The Tower Matrix, an Alternative to deal with Distances and Quasi-distances

Yulin Zhang[1], António Veloso da Costa[2] and J. M. S. Simões-Pereira[3]
(orcids: 0000.0003.1833.7456, 0000.0003.4178.8757 and 0000.0001.8930.5932)

Key words: *Distance; Quasi-distance; Distance matrix; Quasi-distance matrix; Tower matrix.*

Abstract: *To any given $n \times n$ matrix D, associate a so-called tower matrix T with $n \times n$ rows and n columns. This matrix T deserves attention because it gives more information than D: in fact, the tower matrix exhibits, not only the shortest length of the paths from point p to point q, but also, for each third point k, the shortest length of the paths from p to q that pass through k. Interpretations of interest for business management and international commerce, for instance, emphasize the advantage of this supplementary information.*

1. The numerical tower matrix: By definition, a distance matrix is a square matrix D whose entries are non-negative real numbers which satisfy the following three conditions:

$$\forall i : d_{i,i} = 0; \quad \forall i, k : d_{i,k} = d_{k,i}; \quad \forall i, j, k : d_{i,k} \leq d_{i,j} + d_{j,k}.$$

It is usually assumed that $d_{i,k} > 0$ for $i \neq k$. And if the symmetry is not required in the definition, then D will be called a quasi-distance matrix.

Recall that distances and quasi-distances have been studied in terms of graphs and digraphs (see [1], [2], [3], [4] [5], [7]): the length of a shortest path from vertex p to vertex q is the value of entry $d_{p,q}$ of the matrix D. This matrix is associated to a graph (when D is symmetric) or to a digraph (when D is not symmetric).

Defining the numerical tower matrix: Given an arbitrary (that means, not necessarily a distance or quasi-distance) matrix D, of type $n \times n$ and whose entries $d_{p,q}$ are real nonnegative numbers, the tower matrix T will be formed with n columns and $n \times n$ rows denoted $L(p,q)$, where $L(p,q)$ is the sum of $L(p)$, the row p, and $C(q)$, the column q, of D; and the rows $L(p,q)$ are ordered in lexicographic order of p, q, that is to say, by increasing values of pq, this pair pq being seen as a two-digits integer written using the number system of radix $n+1$. By other words, the rows of T appear, from top to bottom, as $L(1,1)$, $L(1,2)$, ..., $L(1,n)$; $L(2,1)$, $L(2,2)$, ... $L(2,n)$; ... ; $L(n,1)$, $L(n,2)$, ..., $L(n,n)$. The reason for the name *tower matrix* is that T can be seen as a pile of n matrices T_1, ..., T_n, each one of type $n \times n$, with $L(1,1)$, ..., $L(1,n)$ forming T_1, $L(2,1)$, ..., $L(2,n)$ forming T_2, and so on.

As immediate consequences of the definition, in the tower matrix T of a matrix D, the entry in position p of the row $L(p,p)$ is equal to $2 \times d_{p,p}$, that is, zero, when D is a distance or quasi-distance matrix; in this case, entries in positions p and q of row $L(p,q)$ are both equal to $d_{p,q}$, and entry in position k, distinct from p and q, of row $L(p,q)$ is equal to $d_{p,k} + d_{k,q}$. This makes the tower matrix really interesting. In comparison with D, the tower matrix T exhibits some pieces of information in a much more direct way: while the entries of D, say $d_{p,q}$, just show the length of the shortest route from p to q, the matrix T, besides exhibiting $d_{p,q}$, it exhibits also the length of the shortest routes from point p to point q which pass through each one of the other points.

It is important to note that all these entries are not necessarily lengths of routes. They may be, for instance, travel times or transportation costs; or phylogenetic distances (see [8]) or even costs of making business (see [9]). Think that, in contemporary global markets, closing a business deal between two players A and B has a cost, say $d_{A,B}$. If there is an intermediary K, the total cost may be $d_{A,K} + d_{K,B}$. The respective tower matrix may be formed and, if some event (say political or economic) makes a direct negotiation between A and B impossible, resorting to an intermediary chosen from a set of intermediaries will be necessary and the important role of a tower matrix becomes obvious in such a situation.

