Telephone Communication Patterns in Austria A Comparison of the IPFP based Graph-Theoretic and the Intramax Approaches

Fischer, Manfred M. and Essletzbichler, Jürgen and Gassler, Helmut and Trichtl, Gerhard

Vienna University of Economics and Business, University of Vienna, University of Vienna, University of Vienna

1992

WSG Diskussionspapiere

WSG 22/92

Telephone Communication Patterns in Austria

A Comparison of the IPFP based Graph-Theoretic and the Intramax Approaches

Manfred M. Fischer, Jürgen Essletzbichler, Helmut Gassler and Gerhard Trichtl

Institut für Wirtschafts- und Sozialgeographie

Wirtschaftsuniversität Wien

Department of Economic and Social Geography

Vienna University of Economics and Business Administration

WSG Discussion Papers

WSG 22/92

Telephone Communication Patterns in Austria

A Comparison of the IPFP based Graph-Theoretic and the Intramax Approaches

Manfred M. Fischer, Jürgen Essletzbichler,
Helmut Gassler and Gerhard Trichtl

Institut für Wirtschafts- und Sozialgeographie
Wirtschaftsuniversität Wien

Vorstand: o.Univ.Prof. Dr. Manfred M. Fischer
A - 1090 Wien, Augasse 2-6, Tel. (0222) 313 36 - 4836

WSG 22/92

Telephone Communication Patterns in Austria

A Comparison of the IPFP based Graph-Theoretic and the Intramax Approaches

Manfred M. Fischer, Jürgen Essletzbichler,
Helmut Gassler and Gerhard Trichtl

WSG-Discussion Paper 22

June 1992

ISBN 3 85037 022 4

Telephone Communication Patterns in Austria

A Comparison of the IPFP based Graph-Theoretic and the Intramax Approaches

Manfred M. Fischer, Jürgen Essletzbichler,
Helmut Gassler and Gerhard Trichtl

Manfred M. Fischer is professor of geography at the Vienna University of Economics and Business Administration. Jürgen Essletzbichler, Helmut Gassler and Gerhard Trichtl are graduate students in geography at the University of Vienna and the Vienna University of Economics and Business Administration.

Revised Version - 9.9.92

Abstract

Two alternative methodological approaches (the IPFP based and the intramax procedures) to the problem of pattern identification in spatial interaction data are compared and evaluated in this paper. After a general discussion of the major characteristics and shortcomings of these methodologies, the paper presents the findings of a case study relying on telecommunication data measured by the Austrian PTT in 1991, in terms of erlangs. The results clearly illustrate the superiority of the intramax approach in the context of medium-sized and relatively centralised flow systems.

1. Introduction

Telecommunication information and communication transfer between different actors in space via electronic media - is among the most rapidly growing fields of spatial interaction. Within this sector, international telecommunication is especially dynamic. Several European countries show annual growth rates in international telecommunication of more than 10 percent for many years. Nevertheless, telecommunication still has a predominantly domestic orientation. For most countries the share of outgoing international telecommunication is below 1 percent (Rietveld and Rossera 1992). This may be partly explained by the fact that short distance communication plays a major role in telecommunication. Annother important reason relates to the barrier effects of national borders which apparently exert a discouraging influence on telecommunication (see Rietveld and Janssen 1990, Fischer et al. 1992, Nijkamp et al. 1990).

Despite the proliferation and diversification of new telecommunication media for transmitting information, including text processing and transmission services such as facsimile transmission, teleconference services and electronic mail, the telephone is still - by far - the most important telecommunication service. A major difficulty in telecommunication research is the availability and quality of data. Many European national telecommunication companies are reluctant to provide data for research purposes.

The current paper results from research undertaken within the Network on European Communication and Transport Activity Research (NECTAR) of the European Science Foundation (ESF) and relies on telephone communication data particularly measured for this purpose by the Austrian PTT in 1991 in terms of erlangs. The data refer to the total telecommunication traffic on the public network. The paper addresses the issue of spatial pattern recognition in the context of domestic telephone communication in Austria. This problem falls into the general category of aggregation and decomposition of relational matrices or the determination of functional regional taxa based on flows through a spatial system which has been a methodological problem of continuing interest in geography and regional science (see Fischer 1982). Two alternative approaches, the iterative proportional fitting based graph-theoretic and the intramax procedure will be used and compared in this paper to identify telephone communication patterns.

