Mathématiques et sciences humaines
Mathematics and social sciences

158 | Été 2002
Varia

Vertex partitioning of a class of digraphs

Partionnement d'une classe de graphes orientés

Louis Ferre and Bertrand Jouve

DOI: 10.4000/msh.2857
ISSN: 1950-6821

Publisher
Centre d'analyse et de mathématique sociales de l'EHESS

Printed version
Date of publication: 1 March 2002
ISSN: 0987-6936

Math. & Sci. hum., (40e année, n° 158, *2002*, p. 59-77)

VERTEX PARTITIONING OF A CLASS OF DIGRAPHS

Louis FERRÉ et Bertrand JOUVE*

RÉSUMÉ – Partitionnement d'une classe de graphes orientés.
Un ensemble de sommets V' d'un graphe orienté $G = (V, A)$ est un pseudo puits si son demi-degré extérieur est faible. La recherche d'un pseudo puits dans un graphe d'ordre élevé est un problème de forte complexité combinatoire. Nous montrons, sur une famille particulière de graphes, que l'utilisation d'une classification hiérarchique des sommets, dont l'ensemble est muni d'une métrique bien choisie, permet la mise en évidence de pseudo puits par leur agrégation en 1er.

MOTS CLÉS – Classification hiérarchique, Dissimilarité, Partitionnement d'un graphe orienté, Graphes de petits mondes.

SUMMARY – *A vertex subset V' of a digraph is a pseudo sink set if its out-degree is low. The research of a pseudo sink set in a digraph is a high complexity combinatory problem. We show, for a particular family of digraphs, that a clustering of the vertex set fitted with a well chosen metric allows to reveal pseudo sink sets by their aggregation in a first level.*

KEYWORDS – Hierarchical clustering, Dissimilarity, Partitioning of a digraph, Small worlds Graphs.

1 INTRODUCTION

In numerous situations (one may cite for instance [5], [15], [11]) scientists are confronted with large netlike objects whose architecture may be symbolized by a directed graph, or digraph, $G = (V, E)$. A step toward understanding the object is to search for hidden structures of the digraph. Two complementary approaches may be run, combinatorial and statistical, that have to be adapted according to the type of graphs: dense or non-dense, symmetrical or non-symmetrical. A combinatorial approach may consist in enumerating all the objects of a combinatorial class (cliques, stable sets, *etc.*) or only the maximal ones. It often comes up against combinatorial explosions as soon as the order of the graph becomes high. Moreover, such an approach is not efficient to reveal objects that are nearly in a class (for instance, a dense subgraph is not but nearly complete). The idea of a statistical approach, within which this paper stands, appears for

*GRIMM, Département de Mathématiques et Informatique, UFR SES, Université Toulouse Le Mirail, 5 allée Antonio Machado, 31058 Toulouse Cedex 09, e-mail : loferre@univ-tlse2.fr, jouve@univ-tlse2.fr

instance in [14], for non-dense and non-directed graphs, in the domain of the pagination of large electrical networks, and is adapted in [12] for dense and directed graphs that modelizes the neuronal cortex. The strategy consists in looking for a clustering of the vertices of the graph G in such a way that the vertices of a same class share some adjacency properties, and that the partition, in its whole, reveals some structure properties of G.

Motivated by a study we lead concerning the topology of the World Wide Web [8], we are interested in automatic search processing of sink sets in "small world" digraphs [17], that is directed graphs of low diameter and made of dense components. The family of digraphs we study in that paper is of that type and we use the previous strategy to present a method for revealing dense components that are pseudo sink sets.

Graphs terminology and notations used in that paper follows [1] with some exceptions: we assume there is a loop at each vertex (even if they will never be drawn in the figures), that is all the arcs (i, i) exist, and, given a vertex i, $\Gamma^+(i)$ and $\Gamma^-(i)$ respectively denote the out-neighbourhood and in-neighbourhood of i. Given a digraph, integers n and m respectively refer to its number of vertices and arcs. An articulation vertex of a connected digraph G is a vertex i such that $G - i$ is not connected. In the same way, we define an articulation set B as a subset of A such that $G - B$ is not connected. A subdigraph $G' = (V', A')$ of G is a dense component if $|V'|^2/|A'|$ is close to 1, where $|\cdot|$ symbolizes the cardinal of a set. A subdigraph G' is a pseudo sink set (*resp.* sink set) of G if its out-degree is low (*resp.* equal to zero).

The family of digraphs we will study in that paper is denoted by $G_{p,q} = (V_{p,q}, A_{p,q})$ and made of p complete digraphs with q vertices (there are two opposite arcs for each pair of vertices and a loop at each vertex), $\overleftrightarrow{K^1}, \overleftrightarrow{K^2}, \cdots, \overleftrightarrow{K^p}$, plus another complete digraph $\overleftrightarrow{K^{p+1}}$, plus p maximum matching between each $\overleftrightarrow{K^i}$ and $\overleftrightarrow{K^{p+1}}$, all the arcs of the matchings directed to $\overleftrightarrow{K^{p+1}}$. They have $q(p+1)$ vertices and $q^2(p+1) + pq$ arcs. The digraph $\overleftrightarrow{K^{p+1}}$ is a sink and the articulation set of $G_{p,q}$ of minimal order. Given p and q, all the isomorphic digraphs $G_{p,q}$ will be confused.

In its main points, that paper follows 3 steps:

1. a metric d is defined on the vertex set V and the finite metric space (V, d) is embedded in $(\mathbb{R}^p, d_2)$, where d_2 is the Euclidean distance associated to the L_2-norm,
2. the digraph $G_{3,3}$ is studied in details as an illustrative example,
3. Mathematical proofs of some empirical results obtained for $G_{3,3}$ are given for the $G_{p,q}$.

2 THE GRAPH EMBEDDING

The adjacency matrix $M(G)$ of a digraph G may be seen as a presence-absence data table. That point of view allows to fit V with a dissimilarity coefficient d for binary variables. We have chosen for d the square root of the half sum of the Czekanovsky-Dice index [7], [9] applied to $M(G)$ and $M'(G)$, where $M'(G)$ is the transpose matrix of $M(G)$:

$$d^2 = \frac{1}{2}\left(d_+^2 + d_-^2\right)$$

where

$$d_+^2 = \frac{|\Gamma^+(i) \,\Delta\, \Gamma^+(j)|}{|\Gamma^+(i)| + |\Gamma^+(j)|} \text{ and } d_-^2 = \frac{|\Gamma^-(i) \,\Delta\, \Gamma^-(j)|}{|\Gamma^-(i)| + |\Gamma^-(j)|}.$$

We denote by $A \,\Delta\, B$ the symmetric difference between the two sets A and B. With this coefficient d, two vertices are close together if and only if they have at the same time a lot of common out – and in – neighbours and few different ones (see Figure 1).

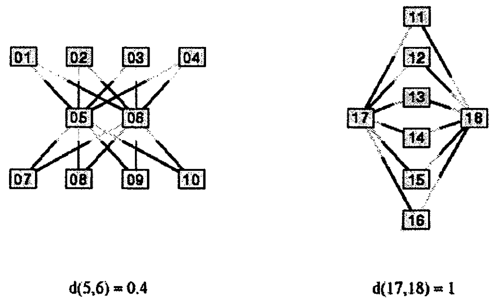

Figure 1. Two examples of calculation of $d(i,j)$. The vertices are numbered. Each line symbolizes an arc whose initial extremity is black and final extremity grey

That choice of d is justified by 3 reasons:

- It is a local index and one just has to know the neighbours of vertices i and j to calculate $d(i,j)$, which is not expensive in computation time for large networks.
- It is an Euclidean semi-distance on V: there exists an isometrical mapping $I_d(G)$ from the metric space (V,d) into the metric space $(\mathbb{R}^p, d_2)$. To prove d is euclidean, just note that d_+ and d_- are Euclidean [7], and that a distance d is Euclidean if and only if the Gram matrix $W(i_0)$ of d at point i_0 is positive for all i_0 [2], where the general entry of $W(i_0)$ is equal to $w_{ij}(i_0) = \frac{1}{2}(d^2(i_0,i) + d^2(i_0,j) - d^2(i,j))$. Then the Gram matrices W_+ and W_- are positive and the Gram matrix $\frac{1}{2}(W_+ + W_-)$ associated with d is also positive so d Euclidean.
- For non-oriented graphs (in that case $d_+ = d_- = d$), empirical results of [14] show that d is efficient to reveal dense components. The adaptation to the directed graphs, by defining d_+ and d_-, lies in a principle of separation between the in – and the out – neighbours, considering the adjacency matrix at the same time as a in-neighbours' table and a out-neighbours' table.

