Analytic combinatorics of graphs

Notes for Alea Young 2016

Élie de Panafieu*
Bell Labs France, Nokia

September 6, 2016

1 Introduction

Analytic combinatorics is the branch of combinatorics that analyzes families of combinatorial objects using their generating functions. Those are series which coefficients contain the combinatorial information on the objects. This field has many qualities that make it great for the analysis of random graphs:

- from the generating functions, many combinatorial information can be extracted, such as precise asymptotics (with as many error terms as wanted), and moments and limit laws of parameters,
- the generating functions can be combined to represent new interesting objects,
- the method is robust: a small perturbation of the model requires often only a small adjustment in the generating functions.

Sadly, I will not have time in those notes to illustrate those qualities. Instead, I will focus on the variety of graph models and constraints that can be investigated using analytic combinatorics.

We first present classic results on connected (multi)graphs and the structure of random (multi)graphs. Then variants of the classic graph model are introduced, with constraints on the degrees of the vertices, the subgraphs allowed, the number of vertices allowed in each edge (hypergraphs), or with colored vertices which connectivity varies according to their color (inhomogeneous graphs). This presentation focuses on the combinatorial decompositions and the generating function manipulations, avoiding most of the analytic technicalities.

1.1 Notations

A *multiset* is a collection of objects, without order, where repetitions are allowed. A *set* is then a multiset without repetitions, and a *sequence*, or *list*, or *tuple*, is an ordered multiset. We denote multisets and sets by the bracket notation $\{2, 3, 7\}$, and sequences by the parenthesis notation $(3, 2, 7)$. The nth coefficient of the generating function

$$f(z) = \sum_{n \geq 0} f_n z^n$$

is denoted by $f_n = [z^n]f(z)$. The derivative of the function f is denoted by ∂f or f'.

*Email: depanafieuelie[at]gmail.com

1.2 Technical lemmas

In those notes, our primary objective is to derive exact expressions for the number of graphs that satisfy some properties. Those numbers will be expressed as the coefficients of generating functions, characterized by various relations. Analytic combinatorics (see Flajolet and Sedgewick (2009)) has developed many “black box” theorems that can be applied to obtain the asymptotics of generating function coefficients. The choice of the theorem depends of the form of the generating function. During the first reading, we suggest not to spend too much time on the technical conditions of those theorems, but rather to recognize the main features.

Theorem 1 (Singularity analysis)**.** *We consider a series $f(z)$ of positive radius of convergence ρ, analytic on the set*

$$\Delta = \{z \mid |z| < R,\ z \neq \rho,\ |\arg(z-\rho)| > \phi\}$$

for some values $R > \rho$ and $0 < \phi < \frac{\pi}{2}$. If

$$f(z) \sim (1 - z/\rho)^{-\alpha} \text{ as } z \to \rho \text{ while } z \in \Delta,$$

for some $\alpha \notin \{0, -1, -2, \ldots\}$, then

$$[z^n]f(z) \sim \frac{\rho^{-n} n^{\alpha-1}}{\Gamma(\alpha)}.$$

The following theorem analyses the singularity of generating functions characterized implicitly. This is in particular the case for trees.

Theorem 2 (implicit functions)**.** *Consider a function $\phi(u)$ analytic at $u = 0$, with nonnegative coefficients, $\phi(0) \neq 0$, and that is not of the form $\phi(u) = \phi_0 + \phi_1 u$. Furthermore, assume that the equation*

$$\phi(\tau) - \tau\phi'(\tau) = 0$$

admits a real positive solution, smaller than the radius of convergence of ϕ. Then the function $y(z)$ defined implicitly by the relation

$$y(z) = z\phi(y(z))$$

has radius of convergence $\rho = \frac{\tau}{\phi(\tau)}$ and is analytic on a set Δ of the form given in Theorem 1.

The Laplace method is a classic analytic technique (see Pemantle and Wilson (2013) for more details and general results).

Theorem 3 (Laplace method)**.** *We consider a neighborhood C of the origin in $\mathbb{R}^d$, and two analytic functions A and ϕ from C to $\mathbb{C}$. Suppose that the real part of $\phi(x) - \phi(0)$ is strictly positive on C except at the origin, and that its Hessian matrix H is nonsingular there. If A does not vanish at the origin, then*

$$\int_{x \in C} A(x) e^{n\phi(x)} dx \sim \frac{A(0)e^{n\phi(0)}}{\sqrt{\det(H)}} \left(\frac{2\pi}{n}\right)^{d/2}.$$

The *Cauchy integral* transforms a coefficient extraction into a complex integral on a small loop around the origin

$$[z^n]f(z) = \frac{1}{2i\pi} \oint \frac{f(z)}{z^n} \frac{dz}{z} \overset{z=\zeta e^{i\theta}}{=} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\zeta e^{i\theta})}{\zeta^n e^{in\theta}} d\theta.$$

A corollary of the Laplace method is then the *large powers theorem*.

Theorem 4 (Large powers theorem)**.** *We consider integers n, N, such that N/n has a positive limit λ. Let $B(z) = \sum_{n \geq 0} b_n z^n$ be a series with nonnegative coefficients and radius of convergence ρ_B, that satisfies $\gcd\{i - j \mid b_i \neq 0,\ b_j \neq 0\} = 1$ (thus $B(z)$ is not a function of the form $z^r C(z^p)$ for some integer $p \geq 2$ and some function C analytic at 0). We introduce the function $L(z) = \frac{zB'(z)}{B(z)}$, and assume that the equation*

$L(z) = \lambda$ has a positive solution $\zeta < \rho_B$ (which is then unique), and that $L'(\zeta) \neq 0$. Let $A(z)$ be a generating function with radius of convergence greater than ζ, that does not vanish at ζ. Then

$$[z^N]A(z)B(z)^n \sim \frac{A(\zeta)}{\sqrt{2\pi n \zeta L'(\zeta)}} \frac{B(\zeta)^n}{\zeta^N}.$$

The value ζ from the large powers theorem is called the *saddle-point*.

