Forthcoming in The Review of Economics and Statistics

DISCUSSION PAPER NO. 163

Technological Similarity and Aggregation in Input-Output Systems: A Cluster-Analytic Approach*

by

Jean-Marie Blin and Claude Cohen

(revised March 1976)

*The authors wish to thank two anonymous referees for their very helpful comments and suggestions on an earlier version of this paper.

ABSTRACT

This paper extends the methodology proposed first by Charnes and Cooper [5] and further specified by Kossov [12], to implement the Hatanaka condition [11] for zero aggregation bias in input-output systems. The notion of technological similarity across industries is defined to determine groupings of 'almost input-homogeneous' industries. Cluster-analysis is used for implementation. Empirical results for the U.S. 1967 (83 x 83) input-output table are analyzed and compared to Chenery and Watanabe's 29 sector aggregation [6]. Application of the method to the Leontief inverse is also demonstrated and implications for testing certain hypotheses in economic history and development theory are briefly discussed.

I. Introduction

1.1. Statement of the Problem

Among the many problems raised by the actual construction of empirical input-output tables two key issues are often singled out: (1) what industrial classification scheme should be adopted and (2) how should the data be structured. Actually, these problems of industry definition and data structuring are two facets of a more fundamental problem: given external constraints on data gathering (availability and format) which make the industry concept mostly unobservable, how can we "best" group the available data into an input-output table. Ideally, each industry would be defined by a single well-defined product and a separate industry should be used for different, but possibly "closely related", products.(1) A widely accepted notion of "best" grouping is one which minimizes the "aggregation bias" i.e., the difference between the gross output forecast obtained with a disaggregated table and the forecast obtained with an aggregated table, for any final demand bill. Historically, a theoretical condition for zero aggregation bias was first derived by Hatanaka [11].(2) Briefly stated, let A denote the $(n \times n)$ disaggregated direct coefficient matrix; x denote the n-dimensional column vector of gross outputs; y denote the n-dimensional column vector

---

(1) The same issues, as in the design of a Standard Industrial Classification System (SIC), are faced when one tries to apply this notion of "closely related" products; and very similar criteria are used, e.g., the sequence of engineering processes involved, or the close substitutability of the products.

(2) See also Aza [2], Balderston and Whitin [3], Fei [7], W.D. Fisher [8], Ghosh [9], Hatanaka [11], MacManus [14], Malinvaud [15], Morimoto [16], Skolka [19], Theil [20], [21].

of final demands; I denote the (n x n) unit matrix; S denote the aggregation operator where S is (M x n) and reads:

$$(1) \qquad S = \begin{bmatrix} 1 \ldots 1 & 0 \ldots & 0 \ldots 0 \\ 0 \ldots 0 & 1 \ldots 1 & 0 \ldots 0 \\ 0 \ldots & 0 \ldots 0 & 1 \ldots 1 \end{bmatrix} = \begin{bmatrix} e_1' \ldots & 0 \\ \vdots \quad e_j' & \vdots \\ 0 \ldots & e_M' \end{bmatrix}$$

e_j' is a row vector of order s(J) whose elements are all 1's; ($j=1 \ldots M$; $\sum_J s(J) = n$); let the same starred symbols denote the aggregated vectors and coefficient matrices. Then we have $(I-A)x = y$ and $(I^*-A^*)x^* = y^*$. The aggregation bias is:

$$(2) \qquad x^* - Sx = [(I^*-A^*)^{-1}S - S(I-A)^{-1}]y = Vy$$

or

$$(3) \qquad Vy = [(I^* + A^* + A^{*2} + \ldots)S - S(I + A + A^2 + \ldots)]y$$
$$= [(A^*S - SA) + (A^{*2}S - SA^2) + \ldots]y$$

A necessary and sufficient condition for zero aggregation bias for any y is that

$$(4) \qquad SA = A^*S$$

For instance if

$$S = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad A = \begin{bmatrix} a_{11} & a_{12} & \vdots & a_{13} \\ a_{21} & a_{22} & \vdots & a_{23} \\ \cdots & \cdots & & \cdots \\ a_{31} & a_{32} & \vdots & a_{33} \end{bmatrix}; \quad A^* = \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}$$

we must simultaneously have

$$a_{11}^* = a_{11} + a_{21} = a_{12} + a_{22} \; ; \; a_{13} + a_{23} = a_{12}^* \; ; \; a_{31} = a_{21}^* = a_{32} \; ; \; a_{22}^* = a_{33}$$

H. Theil [20] offers a useful insight into this condition: if "an additional output of one unit of any firm belonging to macro sector J requires the same demand for the products of sector I, we say we have a 'homogeneous input' structure". In such a case the aggregation bias vanishes. Put another way, the aggregated coefficients of a macro sector are not affected by changes in the production pattern within the macro sector. For applications, of course, such a stringent requirement is of little help and the only feasible route is to devise a method to identify 'almost input-homogeneous' sectors as a basis for grouping.

1.2. Some earlier work in this direction was first reported on by Charnes and Cooper [5]. They observe that what is needed is "a measure of closeness of linear transformations," represented by the original and the aggregated A-matrices. A natural choice is to use a matrix norm, for instance the column sum norm:(3)

(5) $$N_c(A) = ||A|| \equiv \underset{j}{\text{Max}} \sum_{i=1}^{n} |a_{ij}|$$

The goal would then be to find A^* which minimizes $||x^* - Sx||$ the aggregation bias; and Hatanaka's exact fit criterion is obtained when it reaches its lower bound, zero. Alternatively instead of choosing a given level of aggregation (M) expressed by S, and finding some best A^* in this sense, one can set an upper bound, say ε , on the admissible aggregation bias, and find the implied conditions on disaggregated sector coefficients to accept or reject any conceivable sector grouping. Following this route Kossov [12] shows that, if we standardize all coefficients for each sector, (setting $\sum_i a_{ij} = 0$ and $\sum_i a_{ij}^2 = 1$ $\forall j = 1,2,\ldots,n$) to eliminate the influence of value added on the size of the coefficients, the implied condition is simply that the correlation coefficient between any two column vectors of A be at least as large as the value of a function in ε and $||A||$. Using the determinant of the correlation coefficient

---

(3) For a different approach to this problem, see W.D. Fisher [8] who uses a Euclidean norm and introduces a loss function to evaluate the aggregation bias.

matrix as an overall measure of goodness of fit, we can seek the aggregation scheme leading to a minimal increase in the value of this determinant. As Kossov himself admits, however, implementing this objective is still far from straightforward since the order of grouping of industries remains to be determined.

In this paper we propose (1) a notion of technological similarity between industries as a basis for grouping, and (2) a class of cluster-analytic methods for algorithmic implementation (Section II). Section III summarizes our findings for the U.S. 83 x 83 1967 matrix with different clustering methods (including Kossov's suggestions) and various assumptions. We also compare our results with the Chenery-Watanabe [6] 29 cluster classification.

II. Cluster Analysis and Technological Similarity

The classical clustering problem (e.g., see Hartigan [10], Anderberg [1]) can be stated as follows: given a set of n objects defined by the values of q attributes x_{ij} ($i = 1,\ldots,q$; $j = 1,\ldots,n$), find M clusters (subsets of the original object set) such that members of a cluster are similar to each other but do not look much like objects outside the cluster. Interpretations of the term "similar" vary with the type of clustering method chosen. All methods, however, are characterized by the following features: a) the criterion (if any) it seeks to optimize; b) the measure of similarity (or dissimilarity) between all pairs of the sample set; c) the algorithm used to find an optimum partition; and d) the interpretation of the clusters.

Hierarchical fusion algorithms were chosen for our analyses. They start with n clusters and reduce them into a single one. They are based on the methods of Sokal and Michener [18] and Ward [23], as implemented by Wishart [24].

Sokal's method is intuitively appealing and resembles that of Kossov: merge at each stage those two clusters with the most similar mean vectors or centroids. If Euclidean distances are used, the closest (in the least-squares sense) centroid vectors determine the fusion. If correlations are used, the centroid vectors having minimum angle determine the fusion. Centroid methods produce "spherical" clusters: centroids of merged groups are weighted to produce a common centroid for the new group (the weights are not necessarily proportional to the number of objects in each cluster); however the (average) similarity value associated with the fusion of two clusters may rise or fall from stage to stage causing a few reversals.

