DISCUSSION PAPER NO. 33

ON MEASURING ECONOMIC INTERRELATEDNESS

by

J. M. Blin
Frederic Murphy

February 1973

Graduate School of Management
Northwestern University

ON MEASURING ECONOMIC INTERRELATEDNESS

ABSTRACT

This paper addresses a problem first proposed by C. S. Yan and E. Ames in their 1965 paper ("Economic Interrelatedness," Review of Economic Studies, Vol. XXXII, No. 4, pp. 299-310.) In their work these authors had suggested that an economic interrelatedness function be defined for input-output systems by taking some average of the number of positive technological relations existing between any two pairs of sectors. The measure thus defined was then used to assess the validity of various empirical statements made in economic history in connection with the hypotheses of integration vs. diversification in the process of economic growth. In this paper we provide a general theory of interrelatedness concepts in terms of oriented graphs and paths in graphs. After deriving some basic characterization results various classes of functions are suggested to measure economic interrelatedness. For instance instead of considering only minimal paths between pairs of sectors families of paths are introduced; also a natural way to take into account the varying linkage intensities between sectors is provided by considering a valued graph associated with input-output systems. This leads to the definition of a completely general interrelatedness function that can then be used to study such empirical concepts as integration and diversification in economic growth processes, as intended by Yan and Ames.

On Measuring
Economic Interrelatedness

I. The notion of an interrelatedness function

General equilibrium systems in economics draw their success from the fact that they explicitly recognize the interdependence between the decisions of each and every economic agent involved in a certain process of production and/or consumption. The recognition of this fact prompted Schumpeter to declare, some decades ago, that if he were forced to choose one single economist as having contributed the most fundamental insight in economic theory, he would pick Walras. One admittedly very special but yet very rich class of general equilibrium models is afforded by input-output systems. Their success in economic theory and planning theory has been widely acclaimed. Somewhat surprising, however, is the fact that relatively little work has been devoted to an actual study of the notion of interdependence between agents in the input-output framework. One noteworthy paper in this area dealt with this question: in their 1965 paper, Yan and Ames tackled the problem in the following way. Starting from the notion that for any pair of sectors (i,j) $(i,j=1,2,...,n)$ the corresponding entry a_{ij} in the input-output coefficient matrix could take on any value in the closed interval $[0,1]$ they pointed out that whenever $a_{ij} > 0$ for some (i,j) these two sectors are directly connected in the sense that (i) is a supplier of some input incorporated in the production process of (j). Moreover if there exists some finite positive integer K for which the expression

(1) $$a_{ij} = \sum a_{ir_1} a_{r_1 r_2} \cdots a_{r_k j}$$

To each entry b_{ij} of the B matrix assign the smallest value α^* (some positive integer between 1 and ∞) for which the entry $m_{ij}^{\alpha^*} = 1$. The rationale for this choice is best understood in graph-theoretic terms. In particular the following result is quite easily obtained (see for example [2] and [13]).

Theorem 1: Let G be a graph with adjacency matrix M. The α-fold boolean product of M : $M^{\alpha} = M \dot{\times} M \dot{\times} \dots \dot{\times} M$ represents the adjacency matrix of a derived graph G^{α} defined over the same nodes as G and whose arc $(i \to j)$ exists if and only if there exists a path of length α from i to j.[1]

In other words the α-fold product of the adjacency matrix M of G_A provides the following information: does there exist (or not) an ordered sequence of sectors $i \to r_1 \to r_2 \to, \dots, \to r_{\alpha} \to j$? i.e. is sector (j) (indirectly) input-connected to sector (j)?

We can now state and prove a simple characterization theorem for irreducible linear economies.

Theorem 2: A necessary and sufficient condition for A to be an irreducible system is that for all (i,j), i,j=1,2,...,n there exists a finite integer k such that

(3) $\quad m_{ij}^{k} = 1.$

(Note that this value of k is dependent on the particular ordered pair (i,j) that is used.)

(where $\sum$ represents (k)-fold summation) does not vanish then we can say that sector (i) is <u>indirectly</u> related to sector (j), specifically there is a finite sequence of sectors $(r_1, r_2, \ldots, r_k)$ of (k) elements linking the supplying sector (i) to the receving sector (k). The notion of <u>direct</u> vs. <u>indirect</u> interrelatedness hinges upon the existence (or non-existence) of one-step input dependency between sectors (i) and (j). This leads rather naturally to the notion of the length of the path between sector (i) and sector (j): for example, in terms of the above sequence $(r_1, \ldots, r_k)$ we could say that there exists a path of length k between (i) and (j). We are now ready to formalize the Yan and Ames notions of an interrelatedness function and of an order matrix associated with any input-output coefficient matrix A.

