

Analytic Combinatorics of Graphs

Habilitation à Diriger des Recherches

Élie de Panafieu

13. October 2026

Jury composition

Frédérique Bassino	Professeure des Universités	Université Sorbonne Paris Nord
Nicolas Curien (referee)	Professeur des Universités	Université Paris-Saclay
Michael Drmota (referee)	Full Professor	Technische Universität Wien
Antoine Genitrini	Professeur des Universités	Sorbonne Université
Cyril Nicaud (referee)	Professeur des Universités	Université Gustave Eiffel

Defense

November 30th 2026 at 9 a.m.

Jussieu room 25-26 / 105.

Contents

1	Introduction	3
2	Analytic combinatorics	6
2.1	Symbolic method	6
2.2	Asymptotic extraction	7
2.3	Marking a parameter	9
3	Graphs and multigraphs	10
3.1	Simple graphs, multigraphs and compensation factors	10
3.2	Subgraphs	12
4	Degree constraints	15
4.1	Multigraphs with constrained degrees	15
4.2	Subgraphs in multigraphs with degree constraints	17
4.3	Simple graphs with constrained degrees	18
4.4	Regular graphs	20
5	Connected graphs and the phase transition	22
5.1	Fixed excess	23
5.2	Large excess	24
5.3	The critical window	25
5.4	The supercritical regime: an open problem	26
6	Directed graphs	28
6.1	Exact enumeration	28
6.2	Directed acyclic graphs with constrained degrees	30
6.3	The birth of the strong components	32
6.4	Directed acyclic graphs in the $MD(n, m)$ model	34
6.5	Beyond the critical window: an open problem	35
7	Satisfiable 2-SAT formulas	36
8	Spectrum of random graphs and tree walks	39
9	Cluster graphs	43
10	Active clustering	46
11	Perspectives	50
	Bibliography	52

1 Introduction

My field of research is analytic combinatorics: the encoding of discrete structures by generating functions, and the extraction of asymptotics from the analytic behavior of those functions. This manuscript presents my work since the PhD on graph enumeration and on the analysis of graph-related algorithms. It leaves out my work on algorithms unrelated to graphs: computational social choice, over the last three years (Caizergues et al. 2026; Durand, Panafieu, and Perarnau 2026; Panafieu, Durand, and Lang 2026), queueing systems (Bouillard et al. 2018), networks involving the encapsulation and decapsulation of protocols (Panafieu, Lamali, and Wallner 2019; Panafieu and Wallner 2026), and indoor positioning systems (Chen et al. 2024).

One of the joys of my research is the interaction with other fields and researchers. On graph enumeration, those fields are probability, statistical physics and graph theory. The interaction is sometimes a healthy competition over which community will first solve a long-standing problem. Collaborations in which each field brings part of the solution are far more productive and satisfying, as Chapter 9 and Chapter 10 illustrate. The analysis of algorithms also forces me out of my comfort zone in two ways: I have to explore the field the algorithm comes from, and its analysis may require tools I am unfamiliar with. I knew very little about q -analogs or spectral theory before working on the active clustering problem of Chapter 10 and the tree walks of Chapter 8.

I favor proofs that delegate the technical part, the asymptotic extraction, to a *black box theorem*, in the sense of Section 2.2. I suspect that the reputation of the analytic combinatorics of graphs for being highly technical is partly due to long papers built on ad hoc arguments. Such arguments are natural in the first proof of a new result, but I value the derivation of new, simpler proofs. An example is in Section 5.3, where the technical analysis of Janson et al. (1993) is replaced by an application of a variant of the semi-large powers theorem of Banderier et al. (2001). This motivates several of the open problems of Chapter 11.

Contents and publications.

- Chapter 2** The symbolic method, asymptotic extraction, and the link with probability.
- Chapter 3** Simple graphs and multigraphs, and the distribution of small subgraphs. Collet, Gwendal, Élie de Panafieu, Danièle Gardy, Bernhard Gittenberger, and Vlado Ravelomanana. 2020. “Threshold Functions for Small Subgraphs in Simple Graphs and Multigraphs.” *European Journal of Combinatorics* 88:43
- Chapter 4** Degree constraints, and the complete asymptotic expansion of regular graphs, settling a conjecture of McKay (1983). Caizergues, Emma, and Élie de Panafieu. 2023. “Exact Enumeration of Graphs and Bipartite Graphs with Degree Constraints.” In *European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB)*
- Panafieu, Élie de. 2024. “Asymptotic Expansion of Regular and Connected Regular Graphs.” *arXiv:2408.12459*
- Chapter 5** Connected graphs and the phase transition, with work in progress with Sergey Dovgal and Alexandros Singh.

Panafieu, Élie de. 2019. “Analytic Combinatorics of Connected Graphs.” *Random Structures & Algorithms* 55(2):427–495

Chapter 6 Directed graphs, with work in progress with Antoine Genitrini and Khaydar Nurligareev, and with Bernhard Gittenberger, Mehdi Naima and Martin Pépin.

Panafieu, Élie de, and Sergey Dovgal. 2020. “Counting Directed Acyclic and Elementary Digraphs.” *Proceedings of the 32nd Conference on Formal Power Series and Algebraic Combinatorics (FPSAC), Séminaire Lotharingien de Combinatoire* 84B:12

Dovgal, Sergey, Élie de Panafieu, Dimbinaina Ralaivaosana, Vonjy Rasendrasina, and Stephan Wagner. 2024. “The Birth of the Strong Components.” *Random Structures & Algorithms* 64(2):170–266

Chapter 7 Satisfiable 2-SAT formulas and their implication digraphs.

Dovgal, Sergey, Élie de Panafieu, and Vladyslav Ravelomanana. 2023. “Exact Enumeration of Satisfiable 2-SAT Formulae.” *Combinatorial Theory* 3(2):38

Chapter 8 The spectrum of sparse random graphs, through tree walks.

Hainzl, Eva-Maria, and Élie de Panafieu. 2026. “Tree Walks and the Spectrum of Random Graphs.” *European Journal of Combinatorics* 136:26

Chapter 9 Community detection and the random cluster graph.

Gösgens, Martijn, Lukas Lühtrath, Elena Magnanini, Marc Noy, and Élie de Panafieu. 2024. “The Erdős-Rényi Random Graph Conditioned on Every Component Being a Clique.” *arXiv:2405.13454*

Chapter 10 Active clustering, for the labeling of a training set.

Lutz, Quentin, Élie de Panafieu, Alex Scott, and Maya Stein. 2021. “Active Clustering for Labeling Training Data.” *Advances in Neural Information Processing Systems (NeurIPS)* 34:8469–8480

Chapter 11 Open problems and research perspectives.

Acknowledgements. I warmly thank Nicolas Curien, Michael Drmota and Cyril Nicaud for agreeing to referee this manuscript, and Frédérique Bassino and Antoine Genitrini for serving on the jury.

Marc Noy brought me into the RandNET project, visited me in Paris, invited me to Barcelona, and has been a generous collaborator and a friend. Through RandNET, I had the pleasure of working with Alex Scott and Maya Stein, with Martijn Gösgens, Lukas Lühtrath and Elena Magnanini, and with Eva Hainzl, who visited many times and took great initiative in our work. I am grateful to Michael Drmota and Remco van der Hofstad for welcoming me in Vienna and Eindhoven, and for encouraging their students at the time, Eva Hainzl and Martijn Gösgens, to work with me. The PAnDAG project kept inviting me to its meetings, and started two lines of work I am excited about: with Bernhard Gittenberger, Mehdi Naima and Martin Pépin on one side, and with Antoine Genitrini and Khaydar Nurligareev on the other. Alexandros Singh joined Sergey Dovgal and me on the work in progress of Chapter 5. Michael Borinsky offered me fascinating conversations, and shared his draft proving, by a different route, the exact expression for bipartite graphs with degree constraints. I also thank the other coauthors of the work presented here: Gwendal Collet, Danièle Gardy and Bernhard Gittenberger on the distribution of

subgraphs; Dimbinaina Ralaivaosaona, Vonjy Rasendrahasina and Stephan Wagner on directed acyclic graphs; and Laurent Ménard, who suggested with Nicolas Curien the problem of tree walks. More broadly, I thank all my coauthors, named here or not, for the mathematics we did together and for the fun we had doing it, and I am lucky to count many of them as friends.

Two chapters grew out of doctoral theses I co-supervised: Chapter 10, with Quentin Lutz at Télécom Paris, together with Thomas Bonald; and Chapter 4, with Emma Caizergues at Université Paris-Dauphine, together with Jérôme Lang and François Durand. I thank François Durand and Jérôme Lang for bringing me into computational social choice, a topic on which I also had the pleasure of working with Guillem Perarnau and Albert Majó i Rudé. Thanks also to all members of the PROSOC project, submitted to the ANR, who are working with me to bring the social choice and analysis of algorithms communities together. Mohamed (Amine) Lamali and Michael Wallner have walked with me, nondeterministically, for many years. Sergey Dovgal brings an inexhaustible energy to everything we work on.

I thank all my teachers, from Michel Lagouge and Marie-Christine Bras, who gave me in high school a taste for computer science and for mathematics, to Philippe Flajolet and my other teachers of analytic combinatorics, Nicolas Broutin, Michèle Soria, Brigitte Vallée and Gilles Schaeffer. Philippe also helped me find my PhD, and most of my research stands on his shoulders. Vlady Ravelomanana, my PhD advisor, introduced me to graph enumeration, and always had an interesting open problem to share. Manuel Kauers in Linz, and Clémence Magnien and Matthieu Latapy at LIP6, hosted me as a postdoc and widened my scientific horizons. I also thank the colleagues with whom I teach, or have taught recently, Florent Koechlin, Marie Albenque, Pablo Rotondo, Vincent Jugé, Lucas Gerin, Olivier Bodini and Eva Hainzl, for their enthusiasm and for the pedagogical ideas they brought to our courses.

The combinatorics community, at the Aléa meetings at CIRM, at the Journées de Combinatoire de Bordeaux and beyond, felt like a family from the day I joined, and still does. For years, Frédérique Bassino tried to guilt me into finishing this manuscript every single time we met, and her perseverance finally paid off. Antoine Genitrini found the time, in a very full schedule, to guide me through the HDR process.

At Nokia Bell Labs, I thank my colleagues Gérard Burnside, Johann Daigremont, Maria-Laura (Loreta) Maag, Dimitrios Miliotis, Pierre Escamilla, Liuba Tupikina, Davide Rinaldi and Chung Shue (Calvin) Chen for the good mood and the wit they bring to work, and for drawing me into new fields. I thank my managers, Lisa Zhang and Sean Kennedy, for their constant and generous support, and for allowing me to continue theoretical research and teaching.

Finally, I thank my family for passing on to me their taste for learning and teaching. I also thank the family of my brother-in-law, who never fails to ask when I will get my Fields Medal (soon, of course), and my lovely Pauline, who checks that I am alright when she finds me slowly pacing, deep concern painted on my face, happily lost in my maths.

2 Analytic combinatorics

Analytic combinatorics separates the combinatorial and the analytic parts. The combinatorial part decomposes the object under study into constructions that the *symbolic method* translates into generating function specifications. As Robert Sedgewick put it, “if you can specify it, you can analyze it!”. The analytic part then extracts asymptotics from these specifications, or establishes a limit law for an associated parameter. This chapter presents our toolbox: the symbolic method, the asymptotic extraction and the link with probabilities.

2.1 Symbolic method

Our references are Flajolet and Sedgewick (2009) and Bergeron, Labelle, and Leroux (1997). A generating function of a combinatorial family $\mathcal{A}$ equipped with a size function $|\cdot| : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$ is a series

$$\sum_{a \in \mathcal{A}} \gamma(|a|, z),$$

where the fixed function $\gamma(n, z)$ is the *type* of the generating function. The classical types are the ordinary generating functions, $\gamma(n, z) = z^n$, the exponential ones, $\gamma(n, z) = z^n/n!$ – both specializations of the cycle index series of species theory – and the Dirichlet series, $\gamma(n, z) = n^{-z}$, of analytic number theory.

The symbolic method is a dictionary translating combinatorial operations into relations between generating functions. Disjoint union corresponds to the sum, whatever the type. The product is where the types part ways, since the operation it encodes depends on γ : the Cartesian product for ordinary and for Dirichlet generating functions, under the respective conventions $|(a, b)| = |a| + |b|$ and $|(a, b)| = |a||b|$, and the *relabelled* Cartesian product for exponential generating functions, in which the labels of the two components are redistributed over their union in all possible ways. All other constructions follow from the sum and the product. Labeling removes the symmetries, which makes the dictionary both richer and simpler for exponential generating functions. All graphs in this manuscript therefore have labeled vertices. The same argument motivates the labeling of multigraph edges in Chapter 3.

New types can also be designed for a specific problem (Doubilet, Rota, and Stanley 1972). One starts from a combinatorial operation that the product of generating functions should encode, rich enough to carry a central feature of the problem, simple enough for a matching γ to exist. Robinson (1973) built such a type to enumerate directed acyclic graphs. Following Gessel (1995) we call these *graphic generating functions*; they are central to Chapter 6. With Sergey Dovgal and Vlady Ravelomanana, we introduced along the same lines a type suited to satisfiable 2-SAT formulas, presented in Chapter 7.

