Enumerating Graphs

Angel Garrido, Facultad de Ciencias de la UNED

Abstract

We analyze here the most useful tools working with Enumeration, Combinatorics, Cycle Index, Generating Functions, Formal Series, Zeta functions and so on.

Our objective is to reach a consistent mathematical framework, with provide sufficient power to handle adequatly such type of problems.

Keywords: Graph Theory, Combinatorics, Matrix Theory, Group Representation Theory, Mathematical Analysis.

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

1. Enumerative Combinatorics

It is possible to obtain a good, but necessarily classical vision on this subject in the surveys-books of authors as (Harary, 1969, 1973), or (Stanley, 1986). Nevertheless, we introduce some mathematical tools of *Enumerative Combinatorics,* more oriented to study the cardinality of graphs.

It is an area of Mathematics on the number of ways that certain patterns can be formed. In general, given an infinite collection of finite sets, it seeks to describe a *counting function, f,* for the number of objects in every set. The simplest such functions are *closed formulas,* which can be expressed as a composition of elementary functions, such as powers or factorials. Often no closed form is available. Then, we first derive a recurrence relation, then solve the recurrence, and by this, we arrive at the desired closed form.

Given a set, *S,* the most inclusive definition of an *Enumeration* is any surjective application which goes from an arbitrary index set, *I,* onto *S.*

So, every set can be trivially enumerated by the identity function, from *S* onto itself.

Such general definition lends to a *Counting* notion. We are interested, obviously, in "how many things", instead of "in which order". For this reason, it is not necessary initially, but convenient, a well-ordering into the set *S.*

We note some interesting ideas.

- First, that *all finite sets are enumerable.*

- Second, the real numbers, or *real line*, are *not countable enumeration.* It may be proved by Cantor diagonalization argument.

- Third, *there exists an enumeration for a set* if and only if *the set is countable.*

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

There are many mathematical subfields in quick expansion, concerning with enumeration. So, for instance, in *Graph Enumeration,* where the objective is to count graphs that meet certain structural conditions. This connect directly with an important area, called *Extremal Graph Theory,* which studies the graphs which are "extremal" in some sense. That is, among graphs with a certain property. So, with the largest number of edges, the smallest diameter, the largest minimum degree, and so on.

It will be convenient now to allude to the *Formal Power Series,* from which we define *Generating Functions.*

Also we need some very powerful results of Combinatorics, as may be the *Inclusion-Exclusion Principle.*

A *Formal Series,* or *Formal Power Series,* of a field F, is an infinite sequence, $[a_0, a_1, a_2, \ldots]$, defined on F. Usually, it will be denoted with brackets. Equivalently, it is a function from the set of non-negative integers, $\mathbb{N}^* = \mathbb{N} \cup \{0\}$, onto the field F. We write these series as

$$\sum_{i \in \mathbf{N}^*} a_i \, x^i$$

These mathematical tools are very useful to obtain compact representations or to define recursively sequences.

A *generating function, G,* is a formal power series expressible as

$$G(x) = \sum_{i \in \mathbf{N}} a_i \, x^i$$

Roughly, it transforms problems about sequences into problems about functions. So, we may use generating functions to solve many types of “counting” problems.

The more usual of such functions are included into two types, *OGF (Ordinary Generating Function),* as

$$G(x) = \sum_{i \in \mathbf{N}} f(i) \; x^i$$

and *EGF (Exponential Generating Function),* as

$$G(x) = \sum_{i \in \mathbf{N}} f(i) \; \frac{x^i}{i!}$$

As an example, of generating functions giving the first powers of the non-negative integers, we have

$$\frac{x}{1-x} \leftrightarrow x + x^2 + x^3 + ... \leftrightarrow [1, 1, 1, ...]$$
$$\frac{x}{(1-x)^2} \leftrightarrow x + 2x^2 + 3x^3 + ... \leftrightarrow [1, 2, 3, ...]$$
$$\frac{x(x+1)}{(1-x)^3} \leftrightarrow x + 4x^2 + 9x^3 + ... \leftrightarrow [1, 4, 9, ...]$$

and so on.

