Double orthogonality and the nature of networks

P. Van Mieghem*

Delft University of Technology v1: September 2013 v2: October 2013 v3: October 2014

Abstract

Any symmetric real matrix exhibits double orthogonality in its eigenstructure. While the mathematical foundations are crystal clear, the physical meaning of its application to graphs, the underlying topological structure of a network, is surprisingly opaque. This short letter is meant to provoke and inspire in order to understand; in the spirit of Hilbert, "wir müssen wissen".

1 Eigenstructure of a symmetric matrix

1.1 Double orthogonality

Following the notation of [1], we denote by x_k the eigenvector of the symmetric matrix A belonging to the eigenvalue λ_k , normalized so that $x_k^T x_k = 1$. The eigenvalues of an $N \times N$ symmetric matrix $A = A^T$ are real and can be ordered as $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_N$. Let X be the orthogonal matrix with eigenvectors of A in the columns,

$$X = \left[\begin{array}{cccc} x_1 & x_2 & x_3 & \cdots & x_N \end{array} \right]$$

or explicitly in terms of the *m*-th component $(x_j)_m$ of eigenvector x_j ,

$$X = \begin{bmatrix} (x_1)_1 & (x_2)_1 & (x_3)_1 & \cdots & (x_N)_1 \\ (x_1)_2 & (x_2)_2 & (x_3)_2 & \cdots & (x_N)_2 \\ (x_1)_3 & (x_2)_3 & (x_3)_3 & \cdots & (x_N)_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_1)_N & (x_2)_N & (x_3)_N & \cdots & (x_N)_N \end{bmatrix}$$

The eigenvalue equation $Ax_k = \lambda_k x_k$ translates to the matrix equation $A = X\Lambda X^T$, where $\Lambda = \text{diag}(\lambda_k)$.

^{*}Faculty of Electrical Engineering, Mathematics and Computer Science, P.O Box 5031, 2600 GA Delft, The Netherlands; *email*: P.F.A.VanMieghem@tudelft.nl

The relation $X^T X = I = X X^T$ (see e.g. [1, p. 223]) expresses, in fact, double orthogonality. The first equality $X^T X = I$ translates to the well-known orthogonality relation

$$x_{k}^{T}x_{m} = \sum_{j=1}^{N} (x_{k})_{j} (x_{m})_{j} = \delta_{km}$$
(1)

stating that the eigenvector x_k belonging to eigenvalue λ_k is orthogonal to any other eigenvector belonging to a different eigenvalue. The second equality $XX^T = I$, which arises from the commutativity of the inverse matrix $X^{-1} = X^T$ with the matrix X itself, can be written as $\sum_{j=1}^{N} (x_j)_m (x_j)_k = \delta_{mk}$ and suggests us to define the row vector in X as

$$y_m = ((x_1)_m, (x_2)_m, \dots, (x_N)_m)$$
(2)

Then, the second orthogonality condition $XX^T = I$ implies orthogonality of the vectors

$$y_{k}^{T} y_{m} = \sum_{j=1}^{N} (x_{j})_{k} (x_{j})_{m} = \delta_{km}$$
(3)

The third combination, namely $y_k^T x_m = \sum_{j=1}^N (x_j)_k (x_m)_j$, does not seem to possess special properties (see Appendix B).

The sum over j in (3) can be interpreted as the sum over all eigenvalues. Indeed, the eigenvalue equation is

$$Ax\left(\lambda\right) = \lambda \ x\left(\lambda\right)$$

where a non-zero vector $x(\lambda)$ only satisfies this linear equation¹ if λ is an eigenvalue of A such that $x_j = x(\lambda_j)$. We have made the dependence on the parameter λ explicit and can interpret λ as a frequency that ranges continuously over all real numbers. Invoking the Dirac delta-function $\delta(t)$, we can write

$$\sum_{j=1}^{N} (x_j)_m (x_j)_k = \sum_{\lambda \in \{\lambda_1, \lambda_2, \dots, \lambda_N\}}^{N} (x(\lambda))_m (x(\lambda))_k$$
$$= \sum_{j=1}^{N} \int_{-\infty}^{\infty} \delta (\lambda - \lambda_j) (x(\lambda))_m (x(\lambda))_k d\lambda$$

