



Threshold functions for small subgraphs in simple graphs and multigraphs

Gwendal Collet, Elie de Panafieu, Danièle Gardy, Bernhard Gittenberger,
Vlady Ravelomanana

► To cite this version:

Gwendal Collet, Elie de Panafieu, Danièle Gardy, Bernhard Gittenberger, Vlady Ravelomanana.
Threshold functions for small subgraphs in simple graphs and multigraphs. 2018. hal-01847358v2

HAL Id: hal-01847358

<https://hal.science/hal-01847358v2>

Preprint submitted on 1 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Threshold functions for small subgraphs in simple graphs and multigraphs

Gwendal Collet*

Élie de Panafieu†

Danièle Gardy‡

Bernhard Gittenberger§

Vlady Ravelomanana¶

July 15, 2018

Abstract

We revisit the problem of counting the number of copies of a fixed graph in a random graph or multigraph, for various models of random (multi)graphs. For our proofs we introduce the notion of *patchworks* to describe the possible overlappings of copies of subgraphs. Furthermore, the proofs are based on analytic combinatorics to carry out asymptotic computations. The flexibility of our approach allows us to tackle a wide range of problems. We obtain the asymptotic number and the limiting distribution of the number of subgraphs which are isomorphic to a graph from a given set of graphs. The results apply to multigraphs as well as to (multi)graphs with degree constraints. One application is to scale-free multigraphs, where the degree distribution follows a power law, for which we show how to obtain the asymptotic number of copies of a given subgraph and give as an illustration the expected number of small cycles.

Keywords. random graphs, subgraphs, analytic combinatorics, generating functions, power law.

1 Introduction

Since the introduction of the random graphs $G(n, m)$ and $G(n, p)$ by Erdős-Rényi [17] in 1960 one of the most studied parameters is the number of subgraphs isomorphic to a given graph F .

Throughout the paper, for a given graph or subgraph G , $E(G)$ (resp. $V(G)$) denotes the set of edges (resp. vertices). For a given graph F denote by $G[F]$ the number of copies of F contained in the random graph $G(n, p)$ as a subgraph. Observe that by the asymptotic equivalence between $G(n, p)$ and $G(n, m)$ (see for instance [28]) results from one model can be translated into the other rigorously. The distribution of $G[F]$ has been studied extensively since the seminal work of Erdős-Rényi [17] who gave the first results in this direction. A general threshold for $\{G[F] > 0\}$ has been established in 1981 by Bollobás [8] located at $p = n^{-1/\mu}$ where $\mu = \max_H \{|E(H)|/|V(H)| : H \subset F\}$. If the order of magnitude of p is smaller than the threshold, then asymptotically almost surely there is no subgraph F in

A major reference about the distribution of $G[F]$ is the work of Ruciński [40] stating that the number of copies of F is asymptotically normal if and only if $np^\mu \rightarrow \infty$ and $n^2(1-p) \rightarrow \infty$. In the same paper,

*Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstr. 8–10/104, 1040 Wien, Austria. Supported by the Austrian Science Foundation FWF, grant SFB F50-02.

†Nokia-Bell Labs and LINC (France). This work was partially founded by the Austrian Science Fund (FWF) grant F5004, the Amadeus program and the PEPS HYDrATA.

‡DAVID Laboratory, University of Versailles Saint Quentin (France). Partially supported by the Amadeus project 33697ZK *Threshold problems and phase transitions in graph-like structures* (2015–16) and by the ANR-MOST project MetaConc (2015–19).

§Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstr. 8–10/104, 1040 Wien, Austria. Supported by the Austrian Science Foundation FWF, grant SFB F50-03 and the ÖAD grant Amadée F01/2015.

¶IRIF, University of Paris 7 (France). Partially supported by the Amadeus project 33697ZK *Threshold problems and phase transitions in graph-like structures* (2015–16), by the project *Combinatorics in Paris* (2014–17) and by the CNRS-PICS project *Constraint analysis through analytic combinatorics* (2017–19).

Ruciński proved also that at the threshold the number of subgraphs F follows a Poisson law if and only if F is strictly balanced.

In this context, Janson, Oleszkiewicz and Ruciński [29] developed a moment-based method that gives estimates for $\mathbb{P}(G[F] \geq (1 + \epsilon)\mathbb{E}G[F])$ which are best possible up to logarithmic factors in the exponent (the authors proved that their exponential bounds on the upper tail of $G[F]$ are tight). It is important to remark that the notion of strongly balanced graphs, introduced by Ruciński and Vince [41], plays a key role to obtain the results mentioned above.

Recently, there has been an increasing interest in the study of constrained random graphs such as random graphs with given degree sequences or random regular graphs. In these directions, the number of given subgraphs in such structures has been also studied. For instance, Wormald [49] proved in his survey about random regular graphs that the number of short cycles in these structures follows asymptotically a Poisson distribution. McKay, Wormald and Wysocka [35] consider random regular graphs of degree d and show that, when g is such that $(d - 1)^{2g-1} = o(n)$, the numbers of cycles of length up to g are asymptotically distributed as independent Poisson variables. Kim, Sudakov and Vu [30], considering a regular unlabeled graph with n vertices of degree d and a fixed subgraph H , study how the probability that a copy of H occurs varies when d grows, show that this probability gets close to 1 for d around $n^{1-1/E(H)}$ and prove the convergence of the number of copies of H towards a Poisson distribution. Gao and Wormald [22] prove the asymptotic normality of the number of copies of a strictly balanced subgraph H in a random d -regular graph when d grows large.

Instead of constraining the whole graph, we can also require that the *subgraph* is regular. An article by Bollobás, Kim and Verstraëte [9] considers the appearance of such a k -regular subgraph when the density of the graph is around $4k$. Several papers have studied the relation between a k -regular subgraph and a k -core: Prałat, Verstraëte and Wormald [39] prove that the threshold for the appearance of a k -regular subgraph is at most the threshold for the appearance of a non-empty $(k + 2)$ -core, a result improved first by Chan and Molloy [10] to a $(k + 1)$ -core, and further by Gao [20], who showed that the size of a k -regular subgraph is “close” to the size of the k -core. Concurrently, Letzter [31] has obtained the existence of a sharp threshold for the existence of a k -regular subgraph for $k \geq 3$. Very recently, Mitsche, Molloy and Prałat [36] proved that a random graph $G(n, p = c/n)$ typically has a k -regular subgraph if $c > e^{-\Theta(k)}$, which is above the threshold for the appearance of a k -core.

Now extend regular graphs and consider graphs whose degrees form a specified degree sequence. An early reference on the enumeration of such graphs is the article of McKay and Wormald [34], followed by Greenhill and McKay [25] who studied the asymptotic number of sparse multigraphs with degree sequences. Barvinok [3] studies directed and bipartite graphs with prescribed degree, and two papers by Barvinok and Hartigan [4, 5] establish results about (uniform random) graphs with a given degree sequence. In [4], the authors count asymptotically the number of $m \times n$ matrices with prescribed row and column sums, so that their results can be applied to the number of graphs and bipartite graphs with prescribed degrees (on both sides); in [5] they obtain the number of labeled graphs where the degree sequence is fixed. More recently, Gao and Wormald [21] consider sparse graphs and present a survey of enumeration results for graphs with given degree sequences as well.

If we are interested in the appearance of subgraphs in graphs with specified degree sequences, a good survey of the results up to 2010 is by McKay [32]. McKay [33] again studies the structure of a random graph with a given degree sequence, including the probability of a given subgraph or induced subgraph. Chatterjee, Diaconis and Sly [11] consider a general model for (dense) graphs with a given degree sequence and the existence of a limit for sequence of such graphs; this allows them to obtain a general formula from which one might deduce results on the number of triangles (although not explicitly given). Very recent results by Greenhill et al. [24] give the asymptotic expected number of copies of a graph and of induced subgraphs in a random graph with a known degree sequence. As multigraphs model many real-world networks, subgraphs counts have also been derived for random multigraph models with prescribed degrees. For very recent works in these directions we refer to the preprint of Angel, van der Hofstad and Holmgren [1] where the authors consider multigraphs with prescribed degrees and study Poisson approximations of the number of self-loops and multiple edges as well as an estimate on the total variation distance between the number of self-loops and multiple edges and the Poisson limit of their sum. Barbour and Röllin [2] provide a general normal approximation theorem for local graph statistics in the same model.

The next step after fixing the degree sequence is allowing this sequence to follow some probability distribution. An important class of graphs with such a distribution is that of scale-free graphs, *i.e.*,

graphs where the degree distribution follows a power law which means that the probability that a vertex has degree d is proportional to $d^{-\gamma}$ for some γ . Results from the afore-mentioned article by Gao and Wormald [21] can be applied to some power-law sequences. Van der Hofstad [44] gives a nice presentation of the different models for graphs; see also his survey [43] on the configuration model.

Van der Hofstad, Janssen, van Leeuwen and Stegehuis [42, 45] consider triangles, or rather the *clustering coefficient*, in a class of simple graphs with a hidden variables model and a power-law degree distribution. Van der Hofstad, van Leeuwen and Stegehuis [46] consider the number of occurrences of a small connected graph, either as a subgraph or as an induced subgraph; all their results are for the so-called “erased” configuration model, which amounts to a simple graph model. In a companion paper [47] they consider clustering, *i.e.*, the probability of existence of an edge between two neighbours of a given vertex in the configuration model. When the degree of the vertex becomes at least of order $\sqrt{n}$, this probability becomes that of a power law. Their result can be used to derive the expected number of triangles, when a triangle with multiple edges is counted once (this is again the erased configuration model).

Our goal in this paper is to revisit some of these results and to extend them, through analytic combinatorics and extensive use of generating functions for counting graphs with a specified subgraph, or with a given number of subgraphs.

Ours is not the first paper that approaches graph problems with these tools. Roughly at the same time as the pioneer articles of Flajolet, Knuth and Pittel [18] about the appearance of cycles and of Janson *et al.* [27] about the birth of the giant component, higher-dimensional multivariate generating functions where variables are associated to vertices of the graph were used by McKay and Wormald [34]. Such highly multidimensional generating functions appear again in further papers, see McKay [32, 33] and Barvinok and Hartigan [5]. An important development was the study of planar graphs by Gimenez and Noy [23] through analytic combinatorics, followed by several papers in the same direction. E.g., the recent paper by Drmota, Ramos and Rué [15] deals with the limiting distribution of the number of copies of a subgraph in subcritical graphs. Noy, Réquilé and Rué [37] study precise properties of random cubic planar graphs including, most notably for the topics of this article, a proof of asymptotic normality for the number of triangles. Another relevant result, upon which we shall build Section 6, is the enumeration of graphs whose degrees must belong to a specific set, presented by de Panafieu and Ramos [14].

Overview of results

In the next section we give the formal definitions of our model and the objects we are interested in: simple graphs and multigraphs, possibly weighted.

Section 3 presents some combinatorial results on the expected number of subgraphs that can be obtained without resorting to analytic combinatorics tools. We give here the expected number of subgraphs belonging to a given family $\mathcal{F}$ in simple graphs or in multigraphs. The notion of *patchwork* of copies of subgraphs, which is defined there, allows us to study the distribution of the number of occurrences of a subgraph. We are then able to consider the total number (weight) of simple graphs or of multigraphs with a specific number of occurrences of subgraphs in $\mathcal{F}$. We finally derive a Poisson limiting distribution for the number of occurrences of a strictly balanced subgraph in a weighted multigraph.

The tools from analytic combinatorics which we shall use in the rest of the article are generating functions enumerating families of graphs. They are presented in Section 4; we also give there the first generating functions for the families we consider.

The following sections are devoted to probabilistic results under two different random models. In Section 5, (multi)graphs are chosen uniformly at random among all graphs of the same size (number of vertices and edges): This is reminiscent of the Erdős-Rényi $G(n, m)$ model. In Sections 6 and 7, we consider weighted (multi)graphs to study different degree distributions. Both distributions can be achieved by the means of Boltzmann samplers [16], presented in more detail in Section 6.2, which automatically translate our combinatorial decomposition of the (multi)graphs (weighted or not) into a random sampler. In particular, the Boltzmann sampler for multigraphs weighted according to their degrees produces multigraphs following the same distribution as the configuration model does. This equivalence, detailed in Section 6.2, bridges the gap between analytic combinatorics and the “pure” probabilistic setting.

We address the problem of evaluating the number of copies of a given subgraph in Section 5, for both simple graphs and multigraphs. Theorems 1 and 2 give exact and asymptotic expressions for the number

of multigraphs and simple graphs with one distinguished subgraph in $\mathcal{F}$; then Theorems 3 and 4 give the number of multigraphs and simple graphs with exactly t subgraphs in $\mathcal{F}$. The probability that there is at least one copy of a subgraph of $\mathcal{F}$ with high density goes to 0, as shown in Corollary 1 for multigraphs when $m = o(n^{2-1/d(F)})$; Corollary 2 is a more precise variant of this result when we know the essential density of the subgraph. Corollary 3 is the equivalent result for simple graphs; it gives a new proof of the upper bound on the average number of copies of a densest subgraph in a simple graph for $m = \mathcal{O}(n^\alpha)$, α fixed < 2 , and shows that the number of copies tends almost surely to 0 when $\alpha < 2 - 1/d^*(F)$, with $d^*(f)$ the essential density of the subgraph F . Finally, Theorems 5 and 6 prove a Poisson distribution for the number of copies of a strictly balanced subgraph F for a random multigraph and for a random simple graph, respectively, in the range $m \sim cn^{2-1/d(F)}$ where $d(F)$ is the density of the subgraph F .

Section 6 considers how to extend these results to multigraphs with degree constraints. We present our model of randomness for simple graphs or multigraphs with degree constraints in Section 6.1 and examine its relation to the well-known *configuration model* in Section 6.2. Theorem 7 is the analog of Theorems 1 and 2 for weighted graphs; it gives an exact expression for the total weight of multigraphs with one distinguished subgraph in the family $\mathcal{F}$, from which we derive the expected number of subgraphs belonging to $\mathcal{F}$ and the probability that there are t such subgraphs in Corollaries 6 and 7. To get more precise results, we have to consider properties of the set of weights. Weight sets with only finitely many nonzero elements are considered in Section 6.4 for n and m proportional. There we show that the only subgraphs that have a nonzero probability are trees and unicycles, derive the expected number of copies of a tree, and prove a Poisson limiting distribution for the number of occurrences of cycles of given length (Theorem 8). We first obtain a general result when the weights do not grow too fast (in a sense that we make precise), then derive the weighted equivalent of Theorems 5 and 6 (this is Theorem 9).

We consider further examples of weight sets in Section 7, most notably quickly increasing weights. This occurs for instance in the important case of *power-law multigraphs*, also known as *scale-free networks*, which are treated in Section 7.2. We sketch there a general approach to finding the expected number of copies of a sub-multigraph in a multigraph and give a complete answer for small cycles in Theorem 11. We also get results on the threshold for the appearance of trees in the case where $m/n \rightarrow 0$ in Section 7.1 (this is Theorem 10), and on multigraphs where the set of vertex degrees is periodic in Section 7.3.

2 Models and definitions

Most of the following definitions come from Erdős-Rényi [17] and Bollobás [8] for graphs, and from Flajolet, Knuth and Pittel [18], Janson *et al.* [27] or more recently [14] for multigraphs.

Graphs. A simple graph, or graph, G is a pair $(V(G), E(G))$, where $V(G)$ denotes the set of vertices carrying distinct labels, and $E(G)$ the set of edges. Each edge is an unoriented pair of distinct vertices, thus loops and multiple edges are forbidden. An (n, m) -graph is a simple graph with n vertices and m edges. The labels of the vertices are distinct integers. When no other constraint is added, the labeling is said to be *general*. When the vertices are labeled from 1 to n , the labeling is said to be *canonical*. Unless otherwise mentioned, the graphs considered have canonical labeling. The set of all simple graphs with canonical labeling is denoted by $\mathcal{SG}$.

Multigraphs. We define a multigraph as a graph-like object with labeled vertices, and labeled oriented edges, where loops and multiple edges are allowed. More formally, a multigraph G is a pair $(V(G), E(G))$, where $V(G)$ is the set of labeled vertices and $E(G)$ the set of labeled edges (the edge labels are independent from the vertex labels). Each edge is a triple (v, w, e) , where v, w are vertices, and e is the label of the edge (which is oriented from v to w). A loop is a triple (v, v, e) and a multiple edge is a set of at least two edges $\{(u, v, e_1), \dots, (u, v, e_k)\}$. An (n, m) -multigraph is a multigraph with n vertices and m edges. Again, the vertex labels are distinct integers, and the edge labels are distinct integers. When no other constraint is added, the labeling is said to be *general*. When the vertices are labeled from 1 to n and the edges from 1 to m , the labeling is said to be *canonical*. The set of all multigraphs with canonical labeling is denoted by $\mathcal{MG}$, and unless otherwise mentioned, the multigraphs considered have canonical labeling.

Notice that, although a given multigraph may have neither loops nor multiple edges, it would still not be a simple graph, as its edges are oriented and labeled. The orientation of the edges allows a very

simple description of a canonical (n, m) -multigraph G as a sequence $(v_1, \dots, v_{2m})$, where $v_i \in V(G)$ and $e_j = (v_{2j-1}, v_{2j}) \in E(G), \forall 1 \leq j \leq m$. This model of multigraphs can be found in the literature under the name of *quiver* or *multidigraph*, and appears notably in category theory and representation theory (see [12]). It introduces a bias in the enumeration of multigraphs and so differs from the more classical model of vertex-labeled, edge-labeled multigraphs since loops have only one possible orientation, while other edges have two.

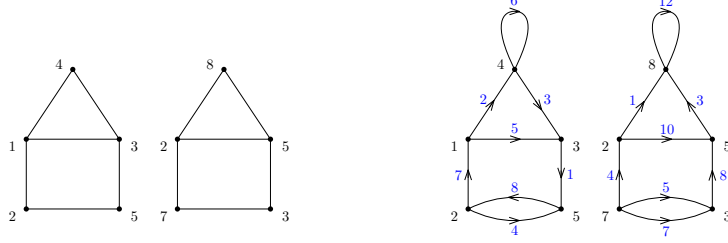


Figure 1: On the left, two isomorphic simple graphs (the first one with canonical labeling). On the right, two isomorphic multigraphs (the first one with canonical labeling).

The following definitions stand for both simple graphs and multigraphs. As such, they are stated for (multi)graphs, which can refer either to simple graphs or to multigraphs depending on the context.

Isomorphic (multi)graphs. Two (multi)graphs G and H with general labeling are *isomorphic* if there exists a bijection α between $V(G)$ and $V(H)$ that induces a bijection β between $E(G)$ and $E(H)$, *i.e.*,

$$(\text{for simple graphs}) \forall \{v, w\} \in E(G), \beta(\{v, w\}) = \{\alpha(v), \alpha(w)\} \in E(H);$$

$$(\text{for multigraphs}) \forall v, w \in V(G) : \{\beta(v, w, e) : (v, w, e) \in E(G)\} = \{(\alpha(v), \alpha(w), e') \in E(H)\}.$$

We also say that H is a G -(multi)graph. Notice that (multi)graph isomorphism is independent from labels. Given a (multi)graph family $\mathcal{F}$, H is a $\mathcal{F}$ -(multi)graph if it is isomorphic to an element of $\mathcal{F}$.

Subgraphs. A (multi)graph F is a *subgraph* of a (multi)graph G if $V(F) \subset V(G)$ and $E(F) \subset E(G)$. We then write $F \subset G$.

Given a (multi)graph family $\mathcal{F}$ and a (multi)graph G , an $\mathcal{F}$ -subgraph of G is a subgraph of G which is isomorphic to an element of $\mathcal{F}$. The number of $\mathcal{F}$ -subgraphs of G is denoted by $G[\mathcal{F}]$. When $\mathcal{F}$ is a singleton $\{H\}$, we simply write $G[H]$ instead of $G[\{H\}]$.

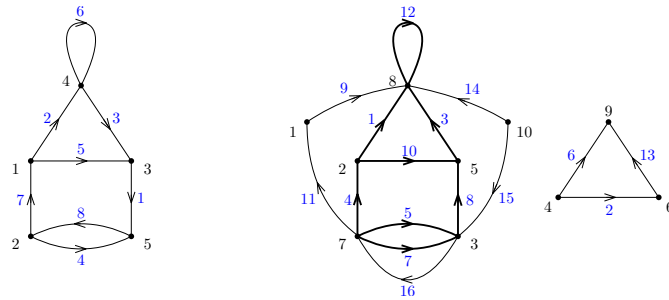


Figure 2: On the left, a multigraph appearing as a subgraph on the (non-connected) multigraph on the right.

