# 9

# A new graph theoretic oriented program for Harris Matrix analysis

Irmela Herzog

Irwin Scollar

*(Rheinisches Amt für Bodendenkmalpflege, Colmantstr. 14, D 5300 Bonn 1, West Germany)*

## 9.1 The Harris diagram

A relative chronology of a site may be obtained by analysing archaeological stratigraphy. This well-known method in archaeology was first put on a systematic basis by Harris (Harris 1975, Harris 1977). He also proposed visualising the relative chronology of a site with a diagram that shows all the stratigraphic relations. Harris called this diagram the Harris Winchester matrix, but mathematically speaking a matrix is a different thing so that we prefer the term Harris diagram.

The following time relationships may exist between two layers 1 and 2 (Orton 1980, p. 65–80):

1. 1 is later than 2 (above)
2. 2 is earlier than 1 (below)
3. 1 and 2 are contemporary based on a priori knowledge
4. There is no direct relationship between 1 and 2.

Another relationship is useful in practice:

5. 1 and 2 are equal (equivalent). An example is a wall observed in two different cuttings.

Fig. 9.1 illustrates how the Harris diagram shows these relationships. Harris does not differentiate between equal and contemporary relationships.

If layer 1 is later than 2 and 2 is later than 3, it follows that 1 is later than 3. In this case the relationship '1 is later than 3' is called indirect, because it is not based on direct observation. If a direct above- or below- relation can also be established indirectly, this relation is called redundant and for reasons of clarity it is normally not included in the Harris diagram.

The layers and their relationships in general do not lead to a unique representation in Harris diagram form. There may be variations in the horizontal sequence of the layers as well as in their depth position, as Fig. 9.2 shows (cf. Dalland 1984). The excavator who draws the diagram tends to choose the horizontal sequence of the layers such that spatially close layers in the field appear near each other in the Harris diagram. Also crossings of lines indicating relationships are avoided as far as possible. However, there are certain situations where crossings cannot be prevented in the Harris diagram.

For layers whose depth position may vary over a range when taking only the time relationships into account, a look at the layers' artefactual content may help to determine their depth position: For example two layers with approximately the same distribution of sherd types should be set approximately on the same horizontal layer.

## 9.2 Existing programs for Harris diagram generation

Shortly after the Harris diagram had been invented first attempts were made to provide computer assisted generation of the diagram, because there is a great amount of manual work needed to establish a diagram for some hundred layers. The computer programs for Harris diagram generation which we have seen all have some disadvantages:

- The STRATA program (Bishop & Wilcock 1976) produces a diagram that only shows the positions of the layers but not their relationships. Interactive modification of the data is not supported.
- GAMP (Day 1987, program version 3.1, 5/4/88) does not support contemporary or equal relationships. No printable diagram output is created.
- GNET (Ryan 1989) is designed for a Sun or a DEC VaxStation and a PostScript laser printer, which are not at many people's disposal. A PC version is available, which requires a graphic card (CGA or better) and a mouse. It supports equal but not contemporary relationships. (The information on the PC version of GNET was supplied by a referee.)
- The ORPHEUS matrix generator (Williams 1989) is a full screen editor for Harris diagrams. It was still being developed when its handbook was written. The example diagram only shows the layer positions but not their relations.

Most of these computer programs require that the layers be identified by numbers only, but in practice it is often convenient to use alphanumeric identifiers like '45a' or 'I-41'. The most serious problem seems to be the layout of the diagram. This is often fixed, and does not take into account that relationships in general do not lead to a unique diagram. The problem of crossings and crossing minimization is not addressed. When the relations of the layers are shown at all, the layers are connected by straight lines so that in a situation such as in Fig. 9.1, (e) a crossing occurs that could be avoided. Cycles are often resolved by deleting an arbitrary relation within the cycle.

## 9.3 The Bonn computer program for Harris diagram generation

The Bonn computer program was developed for IBM-compatible PCs. Dynamic data structures are used so that small data sets need less memory than large data sets. It is independent of the computer's graphic card. Any printer supporting the IBM graphic character set may be used for diagram output.