The paper falls into two major parts. In the first part two alternative methodologies for identifying telephone communication patterns are discussed and compared. The second part presents the empirical findings using the two methodologies, and a comparison is made of the results obtained.

2. Two Alternative Methodologies

Spatial interactions between basic spatial units (bsu's) in general and communication flows in particular may be depicted in form of relation matrices (I_{ij}) in which each score I_{ij} indicates a flow between an origin-destination dyad (i, j) with i and j denoting bsu's. From a methodological point of view the process of pattern identification in spatial interaction data basically requires two-stage strategies. The first stage refers to the data transformation, where the transformed interaction matrix serves as input for the grouping of the bsu's within the second stage. The need for transforming the raw scores of the relation matrix stems from the intractability of the initial data set arising from such problems as the more or less extreme variablility of the marginal totals etc. (see Holmes 1978). The most widely used approaches are the graph-theoretic IPFP based approach originally suggested and applied by Slater (1976a, b, c, 1981, etc.) and the intramax approach proposed by Masser and Brown (1975). These two approaches which do not impose contiguity constraints in the analysis show important differences in terms of both the data transformation and the grouping techniques used.

The graph-theoretic IPFP based approach applies the iterative proportional fitting procedure (IPFP) put forward by Fienberg (1970) in conjunction with a hierarchic clustering procedure founded on the concept of strong components of a directed graph (see Fischer and Slater 1984, inter alia). In the iterative procedure the elements of the initial interaction matrix are scaled to sum up to a given number (e.g. 100) in the row and column totals. The iterative process for successive row and column standardisation is repeated until sufficient convergence to an equilibrium solution is achieved where all the row and column totals of the adjusted matrix are equal to the given number. The entries in this standardised matrix have a number of useful properties. First, the interaction structure of the initial matrix as defined by the cross-product ratios $I_{ij}\, I_{kl} / I_{kj}\, I_{il}$ ($i \neq k$, $j \neq l$) is preserved (see Fienberg 1970). Second, the elements of the matrix can be interpreted as maximum entropy estimates of the initial matrix given the constraint that all the row and column totals must be equal (see Bacharach 1970).

THE IPFP adjusted table can be interpreted as an asymmetric similarity table, the (i, j)-entry of which provides a measure of the functional relationship from i to j (in the context of this paper; a measure of the strength of telephone communication). In the second stage of the approach a directed graph analogue of single linkage clustering is applied to the adjusted matrix. In this procedure, a directed graph is constructed in which each bsu is represented by a node (vertex). Initially, there are no directed arcs connecting the nodes. The first directed arc inserted goes from bsu i to bsu j if the (i, j)-entry of the adjusted table is the largest. The directed arc corresponding to the second largest entry is then drawn, etc. As this procedure proceeds, nodes unite into strong components. Strong components are subsets of nodes between any two members of which a path of directed arcs exists back and forth. They may be considered as clusters or spatial patterns. As additional directed arcs corresponding to lower values in the adjusted table are inserted, distinct strong components previously formed at higher thresholds unify. This process is hierarchical in nature and can be represented in form of a dendrogram where groups of areas unite at the thresholds at which the strong components formed by them fuse (Slater 1976c, Leusman and Slater 1977).

An alternative approach for pattern identification in spatial interaction data, the intramax approach, has been put forward by Masser and Brown (1975) (see also Brown and Pitfield 1990). Like the graph-theoretic IPFP approach the transformation stage of the intramax procedure is concerned with the relative strength of interactions. In this case, however, relative strength is expressed in terms of the difference between observed and expected flows where expected flows are calculated from the products of the observed marginal frequencies divided by the overall total. This measure may be interpreted in terms of the additional information which is gained with respect to the structure of interaction from the elements of the matrix as against its row and column totals (see Masser 1981).