Weighted (multi)graphs. A *weighted (multi)graph family* $\mathcal{F}_\omega$ is a (multi)graph family $\mathcal{F}$ equipped with a *weight* ω , which is a function from $\mathcal{F}$ to a given set. In this article, this set will be either the nonnegative real numbers (see Section 6.1 for weights related to the degree sequence of the multigraph, or Section 7.2 for a concrete example with weights following a power law) or the polynomials in the

variable u with real coefficients (which can be seen as formal weights, useful to track some parameters in (multi)graphs, see Section 4.4). The weight of a (multi)graph G is denoted by $\omega(G)$. The weight F_ω of the family $\mathcal{F}_\omega$ is defined as the sum of the weights of its elements. The *trivial weight* is the one that assigns to each (multi)graph the value 1. When weight ω is the trivial one, it is omitted in the notations, so F denotes the trivial weight of the family $\mathcal{F}$, equal to its cardinality. Hence, the results on weighted (multi)graphs extend the enumerative results. When no other weight is specified, the weight is assumed to be the trivial one.

Weights are usually associated to (multi)graph families which are equipped with a non-uniform distribution. For instance, while with the trivial weight each (multi)graph is equally likely, replacing the weight of a single (multi)graph H by 2 would make it twice as likely to be picked in the new weighted distribution.

We state some general results on the distribution of subgraphs in a weighted (multi)graph family in Section 3. In Section 5, we obtain more precise results for the trivial weight, *i.e.*, for the usual models of graphs and multigraphs. Finally, we consider the case where the weight of a (multi)graph depends on the degrees of its vertices in Section 6.

Counting and probabilities. For enumeration purposes, we will only consider families of (multi)graphs with canonical labeling. Given a family $\mathcal{F}$ of (multi)graphs, the set of elements of $\mathcal{F}$ having n vertices and m edges is denoted by $\mathcal{F}_{n,m}$, and its cardinality by $F_{n,m}$. For instance, $\text{SG}_{n,m}$ is equal to $\binom{n}{m}$, while $\text{MG}_{n,m}$ is equal to n^{2m} , as the labels and orientations of the edges induce a canonical representation of any multigraph as a sequence of $2m$ vertices. When a weight ω is specified, the weight of the set of all (multi)graphs in the (multi)graph family $\mathcal{F}_\omega$ which have n vertices and m edges is denoted by $F_{n,m,\omega}$.

In our model, a random (n, m) -(multi)graph is then a (multi)graph chosen uniformly at random from the set $\mathcal{SG}_{n,m}$ (resp. $\mathcal{MG}_{n,m}$):

$$\mathbb{P}(G \in \mathcal{SG}_{n,m}) = \frac{1}{\text{SG}_{n,m}}, \quad \mathbb{P}(G \in \mathcal{MG}_{n,m}) = \frac{1}{\text{MG}_{n,m}}.$$

This corresponds to the model $\mathcal{G}(n, m)$ introduced by Erdős and Rényi [17]. When a weight ω is specified, then a random (n, m, ω) -(multi)graph is an (n, m) -(multi)graph chosen with probability proportional to its weight:

$$\mathbb{P}(G \in \mathcal{SG}_{n,m,\omega}) = \frac{\omega(G)}{\text{SG}_{n,m,\omega}}, \quad \mathbb{P}(G \in \mathcal{MG}_{n,m,\omega}) = \frac{\omega(G)}{\text{MG}_{n,m,\omega}}.$$

Density and balance. Erdős and Rényi [17] observed first that the analysis of the number of F -subgraphs in a random (n, m) -graph is easier when the graph F is *strictly balanced*. We recall the definition of this property below.

The *density* of a (multi)graph G is defined as the ratio between the numbers of its edges and its vertices, and is denoted by

$$d(G) = \frac{m(G)}{n(G)}.$$

By convention, the empty (multi)graph has density 0. The *essential density* $d^*(G)$ of a (multi)graph G is the density of a subgraph of maximal density:

$$d^*(G) = \max_{H \subset G} d(H).$$

A (multi)graph F is *strictly balanced* if its density is greater than the density of all its strict subgraphs:

$$d(F) > \max_{H \subsetneq F} d(H).$$

It is *balanced* if

$$d(F) \geq \max_{H \subsetneq F} d(H),$$

and *barely balanced* if it is balanced, but not strictly balanced. Equivalently a (multi)graph is balanced if and only if $d(F) = d^*(F)$.

Lemma 1. *Given a connected (multi)graph F , let Pair_F denote the set of (multi)graph such that H is in Pair_F if and only if it is obtained by merging two distinct non-disjoint F -(multi)graphs F_1, F_2 , i.e.,*

$$V(H) = V(F_1) \cup V(F_2), \quad E(H) = E(F_1) \cup E(F_2),$$

and $V(F_1) \cap V(F_2)$ is non-empty. If F is strictly balanced, then the density of any (multi)graph from Pair_F is greater than the density of F .

3 Subgraphs in weighted graphs and multigraphs

In this section, we investigate the distribution of finite subgraphs in a random (n, m, ω) -(multi)graph, for a general weight function ω , and reduce the study of this distribution to the analysis of (n, m, ω) -(multi)graphs where an $\mathcal{F}$ -subgraph is distinguished, for a well chosen (multi)graph family $\mathcal{F}$. The weighted number of all (n, m, ω) -(multi)graphs is denoted by $\text{MG}_{n, m, \omega}$.

A (multi)graph G with a distinguished subgraph F can be represented as a pair (G, F) . Consider two weighted families $\mathcal{G}_{\omega_{\mathcal{G}}}$ and $\mathcal{F}_{\omega_{\mathcal{F}}}$; $\mathcal{G}_{\omega_{\mathcal{G}}}^{[\mathcal{F}_{\omega_{\mathcal{F}}}]}$ denotes the weighted family of all pairs (G, F) with $G \in \mathcal{G}$ and F being an $\mathcal{F}$ -subgraph of G . The weight of (G, F) is then implicitly defined as

$$\omega((G, F)) := \omega_{\mathcal{G}}(G) \omega_{\mathcal{F}}(F).$$

Therefore, the total weight $\text{MG}_{n, m, \omega_{\mathcal{G}}}^{[\mathcal{F}_{\omega_{\mathcal{F}}}]}$ of all (n, m) -(multi)graphs with a distinguished $\mathcal{F}_{\omega_{\mathcal{F}}}$ -subgraph is equal to

$$\text{MG}_{n, m, \omega_{\mathcal{G}}}^{[\mathcal{F}_{\omega_{\mathcal{F}}}]} := \sum_{\substack{(n, m)\text{-(multi)graph } G \\ \mathcal{F}\text{-subgraph } F \text{ of } G}} \omega_{\mathcal{G}}(G) \omega_{\mathcal{F}}(F).$$

When no weight function $\omega_{\mathcal{F}}$ is provided, the trivial weight 1 is used. This total weight plays a central role in the article.

In the rest of this section, we provide three propositions that reduce the study of the number $G[\mathcal{F}]$ of $\mathcal{F}$ -subgraphs in a random (n, m, ω) -(multi)graph G to the analysis of $\text{MG}_{n, m, \omega}^{[\mathcal{H}_{\omega_{\mathcal{H}}}]}$, for well chosen families $\mathcal{H}_{\omega_{\mathcal{H}}}$. The expected value of $G[\mathcal{F}]$ is computed in Proposition 1 using $\text{MG}_{n, m, \omega}^{[\mathcal{H}_{\omega_{\mathcal{H}}}]}$. Proposition 2 then gives an exact expression for the total weight of (n, m, ω) -(multi)graphs with exactly t $\mathcal{F}$ -subgraphs. Finally, Proposition 3 provides technical conditions on the weights so that the limit law of $G[\mathcal{F}]$ is a Poisson law for any strictly balanced (multi)graph F .

3.1 Expected number of subgraphs

Proposition 1 (Expected number of subgraphs, weights, simple graphs and multigraphs). *The expected number of $\mathcal{F}$ -subgraphs in a random (n, m, ω) -graph and a random (n, m, ω) -multigraph is*

$$\frac{\text{SG}_{n, m, \omega}^{[\mathcal{F}]} }{\text{SG}_{n, m, \omega}} \quad \text{and} \quad \frac{\text{MG}_{n, m, \omega}^{[\mathcal{F}]} }{\text{MG}_{n, m, \omega}}, \quad \text{respectively,}$$

where $\text{SG}_{n, m, \omega}$ is the weighted number of all (n, m, ω) -graphs, and $\text{SG}_{n, m, \omega}^{[\mathcal{F}]}$ is the weighted number of all (n, m, ω) -graphs with a distinguished $\mathcal{F}$ -subgraph (and similarly for multigraphs).

Proof. We give the proof for graphs, the proof for multigraphs being identical. By definition, the expected number of $\mathcal{F}$ -subgraphs in a random (n, m, ω) -graph is equal to

$$\frac{1}{\text{SG}_{n, m, \omega}} \sum_{\substack{G \in \text{SG}_{n, m} \\ F \in \mathcal{F}}} G[\mathcal{F}] \omega(G).$$

Since G contains $G[\mathcal{F}]$ $\mathcal{F}$ -subgraphs, there are $G[\mathcal{F}]$ pairs (G, F) in $\text{SG}_{n, m}^{[\mathcal{F}]}$, so the expected number is also equal to

$$\frac{1}{\text{SG}_{n, m, \omega}} \sum_{(G, F) \in \text{SG}_{n, m, \omega}^{[\mathcal{F}]}} \omega(G) = \frac{\text{SG}_{n, m, \omega}^{[\mathcal{F}]} }{\text{SG}_{n, m, \omega}}. \quad \square$$

3.2 Exact number of subgraphs and patchworks

When counting the number of (multi)graphs with n vertices, m edges and containing exactly t $\mathcal{F}$ -subgraphs, we run into a difficulty: those subgraphs may overlap. To describe these overlaps, we use the notion of an $\mathcal{F}$ -patchwork, defined below.

Patchworks for multigraphs. Given a family $\mathcal{F}$ of multigraphs, an $\mathcal{F}$ -patchwork is a finite set of distinct $\mathcal{F}$ -multigraphs (called *pieces*)

$$P = \{(V_1, E_1), \dots, (V_{|P|}, E_{|P|})\},$$

each carrying a general labeling (*i.e.*, the vertex or edge labels need not be consecutive integers starting at 1) which may share vertices and edges, such that if two vertices (from two distinct pieces) share the same labels, then they are merged and similarly, if two edges from distinct pieces share the same label, then they are merged. In particular, this implies that two edges sharing a label must connect the same two vertices with the same orientation. The vertices and edges of P are

$$V(P) = \bigcup_{i=1}^{|P|} V_i, \quad \text{and} \quad E(P) = \bigcup_{i=1}^{|P|} E_i, \quad (1)$$

and their cardinalities are denoted by $n(P)$ and $m(P)$. The size of P is the number of pieces and is denoted by $|P|$. This notion is illustrated in Figure 3.

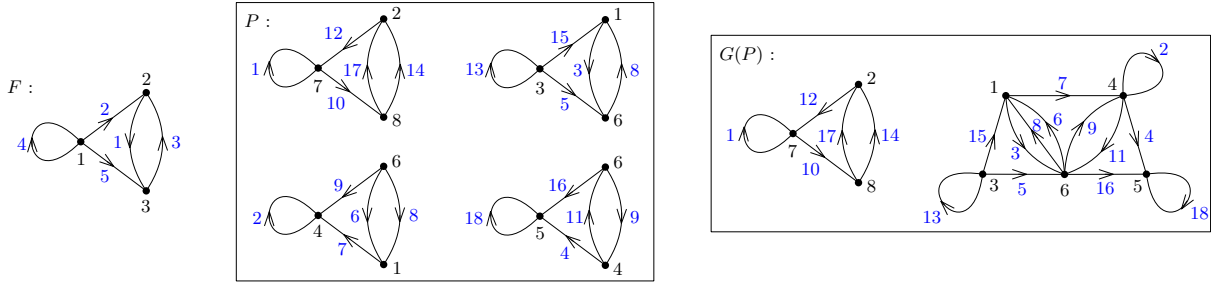


Figure 3: A multigraph F (left) and a F -patchwork P of size 4 (center) with its representation $G(P)$ as a multigraph (right).

A patchwork is called *disjointed* if it contains no pair of pieces sharing one vertex or more.

Patchworks for simple graphs. The notion of a patchwork can be naturally adapted to simple graphs. For a simple graph family $\mathcal{F}$, each piece of an $\mathcal{F}$ -patchwork is then an $\mathcal{F}$ -subgraph and pieces can share vertices and edges. As distinguished from multigraphs, edges connecting the same two vertices in distinct pieces are necessarily merged in the patchwork, so as to avoid multiple edges.

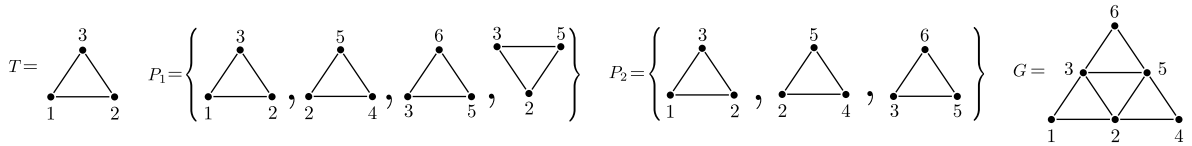


Figure 4: A graph T and two T -patchworks P and P' that have the same graph representation $G(P) = G(P') = G$.

Weights of patchworks and exact number of subgraphs. The family $\mathcal{P}^{\mathcal{F}}(u)$ denotes the set of $\mathcal{F}$ -patchworks equipped with the weight function

$$\omega(P) := u^{|P|},$$

so the weight of a patchwork P is the variable u raised to the power equaling the number of pieces of P .

Similarly, the family $\text{Disj } \mathcal{P}^{\mathcal{F}}$ denotes the subset of disjointed $\mathcal{F}$ -patchworks.

Proposition 2 (Total weight, simple and multigraphs). *The total weight $\text{SG}_{n,m,\omega,t}^{\mathcal{F}}$ of all (n, m, ω) -graphs that contain exactly t $\mathcal{F}$ -subgraphs is*

$$\text{SG}_{n,m,\omega,t}^{\mathcal{F}} := [u^t] \text{SG}_{n,m,\omega}^{[\mathcal{P}^{\mathcal{F}}(u-1)]}.$$

Likewise, for any multigraph family $\mathcal{F}$, the total weight $\text{MG}_{n,m,\omega,t}^{\mathcal{F}}$ of all (n, m, ω) -multigraphs that contain exactly t $\mathcal{F}$ -subgraphs is

$$\text{MG}_{n,m,\omega,t}^{\mathcal{F}} := [u^t] \text{MG}_{n,m,\omega}^{[\mathcal{P}^{\mathcal{F}}(u-1)]}.$$

Proof. We give the proof for graphs, as the proof for multigraphs is identical. It relies on the interpretation of the inclusion-exclusion principle commonly used in analytic combinatorics (see Flajolet and Sedgewick [19, Section III.7.4]). Recall that $G[\mathcal{F}]$ denotes the number of occurrences of $\mathcal{F}$ -subgraphs in the graph G . Let $\mathcal{SG}_{n,m,\omega}^{\mathcal{F}}(u)$ denote the graph family that contains all (n, m) -graphs, equipped with the weight ω_u defined as

$$\omega_u(G) := \omega(G)u^{G[\mathcal{F}]}.$$

Its weight is denoted by $\text{SG}_{n,m,\omega}^{\mathcal{F}}(u)$ and is equal to

$$\text{SG}_{n,m,\omega}^{\mathcal{F}}(u) = \sum_{G \in \mathcal{SG}_{n,m}} \omega(G)u^{G[\mathcal{F}]},$$

and the weight of the (n, m, ω) -graphs that contain exactly t $\mathcal{F}$ -subgraphs is then the t -th coefficient of $\text{SG}_{n,m,\omega}^{\mathcal{F}}(u)$

$$\text{SG}_{n,m,\omega,t}^{\mathcal{F}} = [u^t] \text{SG}_{n,m,\omega}^{\mathcal{F}}(u). \quad (2)$$

When we evaluate $\text{SG}_{n,m,\omega}^{\mathcal{F}}(u)$ at $u + 1$ instead of u and develop the powers of $u + 1$, we obtain

$$\text{SG}_{n,m,\omega}^{\mathcal{F}}(u + 1) = \sum_{G \in \mathcal{SG}_{n,m}} \sum_{j=0}^{G[\mathcal{F}]} \omega(G) \binom{G[\mathcal{F}]}{j} u^j.$$

The binomial coefficient $\binom{G[\mathcal{F}]}{j}$ is equal to the number of $\mathcal{F}$ -patchworks with j pieces contained in G , and u^j is the weight of each of those patchworks. Hence, $\text{SG}_{n,m,\omega}^{\mathcal{F}}(u + 1)$ is also the total weight of the (n, m, ω) -graphs where an $\mathcal{F}$ -patchwork is distinguished:

$$\text{SG}_{n,m,\omega}^{\mathcal{F}}(u + 1) = \sum_{(G,P) \in \mathcal{SG}_{n,m}^{[\mathcal{P}^{\mathcal{F}}(u)]}} \omega(G)u^{|P|} = \text{SG}_{n,m,\omega}^{[\text{Patch}_{\mathcal{F}}(u)]}.$$

Replacing u with $u - 1$, extracting the coefficient $[u^t]$, and injecting the result in Equation (2) finishes the proof. $\square$

3.3 Limit law of the number of subgraphs

Let c denote a real value or a polynomial in the variable u , and F a connected (multi)graph (in this section c will denote the weight of F). For simplification we will deviate from standard notations and use

$$e^{cF} := \text{Set}(c\{F\})$$

to denote the set of (multi)graphs where each component is isomorphic to F . The weight of such a (multi)graph is then defined as c^k , where k is the number of components. Similarly, given a (multi)graph H , we write

$$He^{cF} := \{H\} * \text{Set}(c\{F\}) \quad (3)$$

to denote the set of (multi)graphs where one distinguished connected component is isomorphic to H , while all the others are isomorphic to F . The weight of a (multi)graph from this family is defined as c^k , where k is the number of components isomorphic to F . By convention, when H is the empty (multi)graph (i.e., the graph that has no vertex and no edge), then He^{cF} is equal to e^{cF} .

Lemma 2. *Let F be a connected multigraph, and Pair_F denote the (finite) set of multigraphs obtained by merging two distinct isomorphic copies of F sharing at least one vertex. Consider a sequence $m := m(n)$ and recall that $\text{MG}_{n,m,\omega}^{[G]}$ denotes the cumulative weight of all (n, m, ω) -multigraphs with a distinguished G -subgraph. Assume that the following asymptotic relations hold, as n tends to infinity:*

- (Case $c = 2$) For all $H \in \text{Pair}_F$, we have $\text{MG}_{n,m,\omega}^{[He^{2F}]} = o(\text{MG}_{n,m,\omega}^{[F]})$.
- (Case $c = u - 1$) For t fixed, $[u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]} = \Theta(\text{MG}_{n,m,\omega}^{[F]})$.

Then the weight of all (n, m, ω) -multigraphs that contain exactly t F -subgraphs satisfies

$$\text{MG}_{n,m,\omega,t}^F \sim [u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]}.$$

Proof. Let $\text{Disj } \mathcal{MG}^F$ denote the set of multigraphs G such that the F -subgraphs of G are disjoint (i.e., share no vertex), and $\text{NonDisj } \mathcal{MG}^F$ the complementary set. As usual, let $\text{Disj } \text{MG}_{n,m,\omega,t}^F$ (resp. $\text{NonDisj } \text{MG}_{n,m,\omega,t}^F$) denote the weight of the (n, m, ω) -multigraphs from $\text{Disj } \mathcal{MG}^F$ (resp. $\text{NonDisj } \mathcal{MG}^F$) with exactly t F -subgraphs. Then their sum is equal to the weight of $\mathcal{MG}_{n,m,\omega,t}^F$

$$\text{MG}_{n,m,\omega,t}^F = \text{Disj } \text{MG}_{n,m,\omega,t}^F + \text{NonDisj } \text{MG}_{n,m,\omega,t}^F.$$

Applying the same inclusion-exclusion principle as in the proof of Theorem 2, we obtain

$$\text{Disj } \text{MG}_{n,m,\omega,t}^F = [u^t] \text{Disj } \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]}.$$

The right-hand side can be decomposed using the relation

$$\text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} = \text{Disj } \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} + \text{NonDisj } \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]}.$$

Combining the last three equations, we obtain

$$\text{MG}_{n,m,\omega,t}^F = [u^t] \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} - [u^t] \text{NonDisj } \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} + \text{NonDisj } \text{MG}_{n,m,\omega,t}^F. \quad (4)$$

We now consider the right-hand side, express the first term, and prove that the third term and the absolute value of second term are negligible.

First term. Recall that a disjointed F -patchwork is a set of mutually disjoint multigraphs which are isomorphic to F , so

$$[u^t] \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} = [u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]},$$

which is of order $\Theta(\text{MG}_{n,m,\omega}^{[F]})$.

