Flinders Petrie, the Travelling Salesman Problem, and the Beginning of Mathematical Modeling in Archaeology

Thomas L. Gertzen and Martin Grötschel

Abstract. This article describes one of the first attempts to use mathematical modeling and optimization in archaeology. William Matthew Flinders Petrie (1853–1942), eminent British archaeologist, excavating a large graveyard at Naqada in Upper Egypt suggested in his article "Sequences in Prehistoric Remains" [17] to employ a "distance function" to describe the "closeness of graves in time". Petrie's grave distance is known today as Hamming metric, based on which he proposed to establish the chronology of the graves, i.e., the correct sequence of points in time when the graves were built (briefly called seriation). He achieved this by solving a graph theoretic problem which is called weighted Hamiltonian path problem today and is, of course, equivalent to the symmetric travelling salesman problem. This paper briefly sketches a few aspects of Petrie's biographical background and evaluates the significance of seriation.

2010 Mathematics Subject Classification: 01A55, 05-03, 90-03, 90C27
Keywords and Phrases: Travelling salesman problem, seriation, Hamming metric, archaeology

Introduction

When the second author of this article wrote his PhD thesis on the travelling salesman problem (TSP) more than thirty-five years ago, he came across two articles by D. G. Kendall [12] and A. M. Wilkinson [23], respectively investigating the TSP in connection with archaeological seriation. Since he was interested in solving large-scale TSP instances (and in archaeology), he tried to find the original data of the Naqada-graves, based upon which W. M. Flinders Petrie established the prehistoric chronology of Egypt. His search was unsuccessful.

In 2011, planning this Optimization Stories book, the second author approached the director of the German Archaeological Institute in Cairo, S. Seidlmayer. He suggested contacting the first author, who had recently finished his Master thesis, entitled: "Methodik und wissenschaftsgeschichtlicher Hintergrund der Arbeiten von Sir W. M. Flinders Petrie".

Asking for the original Petrie papers on Naqada, the second author learned from the first that these materials, according to E. Baumgartel, referring to a conversation with M. Murray, were no longer existent:

> *She answered that when they had to give up the Egyptian Department, one room [...] was filled from top to bottom with Petrie's papers. She had worked through them with some students who showed her the papers. She said 'published, destroy, unpublished keep.' Well, Naqada was published.* (See [2, p. 6].)

In order to be absolutely sure, the first author contacted the curator of the Petrie Museum London, S. Quirke, who informed him that certain Petrie materials had been rediscovered within the archives of the museum recently, among others, the original "Naqada-slips", to be explained below. The Petrie Museum staff kindly provided digitized images of the material in spring 2012.

Originally, the two authors planned to jointly reprocess Petrie's data, in order to determine optimum solutions for his seriation problems and to publish their results in this article.

However, it turned out that Petrie's materials only represent a rough sketch and show certain inconsistencies, which require careful additional archaeological investigation and also a certain amount of science historical interpretation. This time consuming work is currently carried out and is going to be published in the near future.

Instead, this paper briefly outlines Petrie's modeling concept and the method he applied to solve the mathematical problem he formulated. This very much resembles the engineering approach to combinatorial optimization still prevailing in industry today: Employ experience/knowledge based heuristics skillfully.

The beginning of mathematical modeling in archaeology

Archeology originally was a field dominated by art historians and linguists. The use of methods from the natural sciences and mathematics began slowly. One of the pioneers of this approach to archaeology was Petrie, one of the most eminent Egyptologists of the late 19^{th} century. To sequence graves in Naqada he developed a mathematical "Ansatz" which has led to mathematical objects such as *matrices with the consecutive ones property*, *Petrie-matrices*, the *travelling salesman problem*, and *data mining*. Petrie outlined his approach in archaeological terms and made no formal mathematical definitions or investigation, but he was aware that he was utilizing mathematical techniques. He already introduced and employed concepts, such as the *Hamming distance*, before they were formally defined in other areas of mathematics and the information sciences and which have completely different applications nowadays.