The aggregation procedure of intramax is simply a modification of Ward's (1963) hierarchical clustering procedure in which the bsu's (groups of bsu's) with the greatest relative strength of interaction in terms of their observed and expected interaction values are merged at each step of the grouping process. The objective function of the aggregation process is based on pairwise comparisons rather than on a variance measure as in the case of Ward's clustering procedure (see Masser and Scheurwater 1980). The objective function of intramax can be expressed as follows:

$$\max_{i \neq j} (I_{ij} - I_{ij}^{*}) + (I_{ji} - I_{ji}^{*}) \qquad i, j = 1, ..., n$$

where I_{ij} represents the observed value of the (i, j) element of the interaction matrix and I_{ij}^{*} the corresponding expected value defined as

$$I_{ij}^{*} = \left(\sum_{j} I_{ij} \sum_{i} I_{ij} \right) / \sum_{i,j} I_{ij} \qquad i, j = 1, ..., n$$

Table 1 summarizes the main characteristics of the two distinct methodological approaches. Each of them has advantages and shortcomings. The intramax approach differs from the graph-theoretic IPFP based one in several ways. First, in the case of the graph-theoretic IPFP approach transformation and aggregation stages are clearly separated, while intramax evaluates the effects of variations in the marginal totals at each step. Second, intramax takes account of the overall strength of the interaction that exists between groups in the objective function at each step of the aggregation process rather than the strength of single links between groups as in the graph-theoretic IPFP case. Third, and closely related, all elements of the interaction matrix are directly considered in the intramax procedure as against a small proportion in the graph-theoretic IPFP (see Masser and Scheurwater 1980). The IPFP is the more effective procedure for completely eliminating the effects of size. But, the IPFP transformation may - in contrast to the intramax transformation - display a systematic bias in awarding exceptionally high scores to flows between those minor centers which are most strongly oriented towards the dominant nodes, and, conversely, in awarding exceptionally low scores to flows between dominant nodes (see Holmes 1978). Consequently the IPFP based approach tends to be more effective when applied to matrices displaying only limited variability in individual interaction values and in row and column totals, and less suitable for analysing sparse matrices containing

Table 1: Some Major Characteristics of the Two Methodological Approaches (see Masser and Scheurwater 1980)

	IPFP Based Graph-Theoretic Approach	Intramax Approach
Objective	Fitting a hierarchic structure to an asymmetric spatial interaction matrix	Delimitation of functional regional taxa of bsu's, i.e. groups of bsu's which have more interaction with each other than with other groups
Transformation Stage	Iterative process for successive row and column standardisation of the initial spatial interaction matrix	Estimation of a matrix of differences between observed and expected flows (procedure repeated at each step of the aggregation process)
Aggregation Stage	Directed graph analogue of single linkage clustering based upon the concept of strong components	Modification of Ward's hierarchic clustering procedure where Ward's conventional objective function is replaced by criteria which take account of the overall effect of interaction across group boundaries

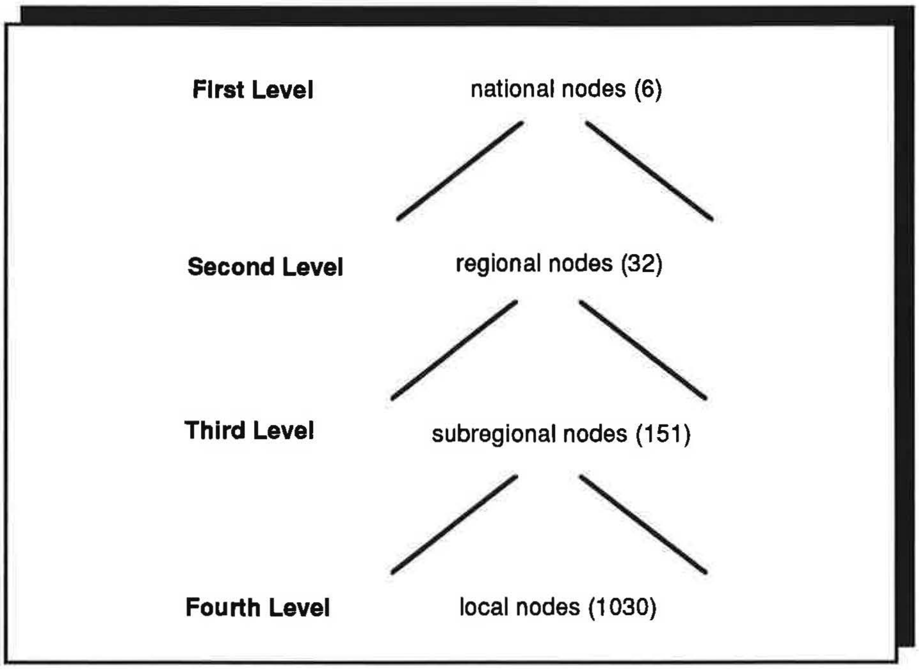
a large number of zero elements indicative of focused interaction systems, especially if these systems are hierarchically structured or highly centralised on one or two dominant nodes.

