

SPECIALTIES AND DISCIPLINES IN THE SOCIAL SCIENCES:
A PRELIMINARY EXAMINATION OF THE STRUCTURE OF THE SOCIAL SCIENCES
USING THE SOCIAL SCIENCES CITATION INDEX[®]

Henry G. Small
Institute for Scientific
Information

Diana Crane
University of Pennsylvania
and Institute for Advanced
Study (1976-77)

Prepared for presentation at the Joint Meeting of The Society for Social Studies of Science and The Sociology of Science Section of the International Sociological Association, Cornell University, Ithaca, New York, November 4-6, 1976.

Concern with mapping the structure of scientific disciplines has been motivated by a number of interests, such as that of understanding the nature of the links between bodies of knowledge and the process and direction of diffusion of knowledge from one area of science to another. Another motivation has been the need for a set of tools for exploring and obtaining an overview of the structure of scientific fields as an aid in selection of specific sub-areas for historical and sociological analysis. Finally, there is a persistent and difficult problem of the relationship between social and cognitive structures in science for which a mapping of disciplines appears to be essential.

Examination of these types of problems requires an enormous data base, much larger than that which it is feasible to generate by means of interviews and questionnaires. Such a data base is provided by the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI).¹ The work reported in this paper had its origins in clustering experiments carried out on the SCI data base using co-citation as a measure of similarity between highly cited documents. From extensive clustering studies of five consecutive years of this file (1970-74) the primary structural unit which emerges in the natural sciences is the narrow subject matter specialty (Small, 1977; Small and Griffith, 1974; Griffith et al, 1974). The mapping of inter-specialty relationships has already been used to examine the structure of the natural sciences (Griffith et al, 1974).

Co-citation analysis provides a means of obtaining an overall view of the interrelationships between scientific areas. In order to understand the significance of the links which occur, other methods such as interviews and content analysis of documents need to be used. By itself, co-citation analysis can suggest relationships which require further exploration using other techniques.

Studies of this kind have implications for understanding the processes underlying the development of knowledge in the social sciences. There is a perennial debate concerning the development and growth of knowledge in the social sciences. Some writers have argued that the growth of social science knowledge is fundamentally different from that of the natural sciences due to a lack of consensus concerning research problems and approaches in the former. It has been suggested (for example, by Price, 1970) that the social sciences tend to draw upon a wide range of materials in the development of new ideas. This mode of utilization of knowledge has been referred to as archival and has been contrasted with the way in which the natural sciences tend to build rapidly upon highly specialized segments of the previous literature. If this is the case, one would expect to find fewer clearly differentiated research areas in the social sciences. At the extreme, the entire literature would be loosely interconnected.

In addition, some writers have developed models of the relationships between research areas in the natural sciences (Holton, 1973; and Mulkey et al., 1975). They suggest that there is a proliferation of new subjects around core areas. In other words, growth of knowledge occurs because theory or techniques which have been useful in one area are subsequently applied to new areas. Thus, if the social science literature is behaving like the natural sciences literature, one ought to be able to observe networks of research areas that are tightly interconnected and that reflect a proliferation of new subjects around core areas.

It is less clear what one might expect concerning the relationships between disciplines, however. Some natural science disciplines are highly interconnected such as the biological sciences. Others such as physics and mathematics are more independent of other disciplines.

We will examine the connections between subject areas in economics, psychology, and sociology as one indication of the extent to which knowledge is diffusing among these disciplines.

This paper presents some of the results of a preliminary investigation of the Social Sciences Citation Index data base (in addition, see Griffith and Small, 1976). We are attempting to answer the following questions: (1) What does the structure of specialties in social sciences imply about the development of knowledge in those fields? (2) Can disciplinary boundaries be discerned between groups of social science specialties? (3) If so, are some disciplines more tightly linked internally than others? Our analysis is primarily concerned with economics, psychology, and sociology.

Methodology

The system which is used to locate research specialties in this analysis has been described in detail elsewhere (Small and Griffith, 1974; Griffith et al., 1974). Three years of the SSCI were merged for the purpose of this study to form a special three-year (1972, 1973, 1974) cumulative file. This file included 1,895,298 citations and 1,170,280 cited items. All items cited 10 or more times in this three-year file were selected ^{2/} for clustering. Since most scientific papers are cited less than five times in their entire history, this means that only very frequently cited papers are being examined. The idiosyncratic citation behavior of some scientific authors is likely to have little effect upon the patterns being observed here.

By examining the citations that link scientific papers, the clustering program identifies research clusters which consist of a relatively small group of co-cited documents and a larger group of documents,

each of which cites two or more of the co-cited documents. These two sets of documents are believed to form the basis of a cluster or research area, a group of researchers engaged in the joint exploration of related problems. Examination of the literature and interviews in a few of the areas seem to confirm this assumption. Informants in two large social science clusters were able to pinpoint major papers among the co-cited documents as having provided the impetus for much of the work in the area. A questionnaire study by Small (1977) of a cluster in the natural sciences (collagen research) revealed the existence of a small group of researchers who were regarded as leading researchers in the area and had extensive informal contacts with one another. A recent study by Mullins et al. (1976) also found considerable evidence that two other clusters in the natural sciences represented groups of researchers who were aware of one another, collaborated with one another, and communicated with one another.

While most of the clusters that emerge in the SSCI file represent substantive areas, a few represent methodological topics such as path analysis and multidimensional scaling. In such cases, groups of researchers are presumably involved in perfecting these techniques. Labeling the clusters appropriately is a difficult task. In general, phrases which recur frequently in the titles of the citing documents have been used but our titles may not be the ones which are in current use among the researchers concerned.

Macrostructure, the structure of links between clusters, is examined using the number of papers citing documents in both clusters.³ Until a detailed analysis of the actual nature of the intellectual links which are represented by co-citing documents in different specialties is undertaken, it is difficult to assess whether a small number of co-citing documents between specialties has very different implications from a

a much larger number. We will assume that the larger the number of co-citing documents, the greater the amount of transfer and exchange of information within and between specialties. We have not as yet attempted to examine the nature of the intellectual exchange that is represented by these linkages. Certainly a somewhat different picture of links would ensue if we examined only unidirectional links from one cluster to another (i.e. including 'unreciprocated' links). It is also possible that some of the links that ensue when the number of papers citing documents in both clusters is examined are not indicative of shared research interests since the linkages may be to techniques or 'classical' papers which are cited pro forma. In other words, further work is needed to assess the content of the links between clusters.