Ward's method adopts Euclidean distances as a measure of dissimilarity. If we define

$$E_p = \sum_{j=1}^{n_p} \sum_{i=1}^{q} (x_{ijp} - \bar{x}_{ip})^2$$

to be the error-sum-of-squares for cluster p, i.e. the within-group-squared deviations about the mean, and $E = \sum_{p=1}^{M} E_p$ to be the total within-group-error-sum-of-squares, then, at each stage, Ward's method finds the two clusters whose fusion yields the minimum increase in E. The function E is non-decreasing and the method is not subject to reversals, i.e. once an object joins a cluster, it is not permitted to leave.

Now to apply these methods to the identification of "quasi-input-homogeneous structures," we consider the n column vectors $(a_{.1}, a_{.2}, \ldots, a_{.n})$ of the original coefficient matrix A as the n taxonomic units. As we know, each such column vector represents the production function of the jth industry in n-dimensional input space (since there is a one-to-one correspondence between the industry set and the product set). The Euclidean distance (correlation) between any two vectors is taken as a measure of overall input-heterogeneity (input-homogeneity) between these two industries. If two sectors are perfectly input-homogeneous, the Euclidean distance between their production function vectors will be zero. As we move away from this extreme case, the distance between the two vectors will increase and their fusion in a common cluster (a "macro-industry") will be postponed until no better (i.e. less distant) candidate for grouping can be found. As the clustering algorithm proceeds, the researcher can stop it at (1) any predetermined level of aggregation, or (2) any level of acceptable aggregation bias. At each step in the fusion process, a set of macro-sectors is determined. At the kth step in the clustering algorithm the aggregation bias $\beta = x^*_{(k)} - S_{(k)}x$ is calculated. Various decision rules can be used to decide when to stop. For instance, for a given c, set k = M whenever there exists an i such that $\epsilon_i = \beta_i/S_{(k)}x_i \geq c$, or $\bar{\epsilon} \geq c$. In this manner, the researcher can make the level of aggregation (the number of clusters) dependent on the mean forecast error or on the maximum error in the final demand bill resulting from aggregation.

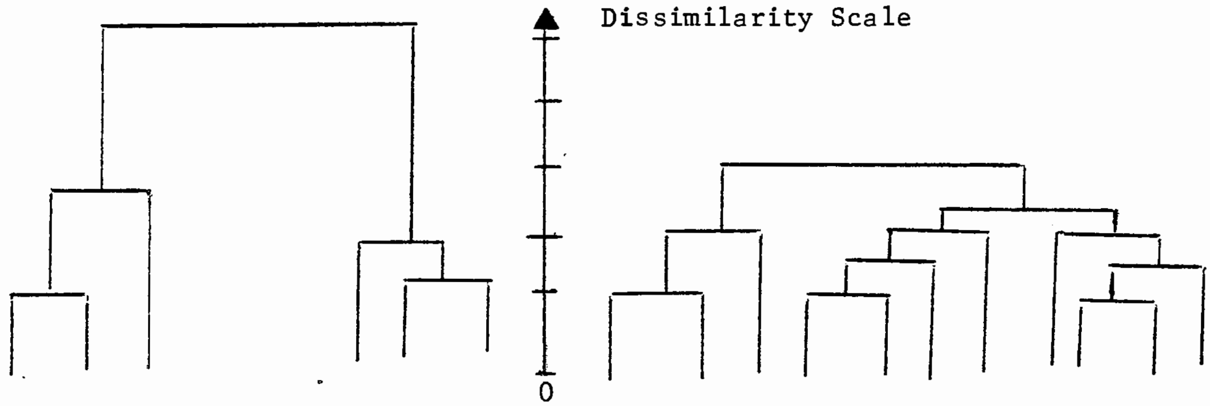
III. Application to the U.S. 1967 (83 x 83) input-output matrix.

3.1. These cluster-analytic procedures for sector grouping have been applied to the U.S. 1967 (83 x 83) direct coefficient table. Successful implementation of these algorithms first requires the choice of a method to correct for the effect of variations in the size of the value added component in each industry.

One approach is to standardize the a_{ij} coefficients as suggested by Kossov, and choose as a measure of technological similarity the correlation coefficient between any two sectors. Alternatively, if a distance function is used as a measure of technological similarity, the value added component is treated as an additional dimension in the production vector of each industry. In any case, it is important to know how sensitive the results are to inclusion or exclusion of the value added component. In view of these requirements, several specifications of two clustering algorithms have been used: (1) Ward's method with the Euclidean distance between sectors as a measure of technological similarity; (2) the centroid method with the correlation coefficient; and (3) the centroid method with the Euclidean distance. Additionally, each of these runs have been performed both with and without the value added coefficient. Our main conclusions are now summarized.

3.2. Dendrogram Analysis

To have a global picture of the order and level of fusion of each sector with the others, it is convenient to set up a dendrogram,i.e. a tree graph in which the distance between a vertex and a node is proportional to the dissimilarity between a given sector and the cluster represented by that node. A basic qualitative property of a dendrogram is the average level of discrimination between clusters. If many "tight" clusters emerge at a low level of dissimilarity, discrimination is facilitated. As fusion of nodes becomes more uniformly distributed over the range of the dissimilarity scale, discrimination becomes much less straightforward since the clusters tend to be diffused in the space. The following graphs illustrate these two cases.



It is intuitively clear that the left-hand graph allows for much easier discrimination between industries than the right-hand one; in the former case, sharp dissimilarities (as measured by the objective function of the clustering algorithm) exist between industries, whereas, in the latter case, overall similarity between all industries is much stronger. Exhibit 1 summarizes the dendrogram obtained with Ward's method applied to the U.S. A-matrix, using Euclidean distances and with value added included (denoted Ward #1).[(4)] We note that a high level of discrimination between sectors exists as the 83 industries merge into only ten clusters for a dissimilarity value of 15 (on a scale ranging from 0 to 260); these tight clusters have a very obvious economic significance and they conform with groupings of industries which we would intuitively feel to be "related". The difference here, of course is that we do not have to rely on intuition or heuristics to decide on this grouping, as we have an objective method to guide the fusion of sectors. A second point worth noting is that, if we compare our results using this method with value added included and with value added excluded, we find that (i) discrimination between clusters is not

(4) Complete dendrograms for each method are available from the authors upon request.

as sharp if we exclude value added--for instance for a dissimilarity level of 15 instead of having 10 clusters there are 19 of them; and (ii) cluster membership seems to obey a vertical integration criterion, as, for instance, in the agriculture-food processing industries sequence. This comparison underscores the need to take account of the value added element by one of the methods we have suggested. Thus, it is interesting to test whether the same results are obtained with a method basing the measure of technological similarity between industries on the coefficient of correlation, which is not affected by the size of the value added coefficients. For this purpose we can compare the dendrogram obtained with the centroid algorithm #6 (i.e. with correlation coefficient as a measure of similarity) with the dendrogram from Ward's method #1 (see Exhibit 1 below): although the discrimination is not quite as high, at the same level of aggregation as before, the clusters are very similar. To facilitate the comparison of the results obtained with the various algorithms and similarity measures, it is helpful to rely on an overall quantitative property of the dendrograms. Hartigan [10] and Sokal and Rohlf [17] have proposed comparing similarity matrices associated with any two dendrograms in terms of distance or correlation coefficient. Table 1 summarizes the Euclidean distances between the similarity matrices obtained for several specifications of the clustering algorithms. Examination of this table reveals that (1) there is, in general, much less difference between algorithms if they treat value added similarly--e.g. both include, or exclude it--than if they treat it differently; (2) the centroid method leads to very close dendrograms whether a distance or a correlation measure of similarity is used. This impression is further confirmed by visual examination of the dendrograms.

