Graph properties and invariants, by their associate matrices

Angel Garrido, Facultad de Ciencias de la UNED

Abstract

It will be very interesting to analyze the invariants and properties of graphs by their different associate matrices. So, we describe now such powerful tools, showing its advantages and disadvantages, and the way to obtain information starting from such mathematical expresions.

Keywords: Graph Theory, Combinatorics, Enumeration of graphs, Graph labeling, Asymptotic Analysis.

Mathematics Subject Classification: 68R10, 68R05, 05C78, 78M35.

1. Adjacency Matrix

The *adjacency matrix* of a finite directed or undirected n-graph (DG or UG) is the $(nxn) - matrix$ where the non-diagonal entry, a_{ij}, is the number of edges that connect from the node i to the node j. And the diagonal entry a_{ii} is either twice the number of loops at i, or just such number of loops, depending on our mathematical requirements.

As there exists a unique adjacency matrix for each graph, up to permutation rows and columns, and it is not the adjacency matrix of any other graph, we dispose by this tool of an algebraic characterization of graphs.

In the special case of a finite simple graph, the adjacency matrix is composed only by ones and zeroes, that is, it will be a (0, 1)-matrix. Its zeroes are present in all the main diagonal.

If the graph is undirected (UG), then its adjacency matrix will be symmetric.

In the case of a complete graph, its adjacency matrix is composed by all ones, except in the diagonal, of zeroes.

The correspondence between a graph and the corresponding eigenvectors and eigenvalues obtained from its adjacency matrix, is analyzed in a relatively new field, named *Spectral Graph Theory.*

The adjacency matrix of a complete bipartite graph, $K_{r,s}$, has the form

$$\begin{pmatrix} 0 & J \\ J^{\ t} & 0 \end{pmatrix}$$

[1] AMO - Advanced Modeling and Optimization. ISSN: 1841-4311

where J will be a *(rxs)*-matrix, and J^{t} its transposed matrix.

Therefore, as we mentioned above, it is a very important fact, from the mathematical viewpoint. These (adjacency matrices) can serve as isomorphism invariants by graphs. So, it permits classify coherently the different types of graphs, and into each class, its different elements.

Let A be the adjacency matrix of a *DG,* or a *UG.* Then, the matrix A^{n} is produced from n copies of A*;* more exactly, it is obtained multiplying the matrix A by itself, i.e. iteratively, n times.

It admits a subsequent combinatorial interpretation, according which the entry in row i and column j gives us the number of (directed or undirected) walks of length n, from the node i until the node j.

We consider now the usual matrix $I - A$, or its opposite, $A - I$, being I the identity matrix. The possibility of invertible character of the matrix $I - A$ is related with the non existence of directed cycles in the graph G.

So, it is the actual situation when we are working with DAGs. In this case, the interpretation would be that the entry in row i and column j give us the number of directed paths from i to j. Such cardinality is always finite, if there are no directed cycles.

This can be explained by geometric series applied to matrices,

$$(I - A)^{-1} = I + A + A^{2} + A^{3} + ... = \sum_{i=0}^{+\infty} A^{i}$$

In our case, it can be interpreted in this way: the cardinal of DAGs from i to j equals the number of DAGs of length zero, plus the number of DAGs of length one, plus the the number of DAGs of length two, plus the number of DAGs of length three, and so on.

The main diagonal of every adjacency matric corresponding to a graph without loops has all zero entries.

For n–regular graphs, n is also an eigenvalue of A, for the vector $v = (1, 1, 1, ...)$.

G is *connected* if and only if the multiplicity order of the eigenvalue n is equal to one. If G is a *connected bipartite graph,* then also $-n$ would be an eigenvalue of A. It is a consequence of the *Perron-Fröbenius Theorem.*

2. Distance Matrix

A *distance matrix* is like a high-level adjacency matrix. But it provide more information. Not only about whether or not two nodes are connected, but also tell us about the distances between them. We assume, for this reason, an unitary distance for each edge.

So, this matrix contains the mutual distances, taken pairwise, of a collection of points-nodes.

Hence, generating a *(nxn)*-matrix. Its elements are non-negative real numbers, given n nodes, or equivalently, given n points in the Euclidean space.

The cardinality of such set of pair of points will be

$\frac{n(n-1)}{2}$

It will be the number of independent elements in the distance matrix.

Note some comparative differences between adjacency and distance matrices.

Firstly, showing only information about connected character, either about metric measure.

