empirical economics, Vol. 2, Issue 1, page 1–9. Physica-Verlag, Vienna.

The Determination of Groups of Functionally Integrated Industries in the United States Using a 1967 Interindustry Flow Table[1])

By *P.B. Slater*, Morgantown[2])

Abstract: A two-stage standardization and grouping algorithm is applied to a 75 × 75 submatrix of the 86 × 86 1967 United States interindustry transaction table. The rows and columns of the submatrix are first adjusted so that they all have uniform sums. Entries of the adjusted table are maximum entropy estimates of flows in an idealized economy, in which all industries produce and consume an equal value of goods. Industries are then aggregated according to whether or not they are connected by chains of links with values equal to adjusted entries, all of which are greater than a given threshold level. Variation of the threshold yields a hierarchical aggregation scheme. Groups of functionally integrated industries can be identified through the examination of a tree diagram (dendrogram) which shows the aggregation results.

Introduction

Interindustry flow tables have been constructed – at great expense – for many nations and regions of the world. These tables are *potential* sources of information on the structural integration of groups of industries. To realize this potential, exploratory data analytical tools must be applied to these matrices [*Slater*, 1974]. Several such methods have been used in recent years to identify industrial complexes, using input-output data [*Fisher/Ghosh; Campbell*, 1972, 1974; *Roepke; Czayka*].

Campbell has suggested that graph-theoretic procedures – in which all interindustrial linkages are either assigned zero or unit weights – be extended to network analyses in which continuous data are employed. Such an extension is achieved in this study of the 1967 United States interindustrial transaction table. This is accomplished by systematically varying thresholds by which the presence or absence of linkages is determined, grouping industries on the basis of these interconnections, and arranging all the distinct groupings obtained at the various thresholds in a hierarchical diagram (dendrogram). The significances of the industrial clusters found can then be informally gauged by the differences (ranges of existence) between the thresholds at which they are formed by the amalgamation of smaller groups and the thresholds at which they combine with others into larger aggregations.

---

[1]) Support for this research was provided by the National Science Foundation under Grant No. SOC 74–23880, "Applications of Graph-Theoretic Clustering and Transaction Flow Analysis in Geography and Regional Science."

[2]) *Paul B. Slater*, Regional Research Institute and Department of Statistics and Computer Science, West Virginia University, Morgantown, West Virginia 26506, USA.

The transaction table is doubly (row and column) standardized before it is used for the clustering. This procedure removes effects of industrial scale and permits attention to be concentrated on interindustrial interaction – i.e., structural integration. The magnitudes of the flows between pairs of industries are confounded with the values of total outputs and inputs. In the interindustry coefficient table, total inputs, but not outputs, are controlled for. Through use of an iterative proportional fitting procedure (IPFP) [*Fienberg*] – also termed the biproportional or RAS method [*Bacharach*] – both total input and output size can be controlled for. Entries of the flôw table after this double standardization are statistical (maximum entropy) estimates of the transactions that would occur if all industries had identical levels of input and output. They can be regarded as measures of direct linkages. The clustering procedure reveals indirect as well as direct ties.

This two-stage procedure – double standardization and hierarchical clustering – has been applied to a variety of flow tables. In particular, it has been used to study regional structures of various nations on the basis of internal migration tables [*Slater,* 1976a–d].

Procedures

A 75 X 75 submatrix of the 86 X 86 1967 United States transactions table [U.S. Department of Commerce] was extracted for analysis. All industries grouped in 1. Agriculture, Forestry, and Fisheries; 2. Manufacturing; 3. Transportation, Communication, Electric, Gas, and Sanitary Services; 4. Wholesale and Retail Trade; 5. Finance, Insurance, and Real Estate; and 6. Services were included. The products of New Construction were not purchased by any of these industries. Therefore, it was eliminated from consideration since the row sum corresponding to it was 0 and could not be adjusted by the IPFP. Government Enterprise, Imports, Dummy Industries, and Special Industries were not studied. The IPFP was employed to adjust this table – with zeros inserted on its diagonal – so that all row as well as all column sums were equal to the same number – here selected to be 1,000. (The IPFP could have been employed to adjust the 1967 table with intraindustrial flows present. Some results would have been different if this had been done. Tobacco Manufactures would not have had the three relatively strong forward linkages it exhibits in Figure 4, for example, since large amounts of its product are sold intraindustrially. Personal consumption expenditure – an unincorporated sector – is the major purchaser of tobacco manufactures.)