Now, the Euclidean finite space (V,d) is isometrically embedded into $\mathbb{R}^s$ by means of Principal Co-ordinates Analysis. If $C = I_n - \frac{1}{n}\mathbf{1}_{n,n}$ is a centering operator and D the distance matrix between the n vertices, a double centering of D, in rows and columns, defines the Torgerson matrix $W = -\frac{1}{2}\, C\, D\, C$ which is the Gram matrix at the center of gravity of (V,d). Let F_k denotes the eigenvectors of $\frac{1}{n}W$, the k^{th} coordinate $\mathbf{i}_k$ of the vertex i in the embedding of (V,d) in $(\mathbb{R}^s, d_2)$ is [16]:

$$\mathbf{i}_k = \sqrt{\lambda_k} F_k^i$$

where λ_k is the k^{th} eigenvalue of W (the eigenvalues are assumed arranged in descending order), and F_k^i the i^{th} coordinate of F_k. Such an embedding is called a $d-$embedding and denoted by $I_d(G)$. At the most, $s = n - 1$, but for some digraphs we may have $s < n - 1$. Because d is a non graduated distance (large distances are all equal to 1 (see Figure 2 for an example), dense components are more or less distributed on a hyper-sphere of $\mathbb{R}^p$ where $p+1$ is the number of dense components. Hence, for directed graphs we find again the results of [14] concerning the fact that each first principal axis will reveal more or less one dense component. We are now interested in knowing how the superior dimensions bear information about the way dense components are linked. The following of the paper explores that question on graphs $G_{p,q}$.

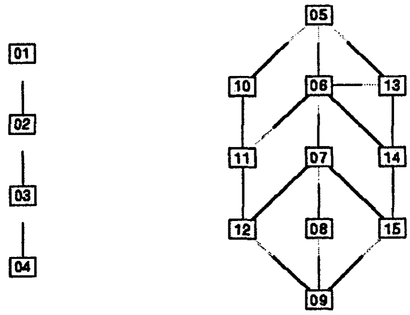

Figure 2. $d(1,4) = d(5,9) = 1$ even if both configurations are noticeably different

3 ILLUSTRATIVE EXAMPLES OF HIERARCHICAL CLUSTERING METHODS TO REVEAL AN ARTICULATION SINK SET

We propose to analyze not $I_d(G_{p,q})$ but some projections of $I_d(G_{p,q})$ onto subspaces generated by some factorial axes. This will allow a better understanding of the role of each factor of the Principal Co-ordinates Analysis for revealing the structure of $G_{p,q}$. We denote by $p_k\,[I_d(G_{p,q})]$ the projection of $I_d(G_{p,q})$ onto the space of the k first factorial axes. The points of $p_k\,[I_d(G_{p,q})]$ are clustered using agglomerative method. We have chosen to use a single linkage method but alternative methods may suit for the $G_{p,q}$ graphs. In the following, the results of the clusterings are represented by dendograms.

In that part, we lead a comparative study between an example of digraph $G_{p,q}$ with $p = q = 3$, and the digraph $\tilde{G}_{3,3}$ constructed from $G_{3,3}$ by reversing some arcs joining the articulation sink set $\overleftrightarrow{K^4}$ to the $\overleftrightarrow{K^i}$s in such a way that $\overleftrightarrow{K^4}$ is no more a sink set nor a pseudo-sink set (see Figure 3).

3.1 Clusterings of $G_{3,3}$

The graph $G_{3,3}$ has a $d-$embedding in $\mathbb{R}^7$ since the 4 last eigenvalues are equal to zero. The agglomerations of the points of the $p_k\,[I_d(G_{3,3})]$ are presented through dendograms of Figure 4 and the scree-graph corresponding to the 7 first eigenvalues associated with the principal axes appears in Figure 5. The scree-graph shows 3 plateaux, corresponding to the 1^{st} and 2^{nd}, 4^{th} and 5^{th}, and 6^{th} and 7^{th} eigenvalues. The relative contribution

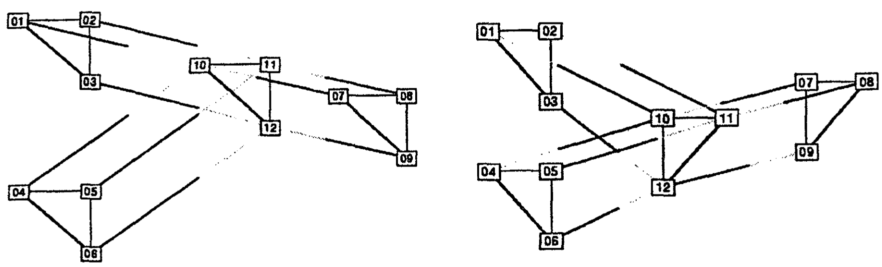

Figure 3. Digraphs $G_{3,3}$ (left) and $\tilde{G}_{3,3}$ (right). The conventions are the same as those of figure 1. Moreover, when there are two reciprocal arcs between two vertices, the line is entirely black. In $G_{3,3}$ the vertices of $\overleftrightarrow{K}^4$ are $\{10, 11, 12\}$.

(CTR) of $\overleftrightarrow{K}^4$ to the inertia of the first two axes (first plateau) is equal to zero although that of each other $\overleftrightarrow{K}^i$ is about $1/3$, and the CTR of $\overleftrightarrow{K}^4$ to the 3^{rd} axis is about 75%. Hence the first plateau corresponds to the two factorial dimensions that reveal the 3 complete digraphs $\overleftrightarrow{K}^i$ for $i \in \{1, 2, 3\}$ (see the first dendograms of the Figure 4) and the

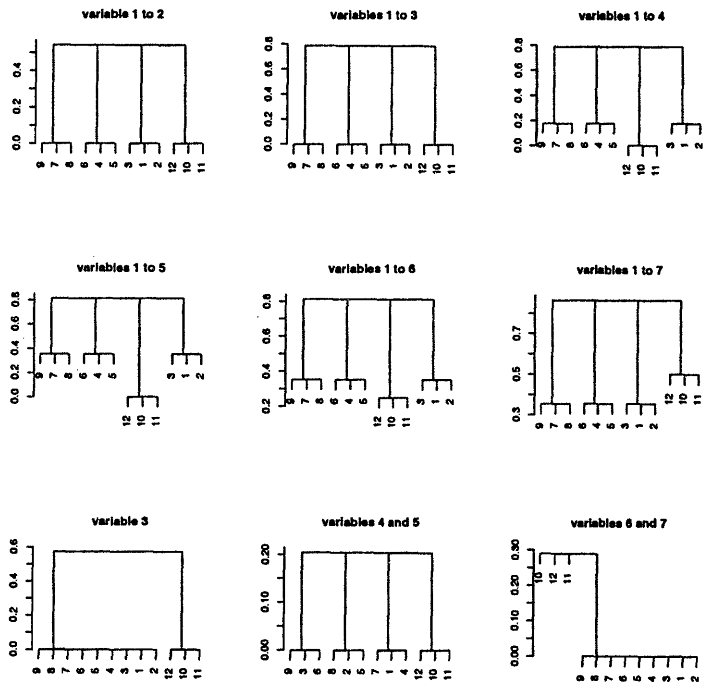

Figure 4. Dendograms of $G_{3,3}$ provided by a single linkage method applied on the vertices of $I_d\,(G_{3,3})$ or on their projections on some principal axes. The variables are the factors of the Principal Co-ordinates Analysis

3^{rd} factor bears $\overleftrightarrow{K}^4$. We verify here, on an example, the results announced at the end of the previous part. If we add one dimension to the space onto which we project $I_d\,(G_{3,3})$, the articulation set is distinguished by its aggregation in a first level. This difference remains until we take 7 variables. As soon as we take 7 variables, the articulation set is aggregated in second. Projections of $I_d\,(G_{3,3})$ onto the space generated by factors of the second plateau perfectly show that these dimensions reflect the way the 3 $\overleftrightarrow{K^i}$s are linked to the articulation set. The clustering of the projection onto that space builds the classes $\{1,4,7\}$, $\{2,5,8\}$, and $\{3,6,9\}$ of the vertices linked to the same vertex of $\overleftrightarrow{K^4}$. Along the factors of the 3^{rd} plateau all the vertices of the $\overleftrightarrow{K^1}$, $\overleftrightarrow{K^2}$ and $\overleftrightarrow{K^3}$ have their coordinates equal to 0, although the vertices of $\overleftrightarrow{K^4}$ are well separated with coordinates (0,-.289), (-.250,.144) and (.250,.144) respectively. That separation of these vertices is a notable difference between these factors and factor 3.

