DOCUMENT RESUME

22 000 002	LI 003 526
AUTHOR TITLE	Nance, Richard E.; And Others Information Networks: Definitions and Message
INSTITUTION	Southern Methodist Univ., Dallas, Tex. Computer
REPORT NO PUB DATE NOTE	TR-CP-710011 Jul 71 29p.;(9 References)
EDRS PRICE DESCRIPTORS	MF-\$0.65 HC-\$3.29 *Definitions; *Electronic Data Processing; *Information Networks, *Madalas Electronic

IDENTIFIERS *Information Transfer

ABSTRACT

ED 060 862

A mathematical definition of an information network is constructed with the purpose of developing a theory useful in answering practical questions concerning information transfer. An information network includes: (1) users, (2) information resources, (3) information centers, and (4) the total information transfer structure linking (1), (2), and (3). Emphasis is placed on the message transfer structure, as distinguished from the document transfer structure, to identify some basic network configurations. Any message transfer structure is shown in graph theory concepts to be either isographic or nonisographic. Among the isographic structures, the cyclic and decentralized networks are defined. The strictly hierarchical network is also defined, and the 2-regular network, reflected in the ARPA design, is identified. Measures of network structure, in particular the accessibility and flexibility in message transfer, are developed. These measures for the basic structures are used to characterize more general structures. While some comparisons of message transfer structures can be made, development of more comprehensive measures is a necessity. (Author)



U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE ÖFFICE OF EDUCATION THIS DOCUMENT HAS BEEN REPRO-DUCED EXACTLY AS RECEIVED FROM THE PERSON OF OFIGANIZATION ORIG-INATING IT, POINTS OF VIEW OR OPIN-IONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDU-CATION POSITION OR POLICY.

۰.

L U	
9	
$\boldsymbol{\omega}$	
0	
9	
0	

5

 ~ 1

INFORMATION NETWORKS: DEFINITIONS AND MESSAGE TRANSFER MODELS (Technical Report CP-710011)

> Richard E. Nance Robert R. Korfhage and U. Narayan Bhat

Computer Science/Operations Research Center Institute of Technology Southern Methodist University Dallas, Texas

July, 1971

ABSTRACT

A mathematical definition of an information network is constructed with the purpose of developing a theory useful in answering practical questions concerning information transfer. An information network includes: (1) users, (2) information resources, (3) information centers, and (4) the total information transfer structure linking (1), (2), and (3). Emphasis is placed on the message transfer structure, as distinguished from the document transfer structure, to identify some basic network configurations.

Any message transfer structure is shown in graph theory concepts to be either isographic or nonisographic. Among the isographic structures, the cyclic and decentralized networks are defined. The strictly hierarchical network is also defined, and the 2-regular network, reflected in the ARPA design, is identified.

Measures of network structure, in particular the accessibility and flexibility in message transfer, are developed. These measures for the basic structures are used to characterize more general structures. While some comparisons of message transfer structures can be made, development of more comprehensive measures is a necessity.

ERIC

I. INTRODUCTION

Contribution of mathematical models

The concept of <u>information network</u> evokes markedly different responses in this period of information deluge and computer panaceas. While librarians have sought for years to understand, organize and control this essential commodity--information, a latent uneasiness, perhaps even distrust, surfaces at the mention of "information network". <u>Library networks</u> seems a more comfortable term, divorced from the esoteric hardware structures implied by "communication networks", "computer networks", and "information networks".

Some maintain that information must be transferred by electronic means in order for an information network to exist. R. C. Swank [1] comments on the conditions imposed by several definitions of "information network" and offers his own definition in terms of the characteristics: (1) information resources, (2) users; (3) intellectual organization of documents or data, (4) methods for delivery of resources, (5) formal organization and (6) bidirectional communications networks. While definition in terms of these characteristics provides a basis for discussion, little can be said definitively about networks in general, the relation of one network's characteristics to another, and the similarities among networks.

We offer a mathematical definition of information network and compare our definition with Swank's. Our definition has the disadvantage, at least viewed by some, of a higher level of abstraction. We accept the abstraction in order to gain rigor that leads to definitive statements about network structures and differences and similarities among networks.