Secondly, an entry of a distance matrix will be smaller, if two elements are closer. Nevertheless, close connected edges may yield larger entries in an adjacency matrix.

Distance matrices have many applications. For instance, in Bioinformatics, where they are used to represent protein structures, in a coordinate-independent manner.

They are also used in sequential and structural alignment, and for the determination of protein structures for Nuclear Magnetic Resonance *(NMR)* or X-ray chrystallography.

3. Similarity Matrix

But sometimes it is more adequate to express data as a *Similarity Matrix.* It will be a matrix of scores which shown the similarity between two data points.

Such matrices are strongly interrelated with both, substitution matrices and the aforementioned distance matrices.

Among its applications, we have

- *Case Based Reasoning*

- *Intelligent Information Retrieval*

- *Content Based Image Retrieval*

- *Sequence Alignment,* where higher scores are given to more similar characters.

Also by lower, or besides negative scores, for dissimilar characters.

And other interesting case is that of so called *Seidel Adjacency Matrices.* They are symmetric matrices, with a row and column for each node. It possess all zeroes on the main diagonal (i.e. when $i = j$), and in the positions corresponding to distinct nodes, i and j, the values

$$\begin{cases} -1, \text{ if the nodes are adjacent} \\ +1, \text{ if they are not} \end{cases}$$

Such matrices are introduced by (Lint and Seidel, 1966).

They are the adjacency matrices of the *Signed Complete Graph*, where the edges of G are negative, being positives the edges which are not in G.

Its eigenvalues properties are very useful in the study of *Strongly Regular Graphs (SRG).*

Recall that the graph G is said to be *Strongly Regular,* if there are two integers, α and β, such that

- every two adjacent nodes have α common neighbours,

and

- every two non-adjacent nodes have β common neighbours.

For this reason, a *strongly d-regular n-graph* will be denoted as

$$srg\,(n, d, \alpha, \beta)$$

being obviously n the number of its nodes, and d the number of edges that incides in each node.

4. Incidence Matrix

The *Incidence Matrix of a digraph* is a *(pxq)*-matrix, which will be denoted by

$$\left(b_{ij}\right)_{\substack{i\in\{1,2,\ldots,p\}\\ j\in\{1,2,\ldots,q\}}}$$

where p denotes the number of its nodes, whereas q denotes the number of its edges.

Its entries will be such that

$$b_{ij} = \begin{cases} -1,\ if\ the\ edge\ e_j\ leaves\ node\ n_i\ (outgoing) \\ +1,\ if\ the\ edge\ e_j\ it\ enters\ node\ n_i\ (ingoing) \\ 0,\ otherwise \end{cases}$$

So, an incidence matrix is a matrix that shows the relationship between two classes of mathematical objects.

If the first class is C and the second class is C', then the matrix has one row for each element of C, and one column for each element of C'.

By this way, the corresponding entry in row i and column j will be equal to one, if i and j are related, in the sense of being incident, and zero otherwise (i.e. if they are not incident).

In Graph theory we can find two kinds of *Incidence Matrices,*

- *unoriented*

and

- *oriented*

Generalizing these concepts, we may to define the *Incidence Matrix* of an Incidence Structure as a *(pxq)*-matrix, where p and q are the number of points and the number of lines, respectively, in such a way that

$$b_{ij} = \begin{cases} +1,\ if\ the\ point\ p_i\ and\ the\ line\ L_j\ are\ incident \\ 0,\ otherwise \end{cases}$$

In this case, the Incidence Matrix will be also a biadjacency matrix of the Levi graph of the structure.

And because there is a Hypergraph for every Levi graph, and a Levi graph for every Hypergraph (or vice versa), it is possible to conclude that

the incidence matrix of an incidence structure describes a hypergraph

5. Degree Matrix

We may introduce the matrix of *valencies,* also called *degrees* (abridged, deg), of the nodes in the graph.

Such *degree matrix* will be denoted by $D(G) \equiv \left(d_{ij}\right)$, and it is defined as

$$d_{ij} \equiv \begin{cases} \deg\left(n_i\right) \equiv \delta_i,\ if\ i = j \\ 0,\ otherwise \end{cases}$$

Do, the Laplacian matrix may be defined by

$$L = D.D^t$$

Hence, it is possible to shown that

$$D.D^t = \Delta - A$$

where

$$\Delta = (\delta_1, \delta_2, ..., \delta_n)$$

As consequence, the Laplacian matrix does not depend of what orientation is chosen for the graph G.