The procedure by which industries are grouped together on the basis of the doubly-standardized values can be represented pictorially. A diagram could be constructed with 75 points in it – one for each industry. Initially, there are no connections between these points. The largest value (a_{ij}) in the adjusted table is found. An arrow is then drawn from the point for industry i to the point for industry j. Next, an arrow is constructed corresponding to the second largest entry, then the third, etc. The arrows serve to unite the points. There are two ways in which they can be considered to do so. In one, two points (l and k) are considered joined, if there is a sequence of directed arrows from l to k and a sequence from k to l. (In graph-theory terminology [*Harary*, et al.], k and l are said to lie in the same strong component of a directed graph [digraph].) In the other, l and k are deemed consolidated if a sequence of arrows – without regard to their direction – connect l and k to one another. (l and k are then said to lie in the same weak component.) Any strong component is, thus, also a weak component, but not vice versa. (These concepts are illustrated in the subsequent discussion of the figures.) As more and more arrows are inserted into the diagram, the components of the graph become fewer in number but larger in size, until eventually all points are united in one component. The process by which the 75 isolated points join in stages together can be represented by a tree diagram (dendrogram). In such a diagram, two industries (i and j) are shown joining together at a certain *threshold*. This is equal to the value of the adjusted entry (a_{pq}), construction of an arrow for which united i and j into the same component. Computationally, these procedures are performed by constructing a 75 X 75 null matrix, the ij-cell of which is changed from zero to one when the threshold is lowered to the value of a_{ij}. The powers of this matrix reveal paths in the diagram, and consequently can be employed to determine the weak and strong components.

It would be of interest to cluster industries both by strong, as well as weak, components. Strong components can be regarded as relatively integrated self-propulsive groups of industries. Increases in the production of one member industry will place added demands on every other constituent. Feedback effects will also induce increased output in the initially expanded industry. The analysis reported here is based on weak components. Chains of interindustrial effects, which need not be cyclic in structure, are found using weak components. The hierarchical clustering of the 75 industries is shown in Figure 1.

Results

The use of any threshold less than 74.5 resulted in a digraph in which all industries were weakly connected to one another. When the forward (supply) linkage of Other Furniture and Fixtures to Other Fabricated Metal Products that had this value and all lesser ones were nullified, the associated digraph had two weak components. These were composed of: 1. all industries except Other Fabricated Metal Products and 2. Other Fabricated Metal Products itself. Thus, Other Fabricated Metal Products – which includes cutlery, hand and edge tools, hardware, wire, safes, steel springs, and pipes – had no forward or backward linkage greater than 74.5. In this sense, it was the most central of the 75 industries, since it had diffuse ties with the others. The six industries listed above Other Fabricated Metal Products in Figure 1 are also relatively central ones.

Many clusters of industries can be found through further examination of the tree diagram. Strong pairs – together with the range of thresholds at which they exist as doublets – are formed by Stone and Clay Products and Stone and Clay Mining and Quarrying (130 to 451); Household Appliances and Service Industry Machines (the latter industry purchases laundry equipment, refrigerators and freezers from the former [133 to 549]); Ordnance and Accessories – establishments engaged in the