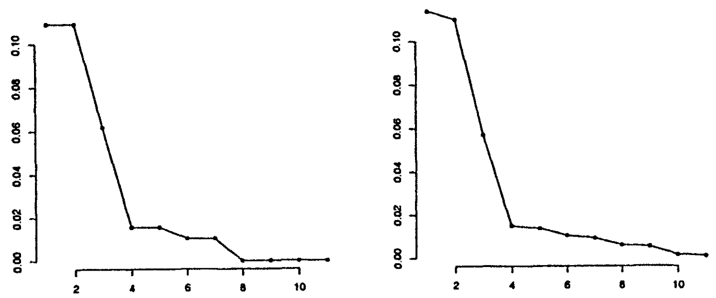

Figure 5. Scree-graphs for the Principal Co-ordinates Analysis of $I_d\,(G_{3,3})$ (left) and $I_d(\tilde{G}_{3,3})$ (right). The p^{th} eigenvalue is equal to the inertia of the d-embedding along the p^{th} principal axis

Let us notice that if we replace the sink set by a source set, the distance table between the vertices are the same, and *a fortiori* the results.

3.2 Study of $\tilde{G}_{3,3}$

The graph $\tilde{G}_{3,3}$ is obtained from $G_{3,3}$ by reversing the arcs $(2,11)$, $(1,10)$, $(4,10)$ and $(9,12)$. The scree-graph of $\tilde{G}_{3,3}$ has the same two important jumps as $G_{3,3}$, between the 2^{nd} and the 3^{rd} eigenvalues, and the 3^{rd} and the 4^{th}. On the other hand, it has no really pronounced plateau after the 4^{th} eigenvalue. The dendograms of Figure 6 reflect these results since the first difference appears when we take 4 dimensions. Moreover, the vertices of the articulation set are never agregated in 1^{st}. Finally, the dendograms relative to the variables $\{4,5\}$ and $\{6,7\}$ have not the form of those of $G_{3,3}$. It reinforces the idea that these dimensions code the matching and its direction.

3.3 A last example mixing $G_{3,3}$ and $\tilde{G}_{3,3}$

In this last example, we combine both digraphs $G_{3,3}$ and $\tilde{G}_{3,3}$ in the form of a "caterpillar" digraph G (see Figure 7).The scree-graph and some dendograms are presented in Figure 8. The scree-graph reveals 5 high eigenvalues corresponding to 6 complete digraphs. The analysis of the contribution of each vertex to different factors shows that inertia of the factors 4 and 5 are greatly due to both articulation sets $\{7,8,9\}$ and $\{10,11,12\}$. Those articulation sets emerge in the clustering using all the dimension of the $d-$embedding by their last aggregation. The sink set is distinguished by its first aggregation in the clustering of $p_7\,[I_d(G)]$.

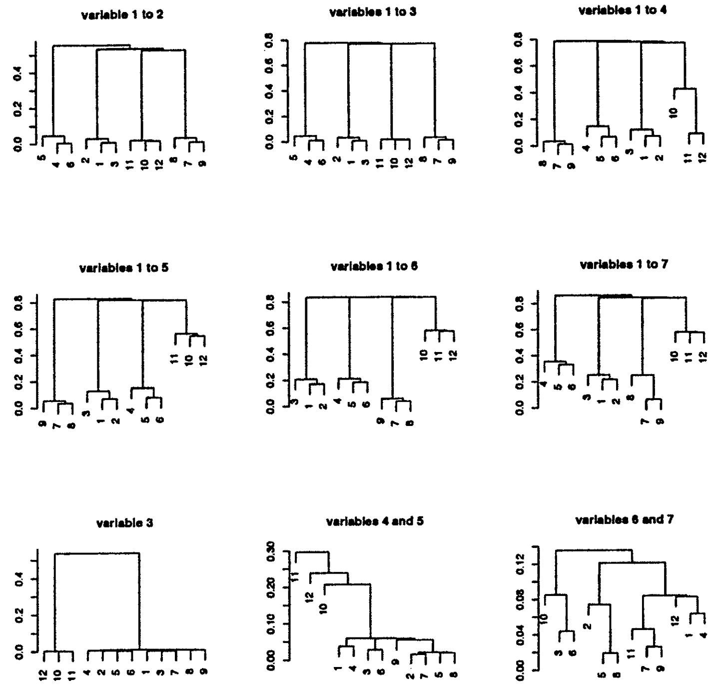

Figure 6. Dendograms of $\tilde{G}_{3,3}$

3.4 Assumptions for the following

That study of $G_{3,3}$ allows to make some assumptions about the way the clustering of $I_d\left(G_{p,q}\right)$ acts on the digraphs $G_{p,q}$:

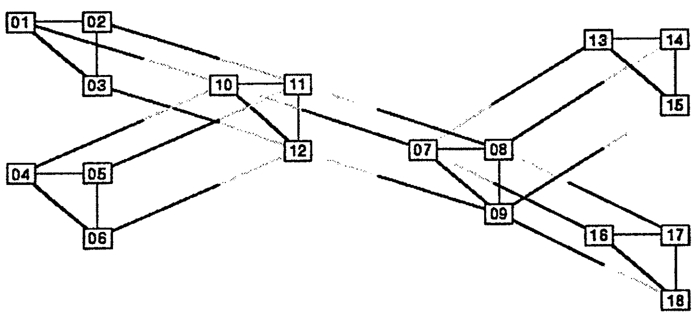

Figure 7. The digraph G is made of 6 complete digraphs with 3 vertices. Both induced by vertices $\{7,8,9\}$ and $\{10,11,12\}$ are articulation sets, only the first is a sink

- The $p-1$ first factors of the Principal Co-ordinates Analysis just allow to separate the $\overleftrightarrow{K^i}$s for $i \in \{1, \cdots, p-1\}$. Those results were revealed in [14]. The number p of complete digraphs is equal to the length of the first plateau of the scree-graph.
- The p^{th} factor reveals the articulation set $\overleftrightarrow{K^{p+1}}$.

- The other factors inform about the way the $\overleftrightarrow{K^i}$s are connected to $\overleftrightarrow{K^{p+1}}$. The scree-graph contains a second plateau of length $q-1$, and the existence of an articulation set $\overleftrightarrow{K^{p+1}}$ which is a sink appears in the aggregative orders of its vertices.

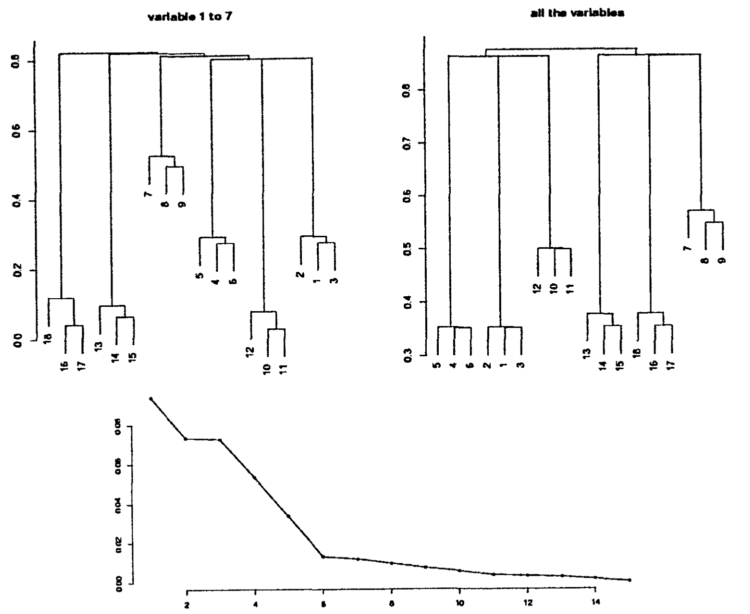

Figure 8. Dendograms for "variables 1 to 7" and "all the variables" are presented. The 7^{th} variable is the end of a light plateau which appears in the scree-graph. Both articulation sets are last aggregated when we take all the variables but the one which is a sink is aggregated in first for 7 variables.