Exercise 1. Using the elementary property

$$[z^n]f(\alpha z) = \alpha^n [z^n] f(z),$$

transform the right hand-side of

$$\frac{1}{n!} = [z^n]e^z,$$

so that the large powers theorem can be applied, and prove Stirling formula.

1.3 Labels

This is an informal introduction to labelled objects and exponential generating functions. For more information, we recommend the first chapters of the book of Flajolet and Sedgewick (2009).

A labelled object is a set of labelled atoms, with some structure on it. In a tree, for example, the atoms are the vertices. The size of an object a, denoted by $|a|$, is then its number of atoms. The labels on the atoms are distinct integers. Given an object of size n, there exists a unique way to relabel its atoms in $\{1, 2, \dots, n\}$, so that the relative order of the atoms stays the same. When counting labelled objects, we thus assume without lost of generality that the labels are consecutive integers starting at 1. For example, a permutation of size n can be represented as a sequence of n distinct integers from $\{1, 2, \dots, n\}$. Therefore, permutations are labelled combinatorial objects.

We use an *exponential generating function* to represent a labelled combinatorial family $\mathcal{F}$

$$F(z) = \sum_{a \in \mathcal{F}} \frac{z^{|a|}}{|a|!} = \sum_{n \geq 0} f_n \frac{z^n}{n!},$$

where f_n denotes the number of objects of size n in $\mathcal{F}$. The reason of this convention is that many natural combinatorial constructions on labelled families translate well into exponential generating function operations:

- the generating function of the disjoint union of two families is the sum of their generating function

$$\mathcal{C} = \mathcal{A} \uplus \mathcal{B}, \text{ implies } C(z) = A(z) + B(z).$$

- The relabelled Cartesian product $\mathcal{C}$ of two labelled families $\mathcal{A}$ and $\mathcal{B}$ is defined as the pairs in $\mathcal{A} \times \mathcal{B}$, where the two objects are relabelled to ensure that the atoms have distinct labels. For example, the pair of permutations $((1, 3, 2), (2, 1))$ is not a proper labelled object, since the atoms 1 and 2 appears twice. The ten corresponding relabelled pairs of permutations are

$$((1,3,2),(5,4)),\ ((1,4,2),(5,3)),\ ((1,4,3),(5,2)),\ ((1,5,2),(4,3)),\ ((1,5,3),(4,2)),$$
$$((1,5,4),(3,2)),\ ((2,4,3),(5,1)),\ ((2,5,3),(4,1)),\ ((2,5,4),(3,1)),\ ((3,5,4),(2,1)).$$

  Observe that the relative orders of the atoms are preserved. The generating function of the relabelled Cartesian product of two labelled families is

$$\mathcal{C} = \mathcal{A} \star \mathcal{B}, \text{ implies } C(z) = A(z)B(z).$$

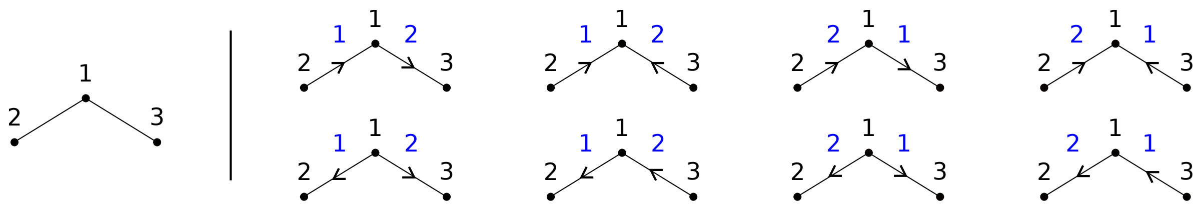

Figure 1: A graph and the corresponding multigraphs.

- The generating function of sequences of labelled objects from $\mathcal{A}$ is $1/(1 - A(z))$ (again, a relabelling occurs). Indeed, the combinatorial family equal to this sequence is $\cup_{n \geq 0} \mathcal{A}^n$, which has generating function $\sum_{n \geq 0} A(z)^n = \frac{1}{1-A(z)}$.
- The generating function of sets of labelled objects from $\mathcal{A}$ is $\exp(A(z))$. Indeed, a set of n objects from $\mathcal{A}$ is a sequence, considered up to any of the $n!$ permutations, so the family of sets of n objects from $\mathcal{A}$ has generating function $\frac{A(z)^n}{n!}$, and the union over n leads to the exponential.
- The generating function of oriented cycles of labelled objects from $\mathcal{A}$ is $\log\left(\frac{1}{1-A(z)}\right)$. A cycle of n objects from $\mathcal{A}$ is a sequence, considered up to any of the n circular permutations, so the generating function of those cycles is $\frac{A(z)^n}{n}$, and the sum over n is the logarithm.

We will see examples of those combinatorial constructions in the following sections. For the analysis of graphs, we will use two different kinds of atoms: the vertices and the edges.

Exercise 2. Prove that $C = A \uplus B$ implies $\mathcal{C} = \mathcal{A} + \mathcal{B}$.

Exercise 3. Let a, b be two labelled objects of size $|a|$ and $|b|$. How many relabelled pairs correspond to (a, b)?

Exercise 4. Prove that if $\mathcal{C} = \mathcal{A} \star \mathcal{B}$, then $C(z) = A(z)B(z)$.

1.4 Models: graphs and multigraphs

We define a multigraph $G = (V, E)$ as a labelled set V of vertices, and a labelled multiset E of edges, where each edge is an oriented pair of vertices. An edge e is then a triplet (u, v, ℓ), where u and v are the vertices linked by e, and ℓ is the label of e.