Third term. The (n, m) -multigraphs with t F -subgraphs form a subset of all (n, m) -multigraphs, so

$$\text{NonDisj } \text{MG}_{n,m,\omega,t}^F \leq \text{NonDisj } \text{MG}_{n,m,\omega}.$$

Any multigraph with a distinguished Pair_F -subgraph is in $\text{NonDisj } \mathcal{MG}^F$, so

$$\text{NonDisj } \text{MG}_{n,m,\omega} \leq \text{MG}_{n,m,\omega}^{[\text{Pair}_F]} = \sum_{H \in \text{Pair}_F} \text{MG}_{n,m,\omega}^{[H]}.$$

Since $\text{MG}_{n,m,\omega}^{[He^{2F}]} = o(\text{MG}_{n,m,\omega}^{[F]})$, so is $\text{MG}_{n,m,\omega}^{[H]}$, and so the third term of the right-hand side of Equation (4) is negligible compared to the first one.

Second term. The absolute value of the second term is bounded by

$$\begin{aligned} \left| [u^t] \text{NonDisj } \text{MG}_{n,m,\omega}^{[\text{Disj } \mathcal{P}^F(u-1)]} \right| &\leq [u^t] \text{NonDisj } \text{MG}_{n,m,\omega}^{[e^{(u+1)F}]} \\ &\leq \sum_{t \geq 0} [u^t] \text{NonDisj } \text{MG}_{n,m,\omega}^{[e^{(u+1)F}]} 1^t = \text{NonDisj } \text{MG}_{n,m,\omega}^{[e^{2F}]}, \end{aligned}$$

which is the weight of the set of all multigraphs from $\text{NonDisj } \mathcal{MG}_{n,m}$ where a disjointed patchwork P is distinguished, and counted with a weight $2^{|P|}$. Let $\mathcal{K}$ be the finite set of graphs composed of a multigraph H from Pair_F and a (possibly empty) set of disjoint copies of F , each sharing at least one vertex with H . Pair_F being finite, and each element having at most $2V(F) - 1$ vertices, the set $\mathcal{K}$ is also finite. There exists at least one element of $\mathcal{K}$ in each graph from $\text{NonDisj } \mathcal{MG}$, thus the following holds:

$$\text{NonDisj } \text{MG}_{n,m,\omega}^{[e^{2F}]} \leq \sum_{K \in \mathcal{K}} \text{MG}_{n,m,\omega}^{[Ke^{2F}]}.$$

Since the sum is finite and each terms is of order $o(\text{MG}_{n,m,\omega}^{[F]})$, the second term of the right-hand side of Equation (4) is also negligible, which concludes the proof. $\square$

The following proposition provides simple conditions on the asymptotic weight of multigraphs with a distinguished subgraph to conclude that the number of F -subgraphs has a Poisson limit when F is strictly balanced.

Proposition 3 (Poisson distribution, strictly balanced, multigraphs). *Let $m := m(n)$ denote an integer sequence going to infinity with n , F be a strictly balanced multigraph, and ω a weight function. Let us assume that there exists a sequence of functions $\lambda_n : \mathcal{MG} \rightarrow \mathbb{R}$ satisfying the following assertions:*

- *The limit $\lambda(F) := \lim_{n \rightarrow \infty} \lambda_n(F)$ exists and is positive.*
- *For any constant c and any multigraph H , we have*

$$\text{MG}_{n,m,\omega}^{[He^{cF}]} \sim \lambda_n(H) e^{c\lambda_n(F)} \text{MG}_{n,m,\omega},$$

where we used the notation introduced in (3).

- *If $d(G) > d(F)$ then $\lambda(G) = 0$.*

Then the number of F -subgraphs in a random (n, m, ω) -multigraph follows a Poisson limit law with parameter $\lambda(F)$.

Proof. Since F is connected and strictly balanced, according to Lemma 1, any multigraph H in Pair_F is denser than F . The third hypothesis of the proposition implies that $\lambda_n(H)$ tends to 0, so

$$\text{MG}_{n,m,\omega}^{[He^{2F}]} \sim \lambda_n(H) e^{2\lambda_n(F)} \text{MG}_{n,m,\omega} = o(\text{MG}_{n,m,\omega}).$$

Thus, the first hypothesis of Lemma 2 is satisfied. In the last part of the proof, we will prove

$$[u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]} \sim \frac{\lambda_n(F)^t}{t!} e^{-\lambda_n(F)} \text{MG}_{n,m,\omega}. \quad (5)$$

This implies

$$[u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]} = \Theta(\text{MG}_{n,m,\omega}),$$

so the second hypothesis of Lemma 2 is satisfied. Applying this lemma, we then obtain

$$\text{MG}_{n,m,\omega,t}^F \sim [u^t] \text{MG}_{n,m,\omega}^{[e^{(u-1)F}]} \sim \frac{\lambda_n(F)^t}{t!} e^{-\lambda_n(F)} \text{MG}_{n,m,\omega},$$

so the number of F -subgraphs in a random (n, m, ω) -multigraph has a Poisson limit law of parameter $\lambda(F)$.

Proof of Equation (5). We follow the proof of [19, Theorem IX.1]. For any u in the unit disk, the function sequence

$$f_n(u) = \frac{\text{MG}_{n,m,\omega}^{[e^{(u-1)F}]}]{\text{MG}_{n,m,\omega}}$$

tends to $e^{(u-1)\lambda(F)}$, and

$$|f_n(u)| \leq \frac{\text{MG}_{n,m,\omega}^{[e^{(|u-1|)F}]}]{\text{MG}_{n,m,\omega}} \leq \frac{\text{MG}_{n,m,\omega}^{[e^{2F}]}]{\text{MG}_{n,m,\omega}}.$$

Since the right-hand side has a finite limit, it is bounded, so $|f_n(u)|$ is uniformly bounded on the unit disk. Then, according to Vitali's Theorem, the sequence of functions $f_n(u)$ converges uniformly in a neighborhood of 0 to $e^{(u-1)\lambda(F)}$, which implies Equation (5). $\square$

4 Generating functions for graphs and multigraphs

4.1 Analytic combinatorics

We recall here briefly the technique of translating combinatorial operations into equations for generating functions.

Symbolic Method. The book of Flajolet and Sedgewick [19] provides an excellent introduction to the techniques of analytic combinatorics. The main idea is to associate to any combinatorial family $\mathcal{A}$ of canonically labeled objects a generating function

$$A(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!},$$

where a_n denotes the number of objects of size n in $\mathcal{A}$. In the present article, to express the generating function of interesting combinatorial families, we apply the following dictionary to translate combinatorial relations between the families into analytic equations on their generating functions.

- **Disjoint union.** If $\mathcal{A} \cap \mathcal{B} = \emptyset$, and $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$, then $C(z) = A(z) + B(z)$.
- **Relabeled Cartesian product.** In the relabeled Cartesian product $\mathcal{C} = \mathcal{A} \times \mathcal{B}$, we consider all relabelings of the pairs (a, b) , with $a \in \mathcal{A}$ and $b \in \mathcal{B}$, so that each label, from 1 to the sum of the sizes of a and b , appears exactly once, *i.e.*, (a, b) is canonically labelled. We then have

$$C(z) = A(z)B(z).$$

- **Set.** The family of sets of objects from $\mathcal{A}$ (where the elements of a set are relabeled such that the set is canonically labeled after all) has generating function $e^{A(z)}$.

4.2 Multigraphs.

Since vertices and edges are labeled, we choose exponential generating functions with respect to both quantities (see Bergeron, Labelle and Leroux [6] or again [19]). Furthermore, a weight $1/2$ is assigned to each edge to take into account the orientation. The generating function of a multigraph family $\mathcal{F}$ is then

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

and the number of multigraphs with n vertices and m edges in the family $\mathcal{F}$ is

$$F_{n,m} = n! 2^m m! [z^n w^m] F(z, w).$$

For example, the generating function of the set $\mathcal{MG}$ of all multigraphs is

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

Weighted multigraphs. A *weighted* multigraph family is a family $\mathcal{F}_\omega$ where each multigraph G comes with a weight $\omega(G)$. For example, a multigraph family is a particular case of a weighted family, where each multigraph has weight 1. The weight of the weighted family $\mathcal{F}$ is

$$F_\omega = \sum_{G \in \mathcal{F}} \omega(G),$$

and its generating function is

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

The weight of $\mathcal{F}_{n,m,\omega}$, the set of multigraphs from $\mathcal{F}_\omega$ with n vertices and m edges, is then

$$F_{n,m,\omega} = n! 2^m m! [z^n w^m] F_\omega(z, w).$$

Subgraphs. In this article, we consider a uniform random multigraph G in $\mathcal{MG}_{n,m}$, where n and m are large integers, and aim at deriving the limit law for the number $G[\mathcal{F}]$ of $\mathcal{F}$ -subgraphs for any finite family $\mathcal{F}$. The generating function of all multigraphs where the number of $\mathcal{F}$ -subgraphs is marked (by the variable u) is denoted by

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

Therefore, the number of multigraphs with n vertices, m edges, and that contain exactly t $\mathcal{F}$ -multigraphs is

$$\text{MG}_{n,m,t}^{\mathcal{F}} = n! 2^m m! [z^n w^m u^t] \text{MG}^{\mathcal{F}}(z, w, u).$$

Notice that this formula corresponds to weighted multigraphs where each graph $G \in \mathcal{MG}$ has weight $\omega(G) = u^{G[\mathcal{F}]}$.

Multivariate vs. weighted series. The generating series $\text{MG}^{\mathcal{F}}(z, w, u)$, in three variables z , w and u , can also be seen as a bivariate series in z and w , for the weighted multigraph family where a graph $G \in \mathcal{MG}$ has weight $\omega(G) = u^{G[\mathcal{F}]}$.

In the following, we will sometimes exploit this equivalence between multivariate generating series and weighted generating series.

4.3 From multigraphs to graphs

Most of our results are first derived for multigraphs, because this model is better suited for generating function manipulations. However, the most common model in the graph and combinatorics communities is the simple graph model.

In a multigraph, a *loop* is an edge linking a vertex to itself. A *multiple edge* is a set of edges between the same two distinct vertices. A *graph* is a multigraph where the edges are neither labeled nor oriented and loops as well as multiple edges are forbidden. Thus, a simple graph on n vertices contains at most $\binom{n}{2}$ edges. Since the edges are non-labeled and non-oriented, we define the generating function of a simple graph family $\mathcal{F}$ as

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

For example, the generating function of all simple graphs is

$$\text{SG}(z, w) = \sum_{n \geq 0} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!}.$$

Lemma 3. *Consider a multigraph family $\mathcal{F}$ which is stable with respect to edge relabeling and change of orientation and containing neither loops nor multiple edges. We build the simple graph family $\mathcal{H}$ from $\mathcal{F}$ by removing the edge labels and orientations. Then, the generating functions of $\mathcal{F}$ and $\mathcal{H}$ are equal:*

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

Proof. There exist $2^{m(G)} m(G)!$ possible orientations and labelings for the edges of any simple graph G . The multigraphs obtained contain neither loops nor multiple edges. Conversely, any multigraph family being stable with respect to edge relabeling and change of orientation and containing neither loops nor multiple edges corresponds to a unique graph family. $\square$

Let $\mathcal{F}$ denote a simple graph family, and $\mathcal{H}$ the corresponding multigraph family, obtained by labeling and orienting the edges in all possible ways. In light of the previous lemma, counting graphs from $\mathcal{SG}_{n,m}$ with t copies from $\mathcal{F}$ is equivalent to counting multigraphs from $\mathcal{MG}_{n,m}$ with no loops and no multiple edges and containing t copies from $\mathcal{H}$. Removing the loops and multiple edges can be realized by forbidding subgraphs from the multigraph family that consists of the following two elements: the multigraph being a single loop and the multigraph which forms a double edge. There are actually four multigraphs comprising solely a double edge (by relabeling and re-orienting the edges), but as the notion of subgraph is independent of the labeling, only one is needed.

4.4 Patchworks

A patchwork P gives rise to a (multi)graph $G(P) := (V(P), E(P))$ (cf. (1)). However, as seen in Figure 4, this underlying (multi)graph is not enough to characterize a patchwork. Some information is lost, e.g., the number of pieces in the patchwork which turns out to be useful for enumeration purposes.

As was already the case for (multi)graphs, a patchwork P has a canonical labeling if its (multi)graph representation $G(P)$ has a canonical labeling. The set of all $\mathcal{F}$ -patchworks with canonical labeling is denoted by $\text{Patch}_{\mathcal{F}}$, and its generating function is

$$\begin{aligned} \text{Patch}_{\mathcal{F}}(z, w, u) &= \sum_{P \in \text{Patch}_{\mathcal{F}}} u^{|P|} \frac{w^{m(P)}}{2^{m(P)} m(P)!} \frac{z^{n(P)}}{n(P)!} \quad (\text{for multigraphs}); \\ \text{Patch}_{\mathcal{F}}(z, w, u) &= \sum_{P \in \text{Patch}_{\mathcal{F}}} u^{|P|} w^{m(P)} \frac{z^{n(P)}}{n(P)!} \quad (\text{for simple graphs}). \end{aligned}$$

Proposition 4 (Generating function, patchworks). *Given a (multi)graph family $\mathcal{F}$, closed by isomorphism and with generating function $F(z, w)$, the generating function of disjoint patchworks is*

$$\text{Disj } \mathcal{P}^{\mathcal{F}}(z, w, u) = e^{uF(z, w)}.$$

The generating function of all patchworks is

$$\text{Patch}_{\mathcal{F}}(z, w, u) = e^{uF(z, w)} (1 + Q^{\mathcal{F}}(z, w, u)),$$

where $Q^{\mathcal{F}}(z, w, u)$ denotes the generating function of patchworks that contain no isolated piece, i.e., each piece shares at least one vertex with another piece.

5 Number of small subgraphs in a random (multi)graph

We consider here simple graphs and multigraphs without degree constraints and examine what can be said about the number of occurrences of subgraphs in the whole graph. First, we derive the exact and asymptotic numbers of (multi)graphs with a distinguished subgraph belonging to a family $\mathcal{F}$ in Sections 5.1 and 5.2.

Next, we obtain an exact formula for the number of subgraphs of $\mathcal{F}$ in Section 5.3. Finally, we apply our results to a variety of problems, some old and some new ones, in Section 5.4: probability of occurrence of a “high-density” subgraph in a multigraph, densest subgraph in a simple graph, limiting Poisson distributions for the number of occurrences of a graph in a simple graph and of a strictly balanced graph in a multigraph.

5.1 Multigraphs with one distinguished subgraph

We first consider here how we can obtain the generating function of the class of all (multi)graphs with one distinguished subgraph, which belongs to the family $\mathcal{F}$.

Given two multigraph families $\mathcal{H}$ and $\mathcal{F}$, we denote by $\mathcal{H}^{[\mathcal{F}]}$ the set of multigraphs from $\mathcal{H}$ where exactly one copy of a multigraph from $\mathcal{F}$ is distinguished. If $\mathcal{F}$ is a weighted family, then the weight of $G \in \mathcal{H}^{[\mathcal{F}]}$ is equal to the weight of the distinguished copy from $\mathcal{F}$.

Theorem 1 (Distinguished, exact/asymptotics, multigraphs). *i) The number (or total weight) of multigraphs with n vertices, m edges, and that contain one distinguished subgraph from $\mathcal{F}$, is equal to*

$$\text{MG}_{n,m}^{[\mathcal{F}]} = n! 2^m m! [z^n w^m] F(z, w) e^z e^{n^2 w/2}. \quad (6)$$

ii) Assume that the generating function F of the family $\mathcal{F}$ is entire in both variables. Let D denote an open set containing the compact unit disk. Let n and $m = m(n)$ be two integers going to infinity such that $\frac{F(nz, 2mw/n^2)}{F(n, 2m/n^2)}$ converges uniformly on each compact $K \in D^2$ to an analytic function on D^2 , $L(z, w)$. Then the set $\text{MG}_{n,m}^{[\mathcal{F}]}$ of multigraphs containing n vertices, m edges, and with one copy from $\mathcal{F}$ distinguished, has weight

$$\text{MG}_{n,m}^{[\mathcal{F}]} \sim n^{2m} F\left(n, \frac{2m}{n^2}\right). \quad (7)$$

Remark. Although assuming that the function F is entire in both variables might seem restrictive, it is in fact quite natural: if the family $\mathcal{F}$ is finite, then the function $F(z, w)$ is a polynomial. An example of an infinite family with infinite radius of convergence in both variables would be the family of cycles with increasing labelling on edges and vertices (the first vertex is labelled by 1, then the edge between 1 and 2 is labelled by 1, and so on). But the family $\mathcal{F}$ of all labelled cycles does not satisfy the assumption of the theorem, as it leads to a generating function satisfying $F(z, 1) = \log \frac{1}{1-z}$, which has radius of convergence 1.

In Section 5.4, Theorem 1 will also be applied when $\mathcal{F}$ is the set of elements from a finite family, whose generating function, the exponential of a polynomial, is indeed entire.

Proof. A multigraph on n vertices, where one $\mathcal{F}$ -subgraph is distinguished, is a copy of a multigraph F from $\mathcal{F}$, a set of additional vertices, and a set of additional edges. The symbolic method (see [19]) translates this combinatorial description into an expression for the generating function, from which we extract the desired coefficient to get the exact expression of i). The asymptotics of ii) is then extracted using a saddle-point method. We now consider each of these points in detail.

Exact expression. To build a multigraph, we start with a distinguished copy of a subgraph from $\mathcal{F}$ and add first a set of vertices. So far, this family is described by the generating function

$$F(z, w)e^z.$$

Now, we add edges. If the multigraph we want to obtain at the end contains n vertices, then the number of possible edges is n^2 . With our convention (an edge has weight $\frac{1}{2}$), the generating function of one edge is $w/2$. Thus adding a set of edges among the n^2 possible ones translates into multiplying the generating function with $e^{n^2 w/2}$. We obtain

$$F(z, w)e^z e^{n^2 w/2}.$$

Finally, we extract the coefficients in z and w to consider only multigraphs with n vertices and m edges and obtain

$$n! 2^m m! [z^n w^m] F(z, w) e^z e^{n^2 w/2}.$$

Asymptotics. We now apply a bivariate saddle-point method to extract the asymptotics (see Section A1.2).

We start from the exact expression (6), to which we apply the following changes of variables

$$z \rightarrow nz, \quad \text{and} \quad w \rightarrow \frac{2m}{n^2} w,$$

and get

$$\text{MG}_{n,m}^{[\mathcal{F}]} = n^{2m} \frac{n!}{n^n} \frac{m!}{m^m} [z^n w^m] F\left(nz, \frac{2m}{n^2} w\right) e^{nz+mw}.$$

We rewrite the coefficient extractions as Cauchy integrals on circles of radii 1

$$\text{MG}_{n,m}^{[\mathcal{F}]} = n^{2m} \frac{n!}{n^n} \frac{m!}{m^m} \frac{1}{(2i\pi)^2} \oint_{|w|=1} \oint_{|z|=1} F\left(nz, \frac{2m}{n^2} w\right) e^{nz+mw} \frac{dz}{z^{n+1}} \frac{dw}{w^{m+1}}.$$

On this contour $\frac{F(nz, \frac{2m}{n^2} w)}{F(n, \frac{2m}{n^2})}$ converges uniformly to $L(z, w)$ which is analytic at $(1, 1)$, where its value is 1. Now there exists a sequence of analytic functions $(\epsilon_n(z, w))_{n \geq 0}$ converging uniformly to 0 such that

$$F\left(nz, \frac{2m}{n^2} w\right) = F\left(n, \frac{2m}{n^2}\right) (L(z, w) + \epsilon_n(z, w)).$$

We thus have

$$\text{MG}_{n,m}^{[\mathcal{F}]} \sim n^{2m} \frac{n!}{n^n} \frac{m!}{m^m} \frac{1}{(2i\pi)^2} \oint_{|w|=1} \oint_{|z|=1} F\left(n, \frac{2m}{n^2}\right) L(z, w) e^{nz+mw} \frac{dz}{z^{n+1}} \frac{dw}{w^{m+1}}.$$

A simple saddle-point method for large powers at $z = 1, w = 1$ (see Theorem VIII.8 from [19] or Lemma 9 in the appendix) and Stirling's formula then lead to

$$\text{MG}_{n,m}^{[\mathcal{F}]} \sim n^{2m} \frac{n!}{n^n} \frac{m!}{m^m} F\left(n, \frac{2m}{n^2}\right) L(1, 1) \frac{e^n}{\sqrt{2\pi n}} \frac{e^m}{\sqrt{2\pi m}} \sim n^{2m} F\left(n, \frac{2m}{n^2}\right). \quad \square$$

5.2 Simple graphs with one distinguished subgraph

We now turn to simple graphs, and establish a result parallel to Theorem 1, which is valid for multigraphs.