4 THEORETICAL RESULTS FOR $G_{p,q}$ GRAPHS

We now give mathematical arguments to confirm and prove the above observations obtained for $p = q = 3$.

Let us first give some additional notations. The general entry of a matrix M will be denoted by m_{ij} and all the matrices we consider are real. The identity squared matrix of order n is denoted by I_n and the full of 1 column matrix of order n by $\mathbf{1}_n$. The transpose of a matrix M is denoted by M' and we set $\mathbf{1}_{n,m} = \mathbf{1}_n \mathbf{1}'_m$. Finally, we write $||\cdot||$ for the L_2−norm.

For such $G_{p,q}$ digraphs, it is easy to compute the distances between any two vertices $i \neq j$:

PROPOSITION 4.1 *Given k and l two distinct integers of* $\{1,\cdots,p\}$,

$$\begin{aligned}
&\textit{if } (i,j) \in \overleftrightarrow{K^{p+1}} \times \overleftrightarrow{K^{p+1}},\ d^2(i,j) = \frac{p}{2(p+q)} \\
&\textit{if } (i,j) \in \overleftrightarrow{K^{p+1}} \times \overleftrightarrow{K^{l}},\ d^2(i,j) = 1 - \left(\frac{1}{2q+1} + \frac{1}{2q+p}\right) \\
&\textit{if } (i,j) \in \overleftrightarrow{K^{l}} \times \overleftrightarrow{K^{l}},\ d^2(i,j) = \frac{1}{2q+2} \\
&\textit{if } (i,j) \in \overleftrightarrow{K^{l}} \times \overleftrightarrow{K^{k}},\ d^2(i,j) = \begin{cases} \frac{1}{2}\left(1+\frac{q}{q+1}\right) & \textit{if } i-j \equiv 0(q) \\ 1 & \textit{otherwise} \end{cases}
\end{aligned}$$

Let us just give, for instance, some elements of the proof for the two first cases. We recall that we have assumed that there is a loop at each vertex. When $(i,j) \in \overleftrightarrow{K^{p+1}} \times \overleftrightarrow{K^{p+1}}$, we have $|\Gamma^+(i) \,\Delta\, \Gamma^+(j)| = 0$, $|\Gamma^-(i) \,\Delta\, \Gamma^-(j)| = 2p$, and $|\Gamma^-(i)| = |\Gamma^-(j)| = q+p$. In the same way, when $(i,j) \in \overleftrightarrow{K^{p+1}} \times \overleftrightarrow{K^{l}}$, $|\Gamma^+(i) \,\Delta\, \Gamma^+(j)| = 2q-1$, $|\Gamma^-(i) \,\Delta\, \Gamma^-(j)| = 2q+p-2$, $|\Gamma^+(i)| = |\Gamma^+(j)| - 1 = q$, $|\Gamma^-(i)| = q+p$ and $|\Gamma^-(j)| = q$. ∎

In order to explain the existence of plateaux in the scree-graph, we now investigate the eigenvalues decomposition of the Torgerson matrix associated with $(V_{p,q}, d)$.

4.1 Eigenvalues of the Torgerson matrix associated with $(V_{p,q}, d)$

We need for that to give some definitions and properties of special matrices. A matrix M will be called of type I, and denoted by M_b^a, if its entries just take two values a and b according to whether it is a diagonal entry or not. In the following $m_{.j}$ will denote the sum of all the entries of the column j of a matrix M. If M is a $q \times q$ matrix of type I, $m_{.j} = a + (q-1)b$. A matrix of type II, denoted by M_a^a, is a matrix whose all entries are equal; it is easy to verify that $\mathbf{1}_q$ is an eigenvector of such a $q \times q$ matrix with $a \neq 0$.

A simple calculus gives the following result:

LEMMA 4.2 *The determinant of a $q \times q$ matrix M_b^a (with $a \neq b$) is*

$$(a + (q-1)\,b)\,(a-b)^{q-1}$$

The matrix M_b^a has one simple eigenvalue equal to $a + (q-1)\,b$ and one of multiplicity $q-1$ equal to $a-b$.

Remark 4.1 *The matrix $M_b^{a-\lambda} = M_b^a - \lambda I_q$ is of type II if and only if λ is an eigenvalue of M_b^a.*

PROPOSITION 4.3 *The Torgerson matrix of the Euclidean metric finite space $(V_{p,q}, d)$ is a $(p+1)$ by $(p+1)$ block matrix of type:*

$$W = \begin{pmatrix} A & B & \cdots & \cdots & B & C \\ B & A & B & \cdots & B & C \\ \vdots & B & A & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ B & \cdots & \cdots & B & A & C \\ C & \cdots & \cdots & \cdots & C & E \end{pmatrix} \quad (1)$$

where each block is a square matrix of order q, E corresponding to $\overleftrightarrow{K^{p+1}}$.
The matrices A, B and E are of type I. The matrices $(A - B)$ *and C are of type II.*

The general entry w_{ij} of W is equal to:

$$w_{ij} = \frac{1}{2}\left(d^2(i,\cdot) + d^2(j,\cdot) - d^2(i,j) - d^2(\cdot,\cdot)\right)$$

where

$$d^2(i,\cdot) = \frac{1}{n}\sum_{j=1}^{n} d^2(i,j) \text{ and } d^2(\cdot,\cdot) = \frac{1}{n}\sum_{j=1}^{n} d^2(i,\cdot)$$

The equality (1) and the form of matrices A, B, C and E then easily result from proposition 4.1. To prove that $(A - B)$ is of type II, one just has to verify that its diagonal and non-diagonal entries are equal. Considering that $d^2(i,.) = d^2(j,.)$ for any vertices i and j not in $\overleftrightarrow{K^{p+1}}$, the common value of the diagonal entries is $\frac{1}{2}d^2(1, q+1)$ and the common value of the non-diagonal ones is $\frac{1}{2}(d^2(1, q+2) - d^2(1, 2))$. Both expressions are equal to $\frac{1}{2}(1 + \frac{q}{q+1})$. ∎

THEOREM 4.4 *The characteristic equation of W is*

$$|\det(W - \lambda I_n)| = \left(\frac{q}{\lambda_1}\right)^{q-1} \lambda^{(q-1)(p-1)} (\lambda - \lambda_1)^{p-1} (\lambda - \lambda_3)^{q-1}$$
$$(\lambda - \lambda_4)^{q-1} (\lambda - \lambda_2)(\lambda - \lambda_5)$$

where $\lambda_1 = \frac{q}{4}\left(1 + \frac{q}{q+1}\right)$, $\lambda_3 = \frac{1}{4}\left(\frac{p}{q+1}\right)$, $\lambda_4 = \frac{1}{4}\left(\frac{p}{p+q}\right)$, $\lambda_5 = 0$ *and*

$$\lambda_2 = \frac{q}{2} - \frac{p^2 + 4p + 1}{4(p+1)} + \frac{p}{2(p+1)(2q+1)} + \frac{1}{4(q+1)} + \frac{p^2}{2(p+1)(2q+p)} + \frac{p^2}{4(q+p)}$$