In the following pages, we will compare three social science disciplines with one another and implicitly with the natural sciences. We are unable as yet to make direct comparisons with the natural sciences since we have not yet analyzed the Science Citation Index file using the same techniques that have been used here. However, since certain statistics describing the numbers and types of clusters are similar for the SCI and SSCI files, the indirect comparison seems reasonable.

We will begin by making comparisons between the three social science disciplines using statistics describing the characteristics of the clusters and the numbers of links between them. We will then use connected graphs to examine the structure of the links and to identify different types of clusters which will in turn be compared using a variety of quantitative measures. Finally, we will use multidimensional

scaling, a technique which can be used to assign documents or clusters of documents positions which reflect their proximity to one another (Kruskal, 1964). This technique provides one means of assessing the relationships between clusters in the file.

Differences between Disciplines

Our first question was whether clusters would emerge from the SSCI file as they had in the SCI file. In fact, 1208 clusters were formed, capturing approximately 46% of the cited items at the initial threshold. The mean size of the clusters is 5.3 cited documents and 55.9 citing documents, while the largest cluster in the file consists of 119 cited documents and 892 citing documents.

Most of the clusters are relatively small. Of the 1208 clusters in the SSCI file at a level of 16%, 517 are cited by less than 30 documents, 548 by 30-99 documents and 143 by 100 or more documents. At any given clustering level, small clusters (less than 30 documents) could represent fragments of larger areas at that same level, the tips of areas which would emerge at lower levels or possibly genuinely isolated groupings. Clusters cited by less than 30 documents are excluded entirely from our analysis since we believe that they are too small to be meaningful indicators of a research or specialty area.

In order to examine differences between clusters and networks of clusters in different disciplines, it was necessary to be able to classify clusters by discipline. For this purpose, we classified the journals in our file using the disciplinary classifications which appear in Ulrich's International Periodicals Directory. We then classified clusters as belonging to a particular discipline if two-thirds or more of the citing articles in the cluster were published

in journals in that discipline (hereafter the disciplinary journal concentration or DJC). A cluster was classified as interdisciplinary psychology, interdisciplinary economics or interdisciplinary sociology, if from one-third to two-thirds of the citing articles were published in journals in that discipline. Surprisingly enough, there was virtually no overlap among these categories (see Table 1). Using these criteria, only one sociology cluster emerged. A substantial number of interdisciplinary sociology clusters were identified, however.

Table 1 About Here

Economics and sociology contained markedly fewer clusters than psychology (see Table 1). This was surprising since the numbers of journals in these disciplines was not substantially lower than in psychology. In addition, there were more clusters in economics and interdisciplinary economics than in interdisciplinary sociology although there are fewer economics journals than sociology journals in the file.

Clusters in these disciplines were on the average substantially smaller than the psychological clusters (see Table 2). Fifty percent of the clusters with 100 or more citing documents were located in psychology and interdisciplinary psychology. Only eight percent of such clusters were located in economics and sociology. In addition, the average numbers of citing documents in clusters in these disciplines were substantially lower than in psychology.

Table 2 about here

Table 2 presents some statistics on the size of clusters in the three disciplines (here we have combined interdisciplinary economics with economics and interdisciplinary psychology with psychology), the age of the items constituting the cited document clusters, and the fraction of cited documents which are books. Cluster age is measured in two ways: by the mean year of publication of cited documents in a cluster and by Price's Index, the percentage of cited documents published during the last five years (price, 1971). For all three disciplines the mean year of publication is substantially older than the mean year of publication for the twenty largest clusters in the 1973 SCI which is 1969.9. Cited publications are clearly most recent in psychology, followed by economics and then sociology. This ordering is preserved when cited books are distinguished from cited articles in almost all instances. Price's Index gives similar results to that of the mean but discriminates more between disciplines and between books and articles. Presumably the more recent the mean date of publication (or the larger Price's Index) the more rapidly new knowledge is being created and old papers are being rendered obsolete.

The percentage of cited items in clusters which are books⁷ is markedly higher for sociology than for economics and economics is higher in turn than psychology. The high percentage of books in sociology does not explain the greater age of cited documents in that field since both articles and books are older than in other disciplines. Reliance on journal articles rather than books is characteristic of the natural sciences and is believed to be a factor in the speed with which knowledge grows and changes in those disciplines.

These statistics suggest that the structure of psychology is most like that of the natural sciences while the structure of sociology is least like that of the natural sciences. A more intensive analysis of economics and sociology suggests additional differences between them. If the DJC (the fraction of papers citing a cluster that appears in journals categorized in

that discipline) can be regarded as a measure of the 'discipline-like' character of a cluster, then it appears that economics behaves more like a discipline than sociology, since it has 16 clusters in which two-thirds of the citing papers are in economics journals compared to only one cluster in which two-thirds of the citing papers are in sociology journals (see Table 1).

This effect can be seen in a more systematic way in Figure 1, where all clusters regardless of size and containing any number of papers in the disciplinary category have been considered. The horizontal axis shows the percentage of articles (grouped in 10% intervals) appearing in journals categorized in the discipline (either economics or sociology). The vertical axis is the percentage of all clusters in the discipline.

The Figure shows clearly that the economics literature is more internally cohesive in terms of appearance in a set of journals, than the sociology literature. In fact, the economics curve has what might be termed a disciplinary "hump" centered on the 60-70 percent range of the DJC. The sociology curve, on the other hand, decreases in a quite regular fashion, with no evidence of special cohesiveness of literature published in sociology journals. Also noteworthy is that fact that the economics curve shows a local minimum in the range of 30-40 DJC, giving some post hoc justification for the cut-off of one-third in the identification of clusters in the interdisciplinary and disciplinary groups discussed earlier.

This analysis of DJC suggests that economics has a higher degree of internal connectivity than sociology. Another measure of the 'discipline-like' character of a set of clusters is the degree to which they are connected to one another. However, the density of the matrices showing relationships between different clusters in the two disciplines is not very different. Twenty-nine percent of the possible ties between economics clusters occur compared to 33% of the possible ties between sociology areas. If one examines the average strength of these ties in terms of numbers of co-citations linking areas, the differences are again not very great (2.77 for economics and 2.68 for sociology).

Networks of Clusters

Is there any evidence in these disciplines that some clusters are closely interrelated in a manner that suggests that these clusters are providing the impetus for the emergence of others? We examined this question using connected graphs.