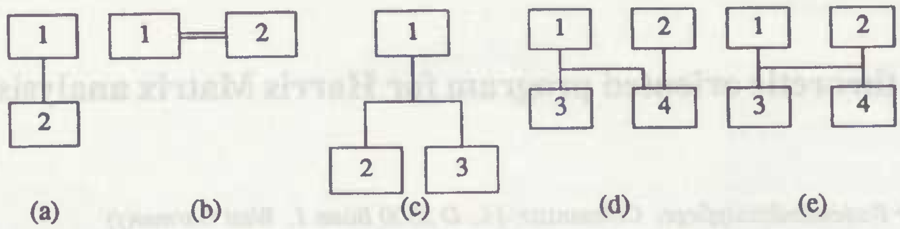



Figure 9.1: Examples of Harris diagram representations of stratigraphic relations:

(a) 1 is above (later than) 2, 2 is below (earlier than) 1.
(b) 1 is equal to (contemporary with) 2.
(c) 1 is above 2 and 3.
(d) 1 is above 3 and 4, 2 is above 4.
(e) both 1 and 2 are above 3 and 4.

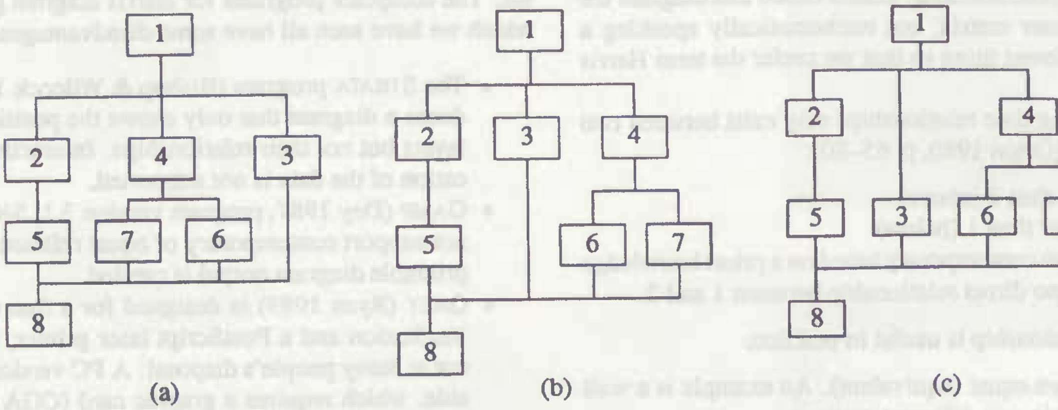



Figure 9.2: Differing representations of the same data

(a) example of a small Harris matrix
(b) same relationships as in (a), but changed horizontal sequences of the layers
(c) same relationships as in (a), but depth of 3 changed.

### 9.3.1 Data entry and change

The program HARRIS developed in Bonn stores the layers in a data base. Each layer must have a unique short identifier (alphanumeric, up to 8 characters) and may have a label (up to 40 characters), for comments. The user of the program may choose to

- define a new layer
- change the name or label of a layer
- delete a layer
- split a layer so that two layers result
- merge two layers.

If the layer names are consecutive numbers then the program can generate the names automatically. The user may choose variable length names (example: 1, 2,. . . , 100) or fixed length names (example: 001, 002,. . . , 100). It is also possible to define the layers as one proceeds, i.e. when establishing a relation.

Above, below, contemporary and equal relationships are supported by the program. Since Harris did not distinguish between contemporary and equal relationships, the difference between the two terms must be explained: if two layers are contemporary they will be set on the same horizontal line of the Harris diagram. If two layers are set equal, their above, below and contemporary relations are merged and the layers are connected with a horizontal double bar, see Fig. 9.3.

If the user enters the relation that 1 is above 2, the program automatically establishes the relation that 2 is below 1. Similarly, if 1 and 2 are already set contemporary and the user sets 3 contemporary with 2, then the program knows that 1 and 3 are contemporary. Any direct relation may be deleted. This is quite easy for above or below relationships. But if a contemporary relation is erased a more complex operation results: for example if 1, 2, 3 and 4 are contemporary and the relation 1 is contemporary with 2 is to be deleted, the user must decide whether 3 and 4 are contemporary with either 1 or 2. If an equal relation say between 1 and 2 is deleted, then the user must decide for each above or below relation of 1 and 2, whether it belongs to only one or both layers. The contemporary relations of two layers whose equal relation is being deleted are dealt with in two different ways depending on whether the layers remain contemporary or not.