Adding all the columns, excepted the last one, to the first column, substracting the first row of blocks from the following $p - 1$ rows of blocks, and substracting $1/p$ of the first column of blocks from the $p - 1$ following ones, we obtain:

$$\det(W - \lambda I_n) = \begin{vmatrix} A_\lambda & B - \frac{1}{p}A_\lambda & \cdots & \cdots & B - \frac{1}{p}A_\lambda & C \\ 0 & A - B - \lambda I_q & 0 & \cdots & 0 & 0 \\ \vdots & 0 & A - B - \lambda I_q & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & \cdots & 0 & A - B - \lambda I_q & 0 \\ pC & 0 & \cdots & \cdots & 0 & E - \lambda I_q \end{vmatrix}$$

where $A_\lambda = A + (p-1)\,B - \lambda I_q$. By reordering the rows and columns of the matrix, the first row and column of blocs becoming the last ones, there exists $\alpha \in \mathbb{N}$ such that:

$$\det\,(W - \lambda I_n) = (-1)^\alpha \begin{vmatrix} A-B-\lambda I_q & 0 & \cdots & \cdots & 0 & 0 \\ 0 & A-B-\lambda I_q & \ddots & \vdots & \vdots & \vdots \\ \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \cdots & \cdots & 0 & A-B-\lambda I_q & 0 & 0 \\ 0 & \cdots & \cdots & 0 & E-\lambda I_q & pC \\ B-\frac{1}{p}A_\lambda & \cdots & \cdots & B-\frac{1}{p}A_\lambda & C & A_\lambda \end{vmatrix}$$

and

$$|\det(W - \lambda I_n)| = |\det(A - B - \lambda I_q)^{p-1}| \cdot |\det((E - \lambda I_q)A_\lambda - pC^2)|.$$

• Let us note $\frac{\lambda_1}{q}$ the common entry of $A-B$, we have $\frac{\lambda_1}{q} = \frac{1}{2}d^2\,(1, q+1) = \frac{1}{4}(1 + \frac{q}{q+1})$ and $|\det\,(A - B - \lambda I_q)\,| = (\frac{q}{\lambda_1})^{q-1}\lambda^{q-1}\,(\lambda - \lambda_1)$.
• The matrix $(E - \lambda I_q)\,A_\lambda - pC^2$ is of type I and is denoted by $M_{b(\lambda)}^{a(\lambda)}$. The polynomial $a\,(\lambda)$ is monic of degree 2 in the indeterminate λ and $b\,(\lambda)$ is a polynomial of degree 1. Since the eigenvalues of W are all real, there exists four reals λ_2, λ_3, λ_4 and λ_5 such that $a\,(\lambda) + (q-1)\,b\,(\lambda) = (\lambda - \lambda_2)\,(\lambda - \lambda_5)$ and $a - b = (\lambda - \lambda_3)\,(\lambda - \lambda_4)$.
• Calculus of λ_3, λ_4 and λ_5 :
A, B, and E are type I matrices, and we respectively denote by a_{ii}, b_{ii} and e_{ii} their diagonal entries, and by a_{ij}, b_{ij} and e_{ij} their non-diagonal entries. The matrix C is of type II with a common entry denoted by c. Since C is of type II, the matrix C^2 is also of type II with entries equal to $c^2 q$. If $M_{\tilde{b}(\lambda)}^{\tilde{a}(\lambda)}$ denotes the matrix $(E - \lambda I_q)\,A_\lambda$, we have

$$\det M_{b(\lambda)}^{a(\lambda)} = \left(\widetilde{a}\,(\lambda) - c^2pq^2 + (q-1)\,\widetilde{b}\,(\lambda)\right)\left(\widetilde{a}\,(\lambda) - \widetilde{b}\,(\lambda)\right)^{q-1}$$

Hence, the eigenvalues of multiplicity $q-1$ of $M_{b(\lambda)}^{a(\lambda)}$ and $M_{\tilde{b}(\lambda)}^{\tilde{a}(\lambda)}$ are the same. Since

$$\det M_{\tilde{b}(\lambda)}^{\tilde{a}(\lambda)} = \det(E - \lambda I_q)\det(A + (p-1)B - \lambda I_q)$$

and E and $A + (p-1)\,B$ are of type I, proposition 4.2 gives

$$\begin{aligned} \lambda_3 &= a_{ii} - a_{ij} + (p-1)\,(b_{ii} - b_{ij}) = p\,(b_{ii} - b_{ij}) \\ \lambda_4 &= e_{ii} - e_{ij} \end{aligned}$$

The result is obtained by using the expressions given by proposition 4.1.

To prove that $\lambda_5 = 0$, it is sufficient to establish that

$$\det M_{b(0)}^{a(0)} = \det[E(A-B) + pEB - pC^2] = 0$$

Since $M_{b(0)}^{a(0)}$ is of type I, it suffices to prove that the sum of all the elements of a column is equal to zero for the sum of all n rows to be equal to 0. Since W is a Torgerson matrice, that property is true on it, and:

$$\begin{aligned} 0 &= w_{\cdot n} = pqc + e_{\cdot 1} && (2) \\ 0 &= w_{\cdot 1} = (p-1)\,b_{\cdot 1} + a_{\cdot 1} + qc = pb_{\cdot 1} + \lambda_1 + qc\,. && (3) \end{aligned}$$

Denote by α_{ij} , β_{ij} and δ_{ij} the entries of matrices $E(A-B)$, pEB and pC^2 respectively, it follows from (2) that

$$\begin{aligned} \alpha_{\cdot 1} &= \lambda_1 e_{\cdot 1} = -\lambda_1 pqc \\ \delta_{\cdot 1} &= pq^2c^2 \\ \beta_{\cdot 1} &= pe_{\cdot 1}b_{\cdot 1} = -p^2qcb_{\cdot 1} \end{aligned}$$

Hence, equation (3) multiplied by pqc becomes:

$$\alpha_{\cdot 1} + \beta_{\cdot 1} - \delta_{\cdot 1} = 0$$

which means that the sum of the elements of the first column of $M_{b(0)}^{a(0)}$ is equal to 0.
• Calculus of λ_2:
Eigenvalues λ_2 and λ_5 are the two zeros of the polynomial

$$P(\lambda) = \tilde{a}\,(\lambda) - c^2pq^2 + (q-1)\,\tilde{b}\,(\lambda)$$

where

$$\begin{aligned} \tilde{a}\,(\lambda) &= \lambda^2 - \lambda\,(e_{ii} + a_{ii} + (p-1)b_{ii}) + (\alpha_{ii} + \beta_{ii}) \\ \tilde{b}\,(\lambda) &= -\lambda\,(e_{ij} + a_{ij} + (p-1)b_{ij}) + (\alpha_{ij} + \beta_{ij}) \end{aligned}$$

Using $\lambda_5 = 0$ it is not difficult to factorize $P(\lambda)$ and write the expression of λ_2:

$$\lambda_2 = e_{\cdot 1} + a_{\cdot 1} + (p-1)\,b_{\cdot 1} = \left(1 + \frac{1}{p}\right) e_{\cdot 1}$$

Moreover,

$$e_{\cdot 1} = d\,(i,\cdot) - \frac{1}{2}d\,(\cdot,\cdot) + (q-1)\left(d\,(i,\cdot) - \frac{1}{2}d\,(i,j) - \frac{1}{2}d\,(\cdot,\cdot)\right)$$

where $i \in \overleftrightarrow{K^{p+1}}$. The proposition 4.1 allows to formulate an expression of λ_2 that is simplified with MAGMA Computational Algebra System V2.7-2, and we obtain:

$$\lambda_2 = \frac{q}{2} - \frac{p^2+4p+1}{4(p+1)} + \frac{p}{2(p+1)(2q+1)} + \frac{1}{4(q+1)} + \frac{p^2}{2(p+1)(2q+p)} + \frac{p^2}{4(q+p)}$$

∎

Consequence 4.5 *If* $q \geq 1$ *then* $\lambda_1 > \lambda_2$ *and* $\lambda_3 \geq \lambda_4$, *and if* $1 \leq p \leq q^2$ *then* $\lambda_2 > \lambda_3$.