And so, there you are some fundamental matrices, which may appears associated with a graph,

- the *Incidence Matrix* encapsulates node-edge relationships.

- the *Adjacency Matrix* encapsulates node-node relationships.

- the *Degree Matrix* encapsulates information about the degrees.

6. Laplacian Matrix

But also there is one last and many times interesting matrix, which will be called the *Laplacian Matrix.*

In Graph Theory, the *Laplacian Matrix* is a matrix representation which may be associated to every graph. And, as we will see more later, it can be used to calculate the number of spanning trees, for a given graph.

This will be one of the properties which can be observed from the Laplacian matrix and its eigenvalues, with support at the Kirchhoff ´s theorem. But it will be applied to find many other properties into the graph.

Let G be a graph.

We can define *the Laplacian Matrix of G,* denoted as

$$L(G) = \left(l_{ij}\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq n}}$$

by

$$L(G) \equiv D(G) - A(G)$$

where $D(G)$ is the degree matrix of G, and $A(G)$ is the adjacency matrix of G. That is, by an equivalent expression,

$$l_{ij} = \begin{cases} \deg(n_i), \ if \ i = j \\ -1, \ if \ i \neq j \ and \ n_i \ is \ adjacent \ to \ n_j \\ 0, \ otherwise, \ i.e.if \ being \ also \ i \neq j, \\ \ n_i \ is \ not \ adjacent \ to \ n_j \end{cases}$$

So, it will be the difference among the degree matrix, $D(G)$, and the adjacency matrix, $A(G)$.

If we consider this on digraphs, either the outdegree or the indegree might be used. It depends on the application.

The *normalized Laplacian matrix* may be expressed by

$$l_{ij}^{N} = \begin{cases} 1, \ if \ i = j \ and \ \deg(n_i) \neq 0 \\ \frac{-1}{\sqrt{\deg(n_i)\deg(n_j)}}, \ if \ i \neq j \ and \ n_i \ is \ adjacent \ to \ n_j \\ 0, \ otherwise, \ i.e.if \ it \ also \ being \ also \ i \neq j, \\ \ n_i \ is \ not \ adjacent \ to \ n_j \end{cases}$$

Among the properties of the eigenvalues, $\{\lambda_i\}_{i=0}^{n-1}$, of a Laplacian matrix, $L(G)$, we can select

1) $L(G)$ will be always positive-semidefinite; that is,

$$\lambda_i \geq 0, \ \forall i \in \{0, 1, 2, ..., n-1\}$$

2) The number of times that zero appears as an of its eigenvalues gives us the number of connected components in the graph.

3) $\lambda_0 = 0$.

4) λ_1 reflects its *algebraic connectivity.*

5) The smallest non-trivial eigenvalue is called its *spectral gap.*

Recall that a *positive-semidefinite (nxn)-matrix* means that for any non-zero vector, *X,* we have

$$X^t \; A \; X \geq 0$$

That is,

$$(x_1 x_2 ... x_n) \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} \geq 0$$

This final result, after multiplying their respective matrices of dimension $(1xn)$, (nxn) and $(nx1)$, will be a scalar, or $(1x1) -$ matrix. More concretely, in this case it will be a non-negative real number.

An important result is the named *Kirchhoff´s Theorem*, according which we can said that:

Let G be a labeled and connected n-graph, and suppose that $\{\lambda_i\}_{i=1}^{n-1}$ are the non-zero eigenvalues of its Laplacian matrix.

Then, the number of its spanning trees will be given by

$$t\,(G) = \frac{\prod_{i=1}^{n-1} \lambda_i}{n}$$

Note that λ_0 must be excluded of this formula, because it is always null.

It is possible to generalize the Laplacian matrix to the infinite case, when the cardinality of the set of edges and nodes is unbounded.

Such generalization is usually known as the *discrete Laplace operator.*

7. Graph and its Genus

An *automorphism of a graph,* $G = (V,\ E)$, will be an isomorphism from G onto itself.

So, a *graph-automorphism* of a simple graph is simply a bijection of its nodes,

$$f : G \to G$$

such that the image of any edge of G is always an edge in G.

That is, if

$$e = \{i,\ j\} \in E$$

then

$$f(e) = \{f(i),\ f(j)\} \in E$$

About another interesting concept in Mathematics, the word *"genus"* has different, but very related, meanings.

So, in Topology, it depends on to consider orientable or non-orientable surfaces.