Most changes in the layer data base affect the corresponding relations. For example, if a layer is erased, all its relations are deleted, too. When two layers are merged, their relations must be merged as well. Conversely, if a layer is split, for each relation of the source layer the user is asked to choose whether it belongs to one or both new layers. Here

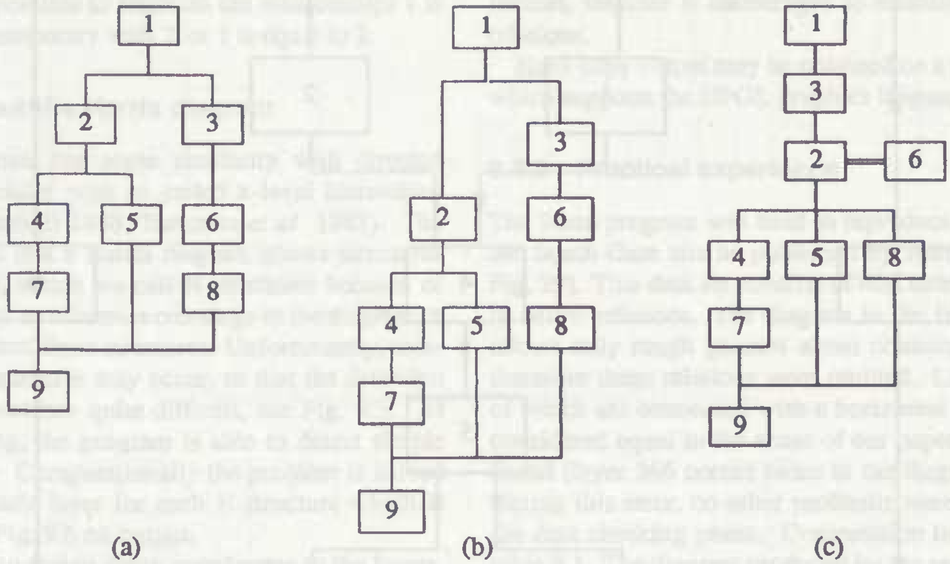



Figure 9.3: Effects of contemporary or equal relations on Harris diagrams

(a) small Harris diagram without contemporary or equal relations.
(b) diagram as in (a), but layers 2 and 6 are set contemporary.
(c) diagram as in (a), but layers 2 and 6 are set equal.

again different measures are taken depending on whether the newly created layers are contemporary. A protocol file listing the layers and their relations can be generated.

A stratigraphic data set is stored by the program in several binary files. Some archaeologists are working with data bases storing layers and their relationships. To avoid entering data twice, a utility is provided for converting Ascii files in the HARRIS protocol file format to files readable by the program.

As Ryan has mentioned, a Harris diagram has some similarity with a mathematical structure called a directed graph, especially if only the above and below relationships are considered. Therefore the data structures proposed for directed graphs were used to store above and below relationships. This enabled us to use standard graph algorithms to solve most problems of data checking.

### 9.3.2 Posing questions about the stratigraphic data set

The user may check the stratigraphic data set by asking the following questions which will be answered by the program:

a) Which layers are directly above, below, contemporary with or equal to a given layer?
b) Which layers are above (below) a given layer?
c) Does a given layer overlie (underlie) another given layer? If so, the direct relations connecting the source layer with the target layer may be listed by the program.
d) Which layers have neither above nor below nor contemporary relations?

To answer questions a) and d) is trivial from the programmer's point of view. The answers of questions b) and c) come in two flavours: the user may or may not take contemporary relations into account. In the former case the methods for answering the questions b) and c) may be found in almost any book with a chapter on graph theory (Aho *et al* 1983, Reingold *et al* 1977), in the latter case these methods must be extended.

### 9.3.3 Automatic data checking

Since it is only possible to lay out the Harris diagram if the data set is consistent, the layout must be preceded by a suite of automatic checks.

