and publications. Even otherwise also, the teacher who trains a student often retains a close relationship with him in later years. A student is also influenced by his teacher and usually adopts the same field of research and maintains linkages with his teacher.

ii. Collaborations in Publications - Many scientists collaborate with other scientists of the same institute, in some other institute or even overseas institute (international collaboration). They bring out joint publications/reports.

iii. Informal Communications - Some scientists discuss their research problems with their colleagues in order to obtain advice as well as suggestions on their work/techniques.

iv. Personality Influence - A number of scientists are influenced by some specific scientists who could be higher ups in hierarchy, their role models or even fellow scientists influence a number of scientists. They select problems and methodologies under such an influence.

v. Attraction for Fashionable Research - The last category of scientists are those who select a problem because it is a `fashion' to work in that area. The examples of pollution and environment could be cited under this category. These scientists have an eye on quick recognition and early reward.

From each of these types of linkages among scientists, one can obtain a picture of the extent to which members of a research area are linked to one another.

4.4.2 : The Research Area as a Social Circle

A social circle in a research is characterized by the presence of direct or indirect linkages among a majority but not necessarily among all of its members. The extent to which members of an area are linked to one another can be measured by the following two relationships:

· By examining the proportion of relationships that actually occurred in relation to the one that could have occurred. In other words, if every scientist were linked to every other scientist in a social circle, than 100 percent of the possible linkages would exist. In practice, however, this percentage would always be less than 100.

· By compiling socio-metric network showing how many of the scientists in the group are linked to one another directly or through a common linkage with other individuals.

A common method of developing a social circle in a research area is the formation of scientific societies or associations. A few examples of such societies are: Society of Environmental Scientists of India, Society of Soil Scientists of India (SSSI), American Association for the Advancement of Science (AAAS), and Science Writers' Association of India.

To foster links among the members, these societies and associations organize their annual meetings. The membership is usually enlarged through personal contacts.

Another way of developing links in a social circle is the organization of symposium/ conference or workshop in a research area. A few strong members of a social circle initiate this activity and bring the majority of members of a research area to a common platform where they exchange views and present their findings. This activity is sometimes undertaken by the concerned society/association also. The opportunity is also utilized to attract newer members to the circle.

The failure to find evidence of social organization in research areas can be interpreted in the following manner:

- It may mean that the categories that were chosen to designate a research area are not the ones that are currently being used to label their works.

- It depicts that the researches are under consideration has not yet experienced the rapid growth that is associated with the development of social organization.

- It could also mean that the research area enjoyed a good growth in the past but was now declining, as it was no longer a `hot' area.

When social contacts among researchers in a discipline are lacking, social organization is unlikely to emerge in the research areas within the discipline (Barton and Wilder 1964).

4.4.3 : Interaction between Cognitive Growth and Social Or

In a developing research area, a lot of questions emerge that need to be answered and a bundle of facts accumulate to test the theory. The area, therefore, attracts a large number of researchers and a social organization emerges. Therefore, during the exponential growth of a research area, a process of social interaction is reflected in which contact between scientists contributes to the cumulative growth of knowledge. In other words, we may say that there is an interaction between cognitive events and social events in the development of a research area that parallels the stages of the logistic curve.

During the period of initial growth of a research area, interesting cognitive event occurs and attracts new scientists to the area that has been sparsely populated. This is followed by a period of cumulative development of knowledge in the area, during which the number of publications and of new authors entering the field grows exponentially.

The shift to exponential growth is marked by the appearance of new groups of scientists within the research area. These groups can be described from two points of view:

i). Structural Aspects: who is linked to whom by socio-metric ties; and

ii). Normative Aspect's: the types of attitudes and behaviour that are expected of their members.

From the structural point of view, two distinct subgroups emerge. One type consists of groups of collaborators, where a collaborative effort is made to generate knowledge and speed up the rate of communication of research findings.