Information network concepts

We do not attempt to cite the extensive body of literature on information networks. An excellent bibliography can be found in the paper by Samuelson [2]. Recent attempts to organize the concept of library or information networks generally have followed one of two approaches. The first approach is to define a network in terms of its functional organization. Swank [1] provides an excellent description using this approach. The second approach is to dwell on the structure for information transfer and be less concerned with the functions served by this transfer. Nance [3] offers an example of this approach.

A few authors have followed both approaches to some degree. Duggan [4, pp. 159-160] describes the "twelve critical components...essential to orderly, planned development of the objectives (of a library network)". She then cites several network configurations; thus recognizing the different possible structures by which information transfer can be accomplished. Duggan suggests six structural forms, and we show these in Figure 1. She also presents the number of "channel links" (C) required by each structure.

Davis [5] uses both approaches in her description of the National Biomedical Communications Network and a design procedure for networks in general. With respect to the structure for information transfer, she identifies four types of network organization, which are shown in Figure 2. The centralized and decentralized structures identified by Davis correspond to the directed and non-directed forms of Duggan; moreover, the composite centralized structure is analogous to Duggan's representation of the interface of two directed networks.

The fourth structure offered by Davis, the hierarchical network, is of

4

-2-



-



(a) Totally Decentralized Network





(c) Composite Centralized Network

Figure 2. Network Classifications Proposed by Davis [5]



considerable interest. She notes that while the control system complexity increases with the hierarchical sturcture, more flexibility in intercommunications and greater reliability of the network are gained [5, p. 35].

In a different context, Kleinrock [6] has used different terminology to identify the identical structures in communication networks. The starnet configuration [6, p. 28] corresponds to the centralized structure shown by Davis, and the fully connectured net [6, p. 101], to Davis' decentralized configuration.

Duggan, Davis and Kleinrock have used the theory of directed graphs, although not explicitly, to enable them to identify different structures. Nance [3] explicitly employs a graph-theoretic approach to model a library network. No one has sought to examine the relationships among the structures in a comprehensive and definitive manner.

To answer questions concerning which structure is best for accomplishing certain objectives, we must first be capable of defining the structures. Our effort is to develop a unifying concept of the term "information network", and from this concept to construct a methodology by which structures for information transfer can be evaluated and compared.

~5~

II. A UNIFYING CONCEPT OF NETWORK STRUCTURES

We define an information network as a set

 $N = \{U, I, C, A, f, f'\}$

where

- U is the set of users (either potential or actual) of information resources (i ϵI);
- I is the set of information resources that are accessed by users (u $\in U$);
- C is the set of information centers and with each $c\in C$ is associated a subset U $\leq U$ and a subset I $\leq I$, either one of which, but not both, may be empty; and
- A is the set of all edges on $U \lor I \lor C$, where an edge from node v_1 to v_2 denotes that v_2 is directly accessible from v_1 , and each edge $(v_1, v_2) \in A \mid C$ (the edges of C) carries one or both of the labels:

m indicating that message transfer, or b indicating that document transfer

can be accomplished from v_1 to v_2 . The distinction between message and document transfer is made in subsequent paragraphs.

Four subgraphs of interest are:

1. $N_{II} = \langle UvC, arcs between U and C \rangle$ and

 $\forall u \in U$ there exists at least one edge (u,c) for $c \in C$,

2. $N_T = \langle IvC \rangle$, arcs between I and C> and

Vi ε I there exists at least one edge (i,c) for $c \varepsilon C$,

3. $G = \langle C, \text{ arcs with label } m \text{ joining nodes in } C \rangle$, and

4. G' = $\langle C$, arcs with label b joining nodes in C > -

Two conditions that are imposed on the relationships among U, I and C:

1. the subgraph $\langle Uv I, A | U_{VI} \rangle$ is totally disconnected, <u>i.e.</u>, no direct access is permitted between users, information resources, or users and information resources; and

2. the subgraphs of G and G' are strongly connected.

