A class of related nonmetric ("monotone invariant") hierarchical grouping methods is presented. The methods are defined in terms of generalized cliques, based on a systematically varying specification of the degree of indirectness of permitted relationships (i.e., degree of "chaining"). This approach to grouping is shown to provide a useful framework for grouping methods based on an \emph{a priori} specification of the properties of the desired subsets, and includes a natural generalization for "complete linkage" and "single linkage" clustering, such as the methods of Johnson [1967]. The central feature of the class of methods is a simple iterative matrix operation on the original disparities ("inverse-proximities" or "dissimilarities") matrix, and one of the methods also constitutes a very efficient single linkage clustering procedure.

Interest in "proximities" or "similarities" data has increased in recent years with the development of powerful and practical models for multidimensional scaling and clustering. While multidimensional scaling models attempt to define underlying dimensions to account for the relative proximities, clustering has a more basic aim, to seek homogeneous groups of data elements. The purpose of both kinds of models is to depict, and generally to simplify, the structure of relations in the data. Much of the recent work in clustering, particularly in the social sciences [see Fillenbaum and Rapoport, 1971; Hubert, 1973a, 1973b, 1974; Jardine and Sibson, 1971] has concentrated on models with the following properties.

1. \textit{Nonmetric}. These models consider only ordinal properties of the data and are invariant under monotonic data transformation. Like nonmetric scaling models, recent clustering models are especially suited to the many kinds of data in the social sciences that do not justify the metric assumptions of traditional models. In areas such as biological taxonomy, where much of the important development of clustering has occurred, constructing the proximities index is itself a major problem. A variety of measures based on statistical distributions and information theory have been proposed, with scale properties ranging from ordinal to ratio [Cole, 1969; Jardine and Sibson, 1971].

† Part of the substantive work for this paper was done while the author was a graduate student at the University of Michigan and was supported by U.S.P.H.S. Grant NIH GM-01231-06. Some of the material was presented at the Meetings of the Midwestern Psychological Association, April, 1970. Gratitude for their helpful comments on an earlier version is expressed to C. H. Coombs and D. Cartwright.
1971]. Here it will simply be assumed that an ordinal measure has been defined on the pairs of points.

No entirely satisfactory descriptive term exists for this measure expressed in its most common form, in which increasing value denotes decreasing degree of relationship. It has been called a “dissimilarity”, which is restrictive in meaning, “inverse-proximity”, which is cumbersome, and “distance”, which is inaccurate because the triangle inequality need not be satisfied. At the risk of introducing even more confusion, the more neutral term disparity will be used here in order to reflect the wide range of possible empirical applications.

2. Hierarchical. Some techniques simply search for subsets of the data points having specified properties [e.g., Needham, 1961]. However, most of the prominent recent models seek a grouping at each of a number of levels (i.e., the requirement of proximity or some derived measure of “closeness” for points in the same subset is systematically varied according to some proximity criterion). A model of this type results in a hierarchical structure of groupings, a sequence of groupings of the points in which the subsets at higher levels are more inclusive than those at lower levels. This scheme is only implicit in some models, but here we will be principally concerned with models in which the levels of groupings are explicitly related to a varying disparity criterion.

3. Partitioning. For most workers in clustering, [e.g. Sneath, 1969; Jardine and Sibson, 1971; Hubert, 1973b] that term implies a partition of the set of data points. While some, particularly in biological taxonomy [and see Hubert, 1974], acknowledge the existence of non-partitioning methods where clusters may overlap, much less attention is devoted to such methods. This is at least partly due to the fact that a hierarchical partitioning method leads to a straightforward and convenient representation (a tree or “dendrogram”) while the resulting structure in the non-partitioning case is more complex. Clustering methods which partition have substantial practical advantages in analyzing large data sets because they generally produce greater information reduction. Many interesting data sets in the social sciences are relatively small, however [see, for example, Fillenbaum and Rapoport, 1971], particularly in exploratory research which is concerned with theoretical rather than practical issues; information reduction in such cases may be unimportant or even disadvantageous. Furthermore, it will be shown in a later section of this paper that the information reduction achieved by partitioning has a theoretical cost that is not generally well understood and that is not obvious in data analyses.

In view of the widespread identification of the terms “clustering” and “grouping” with partitions of the data points, the following definitions seem worthwhile:

A grouping for a set of points is a collection of its subsets identified