And it also appears in Number Theory, as in the case of

$$f(x) = \frac{x}{1-x-x^2} = \sum_{i \in N^*} F_i \; x^i = x + x^2 + 2x^3 + 3x^4 ... \leftrightarrow [1, 1, 2, 3, ...]$$

Where F_i denotes de *i-th Fibonacci Number.*

Recall that it obeys to the recursive formula

$$F_i = F_{i-1} + F_{i+1}, \; \forall i \in \mathbf{N}^*$$

We can introduce operations among generating functions by procedures as

Scaling Rule

If

$$[f_0, \; f_1, \; f_2, ...] \leftrightarrow F(x),$$

then

$$[cf_0, \; cf_1, \; cf_2, ...] \leftrightarrow c \; F(x).$$

Addition Rule

If

$$[f_0, \; f_1, \; f_2, ...] \leftrightarrow F(x),$$

$$\textit{and}$$

$$[g_0, \; g_1, \; g_2, ...] \leftrightarrow G(x),$$

then

$$[f_0 + g_0, \; f_1 + g_1, \; f_2 + g_2, ...] \leftrightarrow F(x) + G(x).$$

Right-shift Rule

If

$$[f_0, \; f_1, \; f_2, ...] \leftrightarrow F(x),$$

then

$$\left[\overbrace{0, 0, ..., 0,} f_0, \; f_1, \; f_2, ...\right] \leftrightarrow x^k \; F(x)$$

Derivative Rule

If

$$[f_0,\ f_1,\ f_2, ...] \leftrightarrow F(x),$$

then

$$[f_1,\ 2f_2, 3\ f_3, ...] \leftrightarrow F\,'(x)$$

Therefore, differentiating a generating function has two effects, on the given sequence. Firstly, each term appear multiplied by its index. And secondly, the entire sequence is shifted left one place.

So, for instance, if we take

$$\tfrac{x}{1-x} \leftrightarrow x + x^2 + x^3 + ... \leftrightarrow [1, 1, 1, ...]$$

it holds

$$\tfrac{d}{dx}\left(\tfrac{x}{1-x}\right) \leftrightarrow \frac{1}{\left(1-x^2\right)} \leftrightarrow [1, 2, 3, 4, ...]$$

2. Inclusion-Exclusion Principle and Sieve Formula

Let $X = \{1, 2, ..., m\}$ and $Y = \{1, 2, ..., n\}$ be two sets, where $m, n \in \mathbb{N}$. Then,

- The *number of applications* between X and Y is given by n^m.
- The *number of injective applications* between such sets would be

$$\tfrac{n!}{(n-m)!} = n(n-1)(n-2)...(n-m+1)$$

- The *number of surjective applications,* S_{nm}, among such sets, will be

$$S_{nm} = m^n - C_{m,1}(m-1)^n + C_{m,2}(m-2)^n + ... + (-1)^{m-1} C_{m,m-1} =$$
$$= m^n - \sum_{k=1}^{m-1} (-1)^{k-1} C_{m,k}(m-k)^n = \sum_{k=0}^{m-1} (-1)^{k-1} C_{m,k}(m-k)^n$$

The *Inclusion-Exclusion Principle* (IEP) is a known result in Combinatorics, being essential in some proofs. Let $c(A)$ be the cardinal of A. We know that given two sets, A and B, the cardinal of their union is equal to the sum of their corresponding cardinals, minus the cardinal of the intersection,

$$c(A \cup\ B) = c(A) + c(B) - c(A \cap\ B)$$

And so,

$$c(\cup\ A_i) \leq \sum c\ (A_i)$$

The inequality is based in that considering the sum of cardinals, when the sets are not mutually disjoint, we surpass the value, and for this it will be corrected.

In general, given a family of sets, $\{A_i\}_{i=1}^{n}$, it holds

$$c(\cup_{i=1}^{n} A_i) = \sum c(A_i) - \sum_{1 \le j < k \le n} c\,(A_j \cap A_k) +$$
$$+ \sum_{1 \le j < k < l \le n} c\,(A_j \cap A_k \cap A_l) - \ldots + (-1)^n\; c\,(\cap_{i=1}^{n}\; A_i)$$

where the summations are extended on *"k-subsets"*, as subsets of k elements into a superset with cardinal n.

Recall that the number of *k-subsets* on n elements is given by the binomial coefficient

$$C_{n,k} = \frac{n!}{k!\;(n-k)!}$$

And the total number of different k-subsets taken from a set of n elements,

$$\sum C_{n,k} = \sum \frac{n!}{k!\;(n-k)!} = 2^n$$

The *IEP* was used by Nicholas Bernoulli, to solve the problem of computing the number of *"derangements"*, i.e. of permutations where the elements do not remain never fixed.

We formalize now this *IEP*.