A natural question is to ask when is D a distance or quasi-distance matrix, by just looking at its rower matrix T. The answer might be understood almost as a necessary and sufficient condition for D to be a distance or quasi-distance matrix. In fact, as an immediate consequence of the definitions, the following result holds:

Theorem: *Let D be a matrix of type $n \times n$ and whose entries $d_{p,q}$ are real nonnegative numbers.*

1\. *A necessary and sufficient condition for D to be a quasi-distance matrix is that, in its tower matrix T, position p of $L(p,p)$ be zero (for every p), positions p and q of $L(p,q)$ be equal to one another and position k of $L(p,q)$ be equal to $d_{p,k} + d_{k,q}$;*

2. A necessary and sufficient condition for D to be a distance matrix is that, besides the conditions in the previous item, the following additional one be also met: in the tower matrix T, for any p and q, position q of $L(p,q)$ is equal to position p of $L(q,p)$.

As an example of matrices D and T, look first at the digraph of Figure 1, with $1, 2, 3, 4$ as main vertices and x as an auxiliary one.

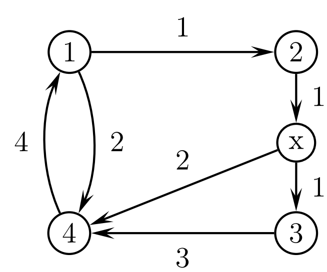

Figure 1: A digraph and its quasi-distances. For the matrices see Figure 2

In Figure 2, see the quasi-distance matrix D and the tower matrix T, yielded by the digraph of Figure 1.

$$D = \begin{bmatrix} 0 & 1 & 3 & 2 \\ 7 & 0 & 2 & 3 \\ 7 & 8 & 0 & 3 \\ 4 & 5 & 7 & 0 \end{bmatrix} \quad \texttt{and} \quad T = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & 8 & 10 & 6 \\ \mathbf{1} & \mathbf{1} & 11 & 7 \\ \mathbf{3} & 3 & \mathbf{3} & 9 \\ \mathbf{2} & 4 & 6 & \mathbf{2} \\ \mathbf{7} & \mathbf{7} & 9 & 7 \\ 8 & \mathbf{0} & 10 & 8 \\ 10 & \mathbf{2} & \mathbf{2} & 10 \\ 9 & \mathbf{3} & 5 & \mathbf{3} \\ \mathbf{7} & 15 & \mathbf{7} & 7 \\ 8 & \mathbf{8} & \mathbf{8} & 8 \\ 10 & 10 & \mathbf{0} & 10 \\ 9 & 11 & \mathbf{3} & \mathbf{3} \\ \mathbf{4} & 12 & 14 & \mathbf{4} \\ 5 & \mathbf{5} & 15 & \mathbf{5} \\ 7 & 7 & \mathbf{7} & \mathbf{7} \\ 6 & 8 & 10 & \mathbf{0} \end{bmatrix}$$

Figure 2: A quasi-distance matrix D and its tower matrix T

Look at T and consider it partitioned into n square sub-matrices T_1, T_2, ..., T_n, where T_1 is formed by the top n rows of T, T_2 formed by the following n rows, and so on. As is easy to verify, in T_i its diagonal and its column $C(i)$ are equal, element to element, to the row $L(i)$ of D. Note that all entries of D appear in T, but they don't use up all positions of T. In Figure 2, we have in bold, the entries of T which are equal to the entries in the rows of D.

2. Defining the Boolean tower matrix: To end up this note consider a variant of the tower matrix defined in boolean terms: call it the *boolean* tower matrix. Where the *numerical* version exhibits path lengths, the boolean variant exhibits simply the existence of paths or links.

Remember (see [6], for instance) that the boolean adjacency matrix D of a graph or digraph G with n vertices is a square matrix of type $n \times n$ whose entries $d_{p,q}$ are defined as follows: - For every p, $d_{p,p} = 1$; for $p \neq q$, it is $d_{p,q} = 1$ when there is an arc (oriented edge) from p to q in the digraph G, or an edge when G is a graph; otherwise, $d_{p,q} = 0$. Note that, in terms of graphs and digraphs, the existence of loops from each vertex to itself is always assumed. These loops are oriented in digraphs, non-oriented in graphs.

Keeping the notation of the previous section, define the *boolean tower matrix* B associated to D as a matrix of type $n^2 \times n$ whose row $L(p, q)$ is the *boolean product (not the boolean sum!)* entry by entry, of the p row $L(p)$ by the q column $C(q)$ of D. This is an unexpected but necessary, or at least convenient difference between the definitions of numerical and boolean tower matrix. Otherwise, the matrix B may also be seen as a pile of square matrices B_1, ...,

B_n, of type $n \times n$, which play here the roles of T_1, ..., T_n, respectively. Thus, the entry k of $L(p,q)$ is 1 if and only if there is, in G, a directed path from p to q consisting of two arcs (p,k) and (k,q). Don't forget that an arc (p,q) may be understood as the path $(p,p),(p,q)$ or as the path $(p,q),(q,q)$; this means that, when the arc (p,q) exists, the entries p and q of $L(p,q)$ are equal to 1. Obviously, when G is a graph, edges will play the role of the (oriented) arcs in digraphs.