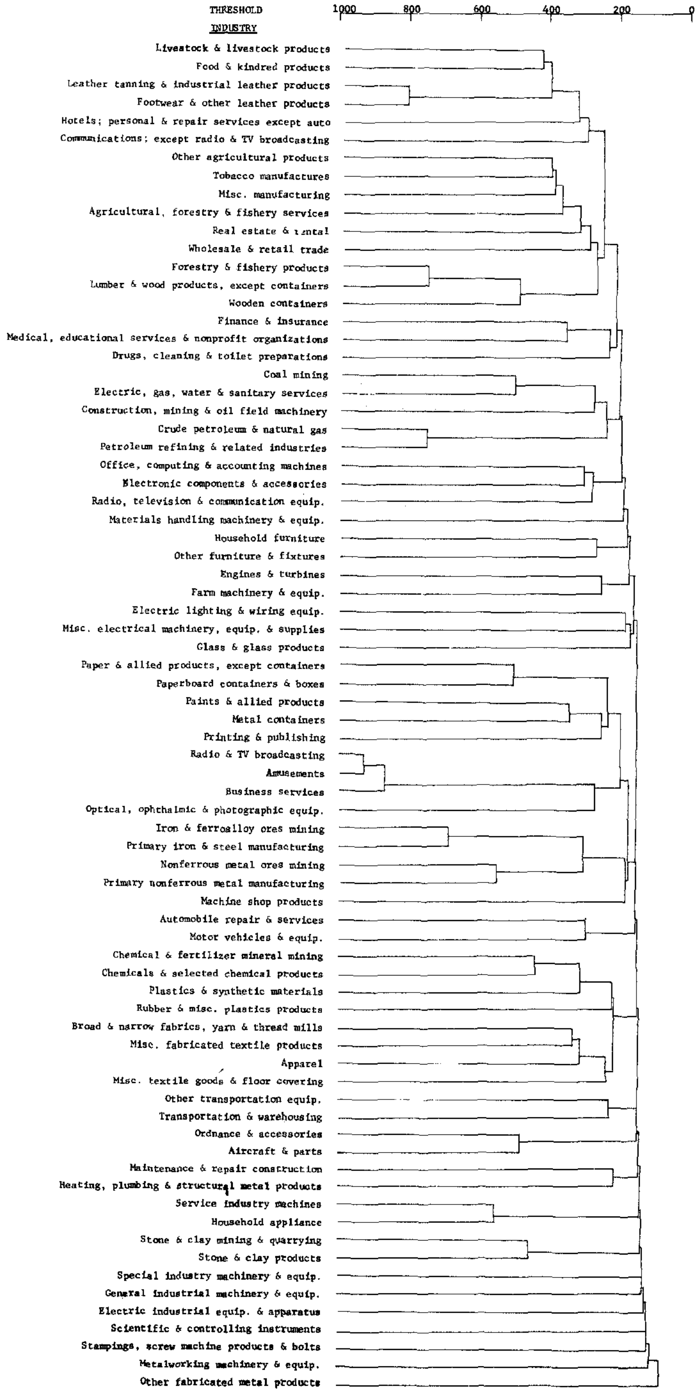
THRESHOLD
1000
800
600
400
200
0
INDUSTRY
Livestock & livestock products
Food & kindred products
Leather tanning & industrial leather products
Footwear & other leather products
Hotels; personal & repair services except auto
Communications; except radio & TV broadcasting
Other agricultural products
Tobacco manufactures
Misc. manufacturing
Agricultural, forestry & fishery services
Real estate & rental
Wholesale & retail trade
Forestry & fishery products
Lumber & wood products, except containers
Wooden containers
Finance & insurance
Medical, educational services & nonprofit organizations
Drugs, cleaning & toilet preparations
Coal mining
Electric, gas, water & sanitary services
Construction, mining & oil field machinery
Crude petroleum & natural gas
Petroleum refining & related industries
Office, computing & accounting machines
Electronic components & accessories
Radio, television & communication equip.
Materials handling machinery & equip.
Household furniture
Other furniture & fixtures
Engines & turbines
Farm machinery & equip.
Electric lighting & wiring equip.
Misc. electrical machinery, equip. & supplies
Glass & glass products
Paper & allied products, except containers
Paperboard containers & boxes
Paints & allied products
Metal containers
Printing & publishing
Radio & TV broadcasting
Amusements
Business services
Optical, ophthalmic & photographic equip.
Iron & ferroalloy ores mining
Primary iron & steel manufacturing
Nonferrous metal ores mining
Primary nonferrous metal manufacturing
Machine shop products
Automobile repair & services
Motor vehicles & equip.
Chemical & fertilizer mineral mining
Chemicals & selected chemical products
Plastics & synthetic materials
Rubber & misc. plastics products
Broad & narrow fabrics, yarn & thread mills
Misc. fabricated textile products
Apparel
Misc. textile goods & floor covering
Other transportation equip.
Transportation & warehousing
Ordnance & accessories
Aircraft & parts
Maintenance & repair construction
Heating, plumbing & structural metal products
Service industry machines
Household appliance
Stone & clay mining & quarrying
Stone & clay products
Special industry machinery & equip.
General industrial machinery & equip.
Electric industrial equip. & apparatus
Scientific & controlling instruments
Stampings, screw machine products & bolts
Metalworking machinery & equip.
Other fabricated metal products