At a minimum of three co-citations between clusters, the economics and inter-disciplinary economics clusters form one large network which includes all but three clusters in the field (see Figure 2). This network consists of two large groups of clusters linked by relatively strong ties (7 co-citations or more) and two small groups of strongly-linked clusters. These groups of tightly linked clusters are all loosely linked to one another and to some peripheral clusters. Each of the groups of clusters contains one or more 'nodal' clusters which are linked to at least 5 clusters in the discipline. One of the large groups of clusters includes areas dealing with investment policies, monetary theory and capital theory. The other large group of clusters includes areas dealing with consumer economics, public goods, income distribution and pollution control. The smaller groups of clusters deal with consumer demand and market structure.

8

When linkage between clusters in interdisciplinary sociology is examined using a minimum of 3 co-citations between clusters, sociology also forms one large network from which only one cluster is excluded. The network consists of three large groups of clusters, each containing a few tightly linked clusters and several peripheral ones. Again, there are nodal clusters which form the cores of these groups of clusters. The large groups of clusters are concerned with: (1) population control and marriage and the family; (2) social and occupational mobility and path analysis and; (3) labeling and deviance. The small group of clusters

deals with religious involvement and belief. Again, the groups of clusters are loosely linked to each other.

This analysis of connected graphs suggests that within each of these disciplines there are sizable networks of closely inter-linked clusters. Although further analysis would be necessary to confirm it, we hypothesize that these networks represent segments of the social sciences which behave like some parts of the natural sciences in terms of the development and utilization of knowledge. In other words, we are suggesting that these areas are organized around clearly defined research goals and exploit the implications of new ideas rapidly.

Certain characteristics of these areas were examined to provide some indirect evidence for this hypothesis. For example, we would expect that what we have called nodal clusters, those that have the most links with other clusters, would contain more cited and citing documents than those that are less tightly linked to other clusters. In other words, these are the most active areas of the discipline, presumably because they contain ideas and problems which are of interest to many members of the discipline. We would also expect that the proportion of books among cited documents would be lower in these fields since the journal article is a more suitable medium for the rapid transmission of findings in a fast-moving field.

We also suggest that nodal clusters represent areas which are providing ideas that are being used in other areas. Adjacent areas may in fact have broken away from these areas. Without doing detailed case studies of these areas, some of which we plan to undertake in the future, some indication of this can be seen from differences in mean dates of publication of cited items in these fields. Mean dates of publication of cited items in nodal areas should be earlier than mean dates of

publication for cited items in clusters which are tightly linked to such nodal clusters.

Table 3 about here

Classifying clusters as nodal, satellite or peripheral,¹⁰ revealed that economics had almost twice as many nodal clusters as sociology. This suggests that even though the overall internal connectivity in that discipline is not higher than that of sociology, certain segments of the discipline are more tightly connected. As expected, nodal areas in economics tended to be substantially larger than satellite and peripheral areas (see Table 3)¹¹ and the percentage of books among cited documents was substantially lower. The mean date of publication of cited documents in nodal clusters was older than that of satellite clusters, suggesting that the latter may have been derived from the former. When these data were examined separately by network, this hypothesis was confirmed in two of the networks and not in the other two. The mean date of publication of cited documents in the peripheral networks was as early or earlier than that of the nodal clusters, a fact which in addition to their smaller size, suggests that these are areas that have not given rise to new developments.

In sociology, the nodal clusters are also substantially larger than the satellite and peripheral clusters but there is little difference in the percentage of books among cited documents in these different types of clusters. Mean date of publication in the nodal areas somewhat preceded that of satellite clusters and a separate analysis by network confirmed this in two out of three cases. Mean date of publication of cited documents in the peripheral clusters was considerably older, again suggesting that these are areas which have not given rise to new developments.

Obviously we are dealing here with very approximate indicators of what is taking place inside these disciplines. Co-citation analysis does, however, suggest areas whose development over time would be worthwhile to study. A long-time member of one of the larger networks in sociology described some aspects of this process of proliferation and growth in the area of family planning. According to him, this area has influenced the development of at least two other areas. In one case, the new area was drawing upon both theory and data from the older area; in the other, just data were borrowed.

Results of Scaling Disciplinary Structures

As a further step in the internal analyses of the economics and sociology cluster sets, nonmetric multidimensional scaling was performed on the cluster co-citations for each set. The two dimensional configurations for each are presented in Figures 3 and 4. Earlier applications of this technique to co-citation data have yielded results which are readily interpretable as maps of a cognitive space, shared by a community of researchers. The economics and sociology maps are no exceptions. In each case, the clusters have been organized on the basis of the cluster co-citation strengths into subject regions which, with more or less precision, contain clusters which have cognitive affinities.

Figure 3 about here

For economics, a triangular array is formed in which the elongated apex comprises an 'international economics' region consisting of clusters concerned with international trade, manufactured exports, and trade models. Moving down the triangle and concentrated to the left of center is a region concerned with the economics of industrial organizations and technological change. Clusters defining this region are concerned with economic regulation, market structure and large firms. Below and to the right of industrial economics we find a region on economic growth and developing countries. Below that and to the left is a region dealing with capital and monetary theory. Central to the triangle, slightly to the left and below the latter, is a group of clusters on business finance and investment. The lower left corner of the triangle is concerned with natural resources and, in a belt surrounding the corner, various topics in welfare economics (property rights, income distribution, public goods, and local government taxes). The lower right-hand corner of the triangle appears to be a continuation of consumer economics with some emphasis on labor supply and demand and pure exchange economies.

In summary, the major axis of variation seems to lie from micro-economics at the bottom of the triangle where clusters are concerned with individual and people-oriented problems, to macro-economics at the top of the triangle where clusters are concerned with industries, economic growth and international trade. Alternatively the axis could be considered an individual to local to national to international dimension. Along the base of the triangle, we have very roughly a transition from liberal to welfare economy to conservative or free market economics, although this trend from left to right does not hold in all regions of the figure.

Clusters in sociology seem to arrange themselves primarily in terms of methodological differences which is not surprising in a discipline where methodological controversies are so widespread. At the top of Figure 4, are three clusters which use demographic methods involving census data. Below are a series of clusters which use survey methods. Some of these use high-powered statistical techniques such as multiple regression and path analysis. The bottom third of the Figure includes areas which primarily use historical and case study methods, such as the sociology of religion and labeling and deviance.

Figure 4 about here

Associated with these regions are long-standing differences in theoretical orientations. Roughly the top two-thirds of the Figure corresponds to what Mullins (1973) has called "standard American sociology" which stands for both a functionalist theoretical approach and the use of survey methodology. The newest form of it, "new causal theory" (what Mullins (1973) calls an elite specialty because it has won ready acceptance in the discipline) is represented by path analysis and ordinal path analysis.