Table 1

COMPARISON OF DENDROGRAMS
(Euclidean Distances)

	#1	#2	#3	#4	#5
#2	14,279.4				
#3	15,164.8	1,986.86			
#4	11,746.4	3,970.6	5,010.6		
#5	14,542.3	608.2	1,703.1	4,176.3	
#6	15,204.5	2,002.4	51.6	5,049.2	1,717.1

#1 ≡ Ward/distance/value added (V.A.)

#2 ≡ Centro/distance/V.A.

#3 ≡ Centro/correlation/V.A.

#4 ≡ Ward/distance/without V.A.

#5 ≡ Centro/distance/without V.A.

#6 ≡ Centro/correlation/without V.A.

3.3. Another useful way of analyzing our results is to compare them with the aggregation used by other authors for a prespecified number of clusters. Chenery and Watanabe [6], in particular, have used a 29 x 29 U.S. input-output table for 1947. Table 2 summarizes the results of this comparison using Ward #1 clustering as reference.(5) Examination of this table reveals much overlap between the clusters obtained by Ward #1 method and those defined heuristically by Chenery and Watanabe. In many cases the clusters are in one to one correspondence; and most discrepancies only consist of a split in two clusters in Ward #1 of a single Chenery-Watanabe cluster. These findings support our initial goal of devising an aggregation method that could economically compress micro-information into operationally meaningful sectors.

---

(5) Similar comparisons can be effected with the classifications obtained with the other methods. But, given the smaller discrimination afforded by the centroid algorithm, as noted earlier, it seems best to concentrate on Ward #1.

Table 2

COMPARISON OF 29 CLUSTER CLASSIFICATION
OBTAINED FROM WARD #1 METHOD
WITH CHENERY-WATANABE CLASSIFICATION

Cluster No.[a]	Ward #1	Chenery-Watanabe[b]
1	1,14	14 (Processed Food)
2	2,7,8,9,10,12,35,47, 50,63,65,70,72,74	
3	3,5,6	5,6 (Metal Mining)
4	4,15	1,2,3,4,15 (Agriculture and forestry)
5	11,22,23,32,36,53	23,51,53,54,55,56,57, 58,62,63 (Industry n.e.c.)
6	13,60	13,59,60 (Transport Equipment)
7	16,19	16,19,17 (Textiles)
8	17	
9	18	18 (Apparel)
10	20,21	20,21 (Lumber and Wood Products)
11	24,29	24,25 (Paper and Products)
12	25	
13	26,73	26,73 (Printing and Publishing)

Table 2 (continued)

Cluster No.	Ward #1	Chenery-Watanabe
14	27,28,30	27,28,29,30 (Chemicals)
15	31	31 (Petroleum Products)
16	33	33,34 (Leather and Products)
17	34	
18	37,39,40,41,42	37,39,40,41,42 (Iron and Steel)
19	38	38 (Nonferrous Metal)
20	43,44,45,46,48,49 52,54,61	43,44,45,46,47,48,49, 50,51,52 (Machinery)
21	51,56,57	(included in Cluster #5)
22	59	(included in Cluster #6)
23	66,69,71,76,77	66,67,69,70,71,72,75, 76,77 (Services)
24	67,75	
25	68,78	68,78 (Electric Power)
26	79	Not regular producing industries
27	80	
28	81	Not regular producing industries
29	82	

Note: (a) Numbers in this column refer to the U.S. 83 industry classification as reproduced in Appendix A.

(b) Industry labels in parenthesis conform with the Chenery-Watanabe terminology.

3.4. Our method is also useful for various structural analyses of production structures. This point is illustrated by applying Ward's method to the 1967 (83 x 83) inverse U.S. matrix $(I-A)^{-1}$ to see how, taking account of all direct <u>and</u> indirect links between sectors (as opposed to considering the "direct" production function of each sector i.e. each column vector of A), changes the notion of technological similarity between industries. Groupings of industries as revealed by the dendrogram (see Exhibit 2) are now quite different from those based on direct technological similarity. If we recall[(6)] that each entry $[z_{ij}] = (I-A)^{-1}$ of the inverse matrix measures the total input-weighted interrelatedness between sector i and j, we can interpret each column of the inverse as the <u>interrelatedness vector</u> between the jth sector and all others. It characterizes how interdependent this sector is with the rest of the economy. The clustering algorithm, applied to these vectors, groups together industries whose interrelatedness patterns are most similar. This information forms the basis for a study of inter-industry transmission of business cycles. If a drop in final demand for the products of a certain industry is expected (due, for instance, to an increase in price of some complementary product or a price cut on some substitute), an important question to be asked is: what will be the repercussions of this final demand variation? An analysis of the dendrogram of the columns of the Leontief inverse matrix reveals which group of sectors will react similarly. This common pattern of repercussions will become increasinly blurred as we consider clusters further away from each other on the dissimilarity

---

(6) See Blin and Murphy [4] on this point.

scale. In other words, sectoral multiplier effects are indirectly compared by considering the overall interrelatedness vector for all pairs of sectors. A dendrogram comparison of the direct and inverse matrix reveals certain basic differences. Firstly, as one might expect, complete fusion of all sectors in the inverse occurs at a much higher level of dissimilarity, about 770 on a scale from 0 to 770, as opposed to 260 for the direct matrix; at that level, there are still 40 separate clusters in the inverse matrix, instead of 1 single cluster in the direct matrix A. Secondly, some of the larger clusters now comprise partially integrated sequences of industries as one would intuitively expect given the meaning of the interrelatedness vector--e.g. new construction, heating plumbing and structural metal products, household furniture, other furniture and fixtures, other fabricated metal products. Thirdly, since the clusters are less tight, a breakdown of certain well-defined clusters of A occurs in the inverse; this is directly due to the hidden interdependence between the sectors in the A matrix which is uncovered by matrix inversion. By considering which clusters have been (almost) preserved in the inverse, we can assess how truly interdependent these industries are overall; only if they remain clustered in the inverse can we conclude that the within-cluster interdependence is stronger than the outside-cluster-interdependence. The implications of such findings are many. We may mention, for instance, that it provides a method for testing the well-known hypothesis of economic historians and development theorists that development entails increasing interdependence [25]. This approach is also being extended to cross-sectional input-output data to compare production structures across countries.