Enumeration by inversion. When a generating function A lies beyond direct application of the symbolic method, one can sometimes enumerate a more general family whose generating function B is accessible, find a relation expressing B in terms of A , and invert it. Inclusion-exclusion is the familiar instance, used e.g. in Section 6.1: distinguishing an arbitrary subset of the occurrences of a pattern gives $A(z, v + 1)$,

and the substitution $v = u - 1$ returns $A(z, u)$. Connected graphs are another, applied in Chapter 5: all graphs are immediate to count, $\text{MG} = \exp(\text{CMG})$, hence $\text{CMG} = \log \text{MG}$. The whole difficulty is choosing a family general enough for the relation to be expressible. We proceed this way for the simple graphs with constrained degrees of Section 4.3.

Choosing the right exact expression. An exact expression is not the end of the combinatorial work, Sedgewick's motto notwithstanding. Sometimes, more effort on the symbolic side provides a new expression with a simpler analysis. Connected multigraphs are a case in point, as discussed in Section 5.2: taking the logarithm of the generating function of all multigraphs gives an exact and elementary expression whose coefficients are dominated by cancellations, whereas discarding the tree and unicyclic components first yields an expression of similar shape whose dominant contribution is easier to locate. Sometimes an expression simplifies unexpectedly; the work then consists in understanding why. Cancellations are a reliable sign that a more direct combinatorial explanation stands behind the expression they leave, as for the positive multigraphs of Equation (5.3), where two factors merge into one. The decomposition one then finds usually reveals structure that the original derivation obscured. Nearly every exact expression in this manuscript has gone through this simplification process.

A construction set. The symbolic method is a construction set: its power comes from the combination of elementary constructors. The following chapters enumerate graph families constrained on their degrees, their subgraphs, their global structure, their orientations. Such constraints combine, so one may count connected regular graphs, or study the distribution of triangles in multigraphs all of whose vertex degrees are prime.

2.2 Asymptotic extraction

Some results of this manuscript are stated as *asymptotic expansions*, in the sense of Poincaré (1886). The general definition expands $f(n)$ along a scale of functions, each negligible with respect to the previous one. We only need the scale of powers of $n^{-1/2}$, and write $f(n) \sim \sum_{k \geq 0} c_k n^{-k/2}$ when, for every K ,

$$f(n) = \sum_{k < K} c_k n^{-k/2} + O(n^{-K/2}).$$

Such a series need not converge, and ours generally do not. An expansion stopping at the first term is what we call the asymptotics, or the asymptotic equivalent, of $f(n)$.

Whenever possible, we reduce the asymptotic analysis to a *black box theorem* stated at the right level of generality: as general as possible, but no more demanding than a direct proof of the special case would be. Such proofs are clearer and modular. Since applications rarely match the clean statement of a theorem, they are also easier to adapt to the variants that applications require. Four black box theorems cover most of the needs of this manuscript.

Large and semi-large powers for coefficients $[z^n]A(z)B(z)^N$ with n/N in a compact interval. Under some technical conditions, the saddle point of $B(z)^N/z^n$

depends only on n/N , and the coefficient is asymptotically Gaussian around it (Gardy 1995).

An Airy-type integral replaces the Gaussian when the saddle point approaches a singularity of B (Banderier et al. 2001). This is the tool of the phase transition of random graphs.

Hayman admissibility typically for entire functions, such as the generating function e^{e^z-1} of set partitions, and functions with an essential dominant singularity, such as $e^{z/(1-z)}$. Hayman (1956) isolated conditions under which the standard saddle point estimate holds, stable under many operations on generating functions.

Singularity analysis for generating functions with a dominant singularity ρ at finite distance, continuable beyond the circle of convergence outside a cone with vertex ρ opening away from the origin. The coefficients are read off the singular expansion: $F(z) \sim C(1 - z/\rho)^{-\alpha}$ gives $[z^n]F(z) \sim C\rho^{-n}n^{\alpha-1}/\Gamma(\alpha)$ (Flajolet and Odlyzko 1990).

Divergent series for generating functions of zero radius of convergence. The divergence is an opportunity, since coefficients growing that fast make the asymptotics of a composed or inverted series depend on a few terms only, extracted by elementary manipulations of formal power series. Foundational results are due to Bender (1975) and Wright (1970). Systematic tools now exist for the *factorially divergent series*, whose coefficients grow like $\Gamma(n + \beta)$ times a geometric factor (Borinsky 2018a), and for the faster-growing series of graph-like structures, such as the series $\sum_n 2^{\binom{n}{2}} z^n/n!$ of simple graphs (Dovgal and Nurligareev 2026).

Several of these theorems rest on two methods that we also use directly. The *Laplace method* gives the asymptotics of real integrals $\int A(x)e^{-n\Phi(x)} dx$, dominated by the neighborhood of the minimum of Φ ; it extends to complex-valued integrands, to several variables, and to full asymptotic expansions (Pemantle, Wilson, and Melczer 2024). The *saddle point method* transports it to complex contour integrals, Cauchy coefficient extractions in particular: the contour is deformed to cross the saddle points along the direction of steepest descent, where the integral becomes a real Laplace integral, and to leave their neighborhood for regions where the integrand is negligible (Bruijn 1981). Finding such a contour is something of an art; the estimates along it must be redone each time. The black box theorems are the recurring situations where this work has been done once and for all.

Sometimes the tools apply as they stand. Sometimes we adapt them: in Panafieu (2019) we designed a bivariate variant of Borinsky’s results, for the asymptotic expansion of connected graphs with n vertices and a number of edges proportional to n ; in Chapter 9 we adapt Hayman admissibility to limit laws. And sometimes precision demands opening the black box and reproving the theorem in the case at hand, at the cost of a less clean and less general proof. At the critical window of random graphs, semi-large powers is the adapted tool inside the window, but merging cleanly with the subcritical phase takes more work. In the digraphs of Chapter 6, the saddle point method is applied by hand and the singularity analysis is an ad hoc version rather than the black box. A simpler derivation, and perhaps a black box theorem behind it, may still be found.

Verification by computer. Some of our results rely on long ad hoc derivations or on technical applications of black box theorems. Arithmetic mistakes then accumulate during the writing and are time-consuming to correct. A solution we favor is to write code verifying all statements, symbolically or numerically, and to publish it alongside the article, so that readers can check the computations themselves. I wrote such code for the asymptotic expansion of connected graphs of large excess of Section 5.2 (Panafieu 2019), Sergey Dovgal for the structure of random digraphs of Section 6.3 (Dovgal et al. 2024), and Martijn Gösgens for the cluster graphs of Chapter 9 (Gösgens et al. 2024).

2.3 Marking a parameter

A parameter of a combinatorial family is recorded by a second variable. Writing $\chi(a)$ for the parameter of an object a – its number of edges, of cliques, of components – the bivariate generating function

$$A(z, u) = \sum_{a \in \mathcal{A}} u^{\chi(a)} \frac{z^{|a|}}{|a|!}$$

carries the joint distribution of size and parameter. It admits three readings, each answering a different probabilistic question.

Extracting a coefficient in u prescribes the parameter. The quantity $n![z^n u^k]A(z, u)$ counts the objects of size n with $\chi = k$, and the associated random model is the uniform distribution on them. The model $\text{MG}(n, m)$ of multigraphs with n vertices and m edges, used throughout Chapter 5, is of this kind.

Evaluating u at a fixed value replaces that conditioning by a random parameter. Normalized by $n![z^n]A(z, u)$, the weight $u^{\chi(a)}$ defines a distribution on the objects of size n that favors large values of χ when $u > 1$ and small ones when $u < 1$, an exponential tilting of the uniform model. For graphs the tilted models are the familiar ones: marking edges and substituting $u = p/(1 - p)$ produces $G(n, p)$, where each edge is present independently with probability p , while for multigraphs with labeled edges the substitution $u = p$ produces the model $\text{MG}(n, p)$ in which the multiplicities are independent Poisson variables. Chapter 3 states both formulas. Tilting makes independent what conditioning had correlated, which is why the analysis of Section 6.3 is carried out in the Poisson model. The same device applied to the size variable rather than to a marking variable is the Poissonization of Jacquet and Szpankowski (1998) and the Boltzmann model of Duchon et al. (2004).

Expanding around $u = 1$ gives the law of the parameter under the uniform model. Let X_n denote χ evaluated on a uniform object of size n . Its probability generating function is a ratio of coefficients,

$$\mathbb{E}(u^{X_n}) = \frac{[z^n]A(z, u)}{[z^n]A(z, 1)},$$

its factorial moments follow by differentiation,

$$\mathbb{E}(X_n(X_n - 1)\cdots(X_n - k + 1)) = \frac{[z^n]\partial_u^k A(z, 1)}{[z^n]A(z, 1)},$$

and the substitution $u = e^t$ turns it into the moment generating function $\mathbb{E}(e^{tX_n})$. The theorem of Curtiss (1942) converts asymptotics into a limit law: if $\mathbb{E}(e^{tX_n})$ converges, for every t in a neighborhood of 0, to the moment generating function of a random variable Y , then X_n converges to Y in distribution. One applies it to the normalized variable $(X_n - \mu_n)/\sigma_n$. When applicable, the quasi-powers theorem of Hwang (1998) is a particularly useful black box theorem encapsulating those calculations.

3 Graphs and multigraphs

This chapter presents our work with Gwendal Collet, Danièle Gardy, Bernhard Gittenberger and Vlado Ravelomanana (Collet et al. 2020), together with the multigraph model on which the manuscript rests. Its contribution is a generating function treatment of the number of copies of a small graph in a random multigraph: it recovers the Poisson limit law at the threshold of strictly balanced graphs, extends it to a complete asymptotic expansion, and introduces the *patchworks*, the tool by which several later chapters transfer results from multigraphs to simple graphs.

The dominant model in the literature is the simple graph, yet multigraphs have long been known to be easier to handle than simple graphs. In random graph theory, the configuration model (Bender and Canfield 1978; Bollobás 1980; Wormald 1999; van der Hofstad 2017a) and the Poissonian random graph (van der Hofstad 2017b) produce multigraphs natively; in enumeration, Flajolet, Knuth, and Pittel (1989) and Janson et al. (1993) derived their results on multigraphs first and transferred them afterwards. Our model is a variant of theirs.

3.1 Simple graphs, multigraphs and compensation factors

A *simple graph* has labeled vertices and unlabeled, unoriented edges, with no loop and no multiple edge. Throughout, n denotes the number of vertices and m the number of edges, and an (n, m) -graph is one with n vertices and m edges. The number of (n, m) -simple graphs is $\binom{\binom{n}{2}}{m}$. Since the vertices are labeled and the edges are not, we associate to a family $\mathcal{H}$ of simple graphs the generating function

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

so that all simple graphs have generating function

$$\text{SG}(z, w) = \sum_{n \geq 0} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!}. \quad (3.1)$$

Multigraphs allow loops and multiple edges. Two conventions coexist for them.

The convention of Flajolet, Knuth, and Pittel (1989) and Janson et al. (1993), which we call *compensated multigraphs*, keeps the labeling of simple graphs: labeled

vertices, unlabeled and unoriented edges, with loops and multiple edges now permitted. Each compensated multigraph G carries a weight, its *compensation factor*,

$$\kappa(G) = 2^{-\sum_u m_{u,u}} \prod_{u \leq v} \frac{1}{m_{u,v}!},$$

where $m_{u,v}$ is the number of edges between u and v , so that $m_{u,u}$ is the number of loops at u . The generating function of a family $\mathcal{H}$ of compensated multigraphs is

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

Our *multigraphs* have labeled vertices and *labeled, oriented* edges, and carry no compensation factor. Formally, an (n, m) -multigraph is a vertex set, taken to be $\{1, \dots, n\}$ when counting, together with an edge set $\{(u_1, v_1, 1), \dots, (u_m, v_m, m)\}$, the edge of label j being oriented from u_j to v_j . Since both vertices and edges are labeled, both are marked by exponential variables, and we associate to a family $\mathcal{F}$ of multigraphs the generating function

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

where edges are marked by $w/2$ rather than the more natural w . That factor makes the two conventions agree: labeling and orienting in all possible ways the edges of a compensated multigraph G produces $2^{m(G)} m(G)! \kappa(G)$ multigraphs, an integer even though $\kappa(G)$ is a fraction, and precisely the factor by which $(w/2)^{m(G)} / m(G)!$ differs from $\kappa(G) w^{m(G)}$. A family of compensated multigraphs and the family of multigraphs it generates therefore have the same generating function. Results transfer freely between the two conventions, so the choice is a matter of convenience. Ours has two advantages.

First, the number of multigraphs in a family is its cardinality. An (n, m) -multigraph is exactly a sequence $(u_1, v_1, \dots, u_m, v_m)$ of $2m$ integers from $\{1, \dots, n\}$, so there are n^{2m} of them, where the corresponding “number” of compensated multigraphs, the sum of their compensation factors, is $n^{2m} / (2^m m!)$.

Second, labeled edges are marked by an exponential variable, which puts the constructions of the symbolic method directly at our disposal. Consider the generating function of all multigraphs,

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

With our convention it is read off: a multigraph on n vertices is a set of labeled edges, each of them one of the n^2 oriented pairs of vertices, and a set of labeled objects marked by $w/2$ has generating function $e^{n^2 w/2}$. With compensation factors the same formula takes a few lines to prove and sits oddly in the symbolic method, since w is then an ordinary variable and no combinatorial operation corresponds to the exponential of an ordinary variable.