It is not sensible to create a Harris diagram showing two or more separate components. Therefore the program first makes sure that the data set is connected (a standard procedure in graph theory) and, if not so, a warning is issued and the components can be listed in a protocol file. With the help of the conversion utility separate data sets may be created for each component. It is possible to proceed with data checking and layout even if the diagram consists of more than one component.

Afterwards the program looks for cycles, i.e. for layers that lie indirectly above or below themselves. If cycles are found, the program lists for each cycle the layers that form it. The program cannot decide which of the relations of the layers in the cycle is erroneous, therefore the user is asked to look for the error in his data and to erase the erroneous relation. A layer may be involved in more than one cycle, see Fig. 9.4.

Again there are methods known in graph theory to solve problems concerning the cycle structure of a graph. The following problems may be solved (listing according to increasing complexity and difficulty of the problem (Reingold *et al* 1977)):

1. Determine whether a given graph contains a cycle
2. Determine a fundamental set of cycles
3. Determine all cycles

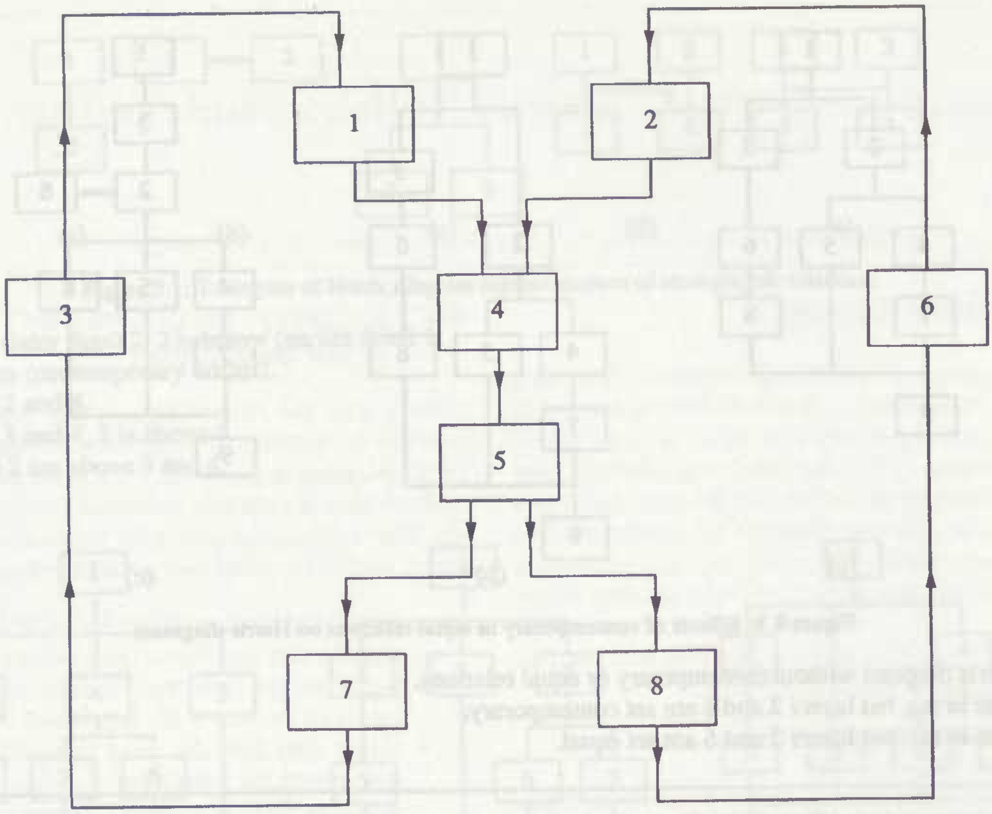



Figure 9.4: Two cycles that have two layers in common:

Cycle a: 1, 4, 5, 7, 3, 1
Cycle b: 2, 4, 5, 8, 6, 2

The program lists the fundamental set of cycles. From these cycles all the other cycles in the graph may be reconstructed. For example, in Fig. 9.4 two fundamental cycles are listed. Another cycle may be generated by combining these two cycles, namely

1, 4, 5, 8, 6, 2, 4, 5, 7, 3, 1.