Consider the boolean matrix M of the same dimension as the A matrix and whose entries are respectively:

$$(2) \qquad m_{ij} = \begin{cases} 1 & \text{iff } a_{ij} > 0 \\ 0 & \text{iff } a_{ji} = 0 \end{cases}$$

This M matrix is simply the adjacency matrix of the graph G_A of the input-output system A. The nodes of G are the sectors 1,2,...,n and the arcs $i \to j$ are assigned to correspond to the non-zero a_{ij} coefficients. In fact these coefficients provide a valuation system for these arcs, which we shall use later on. Yan and Ames propose to associate the following distance matrix B - "order matrix" in their terminology - with any input-output matrix. M^{α}. Let $M^{\alpha} = \underbrace{M \dot{\times} M \dot{\times} \cdots \dot{\times} M}_{\alpha \text{ times}}$ represent the α-fold boolean product (denoted $\dot{\times}$) of the adjancy matrix M corresponding to G_A.

where $\ell(\cdot)$ refers to the length of the corresponding path. Repeating the same argument, if necessary, in a finite number of steps we have a path from i to j of length $k < n$.

A natural consequence of this result is that we need to perform at most n-1 boolean product operations on the M matrix to determine whether or not (j) is directly or indirectly input-dependent on (i). Another simple result can be obtained to characterize the entries of the B matrix proposed by Yan and Ames.

Theorem 4: Let A be an $(n \times n)$ input-output matrix with order matrix B. For any row $i=1,\ldots,n$ of B the following property holds:

$$\text{Let } \gamma = \underset{j}{\text{Max}} \left\{ b_{ij} \mid b_{ij} < \infty \text{ for } j=1,\ldots,n \right\}$$

Then for any integer K with $1 \leq K \leq \gamma$ there is a j with $b_{ij} = K$. Furthermore, if $\gamma = n-1$, $b_{ij} < \infty$ for $j=1,\ldots,n$ and $b_{ij} \neq b_{ij'}$, for $j \neq j'$.

Proof: Pick any row (i) in the B matrix. From Theorems 1 and 3 and the definition of the B matrix, there exists a shortest path from sector (i) to some sector (j) with $b_{ij} = \gamma$. Let this path be

$$(7) \qquad u_{ij} = \{(i,r_1) \; ; \; (r_1,r_2) \; ; \; \ldots \; ; \; (r_\gamma,j)\} \; .$$

Also, let u_{ir_h} be the segment of u_{ij} from (i) to (r_h). Now u_{ir_h} is a shortest path from (i) to (r_h), otherwise the length of u_{ij} could be reduced, which is impossible. This means $b_{ir_1} = \ell(u_{ir_1}) = 1$, $b_{ir_2} = \ell(u_{ir_2}) = 2$, etc. and the first part of the theorem is proved. To see the second part, note that the shortest path from (i) to (j) involves going through every sector if $\gamma = n-1$. Q.E.D.

The reader should be aware that an analogous result can be shown for columns by taking u_{r_hj} segments of u_{ij} .

Proof: It follows at once from the fact that the irreducibility property of a square matrix is equivalent to the strong-connectedness property of its graph (for a proof, see [3]).

An upper bound for the value k i.e. the length of the path between (i) and (j) can be readily found.

Theorem 3: If there exists a path connecting sector (i) to (j) then there exists a path of length $k < n$.

Proof: Let u_{ij} be a path from (i) to (j):

$$u_{ij} = \left\{(i,r_1) \; ; \; (r_1,r_2) \; ; \; \dots \; ; \; (r_{\alpha-1},j)\right\} \tag{4}$$

There are $\alpha-1$ r's and sectors i and j, or $\alpha+1$ sectors listed. If $\alpha \geq n-1$, there must exist at least two of the r's, say r_h and $r_{h'}$ that index the same sector. This means that there is a circuit of length at least one in u_{ij}. Removing that circuit reduces u_{ij} by at least one term. That is,

$$u_{ij} = \left\{(i,r_1) \; ; \; \dots \; ; \; \underbrace{(r_h,r_{h+1}) \; ; \; \dots \; ; (r_{h'}, r_{h'+1})}_{\text{circuit}} \; \dots \; ; \; (r_{\alpha-1},j)\right\} \tag{5}$$