For every finite sequence, $\{A_i\}_{i=1}^{n}$, of at least two subsets ($n \geq 2$), all of them included into a finite set, S, it holds

$$c\left(\cup_{k=1}^{n} A_k\right) = \sum_{\emptyset \neq I \subseteq \{1,2,..,n\}} (-1)^{[c(I)-1]}\; c\left(\cap_{i \in I} A_i\right)$$

Generalizing these ideas, given any measure, $m : S \to \mathbf{R}_+$, for every sequence, $\{A_i\}_{i=1}^{n}$, of at least two subsets from a finite set S, the measure of the set of elements that do not belong to any A_i is given by

$$m\left(\cap_{k=1}^{n} A_k^c\right) = m\,(S) + \sum_{\emptyset \neq I \subseteq \{1,2,..,n\}} (-1)^{[c(I)]}\; m\left(\cap_{i \in I} A_i\right)$$

Note that if we use the convention of emptiness for the indices, i.e. $I = \emptyset$, the previous term $m(S)$ may be included in the summatory. It suffices eliminating such condition of I not empty.

By this, we obtain a version of the *Sylvester Formula,*

$$m\left(\cap_{k=1}^{n} A_k^c\right) = \sum_{I\; \subseteq \{1,2,..,n\}} (-1)^k\; \sum_{I \subseteq \{1,2,..,n\}} m\left(\cap_{i \in I} A_i\right)$$

It is often convenient to aggregate terms relative to subsets with the same cardinality,

$$m\left(\cap_{k=1}^{n} A_k^c\right) = \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,2,..,n\} \\ c(I)=k}} m\left(\cap_{i \in I} A_i\right)$$

Still, it is possible to generalize the equation, to the usually known as *"Sieve Formula".*

Given any measure, $m : S \to \mathbf{R}_+$, for every finite sequence, $\{A_i\}_{i=1,2,...,n}$, with at least two subsets of S, the measure of the set of elements of S that exactly belong to m of such A_i, with $0 \leq m \leq n$, will be

$$\sum_{k=m}^{n} (-1)^{k-m} \; C_{k,m} \sum_{\substack{I \subseteq \{1,2,..,n\} \\ c(I)=k}} m\left(\cap_{i \in I} A_i\right)$$

Note that the Sylvester Formula is a special case of the Sieve Formula, when $m = 0$.

3. Zeta functions

Among the Dirichlet Series, we found a very useful tool, in fields as Number Theory, Probability or Cryptography. It is the so called *Riemann Zeta Function,*

$$\varsigma(s) = \sum_{n \in \mathbf{N}} \frac{1}{n^s} = \prod_{p \ prime} \left(1 - p^{-s}\right)^{-1}, \ with \ s \in \mathbb{C}$$

Where it appears as product of Euler. And because it has a bounded sequence of coefficients, these series converge absolutely to an analytical function, on the complex open half-plane of s such that

$$Re(s) > 1$$

It diverges on the symmetrical open half-plane of s, in the complex plane, $Re(s) < 1$. The defined function on the first region admits analytic continuation to all $\mathbb{C}$, except when $s = 1$. For $s = 1$, this series is formally identical to the Harmonic series, which diverges.

As a consequence, it is a meromorphic function of s, being in particular, holomorphic in a region of the complex plane, showing one pole in $s = 1$, with residue equal to 1. Recall that a function is *holomorphic* when it is complex differentiable, and will be *meromorphic* when it is holomorphic on almost all $\mathbb{C}$, except in a set of isolated points, which are called the *poles* of the function.

Euler found a closed formula for $\zeta(2k)$, when $k \in \mathbf{N}$.

It will be expressed by

$$\zeta(2k) = \frac{(-1)^{k-1}(2\pi)^{2k} B_{2k}}{2\ (2k)!}$$

denoting by B_{2k} the *Bernoulli numbers.*

Such numbers can be defined of different modes.

So,

- as independent terms of *Bernoulli polynomials,* $B_n(x)$,
- by a generating function,

$$G(x) = \frac{x}{e^x - 1} = \sum_{i \in N^*} B_n \frac{x^i}{i!}$$

with

$$|x| < 2\pi$$

where each coefficient of the Taylor Series is the *n-th Bernoulli number.*

- by the recursive formula

$$B_0 = 1$$
$$B_m = -\sum_{j=0}^{m-1} C_{m,j} \frac{B_j}{m-j+1}$$

As the Bernoulli numbers can be expressed in terms of the Riemann Zeta function, they are indeed values of such function to negative arguments.