Exercise 5. How many multigraphs with n vertices and m edges are there?

The combinatorics and graph theory communities usually work on *graphs* instead of multigraphs. The difference is that in a graph, the edges are unlabelled and unoriented. Furthermore, loops (an edge linking a vertex to itself) and multiple edges (set of edges linking the same two vertices) are forbidden. However, multigraphs turn out to be better suited for generating function manipulations than graphs. For simplicity, in this notes, we will therefore focus on multigraphs. All the results presented can be derived for graphs as well.

Examples of graphs and multigraphs are displayed in Figure 1. The number of vertices of a multigraph G is denoted by $n(G)$, and its number of edges by $m(G)$. We also define its excess as $k(G) = m(G) - n(G)$.

Exercise 6. When we erase the edge orientations and labels of a multigraph that contains neither loops nor multiple edges, we obtain a graph. How many multigraphs correspond to a given graph?

Since multigraphs have labelled vertices and labelled edges, we use for them generating functions exponential with respect to both z and w. Furthermore, because their edges are oriented, we introduce a weight $1/2$ on them. The generating function of a multigraph family $\mathcal{F}$ is then defined as

$$F(z, w) = \sum_{G \in \mathcal{F}} \frac{w^{m(G)}}{2^{m(G)} m(G)!} \frac{z^{n(G)}}{n(G)!}.$$

Exercise 7. With this convention, what is the generating function of all multigraphs?

2 Trees and unicycles

Historically, the first graphs families enumerated where trees, in 1860 by Borchardt, and unicycles by Rényi (1959). Recall that the excess of a graph is the difference between its number of edges and vertices. Trees have excess -1, which is the minimum possible excess for a connected graph. A unicycle is a connected multigraph of excess 0.

Theorem 5. *The generating functions of rooted trees, trees, and unicycles are characterized by the relations*

$$T(z) = ze^{T(z)},$$
$$U(z) = T(z) - T(z)^2/2,$$
$$V(z) = \frac{1}{2} \log\left(\frac{1}{1 - T(z)}\right).$$

Proof. A rooted tree is a vertex (the root) and a set of sons, which are themselves rooted trees, so

$$T(z) = ze^{T(z)}.$$

Each tree of size n correspond to n rooted trees (number of possible choices for the root), so

$$zU'(z) = \sum_{n \geq 0} n u_n \frac{z^n}{n!} = T(z),$$

and we can check that $T(z) - T(z)^2/2$ is the unique solution of this differential equation (a more combinatorial proof relies on the *dissymmetry theorem*, from Bergeron et al. (1997)). Any unicycle can be uniquely decomposed as a non-oriented cycle, where each vertex is replaced by a rooted tree so

$$V(z) = \frac{1}{2} \log\left(\frac{1}{1 - T(z)}\right).$$

□

Asymptotics expressions for the number of trees and unicycles with n vertices can be extracted using singularity analysis.

Exercise 8. The Lagrange inversion states that if $y(z)$ is characterized by the relation $y(z) = z\phi(y(z))$ with $\phi(0) \neq 0$, then its coefficients are

$$[z^n]y(z) = \frac{1}{n}[u^{n-1}]\phi(u)^n.$$

Give an exact expression for the number of rooted trees with n vertices.

Exercise 9. Simplify the expression $zT'(z)$.

Exercise 10. What is the generating function of the unicycles that contain a loop or a multiple edge?

We can also prove that random multigraphs with a small number of edges typically contain only trees and unicycles, a result first derived by Erdős and Rényi (1960). To do so, we compare the number of such multigraphs to the total number of multigraphs. Recall that the excess of a multigraph is the difference between its number of edges and vertices.

Theorem 6. *When m/n tends toward a constant smaller than $1/2$, almost all multigraphs with n vertices and m edges contain only trees and unicycles.*

Proof. Trees have excess -1, and unicycles excess 0. Therefore, a multigraph of excess $k = m - n$ (which is negative when $m/n < 1/2$) that contains only trees and unicycles is a set of $-k$ trees and a set of unicycles

$$\frac{U(z)^{-k}}{(-k)!} e^{V(z)}.$$

So the number of such multigraphs with n vertices and m edges is

$$n! 2^m m! [z^n] \frac{U(z)^{n-m}}{(n-m)!} e^{V(z)}.$$

We apply the large powers theorem to extract the asymptotics of this expression

$$n! 2^m m! [z^n] \frac{U(z)^{n-m}}{(n-m)!} e^{V(z)} \sim \frac{n! 2^m m!}{(n-m)!} \frac{U(\zeta)^{n-m} e^{V(\zeta)}}{\zeta^n \sqrt{2\pi(-k)\zeta\phi'(\zeta)}},$$

where $\phi(z) = \frac{zU'(z)}{U(z)}$ and ζ is the unique positive solution of $\phi(\zeta) = \frac{n}{n-m}$. After some computations, we find

$$n! 2^m m! [z^n] \frac{U(z)^{n-m}}{(n-m)!} e^{V(z)} \sim n^{2m},$$

which is the total number of multigraphs with n vertices and m edges. Therefore, when m/n has a limit smaller than $1/2$, almost all multigraphs with n vertices and m edges contain only trees and unicycles. □

Exercise 11. Why can we not reach the same conclusion when m/n has a larger limit than $1/2$?

3 Connected multigraphs with fixed excess

In this section, we present results from Wright (1980). A kernel is a multigraph with minimum degree at least 3.