Let

$$u'_{ij} = \left\{(i,r_1) \; ; \; \dots \; ; \; (r_h,r_{h'+1}) \; ; \; \dots \; ; \; (r_{\alpha-1},j)\right\} \tag{6}$$

and $\quad \ell(u'_{ij}) < \ell(u_{ij}) = \alpha$

However, even though the current measure interrelatedness is not very effective, the approach taken by Yan and Ames offer valuable insights into irreducibility of linear economies as exemplified by the results we have just presented. The methodological contribution is also important: it points towards the use of graph theoretic notions which enable us to reach these results quickly. Having provided these foundations a complete solution to the problem of measuring interrelatedness is now examined.

II. On measuring interrelatedness in a valued graph of input-output relations

The objective of this section is to define a new interrelatedness function meeting the objections raised earlier: (1) it should take into account the strength of the linkage between any two sectors i and j where (j) is input dependent upon (i); (2) in considering only the shortest path μ^*_{ij} between (i) and (j) we choose to ignore the simple fact that other feedback effects also exist; a whole family of paths $\theta_{ij} = \left\{\mu^1_{ij}, \ldots, \mu^\tau_{ij}, \ldots, \mu^t_{ij}\right\}$, can be found between i and j, and each such path reflects a different channel of input-dependence between sectors (i) and (j). To ignore all but the shortest path in θ is bound to give us a very biased view and measure of interrelatedness between any two group of sectors. Furthermore since the a_{ij} coefficients themselves are subject to some bias of their own due to the necessity of aggregation - as we shall examine in the last section of this paper - it is even more important to take into account not only the whole family of paths θ_{ij} linking (i) to (j) - for any pair of sectors i,j - but also the intensity of input-dependence by which each path $\mu^\tau_{ij} \in \theta_{ij}$ must be weighted in the completely general

Once this order matrix B has been obtained it was proposed by the same authors to define a scalar-valued function on it (or any submatrix thereof) that would "measure" the amount of "interrelatedness" between any two subgroups of sectors. Specifically these authors proposed to choose the reciprocal of the harmonic mean of the elements of the appropriate submatrix of B viz

$$(8) \qquad R\begin{pmatrix} i_1 \dots i_r \\ j_1 \dots j_s \end{pmatrix} = \frac{1}{rs} \sum_{v=1}^{r} \sum_{w=1}^{s} \frac{1}{b_{i_v j_w}}$$

For instance if $r = s = 1$ $R\begin{pmatrix} i \\ j \end{pmatrix} = \frac{1}{b_{ij}}$. Also it can be noted that R ranges over the closed interval [0,1]; R = 1 if and only if the corresponding elements of A are all strictly positive; and R = 0 if and only if all the corresponding elements of B are infinite. A natural question to be asked is then: does this interrelatedness function R obtained from this order matrix B fully capture the notion of interdependence that characterizes any general equilibrium system, and, in particular, the input-output system? To answer this question we can simply quote the statement of one of its fundamental limitations as given by one of theauthors: "This measure considers only the existence of input-output relations between industries and disregards the magnitude of transactions" (see [19] p.92).

A second limitation also exists and should be mentioned: in defining the order matrix B, only the shortest paths between any two sectors are considered. However there is a whole family of paths of varying lengths besides that shortest path for each pair of sectors and a forticri for any two group of sectors. Discarding these connections entails a very high informational loss and renders the interrelatedness function much less meaningful in its current formulation.

propose not to ignore the remaining paths but to take them into account and weigh them according to their number - for a given length. To obtain an average measure of interrelatedness based on this multiplicity of paths of various lengths, we could simply look at the following matrix sum:

$$(10) \qquad S_{\lambda} = I + M + M^2 + M^3 + \ldots + M^{\lambda} .$$

The first term in this sum (I) indicates all the paths of length 0 i.e. the loops at each vertex; the second term refers to the paths of length 1, etc. However, for obvious reasons, in any realistic situation this sum would never converge finitely. Then if convergence is ruled out we could simply truncate the series after some preassigned upper bound $\bar{\lambda}$ is reached and apply the $\tilde{R}$ function as above with truncated values $\bar{\bar{m}}_{i_v j_w} = \sum_{\lambda=1}^{\bar{\lambda}} m_{i_v j_w}^{(\lambda)}$.