The second type consists of groups that are linked to one another through their leaders who communicate with each other and transmit information informally across the whole field. This enables them to monitor the rapidly changing research "front" and to keep up informed about the newer developments during the period of exponential growth. This second type of subgroup in a research area is a communication network that links groups of collaborators. It is also known by a commonly used terminology of "Invisible Colleges". As the leader here plays the role of a teacher by communicating information and the remaining members of the group function as students by receiving information.

A number of sub-groups of different sizes may be formed. Some constitute equally active members and are usually not so large in size. Some are larger in size and may contain a few productive scientists and many relatively unproductive ones.

These subgroups or invisible colleges are more commonly formed in developing countries where the leader of the group has links with the international leaders in other countries. The leader makes frequent visits to these places and procures information through correspondence. Other members of a subgroup do not have linkages with the leader or members of other subgroups and depend on their own leader for the needed information.

Analysis of the networks has shown that anyone choosing even one of the most productive members of the research areas means contact development with a large network of individuals. In other words, a high proportion of choices directed towards these individuals means that members of these groups are not so much linked to each other directly but are linked to each other indirectly through these highly influential members. It was suggested that most productive scientists are the leaders of groups of researchers. Their contacts with each other link these groups into a single network. This process has been shown in a number of ways. Crane (1972) has utilized this concept of groups of scientists linked to one another directly or indirectly by published collaborations or student-teacher ties.

It has also been observed that leaders of the large group of collaborators recruit and socialize members and create a sense of commitment to a research area. As these research areas expand in size, the influence of members of the area as compared to scientists who had not published in the area increases. It has been possible to observe the emergence of certain individuals who contribute numerous publications to the area and who exert considerable influence upon its development. Most of these scientists publish their first papers in their fields either before or during the period of exponential growth. These individuals maintain their interest in these areas for considerable periods. Thus, the highly productive scientists are able to have substantial impact upon their research areas in terms of setting the norms-of research and maintaining the cohesion in the area. Thus, during the period of exponential growth, the highly productive scientists play major roles in the direction of theses and selection of research problems. Groups of collaborators that do not have aggressive leadership are unlikely to endure. Differences in the type of communication that occurs may be an important aspect of the research environment provided by large groups of collaborators. It also appears that standards for evaluating the quality of research set by members of a large group of collaborators and that member of small groups do not conform to them. Are those who do conform indeed getting admitted to those groups? This does not seem to be the case, since few members of large groups begin their careers as isolates. Publications by pairs of isolates that are cited frequently are usually published early in the history of the area when standards for research in the area are being formed. Later members learn these standards by socializing from the beginning of their careers in large group of collaborators.

4.5 : Approaches of Scientometrics to Sociology of Scien

Under scientometrics approach to science and technology, a series of new techniques have been developed by scholars such as Solla Price, Eugene Garfield, Henry Small, Belvith Griffith, R. Merton, Diana Crane, N. Mullins, S. Cole, J. Cole, H. Zukerman, W. Mulkay.

This scientomctric approach is based on quantitative characteristics and attributes or objects of documentary flows. It mainly utilizes the bibliographic data on publications, which comprise such attributes as title of paper (including titles and phrases), authors (including their affiliations, co-authors and reputation), the title of book, journal (title, subject and origin of country, and language), etc. The objects of study in scientometrics studies are often grouped under various attributes.

A scientific publication is a written document containing information in respect of scientific activities - either in physical form (hard copy) or electronic form. A scientific paper reveals not only the word-building ability of its author, but also the nature and force of the building blocks derived from the domain of science from which it draws and to which it contributes.

Each publication uniquely contains its own bibliographical information. In addition, each distinct type of publication has its own set of common characteristics and attributes. Two basic units of bibliographic information are generally used:

a) **Items:** These refer to information in or about publications. The principal items of analysis are authors' name, affiliations, citations and keywords, as well as selected words from the title, abstract, or the full text. Some abstracting and indexing services often incorporate additional bibliographical information such as the type of publication (research paper, review, book, etc.), language (English, German, French, etc.), one or more controlled terms (keywords given by data producers), and subject classification terms or codes, and

b) **Entities:** These refer to aggregates of publications. These entities represent a set of publications, starting from the micro-level (a single author) to meso-levels (varying in size from university departments, scientific journals to research specialties) and then to macro-level (scientific sub-fields or countries). The information required for bibliometric analysis is derived from primary journals, indexing and abstracting journals, and databases.