The lower third of the Figure includes a variety of theoretical orientations all of which share a commitment to historical or case study methods. Social interactionism with its social psychological emphasis underlies much of the work on labeling and deviance. Ethnomethodology which is allied in many ways with symbolic interactionism appears as 'phenomenological sociology'. Since the opposition between symbolic interactionism (and now ethnomethodology) and the more empirically oriented functionalist theory has characterized sociology for almost three decades, it is not surprising that it emerges here.

Relationships between Disciplines

A final question of interest to us here is the extent to which these disciplinary networks are linked to other disciplines. There is some indication that sociology clusters have more links with clusters in other fields than do economics clusters. When each cluster is taken as a starting point and all clusters directly linked to it by a cluster co-citation strength of 7 or more are recorded (i.e. its nearest neighbors), Table 4 shows that the mean number of 'nearest neighbors' per cluster is larger for sociology than for economics, indicating that, overall, sociology clusters are involved in more links than economics clusters. However, 42% of sociology's links are to clusters outside the sociology set, while 70% of economics' links are to clusters within the economics set. Therefore we may conclude that economics behaves in a more insular fashion and that sociology is more interconnected with other subject areas. It also suggests that economics is more cohesive as a discipline than sociology.

Table 4 about here

Most of the links to clusters outside economics are weak (i.e. less than 7 co-citations). At a minimum of 3 co-citations between clusters, economics and interdisciplinary economics have ties to 33 clusters outside the field but only two of these belong to interdisciplinary sociology and four to interdisciplinary psychology. On the other hand, in the group of clusters dealing with investment policy, four clusters outside the discipline were strongly linked to the clusters in the field and were apparently important elements in it. A similar phenomenon occurred in groups of clusters dealing with

market structure, consumer demand and consumer economics. In the last group of clusters, a cluster from another discipline was a nodal cluster but this was the only example of a cluster from another discipline performing such a role.

At a minimum of 3 co-citations between clusters, interdisciplinary sociology is linked to 84 clusters outside the discipline. Although not examined systematically, many of the links to fields outside sociology appear to be to clusters in anthropology, political science, and psychiatry. There were seventeen links to interdisciplinary psychology but only one to psychology. Most of these are weak links which disappear when the minimum co-citation level is raised to seven. At this level, interdisciplinary sociology and psychology are only linked indirectly through a couple of 'multi-disciplinary' areas. Each of these areas has small components of psychology and sociology materials (judging from the journals in which the citing references appeared) but not enough to characterize them as belonging to either of these disciplines.

What role did clusters outside sociology seem to be performing in the network? There was no example of a cluster from another discipline that performed a nodal function within a group of sociology clusters. A few clusters from other disciplines seemed to perform the function of linking groups of clusters; these external clusters had several links in the discipline.

The relative absence of links between clusters identified with journals in sociology, economics, and psychology would seem to indicate that information and ideas do not diffuse readily between these three disciplines (see Table 5). It seems clear from these data that economics is more closely allied with fields such as management, law and political science than with sociology and psychology. Sociology has weak affinities

with interdisciplinary psychology but very few with psychology itself. Not surprisingly, psychology has many relatively weak ties with interdisciplinary psychology.¹² Further interpretation of these findings must await a detailed examination of the nature of the links which occur and of the characteristics of the isolated clusters.

Table 5 about here

Conclusion

Co-citation analysis may help to clarify a perennial debate concerning the development of knowledge in the social sciences because it provides information about the interrelationships between hundreds of research areas in the social science disciplines. Is the process of growth in the social sciences similar to that of the natural sciences? If it is similar, we would expect to find both research clusters representing groups of researchers pursuing closely related problems and groups of tightly connected research clusters representing a proliferation of new subjects around core areas.

1208 research clusters were identified in the SSCI file. When these clusters were assigned to economics, psychology, or sociology on the basis of the proportions of citing articles in the clusters which were published in journals in those disciplines, it was clear that the psychology literature shows the greatest propensity and the sociology literature the least propensity to cluster. There were 116 psychology journals in the file; 277 clusters were identified with that discipline. There were 70 economics journals but only 44 economics clusters. Similarly there were 85 sociology journals but only 36 clusters identi-

fied with that discipline.

In all three disciplines, cited publications were less recent than in the largest SCI clusters, a probable indication that new knowledge is being created less rapidly in the social sciences than in the natural sciences. Cited publications were most recent in psychology, followed by economics and then sociology. Another indication of rapid growth is reliance on journal articles rather than books. Here again psychology was most like the natural sciences followed by economics and sociology.

In both economics and sociology, groups of clusters were identified which appear to exhibit a high degree of connectivity and to be subdividing into new areas over time. Within these groups, certain clusters are clearly central and probably provide a source of theory, techniques and data for the others. Further understanding of these processes would require studies of changes in these patterns over time in addition to other types of analyses of the areas concerned.

When the links between clusters in different disciplines were examined, relatively few links between these three disciplines were found, suggesting that information and ideas do not diffuse readily between these three disciplines. There was some indication, however, that some of these disciplines were more closely linked with disciplines other than these three.

As we have already pointed out, the links identified by co-citation analysis have to be examined by other methods such as interviews and content analysis of documents in order to determine whether or not they are meaningful. However, this preliminary analysis suggests that the technique is useful in identifying active areas within scientific disciplines which would merit intensive analysis by other techniques.

FOOTNOTES

1. The SSCI is similar in concept to the older, established Science Citation Index (SCI), which has been described in detail in Weinstock (1971).

Each year about 75,000 articles are entered into the SSCI data base from about 1,000 core journals in all fields of social sciences. In addition, selective coverage is provided of natural science journals which constitute the SCI data base.

2. A threshold for citation frequency is set (usually 10 or 15 times cited per item) and all items satisfying this threshold criterion are selected. In the next step, all co-citations of highly cited items are summarized and a file of pairs of co-cited documents is created. Following the application of some normalization procedure (e.g. the Jaccard coefficient of similarity is used here), a second threshold is set on the document-document similarity for the purpose of input to a single-link clustering algorithm. After all clusters have been formed at the specified level, the linkages between clusters are determined by consideration of document-document similarities lower than the clustering threshold. Clustering of this file was performed at four different levels of the Jaccard coefficient. Although there is considerable evidence that different thresholds are appropriate for different specialties, only clusters obtained at the 11% and 16% levels are discussed here.