In the case of *connected and orientable surfaces,* it is an integer that represents the *maximum number of cuttings,* along closed simple curves, without rendering the resultant manifold disconnected.

For this reason, we said many times that it is the number of *"handles"* on it.

Usually, it is denoted by the letter g.

It will be also definable through the Euler number, or *Euler Characteristic,* denoted by χ.

Such relationship will be expressed, for *closed surfaces,* by

$$\chi = 2 - 2g$$

When the surface has b boundary components, this equation transforms to

$$\chi = 2 - 2g - b$$

which obviously generalizes the above equation.

For example, a *sphere,* or a *disc,* both have genus $g = 0$.

Instead of this, a *torus* has $g = 1$.

In the case of *non-orientable surfaces,* the *genus* of a closed and connected surface is a positive integer, representing the number of cross-caps attached to a sphere.

Recall that a *cross-cap* is a two-dimensional surface that is topologically equivalent to a Möbius string.

As in the precedent analysis, it can be expressed in terms of the Euler characteristic, by

$$\chi = 2 - 2k$$

being k the *non-orientable genus.*

For example, a *projective plane* has non-orientable genus $k = 1$.

And a *Klein bottle* has a non-orientable genus $k = 2$.

Turning to graphs, its corresponding genus will be the minimal integer, n, such that the graph can be drawn without crossing itself on a sphere with n *handles.*

So, a *planar graph* has genus $n = 0$, because it can be drawn on a sphere without self-crossing.

In the *non-orientable case,* the genus will be also the minimal integer, n, such that the graph can be drawn without crossing itself on a sphere with n *cross-caps.*

If we pass now to topological graph theory, we will define as *genus of a group, G,* the minimum genus of any of the undirected and connected Cayley graphs for G.

From the viewpoint of the Computational Complexity, *the problem of "graph genus" is NP-complete* [Thomassen, 1989].

8. Graph Invariants and Graph Properties

We will says either *graph invariant* or *graph property,* when it depends only of the abstract structure, not on graph representations, such as particular labelings or drawings of the graph.

So, we may define a *graph property* as every property that is preserved under all its possible isomorphisms of the graph. Therefore, it will be *a property of the graph itself,* not depending on the representation of the graph.

The semantic difference also consists in its character

- *qualitative*

or

- *quantitative*

For instance, when we said

"this graph does not a perfect graph"

this will be a *property,* because it is a *qualitative* statement.

While when we says

"the number of nodes with degree two in such graph"

this would be an *invariant,* because it is a *quantitative* statement.

From a mathematically strict viewpoint, *a graph property* can be interpreted as *a class of graphs,* composed by the graphs that have in common the accomplishment of the same conditions.

Hence, also can be defined a graph property as a function whose domain would be the set of graphs, and its range will be the bivalued set composed by two options, true and false, $\{T,\ F\}$, according which a determinate condition is either verified or violated for the graph.

A graph property is called *hereditary*, if it is inherited by its induced subgraphs.

And it is *additive,* if it is closed under disjoint union.

For example, the property of a graph to be planar is both additive and hereditary.

Instead of this, the property of being connected is neither.

The computation of certain graph invariants may be very useful, with the purpose to discriminate whether two graphs are isomorphic, or instead of this, they will be non-isomorphic.

The support of these criteria will be that for any invariant at all, two graphs with different values cannot be isomorphic between them.

But however, two graphs with the same invariants either may, or may not, to be isomorphic between them.

So, we will arrive to the notion of *completeness.*

Let $I(G)$ *and* $I(H)$ be invariants of two graphs, G *and* H.

It will be considered *complete*, if the identity of the invariants ever implies the isomorphism of the corresponding graphs, i. e.

$$I(G) = I(H) \Rightarrow G \approx H$$

9. Enumerating DAGs

It will be also interesting to give some previous concepts, as that of *line graph, vertex-transitivity, edge-transitivity,* and so on.

Let G be a graph. Suppose that we denote by *V(G)* its set of nodes, and by *E(G)* its set of edges.

The so-called *line graph of* G will be the graph whose set of nodes is E (therefore, it coincides with the set of edges of G), and whose edges connect all pairs of E which have one common end (or extremity) in G. Usually, it is abridged by *L(G).*

Hence, the *Line Graph of* G is another graph that represents the adjacencies between edges of G.

The Line Graph is sometimes called

- *Adjoint Graph,*

- *Interchange Graph,*

- *Edge Graph*

- *Derived Graph of G,*

and so on.