The functions f and f' constitute the information transfer structure

-6-

(i.t.s.) for the network. To define these functions we use $P_{\mathbf{m}}(C) = \{\text{all open paths in } G\}$ and $P_{\mathbf{b}}(C) = \{\text{all open paths in } G'\}^{\frac{1}{2}}$. For an individual path $P \in P_{\mathbf{b}}(C)$ or $P \in P_{\mathbf{b}}(C)$ let $V(P) = \{\text{nodes } c \in C | c \in P\}$. Then

- f: $P_m \neq 2^{U \times I}$ by 1. if there exists no edge (u,c) with label m, ccV(P), then $f(P) = \phi$;
 - 2. otherwise, f(P) = I for $I = \{i \in I | \text{there exists an edge (i,c) with label m, <math>c \in V(P)\}.$

and

- f': $P_b \rightarrow 2^{I \times U}$ by
- 1. if there exists no edge (u,c) with label b, $c \in V(P)$, then $f'(P) = \phi$;
- 2. otherwise, f'(P) = I for $I = \{i \in I | there exists an edge (i,c) with label b, c \in V(P) \}.$

Note that the condition of strong connectivity placed on G and G' assures that a user in contact with any information center has access, through message and document transfer, to any other center.

One might characterize U and I in various ways, but we believe that this definition of information network enables a mathematical description that captures the intent of the defining characteristics of Swank [1].

The set of users U is considered finite but not necessarily static. Identification of a set member $u \in U$ is accomplished partly by noting the <u>information center</u> utilized by u to place a particular request for information. An information center is defined as an entity where a request for information may be made by users ($U \subseteq U$) and/or information resources ($I \subseteq I$) are located. Thus our description of the user necessarily involves the one or more infor-





An open path is an alternating sequence of distinct nodes and edges such that each directed edge is adjacent from the node preceding it and to the node following it.

mation centers through which he accesses the information network.

Information resources (I) are considered representable by a set of discrete entities, each having a unique identifier. Individual members $i \varepsilon I$ are documents with certain intrinsic labels, including identification, characteristics, information centers where copies of the document are stored, etc. In addition, extrinsic labels including user identification \hat{u} , user information center l, etc., are attached to a copy of the document indicating a transfer of that copy of the document to center l in response to a request by user \hat{u} .

Within an information network two types of information transfer take place [3, p. 59]:

- (1) message transfer (accomplished by f)--information necessary to gain access to the information resources (I), and
- (2) document transfer (accomplished by f!)-supply of the information resources (I) to members of U, the set of users.

10

One can view the first transfer as involving <u>unlabeled</u> information, <u>i.e.</u>, no final destination is prescribed for the transfer; while the latter concerns <u>labeled</u> information, <u>i.e.</u>, associated with the transfer of each item of information is *I* is a final destination, or receiver, us *U*. According to our definition f is the structure imposed to enable <u>access</u>, the <u>message</u> <u>transfer structure</u>, and f' is the structure created to effect <u>delivery</u> or <u>response</u>, the <u>document transfer structure</u>.

-8-

• 2-9

III. STRUCTURAL CLASSIFICATION OF INFORMATION NETWORKS

The graph G is the characteristic used by other authors to classify information networks. For example, Duggan's [4] directed network (see Figure 1(b)) prescribes a specific form of G. By considering the form of G, we can define each specific message transfer structure suggested by Duggan [4] and Davis [5]. Additionally we can characterize structures that are not classified immediately as belonging to any structure class, <u>e.g.</u>, that shown in Figure 3.

~9-

We state the following definitions for information networks, in all cases assuming N = |C| > 1:

- An information network N with N centers is <u>cyclic</u> if and only if id(v_i) = od(v_i) = 1 ¥ v_iεC.
- 2. An information network N with N centers is <u>decentralized</u> if and only if

$$id(v_i) = od(v_i) = N-1 \quad \forall v_i \in C.^2$$

The cyclic and decentralized networks define opposite extremes in the degree of connectivity among strongly connected digraphs. Both of these networks have message transfer structures (G) which are specific examples of an isograph [8, p. 330]. We can refer to the cylcic network as a 1-regular isograph and the decentralized network as an (N-1)-regular isograph. A third isographic network structure is introduced below.