The travelling salesman problem

There is an almost infinite number of articles on the travelling salesman problem, many of these describe some of the origins of the TSP and its great variety of applications. (We recommend Chapters 1 and 2 of [1] for an excellent survey of these two topics.) Since the TSP is usually introduced as the task to find a shortest round trip through a given number of cities, the TSP applications are often associated with vehicle routing, optimal machine control, and the like. One "origin" of the TSP that is often forgotten in overviews is archaeology. That is why we highlight here the independent invention of the TSP in this field. In fact, Petrie also invented a distance measure between graves, which constitutes what we call Hamming metric today.

The Hamming metric

In mathematics, the Hamming distance of two vectors in some vector space is equal to the number of components where the two vectors have different entries. This distance function is clearly non-negative, symmetric, zero only when the two vectors are identical, and obeys the triangle inequality. In other words, it is a *metric*. A computer scientist would say that the Hamming distance between two strings of symbols is the number of positions at which the corresponding symbols disagree. This distance is named after Richard Hamming, who introduced it in his fundamental paper [5] on what we now call Hamming codes. The Hamming distance is, e.g., used in communication to count the number of flipped bits in a transmitted word (in order to estimate errors occurring), and plays an important role in information and coding theory, and cryptography.

Sir William Matthew Flinders Petrie

The excellent biography [3] provides a detailed account of the life and the achievements of Petrie who was born in 1853 near London, died 1942 in Jerusalem and held the first chair of Egyptology (at the University College London) in the United Kingdom. We provide only a few details relevant for the topic addressed here.

Petrie, a grandson of Matthew Flinders, surveyor of the Australian coastline, was tutored at home and had almost no formal education. His father William Petrie, an engineer who held several patents and had great interest in science, taught his son to survey accurately, laying the foundation for his career in archaeology.

William Matthew Flinders Petrie is described by many as a "brilliant" extraordinary individual, one of the leading Egyptologists of his time. Notwithstanding his archaeological discoveries, the fact that he set new standards in painstaking recording of excavations and care of artifacts – thereby inaugurating what might be correctly termed as 'modern' archaeology –, high honors such as a knighthood bestowed upon him and honorary memberships in innumerable

Figure 1: Sir William Matthew Flinders Petrie (© Courtesy of the Egypt Exploration Society, London)

British and international learned societies, Petrie remains a controversial figure due to his right-wing views on social topics and his belief in eugenics, see [19]. Upon his death, he donated his skull to the Royal College of Surgeons London, in particular, to be investigated for its high intellectual capacity in the field of mathematics, see [21].

Petrie and mathematics

William Petrie wrote about his son when Matthew was not yet ten:

> *He continues most energetically studying* [...] *chemicals and minerals.* [...] *we gave him a bit of garden ground to cultivate, to induce him not to spend too long a time in reading his chemical books and making – considering his age – very deep arithmetical calculations ....* (See [3, p. 17].)

Matthew's scientific approach and mathematical mind, basically self-taught, except for two university courses in algebra and trigonometry – but only at the age of twenty-four –, shaped his archaeological career. Having, already at the age of 19, made attempts to understand the geometry of Stonehenge, Petrie applied the same techniques in his 1880–1882 survey of the Pyramids at Giza. His report on his measurements and his analysis of the architecture of the pyramids are till today a prime example of adequate methodology and accuracy. The results of the work published in [14]; [15], and [16] helped to refute a number of mysticism theories linked to ancient monuments.

Petrie's work on the relative chronological ordering of archaeological artifacts showed already a deep understanding of the mathematics behind the seriation problem and was praised in [12, p. 213] as follows:

> *While his writings are not easy to follow, they make fascinating reading for a mathematician,* [...]*, and in my view Petrie should be ranked with the great applied mathematicians of the nineteenth century.* [...] *his writings contain what must surely be the first 'mathematical model'* [...] *in the literature of archaeology.*

Seriation

> *If in some old country mansion one room after another had been locked up untouched at the death of each successive owner, then on comparing all the contents it would easily be seen which rooms were of consecutive dates; and no one could suppose a Regency room to belong between Mary and Anne, or an Elizabethan room to come between others of George III. The order of rooms could be settled to a certainty on comparing all the furniture and objects. Each would have some links of style in common with those next to it, and much less connection with others which were farther from its period. And we should soon frame the rule that the order of the rooms was that in which each variety or article should have as short a range of date as it could. Any error in arranging the rooms would certainly extend the period of a thing over a longer number of generations. This principle applies to graves as well as rooms, to pottery as well as furniture.* (Petrie, 1899 quoted in [3, p. 254])

Below we review and comment Petrie's fundamental publication [18] of 1901. All quotes (written in *italic*) are from this paper.