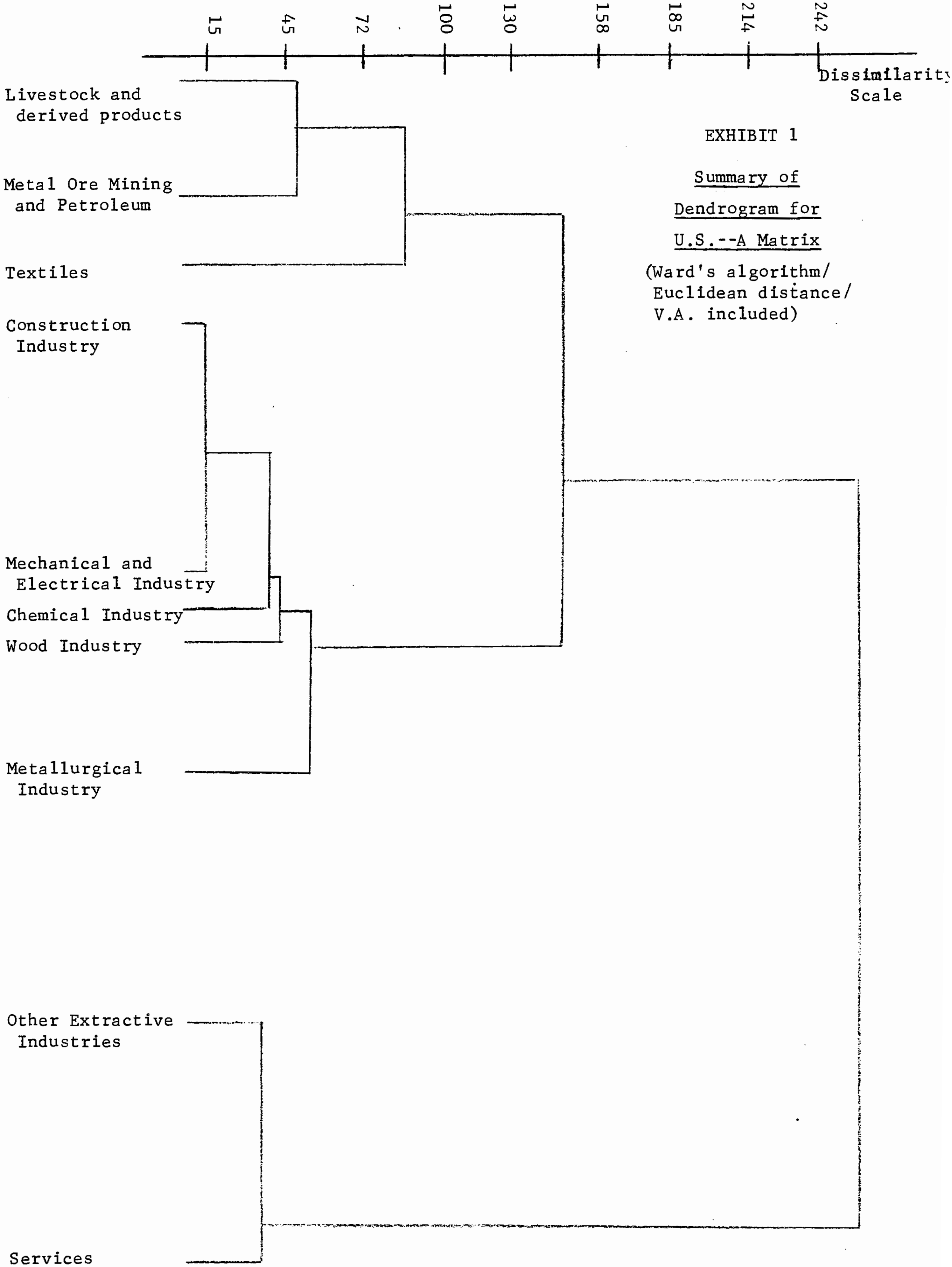


EXHIBIT 1

<u>Summary of</u>

<u>Dendrogram for</u>

<u>U.S.--A Matrix</u>

(Ward's algorithm/
Euclidean distance/
V.A. included)

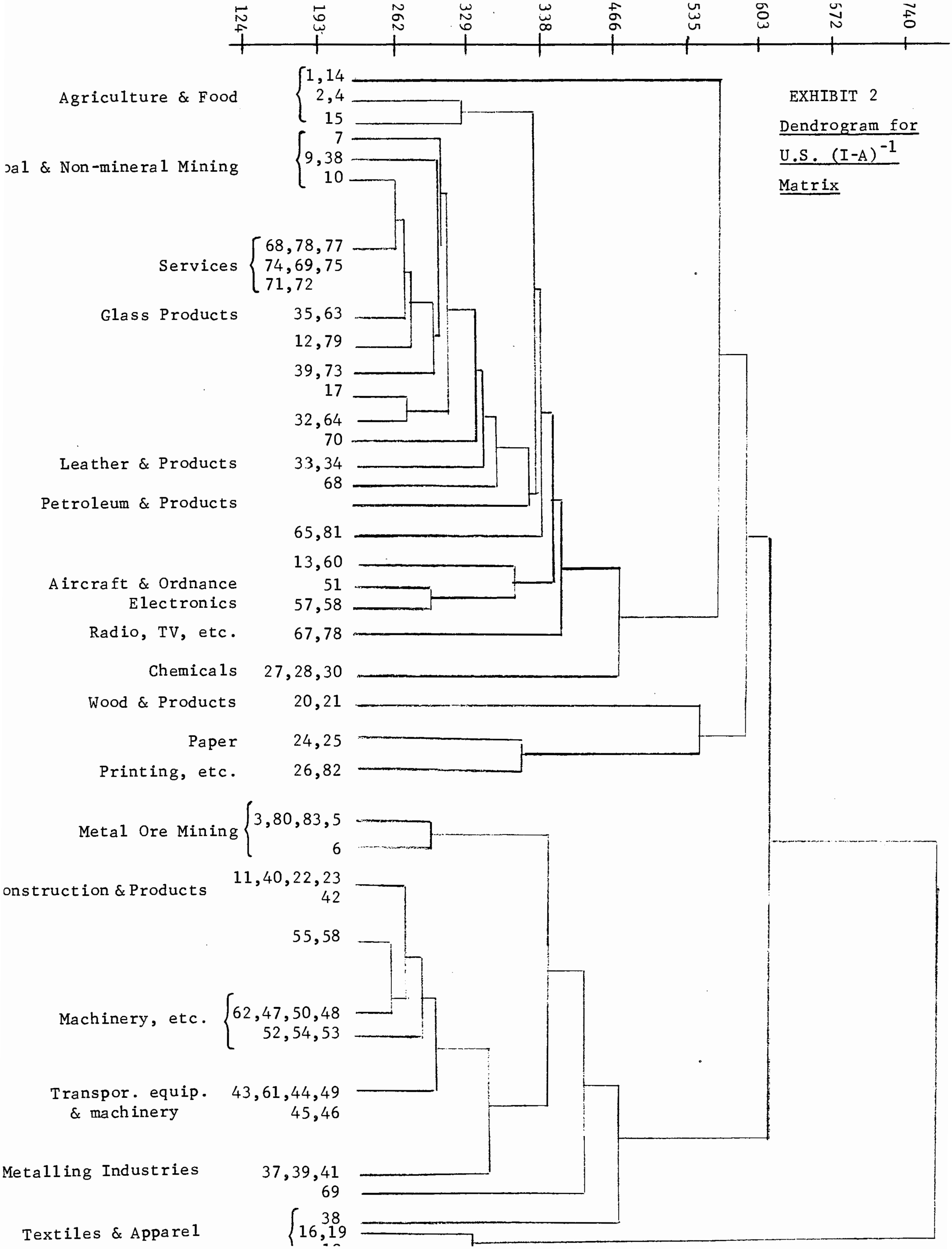


EXHIBIT 2

Dendrogram for U.S. $(I-A)^{-1}$ Matrix

Appendix A

U.S. 83 Industry Classification [22] (1)

Industry Code Number	Industry Name
1	Livestock and livestock products
2	Other agricultural products
3	Forestry and fishery products
4	Agricultural, forestry and fishery services
5	Iron and ferroalloy ores mining
6	Nonferrous metal ores mining
7	Coal mining
8	Crude petroleum and natural gas
9	Stone and clay mining and quarrying
10	Chemical and fertilizer mineral mining
11	New construction
12	Maintenance and repair construction
13	Ordnance and accessories
14	Food and kindred products
15	Tobacco manufactures
16	Broad and narrow fabrics, yarn and thread mills
17	Miscellaneous textile goods and floor coverings
18	Apparel
19	Miscellaneous fabricated textile products
20	Lumber and wood products, except containers
21	Wooden containers
22	Household furniture
23	Other furniture and fixtures
24	Paper and allied products, except containers

Industry Code Number	Industry Name
25	Paperboard containers and boxes
26	Printing and publishing
27	Chemicals and selected chemical products
28	Plastics and synthetic materials
29	Drugs, cleaning and toilet preparations
30	Paints and allied products
31	Petroleum refining and related industries
32	Rubber and miscellaneous plastics products
33	Leather tanning and industrial leather products
34	Footwear and other leather products
35	Glass and glass products
36	Stone and clay products
37	Primary iron and steel manufacturing
38	Primary nonferrous metal manufacturing
39	Metal containers
40	Heating, plumbing and structural metal products
41	Stamping, screw machine products and bolts
42	Other fabricated metal products
43	Engines and turbines
44	Farm machinery and equipment
45	Construction, mining and oil field machinery
46	Materials handling machinery and equipment
47	Metalworking machinery and equipment

Appendix A (continued)

Industry Code Number	Industry Name
48	Special industry machinery and equipment
49	General industrial machinery and equipment
50	Machine shop products
51	Office, computing and accounting machines
52	Service industry machines
53	Electric industrial equipment and apparatus
54	Household appliances
55	Electric lighting and wiring equipment
56	Radio, television and communication equipment
57	Electronic components and accessories
58	Miscellaneous electrical machinery, equipment and supplies
59	Motor vehicles and equipment
60	Aircraft and parts
61	Other transportation equipment
62	Scientific and controlling instruments
63	Optical, ophthalmic and photographic equipment
64	Miscellaneous manufacturing
65	Transportation and warehousing
66	Communications; except radio and TV broadcasting
67	Radio and TV broadcasting
68	Electric, gas, water and sanitary services
69	Wholesale and retail trade
70	Finance and insurance
71	Real estate and rental

Appendix A (continued)

Industry Code Number	Industry Name
72	Hotels; personal and repair services except auto
73	Business services
74	Automobile repair and services
75	Amusements
76	Medical, educational services and nonprofit organizations
77	Federal Government enterprises
78	State and local government enterprises
79	Directly allocated imports
80	Transferred imports
81	Business, travel, entertainment and gifts
82	Office supplies
83	Scrap, used and secondhand goods

Note: (1) There are only 78 regular producing industries; industries 78 through 83 are simply "dummy" or special industries, set up to simplify the estimation procedure. Also sector 74, Research and Development, was eliminated as a separate industry in 1963 on the grounds that "R&D performed for sale is distributed to the purchaser by each of the industries performing the R&D" [19].

