Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

3755 citations in Google Scholar as of 4/1/2016

This was my first published paper; it was written during the summer of 1965 while I was working at Bell Labs, taking a vacation from my mathematics PhD (in category theory). The paper went through many drafts and owes much to the encouragement and criticisms of my colleagues at Bell, especially Roger Shepard and Doug Carroll (Stephen C. Johnson; May 24, 1985; Citation Classic Commentary).

Psychometric Society, Asheville, NC, July, 2016

What Johnson Accomplished in this Paper

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

1) abstracted out from the work of Joe Ward and others, the notion of a *Hierarchical Clustering Scheme* (HCS);

2) characterized, for the first time, a HCS in terms of what is called an *ultrametric*;

3) presented two methods for the construction of a HCS that are now called the *single-link* and *complete-link* hierarchical clustering methods; these two methods are "nonmetric" in the sense of using only the rank order information in the given proximity values;

4) wrote and made available (for free) a Fortran program, **hiclust.f**, to carry out single- and complete-link hierarchical clustering (it is still available at `www.netlib.org`).

Hierarchical Clustering Schemes

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

3755 citations in Google Scholar as of 4/1/2016

Suppose S is a set of n objects, $\{O_1, \ldots, O_n\}$, and between each pair of objects, O_i and O_j, a symmetric proximity measure, p_{ij}, is available.

Any hierarchical clustering strategy produces (using Johnson's terminology) a hierarchical clustering scheme: a sequence or hierarchy of partitions of S, denoted $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{n-1}$, from the information present in the proximity measure.

In particular, the (disjoint) partition, $\mathcal{P}_0$, contains all objects in separate classes (Johnson's "weak" clustering);

$\mathcal{P}_{n-1}$ (the conjoint partition) consists of one all-inclusive object class (Johnson's "strong" clustering);

and $\mathcal{P}_{k+1}$ is defined from $\mathcal{P}_k$ by uniting a single pair of subsets in $\mathcal{P}_k$.

Single-link and Complete-link Hierarchical Clustering

Generally, the two subsets chosen to unite in defining $\mathcal{P}_{k+1}$ from $\mathcal{P}_k$ are those that are "closest," with the characterization of this latter term specifying the particular hierarchical clustering method being used.

(a) complete-link: the maximum proximity value attained for pairs of objects within the union of two sets (thus, the maximum link [or the subset "diameter"] is minimized);

(b) single-link: the minimum proximity value attained for pairs of objects where the two objects from the pair belong to the separate classes (thus, we minimize the minimum link);

(c) average-link: the average proximity over pairs of objects defined across the separate classes (thus, the average link is minimized).

What is an Ultrametric?

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

3755 citations in Google Scholar as of 4/1/2016

Given the partition hierarchies from any of the three criteria mentioned (complete-, single-, or average-link), suppose the values determined when the new subsets were formed (that is, the maximum, minimum, or average proximity between the united subsets) are placed into an $n \times n$ matrix, $\mathbf{U}$.

In general, there are $n - 1$ distinct nonzero values that define the levels at which the $n - 1$ new subsets are formed in the hierarchy;

thus, there are typically $n - 1$ distinct nonzero values present in an appropriately row and column reordered matrix $\mathbf{U}$ that characterizes the identical blocks of matrix entries between subsets united in forming the hierarchy.

The Ultrametric Inequality

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Based on a matrix such as **U**, the partition hierarchy can be retrieved immediately along with the levels at which the new subsets were formed.

In fact, any (strictly) monotone (that is, order preserving) transformation of the $n - 1$ distinct values in such a matrix **U** would serve the same retrieval purposes.

Generally, a matrix **U** that can be used to retrieve a partition hierarchy in this way is called an ultrametric (matrix):

A matrix **U** is called an ultrametric (matrix) if for every triple of subscripts, $i, j, \text{and } k$, $u_{ij} \leq \max(u_{ik}, u_{kj})$;

or equivalently (and much more understandably), among the three terms, $u_{ij}, u_{ik}, \text{ and } u_{kj}$, the largest two values are equal.

Where are We Today Regarding Ultrametrics?

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

3755 citations in Google Scholar as of 4/1/2016

The last paragraph of Johnson's paper is very prescient about where his introduction of the ultrametric concept will lead:

Finally, a different kind of possible extension will be briefly indicated here, in the form of a presently unsolved problem. In Section I we saw that the construction of an HCS is equivalent to finding a metric which satisfies the ultrametric inequality. Given a similarity measure d, we would in general like to find the closest metric D which satisfies the ultrametric inequality — various measures of closeness could be used. ... To the author's knowledge, this problem is unsolved.

Least-Squares Ultrametrics

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

3755 citations in Google Scholar as of 4/1/2016

So, as Johnson foreshadowed, it is now common in the literature to reformulate this hierarchical clustering task as that of locating a best-fitting ultrametric matrix, say, $\mathbf{U}^* = \{u^*_{ij}\}$, to the given proximity matrix, $\mathbf{P}$, such that the least squares criterion

$$\sum_{i<j} (p_{ij} - u^*_{ij})^2 \ ,$$

is minimized.

NP-Hard Ultrametric Problems

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Johnson's earlier comment about being "unsolved" is somewhat of an understatement.

The task of finding the "closest" ultrametric falls within the class of optimization problems now known as NP-hard, which includes all the old combinatorial chestnuts, such as the traveling salesman problem.

Although Johnson was a computer scientist, he cannot be faulted for not proving or knowing this in 1967.

The notion of a problem being NP-hard was not even introduced into the computer science literature until the early 1970s.

A Fortune Cookie From Johnson – for Those Nostalgic for UNIX

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Placing similar objects together into groups is a natural human tendency, one even engaged in by social scientists.

A cluster is such a group formed by someone who wishes to be thought a mathematician.