Lemma 1. *The number of kernels of a given excess is finite: a kernel of excess k contains at most $2k$ vertices and $3k$ edges. Those bounds are reached by cubic multigraphs, i.e. multigraphs where all vertices have degree exactly 3.*

Proof. We consider any kernel with n vertices, m edges, and excess $k = m - n$. The sum of the degrees of all vertices is equal to twice the number of edges, and each vertex has degree at least 3, so

$$2m = \sum_{\text{vertex } v} \deg(v) \geq 3n,$$

which implies $n \leq 2k$ and $m \leq 3k$. Those bounds are reached when $\deg(v) = 3$ for each vertex v. □

In the previous section, we derived the generating functions of connected multigraphs with excess -1 (trees) and 0 (unicycles). We now consider connected multigraphs with positive excess.

Theorem 7. *For any $k \geq 1$, there exists a computable polynomial $Q_k(T)$ such that the generating function of connected multigraphs of excess k is*

$$\mathrm{CMG}_k(z) = \frac{Q_k(T(z))}{(1 - T(z))^{3k}}.$$

Proof. Let us define a path of trees as a sequence

$$(\text{edge}, \text{rooted tree}, \text{edge}, \text{rooted tree}, \ldots, \text{edge}),$$

where the vertices are labelled, and the edges are labelled and oriented. Each edge links the roots of the two neighbor trees in the sequence, except the first and last edges. The generating function of path of trees is

$$\frac{1}{1 - T(z)}.$$

Observe that a path of trees contains one more edge than its number of vertices. As illustrated in Figure **??**, any connected multigraph with positive excess can be uniquely decomposed as a connected kernel where

- vertices are replaced by rooted trees,
- edges are replaced by paths of trees.

By construction, the kernel has the same excess as the multigraph. As a consequence of Lemma 1, the generating function of connected kernels of excess k is a multinomial $CK_k(z, w)$ of power $3k$ in w. Therefore, the generating function of connected multigraphs of excess k is

$$\mathrm{CMG}_k(z) = CK\left(T(z), \frac{1}{1 - T(z)}\right) = \frac{Q_k(T(z))}{(1 - T(z))^{3k}}.$$

□

The asymptotics of connected multigraphs with n vertices and excess k is then derived by application of a singularity analysis.

Exercise 12. According to Kuratowski's Theorem, a multigraph is planar if and only if its kernel contains neither K_5 (the complete multigraph on 5 vertices) nor $K_{3,3}$ (the complete bipartite 3×3 multigraph) as a subgraph. Given an integer k, can we compute the generating function of connected planar multigraphs of excess k?

The generating function of multigraphs of excess k that contain trees and unicycles and exactly c_ℓ components of excess ℓ for all $1 \leq \ell \leq L$ is

$$\frac{U(z)^{-k+K}}{(-k+K)!} e^{V(z)} \frac{\prod_{\ell=1}^{L} Q_\ell(T(z))^{c_\ell}}{(1 - T(z))^{3K}},$$

where $K = \sum_{\ell=1}^{L} \ell c_\ell$. Following the same principle as in the previous section, but pushing the analysis further, Janson et al. (1993) proved that the limit probability for a random graph with n vertices and m edges to contain exactly c_ℓ components of excess ℓ for all $1 \leq \ell \leq L$ is non-zero only when $m = \frac{n}{2}(1 + \mathcal{O}(n^{-1/3}))$, and they computed this limit probability in that case.

Using a probabilistic approach, Erdős and Rényi (1960) proved that when m/n has a limit greater than $1/2$, a typical random graph with n vertices and m edges contains only trees, unicycles, and a unique component of positive excess, called the *giant component*. Analyzing the statistics of this giant component using analytic combinatorics is an open problem.

Figure 2: A multigraph and its representation as a set of vertices with labelled half-edges.

4 Multigraphs with degree constraints

The goal of this section is the enumeration of multigraphs with n vertices, m edges, and where each vertex has its degree in a given set D. We denote by

$$\mathrm{Set}_D(z) = \sum_{d \in D} \frac{z^d}{d!}$$

the exponential generating function of this set, and assume that

- D contains at least 2 elements,
- $\gcd\{d_1 - d_2 \mid d_1, d_2 \in D\} = 1$.

The first assumption discards the enumeration of *regular multigraphs*, where all vertices have the same degree, and that can be analyzed separately. The second assumption just simplifies the analysis, and the general case has been treated by de Panafieu and Ramos (2016).

Theorem 8. *The number of multigraphs with n vertices, m edges, and all vertices having their degree in the set D is*

$$(2m)![x^{2m}] \,\mathrm{Set}_D(x)^n.$$

Proof. Let us consider a multigraph G, and cut each edge into two labelled half-edges. Specifically, an edge labelled ℓ and oriented from the vertex u to the vertex v is replaced by a half-edge labelled $2\ell - 1$ and attached to u, and a half-edge labelled 2ℓ and attached to v. As illustrated in Figure 2, this transforms the multigraph G into a set of vertices, to each of which is attached a set of labelled vertices. The size of each of those sets is the degree of the vertex, and the total number of half-edges is twice the initial number of edges. Therefore, the number of graphs with n vertices, m edges, and having all their degrees in D is

$$(2m)![x^{2m}] \,\mathrm{Set}_D(x)^n.$$

□

Exercise 13. When $m/n \to \pi$, derive the asymptotic number of multigraphs with n vertices, m edges, and where the degree of each vertex is a prime number.

5 Connected multigraphs with large excess

In this section, we derive the asymptotic number of connected graphs when the ratio of the number of edges over the number of vertices has a positive limit. This result has been derived by Bender et al. (1990); Pittel and Wormald (2005); van der Hofstad and Spencer (2006); de Panafieu (2016), using various methods.