Being confronted with the task of establishing a prehistoric chronology of Egypt, based on the finds from his excavations at Naqada, Petrie had to find a way of dealing "*simultaneously with records of some hundreds of graves*" from the cemeteries. He therefore developed a method of abstract classification of objects – mainly ceramics. The pottery was divided into nine distinct categories, subdivided into several type-variations. Fig. 2 shows an example of such a classification. This typology was recorded in alphanumerical codes.

The inventory of the graves Petrie excavated was subsequently written

> *on a separate slip of card for each* [individually numbered] *tomb.* [...] *All the slips were ruled in nine columns, one of each kind of pottery. Every form of pottery found in a given tomb was then expressed by writing the number of that form in the column of that kind of pottery.*

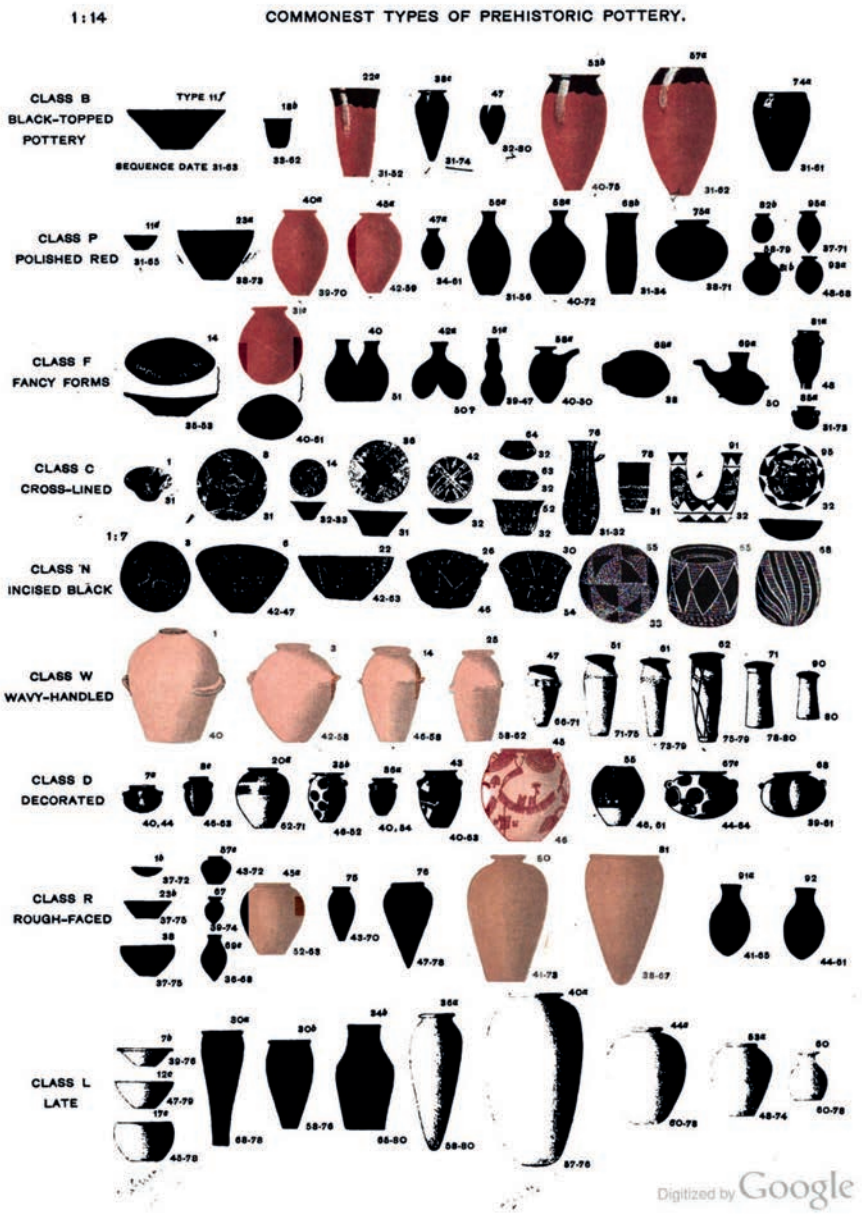

Figure 2: Types of pottery [18] `http://archive.org/stream/diospolisparvac01macegoog#page/n8/mode/2up`