Fig. 1

manufacture of weaponry; in particular, complete guided missiles – and Aircraft and Parts (144 to 478); Transportation and Warehousing and Other Transportation Equipment (146 to 225); Metal Containers and Paints and Allied Products (250 to 341); Crude Petroleum and Natural Gas and Petroleum Refining and Related Industries (241 to 751); Electric, Gas, Water, and Sanitary Services and Coal Mining (277 to 503); and Footwear and Other Leather Products and Leather Tanning and Industrial Leather Products (401 to 806). Several other pairs of varying strengths can be found through examination of the dendrogram.

Several of the larger clusters, together with their salient linkages – those greater than the thresholds at which the clusters separate from larger groups – that bind them as weak components are shown in Figures 2–5. The range of existence of the textile-oriented group of Figure 3 is 150 to 214. At 214 the supply linkage of Plastics and Synthetic Materials to Broad and Narrow Fabrics, Yarn and Thread Mills is nullified, splitting the complex. The four topmost industries in the diagram form a strong component, since a path of directed links exists back and forth between any two of the vertices. In contrast to the other four industries, this group has a high degree of homogeneity – that is, a high density of linkages. Non-homogeneous clusters are characterized by "chaining" effects, by which two industries can be grouped together even if the linkages between them are quite indirect. Figure 4 is an extreme example of this