With the convention of Section 1.4, the generating function of all multigraphs is

$$\mathrm{MG}(z, w) = \sum_{n \geq 0} e^{n^2 w/2} \frac{z^n}{n!},$$

because when ordering the edges according to their labels and orientations, a multigraph with n vertices and m edges becomes a sequence of $2m$ vertices in $\{1, 2, \dots, n\}$, so there are n^{2m} such multigraphs. Since a multigraph is a set of connected multigraphs, the generating function of connected multigraphs $\mathrm{CMG}(z, w)$ satisfies the relation

$$\mathrm{MG}(z, w) = e^{\mathrm{CMG}(z,w)}.$$

Taking the logarithm, we obtain the classic closed form for the generating function of connected multigraphs

$$\mathrm{CMG}(z, w) = \log \left(\sum_{n \geq 0} e^{n^2 w/2} \frac{z^n}{n!} \right).$$

Observe that the argument of the logarithm is a series with a zero radius of convergence. Therefore, we cannot use any analytic property of the logarithm, and the only way to extract the asymptotics seems to be to expand it as a series

$$\mathrm{CMG}(z, w) = \sum_{q \geq 1} \frac{(-1)^{q+1}}{q} \left(\sum_{n \geq 0} e^{n^2 w/2} \frac{z^n}{n!} \right)^q. \tag{1}$$

This expression was the starting point of the analysis of Flajolet et al. (2004), who worked on connected graphs with fixed excess. If we extract the coefficient $n!2^m m![z^n w^m]$, we obtain an exact expression for the number of connected multigraphs with n vertices and m edges

$$\mathrm{CMG}_{n,m} = \sum_{q=1}^{n} \frac{(-1)^{q+1}}{q} \sum_{\substack{n_1 + \cdots + n_q = n \\ \forall j,\ n_j \geq 1}} \binom{n}{n_1, \dots, n_q} \left(n_1^2 + \cdots + n_q^2 \right)^m.$$

However, as already observed by those authors, it is difficult to extract the asymptotics, because of "*magical*" cancellations in the coefficients. In particular, the dominant contribution to the sum does not come from the first value $q = 1$, because the summand is then the number of (non-empty) multigraphs with n vertices and m edges. Those multigraphs are indeed typically not connected, as they contain many trees and unicycles.

Instead of working on this expression using complicated analysis, we will derive a different (although similar) expression, better suited for asymptotics analysis. The main idea, already applied by Pittel and Wormald (2005), is to consider the family $\mathrm{MG}^{>0}$ of multigraphs without trees and unicycles. We call them *positive multigraphs*, since all their components have a positive excess. Let $\mathrm{CMG}^{>0}$ denote the set of connected multigraphs with positive excess. A set of connected multigraphs with positive excess is either empty, or is a positive multigraph, so

$$e^{\mathrm{CMG}^{>0}(z,w)} = 1 + \mathrm{MG}^{>0}(z, w), \qquad \text{which implies} \qquad \mathrm{CMG}^{>0}(z, w) = \log \left(1 + \mathrm{MG}^{>0}(z, w) \right).$$

Working with the excess instead of the number of edges, and denoting by $\mathrm{CMG}_k(z) = [y^k] \mathrm{CMG}(z/y, y)$ the generating function of connected graphs of excess k, we obtain the following expression

$$\mathrm{CMG}_k(z) = [y^k] \log \left(1 + \sum_{\ell > 0} \mathrm{MG}_{\ell}^{>0}(z) y^{\ell} \right) = \sum_{q \geq 1} \frac{(-1)^q}{q} \sum_{\substack{k_1 + \cdots + k_q = k \\ \forall j,\ k_j \geq 1}} \prod_{j=1}^{q} \mathrm{MG}_{k_j}^{>0}(z).$$

This expression looks similar to Equation (1). However, the dominant contribution to the sum will be easy to locate: it comes from the term $q = 1$

$$n![z^n] \mathrm{CMG}_k(z) \sim n![z^n] \mathrm{MG}_k^{>0}(z).$$

This means that a random positive multigraph with n vertices and excess $k = \Theta(n)$ is typically connected, a result first proven by Erdős and Rényi (1960). We skip the proof of this fact on those notes, but it is available in de Panafieu (2016).

Theorem 9. *The generating function of positive multigraphs of excess k is*

$$\mathrm{MG}_k^{>0}(z) = \frac{(2k)!}{2^k k!}[x^{2k}] \frac{e^{-V(z)}}{\left(1 - T(z)\frac{e^x-1-x}{x^2/2}\right)^{k+1/2}}.$$

Proof. A *core* is a multigraph of minimum degree 2. According to Theorem 8, the generating function of cores is

$$\mathrm{Core}(z,w) = \sum_{m \geq 0} (2m)![x^{2m}]e^{z(e^x-1-x)} \frac{w^m}{2^m m!}.$$

In this expression, after developing the exponential as a sum over n and applying the change of variable $m \leftarrow k+n$, we obtain

$$\mathrm{Core}(z,w) = \sum_{k \geq 0}[x^{2k}] \sum_{n \geq 0} \frac{(2(k+n))!}{2^{k+n}(k+n)!} \frac{\left(zw\frac{e^x-1-x}{x^2}\right)^n}{n!} w^k.$$

The sum over n is replaced by its closed form

$$\mathrm{Core}(z,w) = \sum_{k \geq 0}[x^{2k}] \frac{(2k)!}{2^k k!} \frac{w^k}{\left(1 - zw\frac{e^x-1-x}{x^2/2}\right)^{k+1/2}}.$$

The generating function of multicores of excess k is then

$$\mathrm{Core}_k(z) = [y^k]\,\mathrm{Core}(z/y, y) = \frac{(2k)!}{2^k k!}[x^{2k}] \frac{1}{\left(1 - z\frac{e^x-1-x}{x^2/2}\right)^{k+1/2}}.$$