If $q \geq 1$, only the inequality between λ_1 and λ_2 is not evident. We have

$$\lambda_1 - \lambda_2 = -\frac{1}{4} + \frac{p}{p+1}\left[1 - \frac{1}{4}\left(\frac{1}{q+\frac{1}{2}} + \frac{1}{\frac{q}{p}+\frac{1}{2}}\right)\right] + \frac{1}{4}\frac{p^2+1}{p+1} - \frac{1}{4}\frac{p^2}{p+q}$$

Since $q \geq 1$, we have $1 - \frac{1}{4}\left(\frac{1}{q+\frac{1}{2}} + \frac{1}{\frac{q}{p}+\frac{1}{2}}\right) \geq \frac{1}{3}$ and

$$\begin{aligned}\lambda_1 - \lambda_2 &\geq -\frac{1}{4} + \frac{1}{3}\frac{p}{p+1} + \frac{1}{4}\frac{p^2+1}{p+1} - \frac{1}{4}\frac{p^2}{p+q} \\ &\geq \frac{1}{4}\left(\frac{p^2+\frac{1}{3}p}{p+1} - \frac{p^2}{p+q}\right) > 0\end{aligned}$$

If $p \leq q^2$, the difference $\lambda_2 - \lambda_3$ may be rewritten:

$$\lambda_2 - \lambda_3 = \frac{N(\lambda_2 - \lambda_3)}{D(\lambda_2 - \lambda_3)}$$

where

$$\begin{aligned}N(\lambda_2 - \lambda_3) &= \left(2q^3 + 8q^4p^2 + 8q^5p + 2q^3p^3 + 8q^5 + 8q^4 + 8q^4p + p^3\right) \\ &\quad - \left(qp^2 + q^2p + 12q^2p^2 + 5q^2p^3 + 3qp^3 + 4q^3p^2 + 2qp^4 + p^4 + 4q^3p\right)\end{aligned}$$

and

$$D(\lambda_2 - \lambda_3) = 4(p+1)(2q+1)(2q+p)(q+1)(p+q)$$

Let us group some terms of the numerator of $\lambda_2 - \lambda_3$ together:

$$\begin{aligned}N(\lambda_2 - \lambda_3) &= (q^5 - q^2p) + (q^5 - qp^2) + (2q^4p^2 - 2q^2p^2) + (2q^5 + q^5p + 8q^4p - 10q^2p^2) \\ &\quad +(3q^5p - 3qp^3) + (5q^4p^2 - 5q^2p^3) + (4q^5p - 4q^3p^2) + (4q^5 - 4q^3p) \\ &\quad +(2q^3 + q^4p^2 + 8q^4 + p^3 + 2q^3p^3 - 2qp^4 - p^4)\end{aligned}$$

Except for the 4^{th} and the last one, it is easy to verify that each term of $N(\lambda_2 - \lambda_3)$ is positive if $1 \leq p \leq q^2$. Concerning the 4^{th} one just has to prove that $2q^5 + q^5p \geq 2q^2p^2$ since $8q^4p \geq 8q^2p^2$. Assuming that $q \geq 1$, the inequality $2q^5 + q^5p \geq 2q^2p^2$ is equivalent to $P(p) = 2p^2 - pq^3 - 2q^3 \leq 0$. Under the assumption that $1 \leq p \leq q^2$, the condition $P \leq 0$ is true if $4 - q \leq \sqrt{q^2 + \frac{8}{q}}$ which is easily checked. On the other hand,

$$2q^3 + q^4p^2 + 8q^4 + p^3 + 2q^3p^3 > q^4p^2 + 2q^3p^3 \geq p^4 + 2qp^4$$

if $1 \leq p \leq q^2$ and the last expression is strictly positive.

We thus have the required result by noting that the denominator of $\lambda_2 - \lambda_3$ is strictly positive. ∎

In the following we shall assume that the condition $1 \leq p \leq q^2$ is satisfied.

The scree graph of the Torgerson matrix W then has one plateau of length $p-1$ separating the $\overleftrightarrow{K^1}, \cdots, \overleftrightarrow{K^p}$, followed by one dimension that isolates the articulation set $\overleftrightarrow{K^{p+1}}$. Then arrive two plateaux with dimension $q-1$. The dimension of the embedding space of $G_{p,q}$ is $p + 2q - 2$.

4.2 Eigenvectors of the Torgerson matrix associated with $(V_{p,q}, d)$

We then focus on the eigenvectors of W in order to get the coordinates of the points of $I_d\left(G_{p,q}\right)$. The result is given by the following proposition:

PROPOSITION 4.6 *Let*
- $F_1, \cdots, F_n$ *denote the eigenvectors of the Torgerson matrix* W,
- $B_{i,j}$ *be matrices with* $q-1$ *columns and* q *rows,*
- $A_{i,1}$, $A_{i,2}$, $A_{i,3}$, *and* $A_{i,4}$ *be matrices whose entries of a same column are equal, with* q *rows and* $p-1$, 1, $q-1$ *and* $q-1$ *columns respectively,*
- 0 *be the full of zeros matrix with* $(p-1)(q-1)+1$ *columns and* q *rows.*

Then we have:

$$(F_1, \cdots, F_n) = \begin{pmatrix} A_{1,1} & A_{1,2} & B_{1,3} & A_{1,4} & 0 \\ A_{2,1} & A_{1,2} & B_{2,3} & A_{1,4} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & A_{1,2} & B_{p,3} & A_{1,4} & \vdots \\ A_{p+1,1} & A_{2,2} & A_{p+1,3} & B_{2,4} & 0 \end{pmatrix}$$

Assume F is an eigenvector of W corresponding to an eigenvalue λ, and let us write F under the form $(F^1, \cdots, F^{p+1})$ where each F^j is a block of length q we may write $F^j = \left(F^j(1), \cdots, F^j(q)\right)$. The system $W \cdot F = \lambda F$ may be expanded, using the expression of W as a blockmatrix, as follow:

$$\begin{cases} (A-B)\cdot(F^1 - F^2) = \lambda\left(F^1 - F^2\right) \\ (A-B)\cdot(F^1 - F^3) = \lambda\left(F^1 - F^3\right) \\ \vdots \\ (A-B)\cdot(F^1 - F^p) = \lambda\left(F^1 - F^p\right) \\ A \cdot F^1 + B\cdot\left(F^2 + \cdots + F^p\right) + C \cdot F^{p+1} = \lambda F^1 \\ C\cdot\left(F^1 + \cdots + F^p\right) + E \cdot F^{p+1} = \lambda F^{p+1} \end{cases} \tag{4}$$

- 1st case : $\lambda = \lambda_1$

Since $A - B$ is of type II and not null, $\mathrm{Im}(A-B) = \mathrm{sp}\,\mathbf{1}_q$. The first $p-1$ equations of (4) lead to the existence of $p-1$ real constants α^j such that:

$$\forall j \in \{2; \cdots; p\}\ ,\ F^j = F^1 - \alpha^j \mathbf{1}_q$$

The numbers α^j may be chosen for $F_1, \cdots, F_{p-1}$ to be linearly independant, hence corresponding to a base of the eigensubspace associated with λ_1. With $F_i^1 = \alpha_i \mathbf{1}_q$, where α_i are real numbers, the first column of blocks $\{A_{1,1}, A_{2,1}, \cdots, A_{p+1,1}\}$ has the expected form.

- 2nd case : $\lambda = \lambda_3$

The system (4) is equivalent to

$$\begin{cases} F^1 = F^2 = \cdots = F^p \\ -C \cdot F^{p+1} = \left(A + (p-1)B - \lambda I_q\right) \cdot F^1 \\ pC \cdot F^1 = \left(\lambda I_q - E\right) \cdot F^{p+1} \end{cases} \tag{5}$$

The type I matrix $(\lambda_3 I_q - E)$ is not of type II since $\lambda_3 \neq \lambda_4$ (see remark 4.1), and then will be written E_c^b with $b \neq c$. Moreover, $pC \cdot F^1$ has the form $v\mathbf{1}_q$. Finally, the 3^{rd} row of system (5) has the expression:

$$\begin{cases} bF^{p+1}(1) + cF^{p+1}(2) + \cdots + cF^{p+1}(q) = v \\ cF^{p+1}(1) + bF^{p+1}(2) + \cdots + cF^{p+1}(q) = v \\ \vdots \\ cF^{p+1}(1) + cF^{p+1}(2) + \cdots + bF^{p+1}(q) = v \end{cases}$$

The subtraction of a row from its following involves $F^{p+1}(1) = F^{p+1}(2) = \cdots = F^{p+1}(q)$. Hence, by projection onto the second plateau, the vertices of the articulation set merge and the third column of blocks has the expected form.