Figure 3 shows the scan of such slips, provided by the Petrie Museum. The first slip is the "header slip". The first entry indicates that in all "tomb slips"

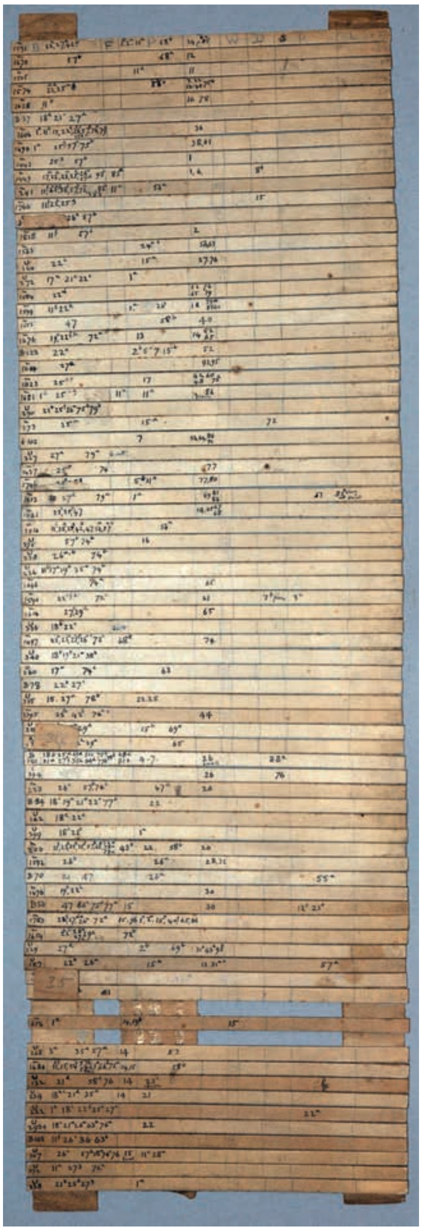

Figure 3: Paper-slips (© Courtesy of the Petrie Museum, London)

the first entry is the individual alphanumerical code of the tomb represented by the slip. The following nine entries of the header slip contain the abbreviated names of Petrie's classification of pottery.

The second slip of Fig. 3 records the inventory of the grave encoded by B 130 (first entry). Six of the following nine entries of the slip are void, indicating that no objects of these six pottery categories were found. The other three entries show that tomb number B 130 contains B(black-topped), F(fancy formed) and N(incised black) pottery, tomb number U 115 contains no N but P(polished red) pottery as well. The entry in column B of row B 130 records the types 22a, 25c and 25f.

What we see here is a data structure which we would call today "sparse matrix representation" or "linked list". Petrie explains that he came up with this representation in order to avoid producing large tables with many empty entries. One can interpret Petrie's data structure as an implicitly defined "grave-pottery type incidence matrix". Each row of this matrix represents a grave. The nine columns B, F, P, ..., L of his slips have to be expanded so that each column corresponds to one type variation of the nine pottery categories. The entry a_{ij} of such an incidence matrix A is equal to "1" if the grave represented by row i contains the pottery type variation represented by column j. In this way every grave is represented by a 0/1-vector describing its pottery contents. Grave B 130, for instance, would have a coefficient "1" in the components representing the pottery type variations B22a, B25c, B25f, F14, N34, and N37, all other components are "0".

In order to pre-arrange the material, Petrie sorted the slips according to stylistic criteria:

> *The most clear series of derived forms is that of the wavy-handled vases* [W]. *Beginning almost globular,* [...] *they next become more upright, then narrower with degraded handles, then the handle becomes a mere wavy line, and lastly an upright cylinder with an arched pattern or a mere cord line around it*

Petrie also knew that: "*there is a class* [...] *we have seen to be later* [L] *than the rest, as it links on to the forms of historic age.*" and arranged his slips accordingly.