Simple graphs enter through the same correspondence. Orienting and labeling in all possible ways the m edges of a simple graph produces $2^m m!$ multigraphs, each contributing $(w/2)^m/m!$, that is w^m in total. The generating function $H(z, w)$ of a family $\mathcal{H}$ of simple graphs is therefore also that of the multigraphs obtained from $\mathcal{H}$ by labeling and orienting the edges in all possible ways. This suggests the strategy used repeatedly in the following chapters: to enumerate the simple graphs satisfying a given property, first count the multigraphs satisfying it, which is easier, then remove loops and multiple edges.

The *excess* of a multigraph is $k = m - n$, and the univariate generating function of the multigraphs of excess k in $\mathcal{F}$ is

$$F_k(z) = [y^k]F(z/y, y),$$

the substitution marking each multigraph by y^{m-n} ; the number of (n, m) -multigraphs in $\mathcal{F}$ is recovered as $2^m m! n! [z^n] F_{m-n}(z)$. A *random (n, m) -multigraph*, denoted $\text{MG}(n, m)$, is drawn uniformly among all (n, m) -multigraphs, so that the probability that it belongs to $\mathcal{F}$ is

$$\mathbb{P}_{G \in \text{MG}(n, m)}(G \in \mathcal{F}) = \frac{2^m m! n!}{n^{2m}} [z^n] F_{m-n}(z).$$

Random models. The uniform model $\text{MG}(n, m)$ is that of the multigraph process of Flajolet, Knuth, and Pittel (1989); Janson et al. (1993), where membership in $\mathcal{F}$ is a coefficient extraction, as above. In the *Poisson multigraph* $\text{MG}(n, p)$, the multiplicity of each of the n^2 oriented pairs of vertices is an independent Poisson variable of mean $p/2$, a homogeneous case of the Poissonian random graph of van der Hofstad (2017b), and membership is an evaluation,

$$\mathbb{P}(\text{MG}(n, p) \in \mathcal{F}) = e^{-n^2 p/2} n! [z^n] F(z, p), \quad (3.2)$$

as explained in Section 2.3. Conditioned on its number of edges, $\text{MG}(n, p)$ is $\text{MG}(n, m)$, so the two models are interchangeable for most purposes, as $G(n, p)$ and $G(n, m)$ are for simple graphs, where the formula reads $\mathbb{P}(G(n, p) \in \mathcal{F}) = (1 - p)^{\binom{n}{2}} n! [z^n] F(z, p/(1 - p))$. We use $\text{MG}(n, m)$ in this chapter and the next two, and the Poisson model, in its directed and its clique-conditioned versions, in Section 6.3 and Chapter 9.

3.2 Subgraphs

The distribution of the number of copies of a small graph in a random graph is a classical topic, opened by Erdős and Rényi (1960) and settled for strictly balanced graphs by Bollobás (1981a) and Karoński and Ruciński (1983) with the probabilistic method. This section presents the generating function treatment of Collet et al. (2020). Given a multigraph family $\mathcal{H}$ closed under isomorphism, an $\mathcal{H}$ -subgraph of G is a subgraph isomorphic to an element of $\mathcal{H}$, and $G[\mathcal{H}]$ denotes the number of them.

One distinguished subgraph. An (n, m) -multigraph with one distinguished $\mathcal{H}$ -subgraph is a copy from $\mathcal{H}$, a set of isolated vertices bringing the vertex count to

n , and a set of edges chosen among the n^2 oriented pairs bringing the edge count to m :

$$\text{MG}_{n,m}^{[\mathcal{H}]} = n!2^m m! [z^n w^m] H(z, w) e^z e^{n^2 w/2}. \quad (3.3)$$

Dividing Equation (3.3) by n^{2m} gives the mean number of $\mathcal{H}$ -subgraphs in a random (n, m) -multigraph. Two linear changes of variables, $[z^n]F(\alpha z) = \alpha^n [z^n]F(z)$, bring the saddle point to $(1, 1)$:

$$\mathbb{E}_{\text{MG}(n,m)}(G[\mathcal{H}]) = \frac{n!}{n^n} \frac{m!}{m^m} [z^n w^m] H\left(nz, \frac{2m}{n^2}w\right) e^{nz} e^{mw}. \quad (3.4)$$

When $\mathcal{H}$ is finite, $H(z, w)$ is a polynomial, the large powers theorem applies, and $\mathbb{E}(G[\mathcal{H}]) \sim H(n, 2m/n^2)$. For a single multigraph H together with its isomorphs, $H(z, w) = (w^{m(H)} z^{n(H)})/|\text{Aut}(H)|$, so that

$$\mathbb{E}_{\text{MG}(n,m)}(G[H]) \sim \frac{1}{|\text{Aut}(H)|} (2mn^{-2+1/d(H)})^{m(H)}, \quad d(H) = \frac{m(H)}{n(H)},$$

the same expression as for the simple graph $G(n, m)$. It tends to 0, to a constant or to infinity according as m is o , Θ or ω of $n^{2-1/d(H)}$. That is the appearance threshold of H when H is *balanced*, no subgraph of it being denser; otherwise a denser subgraph appears later, and the maximal density $d^*(H)$ of a subgraph governs the threshold. The first moment locates the threshold; proving that copies do appear above it requires a second moment, to which we return below.

For an infinite family the exact formula Equation (3.4) still holds, but the asymptotic analysis may change. For cycles, with $H(z, w) = \frac{1}{2} \log \frac{1}{1-zw}$, the large powers theorem gives $\mathbb{E}(G[\mathcal{H}]) \sim \frac{1}{2} \log \frac{1}{1-2m/n}$ for m/n in a compact subinterval of $(0, 1/2)$, and fails as m/n approaches $1/2$, where the saddle point meets the singularity: that collision is the birth of the giant component of Section 5.3.

Patchworks. Occurrences of a pattern are classically counted by inclusion-exclusion, which reduces the problem to counting objects with some distinguished occurrences, so that everything hinges on the ways copies of the pattern can overlap. For words the answer is known: the autocorrelation polynomial of Guibas and Odlyzko (1981), and the cluster method of Goulden and Jackson (1979), which performs the inclusion-exclusion over *clusters*, the connected chains of overlapping occurrences; see also Bassino et al. (2008). Our *patchworks* carry the idea to multigraphs.

A patchwork on $\mathcal{H}$ is a set of multigraphs, each isomorphic to an element of $\mathcal{H}$, that may share vertices and edges and whose union is itself a multigraph, so that two copies sharing an edge label share the whole edge. In graph theory vocabulary, it is a multigraph equipped with an $\mathcal{H}$ -covering in the sense of Harary (1970). Writing $n(P)$, $m(P)$ and $|P|$ for the numbers of vertices, edges and copies of a patchwork P , their generating function is

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

The *connected patchworks* $\text{CPatch}^{(\mathcal{H})}$, those whose union is connected, are the counterpart of the clusters of Goulden and Jackson (1979), and the exponential formula gives $\text{Patch}^{(\mathcal{H})} = \exp(\text{CPatch}^{(\mathcal{H})})$. Unlike the autocorrelation polynomial, they resist computation: we have no expression for them beyond the cases of loops and double edges (Panafieu 2019). Fortunately, as we shall see in Equation (3.6), only a computable truncation of $\text{CPatch}^{(\mathcal{H})}$ is needed under the right circumstances.

Patchworks of two copies give the second moment through Equation (3.3), and patchworks of all sizes give the whole distribution. Let u mark the number of $\mathcal{H}$ -subgraphs in the generating function $\text{MG}_{n,m}^{(\mathcal{H})}(u)$ of (n, m) -multigraphs. Substituting $u = v + 1$ distinguishes an arbitrary subset of the $\mathcal{H}$ -subgraphs, and that subset is a patchwork, so Equation (3.3) applies with $\mathcal{H}$ replaced by the family of patchworks:

$$\text{MG}_{n,m}^{(\mathcal{H})}(v + 1) = n!2^m m! [z^n w^m] \text{Patch}^{(\mathcal{H})}(z, w, v) e^z e^{n^2 w/2}. \quad (3.5)$$

This is inclusion-exclusion in generating function form. The case of most frequent interest, multigraphs containing no $\mathcal{H}$ -subgraph, is obtained at $v = -1$.

Strictly balanced subgraphs. A multigraph H is *strictly balanced* if its density exceeds that of every proper subgraph. We recover the theorem of Bollobás (1981a) and Karoński and Ruciński (1983): at its appearance threshold, the number of copies of a strictly balanced H is asymptotically Poisson. A patchwork of copies of H has density exactly $d(H)$ when its copies are pairwise disjoint, and strictly greater as soon as two copies overlap, so at the threshold $m = \Theta(n^{2-1/d(H)})$ only the disjoint patchworks contribute to leading order. Their generating function is $e^{uH(z,w)}$, hence

$$\begin{aligned} \text{MG}_{n,m}^{(\mathcal{H})}(u) &\approx n!2^m m! [z^n w^m] e^{(u-1)H(z,w)} e^z e^{n^2 w/2} \\ &\sim n^{2m} e^{(u-1)H(n, 2m/n^2)} \end{aligned}$$

by the large powers theorem, $H(n, 2m/n^2)$ being bounded at the threshold; the right-hand side is n^{2m} times the probability generating function of a Poisson law of parameter $H(n, 2m/n^2)$. The rigorous argument is in Collet et al. (2020).

The same reasoning yields the complete asymptotic expansion at the threshold. For a prescribed error term $O(n^{-\alpha})$ with $\alpha > 0$, the connected patchworks carrying more than a bounded number of copies of H are too dense to contribute, so a finite family $\text{Patch}^{(\mathcal{H}, \alpha)}$ suffices:

$$\text{MG}_{n,m}^{(\mathcal{H})}(u) = n^{2m} e^{(u-1)H(n, 2m/n^2)} (\text{Patch}^{(\mathcal{H}, \alpha)}(n, 2m/n^2) + O(n^{-\alpha})). \quad (3.6)$$

The difficulty of triangles. Even triangles are hard: we do not know how to count sets of possibly overlapping triangles with given numbers of vertices, edges and triangles, and no recurrence or implicit equation is known. By Equation (3.5) this is equivalent to counting (n, m) -multigraphs by their number of triangles. We believe the difficulty to be of a different nature from what the usual classification of generating functions expresses: there might be no algorithm computing these numbers in time polynomial in n , though proving it seems out of reach. A heuristic explanation is that loops, double edges and degree constraints are conditions on one or two vertices, whereas a triangle is the first condition on three vertices.

Triangle-free (n, m) -graphs connect to our approach: for $m \leq n^{3/2-\varepsilon}$, their asymptotic number is the exponential of a finite sum whose number of terms grows as ε decreases (Stark and Wormald 2018; Mousset et al. 2020), a truncation of $\text{CPatch}(n, 2m/n^2, -1)$ for the family of triangles. Whether the full series converges is an open question of Jenssen, Perkins, and Potukuchi (2025).

4 Degree constraints

This chapter presents two papers: our work with Emma Caizergues (Caizergues and Panafieu 2023), and a single-author paper (Panafieu 2024). Their contributions are an exact expression for the number of simple graphs with constrained degrees and the complete asymptotic expansion of the number of k -regular graphs, settling a conjecture of McKay (1983), repeated in McKay and Wormald (1991).

Prescribing the degrees is a natural constraint to impose on a graph, as the degree distribution of real networks typically differs from the one of a $G(n, p)$ graph (van der Hofstad 2017b). This is what the configuration model exists for. We start with multigraphs, where half-edges make the problem elementary. Their generating function by excess, derived in our PhD, is used in Chapter 5 and Chapter 6.

4.1 Multigraphs with constrained degrees

The *degree* of a vertex is the number of edges that contain it, a loop counting twice. Let $\delta = (\delta_d)_{d \geq 0}$ be a sequence of weights and

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

Each vertex of degree d receives the weight δ_d , and the weight of a multigraph is the product of the weights of its vertices. The sum of the weights of all (n, m) -multigraphs is then

$$\text{MG}_{n,m}^{(\delta)} = (2m)! [x^{2m}] \Delta(x)^n. \quad (4.1)$$

The proof rests on *half-edges*. Cut each edge (u, v, ℓ) into the labeled half-edges $(u, 2\ell - 1)$ and $(v, 2\ell)$. Since the pairing of $2\ell - 1$ with 2ℓ is fixed in advance, this is a bijection between (n, m) -multigraphs and ordered partitions of $\{1, \dots, 2m\}$ into n possibly empty sets: the set of index u records the half-edges at u , and the edge of label ℓ runs from the vertex holding $2\ell - 1$ to the one holding 2ℓ . A vertex of degree d receives a set of size d and the weight δ_d , so $\Delta(x)^n$ is the exponential generating function of these weighted ordered partitions.

Two choices of weights are of particular interest. When the δ_d lie in $\{0, 1\}$, we count multigraphs whose degrees belong to the set $D = \{d : \delta_d = 1\}$ of allowed degrees, and we write $\text{MG}^{(D)}$ for $\text{MG}^{(\delta)}$. When the δ_d are nonnegative and sum to 1, the weight of a multigraph divided by $\text{MG}_{n,m}^{(\delta)}$ is a probability distribution on (n, m) -multigraphs: the *configuration model* (Bender and Canfield 1978; Bollobás 1980; Wormald 1999) with degrees drawn independently from δ , conditioned on the number

of edges. The half-edge argument is borrowed from that model, whose difficulty lies precisely in conditioning away the loops and multiple edges it produces natively.