The aggregation procedure used in the case of the IPFP based graph-theoretic approach shares most of the deficiencies which have been attributed to conventional single linkage cluster analysis. The most important one is the chaining tendency, i.e. the tendency to group together at a relatively low level bsu's linked by chains of intermediates. Single linkage analysis is a useful tool for identifying optimally connected groups of bsu's which have more interaction with each other than with other groups (see Masser and Scheurwater 1980). Despite these shortcomings, the IPFP based graph-theoretic approach is a useful tool for analysing the patterns which underly spatial interaction data. In particular, its graph-theoretic interpretation which allows one to represent spatial interaction patterns by directed graph diagrams and its use of information concerning asymmetries of interaction between pairs of bsu's are comparative advantages worthwhile to mention (see Slater 1981). In contrast, the intramax approach is more readily applicable to large data sets and can be more easily adopted to deal with large, sparse matrices. The results obtained are directly interpretable in terms of the explained interaction, i.e. the proportion of total flows occurring within groups. The major shortcoming of this approach, however, lies in the fact that the stepwise aggregation procedure can not guarantee an optimal solution to the partitioning problem at any given level.

3. The Empirical Study

The empirical study serves to evaluate the performance of the two methodological approaches in practice, in qualitative rather than quantitative terms. The study relies on telecommunication data measured by the Austrian PTT in 1991, in terms of erlangs, an internationally widely used and reliable measure of telecommunication contact intensity which is defined as number of phone calls multiplied by the average length of the call divided by the duration of measurement. The data refer to the total telecommunication traffic between the 32 telephone districts representing the second level of the hierarchical structure of the Austrian telephone network (see figures 1 and 2). Both approaches, the IPFP based graph-theoretic one and intramax were applied

Figure 1: The Hierarchical Structure of the Austrian Telephone Network



Note: In several cases there are direct lines between regional nodes belonging to different national nodes. The same is true for subregional nodes belonging to different regional nodes.

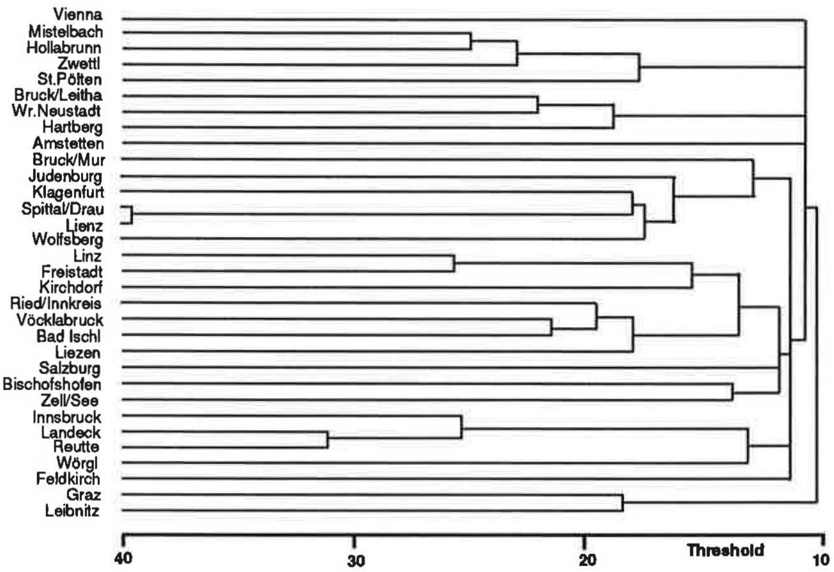
Figure 2: The 32-Zone System Used for Modelling Domestic Telephone Communication

to the (32, 32)-flow matrix with a zero diagonal. The IPFPHC - procedure of the SAS program package (version 5.18) was used in the case of the IPFP based approach and de Jong and Floor's (1991) Flow map 2.1 in the case of intramax.