Some of the *Zeta Function special values* are

$$\varsigma(0) = -1/2$$
$$\varsigma(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + ... = \frac{\pi^2}{6}$$
$$\varsigma(4) = 1 + \frac{1}{2^4} + \frac{1}{3^4} + ... = \frac{\pi^4}{90}$$
$$\varsigma(6) = 1 + \frac{1}{2^6} + \frac{1}{3^6} + ... = \frac{\pi^6}{945}$$
$$\varsigma(8) = 1 + \frac{1}{2^8} + \frac{1}{3^8} + ... = \frac{\pi^8}{9450}$$
$$...$$

Note that we take here *s even.*

Because for *odd* values of *s,* it appears troubles and also irrational numbers; for instance,

$$\varsigma(1) = 1 + \frac{1}{2} + \frac{1}{3} + ... \to \infty \ (\textit{harmonic} \text{ series});$$
$$\varsigma(3) = 1 + \frac{1}{2^3} + \frac{1}{3^3} + ... \simeq 1.2 \ (\textit{Apéry constant})$$
and also
$$\varsigma(1/2) \simeq -1.46; \ \varsigma(3/2) \simeq 2.6;$$
$$\varsigma(5/2) \simeq 0.134; \ \varsigma(7/2) \simeq 1.127, ...$$

The *Logarithm of the Zeta Function* will be

$$\log\varsigma(s) = \sum_{n\geq 2} \left(\frac{\Lambda(n)}{\log\ n}\right)\left(\frac{1}{n^s}\right)$$
being
$$\mathrm{Re}(s) > 1$$

Here, $\Lambda(n)$ denote the *Lambda Function,* also called sometimes *Von Mangoldt function,* defined by

$$\Lambda(n) = \log\ p,\ \ if\ \ n = p^k,\ for\ n \in N$$
$$and\ some\ prime\ number,\ p;$$
$$and\ 0,\ otherwise$$

It is an arithmetic function that is neither additive nor multiplicative,

$$\Lambda(n + n') \neq \Lambda(n) + \Lambda(n')$$
$$\Lambda(n \cdot n') \neq \Lambda(n) \cdot \Lambda(n')$$

Such Lambda function satisfies

$$\log\ n = \sum_{d|n} \Lambda(d)$$

where the summation will be extended to all integers, *d,* dividing to *n.*

Related with the above series, we have the popular *Riemann Hypothesis,* still an important open problem in current Mathematics. It is about the distribution of zeroes of such Zeta Function.

It admits many variations, with different names, Selberg, Ihara, etc.

So, for instance, we may consider its multiplicative inverse, expressible as a series by the *Möbius Function.* It can be reached, from the known series, by tools as the *Möbius Inversion* and the *Dirichlet Convolution.*

The values produced by such function from integer arguments are called *"zeta constants".*

We can observe their convergence to one from the right,

$$\varsigma(s) \to 1^+$$

Also, this *functional equation* is satisfied

$$\varsigma(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right)\ \Gamma(1-s)\ \varsigma(1-s)$$

which is true in all the complex field, relating its values in s and $1 - s$. This equation has a pole simple at s = 1, with residuum equal to one. It was proved by Riemannn (1859).

Euler conjecured an equivalent relation to the function

$$\sum_{n \in N^*} \frac{(-1)^{n+1}}{n^s}$$

Also there exists a symmetric version of the precedent functional equation, reachable by the change

$$s \longmapsto 1 - s$$

This gives

$$\varsigma(s)\ \Gamma\left(\frac{s}{2}\right)\ \pi^{-\frac{s}{2}} = 2^s\ \pi^{-\frac{1-s}{2}}\ \Gamma\left(\frac{1-s}{2}\right)\ \varsigma(1-s)$$

Sometimes, we define the *Eta Function,* denoted by ξ, as

$$\xi(s) \equiv \pi^{-\frac{s}{2}}\ \Gamma\left(\frac{s}{2}\right)\ \zeta(s)$$

It holds

$$\xi(s) = \xi(1-s)$$

The value of the Zeta function for negative even real values is zero.

$$\varsigma(-2) = \varsigma(-4) = \varsigma(-6) = \ ...\ = \varsigma(-2k) = 0$$
$$with\ k \in \mathbf{N}$$

They are called *trivial zeroes of* ζ.

Furthermore, it will be cancelled on values of s that belongs to the critic rang

$$\{s \in C : 0 < \mathrm{Re}(s) < 1\}$$

In this case, we call of *non-trivial zeroes.* It is because the difficulties to find its position into the critical rang.