Regular multigraphs. Taking $D = \{d\}$ in Equation (4.1), the number of d -regular (n, m) -multigraphs is $(2m)! [x^{2m}] (x^d/d!)^n$. It vanishes unless $2m = nd$, the sum of the degrees being twice the number of edges, and otherwise equals the multinomial coefficient $(2m)!/(d!)^n$: there are $(3n)!/6^n$ cubic multigraphs on n vertices. Counting cubic *simple* graphs is another matter. Read (1959) proved that their number on $2n$ vertices satisfies a linear recurrence with polynomial coefficients, and the shortest known form of it, due to Wormald (1979), is of order six, with coefficients of degree five in n . We return to regular simple graphs in Section 4.4.

Asymptotics. For $2m/n$ in a compact subinterval of $(\min D, \max D)$, the asymptotics of Equation (4.1) follow from the large powers theorem, or from its periodic variant when D is contained in an arithmetic progression; with a numerically estimated saddle point, this gives for instance the asymptotic number of (n, m) -multigraphs all of whose degrees are prime.

Generating function by excess. Combining Equation (4.1) with the substitution $(z, w) \mapsto (z/y, y)$ of Section 3.1, the multigraphs of excess k with vertex weights δ have generating function

$$\text{MG}_k^{(\delta)}(z) = \sum_n (2(n+k)-1)!! [x^{2(n+k)}] \Delta(x)^n \frac{z^n}{n!},$$

with $(2m-1)!! := (2m)!/(2^m m!)$. Writing $(2m-1)!! = 2^m \Gamma(m+1/2)/\sqrt{\pi}$ and using $\sum_{n \geq 0} \Gamma(n+\beta) u^n / n! = \Gamma(\beta)(1-u)^{-\beta}$, the sum over n collapses:

$$\text{MG}_k^{(\delta)}(z) = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) [x^{2k}] \frac{1}{(1 - 2zx^{-2}\Delta(x))^{k+1/2}}. \quad (4.2)$$

We derived this expression in Panafieu (2014) and put it to use in Panafieu (2019). It calls for two remarks. Its nature as a formal object depends on the support of δ . When $\delta_0 = \delta_1 = \delta_2 = 0$, the series $x^{-2}\Delta(x)$ has valuation at least 1 in x , and Equation (4.2) is an identity between formal power series in z . Otherwise the valuation is nonpositive and the extraction in x may be undefined; one should then extract in z first, replacing $[x^{2k}]$ by $[x^{2k}]n![z^n]$, which is valid in all cases. Second, the prefactor $2^k \Gamma(k+1/2)/\sqrt{\pi}$ equals $(2k-1)!!$ for $k \geq 0$ and $(-1)^{|k|}/(2|k|-1)!!$ for $k < 0$, so the expression holds at negative excess as well.

Zero-dimensional field theory. The same computation is known in physics, under a different vocabulary. The perturbative expansion of the partition function

$$Z(y) = \frac{1}{\sqrt{2\pi y}} \int e^{(-x^2/2 + z\Delta(x))/y} dx \quad (4.3)$$

of a zero-dimensional quantum field theory is a sum over *Feynman diagrams*, compensated multigraphs with unlabeled vertices, Wick's theorem counting the pairings of half-edges (Bessis, Itzykson, and Zuber 1980; Etingof 2002; Borinsky 2018b). Each vertex contributes a factor y^{-1} and each edge a factor y , so y marks the excess: the

loop expansion of the physicists is our enumeration by excess, and $\text{MG}_k^{(\delta)}(z)$ is the coefficient of y^k in Equation (4.3).

In that treatment, Borinsky (2017) reaches Equation (4.2) by the same resummation, under the hypothesis that no vertex has degree 0, 1 or 2 – the case where, as noted above, no care with formal objects is needed. He then applies Lagrange inversion to turn it into an extraction on the implicitly defined series $x(y)$ solving $y^2/2 = x^2/2 - z\Delta(x)$, and obtains the asymptotics in the excess by singularity analysis of that series.

Lagrange inversion and large powers. We prefer the opposite direction. Lagrange inversion is precisely the passage between an implicitly defined series and a large power: for $A(z) = z\Phi(A(z))$,

$$[z^n]F(A(z)) = \frac{1}{n}[u^{n-1}]F'(u)\Phi(u)^n,$$

and the identity reads as well from right to left. Here $x(y) = y(1 - 2zx^{-2}\Delta(x))^{-1/2}$ evaluated at $x = x(y)$, so the two forms are transforms of one another, and Equation (4.2) is a large power extraction $[x^{2k}]Q(x)^{k+1/2}$ with $Q(x) = (1 - 2zx^{-2}\Delta(x))^{-1}$, at the fixed ratio $2k/k$. The large powers theorem of Section 2.2 delivers the complete asymptotic expansion through the Laplace method, requires no analytic continuation into a Δ -domain, and never asks for the implicit function to be constructed; when the saddle point meets a singularity of Q , its semi-large variant takes over. Whenever a problem presents itself as singularity analysis on an implicitly defined series, it is worth asking whether Lagrange inversion turns it into a large power.

Minimum degree. Define $\exp_{\geq r}(x) = \sum_{d \geq r} x^d/d!$. Multigraphs with minimum degree at least 3 and excess k have generating function

$$\text{MG}_k^{(\geq 3)}(z) = (2k-1)!![x^{2k}] \frac{1}{(1 - 2zx^{-2} \exp_{\geq 3}(x))^{k+1/2}}. \quad (4.4)$$

Since $\exp_{\geq 3}(x)$ has valuation 3, the coefficient of z^n on the right has valuation n in x , so the extraction $[x^{2k}]$ annihilates every term with $n > 2k$: $\text{MG}_k^{(\geq 3)}(z)$ is a polynomial in z of degree exactly $2k$, reflecting the fact that a multigraph of excess k and minimum degree 3 has at most $2k$ vertices, the extremal case being the cubic ones. These multigraphs are the *kernels* of Chapter 5, where they unlock the enumeration of connected multigraphs.

4.2 Subgraphs in multigraphs with degree constraints

This section combines degree constraints and subgraph counting. In Equation (4.1) a vertex of degree d contributes the weight δ_d . Distinguish now a copy of $\mathcal{H}$. A vertex of the copy already carries d half-edges inside it, and the j further half-edges it receives from outside contribute $x^j/j!$ and raise its degree weight to δ_{d+j} , so that vertex contributes $\sum_{j \geq 0} \delta_{d+j} x^j/j! = \Delta^{(d)}(x)$. The degree weights seen by the copy are the derivatives of Δ . Writing $H(z, w, \delta)$ for the generating function of $\mathcal{H}$ with each vertex of degree d marked by δ_d , and $\partial\Delta(x) = (\Delta(x), \Delta'(x), \Delta''(x), \dots)$, the

generating function of multigraphs with vertex weights δ and one distinguished $\mathcal{H}$ -subgraph is

$$\sum_{m \geq 0} (2m)! [x^{2m}] H(z, w, \partial \Delta(x)) e^{z \Delta(x)} \frac{w^m}{2^m m!}, \quad (4.5)$$

the subgraph construction of Section 3.2 and the degree weights of Section 4.1 standing side by side in one formula.

The asymptotics rest on a device we call *linearization*, which recurs throughout the manuscript. Evaluating $\sum_n a_n [x^n] F(x)$ is immediate when $a_n = \alpha^n$: the sum is $F(\alpha)$. When a_n has no such closed form but admits an integral representation $a_n = \int g(t) t^n dt$, the same holds under the integral sign, $\sum_n a_n [x^n] F(x) = \int g(t) F(t) dt$. Here $a_{2m} = (2m)! / (2^m m!) = (2m-1)!!$ is the $2m$ -th moment of the standard Gaussian, and the odd moments vanish, so Equation (4.5) becomes

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} H(z, w, \partial \Delta(t\sqrt{w})) e^{z \Delta(t\sqrt{w})} e^{-t^2/2} dt.$$

The Laplace method applies to that integral, the saddle point method to the coefficient extraction in z and w (Collet et al. 2020). Physicists know the same device as the Hubbard–Stratonovich transformation. We use it again to invert the passage from simple graphs to multigraphs in Section 4.3, to turn the expression for regular graphs into a Gaussian integral of fixed dimension in Section 4.4, and on the Hadamard product of Chapter 6.

To illustrate the technique, we also recover the asymptotic mean number of cycles of a given length in a multigraph with power-law degree weights of exponent in $(2, 3)$. It was first obtained by Bianconi and Marsili (2005), and follows from the much broader study of subgraphs in configuration models with such degrees (van der Hofstad, van Leeuwen, and Stegehuis 2018; 2021).

4.3 Simple graphs with constrained degrees

Simple graphs with degree constraints can be enumerated with the patchworks of Section 3.2, removing loops and double edges. That was our first approach, until unexpected cancellations revealed the simpler path presented below.

The classical expression for the number of simple graphs with prescribed degrees $d_1, \dots, d_n$ is

$$[x_1^{d_1} \dots x_n^{d_n}] \prod_{1 \leq i < j \leq n} (1 + x_i x_j), \quad (4.6)$$

one variable per vertex. It is the starting point of the Fourier-inversion approach to regular graphs of McKay and Wormald (1991); but the number of variables grows with n , which complicates the asymptotic analysis. Our aim is an expression in which the number of variables depends on the degrees allowed, not on n .

A more general model. The strategy is the enumeration by inversion of Section 2.1, and everything turns on the choice of the accessible family. Multigraphs

with weighted degrees will not do: collapsing a bundle of parallel edges into one leaves no trace in a generating function that only counts edges. We therefore enumerate a model in which edges carry weights, where the passage from simple graphs to multigraphs becomes a substitution of variables, and can be inverted.

Edge weights. An *edge-weighted multigraph* is a sequence $(V, E_1, E_2, \dots)$ where E_j is a set of m_j labeled oriented edges of weight j , each weight having its own label set; loops and multiple edges are allowed, and the degree of a vertex is the sum of the weights of its incident edges, loops counted twice. Edges of weight j are marked by $w_j/2$, and the generating function of a family $\mathcal{F}$ with degree weights δ is

$$F(z, \mathbf{w}, \delta) = \sum_{G \in \mathcal{F}} \prod_{u \in V(G)} \delta_{\deg(u)} \prod_{j \geq 1} \frac{(w_j/2)^{m_j(G)}}{m_j(G)!} \frac{z^{n(G)}}{n(G)!}.$$

Cutting each edge of weight j into two half-edges of weight j , marked by x_j , a vertex of degree d receives a set of half-edges whose weights sum to d . Writing $\Xi(y, \mathbf{x}) = \sum_{j \geq 1} x_j y^j$, the generating function of a vertex with its half-edges is

$$P_{\text{WMG}}(\mathbf{x}, \delta) = \sum_{d \geq 0} \delta_d [y^d] e^{\Xi(y, \mathbf{x})} = \Delta(y, \delta) \square_{y=1} e^{\Xi(y, \mathbf{x})},$$

where $\square$ denotes the *exponential Hadamard product*,

$$\sum_n a_n \frac{y^n}{n!} \square_y \sum_n b_n \frac{y^n}{n!} = \sum_n a_n b_n \frac{y^n}{n!}, \quad (4.7)$$

and $\square_{y=1}$ its evaluation at $y = 1$. This is a polynomial when δ has finite support. The half-edge argument of Equation (4.1) then gives, with $\mathbf{m} = (m_1, m_2, \dots)$ and $(2\mathbf{m})! = (2m_1)!(2m_2)!\dots$,

$$\begin{aligned} \text{WMG}(z, \mathbf{w}, \delta) &= \sum_{\mathbf{m}} (2\mathbf{m})! [x^{2\mathbf{m}}] e^{z P_{\text{WMG}}(\mathbf{x}, \delta)} \prod_{j \geq 1} \frac{(w_j/2)^{m_j}}{m_j!} \\ &= e^{x_1^2/2 + x_2^2/2 + \dots} \square_{\mathbf{x}=1} e^{z P_{\text{WMG}}\left(\left(w_j^{1/2} x_j\right)_{j \geq 1}, \delta\right)}, \end{aligned} \quad (4.8)$$

the second form using the exponential Hadamard product in several variables, defined coefficientwise as in Equation (4.7), so that

$$e^{x_1^2/2 + x_2^2/2 + \dots} \square_{\mathbf{x}=1} F(\mathbf{x}) = \sum_{\mathbf{m}} (2\mathbf{m} - 1)!! [x^{2\mathbf{m}}] F(\mathbf{x}),$$

the last line being the linearization of Section 4.2 read backwards.

Two inversions. Two constraints separate simple graphs from these multigraphs; each is removed by an inversion. Loops first: a loop of weight j adds $2j$ to the degree of its vertex, and marking a distinguished subset of the loops is a matter of replacing $\Xi(y, \mathbf{x})$ by $\Xi(y, \mathbf{x}) + v\Xi(y^2, \mathbf{w})/2$ in the vertex polynomial; inclusion-exclusion, $v = -1$, yields loopless edge-weighted multigraphs. Multiple edges next. A loopless edge-weighted multigraph is obtained from an edge-weighted *simple* graph

by replacing each edge of weight j by a set of labeled edges whose weights sum to j , so the two generating functions differ by the change of edge variables $w_j \mapsto v_j(\mathbf{w}) = [y^j] \log(1 + \Xi(y, \mathbf{w}))$. Inverting it and simplifying, the vertex polynomial of edge-weighted simple graphs is

$$P_{\text{WSG}}(\mathbf{x}, \mathbf{w}, \boldsymbol{\delta}) = \Delta(y, \boldsymbol{\delta}) \square_{y=1} \frac{e^{\Xi(y, \mathbf{x})}}{\sqrt{1 + \Xi(y^2, \mathbf{w})}},$$

so that $\text{WSG}(z, \mathbf{w}, \boldsymbol{\delta})$ has the form Equation (4.8) with P_{WMG} replaced by P_{WSG} and w_j by $v_j(\mathbf{w})$ (Caizergues and Panafieu 2023).