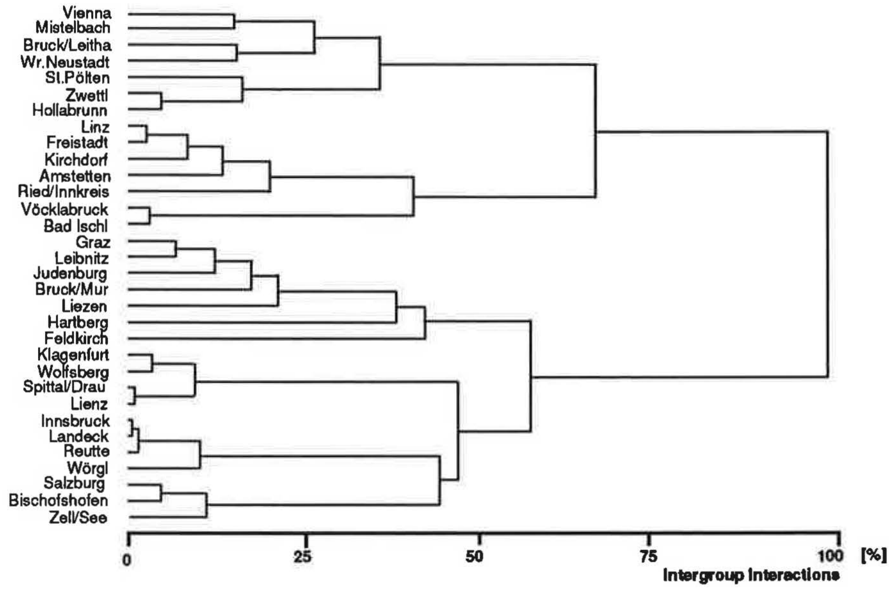
The main findings of the empirical study are summarized in figure 3 in form of dendrograms. Figure 3a displays the result obtained by means of the IPFP based graph-theoretic approach and figure 3b that one by the intramax approach. These figures show that there is a fundamental difference between the results obtained by the two different methodologies. Dendrogram 3a shows most of the deficiencies which have been attributed to the IPFP based graph-theoretic approach. Several instances occur where more than two groups of bsu's merge in a single step, and moreover there are several cases where single bsu's do not fuse with other (groups of) bsu's until the later stages of the aggregation process. The last four steps in the aggregation process involve single zones. Vienna and Amstetten are isolated at a relatively low threshold level of 10.5, Feldkirch at the threshold of 11.6 and Salzburg at the threshold of 11.8. This chaining tendency is not evident in the intramax dendrogram. According to Slater (1976) late entries of single bsu's into the aggregation process should be interpreted in terms of bsu's which have a more national than a regional identity in the telephone communication process. Though this explanation seems reasonable for the case of Vienna, it is hard to see Feldkirch and Amstetten in these terms.

In contrast, the result achieved by means of the intramax approach exhibits a well-defined tree structure which is characteristic of most applications of Ward's clustering procedure. In this case there are no late entries of single bsu's. The principal feature of the communication pattern that emerges from the grouping process can be seen in figure 3b which shows that Austria's telephone communication pattern reflects a North-South dichotomy. The northern communication region includes Vienna, Burgenland, Lower and Upper Austria, while the southern one consists of Vorarlberg, Tyrol, Salzburg, Carinthia and Styria. This division has same correlation with various differences in physical and human geography. The northern region is made up by two functional regional taxa (Upper Austria with the provincial capital of Linz on the one hand, and Vienna, Lower Austria and Burgenland, except Hartberg, on the other). The southern region contains three major subclusters, namely Tyrol-Salzburg, Carinthia, and Styria. With one exception (Feldkirch