Theorem 2 (Distinguished, exact/asymptotics, simple). *i) The number of simple graphs with n vertices, m edges, and that contain one distinguished subgraph from $\mathcal{F}$, is*

$$\text{SG}_{n,m}^{[\mathcal{F}]} = n![z^n w^m] F \left(z, \frac{w}{1+w} \right) e^z (1+w)^{\binom{n}{2}}. \quad (8)$$

ii) Assume that the generating function F of the family $\mathcal{F}$ is entire in both variables. Let D denote an open set containing the compact unit disk. Let n and $m = m(n)$ be two integers going to infinity, such that $m = o(n^2)$ and that $\frac{F(nz, mw/\binom{n}{2})}{F(n, m/\binom{n}{2})}$ converges uniformly on each compact $K \in D^2$ to an analytic function on D^2 , $L(z, w)$. Then the set $\text{SG}_{n,m}^{[\mathcal{F}]}$ of simple graphs containing n vertices, m edges, and with one copy from $\mathcal{F}$ distinguished, has weight

$$\text{SG}_{n,m}^{[\mathcal{F}]} \sim \binom{\binom{n}{2}}{m} F \left(n, \frac{m}{\binom{n}{2}} \right), \quad (9)$$

where $\binom{\binom{n}{2}}{m}$ is the total number of graphs with n vertices and m edges.

Proof. The proof mimics the one we gave for multigraphs, and we concentrate our efforts on the points where it differs.

Exact expression. A simple graph with one distinguished $\mathcal{H}$ -graph is a copy F of a graph from $\mathcal{H}$, a set of additional vertices, and a set of additional edges. Those edges can link any pair of vertices, except those already linked in F .

We first prove part *i)* of the theorem for a family $\mathcal{H}$ that is composed of the simple graphs isomorphic to some (k, ℓ) -graph H ; we shall then extend it to a general family $\mathcal{F}$ of simple graphs.

For the family $\mathcal{H}$ defined above, let $|\text{Aut}(H)|$ be the number of automorphisms of the graph H . The number of H -graphs is $k!/|\text{Aut}(H)|$, and the generating function of $\mathcal{H}$ is

$$H(z, w) = \frac{1}{|\text{Aut}(H)|} w^\ell z^k.$$

The generating function of an H -graph and a set of isolated vertices is $H(z, w)e^z$. If we assume that there are n vertices, we extract the coefficient in z and obtain $n![z^n]H(z, w)e^z$.

Then each pair of the n vertices can be linked by an edge, except the pairs already linked in the H -graph: the number of edges that can be added is $\binom{n}{2} - \ell$. For each of those edges, we decide either to add it, or to not add it. Thus, the generating function of graphs on n vertices with a distinguished H -graph, additional vertices, and additional edges, is

$$n![z^n]H(z, w)e^z(1+w)^{\binom{n}{2}-\ell}.$$

Replacing $H(z, w)$ by its expression, this is equal to

$$n![z^n] \frac{1}{|\text{Aut}(H)|} w^\ell z^k e^z (1+w)^{\binom{n}{2}-\ell} = n![z^n] H \left(z, \frac{w}{1+w} \right) e^z (1+w)^{\binom{n}{2}}.$$

Finally, we fix the number of edges to m , and obtain the number of (n, m) -graphs where one H -graph is distinguished as

$$n![z^n w^m] H \left(z, \frac{w}{1+w} \right) e^z (1+w)^{\binom{n}{2}}.$$

If $\mathcal{F}$ is now a general graph family, its generating function is the sum of the generating functions of all H -graphs for which there exists at least one isomorphic copy of H in $\mathcal{F}$:

$$F(z, w) = \sum_{\text{there is an } H\text{-graph in } \mathcal{F}} H(z, w).$$

The number of (n, m) -graphs where one $\mathcal{F}$ -graph is distinguished is

$$\sum_{\text{there is a } H\text{-graph in } \mathcal{F}} n! [z^n w^m] H \left(z, \frac{w}{1+w} \right) e^z (1+w)^{\binom{n}{2}},$$

and by linearity this is simply

$$n! [z^n w^m] F \left(z, \frac{w}{1+w} \right) e^z (1+w)^{\binom{n}{2}}.$$

Asymptotics. We again apply a bivariate saddle-point method to extract the asymptotics. In expression (8), we apply the changes of variables

$$z \rightarrow nz, \quad w \rightarrow \frac{m}{\binom{n}{2}} w,$$

and obtain

$$\frac{n!}{n^n} \frac{\binom{n}{2}^m}{m^m} [z^n w^m] F \left(nz, \frac{mw/\binom{n}{2}}{1+mw/\binom{n}{2}} \right) e^{nz} \left(1 + \frac{m}{\binom{n}{2}} w \right)^{\binom{n}{2}}.$$

Again the function

$$\frac{F(nz, mw/\binom{n}{2})}{F(n, m/\binom{n}{2})}$$

converges uniformly to an analytic function $L(z, w)$, with $L(1, 1) = 1$. Furthermore, the function

$$\frac{F \left(nz, \frac{mw/\binom{n}{2}}{1+mw/\binom{n}{2}} \right)}{F(n, m/\binom{n}{2})}$$

converges uniformly to $L(z, w)$ as well, because $m = o(\binom{n}{2})$. Thus, there exists a sequence of analytic functions $(\epsilon_n(z, w))_{n \geq 0}$ converging uniformly to 0 such that

$$F \left(nz, \frac{mw/\binom{n}{2}}{1+mw/\binom{n}{2}} \right) = F \left(n, \frac{m}{\binom{n}{2}} \right) (L(z, w) + \epsilon_n(z, w)).$$

The number of (n, m) -graphs, with one $\mathcal{F}$ -subgraph distinguished, becomes

$$\frac{n!}{n^n} \frac{\binom{n}{2}^m}{m^m} F \left(n, \frac{m}{\binom{n}{2}} \right) [z^n w^m] (L(z, w) + \epsilon_n(z, w)) e^{nz} \left(1 + \frac{m}{\binom{n}{2}} w \right)^{\binom{n}{2}}.$$

By an exp-log argument (see Flajolet and Sedgewick [19, p. 29]), there exists a sequence of analytic functions $(\tilde{\epsilon}_n(w))_{n \geq 0}$ converging uniformly to 0 such that

$$\left(1 + \frac{m}{\binom{n}{2}} w \right)^{\binom{n}{2}} = e^{mw} (1 + \tilde{\epsilon}_n(w)),$$

so the number of desired graphs can be written as

$$\frac{n!}{n^n} \frac{\binom{n}{2}^m}{m^m} F \left(n, \frac{m}{\binom{n}{2}} \right) [z^n w^m] (L(z, w) + \epsilon_n(z, w)) (1 + \tilde{\epsilon}_n(w)) e^{nz} e^{mw}.$$

The end of the proof is parallel to the one for Theorem 1: the number of graphs is asymptotically equivalent to

$$\frac{n!}{n^n} \frac{\binom{n}{2}^m}{m^m} F \left(n, \frac{m}{\binom{n}{2}} \right) [z^n w^m] L(1, 1) e^{nz} e^{mw} \sim \binom{\binom{n}{2}}{m} F \left(n, \frac{m}{\binom{n}{2}} \right).$$

□

5.3 Exact enumeration of (multi)graphs with a given number of subgraphs

We now turn our attention to deriving an exact expression for the number

$$\text{MG}_{n,m,t}^{\mathcal{F}} = n!2^m m! [z^n w^m u^t] \text{MG}^{\mathcal{F}}(z, w, u)$$

of multigraphs with n vertices, m edges and containing exactly t $\mathcal{F}$ -subgraphs. The main difficulty is that these subgraphs may overlap. To describe the overlaps, we shall use the notion of *patchwork* defined in Section 3.2, see also Section 4.4.

Theorem 3 (Number of subgraphs, exact, multigraphs). *The number $\text{MG}_{n,m,t}^{\mathcal{F}}$ of (n, m) -multigraphs that contain exactly t $\mathcal{F}$ -subgraphs is*

$$\text{MG}_{n,m,t}^{\mathcal{F}} = n!2^m m! [z^n w^m u^t] \text{Patch}_{\mathcal{F}}(z, w, u-1) e^z e^{n^2 w/2}. \quad (10)$$

Proof. The weighted number of (n, m) -multigraphs with one $\mathcal{F}$ -patchwork distinguished is

$$\text{MG}_{n,m}^{\text{Patch}_{\mathcal{F}}(u)} = n!2^m m! [z^n w^m] \text{Patch}_{\mathcal{F}}(z, w, u) e^z e^{n^2 w/2}.$$

By Proposition 2 we have $\text{MG}_{n,m,t}^{\mathcal{F}} = [u^t] \text{MG}_{n,m}^{\text{Patch}_{\mathcal{F}}(u-1)}$ which completes the proof. $\square$

We defer extracting more detailed information in the case where the subgraphs are strictly balanced until Section 5.4.2, and consider below a result that closely parallels Theorem 3, but now for simple graphs.

Theorem 4 (Number of subgraphs, exact, simple). *The number $\text{SG}_{n,m,t}^{\mathcal{F}}$ of (n, m) -graphs that contain exactly t $\mathcal{F}$ -subgraphs is*

$$\text{SG}_{n,m,t}^{\mathcal{F}} = n! [z^n w^m u^t] \text{Patch}_{\mathcal{F}}\left(z, \frac{w}{1+w}, u-1\right) e^z (1+w)^{\binom{n}{2}}. \quad (11)$$

Proof. If we replace $F(z, w)$ by $\text{Patch}_{\mathcal{F}}(z, w, u)$ in Theorem 2 we obtain

$$\text{SG}_{n,m}^{\text{Patch}_{\mathcal{F}}(u)} = n! [z^n w^m] \text{Patch}_{\mathcal{F}}\left(z, \frac{w}{1+w}, u\right) e^z (1+w)^{\binom{n}{2}}.$$

Then, applying Proposition 2 completes the proof. $\square$

5.4 Asymptotics for the number of occurrences of subgraphs

The problem with the exact expressions derived in Theorems 3 and 4 is that we do not know in general the expression of the generating function of patchworks. Thus, Theorems 1 and 2 cannot be directly applied to extract the asymptotics of the coefficients. However, partial information is enough to address some interesting problems. In Section 5.4.1 below, we give an upper bound for the probability that a high-density subgraph appears in a multigraph of lower density, then derive a bound on the probability of appearance for the densest subgraph of a family $\mathcal{F}$ in a random simple graph. We then consider in Section 5.4.2 the appearance of strictly balanced subgraphs in multigraphs. Finally, we show in Section 5.4.3 that we can easily rederive the well-known Poisson limiting distribution for the number of subgraphs in a simple graph.

5.4.1 Dense subgraphs, and densest subgraph

We investigate here whether a subgraph of high density is likely to appear in a (multi)graph of smaller density.

As a first application of Theorem 1, we prove that subgraphs with high edge density are unlikely to appear in random multigraphs with small edge density. This result was first derived by Erdős and Rényi [17].

Corollary 1 (Probability for high-density subgraph, multigraphs). *Assume that $m = o(n^{2-1/d(F)})$. Then the probability for a uniform random multigraph from $\mathcal{MG}_{n,m}$ to contain one or more copies of the subgraph F tends to 0 at rate $\mathcal{O}(m n^{1/d(F)-2})^{m(F)}$, as $n \rightarrow \infty$.*

Proof. Let F be a given multigraph with generating function

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

By Proposition 1 the expected number of F -subgraphs in a random (n, m) -multigraph is $\mathbb{E}(G[F]) = \text{MG}_{n,m}^{[F]} / \text{MG}_{n,m} = \text{MG}_{n,m}^{[F]} / n^{2m}$. Applying Theorem 1, we obtain for $m = \mathcal{O}(n^\alpha)$ with $0 < \alpha < 2$

$$\mathbb{E}(G[F]) \sim F\left(n, \frac{2m}{n^2}\right) = \mathcal{O}(n^{n(F)-2m(F)+\alpha m(F)}).$$

Finally, this yields

$$\alpha < 2 - \frac{1}{d(F)} \Rightarrow \mathbb{E}(G[F]) \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \mathbb{P}(G[F] > 0) \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

This bound can be further improved by noticing that the relevant parameter is not the density of the multigraph F , but the density of its *densest* subgraph, *i.e.*, its *essential density* $d^*(F)$. We directly obtain the following:

Corollary 2 (Probability for high-essential-density subgraph, multigraphs). *Denote by d^* the density of a maximal densest subgraph of F , and consider a random (n, m) -multigraph G with $m = \mathcal{O}(n^\alpha)$ for some fixed $0 < \alpha < 2$. Then*

$$\begin{cases} G[F] = 0 \text{ asymptotically almost surely} & \text{if } \alpha < 2 - 1/d^*, \\ \mathbb{E}(G[F]) = \mathcal{O}(n^{m(F)(\alpha-2+1/d)}) & \text{if } \alpha \geq 2 - 1/d^*. \end{cases}$$

Proof. Let G denote a random (n, m) -multigraph. If H is a subgraph of F , then G contains F only if it contains H , so

$$\mathbb{P}(G[F] > 0) \leq \mathbb{P}(G[H] > 0).$$

Assume that H is a densest subgraph of F . By Corollary 1, H almost surely does not appear whenever $\alpha < 2 - 1/d(H) = 2 - 1/d^*$, and the same holds for F .

The bound on $\mathbb{E}(G[F])$ when $\alpha \geq 2 - 1/d^*$ follows directly from Theorem 1. $\square$

We now turn to simple graphs, for which we obtain a new proof of the following classic result of Erdős and Rényi [17] and Bollobás [8] by a simple application of Theorem 2.

Corollary 3 (Probability for high-essential-density subgraph, simple graphs). *Denote by d^* the density of a maximal densest subgraph of F , and consider a random (n, m) -graph G with $m = \mathcal{O}(n^\alpha)$ for some fixed $0 < \alpha < 2$. Then*

$$\begin{cases} G[F] = 0 \text{ asymptotically almost surely} & \text{if } \alpha < 2 - 1/d^*, \\ \mathbb{E}(G[F]) = \mathcal{O}(n^{m(F)(\alpha-2+1/d)}) & \text{if } \alpha \geq 2 - 1/d^*. \end{cases}$$

Proof. We use the results from Theorem 2 and the same arguments as in the proofs of Corollaries 1 and 2. $\square$

5.4.2 Strictly balanced subgraphs in multigraphs

We have already noted that extracting exact information on the number of subgraphs from the exact expression derived in Theorem 3 is hindered by the fact that we do not know an explicit expression for the generating function of patchworks. In the case of strictly balanced subgraphs we can nevertheless obtain some information on the distribution of the number of occurrences of these subgraphs, and show that it follows asymptotically a Poisson distribution.

For the notions we use in the context of patchworks recall Section 4.4. Moreover, since patchworks are weighted multigraphs, enriched with further information, we define its density $d(P)$ and its essential density $d^*(P)$ similarly as for multigraphs. If P is a F -patchwork and has essential density $d^*(F)$, we call it an *essential patchwork*. Recall also that a patchwork is disjointed if it contains no pair of pieces sharing one vertex or more. Given a multigraph F , let Pair_F denote the set of F -patchworks that contain exactly two pieces and are not disjointed. The following lemma is due to Erdős and Rényi [17]; we give again its proof for the sake of completeness.

Lemma 4. *A multigraph F is strictly balanced if and only if the density of any patchwork from Pair_F is greater than the density of F .*

Proof. We consider a patchwork P from Pair_F , denote by F_1 and F_2 its two pieces, copies of F , and by H the maximum common subgraph of F_1 and F_2 . Let also I denote the couple $(V(F_1) \setminus V(H), E(F_1) \setminus E(H))$. I is not a multigraph, because some of its edges link a vertex from I to a vertex from H . However, we can define its density as

$$d(I) = \frac{m(F_1) - m(H)}{n(F_1) - n(H)}.$$

The multigraph F_1 is obtained by adding the vertices and edges from I to H . Since F_1 is strictly balanced, we have $d(H) < d(F_1)$, which implies $d(I) > d(F_1)$. The patchwork P from Pair_F is obtained by adding the vertices and edges from I to F_2 . Thus the density of P is greater than the density of F . $\square$

In particular, Lemma 4 entails that only the disjointed patchworks of a strictly balanced multigraph are essential, and Corollary 1 implies that they are the most likely to appear.

We now consider the limit law of $G[F]$, the number of F -subgraphs in G , when F is a fixed strictly balanced multigraph, and G is drawn uniformly at random from $\mathcal{MG}_{n,m}$, for $m = m(n)$. According to Corollary 1, a random multigraph from $\mathcal{MG}_{n,m}$ with $m = o(n^{2-n(F)/m(F)})$ is unlikely to contain any copy of the multigraph F . We now focus on the critical case $m = \Theta(n^{2-n(F)/m(F)})$.

Theorem 5 (Poisson law, strictly-balanced, multigraphs). *Given a strictly balanced multigraph F , setting $\alpha = 2 - 1/d(F)$ and $m \sim cn^\alpha$, and considering a random multigraph G from $\mathcal{MG}_{n,m}$, then $G[F]$ follows in the limit a Poisson law of parameter*

$$\lambda(F) = \frac{c^{m(F)}}{m(F)! n(F)!}.$$

Proof. Our goal is to get the result by applying Proposition 3. Thus for a multigraph F let

$$F(z, w) = \frac{z^{n(F)}}{n(F)!} \frac{w^{m(F)}}{2^{m(F)} m(F)!} \quad (12)$$

and set $\lambda_n(F) = F(n, \frac{2m}{n^2})$. Then

$$\lambda(F) = \lim_{n \rightarrow \infty} F\left(n, \frac{2m}{n^2}\right) = \frac{c^{m(F)}}{m(F)! n(F)!}$$

which is clearly a positive number. Thus the first condition of Proposition 3 is satisfied.

To show that the second condition of Proposition 3 is satisfied, take an arbitrary multigraph G and some constant κ and observe that

$$\text{MG}_{n,m}^{[Ge^{\kappa F}]} = n!m!2^m [z^n w^m] G(z, w) e^{\kappa F(z, w)} e^z e^{n^2 w/2}$$

where $F(z, w)$ is the function given in (12) and $G(z, w)$ the analogue for the multigraph G . By Theorem 1 $\text{MG}_{n,m}^{[Ge^{\kappa F}]}$ can be expressed as

$$n!m!2^m [z^n w^m] G(z, w) e^{\kappa F(z, w)} e^z e^{n^2 w/2} \sim n^{2m} G\left(n, \frac{2m}{n^2}\right) \exp\left(\kappa F\left(n, \frac{2m}{n^2}\right)\right) = \text{MG}_{n,m} \lambda_n(G) e^{\kappa \lambda_n(F)}$$

which matches the requirement from Proposition 3. The only task which is left is showing that $\lambda_n(G)$ tends to zero if $d(G) > d(F)$. Looking at (12) again, we see that (noticing that $\alpha = 2 - \frac{1}{d(F)}$) the exponent of n in

$$G\left(n, \frac{2m}{n^2}\right) = \frac{c^{m(G)}}{n(G)! m(G)!} n^{n(G) - \frac{m(G)}{d(F)}}$$

is negative whenever $d(G) > d(F)$. Thus Proposition 3 is applicable and yields directly the assertion. $\square$

5.4.3 Distribution of the number of occurrences of a subgraph in a simple graph

We now turn to getting information on the distribution of the number of occurrences of a subgraph in a simple graph. The following theorem was first derived by Bollobás [8].

Theorem 6 (Poisson law, strictly-balanced, simple). *Let F denote a strictly balanced graph of density d , with ℓ edges and $|\text{Aut}(F)|$ automorphisms, and assume $m \sim cn^{2-1/d}$ for some positive constant c . Then the number of F -subgraphs in a random (n, m) -graph G follows a Poisson limit law with parameter $\lambda(F) = (2c)^\ell / |\text{Aut}(F)|$, i.e., for any nonnegative integer t ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(G[F] = t) = \frac{\lambda(F)^t}{t!} e^{-\lambda(F)}.$$

Proof. The result can be shown by a reasoning similar to the one presented in the proof of Theorem 5. As in the proof of Proposition 3 it can be shown that the main contribution in (11) at the threshold $\alpha = 2 - 1/d$ comes from disjointed patchworks. Thus an analogue of Proposition 3 for simple graphs is true as well.

Setting $\lambda_n(G) := G(n, m/\binom{n}{2})$, using $\lambda(F) = \lim_{n \rightarrow \infty} \lambda_n(F)$ and the asymptotics provided in Theorem 2 completes the proof. $\square$

5.4.4 Asymptotics of F -free (multi)graphs

Given a sub(multi)graph F , let an F -free (multi)graph be a (multi)graph without F -subgraph. The two preceding theorems allow in particular the enumeration of F -free (multi)graphs for any strictly balanced F at the threshold $m \sim cn^{1-1/d(F)}$.