Simple graphs. Simple graphs are edge-weighted simple graphs all of whose edges have weight 1: setting $\mathbf{w} = (w, 0, 0, \dots)$ gives $v_j(\mathbf{w}) = h_j(w) := -(-w)^j/j$ and

$$\begin{aligned} \text{SG}(z, w, \boldsymbol{\delta}) &= e^{x_1^2/2 + x_2^2/2 + \dots} \square_{\mathbf{x}=\mathbf{1}} e^{z P_{\text{SG}}\left((h_j(w)^{1/2} x_j)_{j \geq 1}, w, \boldsymbol{\delta}\right)}, \\ P_{\text{SG}}(\mathbf{x}, w, \boldsymbol{\delta}) &= \Delta(y, \boldsymbol{\delta}) \square_{y=1} \frac{e^{\Xi(y, \mathbf{x})}}{\sqrt{1 + wy^2}}. \end{aligned} \tag{4.9}$$

The square roots $h_j(w)^{1/2}$ are only apparent, since the Hadamard product with $e^{\sum x_j^2/2}$ retains only the even powers of each x_j . When $\boldsymbol{\delta}$ has finite support S , only the variables $x_1, \dots, x_{\max S}$ occur: the number of variables is the largest allowed degree.

D -finiteness. Read (1959) proved that the number of 3-regular graphs satisfies a linear recurrence with polynomial coefficients, the one described in Section 4.1, and Read and Wormald (1980) reached 4-regular graphs. Goulden, Jackson, and Reilly (1983) conjectured, and Gessel (1990) proved, that the generating function of graphs whose degrees lie in a finite set is D -finite, using the closure of D -finite series under Hadamard product and specialization (Stanley 1980; Lipshitz 1988; Zeilberger 1990). Our expression makes this constructive: when $\boldsymbol{\delta}$ has finite support, $e^{z P_{\text{SG}}}$ is the exponential of a polynomial, hence D -finite; $\sum_m (2m-1)!! x_j^{2m}$ is D -finite since $(2m-1)!!$ is P-recursive; and Equation (4.9) is a Hadamard product of the two, evaluated at $\mathbf{x} = \mathbf{1}$. As a method of computing the differential equation, however, this route is inefficient compared with the creative telescoping of Chyzak, Mishna, and Salvy (2005), who made Gessel's proof effective, and of Chyzak and Mishna (2026), who obtained the equations of k -regular graphs up to $k = 7$.

Bipartite graphs. The construction extends to bipartite graphs, with two vertex types and two degree weight sequences $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$: the Hadamard product is taken over the edge variables shared by the two sides, and the same inversions give a generating function that is D -finite whenever both sequences have finite support. Bipartite multigraphs with degree constraints are also hypergraphs with constrained vertex degrees and hyperedge sizes, and nonnegative integer matrices with prescribed row and column sums; the connection with the *balanced matrices* of Dougherty-Bliss et al. (2025) will be explored in the journal version of Caizergues and Panafieu (2023).

4.4 Regular graphs

This section presents results from Panafieu (2024). The asymptotic enumeration of regular graphs has a history as long as the exact one, surveyed by Wormald (1999)

and Wormald (2018). The dominant term is due to Bender and Canfield (1978), and independently to Bollobás (1980), whose probabilistic proof introduced the configuration model:

$$\text{SG}_n^{(k)} \sim e^{-(k^2-1)/4} \frac{(nk)!}{(nk/2)! 2^{nk/2} k!^n},$$

the fraction being the number of k -regular multigraphs of Section 4.1 divided by the $2^m m!$ labelings and orientations of their edges, and the exponential the limit probability that the configuration model produces a simple graph. McKay (1983) then conjectured, by numerical extrapolation, the first three terms of an expansion in powers of n^{-1} of the ratio between the two sides, a conjecture repeated in McKay and Wormald (1991). An expansion can in principle be read off a P-recursive sequence (Zeilberger 2008; Kauers 2011), but that route needs the recurrence, which restricts it to $k \leq 7$ by Chyzak and Mishna (2026), and rests on a theory (Birkhoff and Trjitzinsky 1932) whose completeness has been questioned (Wong and Li 1992). For general k the conjecture remained open. Confirming it, with all the remaining terms, is the result of this section.

A formula in k variables. Regular graphs are the case $\delta = (0, \dots, 0, 1)$ of Equation (4.9). Introducing the polynomial

$$P_k(\mathbf{x}) = [y^k] \frac{e^{\sum_{j=1}^k (-1)^{(j+1)/2} x_j y^j / \sqrt{j}}}{\sqrt{1+y^2}},$$

the generating function of k -regular graphs is

$$\text{SG}^{(k)}(z) = e^{x_1^2/2 + \dots + x_k^2/2} \square_{\mathbf{x}=1} e^{z P_k(\mathbf{x})}.$$

The point of the formula is its number of variables: k , against n in Equation (4.6). Linearizing the Hadamard product as in Section 4.2 turns it into a Gaussian integral over $\mathbb{R}^k$, of fixed dimension:

$$\text{SG}_n^{(k)} = n! [z^n] \text{SG}^{(k)}(z) = \frac{1}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} P_k(\mathbf{x})^n e^{-(x_1^2 + \dots + x_k^2)/2} d\mathbf{x}. \quad (4.10)$$

Asymptotic expansion. Linear changes of variables bring Equation (4.10) to the form $\int_{\mathbb{R}^k} A(n^{-1/2}, \mathbf{x}) e^{n\Phi(\mathbf{x})} d\mathbf{x}$, where Φ has a unique, non-degenerate critical point at $\mathbf{x} = \mathbf{0}$. Expanding A in powers of $n^{-1/2}$ and applying the multivariate Laplace method (Pemantle, Wilson, and Melczer 2024) to each term gives the expansion to any order; the odd powers cancel, and

$$\begin{aligned} \text{SG}_n^{(k)} &\approx e^{-(k^2-1)/4} \frac{(nk)!}{(nk/2)! 2^{nk/2} k!^n} (1 + c_1(k)n^{-1} + c_2(k)n^{-2} + \dots), \\ c_1(k) &= -\frac{k^4 - 2k^2 + 3k - 1}{12k}, \\ c_2(k) &= \frac{k^8 - 16k^6 + 6k^5 + 50k^4 + 36k^3 - 239k^2 + 234k - 71}{288k^2}, \end{aligned}$$

each $c_r(k)$ being a Laurent polynomial in k . This confirms the conjecture of McKay (1983) and supplies every further term. The coefficients are explicit and computable, and not pretty; simplifying them, and finding for them a combinatorial interpretation in the spirit of Monteil and Nurligareev (2021), are open questions. Connected regular graphs come along for the ride: Bollobás (1981b) and Wormald (1981) showed that a random k -regular graph is connected with high probability, so the dominant terms coincide, and we transferred the full expansion in Panafieu (2024) through the theory of divergent series of Wright (1970) and Bender (1975) recalled in Section 2.2.

Concurrent work. Shortly after Caizergues and Panafieu (2023), Evnin and Horinouchi (2024) obtained the same k -dimensional Gaussian integral by the methods of statistical physics, with a saddle point analysis of the dominant term but no expansion. More recently, Isaev and McKay (2025) proved the existence of an expansion and computed several terms, by a cumulant expansion applied to the Fourier inversion representation, in a framework covering far more general degree sequences. The structural difference between the approaches lies in the dimension of the integral: the Fourier inversion of Equation (4.6) integrates over n variables, ours over k , and the Laplace method applies to the latter to all orders. Combinatorially, we count k -regular graphs by deformation from k -regular multigraphs, which are immediate to enumerate, and the inversion encodes the ways loops and multiple edges overlap, which previous inclusion-exclusion arguments did not track.

5 Connected graphs and the phase transition

This chapter presents the content of (Panafieu 2019) and work in progress with Sergey Dovgal and Alexandros Singh. Its contributions are the complete asymptotic expansion of the number of connected graphs whose excess grows linearly with the number of vertices, the main result of the chapter; an explicit expression for the polynomials of Wright at fixed excess; a short route to the critical window through a black box theorem, designed during our PhD to reach models beyond graphs; and a new exact expression for all multigraphs, from which we hope to extract the statistics of the giant component. Connectivity is understood weakly: the orientation of the edges plays no role.

Since a multigraph is a set of connected multigraphs, $\text{MG}(z, w) = \exp(\text{CMG}(z, w))$ and, by inversion,

$$\text{CMG}(z, w) = \log \sum_{n \geq 0} e^{n^2 w/2} \frac{z^n}{n!}. \quad (5.1)$$

The argument of the logarithm has zero radius of convergence in z for every $w > 0$, so no analytic property of the logarithm is available. Expanding it as a series produces an alternating sum whose dominant contribution is hidden by cancellations: the first term counts all multigraphs, which are typically far from connected. Flajolet, Salvy, and Schaeffer (2004) nonetheless took this expression as the starting point of the complete asymptotic expansion of connected graphs of fixed excess. We follow a different route, through cores and kernels, opened by Wright (1978) and later taken by Pittel and Wormald (2005); it produces expressions whose dominant term is visible.

That route carries over to an excess growing linearly with the number of vertices, the regime of Section 5.2.

Trees and unicycles. Connected multigraphs of excess -1 are the *trees* and those of excess 0 the *unicycles*. The univariate generating functions $T(z)$, $U(z)$ and $V(z)$ of rooted trees, unrooted trees and unicycles have the classical expressions

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

and apply verbatim to our model. Since a tree has one edge less than vertices and a unicycle as many, the bivariate generating functions are $U(z, w) = U(zw)/w$ and $V(z, w) = V(zw)$. Their asymptotics follow from singularity analysis and the expansion $T(z) = 1 - \sqrt{2}\sqrt{1 - ez} + O(1 - ez)$ at the dominant singularity $z = e^{-1}$.

5.1 Fixed excess

Wright introduced the two central objects of this section (Wright 1977; 1978; 1980). A *core* is a multigraph of minimum degree at least 2 and a *kernel* one of minimum degree at least 3. The core of a multigraph is obtained by repeatedly deleting the vertices of degree at most 1, which erases the tree components and prunes the trees hanging from the rest. Its kernel is obtained from the core by suppressing the vertices of degree 2, merging their two incident edges into one, and discarding the components reduced to a single loop, which come from the unicycles. Both operations preserve the excess if the multigraph had no tree component. A multigraph without tree or unicycle component is connected if and only if its kernel is, and a kernel of excess k has at most $2k$ vertices and $3k$ edges, since the sum of its degrees is $2m \geq 3n$: there are finitely many kernels of a given excess, and their generating function is the polynomial $\text{MG}_k^{(\geq 3)}(z)$ of Equation (4.4). Connected kernels of excess k have generating function

$$\text{CMG}_k^{(\geq 3)}(z) = [y^k] \log \sum_{\ell \geq 0} \text{MG}_\ell^{(\geq 3)}(z) y^\ell,$$

a polynomial of degree exactly $2k$ whose leading coefficient counts the connected cubic multigraphs on $2k$ vertices. Hanging a rooted tree at each vertex and replacing each edge by a path of rooted trees, of generating function $1/(1 - T(z))$, reverses the passage to the kernel, so that

$$\text{CMG}_k(z) = \text{CMG}_k^{(\geq 3)} \left(\frac{T(z)}{1 - T(z)} \right) \frac{1}{(1 - T(z))^k} = \frac{P_k(T(z))}{(1 - T(z))^{3k}}, \quad (5.2)$$

where P_k is a polynomial. The generating function of connected multigraphs of fixed excess is thus a rational function of the tree function, and singularity analysis applies through the expansion of $T(z)$ at e^{-1} , the dominant term coming from the pole of highest order, that is, from the cubic kernels. This analysis is due to Wright (1980), who worked with simple graphs and characterized the sequence of polynomials P_k by a differential recurrence. One of our contributions in Panafieu (2019) is an explicit expression for those polynomials in the multigraph case, through Equation (4.4),

together with its transfer to simple graphs by the patchworks of loops and double edges of Section 3.2.

Since Equation (5.2) is a rational function of $T(z)$, singularity analysis would turn that expression into a complete asymptotic expansion of $\text{CMG}_{n,n+k}$ at fixed k . We have not carried out the computation, since the expansion is already known: Flajolet, Salvy, and Schaeffer (2004) obtained it directly, linearizing the quadratic exponent of $(1+w)^{\binom{n}{2}}$ with a Gaussian integral as in Section 4.2, an identity that goes back to Mahler (1940), followed by a delicate analysis. We reuse the identity in Section 6.3 to obtain the typical structure of random digraphs.

5.2 Large excess

This section carries the main result of the chapter: the complete asymptotic expansion of the number of connected graphs whose excess grows linearly with the number of vertices.