After this first arrangement of material (modern algorithmic term: knowledge based preprocessing), Petrie considered the other types of pottery, trying to establish a rough relative chronological order, according to the principles of the Hamming metric, cited above:

> *This rough placing can be further improved by bringing together as close as may be the earliest and the latest examples of any type; as it is clear that any disturbance of the original order will tend to scatter the types wider, therefore the shortest range possible for each type is the probable truth.*

Looking at what Petrie has actually done, one can conclude that this constitutes the simultaneous introduction of the Hamming metric and the TSP. In his chronological arrangement, Petrie considered the closeness of two graves as

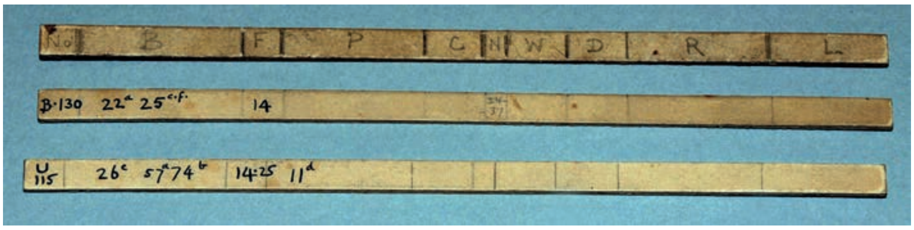

Figure 4: Petrie's arrangement of slips, partial view. (© Courtesy of the Petrie Museum, London)

the number of different entries in the 0/1-vector encoding the graves, which is exactly the Hamming distance of the two grave incidence vectors. Moreover, he claimed that finding an overall arrangement of all graves such that the sum of the Hamming distances between two consecutive graves is as small as possible, would solve his chronological ordering (seriation) problem. And this is nothing but the formulation of the TSP in archaeological terms. Petrie was aware that the available data are imprecise, and that hence the mathematically obtained chronological ordering is only approximate ("probable truth") so that further archaeological "post processing" is necessary.

Having come up with this mathematical model of chronological ordering, Petrie noticed that the amount of data would be outside of his computational capacities. So he applied data reduction and decreased the number of graves according to their statistical relevance: "*In this and all the later stages only graves with at least five different types of pottery were classified, as poorer instances do not give enough ground for study.*"

And thus he began to arrange the 900 remaining paper-slips according to the relative order of appearance of different types of pottery and determined a heuristic solution of a "900-city-TSP". He succeeded in a "satisfactory" arrangement of 700 slips and subsequently made: "*a first division into fifty equal stages, numbered 30 to 80, termed sequence dates or S.D. and then [made] a list of all the types of pottery, stating the sequence date of every example that occurs in theses graves.*" By this he was able to provide a relative chronology, without having to name absolute chronological dates. In other words: Petrie made 49 "cuts" into the list of 700 graves, thereby defining 50 time-periods without giving absolute dates, that are identified by the simultaneous appearance of very similar pottery. This also enabled him to introduce and indicate in his publications periods of appearance of certain pottery types. "*Now on the basis of the list made* [...] *we incorporate all the other graves which contain enough pottery to define their position.*"

In modern TSP-terminology Petrie did the following: He started out with a large number of cities and dispensed those who were irrelevant for the problem, due to insufficient data, to reduce the TSP-instance to a manageable size. (We call this data reduction today). Then he identified a certain subset of cities for which he was able to identify a satisfactory solution (identification of important cities for which a good solution can be found). After that he used a clustering-based insertion-method to produce a feasible and hopefully good solution of the overall problem. A piece of the final sequence of graves (TSP solution) is shown in Fig. 4.

Final remarks

Petrie's sequence dates, which are an outcome of his TSP-approach to seriation, constitute a true paradigm change within the field of archaeology, rendering a scholarly subject, dominated by art historians and linguists, a veritable "scientific" discipline. Pioneering as it was, Petrie's method had and has been further developed and complemented by later archaeologists.