An argument in favor of this truncated solution is simply that it implicitly recognizes the fact that beyond a certain point (say $\bar{\lambda}$) $\bar{\lambda}$-step input-dependence between any two sectors may be of little relative importance simply because some or all of the input coefficients along this path are small, so that the cumulative effect cannot be large. These remarks, however, are rather vague and unquantifiable as they stand. We are now going to show that they can be made more precise and quantified if we allow for differential linkage intensities between the various sectors.

interrelatedness function we are looking for. These two problems will be attacked separately: first a solution will be proposed to take into account the whole family θ_{ij} of paths for any (i,j) pair; and then that solution will be incorporated in a more general scheme that will weigh the paths $\mu_{ij}^{\tau} \in \theta_{ij}$ according to the level of input-dependence along each path.

2.1. <u>Families of paths and interrelatedness of order $1,2,\ldots,\lambda$</u>

Again, let us consider the adjacency matrix M of the graph G_A of a given input-output system A. The following theorem states a standard result in graph theory (see e.g. [2] [13]).

<u>Theorem 4</u>: Let $G = (N,U)$ be a graph with vertex set N and edge set U. Let M be its adjacency matrix. Then each entry p_{ij} of the matrix $P = M^{\lambda}$ indicates the number of distinct paths of length λ from i to j $(i,j \in N)$.

Thus by using Theorem 4 it is a simple matter to take into account the number of elements (paths) μ_{ij}^{τ} in each family θ_{ij} for any (i,j) pair. Algorithms for path enumeration exist and can be used if the knowledge of each element in θ_{ij} is desired. Actually for our purpose, we can simply define a class of multiple interrelatedness functions, one for each value of λ viz:

$$(9) \qquad \tilde{R}_{\lambda}\begin{pmatrix} i_1 & \cdots & i_r \\ j_1 & \cdots & j_s \end{pmatrix} = \frac{1}{rs} \sum_{v=1} \sum_{w=1}^{s} \frac{1}{m_{i_v j_w}^{(\lambda)}}$$

One such function exists for each value of λ. The problem, however, is to find an aggregate measure of multiple interrelatedness. Yan and Ames chose to single out a different λ for each (i,j) pair viz the minimal one necessary to yield a path between (i) and (j). Here we

2.2. Multiple paths, levels of input-dependence and generalized interrelatedness functions

To motivate our discussion let us return for a moment to equation (10) which reads

$$S = I + M + M^2 + M^3 + \ldots + M^\lambda$$

where the M matrix is the adjacency matrix of G_A the graph of a given input-output system A. Each term in S is a matrix whose entries m_{ij}^λ indicate the number of paths of length λ ($\lambda = 0,1,2,\ldots$) between (i) and (j). Now if we simply use the a_{ij} input-output coefficients to take into account the level of input-dependence between (j) and (i) and iterate the A matrix λ times the entry $a_{ij}^{(\lambda)}$ is a measure of the proportion of j's output which is technologically input-dependent on i's output: more precisely given the technological framework reflected by the a_{ij} coefficients after exactly λ steps via intermediate sectors $r_1,r_2,\ldots,r_\lambda$ a proportion $a_{ij}^{(\lambda)}$ of i's output enters as an input to sector j. This matrix A^λ provides us with an answer to both objections that were raised earlier: it takes into account <u>all</u> the paths of length λ between (i) and (j) <u>as well as</u> their valuations in terms of input-dependence. One need not worry whether an aggregate and input-weighted interrelatedness function can be uniquely defined or whether we have to settle for just a class of λ such functions - one for each path of length λ. It is a well-known fact that the following matrix series:

$$\lim_{\lambda\to\infty} T = I + A + A^2 + A^3 + \ldots + A^\lambda$$

converges to the limiting matrix

$$Z = (I-A)^{-1} \quad \text{provided} \sum_i a_{ij} < 1 \text{ for } j=1,\ldots,n.$$

From an empirical standpoint as Yan and Ames have pointed out this substochasticity condition is, of course, met, as soon as we consider the open input-output model so we can simply use the z_{ij} coefficients thus obtained to calculate the generalized - i.e. input-weighted and with multiple paths-interrelatedness function between any two group of secors $(i_1,\ldots,i_r)$ and $(j_1,\ldots,j_s)$. In fact we can simply define this function R* as the arithmetic mean of the appropriate entries in Z

$$R^*\begin{pmatrix} i_1 \ldots i_r \\ \\ j_1 \ldots j_s \end{pmatrix} = \frac{1}{r\cdot s} \sum_{v=1}^{r} \sum_{w=1}^{s} z_{i_v j_w}$$