 An information network N is strictly heirarchical if the graph obtained by replacing all 2-cyles in G by an undirected edge is an undirected tree.

In general any information network can be characterized as <u>isographic</u> or <u>non-isographic</u> depending on the message transfer structure (m.t.s.) G.

The notation $id(v_i)$ and $od(v_i)$ refer to the inward and outward degrees of node v_i respectively, <u>i.e.</u>, the number of edges of inward and outward incidence respectively.





Figure 3. An Arbitrary Message Transfer Structure G* Represented by a Digraph

-10-

•

.

.....

The m.t.s. for cyclic and decentralized information networks with 6 information centers are shown in Figure 4. A strictly hierarchical network m.t.s. is pictured in Figure 5(a). The resulting undirected tree for the m.t.s., shown in 5(b), is obtained by replacing each directed twocycle by a single undirected edge. Note that Figure 3 shows an isographic network that corresponds to none of the three above.

Various other m.t.s. can be proposed. Of particular interest is the 2-regular isographic network shown in Figure 6. This m.t.s. is found in the ARPA network [9].

Structural Measures

Having defined the basic structures, we recognize that any particular information network may offer a message transfer structure that agrees exactly with none of the above. The benefits from defining these basic structures are: (1) precise definitions assure no ambiguity in the concepts, (2) advantages of one structure over another in a particular situation may be revealed, and (3) any network structure may be discussed by comparison with the basic ones. The second claim is explored in a following paper; while the third requires the development of measures comparing any particular structure to the basic forms.

<u>Theorem 1</u>. For a network with N nodes, the minimum (q^0) and maximum (q^1) number of edges are given by

and

13

 $q^{0} = N$ $q^{1} = N(N-1)$

-11-



(a) Cyclic Network for Six Information Centers



(b) Decentralized Network for Six Information Centers

Figure 4. Graphs of the Cyclic and Decentralized Networks for Six Information Centers



.

•



Figure 5. Illustration of the Message Transfer Structures in Strictly Hierarchical Networks

15

.

.



-13--



Figure 6. The 2-Degree Isographic Network for Eight Information Centers (ARPA)

.



. ..



Proof: (Minimum) By assumption, G is a strongly connected digraph. Thus each node must have at least two edges (one inward incident to the other) joining it to G. However, by the definition of an information network, access must be provided to each node and each node must be able to reach any other node.

... $id(v_i) \ge 1$ and $od(v_i \ge 1) \quad \forall i = 1, 2, ..., N$ The minimum occurs obviously when

 $id(v_i) = od(v_i) = 1 \quad \forall i = 1, 2, ..., N.$

This requirement can be met with N nodes only if G is a directed cycle of N edges.

(Maximum) To obtain the maximum number of edges, each node should have an outward degree of N-1. With N nodes, the total number of edges is N(N-1).

Flexibility

We define

$$q(G) = \sum_{i=1}^{N} [id(v_i) + od(v_i)]/2 = \sum_{i=1}^{N} id(v_i) = \sum_{i=1}^{N} od(v_i)$$

and note that q counts the number of edges in a m.t.s. (G) for an information network with N nodes. For the cyclic network $q = q^c = q^0 = N$.

An obvious consequence of the cyclic network structure is that for each node $v_i \in G$, where a message is initiated or referred, there exists one, and only one, node to which the message may be directly referred. This represents the most restrictive structure. The decentralized structure obviously is the least restrictive. Since the lack of restriction, or conversely the presence of flexibility, in originating or referring messages seems intuitively a desirable feature, we label the cyclic network as



0-flexible and the decentralized network as 1-flexible.

The measure

$$Z(N,q) = \frac{q-q^0}{q^1-q^0} = \frac{q-N}{N(N-2)}$$

is proposed to indicate the degree of flexibility in a network m.t.s. with q edges and N nodes. Since

 $z(N,q^0) = 0$ for the cyclic network containing N nodes, and $z(N,q^1) = 1$ for the decentralized network,

we refer to any network as <u>z-flexible</u>, thus reflecting a proportionate degree of flexibility. Several network structures are shown in Figure 7 to illustrate the measure z(N,q).