Figure 3: Dendrograms Produced by the Two Methodological Approaches

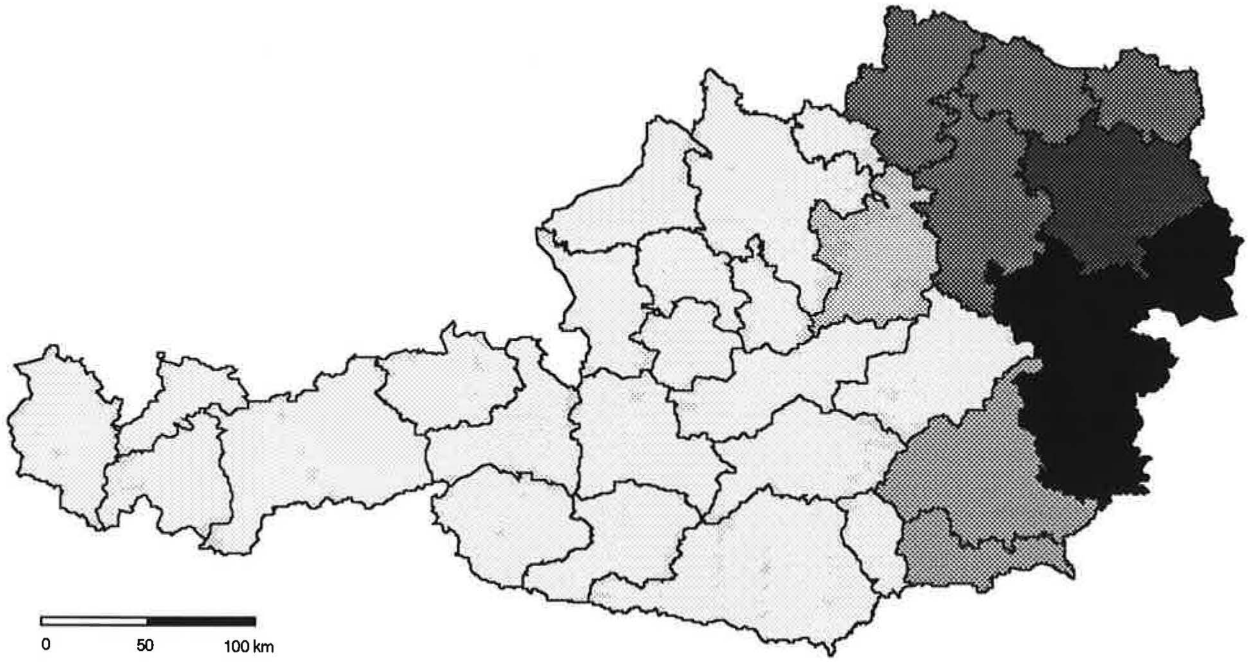


(a) IPFP Based Graph-Theoretic Approach

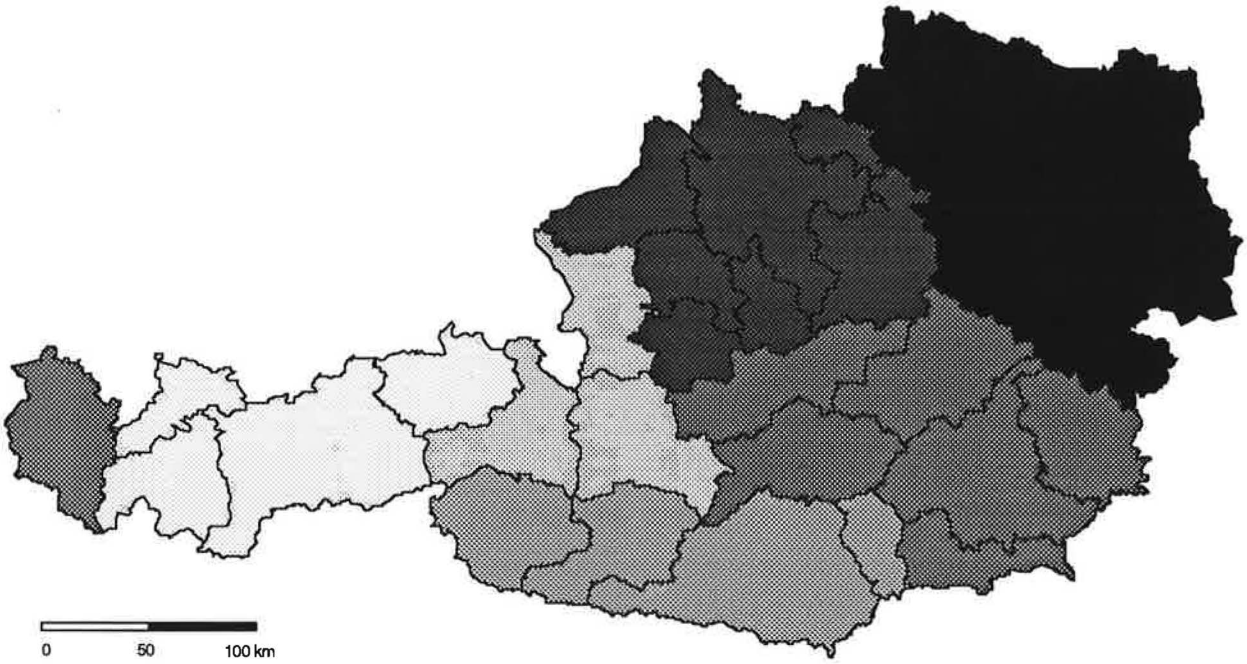


(b) Intramax Procedure

Figure 4: Spatial Patterns Generated at the Level of Six Taxa



(a) IPFP Based Graph-Theoretic Approach



(b) Intramax Approach

merges with the Styrian cluster at the twenty-sixth step). All groupings are between contiguous bsu's as might be expected in view of the literature that exists on the relationship between distance and interaction in general and between distance and telephone communication in particular. Evidently, geographical distance in general and geographical contiguity in particular is a major factor in understanding communication patterns.

Figure 4 summarizes the main features of the two sets of results in spatial terms for the groupings that occur at the level of six taxa of the aggregation process. The spatial configurations show that the application of the IPFP based graph-theoretic approach gives rise to a mixture of a very large region consisting of 21 out of 32 bsu's and five very small regions located in the Eastern part of Austria, while the result obtained by intramax exhibits essentially a pattern of compact regional taxa of bsu's which show more interaction with each other than with other bsu's. The boundaries of the regional regional taxa displayed in figure 4 are indicative for discontinuities in the intensities of telephone communication and point to some sort of barriers to communication.

Finally, it is worthwile to mention that the intramax result conforms extremely closely to an intuitive knowledge of the study area as well as to the grouping of telephone districts used by the Austrian PTT to define the top level of the hierarchical structure of the telephone network. There are only two mismatches which refer to the allocation of Feldkirch (PTT: assignment to Tyrol; intramax: assignment to Styria) and Bad Ischl (PTT: assignment to Salzburg; intramax: assignment to Upper Austria).

4. Summary and Conclusions