References

[1] Anderberg, M. R., Cluster Analysis for Applications, Academic Press, 1973.

[2] Aza, K., "The Aggregation Problem in Input-Output Analysis," Econometrica, 27, 1959, 257-262.

[3] Balderston, J. B. and Whitin, T. M., "Aggregation in the Input-Output Model," Economic Activity Analysis, O. Morgenstern (ed.), John Wiley, New York, 1954, 79-128.

[4] Blin, J. M. and Murphy, F. H., "On Measuring Economic Interrelatedness," The Review of Economic Studies, July, 1974, XLI(3), 437-440.

[5] Charnes, A. and W. Cooper, Management Models and Industrial Applications of Linear Programming, Vol. I, John Wiley, New York, 1961.

[6] Chenery, H. B. and T. Watanabe, "International Comparison of the Structure of Production," Econometrica, 26, October, 1958, 487-521.

[7] Fei, J. C. H., "A Fundamental Theorem for the Aggregation Problem of Input-Output Analysis," Econometrica, 22, 1954, 400-412.

[8] Fisher, W. D., Clustering and Aggregation in Economics, The Johns Hopkins University Press, Baltimore, 1969.

[9] Ghosh, A., "Input-Output Analysis with Substantially Independent Group of Industries," Econometrica, 28, 1960, 88-96.

[10] Hartigan, J. A., Clustering Algorithms, John Wiley, 1975.

[11] Hatanaka, M., "Note on Consolidation Within a Leontief System," Econometrica, 20, 1952, 301-303.

[12] Kossov, V., "The Theory of Aggregation in Input-Output Models" in Contributions to Input-Output Analysis, A. P. Carter and A. Brody (eds.), North-Holland Publishing Co., 1972.

[13] Leontief, W. S., The Structure of American Economy, 1919-1939, 2nd ed., Oxford University Press, New York, 1951.

[14] McManus, M., "General Consistent Aggregation of Leontief Models," Yorkshire Bulletin of Economic and Social Research, 8, 1956, 28-68.

[15] Malinvaud, E., "Aggregation Problems in Input-Output Models," The Structural Interdependence of the Economy, T. Barns (ed.), John Wiley, New York, 1954.

[16] Morimoto, Y., "On Aggregation Problems in Input-Output Analysis," The Review of Economic Studies, XXXVII, 1970, 119-126.

[17] Sokal, R. and Rohlf, F., "The Comparison of Dendrograms by Objective Methods," Taxon, February 1962, 33-40.

[18] Sokal, R. and C. D. Michener, "Statistical Methods for Evaluating Systematic Relationships," Kansas University Science Bulletin, Vol. 38, 1958, 1409.

[19] Skolka, J., The Aggregation Problem in Input-Output Analysis, Czechoslovakian Academy of Sciences, Prague, 1964.

[20] Theil, H., "Linear Aggregation in Input-Output Analysis," Econometrica, 25, 1957, 111-122.

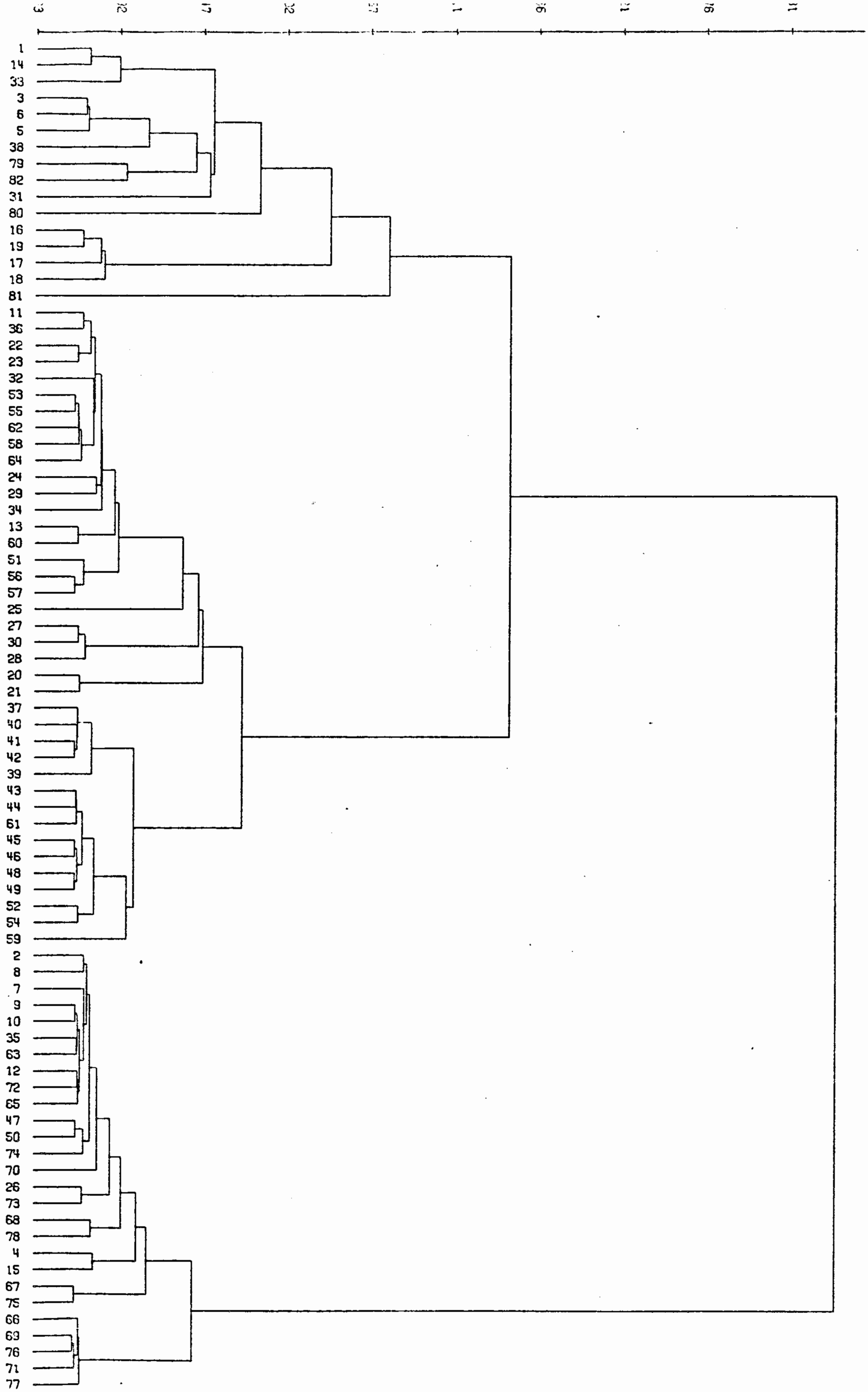
[21] ________, and P. Uribe, "The Information Approach to the Aggregation of Input-Output Tables," The Review of Economics and Statistics, 1969, 451-461.