To obtain zeta function values for negative and non integer arguments, we proceed by

$$\varsigma(-1/2) = \frac{2^{-3/2}\pi^{-1/2}\Gamma(-1/2)}{\sin\left(\frac{\pi s}{2}\right)\varsigma(1-s)} \simeq \frac{-4\pi}{2.6} \simeq -0.2069$$

Γ represents the *Gamma Function of Euler,* defined by an integral expression,

$$\Gamma(s) = \int_0^{\infty} e^{-t} t^{s-1} dt$$

If $n \in N$, then

$$\Gamma(n+1) = n!$$

For this reason, it is considered as an extension of the factorial. It holds

$$\Gamma(s+1) = s\ \Gamma(s)$$

and as

$$\Gamma(1) = 1$$

we have

$$\Gamma(n+1) = n\ \Gamma(n) = ... = n!\ \Gamma(1) = n!$$

Such functional equation also gives an asymptotic limit, proposed by (Nemes, 2007),

$$\varsigma\left(1-s\right)=\left(\frac{s}{2\pi l}\right)^{s}\sqrt{\frac{8\pi}{s}}\cos\left(\frac{\pi s}{2}\right)\left(1+O\left(\frac{1}{s}\right)\right)$$

Among its applications, they are useful to search compact formulas for a sequence given by a recurrence equation; to find relations among sequences, because the form of a generating function may suggest us a recurrence formula; to explore the asymptotic behaviour of sequences, as in our case; to prove identities involving sequences; or to solve enumeration problems in Combinatorics.

The famous *Riemann Hypothesis* says that for every non-trivial zero, *s,* of ζ, it holds

$$\text{Re}\ \ (s)=\frac{1}{2}$$

That is, all non-trivial zeroes are situated on the *critical line,* $x=\frac{1}{2}$.

The *Zeta Function* is *generalizable to graphs,* according to the theory elaborated by (Ihara, 1966). This function was first defined in terms of discrete subgroups.

J. P. Serre suggested can be reinterpreted graph-theoretically. And it was (Sunada, 1985) who put this into practice. In this version, it is denoted by ς_{G}, and defined by

$$\varsigma_{G}\left(s\right)\equiv\left[\prod_{p}\left(1-s^{L(p)}\right)\right]^{-1}$$

This product is taken over all prime walks, *p,* on the graph *G,* being *L(p)* the length of the *prime p.*

Recall that a *closed geodesic* is a closed path such there is no backtracking, if we around twice, i. e. it is a closed proper walk with the initial and final edges different. If γ is a closed geodesic, we denote by γ^{r} the obtained by repeating γ r times.

A closed geodesic which is not the power of another is called a *prime geodesic.*

An equivalence class of prime geodesics is called a *prime geodesic class,* or *prime,* $\wp$.

Given a path, γ, we denote by $L\left(\gamma\right)$ its *length.*

Therefore, two prime geodesics are said to be *equivalent,* if one is obtained from another by a cyclic permutation of edges. The length of a prime, $\wp$, is the length of any of its representatives.

Also ς_{G} is always representable as the reciprocal of a polynomial

$$\varsigma_{G}\left(s\right)\equiv\frac{1}{\det\ \left(I\ -\ T\ s\right)}$$

where T is the *edge adjacency operator* (Hashimoto, 1990).

Recall that the *adjacency operator, A,* is acting on the space of functions defined on the set of nodes of $G = (V,\ E)$.

Being *o(e)* and *t(e)* the origin and terminus of *e,* respectively, it is defined by

$$(Af)(x) = \sum_{e \in E_x} f\ \ [t(e)],$$
$$where$$
$$E_x \equiv \{e \in E : o(e) = x\}$$

(Bass, 1992) also gave a determinant formula involving the adjacency operator.

For any *Complex Network (CN),* or *Graph, G,* the function ς_G can be expressed in terms of ς, for different dimension values, *n.*

So,

If $n = 1,$ *then* $\varsigma_G(s) = 2\varsigma(s)$.

If $n = 2,$ *then* $\varsigma_G(s) = 4\varsigma(s-1)$.

If $n = 3,$ *then* $\varsigma_G(s) = 4\varsigma(s-2) + 2\varsigma(s)$.

If $n = \infty,$ *then* $\varsigma_G(s) = \frac{8}{3}\varsigma(s-3) + \frac{16}{3}\varsigma(s-1)$.

Recall that $\varsigma_G(s)$ is a *decreasing function* of *s.*

That is,

$$\varsigma_G(s_1) > \varsigma_G(s_2),\ if\ s_1 < s_2$$

And in the limit, if $n \to \infty,$ when s is next to the *transition point,* it holds

$$\varsigma_G(s) = \frac{2^n\ \varsigma(s - n + 1)}{\Gamma(n)}$$

If the average degree of nodes, also called *mean coordination number of the graph,* is finite, then there exists exactly a value of *s,* denoted $s_{transition}$, where the Zeta Function changes from infinite to finite, or vice versa.

It is also called *dimension of the Graph,or the Complex Network (CN).*

Also, the ς_G function possesses three properties,

- *monotonicity,*

- *stability,*

and

- *Lipschitz Invariance.*

According to *monotonicity,* a subset has dimension less than or equal to a superset.