In a multigraph, if we remove again and again all vertices of degree 0 and 1, the trees disappear, and the rest of the multigraph is reduced to a core. Conversely, any positive multigraph with a set of unicycles can be uniquely decomposed as a core where each vertex is replaced by a rooted tree. Furthermore, the core and the multigraph have the same excess, so

$$\mathrm{MG}_k^{>0}(z)e^{V(z)} = \mathrm{Core}_k(T(z)).$$

This implies

$$\mathrm{MG}_k^{>0}(z) = \mathrm{Core}_k(T(z))e^{-V(z)} = \frac{(2k)!}{2^k k!}[x^{2k}] \frac{e^{-V(z)}}{\left(1 - T(z)\frac{e^x-1-x}{x^2/2}\right)^{k+1/2}}.$$

□

What we gained with this new expression of the asymptotic number of connected multigraphs with n vertices and excess $k = m - n$

$$n!2^{k+n}(k+n)![z^n]\,\mathrm{CMG}_k(z) \sim n!2^{k+n}(k+n)! \frac{(2k)!}{2^k k!}[z^n x^{2k}] \frac{e^{-V(z)}}{\left(1 - T(z)\frac{e^x-1-x}{x^2/2}\right)^{k+1/2}}$$

is that the right-side expression can be analyzed using a bivariate large powers theorem. We express the coefficient extractions $[z^n x^{2k}]$ as Cauchy integrals and apply the Laplace method.

Exercise 14. Using the expression of the generating function of cores, express the generating function of kernels of excess k.

6 Multigraphs with forbidden subgraphs

We present in this section part of a work in progress of Collet et al. (2016). We consider a connected multigraph H that is not a tree, and assume it is *strictly balanced*, which means that its density is greater than the density of any of its subgraphs

$$\text{for all } G \subsetneq H,\ \frac{m(H)}{n(H)} > \frac{m(G)}{n(G)}.$$

We derive the limit probability for a random multigraph with n vertices and m edges to contain a copy of H as a subgraph. As usual, a copy of H is an isomorphic multigraph where the vertices and edges are relabelled in an increasing way (hence there is only one copy where the vertex and edge labels are consecutive integers starting at 1). This result was first proved by Erdős and Rényi (1960).

Lemma 2. *Let G be a multigraph built from two copies of H sharing at least one vertex, then the density of G is greater than the density of H*

$$\frac{m(G)}{n(G)} > \frac{m(H)}{n(H)}.$$

Proof. Let J denote the largest common subgraph of the two copies of H in G. Then the number of vertices and edges in G are

$$n(G) = 2n(H) - n(J), \qquad m(G) = 2m(H) - m(J).$$

Since H is strictly balanced, the density of J is smaller than the density of H, so

$$\frac{m(J)}{n(J)} < \frac{m(H)}{n(H)}, \quad \text{which implies} \quad \frac{m(G)}{n(G)} > \frac{m(H)}{n(H)}.$$

□

A *patchwork* is a set of copies of H, that might share vertices and edges (however, this is not a multiset, so two elements of a patchwork cannot be identical). This notion is illustrated in Figure 3. The number of distinct vertices of a patchwork P is denoted by $n(P)$, its number of distinct edges by $m(P)$, and the number of multigraphs in P is denoted by $|P|$. The density of a nonempty patchwork is then $m(P)/n(P)$. The generating function of patchworks is defined as

$$P(z, w, u) = \sum_{\text{patchwork } P} u^{|P|} \frac{w^{m(P)}}{2^{m(P)} m(P)!} \frac{z^{n(P)}}{n(P)!}.$$

Lemma 3. *The set of patchworks $P^\star$ that are either empty, or of excess $m(H)/n(H)$, has generating function*

$$P^\star(z, w, u) = \exp\left(u \frac{w^{m(H)}}{2^{m(H)} m(H)!} \frac{z^{n(H)}}{n(H)!} \right).$$

The density of all other patchworks is greater than the density of H.

Proof. A patchwork P is either a set of isolated copies of H, or contains at least two copies of H sharing at least a vertex. In the first case, P is either empty, or its density is equal to the density of H. In the second case, as a consequence of Lemma 2, the density of P is greater than the density of H. The generating function of a set of isolated copies of H is

$$\exp\left(u \frac{w^{m(H)}}{2^{m(H)} m(H)!} \frac{z^{n(H)}}{n(H)!} \right).$$

□

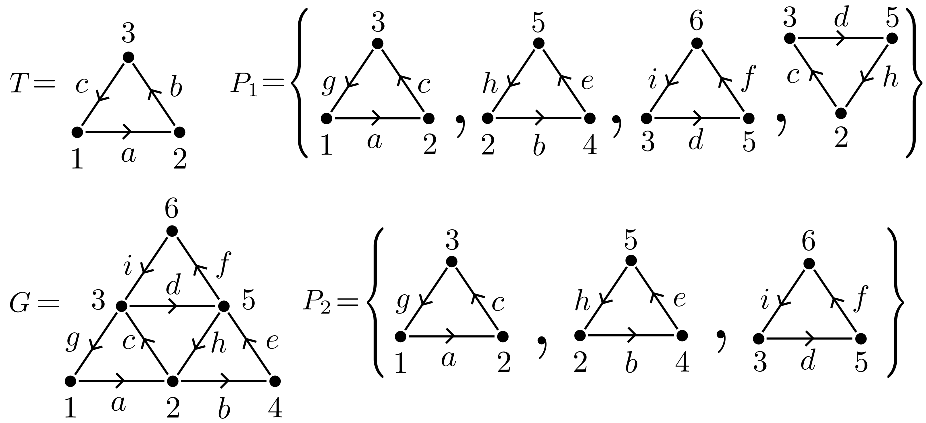

Figure 3: The multigraph H is here denoted by T. Two patchworks P_1 and P_2 are displayed. They both correspond to the same multigraph G.