The paper draws attention to problems associated with the identification of patterns in telephone communication flow data. Two different methodologies are discussed and compared. Each of them has its own particular advantages and shortcomings.

The empirical results obtained illustrate the superiority of the intramax approach compared to the IPFP biased graph-theoretic one. The result achieved by means of intramax is more easily interpretable in terms of functional regions (regional taxa). The combination of single bsu's and a very large group produced by the IPFP based strategy highlights the essential difference between the two strategies, a result which is in accordance with other studies (see, for example, Masser and Scheurwater 1980). The intramax approach leads to spatial groupings of bsu's (hierarchical pattern) which show more interaction with each other than with other bsu's.

It should be emphasized that the empirical study in this paper has been concerned with only one type of spatial interaction data (telephone communication) in the context of medium-sized and relatively centralised flow systems (with six major nodes: Vienna, Salzburg, Linz, Graz, Innsbruck and Klagenfurt). Further investigation is needed of a wide range of data types in association with different types of spatial systems (highly dispersed versus highly centralised flow systems, different numbers of bsu's) before more general conclusions can be drawn with respect to the superiority of the intramax over the IPFP graph-theoretic approach in more general terms than revealed in the context of this paper.

Acknowledgements: This study is part of a research project, funded by the "Fonds zur Förderung der Wissenschaftlichen Forschung" (P 7516). We are extremely grateful to the Director General of the Austrian PTT and to Ing. Robert Totz for the tremendous efforts undertaken in 1991 to measure telephone traffic in Austria at a rather high degree of spatial detail exclusively for this project. We would like to thank all members of the NECTAR - Core Area "Barriers to Communication", and especially Hans Ouwersloot, Piet Rietveld and Fabio Rossera for numerous helpful comments and suggestions on our research efforts. The views expressed and any errors in this paper, however, remain our own.