Corollary 4 (F -free multigraphs). *Let F denote a strictly balanced multigraph of density d and assume $m \sim cn^{2-1/d}$ for some positive constant c . Then the number of F -free (n, m) -multigraphs is given by*

$$\text{MG}_{n,m,0}^F = e^{-\lambda(F)} \text{MG}_{n,m}, \quad \text{with } \lambda(F) := \frac{c^{m(F)}}{n(F)!m(F)!}.$$

Corollary 5 (F -free simple graphs). *Let F denote a strictly balanced graph of density d , and ℓ edges, and assume $m \sim cn^{2-1/d}$ for some positive constant c . Then the number of F -free (n, m) -graphs is given by*

$$\text{SG}_{n,m,0}^F = e^{-\lambda(F)} \text{SG}_{n,m}, \quad \text{with } \lambda(F) := \frac{(2c)^\ell}{|\text{Aut}(F)|}.$$

6 Small subgraphs in multigraphs weighted according to their degrees

We begin with a presentation of our model for multigraphs with degree constraints in Section 6.1, then making explicit its relation to the configuration model in Section 6.2. Section 6.3 then gives an exact expression for the total weight of multigraphs with one specified subgraph, from which we derive the expected number of subgraphs. In the following sections we study the asymptotics of this expected number. We first consider the case where only a finite number of weights is nonzero (Section 6.4), then turn to the case when the weights decrease “fast enough” with the degrees (Section 6.5). There we give a general result (Lemma 6) that holds under quite general conditions and from which we obtain various results on the appearance of some classes of subgraphs. The consideration of infinite families of weights that do not satisfy the conditions of Lemma 6 is deferred to Section 7.

6.1 Model for multigraphs weighted according to their degrees

In many applications, random multigraphs are not necessarily sampled according to the uniform distribution. Instead, we now want to be able to control the degree distribution of a random multigraph.

To that effect, we place weights on vertices of a given degree. More formally, to an infinite sequence $\delta = (\delta_0, \delta_1, \dots)$ of nonnegative weights, we associate the following generating function $\Delta(x)$

$$\Delta(x) = \sum_{d \geq 0} \delta_d \frac{x^d}{d!}.$$

Define $\mathcal{F}_\Delta$ as the multigraph family $\mathcal{F}$ equipped with the following weight: each vertex of degree d carries a weight δ_d , and the weight of a multigraph G is the product of the weights of its vertices

$$\omega_\Delta(G) = \prod_{v \in V(G)} \delta_{\deg(v)} = \prod_{d \geq 0} \delta_d^{n_d(G)},$$

where $n_d(G)$ stands for the number of vertices of G with degree d . The total weight of (n, m) -multigraphs in $\mathcal{F}_\Delta$ is denoted by $F_{n,m,\Delta}$, and the generating function of $\mathcal{F}_\Delta$ is

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

The Δ -multigraphs are the weighted multigraphs from $\mathcal{MG}_\Delta$, and the (n, m, Δ) -multigraphs are the Δ -multigraphs with n vertices and m edges. In particular, the total weight $\text{MG}_{n,m,\Delta}$ of the weighted multigraph family $\mathcal{MG}_{n,m,\Delta}$ is then

$$\text{MG}_{n,m,\Delta} = \sum_{G \in \mathcal{MG}_{n,m,\Delta}} \omega_\Delta(G) = \sum_{G \in \mathcal{MG}_{n,m,\Delta}} \prod_{v \in V(G)} \delta_{\deg(v)}.$$

In this weighted model, it will often prove convenient to work with an additional variable x for (labeled) half-edges of the multigraph. Indeed, since each edge is labeled and oriented, it is represented as a triple (u, v, e) where u and v denote its endpoints, and e its label. This edge can then be cut into two labeled half-edges, one hanging from u with label $2e - 1$, the other hanging from v with label $2e$. Cutting all the edges of a multigraph into half-edges, we obtain a bijection between the (n, m) -multigraphs and the set of n labeled vertices, each vertex of degree d having d dangling labeled half-edges around it, and so that the total number of half-edges sums up to $2m$. This construction, which is reminiscent of the *configuration model* of Bollobás [7] and of Wormald [48] (we explore further the link with this model in Section 6.2 below), yields the following lemma, first derived in [14]:

Lemma 5. *The total weight of $\mathcal{MG}_{n,m,\Delta}$ is equal to*

$$\text{MG}_{n,m,\Delta} = (2m)! [x^{2m}] \Delta(x)^n,$$

and the generating function of $\mathcal{MG}_\Delta$ is equal to

$$\text{MG}_\Delta(z, w) = \sum_{m \geq 0} (2m)! [x^{2m}] e^{z\Delta(x)} \frac{w^m}{2^m m!}.$$

Proof. The generating function of sets of labeled half-edges is $\Delta(x)$, where the set of size d has weight δ_d . Consider the combinatorial family composed of sets of vertices such that

- to each vertex is attached a set of labeled half-edges,
- the weight of the vertex is δ_d if d half-edges are attached to it,
- the weight of the object is the product of the weights of its vertices.

Then the generating function of this combinatorial family is $e^{z\Delta(x)}$. The sum of the weights of the objects from this family that contain $2m$ half-edges in total is then

$$(2m)! [x^{2m}] e^{z\Delta(x)}.$$

Such objects are in bijection with Δ -multigraphs containing m edges, according to the discussion before the statement of the lemma. The generating function of a set of m labeled edges is $\frac{w^m}{2^m m!}$. Summing over m , we obtain the generating function of Δ -multigraphs

$$\text{MG}_\Delta(z, w) = \sum_{m \geq 0} (2m)! [x^{2m}] e^{z\Delta(x)} \frac{w^m}{2^m m!}. \quad \square$$

6.2 Random Δ -multigraphs and the link with the configuration model.

Throughout this section, we compare the distribution of weighted multigraphs given by weighted Boltzmann samplers, denoted by $\mathcal{B}$, coming from analytic combinatorics, and by the purely probabilistic configuration model, denoted by $\mathcal{C}$.

In our weighted model, a random (n, m, Δ) -multigraph is no longer chosen uniformly at random from $\mathcal{MG}_{n,m}$, but rather according to its weight:

$$\mathbb{P}(G \in \mathcal{MG}_{n,m,\Delta}) = \frac{\omega_\Delta(G)}{\text{MG}_{n,m,\Delta}}.$$

Weighted Boltzmann sampling. One can effectively sample according to this distribution thanks to weighted Boltzmann samplers, introduced by Duchon *et al.* [16]. Let us give some classical properties of this random sampler. Define the set $V_\Delta := \{v_{(d)}\}_{d \geq 0}$, where $v_{(d)}$ is a single vertex with label 1, and d pending half-edges labeled from 1 to d . Let the variable x count the number of half-edges, its weighted generating series is thus $\Delta(x)$. Given a real positive value x , let $\Gamma_x^\mathcal{B} V_\Delta$ denote the random sampler for V_Δ that outputs $v_{(d)}$ with a probability equal to

$$\mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta = v_{(d)}) = \frac{\delta_d}{\Delta(x)} \frac{x^d}{d!}.$$

Define the weighted Boltzmann sampler $\Gamma_x^\mathcal{B} V_\Delta^n$, for the weighted family $\mathcal{MG}_{n,\Delta} := \bigcup_{m \geq 0} \mathcal{MG}_{n,m,\Delta}$ as follows:

1. Make n independent calls to the sampler $\Gamma_x^\mathcal{B} V_\Delta$ to produce n vertices labeled from 1 to n , with pending labeled half-edges,
2. If the sum of the degrees is odd, reject and repeat step 1. Otherwise, set $2m$ as the sum of the degrees.
3. Choose uniformly at random a relabeling of the half-edges such that the labels range from 1 to $2m$,
4. For all $j = 1 \dots m$, create a new directed edge labeled by j by connecting the half-edge with label $2j - 1$ to the half-edge with label $2j$.

Hence this sampler outputs a (n, m, Δ) -multigraph G with degree sequence $(d_1, \dots, d_n)$ with probability

$$\begin{aligned} \mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G) &= 0 \quad \text{if } \sum_i d_i \text{ is odd and} \\ \mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G) &= \mathbb{P}(v_1 = d_1, \dots, v_n = d_n) \cdot \mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G | v_1 = d_1, \dots, v_n = d_n) \cdot \mathbb{P}\left(\sum_i d_i \text{ even}\right)^{-1} \\ &= \left(\prod_{i=1}^n \frac{\delta_{d_i}}{\Delta(x)} \frac{x^{d_i}}{d_i!}\right) \cdot \binom{2m}{d_1, \dots, d_n}^{-1} \cdot \left(\frac{\Delta(x)^n + \Delta(-x)^n}{2\Delta(x)^n}\right)^{-1} \\ &= \frac{2}{\Delta(x)^n + \Delta(-x)^n} \frac{x^{2m}}{(2m)!} \omega_\Delta(G) \quad \text{if } \sum_i d_i \text{ is even.} \end{aligned}$$

When conditioned on having m edges, we obtain a weighted Boltzmann sampler for the weighted family $\mathcal{MG}_{n,m,\Delta}$ which outputs a (n, m, Δ) -multigraph G with degree sequence $(d_1, \dots, d_n)$ with probability

$$\begin{aligned} \mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G | \Gamma_x^\mathcal{B} V_\Delta^n \in \mathcal{MG}_{n,m,\Delta}) &= \frac{\mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G)}{\mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n \in \mathcal{MG}_{n,m,\Delta})} = \frac{\mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G)}{\sum_{G' \in \mathcal{MG}_{n,m,\Delta}} \mathbb{P}(\Gamma_x^\mathcal{B} V_\Delta^n = G')} \\ &= \left(\frac{2\omega_\Delta(G)}{\Delta(x)^n + \Delta(-x)^n} \frac{x^{2m}}{(2m)!}\right) \cdot \left(\sum_{G' \in \mathcal{MG}_{n,m,\Delta}} \frac{2\omega_\Delta(G')}{\Delta(x)^n + \Delta(-x)^n} \frac{x^{2m}}{(2m)!}\right)^{-1} \\ &= \frac{\omega_\Delta(G)}{\sum_{G' \in \mathcal{MG}_{n,m,\Delta}} \omega_\Delta(G')} = \frac{\omega_\Delta(G)}{\text{MG}_{n,m,\Delta}}, \end{aligned}$$

as desired.

In the unconditioned sampler the number of edges m is a random variable whose distribution depends on the parameter x . If we condition on a particular value m (*i.e.*, we start over if the number of edges obtained is not m), then the probability to sample a given (n, m, Δ) -multigraph is proportional to its weight. More precisely, one has

$$\begin{aligned}\mathbb{E}(\deg \Gamma_x^{\mathcal{B}} V_{\Delta}) &= \frac{x\Delta'(x)}{\Delta(x)}, \\ \mathbb{V}(\deg \Gamma_x^{\mathcal{B}} V_{\Delta}) &= \frac{x^2\Delta''(x)}{\Delta(x)} + \frac{x\Delta'(x)}{\Delta(x)} - \left(\frac{x\Delta'(x)}{\Delta(x)}\right)^2.\end{aligned}$$

In order to target m edges, we tune the value x such that the total expected number of half-edges is $2m$, so each vertex receives in average $2m/n$ half-edges. Because the mean degree of $\Gamma_x^{\mathcal{B}} V_{\Delta}$ is $\frac{x\Delta'(x)}{\Delta(x)}$, this corresponds to choosing for x the positive solution of the relation

$$\frac{x\Delta'(x)}{\Delta(x)} = \frac{2m}{n}. \quad (13)$$

Let $\text{Support}(\Delta)$ denote the set of indices corresponding to nonzero weights

$$\text{Support}(\Delta) = \{d \mid \delta_d > 0\}.$$

Observe that the equation (13) has a unique positive solution whenever

$$\min(\text{Support}(\Delta)) < \frac{2m}{n} < \max(\text{Support}(\Delta)),$$

because the average value of a random variable stays between its lowest and largest possible values, and because the function $\frac{x\Delta'(x)}{\Delta(x)}$ is increasing, since its derivative is positive as it equals the variance of $\deg \Gamma_x^{\mathcal{B}} V_{\Delta}$ up to a (positive) factor x .

Configuration model. The configuration model has been first introduced by Bollobás [7] to sample multigraphs according to a prescribed degree sequence; see also the presentation given by van der Hofstad [44, Ch. 7]. It was later extended to sample multigraphs with a given degree probabilistic distribution. Let us recall briefly the process.

Given a probability distribution π over the nonnegative integers, a configuration model sampler $\Gamma_{\pi}^{\mathcal{C}} \mathcal{MG}_n$ produces a multigraph $G = (V, E)$ with n vertices as follows:

1. Draw n independent random variables $X_1, \dots, X_n$ with distribution π .
2. Let $\sum X_i := M$: if M is odd, reject and repeat step 1, otherwise let $M := 2m$.
3. Let $V = \{v_1, \dots, v_n\}$, where v_i has X_i unordered pending half-edges, and thus $\deg(v_i) = X_i$.
4. Pick uniformly at random the labels of each half-edges from the set $\{1, \dots, 2m\}$.
5. For all $j = 1 \dots m$, pair the half-edge of label $2j-1$ with the half-edge of label $2j$, the new (directed) edge carries label j .

Proposition 5. *Suppose that, for a given real value x , the probability distribution π_x is defined by:*

$$\forall d \geq 0 : \quad \pi_x(d) = \frac{\delta_d x^d}{\Delta(x)d!}.$$

The distributions over $\mathcal{MG}_n$ under the configuration model with parameter π_x and the weighted Boltzmann model with parameter x are equal. More precisely, given a multigraph $G \in \mathcal{MG}_n$, where $\deg(v_i) = d_i, 1 \leq i \leq n$, we have:

$$\forall G \in \mathcal{MG}_n : \quad \mathbb{P}(\Gamma_{\pi_x}^{\mathcal{C}} \mathcal{MG}_n = G) = \mathbb{P}(\Gamma_x^{\mathcal{B}} V_{\Delta}^n = G).$$

Proof. Let G be a (n, m) -multigraph with vertices $v_1, \dots, v_n$, where $\deg(v_i) = d_i (1 \leq i \leq n)$, we have

$$\begin{aligned}\mathbb{P}(\Gamma_{\pi_x}^{\mathcal{C}} \mathcal{MG}_n = G) &= \mathbb{P}(X_1 = d_1, \dots, X_n = d_n) \cdot \mathbb{P}(\Gamma_{\pi_x}^{\mathcal{C}} \mathcal{MG}_n = G \mid X_i = d_i, \forall 1 \leq i \leq n) \cdot \mathbb{P}\left(\sum_i d_i \text{ even}\right)^{-1} \\ &= \prod_{i=1}^n \frac{\delta_{d_i} x^{d_i}}{\Delta(x)d_i!} \cdot \binom{2m}{d_1, \dots, d_n}^{-1} \cdot \left(\frac{\Delta(x)^n + \Delta(-x)^n}{2\Delta(x)^n}\right)^{-1} = \mathbb{P}(\Gamma_x^{\mathcal{B}} V_{\Delta}^n = G). \quad \square\end{aligned}$$

6.3 Exact number of subgraphs

We consider in this part how to extend the exact and asymptotic enumeration results of Section 5 to take into account the degree constraints we presented in Section 6.1. Theorem 7 gives an exact expression for the total weight of multigraphs with one distinguished subgraph in the family $\mathcal{F}$. This allows us to obtain the expected number of subgraphs belonging to $\mathcal{F}$ and the probability that there are t such subgraphs (Corollaries 6 and 7).

It will prove convenient to add variables to the generating function of multigraph families in order to mark the degrees. Each vertex of degree d is marked by the variable y_d , and the infinite sequence of variables $(y_1, y_2, \dots)$ is denoted by $\mathbf{y}$. The generating function of a multigraph family $\mathcal{F}$ is then

$$F(z, w, \mathbf{y}) = \sum_{G \in \mathcal{F}} \left(\prod_{v \in V(G)} y_{\deg(v)} \right) \frac{w^{m(G)}}{2^{m(G)} m(G)!} \frac{z^{n(G)}}{n(G)!}. \quad (14)$$

Likewise, the generating function of a simple graph family $\mathcal{F}$ is

$$F(z, w, \mathbf{y}) = \sum_{G \in \mathcal{F}} \left(\prod_{v \in V(G)} y_{\deg(v)} \right) w^{m(G)} \frac{z^{n(G)}}{n(G)!}.$$

Observe that the generating function of the (multi)graphs from $\mathcal{F}_\Delta$ is equal to the generating function of $\mathcal{F}$ where $\mathbf{y}$ is replaced by δ :

$$F_\Delta(z, w) = F(z, w, \delta).$$

The variables z and w are redundant in the generating function $F(z, w, \mathbf{y})$, because each vertex has a degree, and the sum of the degrees is twice the number of edges. Thus, given two variables a, b , we have formally

$$F(az, bw, \mathbf{y}) = F\left(z, w, \left(ab^{d/2}y_d\right)_{d \geq 0}\right) \quad (15)$$

Notice that the sum of degrees being even, it ensures that the exponents of b are non-negative integers, defining a proper formal power series.

Theorem 7 (Distinguished, total weight, multigraphs). *Define the operator $\bar{\partial}$ as*

$$\bar{\partial}f(x) = (f(x), f'(x), f''(x), \dots).$$

Given a multigraph family $\mathcal{F}$ with generating function $F(z, w, \mathbf{y})$ (with the convention stated in Equation (14)), the total weight $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$ of (n, m, Δ) -multigraphs where one $\mathcal{F}$ -subgraph is distinguished is equal to

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = n! 2^m m! [z^n w^m] \sum_{j \geq 0} (2j)! [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} \frac{w^j}{2^j j!}. \quad (16)$$

Proof. We combine the half-edges construction of Section 6.1 with the proof of Theorem 2. Let H denote a multigraph from $\mathcal{F}$, with k vertices and ℓ edges, and assume its automorphism group (both on vertices and edges) has size $|\text{Aut}(H)|$. Then the number of H -multigraphs is $2^\ell \ell! k! / |\text{Aut}(H)|$, and the generating function of the H -multigraphs is

$$H(z, w, \mathbf{y}) = \frac{1}{|\text{Aut}(H)|} \left(\prod_{v \in H} y_{\deg(v)} \right) w^k z^\ell.$$

An (n, m, Δ) -multigraph G where an H -subgraph is distinguished can be uniquely decomposed as an H -multigraph, a set of additional vertices, and a set of labeled half-edges, each linked to a vertex. The total number of half-edges must be even, say $2j$. The weight of any vertex linked to d half-edges and e edges in H is then δ_{d+e} , and the weight of G is equal to the product of the weights of the vertices. Equivalently, a vertex (without weight) of degree d in H is substituted by a vertex of arbitrary degree (which is at least d) with adequate weight and with d distinguished half-edges (discounted because already counted by the variable w in $H(z, w, \cdot)$).

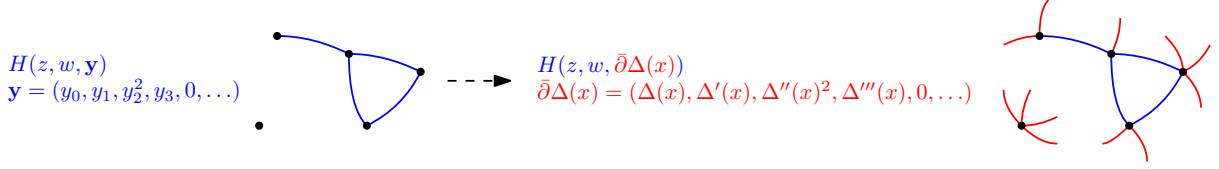


Figure 5: A distinguished graph H with some additional half-edges.

The generating function of Δ -multigraphs, where one H -subgraph is distinguished, is then

$$\text{MG}_{\Delta}^{[\{H\}]}(z, w) = \sum_{j \geq 0} (2j)! [x^{2j}] H(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} \frac{w^j}{2^j j!}.$$

Using the decomposition

$$F(z, w, \mathbf{y}) = \sum_{\text{there is an } H\text{-graph in } \mathcal{F}} H(z, w, \mathbf{y}).$$

and extracting the coefficient $n!2^m m! [z^n w^m]$ concludes the proof. $\square$

We easily derive from Theorem 7 information on the number of occurrences of subgraphs from $\mathcal{F}$.