The isographic networks permit a more general computation of flexibility.

Theorem 2: For a k-regular isographic network with N centers

$$z(N,q) = \frac{k-1}{N-2}$$

<u>Proof</u>: This value results from the fact that there exist exactly kN edges. Obviously the flexibility for a k-degree isographic network is bounded by the flexibility value for the k-regular isographic network, <u>i.e.</u>, for any k-degree isographic network with N centers

$$z(N,q) \leq \frac{k-1}{N-2}$$

For a general network (not necessarily isographic), we define the degree δ by

$$\delta = \max \{ id(v_i), od(v_i) \}$$

i=1,2,...,N

A second characteristic is that for a fixed degree (δ), the network with fewer nodes has greater flexibility.



-18-

One recognizes the difficulty in calculating the measure of flexibility z(N,q) for as N becomes large and many edges are present, computing the value of q by examining the graph becomes quite difficult. Fortunately, the value of q may be determined simply using the adjacency matrix representation of the graph. More on the adjacency matrix is given in Section IV.

20

and the second sec

IV. USING THE GRAPH-THEORETIC CONCEPTS

The diagrammatic representation of graphs proves useful in establishing basic definitions and developing an intuitive appreciation of contrasting information transfer structures. To explore characteristics of individual structures, however, we employ the matrix representation of a graph.

The adjacency matrix

Consider an arbitrary assignment of the integers 1,2,...,N to the N nodes of the graph G depicting the m.t.s. of a information network. We use the integers as subscripts and identify the nodes by v_1, v_2, \dots, v_n to maintain consistency with our previous notation. Let C be an NxN matrix with values c_{ij} where

 $c_{ij} = \begin{cases} 1 \text{ if edge } (v_i, v_j) \in G \\ 0 \text{ otherwise} \end{cases}$

The matrix C is termed the adjacency matrix of the labeled graph G.

Attaching to C the superscripts c and d again to indicate cyclic and decentralized d respectively, we picture a general matrix form for each of these structures in Figure 8 along with the corresponding graph. (In general, different numberings of the nodes yield different matrix representations.) Note that no node is shown to send messages to itself; therefore, in every matrix the main diagonal consists only of zeros. Also note that the matrix representation of G provides a computationally effective means of identifying the decentralized and cyclic structures as substructures in larger, more complicated networks.

Let us examine the matrix representation of G for a strictly hierarchical network shown in Figure 9(a). The matrix is symmetric, and treating only the upper triangular portion, we obtain the block structure illustrated in 9(b).

-19-













(a) Information access structure (G) for strictly hierarchical network



- (b) Matrix representation shows block form of strictly hierarchical structure
- Figure 9. Information Access Structures in a Strictly Hierarchical Network

23



-21-

-22-

Accessibility

An interesting and useful property of the adjacency matrix is noted in the theorem [8, p. 112] stated below:

Theorem. Let C be the adjacency matrix of a digraph H. The i,j entry in C^n specifies the number of walks in H of length n from v_i to v_j .

In more familiar terms, if we let c_{ij}^n be the i, j entry in C^n , then the value of c_{ij}^n represents the number of distinct message transfers involving n referrals that conclude with the message originating at v_i and terminating at v_j . The illustration in Figure 10 can be followed easily. From the original adjacency matrix, C^2 is calculated. The element $c_{11}^2 = 3$ specifies that a message originated at v_1 , referred twice (including the original referral from v_1), and terminated at v_1 can follow three distinct referral paths, <u>i.e.</u>, $v_1v_2v_1$, $v_1v_3v_1$, or $v_1v_4v_1$. Similarly, all other nonzero entries have the value 1, and a message originated at v_1 and terminated at v_2 (likewise for v_3 and v_4) must require an odd number of referrals. By similar examination of C^3 , C^4 , ..., we can conclude that

$$c_{ii}^{n} = 0 \text{ and } c_{jk}^{n} = 0 \quad i=1,2,..., 1 < j < k$$

 $n = 2h-1$

and

$$c_{1j}^{n} = c_{j1}^{n} = 0$$
 $1 < j$
 $n = 2h$
 $h = 1,2,...$

A node $v_j \in G$ is said to be accessible by $v_j \in G$ if there exists at least one directed path from v_j to v_j . Considering the adjacency matrix C this is equivalent to stating

v, is accessible by v, if and only if there exists n (a positive integer) n: C^n has the component $c_{ij}^n > 0$.