Using the non-negative weight function

$$w(\lambda) = \sum_{j=1}^{N} \delta(\lambda - \lambda_j) = \delta\left(\det\left(A - \lambda I\right)\right) \left| \frac{d\det\left(A - xI\right)}{dx} \right|_{x=\lambda}$$

shows that

$$\sum_{j=1}^{N} (x_j)_m (x_j)_k = \int_{-\infty}^{\infty} w(\lambda) (x(\lambda))_m (x(\lambda))_k d\lambda = \delta_{mk}$$
(4)

¹Differentiation with respect to λ yields

$$Ax'(\lambda) = \lambda x'(\lambda) + x(\lambda)$$

illustrating that no derivative (of any order) of $x(\lambda)$ can satisfy $Ax(\lambda) = \lambda x(\lambda)$ for $\lambda \neq 0$.

The right-hand side in (4) is the continuous variant of (3) that expresses orthogonality between functions with respect to the weight function w (see e.g. [1, p. 313]). Specifically², the orthogonality property (4) shows that the set $\{(x(\lambda))_m\}_{1 \le m \le N}$ is a set of N orthogonal polynomials in λ .

1.2 Interpretations of the eigenstructure of a graph

The topology of a network can be represented by a graph G, consisting of a set of nodes connected by a set of links. Many matrices can be associated to a graph, such as the adjacency, Laplacian, signless and normalized Laplacian, modularity, incidence and distance matrix, etc.. We confine ourselves here to a simple, undirected graph G and to its corresponding symmetric adjacency matrix A.

Surprisingly little is known [1, Chapter 1] about the "physical" meaning of the eigenvalue λ_k and its corresponding eigenvector x_k (for each $1 \leq k \leq N$) of the adjacency matrix A. One of the best interpretations follows from the probabilistic matrix, $P = \Delta^{-1}A$, where $\Delta = \text{diag}(d_j)$ and the degree of node j is $d_j = \sum_{k=1}^{N} a_{kj}$. The largest eigenvector component $(x_1(P))_j$ of P, normalized as $u^T x_1(P) = 1$, where u is the all one vector, reflects the probability that a random walk on the graph G visits node j in the long run. Theorem 2.2.4 in Cvetković *et al.* [3] provides the explicit relation between the number $N_k(j)$ of walks of length k starting at node j in a non-bipartite graph and the eigenvector component $(x_1)_j$ of A as

$$\lim_{k \to \infty} \frac{N_k(j)}{\sum_{l=1}^N N_k(l)} = (x_1)_j$$

The case of k = 1 is studied most and the principal eigenvector x_1 corresponding to the spectral radius λ_1 can be regarded as a "dynamic" degree vector (see Appendix A), where each component $(x_1)_j$ reflects all possible walks passing through node j. Most insight and most relations in graph theory (see e.g. [1]) are based on the set $\{x_k\}_{1 \le k \le N}$ of eigenvectors. When studying spectral clustering, Von Luxburg *et al.* [4] show figures of several eigenvectors of the Laplacian as a function of the nodal components.

The vector y_m , defined in (2), reflects the role of the node m over all eigenfrequencies or eigenvalues of A. From a graph metrics point of view, we may argue that y_m specifies how important or "central" node m is with respect to the important or characteristic frequencies (i.e. the eigenfrequencies) of the graph matrix A, that specifies links in G. Perhaps, the following geometric interpretation is daring. Imagine that the graph G is embedded in some geometric structure with negligible mass compared to those of the nodes and links. For example, a planar graph on a large flexible sheet that can be brought into vibration by an external force. Thus, the force bends the sheet up and down so that waves travel over the sheet at certain frequency. The vector y_m may be interpreted as the displacements of node m on the sheet at the eigenfrequencies of the adjacency matrix of the graph G. Another view on the vector $y_m = (x (\lambda_1)_m, x (\lambda_2)_m, \ldots, x (\lambda_N)_m)$ or, more generally, on the set $\{(x (\lambda))_m\}_{1 \le m \le N}$ versus frequency λ , is inspired by our human vision: we perceive the real-world only via the frequency range of visible light, while we know (for example from Röntgen photographs) that additional information is revealed in other frequency bands of the spectrum.