Bender, Canfield, and McKay (1990) were the first to reach this range, by a technical argument starting from Wright's differential recurrence for the polynomials P_k . Two simpler proofs followed. Pittel and Wormald (2005) built on an insight of Łuczak (1992) to prove that when $k = \Theta(n)$ almost all cores without isolated cycles are connected, so that the asymptotics of connected graphs are those of *positive* graphs, whose components all have positive excess. Working at the level of the counting sequences rather than of the generating functions, they then enumerated cores, removed the isolated cycles and hung rooted trees on the vertices. A probabilistic proof was derived by van der Hofstad and Spencer (2006), analyzing the breadth-first exploration of a random graph and estimating the probability that the queue of discovered vertices never empties before all vertices are reached.

We follow the strategy of Pittel and Wormald, in the generating function setting. The generating function of positive multigraphs of excess k is obtained from kernels as in Equation (5.2); substituting Equation (4.4), the two factors merge as announced in Section 2.1:

$$\text{MG}_k^{\text{pos}}(z) = (2k-1)!![x^{2k}] \frac{\sqrt{1-T(z)}}{(1-2T(z)x^{-2}\exp_{\geq 2}(x))^{k+1/2}}, \quad (5.3)$$

the direct reading being that a core with a rooted tree hung at each vertex is a positive multigraph together with a set of unicycles, $\text{MG}_k^{(\geq 2)}(T(z)) = \text{MG}_k^{\text{pos}}(z)e^{V(z)}$. Since positive multigraphs of excess k are sets of connected multigraphs of positive excesses summing to k ,

$$\text{CMG}_{n,n+k} = n!2^{n+k}(n+k)![z^n y^k] \log \left(1 + \sum_{\ell \geq 1} \text{MG}_\ell^{\text{pos}}(z)y^\ell \right),$$

and the multigraph counterpart of the connectivity result of Pittel and Wormald (2005) says that the first term of the logarithm dominates, $\text{CMG}_{n,n+k} \sim \text{MG}_{n,n+k}^{\text{pos}}$, for which we gave an independent analytic proof in Panafieu (2019).

The asymptotics of Equation (5.3) are then a matter of the large powers theorem in its multivariate form, the coefficient extraction $[z^n x^{2k}]$ being of that shape when k/n is fixed. The complete expansion takes more work: it is the composition of the expansion given by the large powers theorem with a bivariate extension of the divergent series techniques of Borinsky (2018a), which we developed for this purpose. As for fixed excess, the patchworks of loops and double edges then transfer the result from multigraphs to simple connected graphs. For $\alpha = k/n$ in a compact interval of $\mathbb{R}_{>0}$ and any fixed order d ,

$$\text{CSG}_{n,n+k} = n!(2k-1)!! \frac{(\lambda/(2\alpha))^k}{2\pi k \zeta^n \lambda^{2k}} (c_0 + c_1 k^{-1} + \dots + O(k^{-d})),$$

where λ is the positive solution of $\frac{\lambda}{2} \frac{e^\lambda + 1}{e^\lambda - 1} = \alpha + 1$, $\zeta = e^{-\alpha-1} \sqrt{(\alpha+1)^2 - \lambda^2/4}$, and the c_r are smooth functions of α given in Panafieu (2019).

5.3 The critical window

The results of this section on graphs and multigraphs are those of Janson et al. (1993), refining Erdős and Rényi (1960) and Bollobás (1984); our contribution is the method, and its extension to other models, both dating from our PhD (Panafieu 2014). We present them nonetheless because Section 6.4 relies on them.

A multigraph decomposes uniquely as a set of trees, a set of unicycles and a *positive* multigraph, whose components all have positive excess. The excess of the multigraph is that of its positive part minus its number of trees, so that

$$\text{MG}_{n,m} = n! 2^m m! \sum_{\ell \geq \max(0,k)} [z^n] \frac{U(z)^{n-m+\ell}}{(n-m+\ell)!} e^{V(z)} \text{MG}_\ell^{\text{pos}}(z), \quad (5.4)$$

where $k = m - n$. The left-hand side is n^{2m} . The point of the formula is that the right-hand side exposes the structure, each of its terms being open to marking and extraction. The regime is decided by the extraction $[z^n] U(z)^{n-m}$, a large power of a function whose dominant singularity is $z = e^{-1}$.

For m/n in a compact subinterval of $(0, 1/2)$, the saddle point lies in $(0, e^{-1})$ and the large powers theorem applies: the term $\ell = 0$ alone is asymptotically n^{2m} , which is the theorem of Erdős and Rényi (1960) that a subcritical random multigraph consists of trees and unicycles with high probability (since $\text{MG}_0^{\text{pos}}(z) = 1$). As m/n tends to $1/2$ the saddle point reaches e^{-1} , the singularity of U , V and T . The breakdown of the method is not an artifact but the signal of a change of combinatorial regime: this is the *critical window*,

$$\frac{m}{n} = \frac{1}{2} + \mu n^{-1/3}, \quad \mu \text{ in a compact interval of } \mathbb{R},$$

in which the positive part is no longer empty, every term of Equation (5.4) contributes, and the limit probability that the positive part has excess ℓ is the limit of the term of index ℓ divided by n^{2m} . Substituting the expression of $\text{MG}_\ell^{\text{pos}}$ from Section 5.1, the dominant contribution to each term comes from the cubic kernels, so that the term reduces to an extraction of the form

$$[z^n]U(z)^{n-m+\ell}e^{V(z)}\frac{T(z)^{2\ell}}{(1-T(z))^{3\ell}}.$$

This is a large power of U whose saddle point coalesces with the singularity, multiplied by a function with an algebraic singularity at the same point, since $1 - T(z) \sim \sqrt{2}\sqrt{1 - ez}$. The tool for such a coalescence is the semi-large powers theorem of Banderier et al. (2001), presented in Flajolet and Sedgewick (2009, Theorem IX.16) in connection with the stable laws of probability theory. It covers $[z^n]H(z)^N$ with N in the critical range and H singular of the form $\sigma - h_1(1 - z/\rho) + h_\lambda(1 - z/\rho)^\lambda + O((1 - z/\rho)^2)$, with $\lambda \in (1, 2)$. The trees contribute the extra factor $(1 - T)^{-3\ell}$, which the original theorem does not accommodate; the variant we designed in Panafieu (2015) admits a factor $(1 - z/\rho)^{-\eta}$ and gives

$$[z^n]\frac{H(z)^N}{(1 - z/\rho)^\eta} \sim \sigma^N \rho^{-n} n^{-(1-\eta)/\lambda} \left(\frac{h_1}{h_\lambda}\right)^{(1-\eta)/\lambda} G\left(\lambda, \eta; \frac{h_1^{1+1/\lambda}}{\sigma h_\lambda^{1/\lambda}} x\right)$$

for $N = \sigma n/h_1 + xn^{1/\lambda}$ with x bounded, in terms of the function

$$G(\lambda, \eta; x) = \frac{1}{\lambda\pi} \sum_{j \geq 0} \sin\left(\pi \frac{1 - \eta + j}{\lambda}\right) \Gamma\left(\frac{1 - \eta + j}{\lambda}\right) \frac{(-x)^j}{j!}, \quad (5.5)$$

which for $\lambda = 3/2$, the case at hand, is an Airy-type integral; the parameter x is a multiple of μ , and the statement is close to Lemma 3 of Janson et al. (1993). The limit probability that a critical random multigraph has h_j components of excess j for $j = 1, \dots, r$, and no other complex component, comes out as

$$3^{-\ell} \prod_{j=1}^r \frac{c_j^{h_j}}{h_j!} \sqrt{\frac{3\pi}{2}} \exp\left(-\frac{\mu^3}{6(1 - 2^{-2/3})^3}\right) G\left(\frac{3}{2}, \frac{1}{4} + \frac{3\ell}{2}; -\frac{3^{2/3}\mu}{2(1 - 2^{-2/3})}\right), \quad (5.6)$$

where $\ell = \sum_j j h_j$ and c_j counts the connected cubic kernels of excess j , normalized as in Section 5.1. The patchworks of loops and double edges carry the same result to simple graphs. This is the distribution of Janson et al. (1993), reached here through a black box theorem.

The reason we designed the variant was not graphs. The route from Equation (5.4) to Equation (5.6) extends to the phase transition of hypergraphs (Panafieu 2015), where hyperedges of arbitrary sizes carry weights, and of inhomogeneous multigraphs (Panafieu 2015; Panafieu and Ravelomanana 2015), where vertices carry colors and a matrix weights the edges according to the colors they join; in the latter case the largest eigenvalue of that matrix locates the critical window. These models were the subject of our PhD; the multigraph case is the one presented here, for simplicity.

5.4 The supercritical regime: an open problem

For m/n in a compact subinterval of $(1/2, \infty)$, a random multigraph consists with high probability of trees, unicycles and a single *giant* component, of excess $\Theta(n)$.

The existence, uniqueness and statistics of the giant component were progressively established (Erdős and Rényi 1960; Bollobás 1984; Stepanov 1987; Łuczak 1990; Pittel 1990; Janson et al. 1993). In particular, Pittel and Wormald (2005) obtained by an enumerative approach the joint Gaussian limit law of the size of the core of the giant component, the excess of that core, and the number of vertices outside it, throughout the supercritical phase.

The decomposition Equation (5.4) still holds; what changes is where its mass sits. In the critical window the dominant terms have bounded index ℓ and can be handled one by one. In the supercritical regime they have index $\ell = \Theta(n)$, the range whose asymptotics Section 5.2 supplies. Substituting those asymptotics into Equation (5.4) and applying the Laplace method for sums would, we believe, yield the statistics of the giant component. We have not pursued that route: the proof would be ad hoc, and the statistics it recovers are already known. We look instead for a single exact expression on which complex analysis can act directly.

A resummation. Unpublished work with Sergey Dovgal and Alexandros Singh provides one. Deleting iteratively the vertices of degree 0 and 1 leaves a multigraph of minimum degree 2, from which the original is recovered by rooting a tree at each vertex and adding tree components. The excess of the whole is that of the minimum-degree-2 part minus the number t of tree components, so that Equation (4.2), taken with $\Delta = \exp_{\geq 2}$ and with z replaced by $T(z)$ to hang the rooted trees, gives

$$\text{MG}_k(z) = \sum_{t \geq 0} \frac{U(z)^t}{t!} (2(t+k)-1)!! [x^{2(t+k)}] \frac{1}{(1 - 2T(z)x^{-2} \exp_{\geq 2}(x))^{t+k+1/2}}.$$

Shifting the extraction to $[x^{2k}]$ brings out a factor $(2x^{-2}U(z))^t$, and the identity $\sum_{t \geq 0} \Gamma(t+\beta)y^t/t! = \Gamma(\beta)(1-y)^{-\beta}$ of Section 4.1 collapses the sum. For the second time in this chapter, two factors merge into one:

$$\text{MG}_k(z) = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) [x^{2k}] \frac{1}{(1 - 2x^{-2}(T(z) \exp_{\geq 2}(x) + U(z)))^{k+1/2}}. \quad (5.7)$$

Two remarks carry over from Equation (4.2). First, this is an identity between formal power series only when the coefficient in z is extracted before the one in x . Second, the prefactor is written with the Gamma function rather than as $(2k-1)!!$ because k here ranges over all of $\mathbb{Z}$. The kernels of Equation (4.4) and the positive multigraphs of Equation (5.3) have nonnegative excess by construction, but a multigraph of excess $k = m - n$ has $k < 0$ whenever $m < n$, which covers the subcritical and critical regimes and the lower half of the supercritical one; at the transition $k \sim -n/2$. The double factorial is undefined there, whereas $2^k \Gamma(k+1/2)/\sqrt{\pi}$ continues it, with the value $(-1)^{|k|}/(2|k|-1)!!$.

The expression has the shape of Equation (4.2) with $z\Delta(x)$ replaced by $T(z)\exp_{\geq 2}(x) + U(z)$. Marking $U(z)$, in the sense of Section 2.3, counts the tree components and hence gives the excess of the giant component.

We expect much from Equation (5.7). It is exact and new, and it encodes the structure of multigraphs. What is currently missing is an asymptotic analysis based on a saddle point method. The large powers theorem does not apply directly, and

a bivariate saddle point is required; constructing a suitable path of integration has so far eluded us. The supercritical enumeration of directed acyclic graphs, open as well, leads in Section 6.5 to an expression of the same shape, so that extracting the asymptotics of Equation (5.7) for $k = \Theta(n)$ would unlock both problems at once. The directed side is the subject of an ongoing collaboration with Bernhard Gittenberger, Mehdi Naima and Martin Pépin within the project PAnDAG.

6 Directed graphs

This chapter presents two published papers and two ongoing collaborations. Its contributions are the structure of a random sparse digraph in the model $D(n, p)$, where each of the $n(n - 1)$ possible arcs is present independently with probability p , obtained with Sergey Dovgal, Dimbinaina Ralaivaosaona, Vonjy Rasendrasasina and Stephan Wagner; the asymptotic enumeration of directed acyclic and elementary digraphs in the model $D(n, m)$, uniform among the digraphs with n vertices and m arcs, published in the subcritical regime and inside the critical window, the supercritical case being ongoing work with Bernhard Gittenberger, Mehdi Naima and Martin Pépin; and the exact enumeration of directed acyclic graphs with constrained degrees, together with its link to parking functions, in progress with Antoine Genitrini and Khaydar Nurligareev.

The division into two models has a story. With Sergey Dovgal we had settled the subcritical regime and the critical window, and were working on the supercritical case, when Vlady Ravelomanana told us that our three future co-authors were asking the same questions in $D(n, p)$. We contacted them. Given the variety of our models and approaches, each group published its own conference paper, ours at FPSAC (Panafieu and Dovgal 2020) and theirs at AofA (Ralaivaosaona, Rasendrasasina, and Wagner 2020) we then joined forces for the journal version (Dovgal et al. 2024).