According to *stability,* the dimension of the union of a family of sets is equal to the maximum cardinal among its members,

$$dim\ (\bigcup_{s=1}^{n} A_s) = \max_{s=1,2,\ldots,n} (\dim\ A_s)$$

And according to *Lipschitz Invariance,* the operations must intervene in the change of distances between nodes only by finite magnitudes, when the size, n, of the graph tends to infinity.

4. Adjacency Matrices

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 needs.

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, a (0, 1)-matrix. Its zeroes are present in all the main diagonal.

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

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

The relation between a graph and the eigenvectors and eigenvalues of 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

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

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 A iteratively n times. It admits the subsequent interpretation: the entry in row i and column j gives 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 caharacter of the matrix $I - A$ is related with the non existence of directed cycles in the graph G. So, it is the case when we are working with DAGs. In this case, the interpretation would be: the entry in row i and column j give 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,

$$(I - A)^{-1} = I + A + A^2 + A^3 + ...$$

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.*

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 alsotells about the distances between them. We assume for this 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 nonnegative real numbers, given n nodes, or equivalently, n points in the Euclidean space.

The cardinality of such set of pair of points will be

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

It is the number of independent elements in the distance matrix.

We may observe some differences between adjacency and distance matrices.

Firstly, showing only information about connected characted, 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.

But sometimes it is more adequate to express data as a *Similarity Matrix.* It is a matrix of scores which shown the similarity between two data points. Such matrices are strongly inter-related with 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. And lower, or besides negative scores, for dissimilar characters.

And other interesting case is that of *Seidel Adjacency Matrices.* They are symmetric matrices, with a row and column for each node.

It possess all zeroes on the main diagonal, and in the positions corresponding to nodes i and j, the values

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

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, and 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.

5. Cycle Index

And now we arrive to some beautiful tools that belongs to Group Theory.

In Combinatorics, Cycle Indices are useful in combinatorial enumeration, when symmetries are taken into account. Because knowing the cycle index of a permutation group, it is possible to enumerate equivalence classes that appears when the group acts on a collection of elements described by a generating function.

In such case, we use the *PET (Pólya Enumeration Theorem).*

The *cycle index,* $Z(P)$, of a permutation group, $P,$ is the average of

$$\prod_{k=1}^{n} a_k^{j_k}$$

over all permutations in the group.

Let P be a permutation group of degree n and order m.

Each element of P has a unique decomposition into disjoint cycles; for instance, $\{c_i\}_{i \in I}$.

Let the *length of a cycle* c_i be denoted by $|c_i|$.

Now, let $j_k(g)$ be the *number of cycles of length k in the permutation g*, being

$$0 \leq j_k(g) \leq \left\lfloor \tfrac{n}{k} \right\rfloor$$
$$\textit{and}$$
$$\sum_{k=1}^{n} k \; j_k(g) = n$$

Then, we may associate to g a monomial in the variables $\{a_i\}_{\{1,2,\ldots,n\}}$

$$\prod_{c \in g} a_{|c|} = \prod_{k=1}^{n} a_k^{j_k(g)}$$

From here, the *Cycle Index of P*, denoted $Z(P)$, is defined by

$$Z(P) \equiv \frac{1}{|P|} \sum_{g \in P} \prod_{k=1}^{n} a_k^{j_k(g)}$$

A typical example may be the *Cyclic Group* C_3, which contains the permutations

$$[1,2,3] \equiv (1)(2)(3)$$
$$[2,3,1] \equiv (123)$$
$$[312] \equiv (132)$$

and so, its cycle index will be

$$Z(C_3) \equiv \tfrac{1}{3}\left(a_1^3 + 2a_3\right)$$

The interpretation of the first permutation would be that it consists of three cycles of length one, also named 1-cycles, or equivalently fixed points.

For this reason, it is represented by a_1^3.

We apply now all this framework to analyze the *edge permutation group of 3-graphs,* i.e. graphs whith order three, or composed by three nodes.

Every permutation in the symmetric group of order three, denoted S_3, of permutations of nodes induces an edge permutation.

To compute the cycle index, we observe its permutations.

They are

- the *identity.* Neither nodes nor edges are permuted. Its contribution is a_1^3.

- *three reflections,* in an axis which pass through a node and the midpoint of the opposite edge. These fix one edge and exchange the remaining two. Therefore, its contribution will be $3a_1a_2$.

- *two rotations,* each one in reverse orientation, clockwise and counter-clockwise, respectively. So, it creates a cycle of three edges, being its contibution $2a_3$.