There are two common approaches to quantification of information flows: (a) the first is built directly on the work of Price and his predecessors. It follows the dynamics of separate items or objects. These publications as a whole or its attributes such as authors' names, affiliations, citations and keywords, as well as selected words from title, abstract, and in text, journal title, its country and subject, etc. and (b) the second approach consists in identification of links between objects, their correlation and classification.

4.5.1 : Different Approaches to Quantification of Informat

First Approach

The first, based directly on the work of Price and his predecessors. It follows the dynamics of separate items or objects. These publications as a whole or its attributes such as author's names, affiliations, citations and keywords, as well as selected words from title, abstract, and in text, journal title, its country and subject, etc.

In this approach, different pictures of dynamics of publications are obtained as regards to the state of science and technology. For example, one can model the growth of a scientific field and the direction it is likely to follow, by measuring its annual publications data. Similarly, if we choose keywords as the object of study, then alternations in keyword frequencies can reflect both some substantial regrouping of research directions in science. Similarly, a study of the distribution of productivity of authors and of articles in journals in different time periods can be studied in terms of such indicators, which can throw light on the level of development in the scientific specialties. The use of citation data and their age distribution over different time periods can also reveal the level of obsolescence and the structural changes taking place in the field from time to time.

Second Approach

The second approach consists in identification of links between objects, their correlation and classification. Under this approach, the analytical scientometric procedures are not directed at obtaining characteristics, but to gain statistics relating to quantitative indicators of information files and also to identify with their help quantitative structures of science and its individual subject fields, by mapping them. The concept "structure" implies information on relations among constituting elements. Scientometrics has a variety of information at its disposal to study the relational aspects of study. The examples under this category are: the authors in co-authored publications, their affiliations, countries, keywords from the text, and the citation list appended to it. Each attribute reflects a different aspect of science. For example, the addresses can be used for assessment of collaboration among nations, while citation relations may indicate intellectual links between the set of publications.

4.5.2 : Importance of Citations and Age Distribution

Different forms of citation analysis comprise an important component of scientometric analysis. In scientometrics, citation serves as an object of investigation, the reliability of which is determined by the tradition of science as a social institution. The phenomenon of citation represents an important ethical norm in science, a general regulator, and a means of communication.

Merton has suggested that it is the recognition through citations received which is one of the principal rewards in science. He describes the distinctive and seemingly paradoxical circumstances that in science the more freely a scientist gives his intellectual property away, the more surely he is to win. Apart from cumulative nature, Kaplan has viewed reference/citation as a distinct entity in a social system of science. He has also proposed that citation should be viewed as a social control mechanism and publications as a property. He states that citation is an important institutional device for copying with problem of individual copyrights with respect of recognition and priority. While Kaplan has brought the concept of individual copyrights, Ravetz has introduced the idea of intellectual rights based on the interpretation that the publications and citations jointly bring reward and recognition.

Gilbert has proposed an alternative theory of citing, which considers scientific publication as a "tool of persuasion". He has stated that a scientist who has obtained results, which he believes, arc true and important, has to persuade the scientific community to share his opinions on the value of his work. Brooks, who studied the motivations of academic authors in giving references, also has confirmed this view. Seven motivational scales have been identified (currency, negative credit, operational information, persuasiveness, positive credit, reader alert and social consensus). Among these persuasiveness has been regarded as the major motivator. Citation due to social consensus or negative credit is very weak. Persuasive communities refer to the scientists whose works are cited in a particular corpus of scientific writings. An author refers to other works that will support knowledge claims thus gathering an environment of supporting opinion. Moravcsik and Murugesan (1975) have found that negative references are quite rare in science and this supports the idea that references serve to create communities of supporting opinion. Once a corpus of formal communications that defines a scientific community has been identified, one can aggregate together the references made in that corpus to form the picture of the persuasive community. This community is the environment surrounding a scientific community that has persuasive power for the members of that community (Allen, 1994). On the basis of the above discussion, we can say that the references/citations have been used for various purposes in the past, such as scholarly bricklaying, property right protection, priority claims, and as a persuasive tool.