[22] U.S. Dept. of Commerce, Bureau of Economic Analysis, "The Input-Output Structure of the U.S. Economy, 1967," Survey of Current Business, 1975, 24-56.

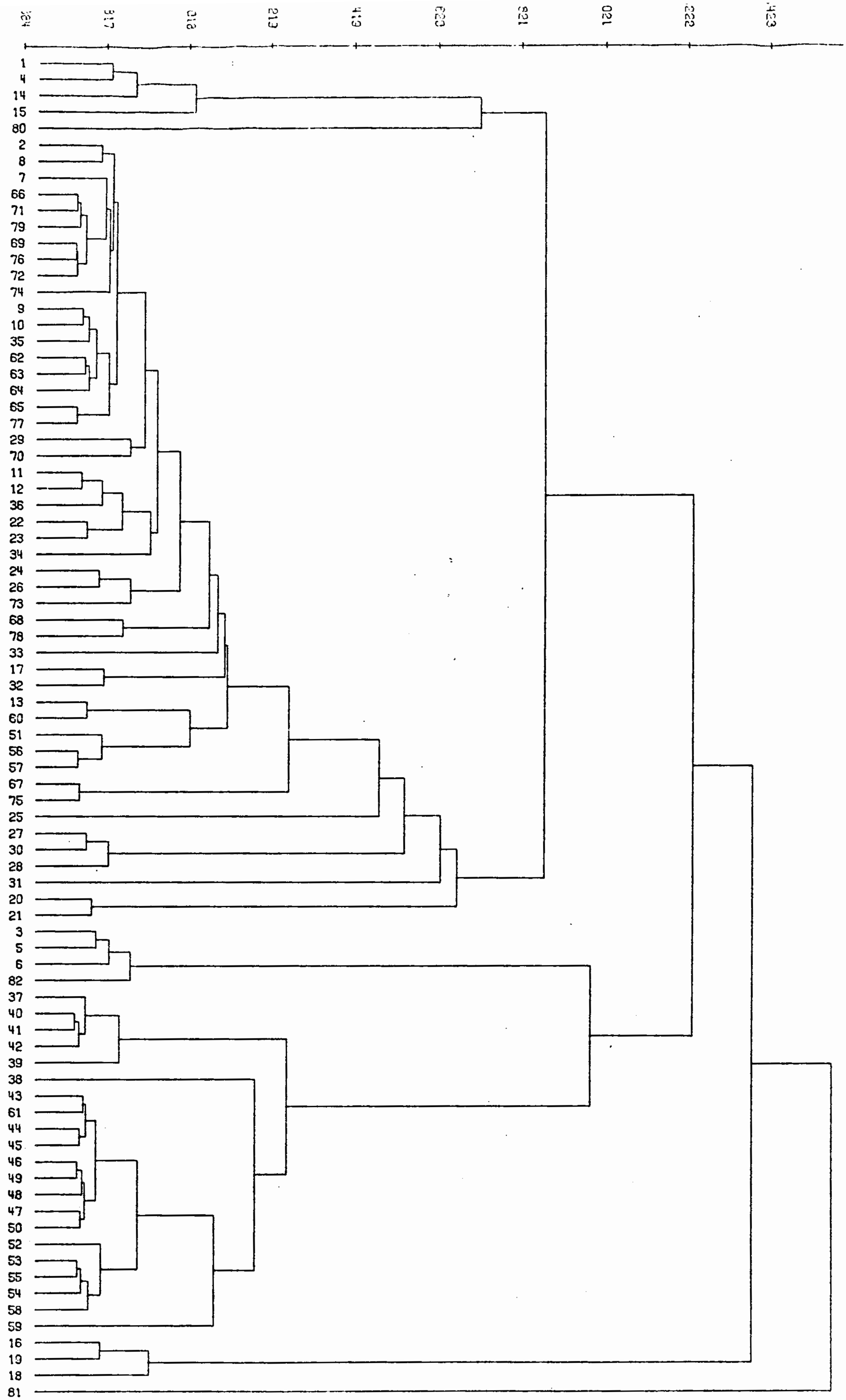
[23] Ward, J. H., "Hierarchical Grouping to Optimize an Objective Function," JASA, 58 (1963), 236-244.

[24] Wishart, D., CLUSTAN 1A, FORTRAN Programs for Cluster Analysis, Vogelback Computing Center, Northwestern University.

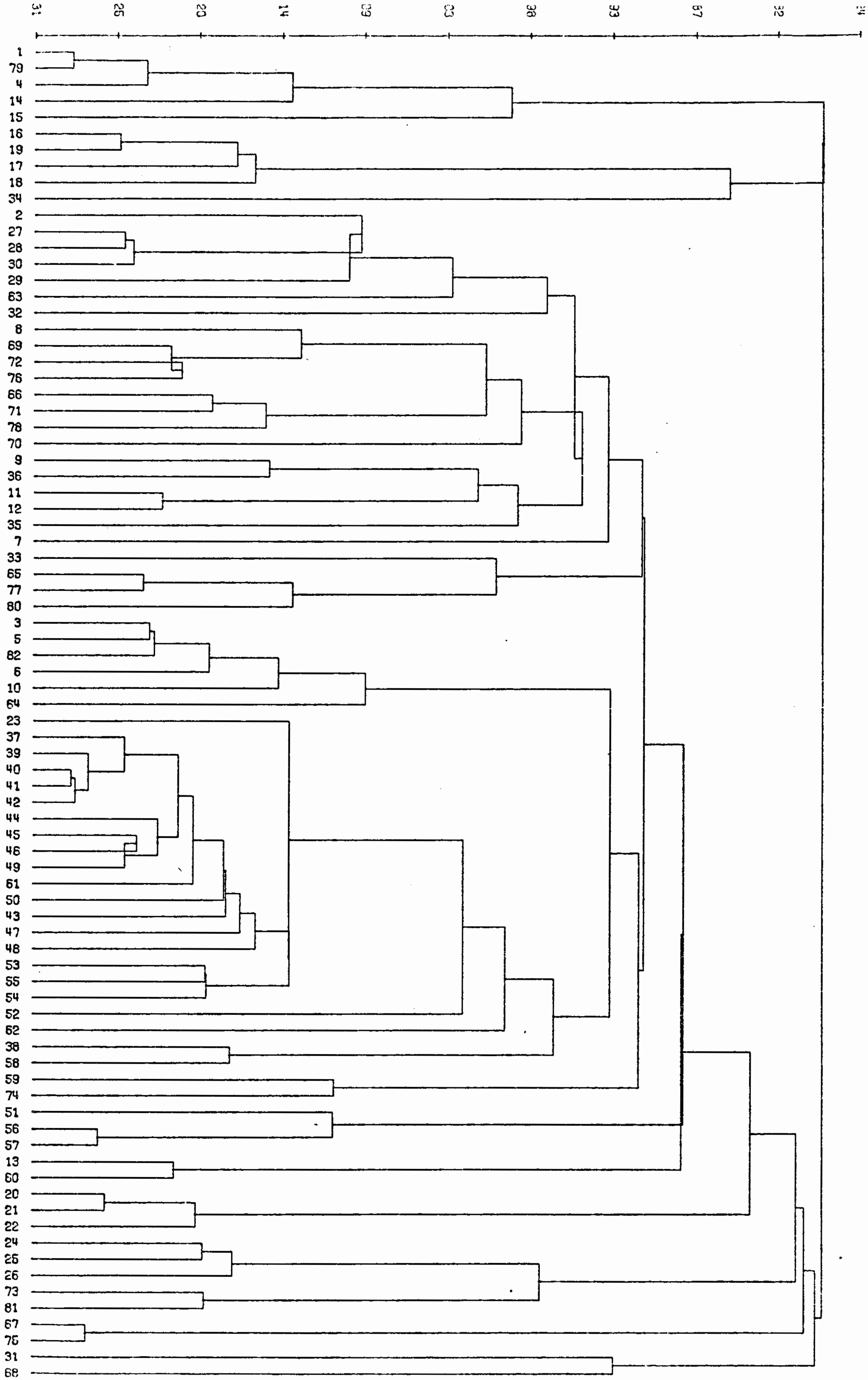
[25] Yan, C. and Ames, E., "Economic Interrelatedness," The Review of Economic Studies, XXXII, October 1965, 299-310.