Therefore, the *Cycle Index* of this group, E_3, of edge permutations, will be

$$Z(E) = \frac{1}{6}\left(a_1^3 + 3a_1a_2 + 2a_3\right)$$

But the case of the edge permutation group for n-graphs is very different when $n > 3$, because in such case

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

Then the permutations are induced by

- the *identity.* It maps all nodes to themselves. Therefore, itlaso maps all edges to themselves. As consequence, its contribution will be a_1^6.

- *six permutations* that exchange two nodes, preserving the edge that connects such two nodes, and also the edge that connects the two nodes not excahnged. Its contribution to cycle index is $6a_1^2a_2^2$.

- *eigth permutations* that fix one node, and produce a 3-cycle for the remaining three nodes not fixed. Its contribution will be $8a_3^2$.

- *three permutations* that exchange, simultaneously, two node pairs, preserving the two pairs that connect the two pairs. The remaining edges constitutes two 2-cycles. So, its contribute by $3a_1^2a_2^2$.

- *six permutations* that produce rotations of nodes along a 4-cycle. So, it creates a 4-cycle of edges, and exchange the remaining two edges. Its contribute by $6a_2a_4$.

Hence, the Cycle Index of such *edge permutation group of 4-graphs,* which may be denoted by E_4, will be

$$Z(E_4) = \frac{1}{24}\left(a_1^6 + 6a_1^2a_2^2 + 8a_3^2 + 3a_1^2a_2^2 + 6a_2a_4\right) =$$
$$= \frac{1}{24}\left(a_1^6 + 9a_1^2a_2^2 + 8a_3^2 + 6a_2a_4\right)$$

6. 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 graphical tools containing both, directed or/and undirected edges; hence, producing respectively Directed Graphs (DGs), 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* (in acronym, *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 Markov equivalence class. In the EG, each directed edge would have the same direction, in all the graphs that form its equivalence class. There is a bijective correspondence (one-to-one) among the set of Markov equivalence classes and the set of essential graphs.

The labeled or unlabeled character of the graph means whether its nodes or edges are distinguishable or not.

For this, we say it is

- *vertex-labeled,*
- *vertex-unlabeled,*
- *edge-labeled,*

or

- *edge-unlabeled.*

The labeling is a mathematical function, referred to a value or name assigned to its elements, either nodes, edges, or both, which makes them distinguishable.

DAGs are studied by Stanley and Robinson, among other authors. And also its necessary mathematical framework, not only by classical combinatorial methods, but by new tools, as can be the Ihara-Selberg function for graphs.

Because it would be a long sequence of mathematicians, as Erdös, Rényi, Harary, Steinsky, Gillispie, Perlman, Andersson, Madigan, Studený, Volf, Hashimoto, Sunada, and so on.

Let

$$F(s) \equiv \sum_{i\in N^*} \frac{c_i\ s^i}{i!}$$

(*Robinson*, 1973) denotes as Δ the linear operation on exponential generating functions which divides by $\exp_2 C_{i,2}$, i.e.

$$\Delta F(s) \equiv \sum_{i\in N^*} \frac{c_i\ s^i}{i!\ \exp_2 C_{i,2}}$$

Then, we can use the function $\Delta F(s)$ to count labeled DAGs. It will be called as the *Special Generating Function for F*.

If we denote by a_n the number of labeled n-DAGs, Grabner and Steinsky have analyzed the zeroes of the function

$$\Delta \exp(-s) = \sum_{i\in N^*} \frac{(-s)^i}{i!\ \exp_2 C_{i,2}}$$

by Mathematical Analysis, more exactly by *Theory of Residua.*

It is possible to use generating functions to count labeled DAGs. For these, it is necessary to make intervene the *Inclusion-Exclusion Principle (IEP).*

So, if we take the set of *n-essential graphs,* and denote its cardinal by a_n, applying the IEP, we obtain

$$a_n = \sum_{s=1,\dots,p} (-1)^{s+1} \sum_{\substack{i_j \\ j\in\{1,\dots,s\}}} c\left(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_s}\right)$$

where

$$A_k = \{G \in E : k \text{ is a } \textit{terminal node} \text{ of } G\},$$
$$\textit{with } k = 1,\ 2,\ \dots,\ n \quad [*]$$

7. Conclusion

So, we are analyzed here the most useful tools to work with problems of

- *Enumeration,*
- *Combinatorics,*
- *Cycle Index,*
- *Generating Functions,*
- *Formal Series,*

and

- *Zeta functions,*

basically related with different graphs, as

- *UGs (Undirected Graphs),*
- *DAGs (Directed Acyclic Graphs),*

- *CGs (Chain Graphs),*

a set which subsumes the precedent subclasses, and so on.