• 3rd case : $\lambda = \lambda_4$

We obtain the same system as (5) with $(A + (p-1)B - \lambda_4 I_q)$ which is of type I but not II. Expanding the second equation, we find $F^1(1) = F^1(2) = \cdots = F^1(q)$. That result combined with the first equation of system (5) gives the 4^{th} column of blocks.

• 4th case : $\lambda = \lambda_2$

The first row of system (5) is again true. Within each of the 2^{nd} and the 3^{rd} row, the left term has the form $v\mathbf{1}_q$. Since $(A + (p-1)B - \lambda_2 I_q)$ and $(\lambda_2 I_q - E)$ are not of type I, it involves that $F^1(1) = F^1(2) = \cdots = F^1(q)$ and $F^{p+1}(1) = F^{p+1}(2) = \cdots = F^{p+1}(q)$. That gives the form of the 2^{nd} column of blocks. ∎

The form of the coordinates matrix of the points of $I_d(G_{p,q})$ then proves the preliminary aggregation of the vertices of the sink set with an agglomerative method applied on the points of a $p_k[I_d(G_{p,q})]$ where $k \in \{p+2, \cdots, p+q-1\}$, indexes of the 2^{nd} plateau eigenvalues.

4.3 Toward an extension to a $G_{p,q}$ perturbed graph

We said in introduction that an advantage of the statistical approach is its efficiency to reveal a vertex set that is not a sink set but a pseudo sink set. We consider digraphs $G_{p,q}(\epsilon)$ obtained from $G_{p,q}$ by reversing arcs of the matching. We investigate only this case to keep in the scope of the paper. However, the following discussion also applies when the whole number of arcs is modified. We assume that the number of reversed arcs is low to use the perturbation theory for matrix. For convenience we write M for $M(G_{p,q})$. Let $M(\varepsilon)$ be the adjacency matrix of $G_{p,q}(\epsilon)$ and take $\varepsilon = \frac{||M(\varepsilon)-M||}{||M(\varepsilon)-M||+||M||}$, we have:

$$M(\varepsilon) = M + T(\varepsilon) = M + \varepsilon U$$

where $T(\varepsilon)$ is a perturbation of M in terms of a $n \times n$ matrix with entries in $[-1;1]$ and U is a matrix depending on ε. Hence $\varepsilon \in [0;1]$ and $\varepsilon = 0$ if and only if M is without perturbation. Note that $||M(\varepsilon) - M||^2$ is equal to the number of reversed arcs. For a small perturbation, that is for ε in the neighborhood of 0, the matrix U is an holomorphic function of ε.

The perturbation of M affects D and W in terms of two matrices:

$$D(\varepsilon) = D + \varepsilon D^{(1)} \quad \text{and} \quad W(\varepsilon) = W + \varepsilon W^{(1)} .$$

Under the conditions of applicability of the perturbation theory, the eigenvalues and eigenvectors of W are holomorphic functions at $\varepsilon = 0$. The plateaux of the scree-graph of W are then transformed in pseudo plateaux of $W(\varepsilon)$.

The Czekanovsky-Dice index defined in part 2 may be generalized to quantitative variables in the following way:

PROPOSITION 4.7 *Let m_i and m_j be quantitative positive vectors of dimension n (the rows of $M(\varepsilon)$) and let $s(i,j)$ be the similarity index between i and j defined by:*

$$s(i,j) = \frac{2\,\langle m_i, m_j\rangle}{\|m_i\|^2 + \|m_j\|^2}$$

where $\langle\cdot,\cdot\rangle$ is a scalar product, the dissimilarity $d = \sqrt{1-s}$ is Euclidean.
If m_i and m_j are binary, d is the Czekanovsky-Dice index.

The different steps of the proof used in the binary case to show that d is Euclidean remains valid for quantitative positive variables and the n points with dissimilarities table $D(\varepsilon)$ may be isometrically embedded in $\mathbb{R}^n$. Because these coordinates are continuous functions of the eigenvectors and eigenvalues of $W(\varepsilon)$, they also move continuously with ε around $\varepsilon = 0$. If ε is small enough, the aggregation order of the vertices of the perturbed graph $G_{p,q}(\epsilon)$ is then the same as $G_{p,q}$.

It remains to clarify the conditions of application of the perturbation theory in the specific case we are interested in. These conditions are the convergence of various power series, which, in the particular case of a symmetrical matrix, may be reduced to:

$$\|\varepsilon\, W^{(1)}\| \leq \frac{e}{2}$$

where

$$e = \min_{i\in\{2,3,4\}} \{|\lambda_i - \lambda_{i-1}|\,;|\lambda_{i+1} - \lambda_i|\} \tag{6}$$

One may verify that $\|W^{(1)}\| \leq \sqrt{2}\|D^{(1)}\|$, and then if Ω is an upper bound of $\|D^{(1)}\|$, the theory is applicable if $\varepsilon \leq \frac{e}{2\sqrt{2}\Omega}$.

That inequality notably provides a sufficient condition on the maximum number of arcs that may be reversed from $G_{p,q}$ for a pseudo-sink set to be revealed. In the general case of graphs $G_{p,q}$, to find an efficient constant Ω carries on big calculations. We may however notice that if

$$p \sim q^{1+a}$$

where $a \in [0, 1]$, we have:

$$4(\lambda_1 - \lambda_3) \sim q$$
$$4(\lambda_1 - \lambda_2) \sim q$$
$$4(\lambda_3 - \lambda_4) \sim q^a$$
$$4(\lambda_4 - \lambda_5) \sim 1$$

Hence, as equation 6 may be used to each eigenvalue separately, for instance if a is low, e may be taken larger for λ_1 and λ_2 than for λ_3 and λ_4. So the first two eigenvalues, revealing the $\overleftrightarrow{K^i}$'s, are more resistant to perturbations. This fact is consistent with the capability of the method to reveal the $\overleftrightarrow{K^i}$'s whatever the perturbation. Not surprisingly, the last eigenvalues are the more affected: revealing articulation sink set is a more sensible task.

When $a < 0$, similar calculations lead to more restricting conditions since $4(\lambda_3 - \lambda_4) \sim q^{2a}$. To get an insight of the behavior of the method in that case, we give below (table 1) some simulated results. The following table presents the results of our simulations for revealing pseudo-sink sets on digraphs $G_{p,q}(\epsilon)$. The simulations are built from 3 parameters: p, q and the number N of arcs of the matching (that is entering $\overleftrightarrow{K^{p+1}}$) whose direction is reversed. These 3 parameters respectively varies in $\{3, \cdots, 9\}$, $\{3, \cdots, 10\}$ and $\{0, \cdots, min(\frac{pq}{2}, 23)\}$, $\frac{pq}{2}$ being the half of the number of arcs of the matching. For each value of the triple (p, q, N) we run 100 simulations (the choice of the arcs that are reversed is a random processing) and we give the percentage of time the vertices of $\overleftrightarrow{K^{p+1}}$ are first aggregated in the clustering of $p_{p+q-1}[I_d(G_{p,q})]$ where $p+q-1$ is the index of the end of the 2^{nd} plateau (we uses single linkage method for the clustering). For example, the framed number has to be read as follow: for 100 digraphs $G_{5,6}$ from which one has reversed at random 7 arcs of the matching, 85 present an aggregation of the vertices of $\overleftrightarrow{K^6}$ (the articulation set) on a first level, within a clustering of $p_{10}[I_d(G_{5,6})]$. No surprising is the fact that the results are closely related to the values of the parameters. Particularly and roughtly speaking, the larger p is, the larger the number of allowed inversion is. These simulations give encouraging results on the behaviour in practice of the method.

NOTES

1. The graphs are drawn with the PIGALE Toolkit (Copyright (C) 2001 Hubert de Fraysseix, Patrice Ossona de Mendez) downloaded from: "http://www.ehess.fr/centres/cams/person/hf/".
2. Data Analysis has been made with R ([10]).
3. We thanks both referee for their helpfull comments principally about the 3^{rd} part.
4. This work was partially supported by the research french ministry under contract ACI Cognitique COG85.