3. This is a slight modification of the original definition of cluster co-citation given in Griffith et al. (1974). Clusters at a particular level of association (16%) are used as input to a second clustering step, in effect, a clustering of clusters. To do this, a suitable linkage measure between clusters is needed, and the one utilized here is the number of papers citing documents in both clusters. This linkage measure

is then used as input to the same single-link clustering procedure used originally to cluster the cited documents. By lowering the inter-cluster linkage threshold, a successively more inclusive perspective of the clusters is obtained until, at a sufficiently low threshold, all or nearly all clusters have been aggregated. At any step in the process the clusters of clusters can be represented as either a connected graph or as a configuration in two dimensions using multidimensional scaling.

In our discussion of connected graphs, we speak of the number of links that a cluster has with other clusters and the strength of these links. The latter represents the number of papers citing documents in both clusters. We discuss primarily strengths of 3 co-citations between clusters and 7 co-citations between clusters.

4. For example, the mean cluster size obtained in the SSCI (5.3 cited documents per cluster) is very similar to the mean cluster size obtained in the SCI (5.0 cited documents per cluster). Also the mean numbers of documents linked to the cited documents are similar: 117.8 for SSCI and 89.7 for SCI. The amount of connectedness between cited items which result from co-citation are also similar: .83% for SSCI and .56% for SCI. The mean numbers of clusters linked to a cluster are also comparable: 41.2 for SSCI and 32.0 for SCI. Finally, the percentages of clusters connected to one another are quite close: 3.4% for SSCI and 2.0% for SCI.
5. Each source journal listed in the SSCI front matter in the categories of psychology, sociology, economics and several related headings were checked in the 1973 edition of Ulrichs. Our final disciplinary journal sets consisted of 116 journals in psychology (including social psychology but not psychiatry or psychoanalysis), 85 journals in sociology (including criminology and demography), and 70 journals in economics.

6. The average numbers of citing documents per cluster were: economics: 63; interdisciplinary economics: 68; interdisciplinary sociology: 69; interdisciplinary psychology: 98; and psychology: 95.
7. The citing items consist entirely of journal articles since only journals are searched for citations for the SSCI.
8. Including the one sociology cluster and the cluster which contained one-third or more articles from both psychology and sociology journals.
9. Although our analysis of psychology and interdisciplinary psychology is incomplete, a preliminary analysis indicates that, at a level of 7 co-citations between clusters, 126 of the 197 interdisciplinary psychology clusters form one large network. In addition, there are 16 pairs, triples, and quadruples, 42 isolated clusters, and 13 clusters linked to other fields only. At the same level, the psychology clusters form one large network with 33 clusters, a medium-size network with 14 clusters and two small networks with five clusters each. Twenty-six of the clusters are isolated at this level.
10. A nodal cluster had at least five ties, each representing 3 or more co-citations, with other clusters in the same discipline. A satellite cluster had at least one strong (≥ 7) tie with such a cluster. The remaining clusters were considered peripheral.
11. Not all large clusters were nodal, however. Size of cluster and the tendency of a cluster to be linked to other clusters are closely inter-related. A cluster in which exciting ideas are developed is one which is likely to attract researchers and thus to expand rapidly. At a somewhat later stage, researchers will begin to apply theory or techniques or both to other problems, thus creating new research areas.
12. Psychology was linked to 69 clusters in interdisciplinary psychology.

TABLE 1

NUMBER OF CLUSTERS BY DISCIPLINE

Economics	16
Interdisciplinary economics	28
Psychology	80
Interdisciplinary psychology	197
Sociology	1
Interdisciplinary sociology	34
Interdisciplinary sociology & psychology	1
Clusters in other disciplines	334
Clusters with less than 30 citing references	<u>517</u>
<u>TOTAL</u>	1208

TABLE 2

CLUSTER CHARACTERISTICS FOR PSYCHOLOGY, ECONOMICS AND SOCIOLOGY
(CLUSTERS WITH 30 OR MORE CITING PAPERS AND ONE-THIRD OR MORE DJC)

	<u>Psychology Clusters</u>	<u>Economics Clusters</u>	<u>Sociology Clusters</u>
Number of clusters	277	44	36
Mean size (cited items)	9.1	5.3	5.5
Mean size (citing items)	96.8	65.9	69.1
Mean year of publication of cited items in clusters	1965.8	1965.4	1963.6
Mean year of publication of books cited in clusters	1965.1	1965.6	1962.9
Mean year of publication of articles cited in clusters	1965.9	1965.3	1964.0
Percentage of cited items that are books	14.5	24.5	39.0
Percentage of cited items published in last five years (Price's Index)	57.9	52.8	46.5
Books	54.6	43.9	43.6
Articles	58.6	55.7	48.4

TABLE 3

NODAL CLUSTERS AND THEIR SATELLITES IN ECONOMICS AND SOCIOLOGY

	Economics			Sociology				
	Nodal	Satellite	Peripheral	All Clusters	Nodal	Satellite	Peripheral	All Clusters
Mean number of cited documents	8.8	3.7	3.8	5.3	10.7	4.1	4.4	5.5
Mean number of citing documents	97.8	62.5	47.7	65.9	140.0	49.3	52.7	69.1
Percentage of books among cited documents	12	49	30	24	35	31	42	37
Mean dates of publication of cited items:	1964.9	1967.4	1964.0	1964.9	1965.1	1966.9	1962.3	1963.8
Network 1	1967.0	1965.8	1964.3		1964.9	1968.5	1961.0	
Network 2	1965.5	1965.7	1961.6		1966.5	1965.9	1962.6	
Network 3	1962.6	1969.3	1961.7		1964.5	1966.9	1962.6	
Network 4	1965.2	1970.5	1967.6					
N	13	10	21	44	7	7	22	36

TABLE 4

'NEAREST NEIGHBOR' ANALYSIS: ECONOMICS AND SOCIOLOGY*

	<u>Economics</u>	<u>Sociology</u>
Total number of nearest neighbors	57	62
Mean number of nearest neighbors per cluster	1.3	1.7
Number of nearest neighbors not in discipline (percent of total)	17 (30%)	26 (42%)
Number of nearest neighbors in same discipline (percent of total)	40 (70%)	36 (58%)