Our objective, to obtain a consistent mathematical framework, with provide sufficient power to handle adequatly such type of problems, is (possibly) now reached.

References

S. A. Anderson, D. Madigan, and M. D. Perlman (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25**: 505-541.

H. Bass (1992). The Ihara-Selberg Zeta Function of a tree lattice. *Int. J. Math.*, **3**(6): 717-797.

H. Bass (1989). Zeta functions of finite graphs and representations of p-adic groups. *Advanced Studies in Pure Math.*, **15**: 211-280.

E. A. Bender, and S. Gill Williamson (2006). *Foundations of Combinatorics with Applications.* Dover Publ.

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 **B44** (3): 363-369.

B. Bollobás (1978). *Extremal Graph Theory.* Academic Press, New York. Reedited by Dover Publ.

B. Bollobás (1998). *Modern Graph Theory.* Springer Verlag, New York.

Ch. A. Charalambides (2002). *Enumerative Combinatorics.* Chapman and Hall/CRC.

P. Flajolet , and R. Sedgewick (2009). *Analytic Combinatorics.* Cambridge University Press.

A. Garrido (2009). Bayesian Networks and Essential Graphs. *IEEE Computer Society Press.* Enlarged and modified paper of *Proc. CANS 2008,* Petru Maior University Press, Tirgu Mures, 2008. To appear, 12 pp.

S. Gill Williamson (2002). *Combinatoris for Computer Science.* Dover Publ.

S. B. Gillispie, and M. D. Perlman (2002). The size distribution for Markov equivalence classes of acyclic digraph models. *AI* **141** (1/2): 137-155.

S. B. Gillispie, and M. D. Perlman (2001). Enumerating Markov equivalence classes of acyclic digraph models. *UAI 2001:* 171-177.

I. P. Goulden, and D. M. Jackson (1983). *Combinatorial Enumeration.* Wiley, New York.

P. J. Grabner and B. Steinsky (2005). Asymptotic behaviour of the poles of a special generating function for acyclic digraphs. *Aequationes Mathematicae* **70** (3): 268-278.

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.

F. Harary (1969). *Graph Theory.* Addison-Wesley, Reading, Mass., USA.

F. Harary, and E.M. Palmer (1973). *Graphical Enumeration.* Academic Press. New York.

F. Harary (1965). *Structural Models: An Introduction to the Theory of Graphs.* Wiley, New York.

K. Hashimoto (1990). On Zeta and L-functions of finite graphs. *I. J. Math.* **1**: 381-396.

Y. Ihara (1966). Discrete subgroups of PL (2, k). *Proc. Sympos. Pure MathBoulder, Colorado*: 219-235.

Y. Ihara (1996). On discrete subgroups of the two by two projective linear groups over p-adic fields. *J. Math. Soc. Japan* **18**: 219-235.

J.W. Kennedy, K. A. Mc Keon, E. M. Palmer, and R. W. Robinson (1990). Asymptotic number of symmetries in locally restricted trees. *Discrete Applied Mathematics* **26 (1):** 121-124.

M. Kotani, and T. Sunada (2000). Zeta Functions of Finite Graphs. *J. Math. Si. Univ. Tokio,* **7**: 7-25.

S. Lando (2003). *Lectures on Generating Functions.* AMS Series.

G. Pólya (1937, 1987). Kombinatorische Anzalhbestimmungen für Grüppen, Graphen und chemische Verbindungen. *Acta Mathematica* **68:** 145-254. Translated with commentaries by R. C. Read, in *Combinatorial enumeration of groups, graphs...* Springer.

M. Riedel (2006). Pólya´s enumeration theorem and the symbolic method. PostScript file.

J. Riordan (2002). *Introduction to Combinatorial Analysis.* Dover Publ.

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

R. W. Robinson (1970). Enumeration of acyclic digraphs. *Proc. Second Chapel Hill Conf. on Combinatorial Mathematics and Its Applications:* 391-399. Univ. North Carolina, Chapel Hill.

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

R. P. Stanley (1986). *Enumerative Combinatorics.* Vols. I and II. Wadsworth and Brooks/Cole Advanced Boks and Software.

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.

M. Studeny (1987). Asymptotic behaviour of empirical multiinformation. *Kybernetika* **23** (2): 124-135.

M. Studeny and M. Volf (1999). A graphical characterization of the largest chain graphs. *IJAR* **20**: 209-236.

T. Sunada (1985). L-functions in geometry and some applications. *Lecture Notes in Mathematics* **1201**: 266-284.

H. S.Wilf (2006), *Generatingfunctionology.* A K Peters, 3rd edition.