²The eigendecomposition of a general tri-diagonal stochastic matrix in [2, Appendix] exemplifies how orthogonal polynomials as a function of λ enter.

Another thought is that, if the frequency interpretation of λ is correct, then we may use Fourier analysis to rigorously define localization in space or in frequency. For example, if all components of y_m are constant (or more or less of the same magnitude), then the Fourier (or Laplace) transform indicates that node m is very localized in space³. The opposite is also true, if only a few components of y_m in successive order are significant, hence localized in the frequency domain, then node m is globally connected in the spacial domain.

The case $y_m^T y_m = \sum_{j=1}^N (x_j)_m^2 = 1$ in second orthogonality relation (3) means that, considered over all eigenvalues (eigenfrequencies or eigenmodes) of a graph, each node in *G* is equally important. In other words, the often associated importance to high-degree nodes seems only partially true, i.e. only for certain eigenfrequencies and certain types of graphs (the largest eigenvalue and graphs with large spectral gap; see Section A below).

The equations (1) and (3) constitute in total N^2 orthogonality conditions associated to the matrix X, whose rank is N (since $|\det X| = 1$). Although the N^2 elements of X represent a set of independent row and/or column vectors, curiously, an equal number of additional conditions is embedded in them. The latter seems to indicate that considerable information condensation can be attained, raising the question how many bits are minimally needed to reconstruct X exactly. Any orthogonal matrix describes a rotation of an orthogonal set of basis vectors in the N-dimensional space, which suggests that, beside N bits (corresponding to the basis vectors e_j), N rotation angles (N real numbers) are needed. Similar considerations have likely led Cvetković *et al.* [3] to define graph angles. The fundamental weight $w_k = u^T x_k = \sum_{j=1}^N (x_k)_j$ provides additional information to determine the eigenvector components, illustrating its important role⁴. The graph angle γ_k is related to fundamental weight by $\cos \gamma_k = \frac{w_k}{\sqrt{N}}$. These considerations may hint that only 2N real numbers (eigenvalues and fundamental weights or graph angles) are needed to construct the graph G exactly. A rigorous proof of the minimum number of bits needed to construct a graph is currently an open problem.

2 Summary

The key relation $A = X\Lambda X^T$ shows that all "topological" information about the graph (left-hand side) is contained in the "spectral space" (right-hand side). While linear algebra, in particular eigenvalue decomposition, is a mature branch of mathematics, the physical meaning of its application to networks and graphs remains puzzling. Many articles have been written on graph metrics and many will still appear. Nearly all metrics are correlated and so far there is no generally accepted set that characterizes a graph without loosing too much information. The set of orthogonal – thus uncorrelated – vectors y_1, y_2, \ldots, y_N mathematically reflects the *complete* information about the importance of each node in G over all eigenfrequencies $\lambda_1, \lambda_2, \ldots, \lambda_N$ and can thus be considered as an "ideal" nodal centrality set of metrics.

³Recall that $\sin(ar) = \frac{e^{iar} - e^{-iar}}{2i}$ (broad in space) has a Fourier transform proportional to $\delta(f-a) - \delta(f+a)$ (peaked in frequency domain).

⁴We mention [1, p. 41] that the vector $w = (w_1^2, w_2^2, \ldots, w_N^2)$ satisfies $V_N(\lambda) w = \mathbf{N}$, where $V_N(\lambda)$ is the Vandermonde matrix of the vector with eigenvalues $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ and where the vector $\mathbf{N} = (N_1, N_2, \ldots, N_N)$ has as components the total number of walks with k hops, $N_k = u^T A^k u$. Since $V_N(\lambda)$ can be inverted when all eigenvalues are different, the fundamental weights w can be expressed in terms of the eigenvalues and the number of walks in the graph G.