Corollary 6 (Expected number of subgraphs, degree constraints, multigraphs). *Given a multigraph family $\mathcal{F}$, the expected number of $\mathcal{F}$ -subgraphs in a random (n, m, Δ) -multigraph is*

$$\frac{\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}}{\text{MG}_{n,m,\Delta}} = \frac{n!2^m m! [z^n w^m] \sum_{j \geq 0} (2j)! [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} \frac{w^j}{2^j j!}}{(2m)! [x^{2m}] \Delta(x)^n}.$$

Proof. According to Proposition 1, the average number of $\mathcal{F}$ -subgraphs in a random (n, m, Δ) -multigraph is

$$\frac{\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}}{\text{MG}_{n,m,\Delta}}.$$

The expression of $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$ has been derived in Theorem 7, while Lemma 5 provides the expression of $\text{MG}_{n,m,\Delta}$. $\square$

Corollary 6 is useful to find the threshold for the emergence of a given multigraph F as a subgraph in a random (n, m, Δ) -multigraph, *i.e.*, the rate at which $m = m(n)$ should go to infinity with n so that a random (n, m, Δ) -multigraph typically contains a bounded, but positive number of F -subgraphs. In order to derive more information on the limit law of the number of F -subgraphs in a random (n, m, Δ) -multigraph, we extend the notion of patchworks (see Sections 3.2 and 4.4) to Δ -multigraphs. To take into account the degrees, the generating function of patchworks of a multigraph family $\mathcal{F}$ becomes

$$\text{Patch}_{\mathcal{F}}(z, w, \mathbf{y}, u) = \sum_{P \in \mathcal{P}^{\mathcal{F}}} u^{|P|} \left(\prod_{v \in V(P)} y_{\deg(v)} \right) \frac{w^{m(P)} z^{n(P)}}{2^{m(P)} m(P)! n(P)!},$$

where $\deg(v)$ denotes the degree of the vertex v in the patchwork P . We also denote by $\mathcal{P}^{\mathcal{F}}(u)$ the set of $\mathcal{F}$ -patchworks, where the weight of a patchwork P is defined as

$$\omega(P) = u^{|P|}.$$

Corollary 7 (Probability for number of subgraphs, degree constraints, multigraphs). *Given a multigraph family $\mathcal{F}$, the probability for a random (n, m, Δ) -multigraph to contain exactly t $\mathcal{F}$ -subgraphs is equal to*

$$\frac{[u^t] \text{MG}_{n,m,\Delta}^{[\mathcal{P}^{\mathcal{F}}(u-1)]}}{\text{MG}_{n,m,\Delta}},$$

where the family $\mathcal{P}^{\mathcal{F}}(u-1)$ is defined just before the corollary, the term $\text{MG}_{n,m,\Delta}^{[\mathcal{P}^{\mathcal{F}}(u-1)]}$ is expressed in Theorem 7, and the value of $\text{MG}_{n,m,\Delta}$ is provided by Lemma 5.

Proof. This is a direct application of Proposition 2 to multigraphs weighted according to their degrees. $\square$

Corollaries 6 and 7 reduce the study of the number of $\mathcal{F}$ -subgraphs in a random (n, m, Δ) -multigraph to the asymptotic analysis of $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$, the total weight of (n, m, Δ) -multigraphs where one $\mathcal{F}$ -subgraph is distinguished. This asymptotic analysis widely depends on the weight vector δ and the analytic properties of its generating function $\Delta(x)$. In the rest of Section 6 we start with the cases in which either only finitely many weights are nonzero or the weights do not increase too fast (this is Lemma 6 followed by its applications). In Section 7 some cases in which the conditions of Lemma 6 fail are discussed.

6.4 Finite set of nonzero weights

When there is only a finite number of nonzero weights, $\Delta(x)$ is a polynomial and the Δ -multigraphs have bounded degrees. More precisely, since the sum of the degrees is twice the number of edges, for any (n, m, Δ) -multigraph with nonzero weight,

$$n \min(\text{Support}(\Delta)) \leq 2m \leq n \max(\text{Support}(\Delta)),$$

which implies that $2m/n$ is bounded and in $[\min(\text{Support}(\Delta)), \max(\text{Support}(\Delta))]$. When $2m/n$ reaches one of those bounds, then the multigraph is regular (all its vertices have the same degree). In particular, if there is only one nonzero weight, *i.e.*, $\Delta(x)$ is a monomial, then $2m/n$ is equal to the corresponding degree and the multigraph is regular. The analysis of subgraphs in regular multigraphs has been achieved previously (we refer the reader to Bollobás [7] and to McKay, Wormald and Wysocka [35]), so we omit this case here and consider $2m/n$ having a limit in $]\min(\text{Support}(\Delta)), \max(\text{Support}(\Delta))]$.

Lemma 6. *Consider a weight generating function $\Delta(x)$ of infinite radius of convergence and support $\text{Support}(\Delta)$. Let the integers n and $m := m(n)$ tend to infinity and assume*

(C1)

$$\inf_{n \geq 0} \frac{2m}{n} > \min(\text{Support}(\Delta)), \quad \sup_{n \geq 0} \frac{2m}{n} < \max(\text{Support}(\Delta)).$$

Let $\chi := \chi(n)$ denote the unique positive solution of

$$\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \frac{2m}{n}.$$

Consider a multigraph family $\mathcal{F}$ with generating function $F(z, w, \mathbf{y})$, and let $\text{MG}_{n,m,\Delta}$ denote the total weight of (n, m, Δ) -multigraphs. Moreover, set

$$G_n(z, x, t) = F\left(nz, \frac{1}{2mt^2}, \left(\frac{(x\chi)^d \Delta^{(d)}(x\chi)}{\Delta(x\chi)}\right)_{d \geq 0}\right) \quad (17)$$

and for a given $\varepsilon > 0$,

$$L_n(z, x, t) = \begin{cases} \frac{G_n(z, x, t)}{G_n(1, 1, 1)} & \text{if } t \geq \varepsilon, \\ \frac{G_n(z, x, \varepsilon)}{G_n(1, 1, 1)} & \text{if } 0 \leq t \leq \varepsilon. \end{cases}$$

Assume moreover:

(C2) $L_n(z, x, t)$ converges to some limit $L(z, x, t)$, as $n \rightarrow \infty$ and uniformly for $|z| = 1$, $|x| = 1$ such that $\arg x \leq \varepsilon$ and t in any compact sub-interval of $[0, \infty)$. Moreover, $L_n(z, x, t)$ is uniformly bounded for z and t as above and $|x| = 1$.

(C3) For all $0 < \varepsilon < \frac{\pi}{2}$, we have

$$\int_{-\pi}^{\pi} \frac{\Delta(\chi e^{i\theta})^n}{e^{2mi\theta}} d\theta \sim \int_{-\varepsilon}^{\varepsilon} \frac{\Delta(\chi e^{i\theta})^n}{e^{2mi\theta}} d\theta,$$

as $n \rightarrow \infty$.

Then the number $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$ of (n, m, Δ) -multigraphs with one distinguished $\mathcal{F}$ -subgraph is asymptotically equal to $\text{MG}_{n,m,\Delta} \cdot G_n(1, 1, 1)$.

Remark. The conditions of this lemma may appear to be very strong requirements. Nevertheless there are some “natural” assumptions on the weight sequence under which the conditions of Lemma 6 are satisfied: Assume that the weights are bounded in the sense

$$\delta_j \leq K^j \quad (18)$$

for some positive constant K and that we do not have any periodicities, *i.e.*, the set of allowed degrees is not a subset of some lattice $r + \ell\mathbb{N}$ with integers $r \geq 0$ and $\ell \geq 2$. The bound (18) guarantees that $\Delta(x)$ has infinite radius of convergence. If we additionally assume that $\ell\mathbb{N} \subseteq \text{Support}(\Delta)$ for some positive integer ℓ , then $\Delta(x)$ is a linear combination of terms of the form $e^{Kx+2\pi i/\ell}$ plus smaller order terms and thus behaves essentially like e^{Kx} . This implies that locally around $x = 1$, all terms of the form $x^d \Delta^{(d)}(x\chi)/\Delta(x\chi)$ converge uniformly even if χ tends to infinity. Consequently, Condition (C2) is satisfied. To show Condition (C3), we first observe that the aperiodicity condition guarantees that the only region that matters is x near 1. That means that we need the uniform convergence of Condition (C2) only for x near 1. Furthermore, note that e^{Kx} is Hayman-admissible (*cf.* [26]), which essentially means that the saddle point method directly applies. Hayman’s proof uses certain estimates of Cauchy-like integrals. One of them straight-forwardly implies Condition (C3), if there is no periodicity.

In case of periodicity, we can still apply Lemma 6 in a slightly modified form. Then there are more, say s , crucial regions, one near 1 and the others distributed along a regular s -gon. Conditions (C2) and (C3) have to be extended accordingly. Cases where periodicity occurs will be discussed in Section 7.

Proof. Recall Equation (16):

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = n!2^m m! [z^n w^m] \sum_{j \geq 0} (2j)! [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} \frac{w^j}{2^j j!}.$$

Using the classical formula (moments of even order of the standard normal distribution)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^{2j} e^{-t^2/2} dt = \sqrt{\frac{2}{\pi}} \int_0^{\infty} t^{2j} e^{-t^2/2} dt = \frac{(2j)!}{2^j j!},$$

we rewrite the sum on the right-hand side of (16) as

$$\begin{aligned} \sum_{j \geq 0} \frac{(2j)!}{2^j j!} [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} w^j &= \sum_{j \geq 0} \sqrt{\frac{2}{\pi}} \int_0^{\infty} w^j t^{2j} e^{-t^2/2} dt [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) e^{z\Delta(x)} \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \left(\sum_{j \geq 0} [x^{2j}] F(z, w, \bar{\partial}\Delta(x)) (t\sqrt{w})^{2j} \right) e^{z\Delta(t\sqrt{w})} e^{-t^2/2} dt \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} F(z, w, (\bar{\partial}\Delta)(t\sqrt{w})) e^{z\Delta(t\sqrt{w})} e^{-t^2/2} dt, \end{aligned}$$

where interchanging the sum and the integral is licit, because for sufficiently small x the series converges uniformly. For the evaluation of the sum in the last step we used the fact that the series F has only even powers of x . This follows from the fact that x counts the number of half-edges in the graph, which is always even (*cf.* Theorem 7). To obtain $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$, we extract the coefficient $[z^n w^m]$ of the previous expression and multiply by $n!2^m m!$:

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = n! \frac{2^{m+\frac{1}{2}}}{\sqrt{\pi}} m! [z^n w^m] \int_0^{\infty} F(z, w, (\bar{\partial}\Delta)(t\sqrt{w})) e^{z\Delta(t\sqrt{w})} e^{-t^2/2} dt. \quad (19)$$

Evaluation of the coefficients of z^n and w^m is based on the saddle-point method. This leads to

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = \frac{n!2^{m+\frac{1}{2}} m!}{\sqrt{\pi}(2\pi i)^2} \int_0^{\infty} \int_{|w|=g(t)} \int_{|z|=n/\Delta(t\sqrt{w})} \frac{F(z, w, (\bar{\partial}\Delta)(t\sqrt{w})) e^{z\Delta(t\sqrt{w})}}{z^{n+1} w^{m+1}} dz dw e^{-t^2/2} dt,$$

where

$$g(t) = \begin{cases} \chi^2/t^2 & \text{if } t \geq \varepsilon, \\ \chi^2/\varepsilon^2 & \text{if } 0 \leq t \leq \varepsilon, \end{cases}$$

and by Condition (C1) the saddle-point χ remains in a compact subinterval of $]0, +\infty[$. To simplify we apply successively the following changes of variables:

$$z \rightarrow \frac{nz}{\Delta(\sqrt{wt})}, \quad w = (x\chi/t)^2, \quad t \rightarrow \sqrt{2mt}.$$

The expression becomes

$$\begin{aligned} \frac{\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}}{G(1,1,1)} &= \frac{n!2^{m+\frac{1}{2}}m!(2m)^{m+\frac{1}{2}}}{n^n\chi^{2m}\sqrt{\pi}(2\pi i)^2} \left(\int_{\varepsilon}^{\infty} \int_{|x|=1} \int_{|z|=1} \frac{L_n(z,x,t)e^{nz}\Delta(x\chi)^n}{z^{n+1}x^{2m+1}} dz dx t^{2m}e^{-mt^2} dt \right. \\ &\quad \left. + \varepsilon^{2m} \int_0^{\varepsilon} \int_{|x|=1} \int_{|z|=1} \frac{L_n(z,x,t)e^{nz}\Delta(x\chi)^n}{z^{n+1}x^{2m+1}} dz dx e^{-mt^2} dt \right). \end{aligned} \quad (20)$$

Since $L_n(z,x,t)$ is uniformly bounded, by Condition (C3) truncating the circle $|x| = 1$ to $|x| = 1, \arg(x) \leq \varepsilon$ in (20) causes an asymptotically negligible error. Next, we start with the evaluation of the inner integral. Note that $\arg(x) \leq \varepsilon$ and thus Condition (C2) applies: $L_n(z,x,t)$ converges uniformly to $L(z,x,t)$. Thus we obtain

$$\frac{1}{2\pi i} \int_{|z|=1} \frac{L_n(z,x,t)e^{nz}}{z^{n+1}} dz \sim \frac{1}{2\pi i} \int_{|z|=1} \frac{L(z,x,t)e^{nz}}{z^{n+1}} dz \sim L(1,x,t)[z^n]e^{nz} = L(1,x,t)\frac{n^n}{n!}, \quad (21)$$

where we used Lemma 9 in the last asymptotic estimate.

The next integral is

$$\frac{1}{2\pi i} \int_{|x|=1} \frac{L(1,x,t)\Delta(x\chi)^n}{x^{2m+1}} dx \sim \frac{1}{2\pi i} \int_{|x|=1, \arg(x) \leq \varepsilon} \frac{L(1,x,t)\Delta(x\chi)^n}{x^{2m+1}} dx \sim L(1,1,t)[x^{2m}]\Delta(x\chi)^n \quad (22)$$

which follows from the analyticity of $L(1,x,t)$, Condition (C3) and the Taylor expansion $L(1,x,t) \sim L(1,1,t) + L_x(1,1,t)(x-1)$ together with $x-1 = \mathcal{O}(\varepsilon)$.

For the last integral we use the Laplace method (see [19, Theorem B.7]). First observe that the function $t \mapsto L(1,1,t)$ is positive and decreasing on the positive real line and that $t^{2m}e^{-mt^2}$ attains its maximum at $t = 1$. Thus, [19, Theorem B.7] tells us that

$$\begin{aligned} \int_{\varepsilon}^{\infty} L(1,1,t)t^{2m}e^{-mt^2} dt &\sim L(1,1,1) \int_{\varepsilon}^{\infty} t^{2m}e^{-mt^2} dt \\ &\sim L(1,1,1) \int_0^{\infty} t^{2m}e^{-mt^2} dt = L(1,1,1) \sqrt{\frac{\pi}{2}} \frac{1}{(2m)^{m+\frac{1}{2}}} \frac{(2m)!}{2^m m!}, \end{aligned} \quad (23)$$

as $m \rightarrow \infty$ and for $\varepsilon < 1$. Finally, we have to show that

$$\int_0^{\varepsilon} L(1,1,t)t^{2m}e^{-mt^2} dt$$

is negligible. Note that $L_n(1,1,t)$ is constant on the integration interval, and so its limit $L(1,1,t)$ is constant as well. The exponential is bounded by 1 and thus the integral is $\mathcal{O}(\varepsilon^{2m})$. Since $\int_0^{\infty} t^{2m}e^{-mt^2} dt = \Theta(\sqrt{m}e^{-m})$, ε^{2m} is negligible for $\varepsilon < 1/\sqrt{e}$.

Collecting the contributions from (21)–(23) and inserting them into (20) gives finally

$$\frac{\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}}{G(1,1,1)} \sim (2m)! [x^{2m}] \Delta(x)^n = \text{MG}_{n,m,\Delta}$$

which concludes the proof. $\square$

In particular, if all multigraphs from $\mathcal{F}$ have maximal degree d greater than $\max(\text{Support}(\Delta(x)))$, then $\Delta^{(d)}(\chi)$ vanishes, so the lemma claims that $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$ tends to 0. Indeed, $\text{MG}_{n,m,\Delta}^{[\mathcal{F}]}$ vanishes, because the maximum degree of a subgraph of a multigraph G is never greater than the maximum degree of G .

Theorem 8 (Finite set of weights, trees and cycles). *Consider a polynomial $\Delta(x)$, integers n and $m := m(n)$ going to infinity such that $2m/n$ has a limit in $]\min(\text{Support}(\Delta(x))), \max(\text{Support}(\Delta(x)))[,$ the unique positive solution χ of the equation*

$$\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \lim_{n \rightarrow \infty} \frac{2m}{n},$$

and a random (n, m, Δ) -multigraph G . We also assume that $\Delta(x)$ is aperiodic, meaning that

$$\gcd\{d_1 - d_2 \mid \delta_{d_1} \neq 0, \delta_{d_2} \neq 0\} = 1$$

(the periodic case is treated in Section 7.3).

- i) Any connected multigraph that is neither a tree nor a unicycle is asymptotically almost surely not a subgraph of G .
- ii) Let T be a tree with k vertices and $|\text{Aut}(T)|$ automorphisms. Then the expected number of T -subgraphs in G is asymptotically equal to

$$\frac{1}{|\text{Aut}(T)|} \frac{n^k}{(2m)^{k-1}} \prod_{v \in V(T)} \frac{\chi^{\deg(v)} \Delta(\deg(v))(\chi)}{\Delta(\chi)}.$$

- iii) Denoting by C_ℓ a cycle of length ℓ , then $G[C_\ell]$ follows asymptotically a Poisson law of mean

$$\frac{1}{2\ell} \left(\frac{n}{m} \frac{\chi^2 \Delta''(\chi)}{\Delta(\chi)} \right)^\ell.$$

Proof. i) Any connected multigraph H that is neither a tree nor a cycle has a essential density greater than 1. Let F denote a subgraph of H of density greater than 1. Then, when $2m/n$ and χ have finite limits,

$$F \left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right)$$

tends to zero. According to Proposition 1, the expected number of F -subgraphs in a random (n, m, Δ) -multigraph is

$$\frac{\text{MG}_{n,m,\Delta}^{[F]}}{\text{MG}_{n,m,\Delta}}.$$

Condition (C1) of Lemma 6 is satisfied. The polynomial $F(z, w, \mathbf{y})$ satisfies Condition (C2). Finally, Condition (C3) is satisfied by the aperiodic polynomial $\Delta(x)$, by application of the Daffodil Lemma [19]. Applying Lemma 6, we obtain

$$\text{MG}_{n,m,\Delta}^{[F]} \sim F \left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right) \text{MG}_{n,m,\Delta},$$

so the expected number of F -subgraph in a random (n, m, Δ) -multigraph G tends to 0. Therefore, with high probability, G contains no H -subgraph.

- ii) The second result of the theorem is also a direct consequence of Lemma 6.
- iii) The last result is proven by application of Proposition 3. We chose for α_n the function

$$\alpha_n(F) = F \left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right).$$

When C is a cycle, its number of vertices is the same as its number of edges, so

$$\alpha_n(C) = C \left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right)$$

has a finite limit (as m is proportional to n , and χ has a finite limit). Consequently, the first condition of Proposition 3 is satisfied. The second condition is obtained by application of Lemma 6, and the third condition is trivial. Thus, the number of cycles of length ℓ follows a Poisson limit law of parameter

$$\alpha_n(C) = \frac{1}{2\ell} \left(\frac{n}{m} \frac{\chi^2 \Delta''(\chi)}{\Delta(\chi)} \right)^\ell. \quad \square$$

Remark. The case of p -regular multigraphs, which is a trivial example of periodic Δ , can be dealt with in a simple way. Notice that $\Delta(x) = \frac{x^p}{p!}$, and thus we have

$$\frac{x^d \Delta^{(d)}(x)}{\Delta(x)} = p(p-1) \dots (p-d+1) = \frac{p!}{(p-d)!}$$

for $0 \leq d \leq p$. These quantities do not depend on x . Hence, the expression of $G(z, x, t)$ becomes

$$G(z, x, t) = F \left(nz, \frac{1}{2mt^2}, (\mu_d)_{d \geq 0} \right) \quad \text{where } \mu_d = \begin{cases} p(p-1) \dots (p-d+1), & \forall 0 \leq d \leq p, \\ 0, & \forall d > p. \end{cases}$$

Condition (C2) of Lemma 6 is trivially satisfied, while Conditions (C1) and (C3) are no longer needed. It suffices to notice that the total number of half-edges must be a multiple of p and that

$$[x^{pn}] \Delta^n(x) = \frac{1}{(p!)^n}.$$

Proposition 6. *Let $\mathcal{F}$ be a multigraph family with generating function $F(z, w, \mathbf{y})$, and let $\text{Reg}_{p,n}$ denote the total weight of p -regular multigraphs with n vertices (and $pn/2$ edges). The expected number of copies of elements of $\mathcal{F}$ is asymptotically equal to*

$$\frac{\text{Reg}_{p,n}^{[\mathcal{F}]}}{\text{Reg}_{p,n}} \sim F \left(n, \frac{1}{np}, (\mu_d)_{d \geq 0} \right) \quad \text{where } \mu_d = \begin{cases} p(p-1) \dots (p-d+1), & \text{if } 0 \leq d \leq p, \\ 0, & \text{if } d > p. \end{cases}$$

6.5 Infinite set of nonzero weights

We now turn to the case when there is an infinite number of nonzero weights, but these weights do not grow “too fast”; in particular we assume that their generating function has infinite radius of convergence, so Lemma 6 is applicable. Theorem 9 below is the equivalent of Theorem 6, in the case of degree constraints.