This example shows that a listing of the fundamental cycles is adequate. Listing all cycles will lead in general to less clarity.

When checking for cycles, preliminary depth coordinates are assigned to the layers without taking the contemporary relations into account. These preliminary depths help to reduce the effort when looking for redundant links. For each below link (from layer A down to layer B) a check is made to find out if another path connects these two layers. The program follows all paths from source layer A downwards (except the direct path to B) until these paths hit the preliminary depth of B. If B is encountered, the relationship A above B is redundant and therefore erased.

The program looks for bad contemporary relations that are in conflict with (indirect) above or below relations. For example, if the user defines the relations 1 above 2, 2 above 3, and 1 contemporary with 3, then the contemporary relationship will be considered bad by the program and will be deleted. Above or below relationships have higher priority than contemporary relations because the former can be observed directly by the excavator, whereas the latter are only subjective conclusions and therefore more open to error. When performing this check, the program again benefits from the preliminary depth coordinates. When checking whether A contemporary with B is a bad contemporary relation, then either:

- A and B have the same depth coordinates. Then A cannot be above B and B cannot be above A, therefore the contemporary relation is not bad

or

- the depth coordinates are different, e.g. A's depth is less than B's depth. Then, as in redundancy checking, all paths from A down to the level of B are generated. If B is encountered, the contemporary relation is considered bad.

Conflicting contemporary relationships must also be resolved. If, as in Fig. 9.2, (c), it is requested that 2 be contemporary with 6 and 4 be contemporary with 5, then it is not possible to draw a diagram which allows these conditions. In this case the two contemporary relations are called conflicting and again the user is asked to choose the erroneous relation. In our example a cycle including contemporary relations results: 2 is above 5, 5 is contemporary with 4, 4 is above 6, and 6 is contemporary with 2. Therefore it is obvious that conflicting contemporary relations can be found via an extension of the normal fundamental cycle finder.

The data checking phase may take several minutes for a data set with several hundred layers. Therefore it is not possible to check after the establishment of each new relation whether or not the data set remains consistent. Only the direct relations are checked on entry, i.e. if 1 is directly above 2, it is not possible to establish the relationships 1 is below 2, 1 is contemporary with 2, or 1 is equal to 2.

### 9.3.4 Laying out the Harris diagram

The Harris diagram has some similarity with directed graphs, and especially with so called k-level hierarchies (Di Battista & Nardelli 1988, Tamassia *et al* 1988). The main difference is that a Harris diagram allows structures as in Fig. 9.1, (e), which we call H-structures because of their form. In order to minimise crossings in the diagram, it is necessary to detect these structures. Unfortunately, combinations of H- structures may occur, so that the detection of H-structures becomes quite difficult, see Fig. 9.5. At the time of writing, the program is able to detect simple H-structures only. Computationally the problem is solved by creating a pseudo-layer for each H-structure which is represented as in Fig. 9.6 on output.

It is quite easy to assign depth coordinates to the layers. There are methods known in the literature (Dalland 1984) which can be readily extended so that contemporary relations are taken into account. In general, more than one depth coordinate configuration is valid. The program positions the layers as high as possible in the diagram.

The assignment of horizontal layer sequences in the diagram is by far the most difficult problem in automatic layout. If the diagram cannot be drawn without crossings, one may want to minimise the number of crossings. But it has been shown that this problem is NP-complete (Di Battista & Nardelli 1988). Therefore, we can only use heuristic methods to achieve this goal approximately. The user may choose to determine the horizontal sequences of the layers manually or automatically. The manual layout is assisted by the program in that the user-selected sequence is saved and becomes the default in the next iteration. Automatic layout is achieved via the PQ-tree concept as proposed by Booth and Lueker (1976) and applied to k-level hierarchies by Di Battista and Nardelli (1988).

Finally, the program computes horizontal positions for each layer. Then the printout showing the layers and their relations is prepared. The diagram is created using only the IBM character set, so that the program works independently of the graphic card, and the diagram may be printed on almost any printer. With most large data sets the breadth of the diagram will be greater than the maximum printer line length. Therefore sidewise output of the diagram is supported.