Let A and $A_{G\setminus\{j\}}$ denote the adjacency matrix of the graph G and of the graph $G_{\setminus\{j\}}$ in which node j is removed from G, respectively. The square of the j-th component of eigenvector x_k of A belonging to eigenvalue λ_k with multiplicity 1 equals [5]

$$(x_k)_j^2 = -\frac{1}{c'_A(\lambda_k)} \det\left(A_{G\setminus\{j\}} - \lambda_k I\right)$$
(5)

where $c_A(\lambda) = \det(A - \lambda I)$ is the characteristic polynomial of A and $c'_A(\lambda) = \frac{dc_A(\lambda)}{d\lambda}$. Akin to sensitivity or robustness analyses on graphs [6], the elegant expression (5) associates $(x_k)_j^2$ to the *impact of the removal* of node j in the graph G at the characteristic frequency λ_k corresponding to eigenvector x_k of a graph matrix.

Unfortunately, this positive view also has darker sides. First, each graph matrix (such as e.g. the Laplacian, adjacency and modularity matrix) possesses such doubly orthogonal eigenvectors. Thus, the N nodal frequency centrality metrics for node m, i.e. the components or squared components of y_m , are also dependent on the specific graph matrix. Hence, there does not seem be a single ideal set of centrality metrics per graph. Still, we are confronted with the choice of the "ideal" centrality metrics over the space of all possible (symmetric) graph matrices. Second and as mentioned several times, the meaning of both the impact (or amplitude) and characteristic frequency are waiting for an explanation useful to networks. Third, from a practical point of view, the vectors y_1, y_2, \ldots, y_N , though complete and uncorrelated, require global information, i.e. the full knowledge of the adjacency matrix of G. Even if we would understand the meaning of each vector y_m or of its squared components $(x_k)_m^2$, their complete use (i.e. all components over all frequencies λ_k) as a centrality metrics set is hardly feasible for large networks. Fourth, besides this computationally infeasibility objection, we can question the normalizations $x_k^T x_k = 1$ for each $1 \le k \le N$ that consider each eigenvalue as equally important. Indeed, for a positive semi-definite symmetric matrix A (whose eigenvalues are non-negative), we can write the spectral decomposition

$$A = \sum_{k=1}^{N} \lambda_k x_k x_k^T = \sum_{k=1}^{N} \left(\sqrt{\lambda_k} x_k \right) \left(\sqrt{\lambda_k} x_k \right)^T$$

which suggests to scale the importance of an eigenvector as $v_k = \sqrt{|\lambda_k|} x_k$. Clearly, the eigenvectors corresponding to the larger (in absolute value) eigenvalues deserve more weight, as earlier was exploited in graph reconstructability [7] and only a few of the larger ones may be sufficient as centrality metrics.

Acknowledgement. The work is supported by EU project CONGAS (Grant No. FP7-ICT-2011-8-317672) and, partly by EU NoE EINS (FP7-ICT-288021). I am grateful to Mark Newman, Dragan Stevanović and Jil Meijer for their constructive comments.

References

- [1] P. Van Mieghem. Graph Spectra for Complex Networks. Cambridge University Press, Cambridge, U.K., 2011.
- [2] P. Van Mieghem. Decay towards the overall-healthy state in SIS epidemics on networks. arXiv:1310.3980, 2013.
- [3] D. Cvetković, P. Rowlinson, and S. Simić. *Eigenspaces of graphs*. Cambridge University Press, Cambridge, U.K., 1997.
- [4] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. The Annals of Statistics, 36(2):555– 586, 2008.

- [5] P. Van Mieghem. Graph eigenvectors, fundamental weights and centrality metrics for nodes in networks. arXiv:1401.4580, 2014.
- [6] S. Trajanovski, J. Martin-Hernandez, W. Winterbach, and P. Van Mieghem. Robustness envelopes of networks. *Journal of Complex Networks*, 1:44–62, 2013.
- [7] D. Liu, H. Wang, and P. Van Mieghem. Spectral perturbation and reconstructability of complex networks. *Physical Review E*, 81(1):016101, January 2010.
- [8] C. D. Meyer. Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000.
- [9] P. Van Mieghem. Epidemic phase transition of the SIS-type in networks. Europhysics Letters (EPL), 97:48004, 2012.