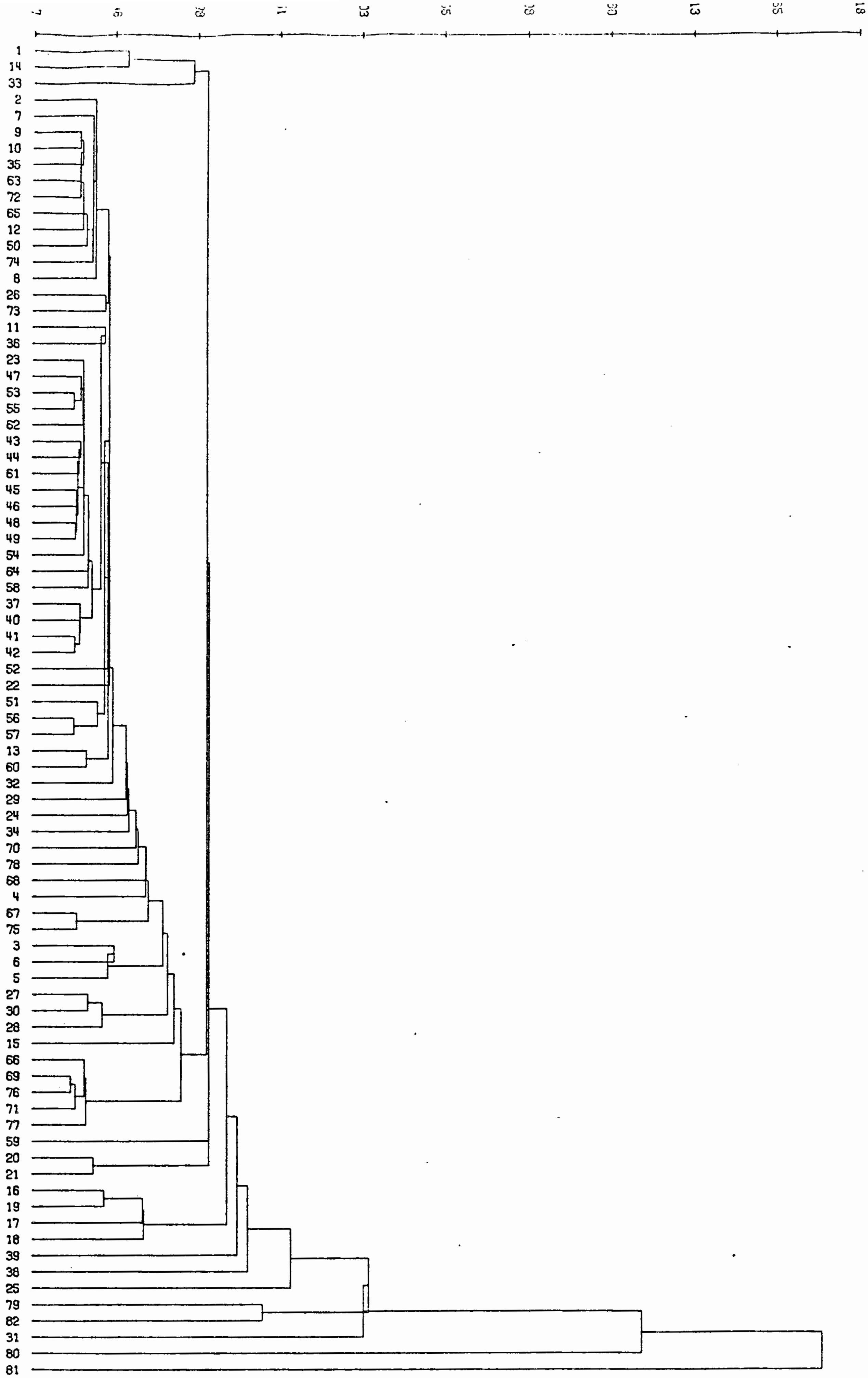
CLUSTERING THE COLS. OF A FOR THE U.S. /DIST/VA INCL.



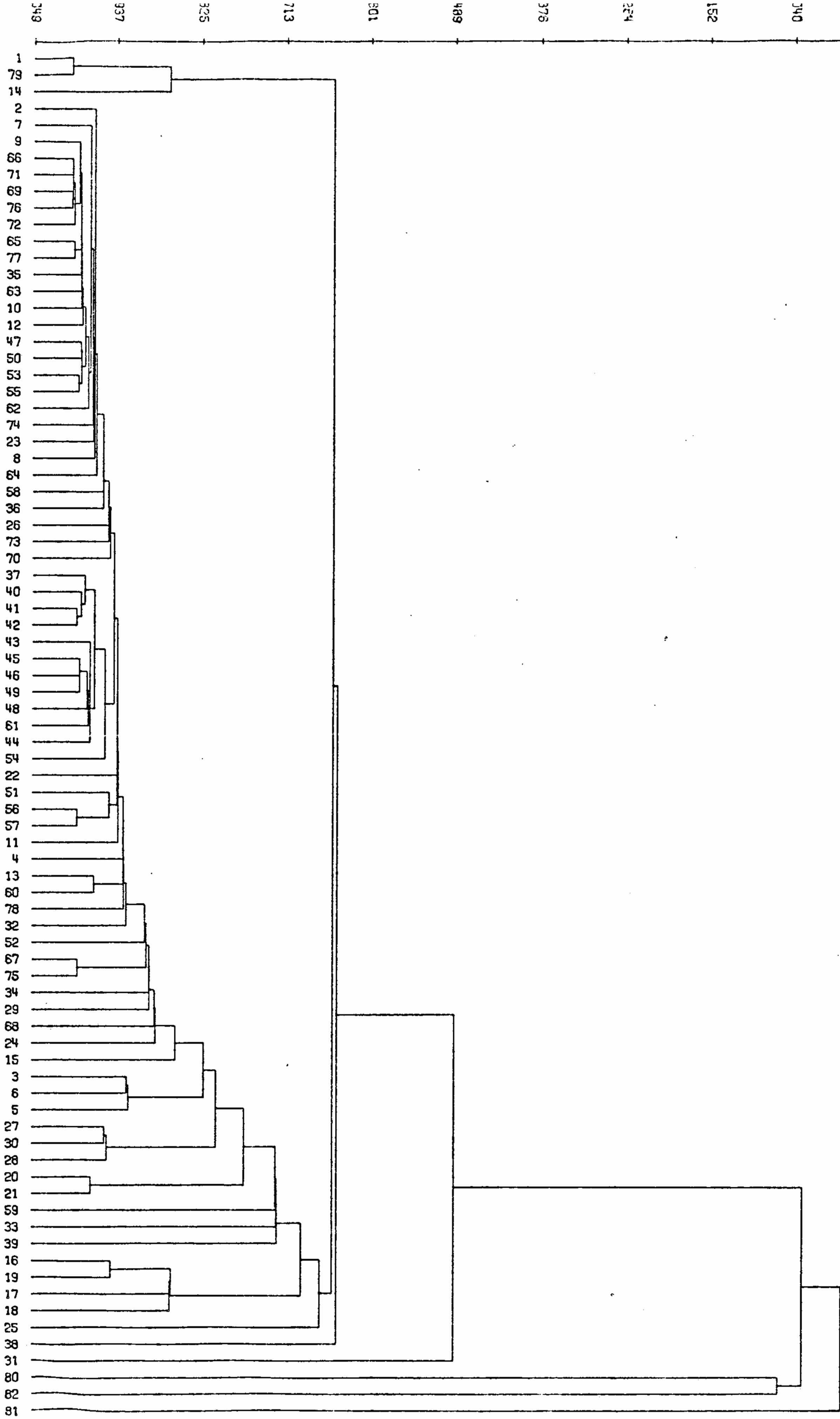
CLUSTERING THE COLS. OF A FOR THE U.S. /DIST/VA EXCL.