It is quite difficult to include crossings in the diagram. There are cases when quite a few lines corresponding to crossing relations extend over the full breadth of the diagram. This certainly does not enhance the readability of the diagram. Therefore crossings are currently not plotted. Any layer that has a crossing relation is marked with a double frame. A list of relations which could not be displayed is generated at the bottom of the diagram.

Once a diagram has been created, the user may list all layers at a given depth horizon. Also, the depth of layers may be changed. This means that from Fig. 9.2(b) 9.2(c) may be generated and vice versa. It is not recommended to use this feature extensively, because after a new layer or new relations have been entered, this depth information gets lost. Instead, the user is encouraged to establish contemporary relations.

Hard-copy output may be obtained on a plotter or printer which supports the HPGL graphics language.

### 9.3.5 Practical experience

The Bonn program was used to reproduce the diagram for the South Gate site as published by Harris (Harris 1975, Fig. 29). This data set consists of 406 strata and 856 above or below relations. The diagram in the Harris publication allows only rough guesses about contemporary relations, therefore these relations were omitted. Layers the frames of which are connected with a horizontal double bar were considered equal in the sense of our paper. One error was found (layer 266 occurs twice in the diagram). After correcting this error, no other problems were detected during the data checking phase. Computation times are given in table 9.1. The diagram produced by the program is 110 cm long and 40 cm wide.

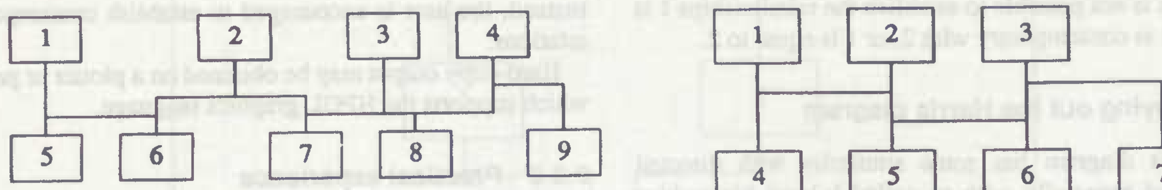
Additionally, an excavation in the city of Xanten in the lower Rhineland was analysed. The data set consisted of about 1000 strata and 6900 relations. Some hundred redundant links were erased. Several cycles were found, and from one erroneous relation several cycles often resulted. Also about 50 bad contemporary and a very few conflicting contemporary relations were found. Because of the excellent standard of the excavation records, these errors could be corrected within a few hours. During the layout phase, the program detected 22 H-structures. Computation times for a version of the Xanten data set with 881 redundant links and 40 bad contemporary relations are given in table 9.1. The resulting Harris diagram shows 67 depth horizons and is about 3.5 metres wide.

In general, computation time depends very much on the speed of the disk storing the stratigraphic data set. Layout preparation speed is highly correlated with the number of layers having more than one above relation and on the number of crossings. A mathematic coprocessor does not decrease processing time.

### 9.3.6 Planned extensions

A remaining problem is the detection of combined H-structures. Also it may be possible to show at least some local crossings in the diagram. If cycles or conflicting contemporary relations are encountered, the program now lists the layers that are part of the cycle and asks the user to cut one of the relations. Only after all cycles have been resolved successfully, is it possible to start the diagram layout. An alternative currently not supported would be to lay out the diagram and replace the cycles with a special symbol.

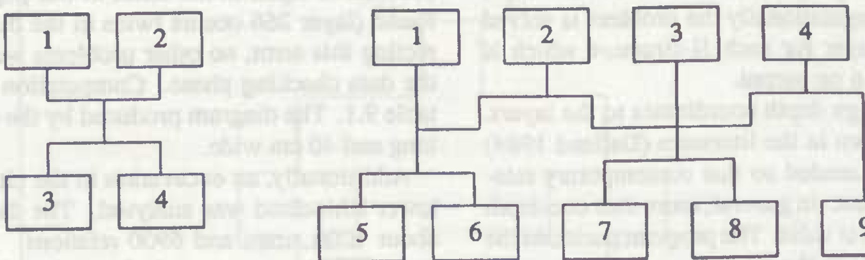
Another planned extension is the introduction of a phase concept. This enables the user to incorporate his knowledge and theories about the phases of the stratigraphic data set into the analysis. The administration of phases will include the generation of new phases, deletion, splitting and merging of existing phases and, of course, the attachment or detachment of phases to layers. The phases have to be checked, too, in that any phase cycles must be removed. The indication of phases will be part of the final diagram.