*

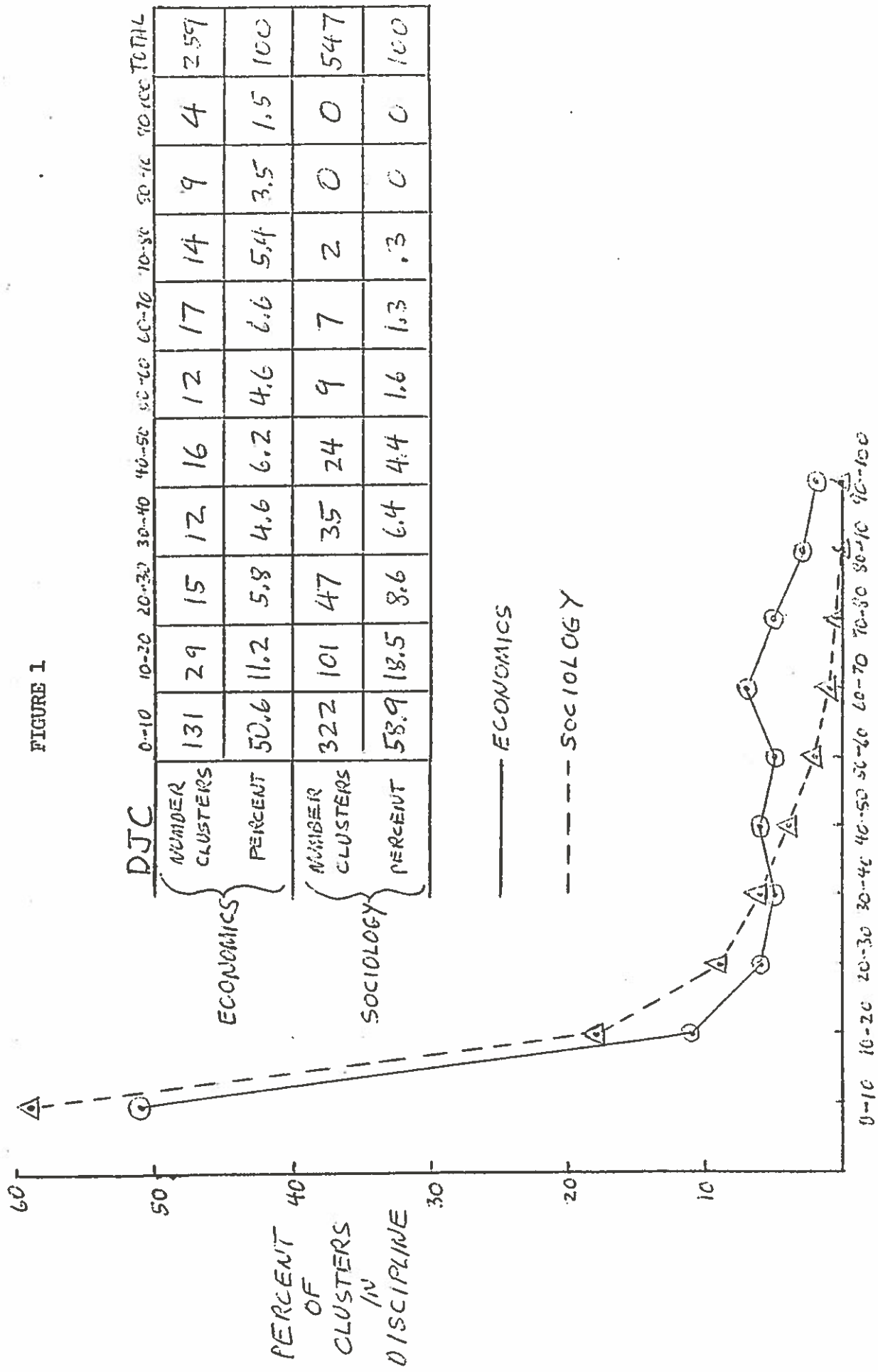
At ≥ 7 co-citations in each link.

TABLE 5

NUMBER OF LINKS BETWEEN ECONOMICS, SOCIOLOGY AND OTHER DISCIPLINESAT 23 CO-CITATIONS

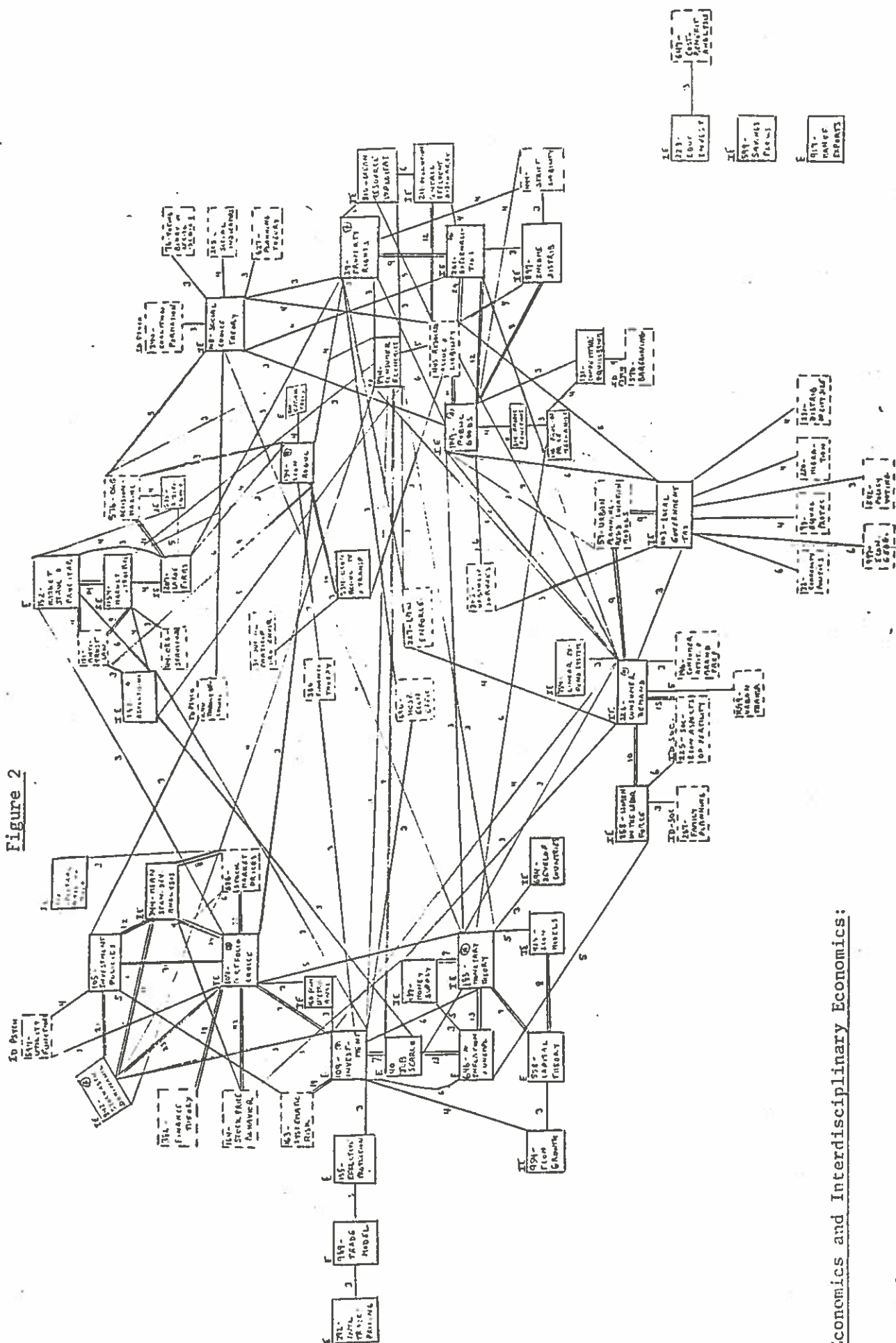
	Economics	Sociology	Psychology	Interdisciplinary Psychology	Other disciplines	Total
Economics	-	2	0	4	27	33
Sociology	2	-	1	17	64	84