A The degree vector d

Any vector in the N-dimensional space can be written as a linear combination of orthogonal vectors x_1, x_2, \ldots, x_N that span that space. Hence, the degree vector $d = (d_1, d_2, \ldots, d_N)$ of a graph G, where d_j denotes the degree of node j, can be written as

$$d = \sum_{j=1}^{N} r_j x_j$$

where the scalars r_j are computed using orthogonality of the vectors x_1, x_2, \ldots, x_N . Multiplying both sides by x_m^T and using the orthogonality $x_m^T x_j = \delta_{mj}$ yields

$$x_m^T d = r_m$$

Further, d = Au (see e.g. [1, p. 15]), where the vector u = (1, 1, ..., 1) is the all-one vector so that

$$x_m^T d = x_m^T A u = (Ax_m)^T u = \lambda_m x_m^T u = \lambda_m u^T x_m$$

Hence, we find that⁵ the spectral decomposition of the degree vector,

$$d = \sum_{j=1}^{N} \lambda_j \left(u^T x_j \right) x_j \tag{6}$$

where we call the real numbers

$$w_j = u^T x_j = \sum_{k=1}^N (x_j)_k$$

fundamental weights, that turn out to be equally important as the eigenvalues. Further (see e.g. [1, p. 33]), we have that

$$N = \sum_{j=1}^{N} (u^{T} x_{j})^{2} \text{ and } 2L = \sum_{j=1}^{N} \lambda_{j} (u^{T} x_{j})^{2}$$

⁵When we write the spectral decomposition (6) of the k-th component of the degree vector as

$$d_{k} = \sum_{j=1}^{N} \left(d^{T} x_{j} \right) \left(x_{j} \right)_{k} = \sum_{j=1}^{N} \sum_{m=1}^{N} d_{m} \left(x_{j} \right)_{m} \left(x_{j} \right)_{k} = \sum_{m=1}^{N} d_{m} \left\{ \sum_{j=1}^{N} \left(x_{j} \right)_{m} \left(x_{j} \right)_{k} \right\}$$

Since such a spectral decomposition holds for any N-dimensional vector, we naturally find the second type (3) of orthogonality.

and, in general, the total number of walks with k hops in the graph G equals

$$N_k = \sum_{j=1}^N \lambda_j^k \left(u^T x_j \right)^2$$

The general relation for diagonizable matrices, $f(A) = \sum_{k=1}^{N} f(\lambda_k) x_k x_k^T$, valid for a function f defined on the eigenvalues $\{\lambda_k\}_{1 \le k \le N}$ (see e.g. [8, p. 526]), reduces for the element for node j to

$$(f(A))_{jj} = \sum_{k=1}^{N} f(\lambda_k) (x_k)_j^2$$

illustrating that the squares of the eigenvector component arise as weights for $f(\lambda_k)$ to specify a function of the adjacency matrix A at node j. In particular, for powers $f(z) = z^n$, nice formulae appear: for n = 0, we find the second orthogonality relation (3) and for n = 1 (since $A_{jj} = 0$, from which trace $(A) = \sum_{k=1}^{N} \lambda_k = 0$)

$$0 = \sum_{k=1}^{N} \lambda_k \left(x_k \right)_j^2$$

while for n = 2 (since $(A^2)_{jj} = d_j$)

$$d_j = \sum_{k=1}^N \lambda_k^2 \left(x_k \right)_j^2 \tag{7}$$

From (6) and (7), we find that the degree d_j of node j can be expressed as

C

$$d_{j} = \sum_{k=1}^{N} \lambda_{k} w_{k} (x_{k})_{j} = \sum_{k=1}^{N} \lambda_{k}^{2} (x_{k})_{j}^{2}$$