(a)
1 overlies: 5, 6.
2 overlies: 5, 6, 7, 8.
3 overlies: 7, 8.
4 overlies: 7, 8, 9.

(b)
1 overlies: 4, 5, 6.
2 overlies: 4, 5, 6.
3 overlies: 5, 6, 7.

Figure 9.5: Two examples of combinations of H-structures.




(a) (b)

Figure 9.6: Program representation of H-structures:

(a) Fig. 9.1(e)
(b) Fig. 9.5(a)

A very useful extension would be the combination of the program with a finds' data base, so that the finds from the layers may be visualised in the Harris diagram.

The program is distributed with Version 4 of the Bonn Seriation and Archaeological Statistics Package at no extra charge.

## Bibliography

AHO, A. V., J. E. HOPCROFT, & J. D. ULLMAN 1983. *Data Structures and Algorithms*. Addison-Wesley.

BISHOP, S. & J. D. WILCOCK 1976. "Archaeological Context Sorting by Computer: The Strata Program", *Science and Archaeology*, 17: 3–12.

BOOTH, K. S. & G. S. LUEKER 1976. "Testing for the Consecutive Ones Property, Interval Graphs, and Graph Planarity Using PQ-Tree Algorithms", *Journal of Computer and System Sciences*, 13: 335–379.

DALLAND, M. 1984. "A procedure for use in stratigraphical analysis", *Scottish Archaeological Review*, 3 (2): 116–127. with comment by E.C. Harris: 127–133.

DAY, G. A. N. 1987. "GAMP — an archaeological matrix program". Computer Science BSc dissertation, Department of Electronics and Computer Science, University of Southampton.

DI BATTISTA, G. & E. NARDELLI 1988. "Hierarchies and Planarity Theory", *IEEE Transactions on Systems, Man, and Cybernetics*, 18 (6): 1035–1046.

HARRIS, E. C. 1975. "The Stratigraphic Sequence: A Question of Time", *World Archaeology*, 7: 109–121.

HARRIS, E. C. 1977. "Units of archaeological stratification", *Norw. Arch. Rev.*, 10: 84–94. with comments: 95–106.

ORTON, C. 1980. *Mathematics in Archaeology*. Collins, London.

REINGOLD, E. M., J. NIEVERGELT, & N. DEO 1977. *Combinatorial Algorithms — Theory and Practice*. Prentice Hall, Englewood Cliffs.

RYAN, N. S. 1989. "Browsing through the Stratigraphic Record", in Rahtz, S. P. Q. & Richards, J. D., (eds.), *Computer Applications and Quantitative Methods in Archaeology 1989*, International Series 548, pp. 327–334. British Archaeological Reports, Oxford.

TAMASSIA, R., G. DI BATTISTA, & C. BATINI 1988. "Automatic Graph Drawing and Readability of Diagrams", *IEEE Transactions on Systems, Man, and Cybernetics*, 18 (1): 61–79.

WILLIAMS, P. 1989. "Orpheus Matrix Generator. Handbook of Orpheus/Delilah".

| | South Gate | | Xanten | |
|---|---|---|---|---|
| | AT-286, 8 Mhz. Data on Harddisk with 40 ms seek | AT-386, 20 Mhz. cached Data on Ramdisk | AT-286 Harddisk | AT-386 Ramdisk |
| Automatic data check | 1 min 17 s | 10 s | 28 min | 6 min |
| Layout preparation | 5 min 25 s | 45 s | 1 h 32 min | 17 min |
| Creating matrix file | 44 s | 9 s | 36 min | 5 min |
| Total | 7 min 26 s | 1 min 4 s | 2 h 36 min | 28 min |

**Table 9.1: Computation times of the program HARRIS for the data sets Xanten and South Gate, Xanten consisting of about 1000 layers and 6900 relations and South Gate of 406 layers and 856 relations.**