Theorem 10. *The generating function of multigraphs where a variable u marks the number of occurrences of subgraphs copies of H is*

$$\mathrm{MG}(z,w,u)=\sum_{m\geq 0}(2m)![x^{2m}]P(ze^x,w,u-1)e^{z\exp(x)}\frac{w^m}{2^m m!}.$$

Proof. The generating function of multigraphs where each occurrence of the subgraph H is either marked with the variable u, or left unmarked, is $\mathrm{MG}(z,w,u+1)$. In such a multigraph G, by construction the set of marked subgraphs form a patchwork P. If we cut each edge that is not in P into two labelled half-edges, we obtain a representation of the multigraph G as

- a patchwork P, where each vertex comes with a set of half-edges,
- a set of vertices (the vertices of G that do not belong to P), each attached to a set of half-edges.

The total number of half-edges must be even. Denoting this number by m, and using the variable x to mark the half-edges, we obtain

$$\mathrm{MG}(z,w,u+1)=\sum_{m\geq 0}(2m)![x^{2m}]P(ze^x,w,u)e^{z\exp(x)}\frac{w^m}{2^m m!}.$$

□

The previous expression is not in a shape allowing the application of one of the theorems from Section 1.2, so we have to work a bit more to obtain asymptotics.

Lemma 4. *The number of multigraphs with n vertices, m edges, and that have no subgraph that is a copy of H is*

$$n!2^m m![z^n w^m]\,\mathrm{MG}(z,w,0)=n^{2m}\frac{n!}{n^n}\frac{m!}{m^m}\sqrt{2m}[z^n x^{2m}]\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}P\left(nz,\frac{2m}{n^2}\frac{x^2}{t^2},-1\right)e^{nz}e^{2mx}t^{2m}e^{-mt^2}dt. \tag{2}$$

Proof. Applying the classic identity

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^q e^{-t^2/2}dt=\begin{cases}0 & \text{if } q \text{ is odd,}\\ \frac{(2m)!}{2^m m!} & \text{if } q=2m,\end{cases}$$

we obtain that for any entire function f, we have

$$\sum_{m \geq 0} (2m)![x^{2m}]f(x) \frac{w^m}{2^m m!} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\sqrt{w}t) e^{-t^2/2} dt.$$

We apply this relation to the expression of $\mathrm{MG}(z, w, 0)$, derived in Theorem 10. The number of multigraphs with n vertices, m edges, and without any subgraph copy of H is then

$$n! 2^m m! [z^n w^m] \, \mathrm{MG}(z, w, 0) = n! 2^m m! [z^n w^m] \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} P\left(z e^{\sqrt{w}t}, w, -1 \right) e^{z \exp(\sqrt{w}t)} e^{-t^2/2} dt.$$

In order to apply the Laplace method and the large powers theorem with saddle-point 1, we transform this expression and apply successively the changes of variables

$$z \to n e^{-\sqrt{w}t} z, \quad w \to \left(\frac{2m}{n} \frac{x}{t} \right)^2, \quad t \to \sqrt{2m} t,$$

The expression becomes

$$n^{2m} \frac{n!}{n^n} \frac{m!}{m^m} \sqrt{2m} [z^n x^{2m}] \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} P\left(nz, \frac{2m}{n^2} \frac{x^2}{t^2}, -1 \right) e^{nz} e^{2mx} t^{2m} e^{-mt^2} dt.$$

□

Theorem 11. *Set* $\alpha = 2 - \frac{n(H)}{m(H)}$*, and consider integers* n *and* m *such that* n/m^α *has a positive limit* c*. Then the limit probability for a random multigraph with* n *vertices and* m *edges to contain no copy of* H *as a subgraph is*

$$\exp\left(- \frac{(2c)^{m(H)}}{m(H)! n(H)!} \right).$$

Proof. For simplicity, in this notes, we will assume that the generating function of patchworks satisfy the conditions of the Laplace method and the large powers theorems. Observe that when applying those techniques with saddle-points at 1, we have

$$\int_C A(x) e^{n\phi(x)} \sim A(0) \int_C e^{n\phi(x)}, \quad \text{and} \quad [z^N] A(z) B(z)^n \sim A(1) [z^N] B(z)^n.$$

Since we chose our changes of variables to ensure that the saddle-points are located at 1, we then obtain

$$n! 2^m m! [z^n w^m] \, \mathrm{MG}(z, w, 0) \sim n^{2m} P\left(n, \frac{2m}{n^2}, -1 \right),$$

where n^{2m} is the total number of multigraphs with n vertices and m edges. Therefore, the probability for a random multigraph with n vertices and m edges to contain no subgraph that is a copy of H has the same limit as $P\left(n, \frac{2m}{n^2}, -1\right)$.

Since m is negligible compared to n^2, the dominant contribution comes from patchworks with a small density, so we consider only the contribution of patchworks of density smaller or equal to the density of H. According to Lemma 3, it is equal to

$$\exp\left(- \frac{(2m/n^2)^{m(H)}}{m(H)!} \frac{n^{n(H)}}{n(H)!} \right).$$

It is then natural to consider m of the form cn^α, and we obtain

$$\exp\left(- \frac{(2c)^{m(H)}}{m(H)! n(H)!} n^{(\alpha - 2) m(H) + n(H)} \right).$$

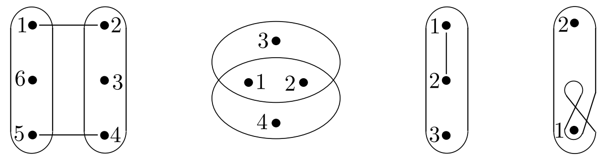

Figure 4: Four examples of hypergraphs.