Theorem 9 (Poisson law, weights, strictly balanced, multigraphs). *Consider a weight generating function $\Delta(x)$ with nonnegative coefficients, infinite radius of convergence, that is not a polynomial, and such that for all $d \geq 0$, the ratio $\Delta^{(d)}(x)/\Delta(x)$ has a finite positive limit when x goes to infinity. Let F be a strictly balanced multigraph of density d . Consider an integer sequence $m = \Theta(n^{2-1/d})$, and let $\chi := \chi(n)$ denote the unique positive solution of the relation*

$$\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \frac{2m}{n}.$$

Suppose that $\Delta(x)$ satisfies Condition (C3) of Lemma 6. Then the number of F -subgraphs in a random (n, m, Δ) -multigraph has a limit Poisson law of parameter

$$\lim_{n \rightarrow \infty} F \left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right).$$

Proof. The only strictly balanced multigraphs of density 1 are the cycles. Their number follows a Poisson law, and the proof is the same as for Theorem 8. So we focus on the case of a strictly balanced multigraph F with density greater than 1, so $2m/n$ tends to infinity. The relation

$$\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \frac{2m}{n}$$

implies that χ tends to infinity with n . By assumption, $\frac{\Delta'(\chi)}{\Delta(\chi)}$ has a finite limit ℓ , which gives

$$\chi \sim \frac{2m}{\ell n}.$$

This implies that

$$\alpha_n(F) = F\left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)}\right)_{d \geq 0}\right)$$

has a finite limit. Also, for any multigraph H of density higher than d ,

$$\alpha_n(H) = \Theta\left(\frac{n^{n(H)}}{(2m)^{m(H)}} \prod_{v \in V(H)} \frac{\chi^{\deg(v)} \Delta^{(d)}(\chi)}{\Delta(\chi)}\right) = o(1),$$

because $m = \Theta(n^{2-1/d})$, $\frac{m(H)}{n(H)} > d$ and $\sum_{v \in V(H)} \deg(v) = 2m(H)$. By application of Lemma 6 the second assumption of Proposition 3 is satisfied, so the number of F -subgraphs in a random (n, m, Δ) -multigraph follows asymptotically a Poisson law of parameter $\alpha_n(F)$. $\square$

7 Extensions: when the conditions of Lemma 6 do not hold

We consider in this section an example for each case when one of the three conditions of Lemma 6 fails.

7.1 When Condition (C1) does not hold: $\frac{2m}{n} \rightarrow 0$, threshold for trees

We look here at what happens when $m/n \rightarrow 0$, and study the appearance of trees.

Given a family of degrees defined by its generating function $\Delta(x) = \sum_{d \geq 0} \delta_d \frac{x^d}{d!}$, set

$$\mu(j) := \min\{d : d \geq j, \delta_d > 0\}, \text{ for any } j \geq 0.$$

In particular, we have $\mu(0) = 0$, i.e., $\delta_0 > 0$. Otherwise there would be no isolated vertex, and the number of edges would grow at least linearly in the number of vertices ($2m \geq n$), contradicting $m/n \rightarrow 0$. Define also, for a tree T ,

$$\gamma_\Delta(T) := \sum_{v \in T} \mu(\deg(v)).$$

Theorem 10 (Trees in sparse multigraphs). *Let T be a given tree with k vertices. Suppose that Δ satisfies Condition (C3), and consider a random (n, m, Δ) -multigraph G with $m = \Theta(n\epsilon_n)$, where $\epsilon_n \rightarrow 0$. Then*

$$\mathbb{E}(G[T]) = \Theta\left(n\epsilon_n^{-(k-1) + \frac{\gamma_\Delta(T)}{\mu(1)}}\right).$$

Thus, when $\epsilon_n = o(n^\lambda)$ with $\lambda = \frac{\mu(1)}{(k-1)\mu(1) - \gamma_\Delta(T)}$, $G[T] = 0$ almost surely.

Proof. As $\frac{2m}{n}$ tends to 0, so does the solution χ of the saddle-point equation (13), and we can use the saddle-point equation to estimate the order of magnitude of the saddle-point and of the derivatives of Δ at χ :

$$\Delta(\chi) = \delta_0 + o(1) = \Theta(1), \quad \Delta^{(j)}(\chi) = \delta_{\mu(j)} \frac{\chi^{\mu(j)-j}}{\mu(j)!} + o(\chi^{\mu(j)-j}) = \Theta(\chi^{\mu(j)-j}), \quad \text{for all } j \geq 1.$$

This implies

$$\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \Theta(\chi^{\mu(1)}) = \frac{2m}{n} = \Theta(\epsilon_n).$$

Then we obtain

$$\chi = \Theta\left(\epsilon_n^{\frac{1}{\mu(1)}}\right), \quad \frac{\chi^j \Delta^{(j)}(\chi)}{\Delta(\chi)} = \Theta\left(\epsilon_n^{\frac{\mu(j)}{\mu(1)}}\right).$$

Although Condition (C1) of Lemma 6 does not hold, as long as Condition (C3) is valid the proof can be adapted, using the fact that $T(z, w, \mathbf{y})$ is simply a monomial in this case:

$$T(z, w, \mathbf{y}) = \frac{z^{k+1}}{(k+1)!} \frac{w^k}{2^k k!} \prod_{v \in V(T)} y_{\deg(v)}, \quad \text{where } k = |E(T)|.$$

First, we directly get Equation (19):

$$\text{MG}_{n,m,\Delta}^{[T]} = n! \frac{2^{m+\frac{1}{2}}}{\sqrt{\pi}} m! [z^n w^m] \int_0^\infty T(z, w, (\bar{\partial}\Delta)(t\sqrt{w})) e^{z\Delta(t\sqrt{w})} e^{-t^2/2} dt.$$

Applying the changes of variables

$$z \rightarrow \frac{nz}{\Delta(t\sqrt{w})}, \quad w \rightarrow \frac{x^2}{t^2}, \quad t \rightarrow \sqrt{2mt}$$

we get

$$\text{MG}_{n,m,\Delta}^{[T]} = \frac{n!}{n^n} \frac{(4m)^{m+\frac{1}{2}}}{\sqrt{\pi}} m! [z^n x^{2m}] \int_0^\infty T\left(nz, \frac{1}{2mt^2}, \left(\frac{x^d \Delta^{(d)}(x)}{\Delta(x)}\right)_d\right) \Delta^n(x) e^{nz} t^{2m} e^{-mt^2} dt.$$

Similarly, setting

$$G_n(z, x, t) := T\left(nz, \frac{1}{2mt^2}, \left(\frac{x^d \Delta^{(d)}(x)}{\Delta(x)}\right)_d\right), \quad L_n(z, x, t) := \frac{G_n(z, x, t)}{G_n(1, \chi, 1)},$$

we obtain

$$\text{MG}_{n,m,\Delta}^{[T]} = G_n(1, \chi, 1) \cdot \frac{n!}{n^n} \frac{(4m)^{m+\frac{1}{2}}}{\sqrt{\pi}} m! [z^n x^{2m}] \int_0^\infty L_n(z, x, t) \Delta^n(x) e^{nz} t^{2m} e^{-mt^2} dt.$$

Since $L_n(z, x, t)$ is a monomial, we can apply Laplace's method for $t = 1$ and obtain

$$\int_0^\infty L_n(z, x, t) t^{2m} e^{-mt^2} dt \sim_{n \rightarrow \infty} L_n(z, x, 1) \int_0^\infty t^{2m} e^{-mt^2} dt = L_n(z, x, 1) \sqrt{\frac{\pi}{2}} \frac{1}{(2m)^{m+\frac{1}{2}}} \frac{(2m)!}{2^m m!},$$

uniformly for $|z| = 1$ and x in some neighbourhood of the origin. Then we extract the coefficient in z with a simple saddle-point method at $z = 1$:

$$[z^n] L_n(z, x, 1) e^{nz} \sim L_n(1, x, 1) \frac{n^n}{n!}.$$

We are left with the extraction of the coefficient in x :

$$[x^{2m}] L_n(1, x, 1) \Delta^n(x) = \frac{1}{2i\pi} \int_{|x|=\chi} L_n(1, x, 1) \frac{\Delta^n(x)}{x^{2m+1}} dx = \frac{1}{2\pi \chi^{2m}} \int_{-\pi}^{\pi} L_n(1, \chi e^{i\theta}, 1) \frac{\Delta^n(\chi e^{i\theta})}{e^{2im\theta}} d\theta.$$

By Condition (C3) and the analyticity of L_n , the main contribution originates from the region around 0. We thus have

$$[x^{2m}] L_n(1, x, 1) \Delta^n(x) \sim L(1, \chi, 1) [x^{2m}] \Delta^n(x) = L(1, \chi, 1) \frac{[x^{2m}] \Delta^n(x)}{(2m)!}.$$

Collecting everything, we obtain the asymptotics of the expected number of occurrences of T :

$$\mathbb{E}(G[T]) = \frac{\text{MG}_{n,m,\Delta}^{[F]}}{\text{MG}_{n,m,\Delta}} \sim G_n(1, \chi, 1).$$

Substituting $m = n\epsilon_n$, we finally have

$$\mathbb{E}(G[T]) = G_n(1, \chi, 1) = T\left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)}\right)_d\right) = \Theta\left(n\epsilon_n^{-(k-1) + \frac{\gamma_\Delta(T)}{\mu(1)}}\right).$$

Again, the random variable $G[T]$ exhibits a threshold behaviour at the value

$$\epsilon_n := \Theta\left(n^{\frac{\mu(1)}{(k-1)\mu(1) - \gamma_\Delta(T)}}\right) \rightarrow 0. \quad \square$$

Notice that we recover the threshold obtained in Corollary 2 when all degrees are allowed, which implies that $\mu(j) = j$, for $j \geq 0$, and that $\gamma_\Delta(T)$ is simply twice the number of edges of T :

$$\epsilon_n := n^{\alpha_0 - 1} \Rightarrow \alpha_0 = 1 + \frac{1}{(k-1) - \gamma_\Delta(T)} = 1 - \frac{1}{k-1} = 2 - \frac{1}{d(T)}.$$

7.2 When Condition (C2) does not hold: Power law

This subsection is devoted to the study of weighted multigraphs where the distribution of the vertex degrees in a random multigraph G follows a power law with parameter $\beta > 1$, *i.e.*:

$$\exists C > 0, \forall d \geq 1, \forall v \in V(G) : \mathbb{P}(\deg(v) = d) = Cd^{-\beta}.$$

Many real-world networks exhibit a power-law degree distribution, usually with a parameter β located between 2 and 3. The degree distribution being fixed, the ratio $2m/n$ is then determined by β . In our model, this corresponds to weights $\delta_d = d^{-\beta}d!$ in the expression of Δ and to evaluating it at $x = 1$. We obtain

$$\Delta_\beta(x) := \sum_{d \geq 1} d^{-\beta} x^d \quad \text{and} \quad \mathbb{P}(\deg(\Gamma_{x=1}^{\mathcal{B}} V_{\Delta_\beta}) = d) = d^{-\beta}.$$

Notice that Δ_β now has finite radius of convergence equal to 1, and at $x = 1$ we have

$$\Delta_\beta(1) = \zeta(\beta), \quad \Delta'_\beta(1) = \zeta(\beta - 1),$$

where $\zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}$ is the Riemann zeta function. In this case, with the saddle-point located at $\chi = 1$, the saddle-point equation has a particular shape:

$$\frac{\Delta'_\beta(1)}{\Delta_\beta(1)} = \frac{\zeta(\beta - 1)}{\zeta(\beta)} = \frac{2m}{n}.$$

Hence the limit of $2m/n$ is prescribed by the parameter β of the power law.

Lemma 6 does not apply here as Condition (C2) is not satisfied: $L_n(z, x, t)$ is in general not analytic around $x = 1$, because of the higher derivatives of Δ_β involved in the formula.

Nonetheless, most of the steps are still valid. After a similar reasoning to extract the coefficient $[z^n]$ and estimate the integral with respect to t , one obtains

$$\text{MG}_{n,m,\Delta_\beta}^{[F]} \sim (2m)! [x^{2m}] F\left(n, \frac{1}{2m}, \left(\frac{x^d \Delta_\beta^{(d)}(x)}{\Delta_\beta(x)}\right)_{d \geq 0}\right) \Delta_\beta(x)^n.$$

Let us study a simple family F with a unique element, with n_F vertices and m_F edges. Then

$$F(z, w, \mathbf{y}) = \frac{1}{|\text{Aut}(F)|} \frac{z^{n_F}}{n_F!} \frac{w^{m_F}}{2^{m_F} m_F!} \prod_{v \in F} y_{\deg(v)}.$$

Replacing F by its expression in the above formula yields

$$\text{MG}_{n,m,\Delta_\beta}^{[F]} \sim \frac{(2m)!n^{n_F}}{(2m)^{m_F}n_F!2^{m_F}m_F!} [x^{2m}] \prod_{v \in F} \frac{x^{\deg(v)} \Delta_\beta^{(\deg(v))}(x)}{\Delta_\beta(x)} \Delta_\beta(x)^n.$$

A standard saddle-point technique cannot be applied directly here, as there is a conflict between the saddle-point at $x = 1$ with the large power $\Delta_\beta(x)^n$ and the singularities of $\Delta_\beta^{(j)}(x)$, $j \geq 2$.

To circumvent this problem, we use a reformulation of Lagrange inversion (*cf.* Lemma 7 in the appendix). In our setting, after some rewriting, we obtain

$$[x^{2m}] \prod_{v \in F} \frac{x^{\deg(v)} \Delta_\beta^{(\deg(v))}(x)}{\Delta_\beta(x)} \Delta_\beta(x)^n = [x^{2m-n}] \prod_{v \in F} \frac{x^{\deg(v)} \Delta_\beta^{(\deg(v))}(x)}{\Delta_\beta(x)} \left(\left(\frac{\Delta_\beta(x)}{x} \right)^{\frac{n}{2m-n}} \right)^{2m-n}.$$

Lemma 7 can now be applied with

$$\Phi(x) := \left(\frac{\Delta_\beta(x)}{x} \right)^{\frac{n}{2m-n}}, \quad H(x) := \prod_{v \in F} \frac{x^{\deg(v)} \Delta_\beta^{(\deg(v))}(x)}{\Delta_\beta(x)}$$

and with $T(z)$ defined implicitly by the equation $T(z) = z\Phi(T(z))$. We get

$$[x^{2m}] \prod_{v \in F} \frac{x^{\deg(v)} \Delta_\beta^{(\deg(v))}(x)}{\Delta_\beta(x)} \Delta_\beta(x)^n = [z^{2m-n}] \frac{zT'(z)}{T(z)} \prod_{v \in F} \frac{T(z)^{\deg(v)} \Delta_\beta^{(\deg(v))}(T(z))}{\Delta_\beta(T(z))}. \quad (24)$$

The right part of this equation is now amenable to singularity analysis.

Asymptotic expansions of Δ_β , Φ and T when $2 < \beta < 3$. The function $\Delta(x)$ is a simple case of a polylogarithm, whose asymptotics around $x = 1$ is known (see Flajolet and Sedgewick [19, Theorem VI.7, p. 408]):

$$\Delta_\beta(x) \sim \Gamma(1-\beta)(-\log(x))^{\beta-1} + \sum_{j \geq 0} \frac{(-1)^j}{j!} \zeta(\beta-j)(-\log(x))^j.$$

When $2 < \beta < 3$, one has more precisely

$$\Delta_\beta(x) = \zeta(\beta) - \zeta(\beta-1)(1-x) + \Gamma(1-\beta)(1-x)^{\beta-1} + (\zeta(\beta-2) - \zeta(\beta-1)) \frac{(1-x)^2}{2} + O(1-x)^\beta,$$

which leads to the following asymptotics for $\Phi(x)$:

$$\Phi(x) = \zeta(\beta)^{\frac{n}{2m-n}} + \frac{n}{2m-n} \frac{\zeta(\beta) - \zeta(\beta-1)}{\zeta(\beta)} \zeta(\beta)^{\frac{n}{2m-n}-1} (1-x) + \frac{n\Gamma(1-\beta)}{2m-n} \zeta(\beta)^{\frac{n}{2m-n}-1} (1-x)^{\beta-1} + O(1-x)^2.$$

Notice that

$$\frac{n}{2m-n} = \frac{1}{2m/n-1} = \frac{1}{\Delta'_\beta(1)/\Delta_\beta(1)-1} = \frac{1}{\zeta(\beta-1)/\zeta(\beta)-1} = \frac{\zeta(\beta)}{\zeta(\beta-1)-\zeta(\beta)}.$$

Finally one gets

$$\Phi(x) = \zeta(\beta)^{\frac{n}{2m-n}} \left(x + \frac{\Gamma(1-\beta)}{\zeta(\beta-1)-\zeta(\beta)} (1-x)^{\beta-1} \right) + O(1-x)^2.$$

To obtain the asymptotics of $T(z) = z\Phi(T(z))$, one might want to use a classical Lagrange inversion. Here the saddle-point lies on the circle of convergence of $\Phi(x)$, as they are both equal to 1. Fortunately, with this remarkable expansion of $\Phi(x)$, it falls into the setting presented by Flajolet and Sedgewick in [19, Section VI.18, p. 407], yielding directly the Δ -analyticity of $T(z)$ as well as its expansion:

$$T(z) = 1 - \tau^{\frac{1}{\beta-1}} \left(1 - \frac{z}{\rho} \right)^{\frac{1}{\beta-1}} + o \left(1 - \frac{z}{\rho} \right)^{\frac{1}{\beta-1}}, \text{ where } \tau := \frac{\zeta(\beta-1) - \zeta(\beta)}{\Gamma(1-\beta)}, \text{ and } \rho := \zeta(\beta)^{-\frac{n}{2m-n}}. \quad (25)$$

Moreover, we have

$$T'(z) = \frac{\tau^{\frac{1}{\beta-1}}}{\rho(\beta-1)} \left(1 - \frac{z}{\rho}\right)^{\frac{1}{\beta-1}-1} + o\left(1 - \frac{z}{\rho}\right)^{\frac{1}{\beta-1}-1}.$$

We do not pursue the computation of the asymptotics in the general case, as it becomes somewhat intricate, but consider below the specific case of cycles.

Cycles of fixed length

When F is a cycle of length ℓ , ($\ell \geq 3$), we have

$$F(z, w, \mathbf{y}) = \frac{1}{2\ell} \frac{z^\ell}{\ell!} \frac{w^\ell}{2^\ell \ell!} y_2^\ell.$$

Theorem 11 (Cycles in power-law multigraphs). *When the weights follow a power law of parameter β ($2 < \beta < 3$), the expected number of ℓ -cycles in a random (n, m, Δ_β) -multigraph is equal to*

$$\frac{\text{MG}_{n,m,\Delta_\beta}^{[F]}}{\text{MG}_{n,m,\Delta_\beta}} \sim \kappa_{\beta,\ell} n^{\frac{3-\beta}{\beta-1}\ell},$$

where $\kappa_{\beta,\ell}$ is independent of n and can be explicitly computed.