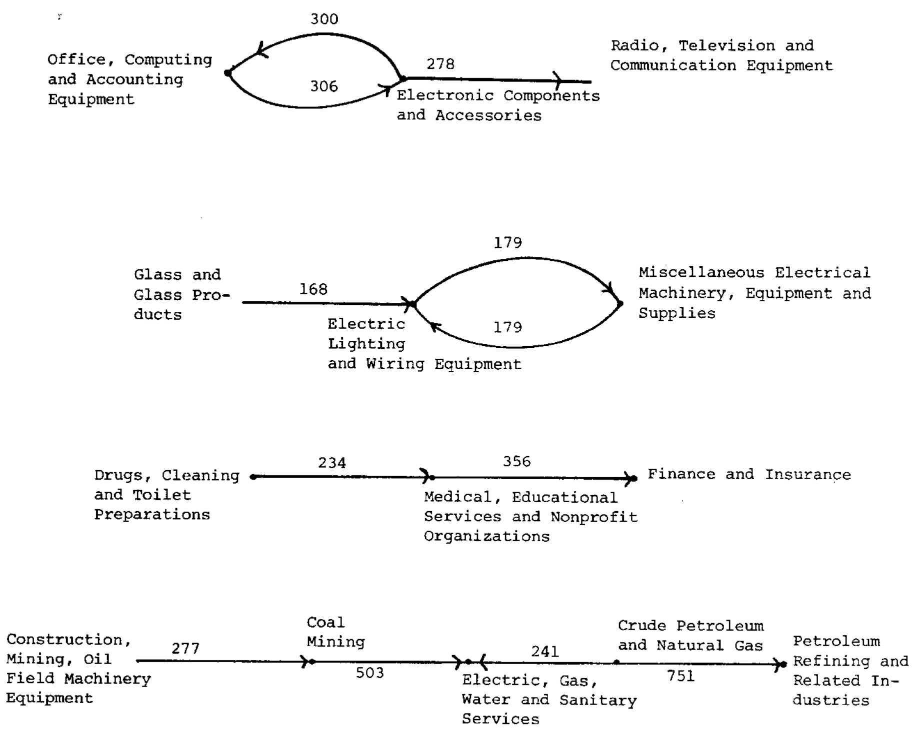

Fig. 2

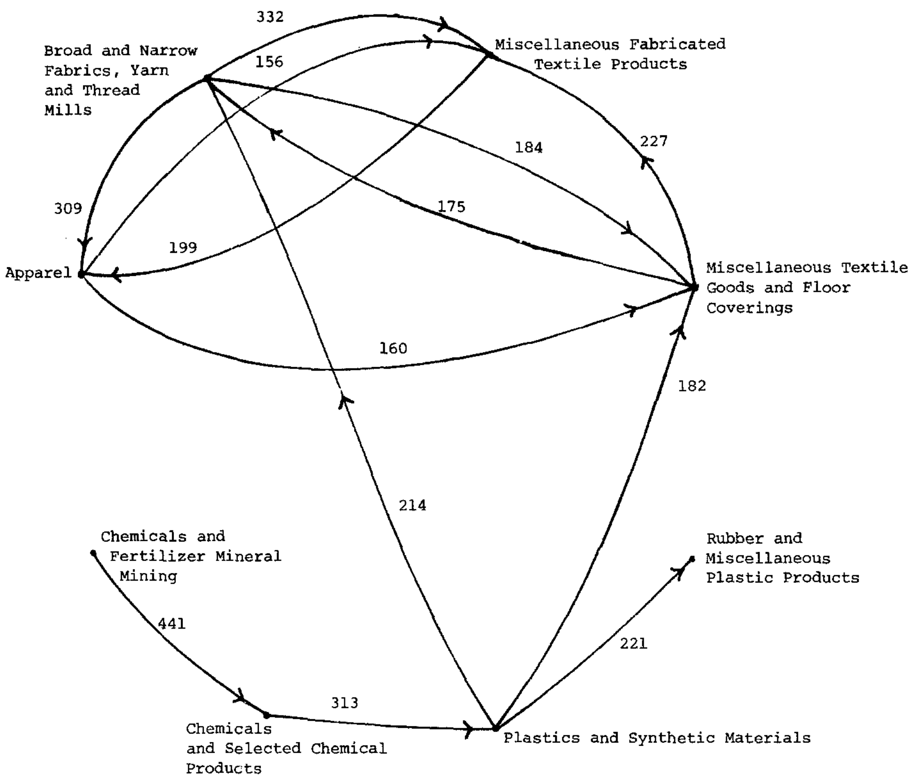

Fig. 3

phenomenon, while Figure 5 also exhibits it, although it does have two strong components which increase homogeneity. These are Food and Kindred Products and Livestock and Livestock Products, and Agricultural, Forestry and Fishery Services, Forestry and Fishery Products, Lumber and Wood Products, except Containers and Wooden Containers.

Discussion

A method for effectively summarizing the intricate interrelationships between industries has been presented. It has been suggested that the efficiency of the spatial organization of the economy can be evaluated by comparing aspatial industrial complexes, such as those identified here, with actual industrial conglomerations [*Roepke*].

The results given here can be extended in several directions. Perhaps the most interesting would be an analogous study of the 1967 flow table on the 484-industry level. Industrial classification procedures could be critically re-examined by doing so. Comparisons between groupings on the basis of weak and strong components should also