Using citation data, Cole and Cole have studied the social stratification in science. They regarded the number of references to an author as a rough measure of his recognition and importance of his scientific work. A high citation frequency of publication may serve as a marker of innovation (a new idea, method, and discovery). Some scientists regard the group of papers frequently cited within a specialty, as a quiet carrier of its paradigm. Kuhn possibly had this concept in mind when he wrote that changes in citation pattern of literature, as found in research publications, could be accepted as a symptom of scientific revolutions.

Citation practices vary from one subject area to another. These also depend on the nature of work under consideration. The recency of citations appearing in scientific publications depends upon the rate of growth as well as its level of development. In a fast growing scientific specialty, the cited references are mainly of recent publications, in slow growing specialties, the proportion of references cited to older publications is significantly higher. The rate of obsolescence is clearly related to the "degree of paradigmaticity of a specialty" On studying the citation decay curve, it is observed that: (a) second part of this citation decay curve, which accounts for 70 per cent of all cited publications bears some resemblance to the literature growth curve and (b) the first part of citation decay curve, comprising 30 per cent of highly cited references to recent publications, and distribution of these citations is defined by shape curve. Price has made a distinction between archival literature and research front literature based on this classification. The archival literature, constituting a substantial of all the citations made, can be compared to the body of science, while the research front literature, though relatively small in size, is referred to as the growing skin. According to this image, science not only grows like a tree, in the sense of branching off in twigs, but it also resembles a tree in growing only on the outside being supported by a lifeless stable structure from within. The archival body of literature is rarely used in the construction of new ideas in a scientific specialty, the research front literature is, as far as can be determined from the recent publication cited, and is most heavily used in that respect.

4.5.3 : Mapping of Science

Scientometric maps can be built for exploring. the underlying structure of similarities and interrelationships between items and / or entities. Price had initiated work on dynamic mapping of data using Science Citation Index in early 1960s. Studies on different levels (i.e. word, articles and journals) employing different methodologies (cluster analysis, factorial analysis, graph analysis and multidimensional scaling) have been used to map the dynamic structure of science. Techniques related to coword/co-classification, co-citation/bibliographical coupling and journal-to-journal citation network have been used to generate cartography of science. They are primarily intended to show the practical as well as more cognitive values. Since the number of variables used in these studies is multivariate, suitable dimensionality-reduction techniques are required to simply what might first appear to be a complex pattern of associations among many entities. Through these spatial representations, one is able to observe the linkages among research areas/sub-fields and time series depiction that shows the dynamics in the research fields. The intertwining and the complexity of research fields can thus be understood (to some extent) through these maps. These have been employed to identify science maps, maps of disciplines, research fronts, scientific journal networks, epistemic and conceptual networks, invisible colleges and author networks.

The following types of scientometric maps are frequently used:

1) Journal-to-journal citation network maps - Journals are the most important communication channels of science. Each journal represents a distinct domain pf science, covering one or more topic areas. The current state and developments to science is partially reflected in the emergence and disappearance of journals av well as changes in the strength of linkages within a journal citation network. Eacp article in a particular journal links itself to the main research theme of that journa . Similarly, an article when it cites another articles in another journal links at the macro level the themes of two journals together. Higher the citation density implies stronger the linkage among the said journals, and vice versa. Thus, the citation network among journals provides us with a macro level overview of science;