References

Bacharach, M. (1970): **Biproportional Matrices and Input-Output Change.** New York. Cambridge University.

Brown, P.J.B. and Pitfield, D.E. (1990): The intramax deriviation of commodity market structures from freight flow data, **Transportation Planning and Technology** 15, pp. 59-81.

Fienberg, S.E. (1970): An iterative procedure for estimation in contingency tables, **The Annals of Mathematical Statistics**, vol. 41, pp. 907-917.

Fischer, M.M. (1982): **Eine Methodologie der Regionaltaxonomie: Probleme und Verfahren der Klassifikation und Regionalisier-ung in der Geographie und Regionalforschung**. Bremen. Universität Bremen.

Fischer, M.M. and Slater, P.B. (1984): Structuring relation matrices: Hierarchical migration patterns in the Federal Republic of Germany, in Bahrenberg, G. and Fischer, M.M. (eds.): **Theorie und Quantitative Methodik in der Geographie**; pp. 174-201. Bremen. Universität Bremen.

Fischer, M.M., Essletzbichler, J., Gassler, H. and Trichtl, G. (1992): Interregional and international telephone communication. Aggregate traffic models and empirical evidence for Austria, **Sistemi Urbani** (in press).

Hirst, M.A. (1977): Hierarchical aggregation procedures for interaction data: A comment, **Environment and Planning A**, vol. 9, pp.99-103.

Holmes, J.M. (1978): Transformation of flow matrices to eliminate the effects of differing sizes of origin-destination units: A further comment, **IEEE Transactions on Systems, Man, and Cybernetics**, vol. 8, pp. 325-332.

Jong, T. de and Floor, M. (1991): Using Flowmap 2.1. A program for the display and analysis of interaction data, Regional and Rural Development Planning, Faculty of Geography University of Utrecht.

Leusman, C.S. and Slater, P.B. (1977): A functional regionalization program based on the standardization and hierarchical clustering of transactions flow tables, **Computer Applications 4**, pp. 769-777.

Masser, I. (1981): The analysis of spatial interaction data, TRP 31, Department of Town and Regional Planning, University of Sheffield.

Masser, I. and Brown, P.J.B. (1978): Hierarchical aggregation procedures for interaction data, **Environment and Planning A**, vol. 7, pp. 509-523.

Masser, I. and Scheurwater, J. (1978): The specification of multi-level systems for spatial analysis, in Masser, I. and Brown, P. (eds.): **Spatial Representation and Spatial Interaction**, pp. 151-172. Leiden and Boston: Martinus Nijhoff.

Masser, I. and Scheurwater, J. (1980): Functional regionalisation of spatial interaction data: An evaluation of some suggested strategies, **Environment and Planning A**, vol. 12, pp. 1357-1382.

Nijkamp, P., Rietveld, P. and Salomon, I. (1990): Barriers in spatial interactions and communications. A conceptual exploration, **The Annals of Regional Science**, vol. 24, pp. 237-252.

Rietveld, P. and Janssen, L. (1990): Telephone calls and communication barriers. The case of The Netherlands, **The Annals of Regional Science**, vol. 24, pp. 307-318.

Rietveld, P. and Rossera, F. (1992): Telecommunications demand; the role of barriers, Paper presented at the NECTAR International Symposium, 18-21 March 1992, Amsterdam.

Slater, P.B. (1976a): A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows, **Regional Studies**, vol. 10, pp. 123-132.

Slater, P.B. (1976b): Hierarchical internal migration regions of France, **IEEE Transactions on Systems, Man, and Cybernetics**, vol. 9, pp. 321-324.

Slater, P.B. (1976c): A hierarchical regionalization of State Economic Areas based upon 1965-1970 migration flows, **Review of Public Data Use**, vol. 4, pp. 32-56.

Slater, P.B. (1981): Comparisons of aggregation procedures for interaction data: An illustration using a college student international flow table, **Socio-Economic Planning Sciences**, vol. 15, pp. 1-8.

Ward, J.M. (1963): Hierarchical grouping to optimise an objective function, **Journal of the American Statistical Association**, vol. 58, pp. 236-244.