Now, we rewrite (6) as

$$l = \lambda_1 \left(u^T x_1 \right) x_1 + c$$

where the correction vector c equals

$$c = \sum_{j=2}^{N} \lambda_j \left(u^T x_j \right) x_j$$

The correction vector is c = 0 only for regular graphs, where the principal eigenvalue is $x_1 = \frac{1}{\sqrt{N}}u$ and $u^T x_k = 0$ for each $2 \le k \le N$ because eigenvectors are orthogonal. If the correction vector c is negligibly small (e.g. when the spectral gap is large $(\lambda_1 >> \lambda_2)$ or for almost regular graphs or in other cases that we still need to investigate), then

$$d \approx \lambda_1 \left(u^T x_1 \right) x_1 \tag{8}$$

In simple dynamic processes on a network, such as SIS epidemics⁶, the vector v with the nodal infection probabilities is proportional to x_1 close to the epidemic phase transition. Only in those graphs obeying (8) where the degree vector d is (approximately) proportional to the principal eigenvector x_1 , the dynamics (e.g. v) is directly proportional to the graph's topological structure.

⁶Also Kuramoto synchronization, see e.g. [9].

B The set of orthogonal vectors y_1, y_2, \ldots, y_N and the network's topology

We now proceed by writing the "classical" eigenvector x_j as a linear combination of the set of orthogonal vectors y_1, y_2, \ldots, y_N that span the N-dimensional space,

$$x_k = \sum_{j=1}^N b_{kj} y_j$$

where the real number

$$b_{km} = x_k^T y_m = \sum_{l=1}^N (x_k)_l (x_l)_m$$

Written in matrix form

$$X = BY \tag{9}$$

where the matrix Y is

$$Y = \begin{bmatrix} (y_1)_1 & (y_2)_1 & (y_3)_1 & \cdots & (y_N)_1 \\ (y_1)_2 & (y_2)_2 & (y_3)_2 & \cdots & (y_N)_2 \\ (y_1)_3 & (y_2)_3 & (y_3)_3 & \cdots & (y_N)_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (y_1)_N & (y_2)_N & (y_3)_N & \cdots & (y_N)_N \end{bmatrix} = \begin{bmatrix} (x_1)_1 & (x_1)_2 & (x_1)_3 & \cdots & (x_1)_N \\ (x_2)_1 & (x_2)_2 & (x_2)_3 & \cdots & (x_2)_N \\ (x_3)_1 & (x_3)_2 & (x_3)_3 & \cdots & (x_3)_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_N)_1 & (x_N)_2 & (x_N)_3 & \cdots & (x_N)_N \end{bmatrix} = X^T$$

Hence, we have that

$$X = BX^T$$

and, after right-multiplying by X and using $X^T X = I$, we find that

$$B = X^2$$

Since the inverse $X^{-1} = X^T$, consistency is found by inverting (9),

$$Y = B^{-1}X = (X^T)^2 X = X^T$$

We are now prepared to write the adjacency matrix A in terms of the orthogonal vectors y_1, y_2, \ldots, y_N . Starting from (see e.g. [1, p. 2])

$$A = X\Lambda X^T = BY\Lambda (BY)^T = BY\Lambda Y^T B^T$$

or from (see e.g. [1, p. 226])

$$A = \sum_{k=1}^{N} \lambda_k x_k x_k^T = \sum_{k=1}^{N} \lambda_k \left(\sum_{j=1}^{N} b_{kj} y_j \right) \left(\sum_{l=1}^{N} b_{kl} y_l \right)^T$$
$$= \sum_{j=1}^{N} \sum_{l=1}^{N} \left(\sum_{k=1}^{N} \lambda_k b_{kj} b_{kl} \right) y_j y_l^T$$

shows that the orthogonal vectors y_1, y_2, \ldots, y_N are, by no means, naturally related to the topology of the graph G, because the expressions in terms of orthogonal vectors y_1, y_2, \ldots, y_N are more complicated and less transparent than expressing A in terms of its eigenvectors.