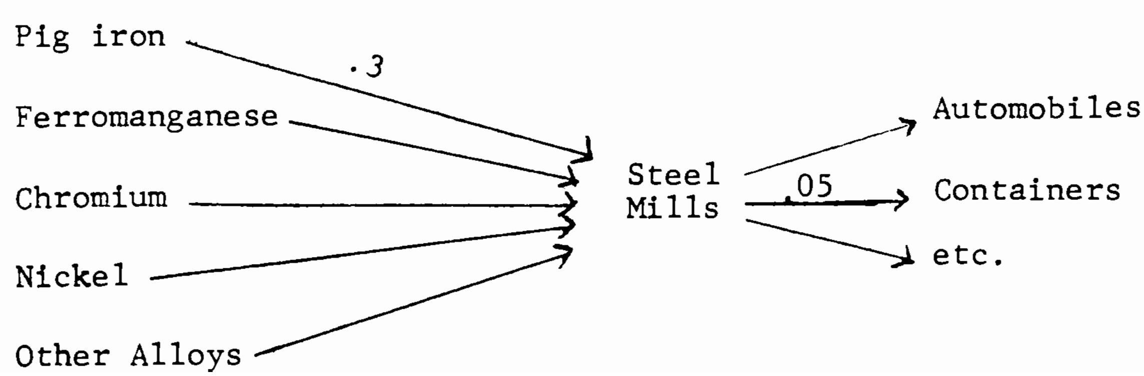
For instance, if v = n and s = 1 the jth sector's interrelatedness with all sectors in the economy is simply the arithmetic mean of the entries in the jth column vector z_j in Z.

Thus, as we had mentioned at the end of the preceding section, we can now see more clearly how a simultaneous solution to the intensity of input dependence problem and the existence of multiple paths problem is simply achieved by considering the non-negative inverse (Z) matrix. An interesting illustration of the kind of solution we obtain with this new measure R* is provided by the case of a perfectly triangular economy:

$$A = \begin{bmatrix} a_{11} & \cdots & & a_{1n} \\ 0 & \ddots & & \\ \vdots & & a_{ij} & a_{in} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & a_{nn} \end{bmatrix}$$

III Some remarks on aggregation in input-output studies and its effect on interrelatedness measures

A word of caution is now in order. Although most researchers in input-output economics are fully aware of the problems raised by aggregating data from various multi-product firms into sectoral data, this notion of sector, as defined by a single homogeneous product, is hardly applicable in practice. The problem we encounter in empirical input-output studies is simply that given the immense diversity of products which characterize modern industrialized economies, the sectors must be defined as "aggregates" of closely related products. This notion of "closeness" is seldom explicitly elaborated upon: more often than not this preliminarly aggregation takes place through a certain choice of terminology; picking a name for a sector involves such an aggregation and this is enough to introduce some spurious relationships in the model. This kind of aggregation may of course be compounded by some later explicit aggregation meant to reduce the size of the matrix and hence the computational difficulties involved. An illustration of this notion can be readily given; consider the following diagram (taken from Noble [12])



Then Z is also triangular.

$$Z = \begin{bmatrix} z_{11} & \cdot & \cdot & \cdot & \cdot & z_{1n} \\ \cdot & \ddots & & & & \\ \cdot & & z_{ii} & & & z_{in} \\ \cdot & & & \ddots & & \\ 0 & & & & & z_{nn} \end{bmatrix}$$

and for instance

$$R^* \begin{pmatrix} & 1 & \\ 1 & \ldots & n \end{pmatrix} = \frac{z_{11}}{n}$$ since the first sector is completely

self-sufficient.

Sequentially we get

$$R^{**} \begin{pmatrix} & 2 & \\ 1 & \ldots & n \end{pmatrix} = \frac{1}{n} \left[z_{12} + z_{22} \right]$$

since sector 2 is self-sufficient with the exception of its reliance upon some proportion of one's output as its first input. This argument applies in the same manner to the remaining sectors.

An empirical study and intertemporal as well as international comparisons of the behavior of this R* function will be provided elsewhere. For the time being we would like to turn to a sensitivity analysis under aggregation of the interrelatedness measures we have studied.

not because of an arbitrary aggregation scheme but because of the technology involved. Clearly a good measure of interrelatedness should try to minimize the effects of such spurious links. This is precisely the fundamental advantage of a weighted measure of interrelatedness as we have proposed here: even though it will not be able to distinguish the amount of spurious input-dependence in the actual amount given by the a_{ij}'s (and, once aggregation has taken place, this information is lost anyway, so no method can hope to retrieve it in full) it will at least assign different weights to these connections according to the length of each path. More precisely the longer the path, the more likely are we to encounter a spurious connection along one of its arcs; but the weights assigned to the paths being products of a_{ij} coefficients which are all less than one, the effect of a spurious coefficient in that product becomes increasingly smaller as the path grows larger.