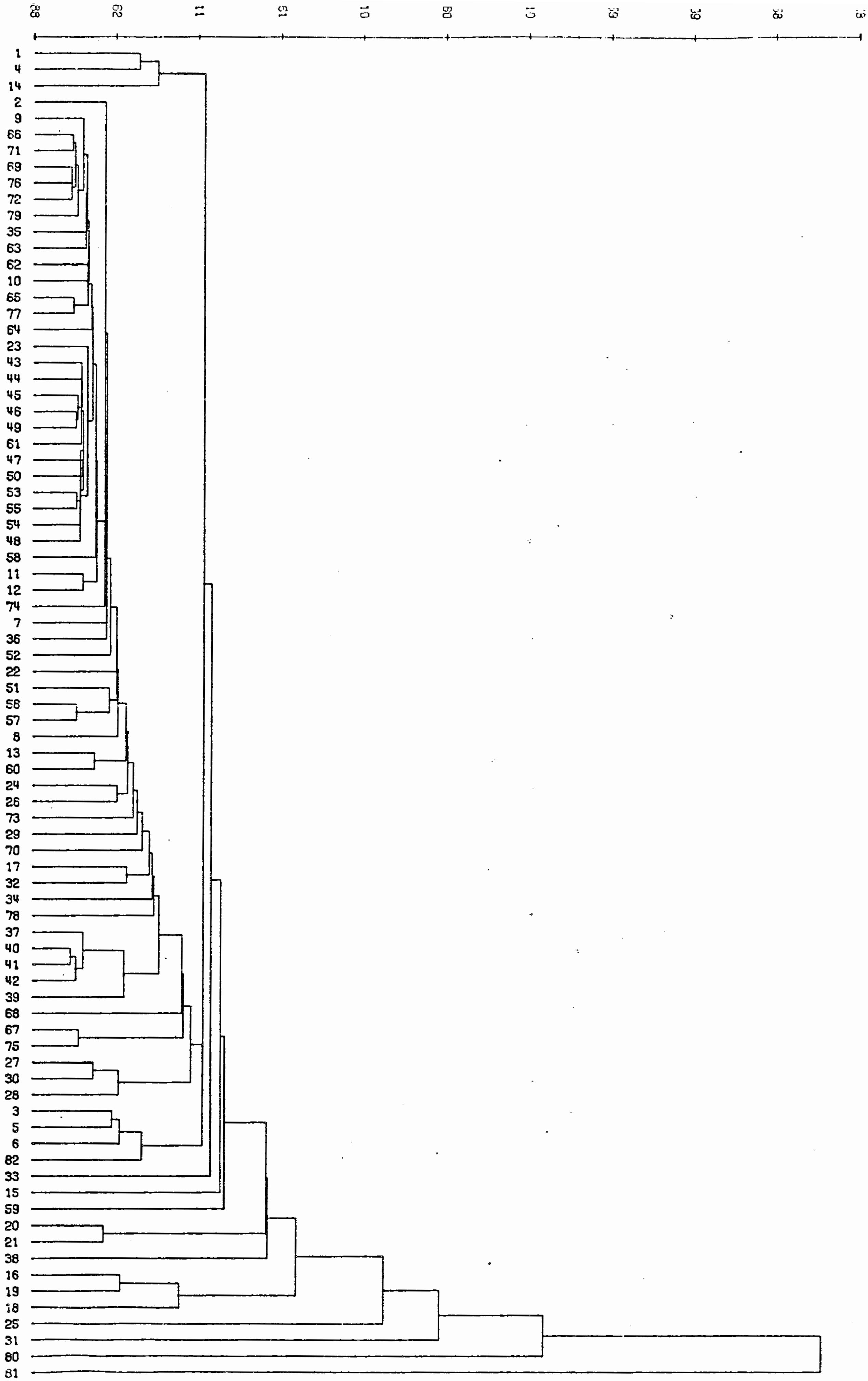
CLUSTERING THE COLS. OF THE A FOR THE U.S. CORR/CENTRO/VA EXCL.



CLUSTERING THE COLS. OF A FOR THE U.S./DIST/CENTRO/V.A. INCL.



CLUSTERING THE COLS. OF A FOR THE U.S./CORR./CENTRO/V.A. INCL.



CLUSTERING THE COLS. OF THE A FOR THE U.S. /DIST/CENTRO/VA EXCL.

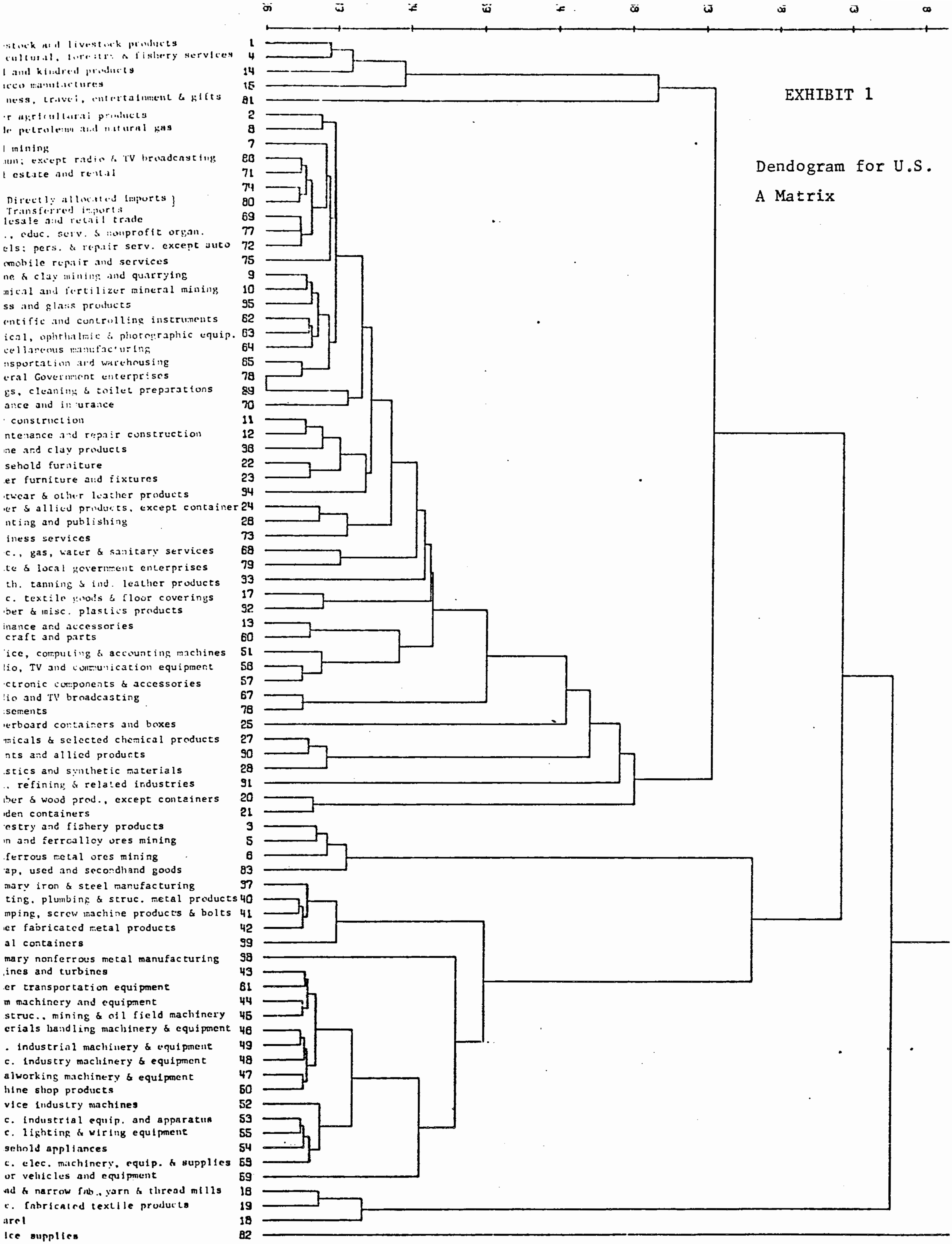


EXHIBIT 1

Dendogram for U.S. A Matrix