2) Co-classification network - By adopting the degree of co-occurrence of subject classification headings as the organizing principle, content analysis opens up the whole area of a particular subject discipline to a quantitative study of inter-disciplinarity. When aggregated over a set of publications, the frequency in which headings are jointly used can be operationalised to function as a direct quantitative proxy of the degree of interdisciplinarity between the respective fields. Moreover, the

internal structure of a subject discipline can be expressed quantitatively via the network of co-occurrences between all pairs of headings. Although Price started work on this type of methodology in the 1960's, Peters et al (1996) is probably the first well documented quantitative analysis of such , co-occurrence data in the science studies context;

3) Bibliographical coupling - Bibliographical coupling developed by Kessler (1963), is the number of common references cited in two documents, and indicate the degree of similarity of contents of the cited papers;

4) Co-citation network -- The concept of co-citation was introduced by Small in 1973 to generate clusters of related papers. The number of times the two papers are cited together in subsequent literature determines the co-citation strength of the two cited papers. Small and Griffith (1974) and others have used this concept in studies on structure of science and scientific literature. Using this technique one can map out the relationships between scientists in various recognized specialties of science, discover sub-specialties, follow the time development of new ideas, etc. The co-citation methodology created a realistic picture of differentiation in science and the speed and extensiveness of change. This discovery of a bibliographic information structure that parallels social and intellectual structure is of major importance. The changes in this structure depict integration in the groupings 'and the overall changes in the scientific community. By furnishing a method of grouping documents and persons, the co-citation produced a technique for the analysis of scientific behaviour and for the development of policy analysis; and

5) Co-authorships networks - Collaboration, as evinced in the frequency of co-authorship among the set of authors, has been used as a method of mapping linkages among authors in different fields. Groups of scientists who frequently communicate and exchange information are also found to be collaborating among themselves. Such **studies** have increasingly played an important role in the identification of invisible colleges in major fields. Studies also depict the level of productivity (measured by rank) at which authors prefer to collaborate among themselves. According to Kretschmer (1999), scientists with a particular rank order keep less distance from members with the same rank than from members with different ranks. It means that 'relative frequency of contacts (as reflected through co-authorship) between partners having the same rank is higher than between partners with different ranks, also interpreted as "Birds of same feather flock together". The behaviour of authors in choosing partners in collaboration is similar to classification of friendship by occupational status studied by Mardson, where preference was given to insiders against outsiders. Kretschmer found that the structure of co-authorship networks was on the line with the social psychological theory. She has used multivariate statistics to depict the entire pictu'e of co-authorship relationships through two- and three-dimensional graphic presentations. Price and Beaver(1966) found that productivity per author increased as the number of collaborators increased. They noticed the existence of a small core group of active authors, surrounded by large floating population of authors, who collaborated with leaders on only one or two projects and then disappeared. In this study they have found that out of 555 authors **in** 1966 high producers with 14 **or** more publications were those who had 12 or more collaborators. Those authors who worked alone or with one collaborator not able to produce more than 4 publications in the same period.

4.6 : Summary

In this Unit, the newly emerged field of sociology of science has been introduced and its relation with scientometrics has been explained through following points:

- The definition of sociology of science has been provided and its domain has been spelt out;
- Different approaches to the sociology of science have been outlined;
- Four basic rules that control science as a social activity enumerated are: Universalism, Communism, Disinterestedness, and Organized skepticism;
- Major contributions to the development of sociology of science in different periods arc provided;
- Sociology of science within science has been depicted;
- Different approaches for studying the growth of knowledge have been discussed;
- Three growth models have been described for studying the growth of knowledge;
- Different versions of the "Cumulated growth model" are discussed;
- Selection of the model for the growth of knowledge is discussed;
- Five indicators of social organizations in research described are: Teacher-student relationship, Collaborations in publications, Informal communication, Personality influence, and Attraction towards fashionable research;
- The linkages between the members of a research area as a social circle are outlined;
- The interaction between cognitive growth and social organization is discussed;
- The formation of invisible college is described;
- Different approaches of scientometrics to sociology of science are discussed; and
- Different maps used for scientometric studies described are: Journal-to-Journal network maps, Co-classification networks, Bibliographical coupling, Co-citation network and Co-authorship network..