BIBLIOGRAPHY

[1] BANG-JENSEN J., GUTIN G., *Digraphs: Theory, Algorithms and Applications*, London, Springer, 2001.

[2] BARTHÉLEMY J.P., GUÉNOCHE A., *Les arbres et les représentations des proximités*, Paris, Masson, 1988.

[3] BENZECRI J.P.F., *Pratique de l'Analyse des Données*, Paris, Dunod, 1984.

[4] BERGE C., *Théorie des graphes*, Paris, Bordas, 1983.

[5] BRODER A., KUMAR R., et *al.*, "Graph structure in the Web", *9th International World Wide Web Conference*, Foretec Seminar, 2000.

[6] FERRE L., JOUVE B., "Partitioning of directed networks", *Second International Workshop Network Sampling*, Maastricht (Pays-Bas), 1-4 Mars 2000.

[7] FICHET B., LE CALVE G., "Structure géométrique des principaux indices de dissimilarité sur signe de présence-abscence", *Statistique et Analyse des Données* 9 (3), 1984, p. 11-44.

[8] GARGALLO Y, JOUVE B., "Le Web vu comme un réseau orienté", Colloque *Comprendre les Usages d'Internet* , ENS Ulm, Paris, 3-4 Décembre 1999.

[9] GOWER J.C., LEGENDRE P., "Metric and Euclidean Properties of dissilarity coefficients", *Journal of Classification* 3, 1986, p. 5-48.

[10] IHAKA R. and GENTLEMAN R., "R: a langage for Data Analysis and Graphics", *Journal of Computational and Graphical Statistics* 5 (3), 1996, p. 299-314.

[11] JOUVE B., ROSENSTIEHL P., IMBERT M., "A mathematical approach to the connectivity between the cortical visual areas of the macaque monkey", *Cerebral Cortex* (8), 1998, p. 28-39.

[12] JOUVE B., *La multiplicité des aires visuelles du cortex : approche par la théorie des graphes*, Thèse de Sciences Cognitives, EHESS, 1999.

[13] KATO T., *Perturbation theory for linear operators*, New York, Springer, 1966.

[14] KUNTZ P., *Représentation euclidienne d'un graphe abstrait en vue de sa segmentation*, Thèse de Mathématiques Appliquées, EHESS, 1992.

[15] PLOUX S., VICTORRI B., "Construction d'espaces sémantiques à l'aide de diction-naires de synonymes", *Traitement Automatique des Langues* 39 (1), 1998, p. 161-182.

[16] SAPORTA G., *Probabilités, analyse des données et statistique*, Paris, Technips, 1990.

[17] WATTS D.J., *Small Worlds*, Princeton, Princeton University Press, 1999.

Table 1.

N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
q=3																								
p=3	100	100	77	0																				
p=4	100	100	100	40	0	0	0																	
p=5	100	100	100	85	93	19	0	0																
p=6	100	100	100	85	97	67	18	0	0	0														
p=7	100	100	100	92	90	85	64	24	5	1	0													
p=8	100	100	100	95	96	93	78	61	28	5	1	0	3											
p=9	100	100	100	91	94	98	85	76	47	36	33	7	0	0										
q=4																								
3	100	100	82	54	12	16	26																	
4	100	100	100	96	84	35	24	7	17															
5	100	100	100	100	86	79	71	20	25	3	12													
6	100	100	100	100	95	75	90	68	26	6	13	1	10											
7	100	100	100	100	93	78	88	89	64	33	14	7	10	0	4									
8	100	100	100	100	91	76	85	93	77	70	57	31	10	4	10	1	17							
9	100	100	100	100	97	76	86	85	62	72	71	47	43	22	13	8	5	0	9					
q=5																								
3	100	100	87	61	32	2	13	41																
4	100	100	100	98	89	81	46	32	3	5	5													
5	100	100	100	100	99	92	73	58	41	25	12	13	8											
6	100	100	100	100	100	99	92	83	75	72	41	25	19	7	11	0								
7	100	100	100	100	100	98	92	78	75	82	76	63	30	15	12	2	6	4						
8	100	100	100	100	100	97	89	84	88	86	80	70	61	38	29	12	10	6	4	4	2			
9	100	100	100	100	100	97	84	76	80	83	75	65	67	66	55	44	22	10	7	5	2	8	1	
q=6																								
3	100	100	86	67	42	17	36	4	12	32														
4	100	100	100	100	96	91	78	47	47	16	12	22	36											
5	100	100	99	99	100	99	96	85	73	51	39	28	38	5	9	5								
6	100	100	100	100	100	100	100	95	83	78	73	55	49	22	22	12	15	7	13					
7	100	100	100	100	100	100	98	92	85	77	77	76	66	59	37	18	8	4	16	1	11	4		
8	100	98	98	98	100	100	99	94	84	78	77	77	81	70	61	46	30	16	13	4	9	2	7	1
9	100	98	98	100	100	100	99	91	90	84	85	78	82	78	69	64	57	50	46	29	15	8	8	2
q=7																								
3	100	100	86	73	37	21	28	38	30	11	17													
4	100	100	100	100	99	81	92	71	42	32	24	10	17	33	46									
5	100	100	100	100	100	100	99	96	91	75	60	39	32	30	26	21	8	7						
6	100	100	98	100	99	100	100	97	95	88	82	77	66	55	50	40	22	16	13	7	14	10		
7	100	100	96	100	100	100	99	99	97	94	90	84	82	81	74	63	37	25	21	15	16	12	10	
8	100	98	97	100	99	100	100	98	94	89	87	86	81	82	75	73	64	53	40	26	18	10	9	5
9	100	99	96	98	98	98	100	99	98	96	90	84	79	76	73	67	66	64	64	59	47	33	21	14
q=8																								
3	100	100	88	73	52	37	32	35	45	29	16	14	19											
4	100	100	100	100	98	77	73	69	57	37	44	25	22	7	19	34	56							
5	100	100	100	100	100	100	100	96	88	87	76	61	51	31	34	39	38	20	16	12	19			
6	100	100	99	100	100	100	99	100	100	100	96	89	79	70	72	58	50	28	24	17	26	6	18	16
7	100	96	100	99	99	99	100	100	100	99	98	92	84	85	84	75	67	54	42	30	29	18	23	16
8	100	96	97	98	98	98	100	98	99	98	97	96	91	87	87	84	87	77	70	62	56	36	33	24
9	100	96	99	99	97	100	100	99	100	98	95	89	85	83	77	84	83	86	89	81	74	64	56	40
q=9																								
3	100	100	93	78	60	38	29	28	38	45	31	18	10	22										
4	100	100	99	100	99	92	83	55	67	43	34	29	23	15	18	14	21	38	52					
5	100	100	100	100	100	98	96	94	95	81	79	59	69	56	54	38	36	35	27	22	12	22	19	
6	100	100	99	100	100	97	100	100	100	99	97	92	84	77	64	61	57	51	41	37	34	28	26	31
7	100	100	100	100	100	100	98	99	100	100	100	98	97	96	90	87	73	68	66	56	47	39	30	18
8	100	97	98	99	100	100	99	100	100	100	100	100	97	95	95	89	89	86	84	79	72	67	58	45
9	100	95	97	99	100	97	98	98	100	100	100	98	97	96	91	85	82	78	77	78	76	74	65	63
q=10																								
3	100	100	98	80	59	46	38	42	34	35	37	30	24	19	14	14								
4	100	100	100	100	99	89	87	66	56	38	40	24	23	20	30	21	27	20	33	33	44			
5	100	100	100	100	100	98	97	89	86	88	75	76	62	67	64	54	47	39	39	36	24	20	13	20
6	100	99	100	100	100	100	100	99	99	99	98	98	94	92	84	77	69	65	60	45	39	38	33	39
7	100	98	100	100	100	100	99	100	100	100	100	100	99	99	96	92	91	83	80	72	67	55	50	39
8	100	100	97	99	100	99	100	99	100	100	100	100	99	99	97	93	88	83	84	79	76	71	60	53
9	100	97	100	100	97	99	99	100	100	100	99	100	100	98	99	99	95	90	87	87	84	79	75	75