Throughout, the *structure* of a digraph means the description of its strongly connected components: a digraph is *acyclic* when each of them is a single vertex, and *elementary* when each of them is a single vertex or a cycle. The papers work with simple digraphs. We present everything here in the multidigraph model of Chapter 3, for consistency with the rest of the manuscript and because the asymptotic analysis is simplest there; results transfer as they did for graphs. The two models become $MD(n, m)$, uniform among the multidigraphs with n vertices and m edges, and the *Poisson multidigraph* $MD(n, p)$, in which the multiplicity of each of the n^2 oriented pairs of vertices is an independent Poisson variable of mean p ; the latter is the Poisson multigraph $MG(n, 2p)$ of Equation (3.2).

6.1 Exact enumeration

Graphic generating functions. The exact enumeration of digraph families, of DAGs and strongly connected digraphs in particular, has a rich history (Liskovets 1969; 1970; Wright 1971). The unifying tool is due to Robinson (1973), who introduced a new type of generating function, named *graphic generating function* by Gessel (1995), together with the combinatorial operation its product encodes. We have no

contribution here, and present the tool because the rest of the chapter rests on it. The exponential and *graphic* generating functions of a multidigraph family $\mathcal{H}$ are

$$H(z, w) = \sum_{G \in \mathcal{H}} \frac{w^{m(G)} z^{n(G)}}{m(G)! n(G)!}, \quad \hat{H}(z, w) = \sum_{G \in \mathcal{H}} e^{-n(G)^2 w/2} \frac{w^{m(G)} z^{n(G)}}{m(G)! n(G)!},$$

the hat distinguishing the two types; for simple digraphs the factor $e^{-n^2 w/2}$ becomes $(1+w)^{-\binom{n}{2}}$. Two examples recur. The multidigraphs on n vertices have generating function $e^{n^2 w} z^n / n!$, so that, calling *sets* the edgeless multidigraphs,

$$\widehat{\text{MD}}(z, w) = \sum_{n \geq 0} e^{n^2 w/2} \frac{z^n}{n!} = \text{MG}(z, w), \quad \widehat{\text{Set}}(z, w) = \sum_{n \geq 0} e^{-n^2 w/2} \frac{z^n}{n!} = \text{MG}(z, -w).$$

The exponential Hadamard product of Equation (4.7), taken in the variable z , translates each type into the other:

$$\hat{H}(z, w) = H(z, w) \square_z \widehat{\text{Set}}(z, w), \quad H(z, w) = \hat{H}(z, w) \square_z \text{MG}(z, w). \quad (6.1)$$

Arrow product. As discussed in Section 2.1, a type of generating function earns its place through the combinatorial operation that matches the product. The *arrow product* $\mathcal{A} \otimes \mathcal{B}$ is obtained by taking the relabeled Cartesian product of two multidigraphs and adding an arbitrary set of labeled edges, each directed from a vertex of the first to a vertex of the second; it extends to families by union. If $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$, then

$$\hat{C}(z, w) = \hat{A}(z, w) \hat{B}(z, w).$$

Together with the sum for disjoint union, this is all the symbolic method we need for digraphs.

Directed acyclic graphs. Robinson (1973) obtained the enumeration of DAGs by inclusion-exclusion on the *sources*, the vertices of in-degree 0. Let $\text{DAG}(z, w, u)$ be their generating function with u marking the sources. Then $\text{DAG}(z, w, v+1)$ counts DAGs with an arbitrary subset of their sources distinguished, and such an object is uniquely the arrow product of its distinguished sources, a set, with the remaining DAG, so

$$\widehat{\text{DAG}}(z, w, v+1) = \widehat{\text{Set}}(vz, w) \widehat{\text{DAG}}(z, w). \quad (6.2)$$

At $v = -1$ the left-hand side counts DAGs without sources, of which only the empty one exists, of generating function 1, hence

$$\widehat{\text{DAG}}(z, w) = \frac{1}{\text{MG}(-z, -w)}, \quad \widehat{\text{DAG}}(z, w, u) = \frac{\text{MG}((u-1)z, -w)}{\text{MG}(-z, -w)}. \quad (6.3)$$

Gessel (1995) refined this to mark sources and sinks simultaneously.

Strongly connected components. The same reasoning yields the generating function $\text{SCC}(z, w)$ of strongly connected multidigraphs, the building blocks of every multidigraph: contracting each strongly connected component to a vertex produces a DAG, since two mutually reachable components would be one. Call a component

source-like when no other component points to it, and let u mark the source-like components in the generating function $\text{MD}(z, w, u)$ of all multidigraphs. Distinguishing an arbitrary subset of them decomposes a multidigraph as the arrow product of a set of strongly connected multidigraphs, of exponential generating function $e^{v\text{SCC}(z, w)}$, with an arbitrary multidigraph, so that

$$\widehat{\text{MD}}(z, w, v+1) = \left(e^{v\text{SCC}(z, w)} \boxtimes_z \widehat{\text{Set}}(z, w) \right) \text{MG}(z, w).$$

At $v = -1$ the left-hand side is 1, and undoing the Hadamard product gives the closed form

$$\text{SCC}(z, w) = -\log \left(\text{MG}(z, w) \boxtimes_z \frac{1}{\text{MG}(z, w)} \right), \quad (6.4)$$

which packs the recurrences of Liskovets (1970) into a single identity. These are the exact expressions on which the next sections build.

6.2 Directed acyclic graphs with constrained degrees

The enumeration of DAGs with constrained or weighted degrees is the common refinement of Section 6.1 and Section 4.1. We include this work in progress with Antoine Genitrini and Khaydar Nurligareev, young as it is, because the tools of the manuscript (half-edges, the exponential Hadamard product and the inclusion-exclusion on sources) combine here with no friction, as the degrees and the subgraphs did in Section 4.2.

Model. Two sequences $\delta^- = (\delta_d^-)_{d \geq 0}$ and $\delta^+ = (\delta_d^+)_{d \geq 0}$, with generating functions

$$\Delta^-(x) = \sum_d \delta_d^- \frac{x^d}{d!}, \quad \Delta^+(x) = \sum_d \delta_d^+ \frac{x^d}{d!},$$

weight the in- and out-degrees. A DAG is weighted by the product of the weights of its vertices:

$$\text{DAG}(z, w, \delta^-, \delta^+) = \sum_{G \in \mathcal{DAG}} \left(\prod_{v \in V(G)} \delta_{\deg^-(v)}^- \delta_{\deg^+(v)}^+ \right) \frac{w^{m(G)} z^{n(G)}}{m(G)! n(G)!}.$$

We take $\delta_0^- \neq 0 \neq \delta_0^+$, since every DAG has a source and a sink, and write $\text{DAG}_n(w, \delta^-, \delta^+) = n! [z^n] \text{DAG}(z, w, \delta^-, \delta^+)$.

Weighted out-degrees. Let the in-degrees run free first, $\delta^- = \mathbf{1}$. The derivation is the inclusion-exclusion on sources of Equations (6.2) and (6.3), with half-edges doing the bookkeeping of the degrees. A DAG with a distinguished subset of its sources, marked by v , is a set of distinguished sources together with a DAG, the edges leaving the distinguished sources landing anywhere in the latter. Cutting each such edge into two half-edges marked by an exponential variable x , a distinguished source of out-degree d contributes δ_d^+ and d outgoing half-edges, so the sources contribute $e^{vz\Delta^+(xw)}$; each vertex of the remaining DAG receives an arbitrary set of incoming

half-edges, contributing e^x . The half-edges are matched by the exponential Hadamard product, as in the bipartite construction of Section 4.3:

$$\text{DAG}(z, w, \mathbf{1}, \delta^+, v+1) = e^{vz\Delta^+(xw)} \square_{x=1} \text{DAG}(ze^x, w, \mathbf{1}, \delta^+).$$

At $v = -1$ the left-hand side counts sourceless DAGs, that is, the empty one. Since $\text{DAG}(ze^x, w, \mathbf{1}, \delta^+) = \sum_n \text{DAG}_n e^{nx} z^n / n!$, the Hadamard product collapses to evaluations through $e^{\alpha x} \square_{x=1} F(x) = F(\alpha)$, leaving

$$1 = \sum_{n \geq 0} \text{DAG}_n(w, \mathbf{1}, \delta^+) e^{-z\Delta^+(nw)} \frac{z^n}{n!}. \quad (6.5)$$

Given a sequence $\mathbf{a} = (a_0, a_1, \dots)$, the *Gončarov polynomial* $g_n(x; \mathbf{a})$ is the unique monic polynomial of degree n with $g_n^{(k)}(a_k; \mathbf{a}) = 0$ for $0 \leq k < n$; the family is characterized by the *Appell relation*

$$\sum_{n \geq 0} g_n(x; \mathbf{a}) e^{a_n z} \frac{z^n}{n!} = e^{xz}, \quad (6.6)$$

and was brought to combinatorics by Kung and Yan (2003). Comparing Equation (6.5) with Equation (6.6) at $x = 0$ and $a_n = -\Delta^+(nw)$ identifies the coefficients, so that the total weight of DAGs on n vertices with out-degrees weighted by δ^+ and free in-degrees is

$$\begin{aligned} \text{DAG}_n(w, \mathbf{1}, \delta^+) &= g_n\left(0; (-\Delta^+(kw))_{k \geq 0}\right) \\ &= g_n\left(\delta_0^+; (\delta_0^+ - \Delta^+(kw))_{k \geq 0}\right), \end{aligned}$$

the second form following from the shift identity $g_n(x+c; \mathbf{a}+c\mathbf{1}) = g_n(x; \mathbf{a})$.

Two special cases of Gončarov polynomials, $g_n(x; c\mathbf{1}) = (x-c)^n$ and $g_n(x; (a+bk)_k) = (x-a)(x-a-bn)^{n-1}$, give two sanity checks. If Δ^+ is constant, all out-degrees are 0 and $\text{DAG}_n = (\delta_0^+)^n$. If $\Delta^+(x) = \delta_0^+ + \delta_1^+ x$, only out-degrees 0 and 1 are allowed, the sequence $\delta_0^+ - \Delta^+(kw) = -\delta_1^+ kw$ is arithmetic, and

$$\text{DAG}_n(w, \mathbf{1}, \delta^+) = \delta_0^+ (\delta_0^+ + \delta_1^+ nw)^{n-1},$$

a Cayley formula, as it should be: a DAG with out-degrees at most 1 is a forest oriented toward its roots.

Parking functions. For a nondecreasing sequence $\mathbf{u} = (u_0, \dots, u_{n-1})$ of positive integers, a *$\mathbf{u}$ -parking function* of length n is a sequence $(p_1, \dots, p_n)$ of positive integers whose nondecreasing rearrangement satisfies $p_{(i)} \leq u_{i-1}$ for every i , and Kung and Yan (2003) proved that there are $(-1)^n g_n(0; \mathbf{u})$ of them. Hence, whenever $(\Delta^+(kw))_{k \geq 0}$ is a nondecreasing sequence of positive integers, $\text{DAG}_n(w, \mathbf{1}, \delta^+)$ counts the $(\Delta^+(kw))_k$ -parking functions of length n . Finding a bijection between DAGs with out-degree constraints and parking functions is an appealing line of research. The case of out-degrees at most 1 is suggestive, since $\delta_0^+ (\delta_0^+ + \delta_1^+ nw)^{n-1}$ is the shape of the classical count of parking functions and of forests alike, and the bijection sought

would specialize to one of the known correspondences between the two (Foata and Riordan 1974; Kreweras 1980; Shin 2008).

Weighted in- and out-degrees. The argument accommodates weights on both degrees at the price of treating δ^- as formal. A vertex of the undistinguished part that receives e additional incoming edges has its in-degree weight shifted from δ_d^- to δ_{d+e}^- , and differentiation performs the shift – the device already used for subgraphs in Equation (4.5). The inclusion-exclusion then gives

$$\text{DAG}_n(w, \delta^-, \delta^+) = - \sum_{k=0}^{n-1} \binom{n}{k} (-\delta_0^- \Delta^+(xw))^{n-k} \square_{x=1} \text{DAG}_k\left(w, (\partial_x^d \Delta^-(x))_{d \geq 0}, \delta^+\right),$$

which is effective but no longer collapses to an evaluation. The Gončarov identification does not survive, the coefficients being polynomials in the shifted weights rather than numbers.

6.3 The birth of the strong components

This section presents the results of Dovgal et al. (2024) on the phase transition of random digraphs.

The transition. The random digraph $D(n, p)$ undergoes a phase transition at $p = 1/n$, discovered by Karp (1990) and Łuczak (1990): for $np < 1$ fixed it is with high probability elementary, while for $np > 1$ a unique strong component of linear size appears. Call *complex* a strong component of positive excess. Łuczak and Seierstad (2009) located the critical window at $np = 1 + \mu n^{-1/3}$ with μ fixed, and proved that the probability of being elementary decreases from 1 to 0 as μ runs from $-\infty$ to $+\infty$; Goldschmidt and Stephenson (2023) later described the scaling limit of the strong components inside the window. We obtain the limit of that probability as an explicit function of μ , together with the probability of being acyclic and the probability of containing a single complex component of prescribed kernel, across the three regimes. While Dovgal et al. (2024) covers all models, we present the results in the Poisson multidigraph $\text{MD}(n, p)$.