FIGURE 1



DISCIPLINARY JOURNAL CONCENTRATION (DJC)

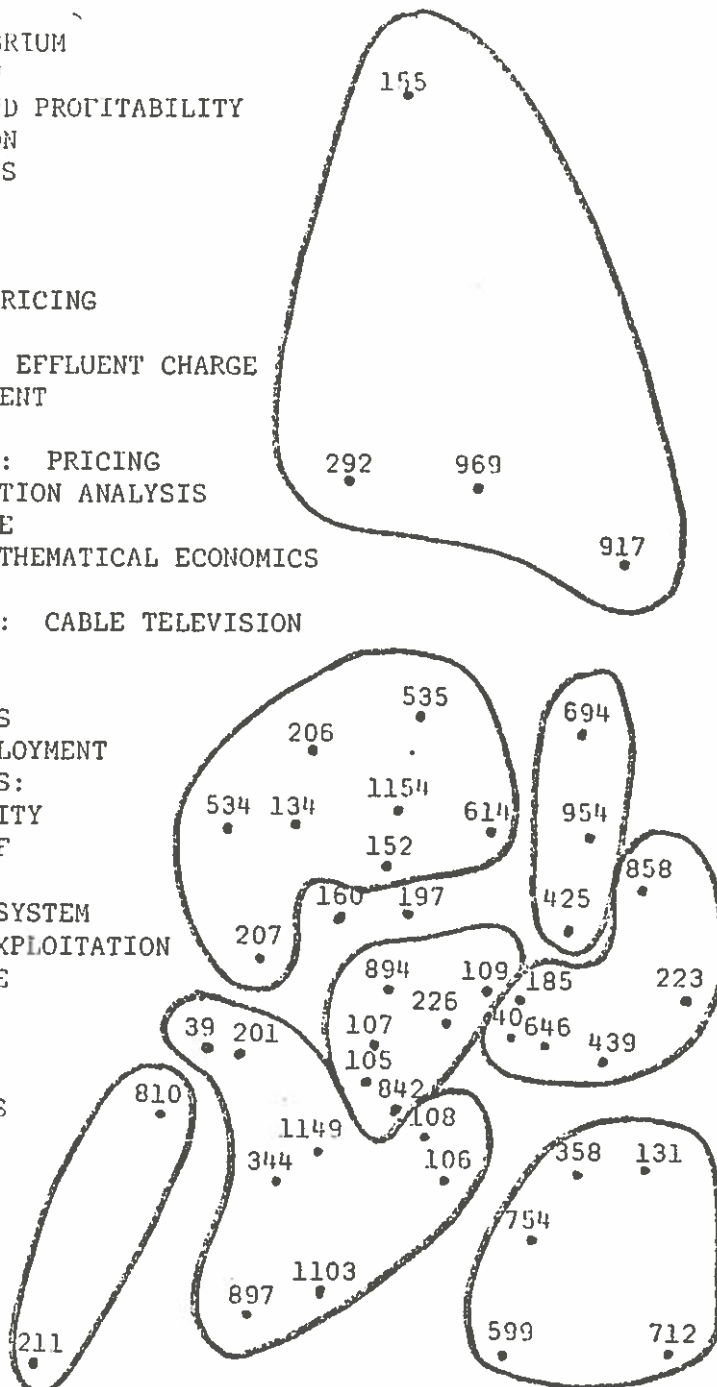
Figure 2



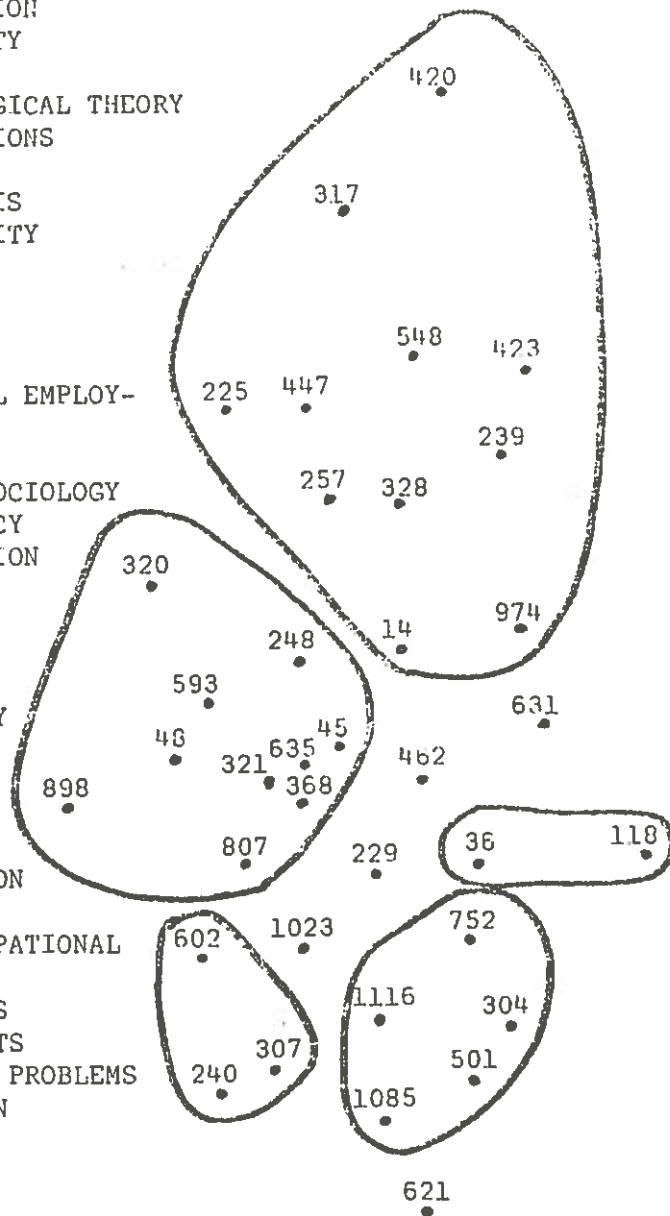
Economics and Interdisciplinary Economics:

at :3 co-citations between clusters

- 39 PROPERTY RIGHTS
- 40 JOB SEARCH
- 105 INVESTMENT POLICIES
- 106 PROCESSES OF PRICE MECHANISM
- 107 PORTFOLIO CHOICE
- 108 SOCIAL CHOICE THEORY
- 109 INVESTMENT
- 131 COMPETITIVE EQUILIBRIUM
- 134 ECONOMIC REGULATION
- 152 MARKET STRUCTURE AND PROFITABILITY
- 155 EFFECTIVE PROTECTION
- 160 ECONOMETRIC ANALYSIS
- 185 MONETARY THEORY
- 197 ADVERTISING
- 201 EXTERNALITIES
- 206 OPTIMAL TAXES AND PRICING
- 207 LARGE FIRMS
- 211 POLLUTION CONTROL: EFFLUENT CHARGE
- 223 EDUCATIONAL INVESTMENT
- 226 CONSUMER DEMAND
- 292 INTERNATIONAL TRADE: PRICING
- 344 MEAN STANDARD DEVIATION ANALYSIS
- 358 WOMEN IN LABOR FORCE
- 425 GROWTH MODELS IN MATHEMATICAL ECONOMICS
- 439 MONEY SUPPLY
- 534 ECONOMIC REGULATION: CABLE TELEVISION
- 535 X-EFFICIENCY
- 599 SAVINGS FLOWS
- 614 PRODUCTION FUNCTIONS
- 646 INFLATION AND UNEMPLOYMENT
- 694 DEVELOPING COUNTRIES:
MARGINAL PRODUCTIVITY
- 712 SPECTRAL ANALYSIS OF
TIME SERIES DATA
- 754 LINEAR EXPENDITURE SYSTEM
- 810 NATURAL RESOURCES EXPLOITATION
- 842 STOCHASTIC DOMINANCE
- 858 CAPITAL THEORY
- 894 CONSUMER ECONOMICS
- 897 INCOME DISTRIBUTION
- 917 MANUFACTURED EXPORTS
- 954 ECONOMIC GROWTH
- 969 TRADE MODEL
- 1103 LOCAL GOVERNMENT
TAXATION AND
EXPENDITURE
- 1149 PUBLIC GOODS
- 1154 MARKET STRUCTURE



- 14 KINSHIP NETWORKS
- 36 LABELING & DEVIANCE
- 45 CAREER MOBILITY
- 48 ECOLOGICAL CORRELATION
- 118 PRISON
- 225 SOCIOECONOMICS ASPECTS OF FERTILITY
- 229 ROLE COMMITMENT
- 239 PREMARITAL SEX
- 240 SOCIOLOGY OF RELIGION
- 248 INDIVIDUAL MODERNITY
- 257 FAMILY PLANNING
- 304 EVALUATING SOCIOLOGICAL THEORY
- 307 RELIGIOUS INSTITUTIONS
- 317 POPULATION CONTROL
- 320 SURVEY DATA ANALYSIS
- 321 OCCUPATIONAL MOBILITY
- 328 FAMILY PROCESS
- 368 CAUSAL ANALYSIS
- 420 FAMILY PLANNING
- 423 MARITAL ADJUSTMENT
- 447 EFFECTS OF MATERNAL EMPLOY-
MENT ON CHILDREN
- 462 COHORT ANALYSIS
- 501 PHENOMENOLOGICAL SOCIOLOGY
- 548 ADOLESCENT PREGNANCY
- 593 SOCIAL STRATIFICATION
- 602 PARTICIPATION IN
RELIGIOUS GROUPS
- 621 LABELING IN SOCIAL
DEVIANCE
- 631 SOCIOLOGICAL THEORY
CONSTRUCTION
- 635 PATH ANALYSIS
- 752 SOCIOLOGY OF
SOCIOLOGY
- 807 EMPIRICAL EVALUATION
OF FORMAL THEORY
- 898 INTERNATIONAL OCCUPATIONAL
PRESTIGE
- 974 POPULATION DYNAMICS
- 1023 GROUP CONFLICT: RIOTS
- 1085 ANALYSIS OF SOCIAL PROBLEMS
- 1116 SOCIETAL TRANSITION



REFERENCES

- Baughman, J. C. A structural analysis of the literature of sociology, The Library Quarterly, 44 (October 1974), 293-308.
- Ben-David, J. The state of sociological theory and the sociological community. Comparative Studies in Society and History, 15 (1973), 448-472.
- Griffith, B. C., H. G. Small, J. A. Stonehill, and S. Dey. The structure of scientific literature II: towards a macro- and microstructure for science. Science Studies 4 (1974), 339-365.
- Griffith, B. C. and H. S. Small. The structure of the social and behavioral sciences literature. Unpublished paper, 1976.
- Holton, G. Models for understanding the growth and excellence of scientific research. In G. Holton, Thematic Origins of Scientific Thought, Cambridge: Harvard University Press, 1973.
- Kruskal, J. B. Multi-dimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis. Psychometrika, 29 (1964), 1-37.
- Mulkay, M. et al. Problem areas and research networks in science. Sociology 9 (1975), 187-203.
- Mullins, N. Theories and Theory Groups in Contemporary American Sociology, New York: Harper and Row, 1973.
- Mullins, N. et al. The group structures of two scientific specialties: A comparative study. Paper presented at the Joint Meeting of The Society for Social Studies of Science and The Sociology of Science Section of the International Sociological

Association, Cornell University, Ithaca, New York, November 4-6, 1976.

Price, D. Citation measures of hard science, soft science, technology and nonscience. In C. E. Nelson and D. Pollack (eds.) Communications among Scientists and Engineers. Lexington, Mass: D. C. Heath and Company, 1970.

Small, H. G. and B. C. Griffith. The structure of scientific literature I: identifying and graphing specialties. Science Studies, 4 (1974), 17-40.

Small, H. G. Citation model for specialty development: a case study of collagen. Social Studies of Science, to appear in 1977.

Weinstock, M. Citation indexes. In Encyclopedia of Library and Information Science, Volume 5. New York: Marcel Dekker, 1971, 16-40.