Proof. We want to extract the coefficient $[z^{2m-n}]$ in the following expression, coming from (24):

$$\frac{zT'(z)}{T(z)} \left(\frac{T(z)^2 \Delta_\beta''(T(z))}{\Delta_\beta(T(z))} \right)^\ell.$$

Using [19, Theorem VI.7, p. 408] again, we first obtain the asymptotics of Δ_β'' around $x = 1$:

$$\begin{aligned} \Delta_\beta''(x) &= \frac{\Delta_{\beta-2}(x) - \Delta_{\beta-1}(x)}{x^2} \\ &= \Gamma(3-\beta)(-\log(x))^{\beta-3} + o(\log^{\beta-3}(x)), \end{aligned}$$

where the second equation holds because $\beta - 3 < 0$, all the other powers of ξ being positive. With the expansion (25) for $T(z)$, we get

$$\begin{aligned} \Delta_\beta(T(z)) &= \zeta(\beta) + o(1); \\ \Delta_\beta''(T(z)) &= \Gamma(3-\beta) \left(\tau \left(1 - \frac{z}{\rho}\right) \right)^{\frac{\beta-3}{\beta-1}} + o\left(1 - \frac{z}{\rho}\right)^{\frac{\beta-3}{\beta-1}}. \end{aligned}$$

We are in the critical composition scheme (cf. [p. 412][19]): The value of T at its singularity is equal to the singularity of Δ ,

$$T(\rho) = 1.$$

We now collect every expansion and obtain

$$\frac{zT'(z)}{T(z)} \left(\frac{T(z)^2 \Delta_\beta''(T(z))}{\Delta_\beta(T(z))} \right)^\ell \sim z^{\tau^{\frac{\beta-3}{\beta-1}\ell + \frac{1}{\beta-1}} \Gamma(3-\beta)^\ell} \frac{\Gamma(3-\beta)^\ell}{\rho(\beta-1)\zeta(\beta)^\ell} \left(1 - \frac{z}{\rho}\right)^{\frac{\beta-3}{\beta-1}\ell + \frac{1}{\beta-1} - 1}.$$

Applying the transfer lemma presented in Flajolet and Sedgewick [19], we get finally

$$\begin{aligned} [z^{2m-n}] \frac{zT'(z)}{T(z)} \left(\frac{T(z)^2 \Delta_\beta''(T(z))}{\Delta_\beta(T(z))} \right)^\ell &\sim [z^{2m-n-1}] \frac{\tau^{\frac{(\beta-3)\ell+1}{\beta-1}} \Gamma(3-\beta)^\ell}{\rho(\beta-1)\zeta(\beta)^\ell} \left(1 - \frac{z}{\rho}\right)^{\frac{(\beta-3)\ell-\beta+2}{\beta-1}} \\ &\sim \frac{\tau^{\frac{(\beta-3)\ell+1}{\beta-1}} \Gamma(3-\beta)^\ell}{(\beta-1)\zeta(\beta)^\ell} \frac{(2m-n-1)^{-\frac{(\beta-3)\ell-\beta+2}{\beta-1}-1}}{\Gamma\left(-\frac{(\beta-3)\ell-\beta+2}{\beta-1}\right)} \rho^{-(2m-n)} \\ &\sim \frac{\tau^{\frac{(\beta-3)\ell+1}{\beta-1}} \Gamma(3-\beta)^\ell}{(\beta-1)\zeta(\beta)^\ell \Gamma\left(\frac{(3-\beta)\ell+\beta-2}{\beta-1}\right)} (2m-n)^{\frac{(3-\beta)\ell-1}{\beta-1}} \zeta(\beta)^n. \end{aligned}$$

Now, from the saddle-point equation the term $(2m - n)$ can also be written as

$$\left(\frac{\zeta(\beta - 1)}{\zeta(\beta)} - 1 \right) n.$$

We obtain the asymptotics of $\text{MG}_{n,m,\Delta_\beta}^{[F]}$:

$$\text{MG}_{n,m,\Delta_\beta}^{[F]} \sim \frac{(2m)!n^\ell}{(2m)^\ell \ell! 2^{2\ell}} \frac{\tau^{\frac{(\beta-3)\ell+1}{\beta-1}} \Gamma(3-\beta)^\ell}{(\beta-1)\zeta(\beta)^\ell \Gamma\left(\frac{(3-\beta)\ell+\beta-2}{\beta-1}\right)} \left(\frac{\zeta(\beta-1)}{\zeta(\beta)} - 1 \right)^{\frac{(3-\beta)\ell-1}{\beta-1}} n^{\frac{(3-\beta)\ell-1}{\beta-1}} \zeta(\beta)^n.$$

Similarly, we can obtain the asymptotics of $\text{MG}_{n,m,\Delta_\beta}$ (by simply setting $\ell = 0$ in the above formula):

$$\text{MG}_{n,m,\Delta_\beta} \sim (2m)! \frac{\tau^{\frac{1}{\beta-1}}}{(\beta-1) \Gamma\left(\frac{\beta-2}{\beta-1}\right)} \left(\frac{\zeta(\beta-1)}{\zeta(\beta)} - 1 \right)^{-\frac{1}{\beta-1}} n^{-\frac{1}{\beta-1}} \zeta(\beta)^n.$$

Setting

$$\begin{aligned} \kappa_\beta &:= \frac{\tau^{\frac{\beta-3}{\beta-1}\ell}}{\ell! 2^{2\ell}} \left(\frac{\zeta(\beta-1)}{\zeta(\beta)} - 1 \right)^{\frac{3-\beta}{\beta-1}\ell} \frac{\Gamma\left(\frac{\beta-2}{\beta-1}\right)}{\Gamma\left(\frac{(3-\beta)\ell+\beta-2}{\beta-1}\right)} \left(\frac{\Gamma(3-\beta)}{\zeta(\beta-1)} \right)^\ell \\ &= \frac{1}{\ell! 2^{2\ell}} \left(\frac{\zeta(\beta)}{\Gamma(1-\beta)} \right)^{\frac{\beta-3}{\beta-1}\ell} \frac{\Gamma\left(\frac{\beta-2}{\beta-1}\right)}{\Gamma\left(\frac{(3-\beta)\ell+\beta-2}{\beta-1}\right)} \left(\frac{\Gamma(3-\beta)}{\zeta(\beta-1)} \right)^\ell \end{aligned}$$

finishes the proof of Theorem 11. □

7.3 When Condition (C3) does not hold: Eulerian multigraphs and graphs with periodic degree sequence

When there exist integers $r \geq 0$ and $p \geq 2$ such that the index of any nonzero coefficient of Δ is equal to r modulo p , the function Δ is said to be *periodic*. When p is the largest integer satisfying this property, then Δ is said to be *p-periodic*. This is in particular the case of Eulerian graphs, where each vertex has an even degree, so

$$\Delta(x) = \cosh(x).$$

According to the Daffodil Lemma ([19, Lemma IV.1]), there is a function Ω analytic at 0 and aperiodic (*i.e.*, its largest period is 1) such that

$$\Delta(x) = x^r \Omega(x^p).$$

Condition (C2), which states that the main contribution of the integral of Δ on a circle centered at the origin comes from the part close to the real axis, no longer holds. Indeed, for any positive real value χ and p th root ρ of unity, we have

$$|\Delta(\rho\chi)| = |(\rho\chi)^r \Omega(\rho^p \chi^p)| = \Delta(\chi).$$

The number of edges of an (n, m, Δ) -multigraph must then satisfy

$$2m = \sum_{v \in V(G)} \deg(v) = nr + p \sum_{v \in V(G)} \frac{\deg(v) - r}{p},$$

so $2m - nr$ must be divisible by p .

According to Lemma 5, the total weight of (n, m, Δ) -multigraphs is

$$\text{MG}_{n,m,\Delta} = (2m)! [x^{2m}] \Delta(x)^n = (2m)! [x^{2m}] x^{nr} \Omega(x^p)^n.$$

This implies that p must divide $2m - nr$ (otherwise, there are no (n, m, Δ) -multigraphs, so the total weight is 0), and

$$\text{MG}_{n,m,\Delta} = (2m)! [x^{\frac{2m-nr}{p}}] \Omega(x)^n.$$

Likewise, for each integer d , the function $x^d \Delta^{(d)}(x)$ is also p -periodic, so there is a function Ω_d such that

$$x^d \Delta^{(d)}(x) = x^r \Omega_d(x^p).$$

The equation

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = \frac{n! 2^{m+\frac{1}{2}} m! (2m)^{m+\frac{1}{2}}}{n^n \sqrt{\pi}} \int_{t=0}^{\infty} [z^n x^{2m}] F \left(nz, \frac{1}{2mt^2}, \left(\frac{x^d \Delta^{(d)}(x)}{\Delta(x)} \right)_{d \geq 0} \right) e^{nz} \Delta(x)^n t^{2m} e^{-mt^2} dt$$

becomes

$$\text{MG}_{n,m,\Delta}^{[\mathcal{F}]} = \frac{n! 2^{m+\frac{1}{2}} m! (2m)^{m+\frac{1}{2}}}{n^n \sqrt{\pi}} \int_{t=0}^{\infty} [z^n x^{\frac{2m-nr}{p}}] F \left(nz, \frac{1}{2mt^2}, \left(\frac{\Omega_d(x)}{\Omega(x)} \right)_{d \geq 0} \right) e^{nz} \Omega(x)^n t^{2m} e^{-mt^2} dt.$$

The rest of the analysis of subgraphs in (n, m, Δ) -multigraphs is then the same, and we conclude that any multigraph F with generating function $F(z, w, \mathbf{y})$ satisfies the following assertion: If $m := m(n)$ tends to infinity with n in such a way that

- p divides $2m - nr$,
- and the term

$$E_n = F \left(n, \frac{1}{2m}, \left(\frac{\Omega_d(\chi)}{\Omega(\chi)} \right)_{d \geq 0} \right),$$

where $\chi := \chi(n)$ is the unique positive solution of the equation

$$\frac{\chi \Omega'(\chi)}{\Omega(\chi)} = \frac{2m - nr}{pn},$$

has a finite limit, as n tends to infinity,

then the limit of E_n is equal to the asymptotic expected number of F -subgraphs in a random (n, m, Δ) -multigraph.

When $2m/n$ tends to infinity, then χ tends to infinity with n . In that case, there are other more subtle ways for Condition (C2) to fail. For example, Δ could be aperiodic, but may be close to a periodic function on circles of large radius. For example, Condition (C2) is not valid for

$$\Delta(x) = \sinh(x) + 1$$

for any $2m/n$ tending to infinity. Those weights correspond to multigraphs where each vertex has either odd degree or degree 0. Those other examples require a case-by-case refinement of our analysis to determine the contributions on the circle of radius χ .

8 Perspectives, future works and limitations

We presented an approach via analytic combinatorics to the problem of counting subgraphs in various models of random graphs. The approach gives precise expressions for subgraph counts and distributional results on the number of subgraphs belonging to an *a priori* given family of graphs. The notion of patchworks was the crucial concept to keep track of vertices and edges even if several copies of subgraphs overlap.

This approach can certainly be enhanced into several directions. One is the treatment of strictly balanced *induced* subgraphs. There are already several examples in the literature treating “fixed” as well as induced subgraphs.

Another direction is weakening the balancedness property. In order to cope with overlappings, it proved convenient to require that the subgraphs under consideration must be strictly balanced. It

seems hard to drop this requirement completely, but the condition can be weakened to barely balanced subgraphs (for the definition see the end of Section 2). Certainly, some additional technicalities have to be overcome, but we are confident that the methodology presented in this paper can be adapted to cover subgraph counts for barely balanced graphs. This problem is the topic of a forthcoming paper, see [13].

What is currently not covered by our approach are random graphs with other constraints than constraints on their node degrees. It seems not out of reach to extend the approach to graph families with properties being amenable to specifications with generating functions. Examples for such graph families are probably connected graphs or multi-partite graphs.

Except for varying the graph family we may think of modifying the stage in the random graph process. It is well known that there are many copies of a given subgraph if we are beyond the threshold. So far, we have no results on what *exactly* happens beyond the threshold, *i.e.*, asymptotics for subgraph counts, Gaussian limiting distributions, *etc.* As mentioned at the end of the last section, certain subtleties may arise which seem to require a case-by-case study.

What seems currently out of reach is the counting problem for graph families defined via topological properties like forbidding certain minors. In this case, we would need the series of minor patchworks. Our approach enables us to count graphs with a distinguished occurrence of a minor, but this does not allow us to count graphs with at least one occurrence, since the graphs with many occurrences will appear more often than the ones with few occurrences. The resulting distribution is therefore not uniform.

Another problem we currently cannot solve is the case of infinite families of subgraphs. For instance, when the subgraph family is the set of all cycles (of arbitrary length) and we are asking for the number of $\mathcal{F}$ -free graphs, then the resulting counting problem is that for the number of trees on n vertices. Although our approach is quite powerful and general, it is currently not allowing to cover this simple problem. Since having a certain minor or not may also be described in our setting by means of an infinity (multi)graph family, an extension to infinite families seems particularly desirable.

Acknowledgment. We thank a referee for careful reading which led to substantial improvements of the original paper.

References

- [1] Omer Angel, Remco van de Hofstad, and Cecilia Holmgren. Limit laws for self-loops and multiple edges in the configuration model. *Preprint*, 2016. arXiv:1603.07172.
- [2] Andrew Barbour and Adrian Röllin. Central limit theorems in the configuration model. *Preprint*, 2017. arXiv:1710.02644.
- [3] Alexander Barvinok. On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries. *Advances in Mathematics*, 224:316–339, 2010.
- [4] Alexander Barvinok and John Hartigan. An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Transactions of the American Mathematical Society*, 364:4323–4368, 2012.
- [5] Alexander Barvinok and John Hartigan. The number of graphs and a random graph with a given degree sequence. *Random Structures & Algorithms*, 42(3):301–348, 2013.
- [6] François Bergeron, Gilbert Labelle, and Pierre Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1997.
- [7] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316, 1980.
- [8] Béla Bollobás. Threshold functions for small subgraphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 9:197–206, 1981.
- [9] Béla Bollobás, Jeong Han Kim, and Jacques Verstraëte. Regular subgraphs of random graphs. *Random Structures and Algorithms*, 29(1), 2006.

- [10] Siu On Chan and Michael Molloy. $(k + 1)$ -cores have k -factors. *Combinatorics, Probability and Computing*, 21(6), November 2012.
- [11] Sourav Chatterjee, Persi Diaconis, and Allan Sly. Random graphs with a given degree sequence. *The Annals of Applied Probability*, pages 1400–1435, 2011.
- [12] Xueqing Chen, Ki-Bong Nam, and Tomáš Pospíchal. Quivers and representations. In *Handbook of Algebra. Vol. 6*, volume 6 of *Handb. Algebr.*, pages 507–561. Elsevier/North-Holland, Amsterdam, 2009.
- [13] Gwendal Collet and Danièle Gardy. Patchworks of barely balanced graphs. In preparation.
- [14] Élie de Panafieu and Lander Ramos. Graphs with degree constraints. *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics (Analco16)*, <http://epubs.siam.org/doi/book/10.1137/1.9781611974324>, 2016.
- [15] Michael Drmota, Lander Ramos, and Juanjo Rué. Subgraph statistics in subcritical graph classes. *Random Structures and Algorithms*, 51:631–673, 2017.
- [16] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13:2004, 2004.
- [17] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.
- [18] Philippe Flajolet, Donald E. Knuth, and Boris Pittel. The first cycles in an evolving graph. *Discrete Mathematics*, 75(1–3):167–215, 1989.
- [19] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [20] Pu Gao. The first k -regular subgraph is large. *Combinatorics, Probability and Computing*, 23(3):412–433, May 2014.
- [21] Pu Gao and Nicholas Wormald. Enumeration of graphs with a heavy-tailed degree sequence. *Advances in Mathematics*, 287:412–450, 2016.
- [22] Zhicheng Gao and Nicholas Wormald. Distribution of subgraphs of random regular graphs. *Random Structures & Algorithms*, 32(1):38–48, 2008.
- [23] Omer Giménez and Marc Noy. Asymptotic enumeration and limit laws of planar graphs. *Journal of the American Mathematical Society*, 22(2):309–329, 2009.
- [24] Catherine Greenhill, Michael Isaev, and Brendan McKay. Subgraph counts for dense random graphs with specified degrees. 2018. arXiv:1801.09813.
- [25] Catherine Greenhill and Brendan McKay. Asymptotic enumeration of sparse multigraphs with given degrees. *SIAM Journal on Discrete Mathematics*, 27:2064–2089, 2013.
- [26] Walter K. Hayman. A generalisation of Stirling’s formula. *J. Reine Angew. Math.*, 196:67–95, 1956.
- [27] Svante Janson, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel. The birth of the giant component. *Random Structures and Algorithms*, 4(3):233–358, 1993.
- [28] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. John Wiley & Sons, 2000.
- [29] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142:61–92, 2004.
- [30] Jeong Han Kim, Benny Sudakov, and Van Vu. Small subgraphs of random regular graphs. *Discrete Mathematics*, 307:1961–1967, 2007.

- [31] Shoham Letzter. The property of having a k -regular subgraph has a sharp threshold. *Random Structures and Algorithms*, 42(4):509–519, July 2013.
- [32] Brendan McKay. Subgraphs of random graphs with specified degrees. *Proceedings of ICM*, 9, 2010.
- [33] Brendan McKay. Subgraphs of dense random graphs with specified degrees. *Combinatorics, Probability and Computing*, 20(3):413–433, 2011.
- [34] Brendan McKay and Nicholas Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *European Journal of Combinatorics*, 11(6):565–580, 1990.
- [35] Brendan McKay, Nicholas Wormald, and Beata Wysocka. Short cycles in random regular graphs. *Electronic J. of Combinatorics*, 11(1):66, 2004.
- [36] Dieter Mitsche, Michael Molloy, and Pawel Prałat. k -regular subgraphs near the k -core threshold of a random graph. 2018. arXiv:1804.04173.
- [37] Marc Noy, Clément Requilé, and Juanjo Rué. Further results on random cubic planar graphs. 2018. arXiv:1802.06679.
- [38] Robin Pemantle and Mark C. Wilson. *Analytic Combinatorics in Several Variables*. Cambridge University Press, New York, NY, USA, 2013.
- [39] Pawel Prałat, Jacques Verstraëte, and Nicholas Wormald. On the threshold for k -regular subgraphs of random graphs. *Combinatorica*, 31(5):565–581, 2011.
- [40] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields*, 78:1–10, 1988.
- [41] Andrzej Ruciński and Andrew Vince. Strongly Balanced Graphs and Random Graphs. *Journal of Graph Theory*, 10:251–264, 1986.
- [42] Clara Stegehuis, Remco van der Hofstad, A.J.E.M. Janssen, and Johan van Leeuwen. Clustering spectrum of scale-free networks. *Physical Review E*, 96 (4),042309, 2017.
- [43] Remco van der Hofstad. Counting graphs and null models of complex networks: Configuration model and extensions. In *Graph-Theoretic Concepts in Computer Science - 43rd International Workshop, WG 2017, Eindhoven, The Netherlands, June 21-23, 2017*, pages 1–17, November 2017.
- [44] Remco van der Hofstad. *Random Graphs and Complex Networks. Vol. I*. Cambridge Series on Statistical and Probabilistic Mathematics, 2017.
- [45] Remco van der Hofstad, A.J.E.M. Janssen, Johan van Leeuwen, and Clara Stegehuis. Local clustering in scale-free networks with hidden variables. *Physical review E*, 95(022307), 2017.
- [46] Remco van der Hofstad, Johan van Leeuwen, and Clara Stegehuis. Optimal subgraph structures in scale-free configuration models. 2017. arXiv:1709.03466.
- [47] Remco van der Hofstad, Johan van Leeuwen, and Clara Stegehuis. Triadic closures in configuration models with unbounded degree fluctuations. 2017. arXiv:1710.02027. To appear in *Journal of Statistical Physics*.
- [48] Nicholas Wormald. *Some problems in the enumeration of labelled graphs*. Newcastle University, 1978.
- [49] Nicholas Wormald. Models of random regular graphs. In J.D. Lamb and D.A. Preece, editors, *Surveys in Combinatorics*, volume 276 of *London Mathematical Society Lecture Note Series*, pages 239–298. Cambridge University Press, 1999.

A1 Appendix: analytic and combinatorial tools

A1.1 Lagrange inversion

The Lagrange inversion theorem allows us to obtain the coefficients of a function defined by an implicit equation; see for example [19, p. 732-33]. Here we use it in the following form:

Lemma 7. *Let $\Phi(t), H(t)$ be formal power series, such that: $\Phi(0) \neq 0$. Let $T(z)$ be a function implicitly defined by: $T(z) = z\Phi(T(z))$. Then, for all $n \geq 0$:*

$$[t^n]H(t)\Phi(t)^n = [z^n]\frac{zT'(z)}{T(z)}H(T(z)).$$

Proof. It follows from a simple change of variable $t = T(z)$, i.e. $z = t/\Phi(t)$:

$$[z^n]\frac{zT'(z)}{T(z)}H(T(z)) = \frac{1}{2i\pi} \oint \frac{zT'(z)}{T(z)}H(T(z))\frac{dz}{z^{n+1}} = \frac{1}{2i\pi} \oint \frac{H(t)}{t} \frac{dt}{(t/\Phi(t))^n} = [t^n]H(t)\Phi(t)^n. \quad \square$$

A1.2 The saddle-point heuristic

We use in our proofs a simple case of the Laplace method (see e.g. the book of Pemantle and Wilson [38]).

Lemma 8 (Laplace). *Consider two entire functions $A(t)$ and $B(t)$, where $B(t)$ is a positive function that reaches its unique maximum on the real open interval I at a point r , and $A(r) \neq 0$. Then we have*

$$\int_I A(t)B(t)^n dt \sim A(r) \int_I B(t)^n dt$$

whenever the integral is well defined.

The saddle-point method is a technique to compute the asymptotics of the coefficients of a generating function. The coefficient extraction is written as a Cauchy integral, on which a Laplace method is applied. There exist many variations of this technique. We will use here the following lemma, which is a particular case of Theorem VIII.8 from [19].

Lemma 9 (Saddle point). *Consider two entire functions $A(z)$ and $B(z)$, and a sequence of integers $N(n)$ such that $N(n)/n$ has a positive finite limit λ . Assume there exists a unique positive solution r to the equation*

$$\frac{rB'(r)}{B(r)} = \lambda,$$

such that $A(r) \neq 0$ and $\left(\frac{rB'(r)}{B(r)}\right)' \neq 0$. Then

$$[z^{N(n)}]A(z)B(z)^n \sim A(r)[z^{N(n)}]B(z)^n.$$