Looking at the above graph and assuming for instance that the arcs are valued with the appropriate input coefficients of the input-output matrix, it is clear that in order to use it for predictive purposes one fundamental assumption must hold, viz. that the multiplication of the input coefficients yields a predictive functional relationship. In the above example the production of containers requires a production of pig iron of 3/10 times 5/10, i.e. 15/100. In order to give good predictive results the inputs must be independent of the composition of the output. Independence exists for the flows of pig iron and manganese into steel mills but not for chromium and nickel, because they go into stainless steel but not into carbon steel. In short, a purchase order for carbon steel <u>does not</u> generate requirements for production of chromium or nickel. This example is by no means unique. In fact, the same problem occurs for most sectors; and, thus, we must be careful to keep in mind that most empirical input-output coefficients contain such distortion factors of the level of sectoral interdependence beyond the first level of demand as represented by the A matrix.

In regard to the problem of measuring interrelatedness between groups of sectors in an economy, the difficulty we have just described provides one of the strongest arguments in favor of the generalized interrelatedness measure R* proposed in this paper. The reason is simple: when we ignore the intensity of input-dependence between sectors and collapse this information into an adjacency matrix we may actually be taking into account connections that are, say, 90%, spurious and only due to the aggregation format used. These connections, however, carry exactly as much weight in the Yan and Ames interrelatedness index (R) as the "genuine" ones, i.e. the ones that arise

BIBLIOGRAPHY

[1] Augac, H., "La hierarchie des industries dan un tableau des echanges interindustriels, <u>Revue Economique</u>, Vol. 2, 1960.

[2] Berge, C., <u>Théorie des Graphes et ses Applications</u>, Dunod, Paris, 1958.

[3] Blin, J. M., <u>Patterns and Configurations in Economic Science</u>, D. Reidel Publishing Company, Dordrecht, Holland, (in press).

[4] Debreu, G., and I. N. Herstein, "Nonnegative Square Matrices," <u>Econometrica</u> 21, 1953, pp. 597-607.

[5] Dulmage, A. L. and N. S. Mendelsohn, "The Structure of Powers of Non-Negative Matrices," <u>Canadian Journal of Mathematics</u>, 17, 1965, pp. 318-330.

[6] Kemeny, J. C., and J. L. Snell, <u>Finite Markov Chains</u>, Princeton C. Van Norstrand, 1960.

[7] Koopmans, T. C., (ed.) <u>Activity Analysis of Production and Allocation</u>.

[8] Korte, B., and W. Oberhofer, Triangularizing input-output matrices and the structure of Production, <u>European Economic Review</u>, 1, 4, 1971.

[9] Leontief, W., <u>The Structure of American Economy</u>, 1919-1939, 2nd ed., Oxford University Press, N. Y., 1951.

[10] ____________, (ed.) <u>Studies in the Structure of the American Economy</u>, Oxford University Press, N. Y. 1953.

[11] Masson, D., "Methode de triangulation de tableau European des echanges interindustriels," <u>Revue Economique</u>, Vol. 2, 1960.

[12] Noble, S. B., Some Flow Models of Production constraints, <u>Naval Research Logistics Quarterly</u>, Vol. 7, #4, pp. 401-419.

[13] Ore, O., Theory of Graphs, A. M. S., Providence, 1962.

[14] Rosenblatt, D., "On Linear Models and the Graphs of Minkowsky-Leontief Matrices," Econometrica, 25, 1957, pp. 325-338.

[15] Schumpeter, J., History of Economic Analysis, G. Allen and Unwin Ltd., London, 1963.

[16] Simpson, D., and J. Tsukui, "The Fundamental Structure of Input-Output Tables: an International Comparison," The Review of Economics and Statistics, Vol. 47, 1965, pp. 434-446.

[17] Varga, R. S., Matrix Iterative Analysis, Prentice-Hall, Inc. Englewood Cliffs, 1962.

[18] Yan, C., and E. Ames, Economic Interrelatedness, Review of Economic Studies, Oct. 1965, Vol. XXXII, No. 4, pp. 299-310.