Mathematically speaking, other researchers suggested to replace the Hamming distance by weighted versions and other metrics, taking for instance into account spatial distribution, by dissimilarity coefficients, obtained from statistical analysis of grave contents, and so on. In most of these cases the result was a mathematical model that is equivalent to the TSP with an objective function describing some grave-relationship. A brief survey of these and other approaches, the definition of Petrie matrices, and related concepts can be found in [20].

Literature and further reading

[1] D. L. Applegate, R. E. Bixby, V. Chvátal and W. J. Cook , *The Traveling Salesman Problem: A Computational Study*, Princeton University Press, Princeton, 2006.

[2] E. J. Baumgartel, *Petrie's Naqada Excavation. A Supplement*, London, 1970.

[3] M. Drower, *Flinders Petrie. A Life in Archaeology*, 2nd edition, University of Wisconsin Press, Madison, 1996.

[4] M. K. H. Eggert, *Prähistorische Archäologie. Konzepte und Methoden*, A. Francke Verlag, Tübingen und Basel, 2001.

[5] R. W. Hamming, Error detecting and error correcting codes, *Bell System Technical Journal* 29 (1950) 147–160.

[6] S. Hendrickx, The Relative Chronology of the Naqada Culture. Problems and Possibilities, in J. Spencer(ed.) *Aspects of Early Egypt*, London, 1999, 36–69.

[7] S. Hendrickx, La Chronologie de la préhistoire tardive et des débuts de l'histoire de l'Egypte, *Archéo-Nil* 9 (1999) 13–33.

[8] F. R. Hodson, D. G. Kendall and P. Tautu (eds.), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University Press, Edinburgh, 1971.

[9] W. Kaiser, *Studien zur Vorgeschichte Ägyptens*, 1955.

[10] W. Kaiser, Stand und Probleme der ägyptischen Vorgeschichtsforschung, *Zeitschrift für Ägyptische Sprache und Altertumskunde* 81 (1956) 87–109.

[11] W. Kaiser, Zur inneren Chronologie der Naqadakultur, *Archaeologia Geographica* 6 (1957) 69–77.

[12] D. G. Kendall, Seriation from abundance matrices, *Mathematics in the Archaeological and Historical Sciences*, in [7], pp. 215–252.

[13] M. J. O' Brien and R. L. Lyman, *Seriation, Stratigraphy and Index Fossils. The Backbone of Archaeological Dating*, Kluwer Academic/Plenum Publishers, New York, 1999.

[14] W. M. Flinders Petrie, *Researches on the Great Pyramid, Or Fresh Connections: Being a Preliminary Notice of some Facts*, London, 1874.

[15] W. M. Flinders Petrie, *Inductive Metrology. Or the Recovery of Ancient Measures from the Monuments*, London, 1877.

[16] W. M. Flinders Petrie, *Stonehenge. Plans, Desriptions and Theories*, London, 1880.

[17] W. M. Flinders Petrie, Sequences in Prehistoric Remains, *The Journal of the Anthropological Institute of Great Britain and Ireland* 29 (1899) 295–301.

[18] W. M. Flinders Petrie, *Diospolis Parva. The Cemeteries of Abadiyeh and Hu*, London, 1901.

[19] N. A. Silberman, Petrie's Head. Eugenics and Near Eastern Archaeology, in: A. B. Kehoe and M. B. Emmerichs (eds.), *Assembling the Past: Studies in the Professionalization of Archaeology,* University of New Mexico Press, Albuquerque, 1999, pp. 69–79.

[20] A. Shuchat, Matrix and network Models in Archaeology, *Mathematics Magazine*, 57 (1984) 3–14.

[21] P. J. Ucko, The Biography of a Collection. The Sir Flinders Petrie Palestinian Collection and the Role of University Museums. Appendix A: Donation of Remains of Sir William Flinders Petrie by Simon Chaplin, *Museum Management and Curatorship* 17 (1998) 391–394.

[22] E. M. Wilkinson, Archaeological Seriation and the Travelling Salesman Problem, in [7], pp. 276–283.

Thomas Gertzen
Wehnertstraße 3
12277 Berlin-Marienfelde
Germany
thomasgertzen@aol.com

Martin Grötschel
Konrad-Zuse-Zentrum
für Informationstechnik
Berlin (ZIB)
Takustraße 7
14195 Berlin
Germany
groetschel@zib.de