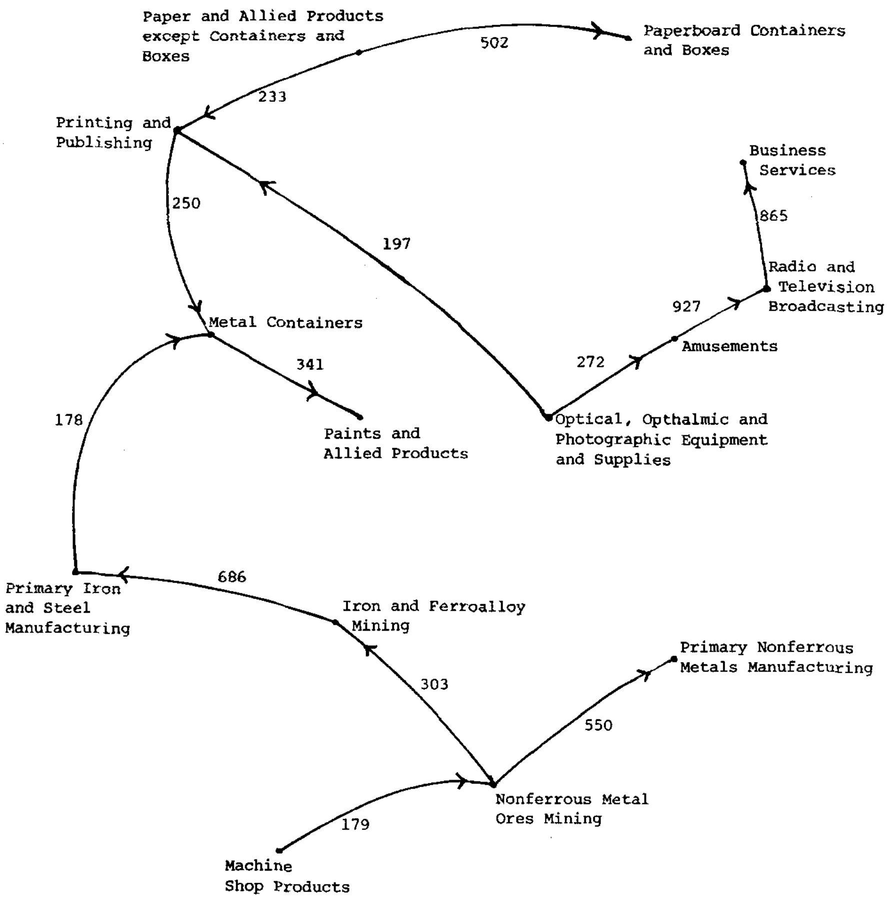

Fig. 4

be made. It would also be of value to conduct analyses in which intraindustrial flows are taken into account through adjustment of the flow table with them present on the diagonal. Industries which purchase relatively large amounts of their own products will then have relatively weaker linkages than in the zero-diagonal analysis. The degree of structural integration of industries in various nations could be compared, using flow tables with identical classification schemes.

Consideration needs to be given in any of these projected works so as to the appropriate submatrices of the transaction table to be analyzed. Should government enterprises, dummy and/or special industries be disregarded? Relevant to this question is the fact that the biproportional adjustment procedure may not converge if the table to be standardized is sparse; that is, most interindustry flows are recorded as zero. It would then be necessary to decompose the matrix into subtables to be separately analyzed by the two-stage standardization and clustering algorithm.