A graph, *G,* is said to be *node-transitive* (or *vertex-transitive*), if for any two of its nodes, n_i and n_j, there is an automorphism which maps n_i to n_j.

A simple graph, *G,* is said to be *edge-transitive* (or *link-transitive*), if for any two of its edges, e and e', there is an automorphism which maps e into e'.

A simple graph, *G,* is said to be *symmetric,* when it is both, node-transitive and edge-transitive graph.

But a simple graph, *G,* which is edge-transitive, but not node-transitive, will be said *semi-symmetric.*

Obviously, such a graph will be necessarily a bipartite graph.

10. Enumerating Bayesian Networks

Bayesian Networks are the most sucessful class of models to represent uncertain knowledge. But the representation of conditional independencies (CIs, in acronym) does not have uniqueness. The reason is that probabilistically equivalent models may have different representations.

And this problem is overcome by the introduction of the concept of *Essential Graph,* as unique representant of each equivalence class. They represent *CI models* by *graphs.* Such mathematical and graphical tools containing both, directed or/and undirected edges; hence, producing respectively *Directed Graphs (DGs),* in particular acyclic elements, or *Directed Acyclic Graphs (DAGs),* either *Undirected Graphs (UGs),* or *Chain Graphs (CGs),* in the mixed case. So, DAG models are generally represented as Essential Graphs (EGs).

Knowing the ratio of EGs to DAGs is a valuable tool, because through this information we may decide in which space to search. If the ratio is low, we may prefer to search the space of DAG models, rather than the space of DAGs directly, as it was usual until now.

The most common approach to learning DAG models is that of performing a search into the space of either DAGS or DAG models (EGs).

It is preferable, from a mathematical point of view, to obtain the more exact solution possible, studying its asymptotic behaviour. But also it is feasible to propose a Monte Carlo Chain Method (MCMC) to approach the ratio, avoiding the straightforward enumeration of EGs. And a many more elegant construct, if very difficult, through the Ihara Zeta function for counting graphs.

Recall that a DAG, *G,* is *essential,* if every directed edge of G is protected. So, an *Essential Graph (EG)* is a graphical representation of a Markovian equivalence class.

In relation with the Essential Graph, each directed edge would have the same direction in all the graphs that belongs to its equivalence class.

There is a bijective correspondence (one-to-one) among the set of Markovian equivalence classes and the set of essential graphs, its representatives.

The labeled or unlabeled character of the graph means whether its nodes or edges are distinguishable or not. For this, we will say that it is *vertex-labeled, vertex-unlabeled, edge-labeled,* or *edge-unlabeled.*

The labeling will be considered as a mathematical function, referred to a value or name (label), assigned to its elements, either nodes, edges, or both, which makes them distinguishable.

Conclusion

In this way, it is reached a general vision of the situation, reflecting the degree of fitness to the adequated modelings, and we may to reflect the properties and invariants through its corresponding matrices.

So, we will establish a more powerful analytical framework, improving some more our theoretical basis.

References

E. A. Bender, L. B. Richmond, R. W. Robinson, and N. C. Wormald (1986). The asymptotic number of acyclic digraphs I. *Combinatorica* **6** (1): 15-22.

E. A. Bender, and R. W. Robinson (1988). The asymptotic number of acyclic digraphs II. *J. Comb.Theory,* Serie B **44** (3): 363-369.

F. Harary (1955). The number of linear, directed, rooted, and connected graphs. *Proc. Amer. Math. Soc.* (AMS) **78**: 445-463.

F. Harary (1957). The number of oriented graphs. *Michigan Math. J.* Volume 4, Issue 3: 221-224.

R. W. Robinson (1973). Counting labeled acyclic digraphs, in Harary, F. (ed.), *New Directions in the Theory of Graphs:* 239-279. Academic Press. New York.

R. W. Robinson (1977). Counting unlabeled acyclic digraphs. *Combinatorial Mathematics V* (C. H. C. Little Ed.). Springer. *Lecture Notes in Mathematics* **622:** 28-43.

B. Steinsky (2004). Asymptotic behaviour of the number of labeled essential acyclic digraphs and labeled chain graphs. *Graphs and Combinatorics* **20** (3): 399-411.

B. Steinsky (2003). Enumeration of labeled chain graphs and labeled essential directed acyclic graphs. *Discrete Mathematics* **270** (1-3): 266-277.