If $\alpha < 2 - \frac{n(H)}{m(H)}$, then this exponential tends to 1, and almost no random multigraph with n vertices and m edges contains H as a subgraph. On the other hand, if $\alpha > 2 - \frac{n(H)}{m(H)}$, then this exponential tends to 0, and almost all multigraphs contain H as a subgraph. The value of interest is thus $\alpha = 2 - \frac{n(H)}{m(H)}$. In this case, the limit probability for a random multigraph to not contain any subgraph copy of H is

$$\exp\left(-\frac{(2c)^{m(H)}}{m(H)!n(H)!}\right).$$

□

We saw in Section 1.4 that each graph with m edges corresponds to exactly $2^m m!$ multigraphs (the number of possible edge orientations and labelling). Conversely, any set $\mathcal{F}$ of multigraphs each with m edges, stable by edge relabelling and change of orientation, and that contain neither loops nor multiple edges, can be reduced to a set of $|\mathcal{F}|/(2^m m!)$ graphs. The technique we presented in this chapter allows to remove loops and double edges (hence also multiple edges) from a multigraph family. This is how we translated many results on multigraphs to graphs.

Exercise 15. When n and m are proportional, find two expressions for the limit probability that a random multigraph with n vertices and m edges contains neither loops nor multiple edges.

7 Hypergraphs and inhomogeneous multigraphs

Hypergraphs are a generalization of graphs, where each hyperedge can contain 2 or more vertices, as illustrated in Figure 4. They are used to represent databases: each vertex represents an object, and each hyperedge an attribute. Most of the work on hypergraphs focuses on the *uniform* case, where all hyperedges contain the same number of vertices. Using analytic combinatorics, de Panafieu (2015b) adopted a more general setting, allowing any size of hyperedge in a given set D, and generalized to hypergraphs the results presented in Sections 2 and 3.

The inhomogeneous graph model is also known as the stochastic graph model, and is related to the Ising model (see Söderberg (2002); van der Hofstad (2014)). In this model, each vertex has a color, taken in a finite set, and only some colors are allowed to be linked by an edge. Those rules are encoded into a $\{0, 1\}$ symmetric matrix R. Properly k-colored graphs are a particular case, where each color can be linked to any different color. The matrix R is then the $k \times k$ matrix with 0 on the diagonal and 1's everywhere else. Any other Constraint Satisfaction Problem (CSP) where the constraints contain only two variables can be modeled by an inhomogeneous graph as well. The results of Sections 2 and 3 have been extended to inhomogeneous graphs by de Panafieu (2015a); de Panafieu and Ravelomanana (2015), using analytic combinatorics.

8 Conclusion

We could as well consider graphs where some subgraph is forbidden, and with degree constraints. Or hypergraphs with degree constraints. Or inhomogeneous hypergraphs (those count the satisfied instances of Constraints Satisfaction Problems, which is nice!). The properties we analyzed one by one can be combined at will.

The techniques used to analyze the structure of random multigraphs can be applied to the other models, but only few results in this direction have been published so far. So, if you are interested, do not hesitate to contact me.

In the models of multigraphs with degree constraints, hypergraphs and inhomogeneous multigraphs, we considered a set of admissible integer values. This is equivalent with assigning a weight 0 or 1 to each integer. We can as well consider real-valued weights. The hypergraphs, for example, are then counted with a weight, equal to the product of the weights of the sizes of the edges. This induces a random distribution on the sets of hypergraphs that is similar to the models introduced by the probabilists. Designing formal transfer theorems would be nice (a general theorem might already exist? I am interested).

References

E. A. Bender, E. R. Canfield, and B. D. McKay. The asymptotic number of labeled connected graphs with a given number of vertices and edges. *Random Structures and Algorithm*, 1:129–169, 1990.

F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1997.

G. Collet, E. de Panafieu, D. Gardy, B. Gittenberger, and V. Ravelomanana. Counting graphs with forbidden subgraphs. *work in progress*, 2016.

E. de Panafieu. Enumeration and structure of inhomogeneous graphs. *proceedings of the International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC), poster presentation*, page 12, 2015a.

E. de Panafieu. Phase transition of random non-uniform hypergraphs. *Journal of Discrete Algorithms*, 31 (0):26 – 39, 2015b.

E. de Panafieu. Counting connected graphs with large excess. *proceedings of the International Conference on Formal Power Series and Algebraic Combinatorics (Fpsac16)*, 2016.

E. de Panafieu and L. Ramos. Graphs with degree constraints. *proceedings of the Meeting on Analytic Algorithmics and Combinatorics (Analco16)*, 2016.

E. de Panafieu and V. Ravelomanana. Analytic description of the phase transition of inhomogeneous multigraphs. *European Journal of Combinatorics*, 2015.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17, 1960.

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

P. Flajolet, B. Salvy, and G. Schaeffer. Airy phenomena and analytic combinatorics of connected graphs. *Electronic Journal of Combinatorics*, 11(1), 2004.

S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures and Algorithms*, 4(3):233–358, 1993.

R. Pemantle and M. C. Wilson. *Analytic Combinatorics in Several Variables*. Cambridge University Press, New York, NY, USA, 2013.

B. Pittel and N. C. Wormald. Counting connected graphs inside-out. *Journal of Combinatorial Theory, Series B*, 93(2):127–172, 2005.

A. Rényi. On connected graphs I. *publication of the mathematical institute of the hungarian academy of sciences*, 4(159):385–388, 1959.

B. Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E*, 66:066–121, Dec 2002.

R. van der Hofstad. *Random Graphs and Complex Networks. Vol. I.* 2014.

R. van der Hofstad and J. Spencer. Counting connected graphs asymptotically. *European Journal on Combinatorics*, 26(8):1294–1320, 2006.

E. M. Wright. The number of connected sparsely edged graphs III: Asymptotic results. *Journal of Graph Theory*, 4(4):393–407, 1980.