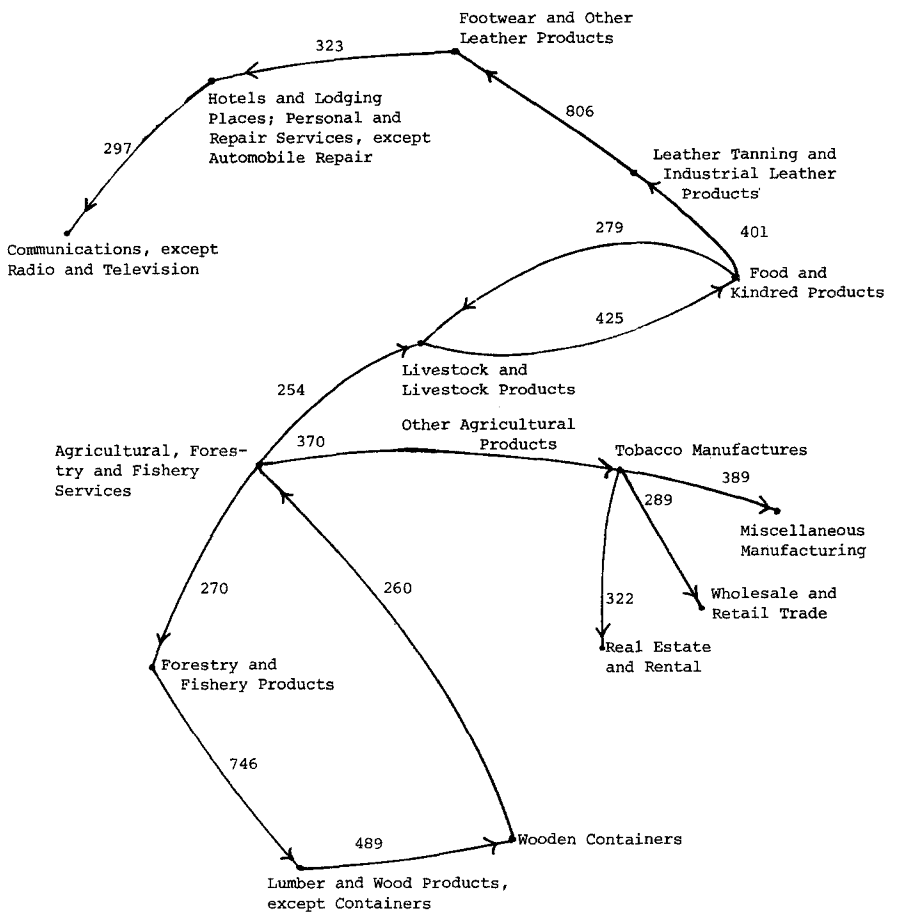

Fig. 5

References

Bacharach, M.A.: Biproportional Matrices and Input-Output Change, Cambridge 1970.

Campbell, J.: Growth Pole Theory, Digraph Analysis, and Interindustry Relationships, Tijdschrift voor Economische en Sociale Geografie **63**, 1972, 79–87.

– : Selected Aspects of the Interindustry Structure of the State of Washington, Economic Geography **50**, 1974, 35–46.

– : Application of Graph Theoretic Analysis to Interindustry Relationships, Regional Science and Urban Economics **5**, 1975, 91–106.

Czayka, L.: Qualitative Input-Output Analyses, Meisenheim am Glan 1972.

Fienberg, St.E.: An Iterative Procedure of Estimation in Contingency Tables, Annals of Mathematical Statistics **41**, 1970, 907–917.

Fisher, W.D.: Clustering and Aggregation in Economics, Baltimore 1969.

Ghosh, A., and *H. Sarkar*: An Input-Output Matrix as a Spatial Configuration, Economics of Planning **10**, 1970, 133–142.

Harary, F., R.Z. Normann, and *D. Cartwright*: Structural Models: An Introduction to the Theory of Directed Graphs, New York 1965.

Roepke, H., D. Adams, and *R. Wiseman*: A New Approach to the Identification of Industrial Complexes Using Input-Output Data, Journal of Regional Science **14**, 1974, 15–29.

Slater, P.B.: Exploratory Analyses of Interindustrial Transaction Tables, Environment and Planning **6**, 1974, 207–214.

– : Hierarchical Internal Migration Regions of France, IEEE Transactions on Systems, Man, and Cybernetics **6**, 1976a, 321–324.

– : The Use of State-to-State College Migration Data in Developing a Hierarchy of Higher Educational Regions, Research in Higher Education **4**, 1976b, 305–315.

– : A Hierarchical Regionalization of Japanese Prefectures Using 1972 Interprefectural Migration Flows, Regional Studies **10**, 1976c, 123–132.

– : A Hierarchical Regionalization of State Economic Areas Using 1965–70 Migration Data, Review of Public Data Use **4**, November, 1976d, 32–56.

U.S. Department of Commerce: Input-Output Structure of the U.S. Economy, Survey of Current Business **15**, 1974, 24–56.

Received October 5, 1976