Figure 10. Powers of the Adjacency Matrix Reveal Characteristics of Referrals and Information Transfer

25



~23-

By definition, every node $v_j \in G$ is accessible by every node $v_i \in G$ if G represents the m.t.s. for an information network. An information network is defined to be ρ -accessible where ρ indicates the minimum number of referrals necessary to enable complete accessibility, <u>i.e.</u>, every node has access to every other node.

$$\rho = \min n : \Xi(n) = \sum_{j=1}^{n} C^{j}$$
 has

elements $\xi_{ij} > 0 \quad \forall i \neq j = 1, 2, \dots, N$.

Complete accessibility for any decentralized information network is accomplished by a single referral; however, the cyclic network with N centers is N-accessible.

V. SUMMARY

Working from a verbal definition of an information network, we offer a mathematical definition in terms of: (1) users, (2) information resources, (3) information centers, and (4) the total information transfer structure linking (1), (2) and (3). We concentrate on the message transfer structure, as distinguished from the document transfer structure, to identify some basic information networks. A more comprehensive classification should include the document transfer structure as well.

Based on graph theory concepts, any message transfer structure can be classified as isographic or nonisographic. Among the isographic structures, characteristics of the cyclic and decentralized networks are explored, and a flexibility measure for a general network is defined using these basic forms. The adjacency matrix representation provides a means of assessing the accessibility of the information centers in the network. The property of complete accessibility is shown to be markedly different for the cyclic and decentralized structures.

Future research is focused on extending the classification to include both document and message transfer structures. Methods of evaluation considering the information transfer structure and the information resources resident at each center are presently being investigated. The concepts developed here are being tested in applications to the public library access network (PLAN) and to existing university and medical library networks. The ARPA computer network offers yet another promising area of application.

-25-

REFERENCES

- Swank, R. C., "Interlibrary Cooperation, Interlibrary Communications, and Information Networks--Explanation and Definition," <u>Conference on</u> <u>Interlibrary Communications and Information Networks</u>, Warrenton, Virginia, Sept. 27-Oct. 2, 1970.
- Samuelson, Kjell, "World-Wide Infomation Networks," <u>Conference on Inter-</u> <u>library Communications and Information Networks</u>, Warrenton, Virginia, Sept. 27-Oct. 2, 1970.
- Nance, Richard E., "An Analytical Model of a Library Network," <u>J. ASIS</u>, <u>21</u>(1): Jan.-Feb. 1970, pp. 58-66.
- Duggan, Maryann, "Library Network Analysis and Planning (LIB-NAT)," J. Library Automation, 2(3): Sept. 1969, pp. 157-175.
- Davis, Ruth M., "The National Biomedical Communications Network as a Developing Structure," <u>Conference on Interlibrary Communications and</u> <u>Information Networks</u>, Warrenton, Virginia, Sept. 27-Oct. 2, 1970.
- 6. Kleinrock, Leonard, <u>Communication Nets:</u> Stochastic Message Flow and <u>Delay</u>, McGraw-Hill, 1964.
- 7. Harary, Frank, Graph Theory, Addison-Wesley, 1969.

.7. 14

- 8. Marary, Frank, R. Z. Norman, and Dorwin Cartwright, <u>Structural Modelst</u> An Introduction of the Theory of Directed Graphs, John Wiley, 1965
- Roberts, Lawrence G. and Barry D. Wessler, "Computer Network Development to Achieve Resource Sharing," <u>AFIPS Proceedings</u>, <u>36</u>: May 5-7, 1970, pp. 543-549.







ERIC JTER SCIENCE/ INTIONS RESEARCH CENTER SMU INSTITUTE OF TECHNOLOGY DALLAS, TEXAS, 75222