In Figure 3, see, as an example, the boolean adjacency matrix D and the boolean tower matrix B of the digraph pictured in Figure 4.

$$D = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \texttt{and} \quad B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 1 \\ \mathbf{1} & 0 & \mathbf{1} & 1 \\ \mathbf{1} & 0 & 0 & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & \mathbf{0} & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{1} & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{0} & \mathbf{0} & 0 \\ 1 & 0 & \mathbf{1} & 0 \\ 1 & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 1 & \mathbf{0} \\ 0 & \mathbf{1} & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & \mathbf{1} \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix}$$

Figure 3: A boolean adjacency matrix D and its boolean tower matrix B

Recalling what was pointed out for the matrices T_i, it holds also here that, in each B_i, the main diagonal and the column C_i (in bold) are equal, entry by entry, to the row $L(i)$ of D.

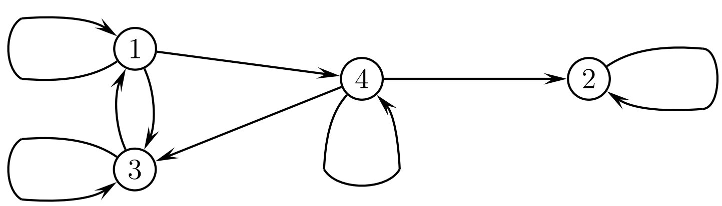

Figure 4: The digraph whose associated matrices are in Figure 3

3. Some conclusive remarks: The tower matrix defined in this paper may be associated to any square matrix whose entries take numerical or Boolean values.

The focus of the analysis in this paper is on tower matrices which are associated with distance, quasi-distance or adjacency matrices. Their huge potential for applications is pointed out. In fact, transport costs, travel times, genetic similarities and negotiation difficulties, among many others, all these variables may play the role of traditional distances, quasi-distances, existence or non-existence of links.

Note also that tower matrices may be naturally associated to graphs and digraphs. Plus, an unexpected difference in the definitions of the numerical versus the Boolean versions of the tower matrix is impressive and pointed out in this note.

References

[1] Agnese Baldisserri and Elena Rubei: *Distance matrices of some positive-weighted graphs,* Australasian Journal of Combinatorics, 70-2 (2018), pp. 185-201.

[2] P. Buneman: *A note on the metric properties of trees,* J. Comb. Theory Ser. B, 17 (1974), pp. 48-50.

[3] S. L. Hakimi and S. S. Yau: *Distance matrix of a graph and its realizability,* Quart. Applied Math. 22 (1964-65), pp 305-317.

[4] J. M. S. Simões-Pereira: *A note on the tree realizability of a distance matrix,* J. Combin. Theory, 6 (1969) pp. 359-374.

[5] J. M. S. Simões-Pereira: *A note on optimal and suboptimal digraph realizations of quasidistance matrices,* SIAM J. Alg. DIsc. Math, 5 (1984) pp.117-132.

[6] J. M. S. Simões-Pereira: *Matemática Discreta: Grafos, Redes, Aplicações*, (*Portuguese*), Editora Luz da Vida, Coimbra, Portugal, 2009.

[7] K. A. Zaretzki: *Constructing a tree on the basis of a set of distances between pendant vertices,* Uspekhi Mat. Nauk, 20-6 (1965), pp.90-92 (in Russian).

[8] Seen on 2018, December 29: *Distance Matrices in Phylogeny,* in https://en.wikipedia.org/wiki/Distance Matrices in Phylogeny.

[9] Seen on 2018.12.29: *Mediation in Politics and the Politics of Mediation,* https://www.mediate.com/articles/benjamiminpolitics.cfm.

– – – – – – – – – –

Research by authors *Zhang* and *da Costa* was supported by Portuguese Funds through *FCT - Fundação para a Ciência e Tecnologia*, within Project UID/MAT/00013/2013.

Simões-Pereira received a grant from the *Divine Cities Foundation*.

– – – – – – – –

[1,2]Dept. of Math., University of Minho, Braga, Portugal.

E-mails: zhang@ math.uminho.pt, aveloso@math.uminho.pt,

[3]Dept. of Math., University of Coimbra, Coimbra, Portugal.

E-mail: siper@mat.uc.pt