From exact expressions to integrals. By the digraph counterpart of Equation (3.2), the probability that an $\text{MD}(n, p)$ multidigraph belongs to a family $\mathcal{F}$ is $e^{-n^2 p/2} n! [z^n] \hat{F}(z, p)$, through the graphic generating function of Section 6.1. Our families are described by their strong components: if $\mathcal{S}$ is a family of strongly connected multidigraphs with exponential generating function $S(z, w)$, the inclusion-exclusion on source-like components gives, for the multidigraphs all of whose strong components lie in $\mathcal{S}$,

$$\hat{D}_{\mathcal{S}}(z, w) = \frac{1}{e^{-S(z, w)} \square_z \widehat{\text{Set}}(z, w)}, \quad (6.7)$$

with $S = 0$ for acyclic multidigraphs, $S = V(z, w)$ for elementary ones, and a marking variable for a distinguished complex component. The Hadamard product is linearized into an integral by the Gaussian representation of $\widehat{\text{Set}}$, the Hubbard–Stratonovich device of Section 4.2:

$$\begin{aligned}\varphi(z, w; F) &:= e^{-S(z, w)} \square_z \widehat{\text{Set}}(z, w) \\ &= \frac{1}{\sqrt{2\pi w}} \int_{-\infty}^{+\infty} F(zwe^{-ix}) \exp\left(-\frac{x^2}{2w} - ze^{-ix}\right) dx,\end{aligned}\tag{6.8}$$

where F collects the contribution of the components. Each of the three families corresponds to one variant of φ , and each probability to a coefficient of a simple rational expression in those variants.

Strategy. The coefficient extraction in z is a singularity analysis, the singularities being the zeros of $\varphi(z, p)$. Two features take it outside the black box. First, $p = \lambda/n$ tends to 0, so the zeros move with n : a saddle point analysis of Equation (6.8), carried out by hand, places the j -th zero at $(ew)^{-1}(1 - a_j 2^{-1/3} w^{2/3} + O(w))$, where a_j is the j -th zero of the Airy function Ai , and the zeros of the elementary variant are governed by those of Ai' . Second, the number of contributing zeros decides the regime. For $\lambda \neq 1$ the first zero dominates. In the window $p = (1 + \mu n^{-1/3})/n$ the zeros become equally spaced on the scale $n^{-1/3}$ and all contribute comparably; summing them produces the contour integrals of the results, the directed counterparts of the function $G(3/2, \eta; x)$ of Equation (5.5).

Results. For the probability that a random multidigraph is acyclic, with $p = \lambda/n$,

$$\begin{aligned}\mathbb{P}_{\text{MD}(n, p)}(\text{acyclic}) &\sim \begin{cases} 1 - \lambda, & \lambda < 1, \\ \gamma_1 n^{-1/3}, & \lambda = 1, \\ \gamma_2(\lambda) n^{-1/3} \exp(-\alpha(\lambda)n + a_1 \beta(\lambda) n^{1/3}), & \lambda > 1, \end{cases} \\ \gamma_1 &= \frac{2^{-1/3}}{2i\pi} \int_{-i\infty}^{i\infty} \frac{dt}{\text{Ai}(-2^{1/3}t)} \approx 0.488737, \quad \gamma_2(\lambda) = \frac{2^{-2/3}}{\text{Ai}'(a_1)} \lambda^{5/6} e^{(\lambda-1)/6}, \\ \alpha(\lambda) &= \frac{\lambda^2 - 1}{2\lambda} - \log \lambda \quad \beta(\lambda) = 2^{-1/3} \lambda^{-1/3} (\lambda - 1)\end{aligned}$$

and $a_1 \approx -2.338107$ is the zero of Ai of smallest modulus. Inside the window, for $\mu = o(n^{1/12})$,

$$\mathbb{P}_{\text{MD}(n, p)}(\text{acyclic}) \sim n^{-1/3} 2^{-1/3} e^{-\mu^3/6} \frac{1}{2i\pi} \int_{-i\infty}^{i\infty} \frac{e^{-\mu s}}{\text{Ai}(-2^{1/3}s)} ds,$$

which matches $|\mu|n^{-1/3}$ as $\mu \rightarrow -\infty$ and the supercritical line as $\mu \rightarrow +\infty$. The probability of being elementary has the same shape with Ai' in place of Ai : it tends to 1 for $\lambda < 1$, is a similar contour integral in the window, and is exponentially small for $\lambda > 1$. At the center of the window a random multidigraph is acyclic with probability $\gamma_1 n^{-1/3}$, elementary with probability about 0.6997, and carries exactly one complex component, bicyclic, with probability tending to 1/8. Prescribing the kernel of that component replaces the Airy function by a generalized one, and several complex components lead to products of these.

The singularity analysis with moving singularities and the saddle point analysis of Equation (6.8) were both done by hand, and the resulting proofs are long and

specific to the families treated. Finding general theorems that package them, so that the next digraph family costs a substitution rather than a paper, is a worthy line of research. A concrete motivation is the derivation of the statistics of the giant strong component above the transition.

6.4 Directed acyclic graphs in the $\text{MD}(n, m)$ model

This section presents the results of Panafieu and Dovgal (2020), which predate the previous section and approach the same transition in the uniform model $\text{MD}(n, m)$ of multidigraphs with n vertices and m edges, by a completely different combinatorial route. The paper treats simple digraphs; we present the lighter multidigraph version. Before this work the asymptotics of DAGs were known in the dense case, p constant, from Bender et al. (1986), and in the subcritical regime from the Poisson law of the number of short cycles implicit in Łuczak (1990); the critical window was open.

DAGs as pairs of multigraphs. The exact expression Equation (6.3) is not immediately amenable to asymptotic analysis: $\text{MG}(-z, -w)$ is a divergent series whose reciprocal is an alternating sum, the situation of Equation (5.1). As there, the way forward is another exact expression, whose dominant contribution is visible. Since $e^{n^2 w/2} = n![z^n] \text{MG}(z, w)$, Robinson's formula reads

$$\text{DAG}_{n,m} = n!m![z^n w^m] \frac{e^{n^2 w/2}}{\text{MG}(-z, -w)} = n!^2 m![z_0^n z_1^n w^m] \frac{\text{MG}(z_0, w)}{\text{MG}(-z_1, -w)} : \quad (6.9)$$

the number of DAGs is a signed count of *pairs* of multigraphs on n vertices each, but sharing m edges, the second carrying the sign (-1) raised to its numbers of vertices, edges and components. We obtained this identity by manipulating generating functions and have no combinatorial explanation for it; finding one is an appealing problem. Intuitively, the two multigraphs should share the m edges roughly equally, each receiving about $m/2$ of them on n vertices, so that both become critical when $m/2 = n/2$: the transition of DAGs at $m = n$ is that of multigraphs at $m = n/2$, seen twice.

We substitute for both occurrences of MG the decomposition of Equation (5.4) into trees, unicycles and a positive part. After a change of variables, Equation (6.9) becomes

$$\begin{aligned} \text{DAG}_{n,m} = \sum_{t \geq \max(0, m-2n)} n!^2 m![z_0^n z_1^n] \frac{(U(z_0) + U(z_1))^{2n-m+t}}{(2n-m+t)!} e^{V(z_0)-V(z_1)} \\ [y^t] \frac{1 + \sum_{k \geq 0} \text{MG}_k^{\text{pos}}(z_0) y^k}{1 + \sum_{\ell \geq 0} \text{MG}_\ell^{\text{pos}}(z_1) (-y)^\ell}, \end{aligned} \quad (6.10)$$

an infinite sum indexed by the total excess t of the two positive parts. It is the counterpart of Equation (5.4) and behaves in the same way: for m/n in a compact subinterval of $(0, 1)$ the term $t = 0$ dominates, while in the critical window $m = n(1 + \mu n^{-1/3})$ all terms contribute comparably.

A bivariate semi-large powers lemma. Replacing each MG_k^{pos} by its dominant term, the cubic kernels $e_k T(z)^{5k} / (1 - T(z))^{3k}$ with $e_k = \frac{(6k)!}{2^{5k} 3^{2k} (2k)! (3k)!}$, each

term of Equation (6.10) becomes an extraction of a large power of $U(z_0) + U(z_1)$ whose saddle point coalesces with the singularity e^{-1} in both variables, times factors singular at the same point: the situation of the semi-large powers theorem of Banderier et al. (2001), in two variables, for which we proved a bivariate variant. The asymptotics of each term are then, up to an explicit prefactor, a product of two copies of $G(\lambda, \eta; x)$ of Equation (5.5). The critical structure of DAGs is governed by products of two copies of the Airy-type functions that govern critical multigraphs.

Results. The end results are the same as in the previous section, in the subcritical regime and inside the critical window, with λ replaced by m/n , showing that a random $\text{MD}(n, p)$ multidigraph behaves as if its number of edges were exactly the expected number of edges. The two papers give different expressions for the limit probability in the window; their equality was a reassuring confirmation of our computations. Since acyclicity and elementarity are monotone properties and the number of edges of $\text{MD}(n, p)$ fluctuates on a smaller scale than the critical window, the classical coupling of the two models transfers these limits from one to the other, which explains their equality. The transfer fails beyond the window, where the probabilities are exponentially small. The two approaches part company beyond the window: in $\text{MD}(n, p)$, Section 6.3 covers the supercritical regime, where the probabilities of being acyclic or elementary are exponentially small; in $\text{MD}(n, m)$, the number of DAGs with $m/n > 1$ fixed is still open, for the reason discussed next.

6.5 Beyond the critical window: an open problem

In $\text{MD}(n, m)$, the number of DAGs with $m/n > 1$ fixed is not known asymptotically. This section reports on work in progress on this problem with Bernhard Gittenberger, Mehdi Naima and Martin Pépin.

In Equation (6.10) the dominant contribution above the transition comes from summands whose index t grows with n , and we do not know how to control them: the obstruction of Section 5.4, again. The heuristic of Section 6.4 also stops being reliable. Numerical experiments by Bernhard Gittenberger and Martin Pépin, together with a comparison with the $\text{MD}(n, p)$ results, suggest that above the transition the m edges no longer distribute equally between the two multigraphs of Equation (6.9): one appears to remain critical while the other becomes supercritical.

The promising path is the one left aside in Section 5.4. The resummation that produced Equation (5.7), gathering the trees hung on a core and the tree components into a single factor, applies to DAGs as well. Passing Equation (6.9) to the excess grading by $(z, w) \mapsto (z/y, y)$ and inserting the decomposition into trees, unicycles and a positive part in both numerator and denominator,

$$\text{DAG}_{n,m} = n!^2 m! [z_0^n z_1^n y^{m-2n}] e^{(U(z_0)+U(z_1))/y} e^{V(z_0)-V(z_1)} \frac{\text{MG}^{\text{pos}}(z_0/y, y)}{\text{MG}^{\text{pos}}(-z_1/y, -y)}.$$

The trees of the two multigraphs enter only through $U(z_0) + U(z_1)$. That fact let the bivariate lemma of Section 6.4 decouple the variables; it is also what makes the resummation possible. Expanding the exponential and extracting $[y^{m-2n}]$, the sum over the number of tree components is the one resummed in Equation (5.7), with $U(z_0) + U(z_1)$ in place of $U(z)$, so that

$$\begin{aligned}
\text{DAG}_{n,m} &= n!^2 m! [z_0^n z_1^n] e^{V(z_0) - V(z_1)} \sum_{\ell \geq 0} \left([y^\ell] \frac{1}{\text{MG}^{\text{pos}}(-z_1/y, -y)} \right) \\
&\quad \frac{2^{m-2n-\ell}}{\sqrt{\pi}} \Gamma\left(m - 2n - \ell + \frac{1}{2}\right) \\
&\quad [x^{2(m-2n-\ell)}] \frac{1}{(1 - 2x^{-2}(T(z_0) \exp_{\geq 2}(x) + U(z_0) + U(z_1)))^{m-2n-\ell+1/2}}.
\end{aligned} \tag{6.11}$$

We conjecture that every summand contributes to the same order, so that the asymptotics of $\text{DAG}_{n,m}$ are the sum of theirs. The last line of Equation (6.11) differs from Equation (5.7) only by the replacement of $U(z_0)$ with $U(z_0) + U(z_1)$ and by the presence of a second variable, which is why the extraction called for in Section 5.4 would settle this problem too.

7 Satisfiable 2-SAT formulas

This chapter presents our work with Sergey Dovgal and Vladyslav Ravelomanana (Dovgal, Panafieu, and Ravelomanana 2023). Its contributions are the exact enumeration of satisfiable 2-SAT formulas, which relies on a new type of generating function, and a prediction of the limit probability of satisfiability across the critical window, to eleven digits at its center, where simulations had reached three.

Random constraint satisfaction problems are graph-like objects: variables are vertices, and clauses on two variables are edges, on more variables hyperedges. Our first work in this direction, during our PhD, concerned 2-Xor formulas (Panafieu et al. 2014; 2016), multigraphs with colored edges whose satisfiability threshold at $m/n = 1/2$ is tied to the birth of the giant component of Section 5.3. For 2-SAT the relevant structure is directed, so we adapt the tools of Chapter 6.

2-SAT. A 2-CNF on the variables $x_1, \dots, x_n$ is a set of clauses $(u \vee v)$, where u and v are literals, variables or their negations, on distinct variables; it is *satisfiable* if some assignment of the variables makes every clause true. Its *implication digraph* has the $2n$ literals as vertices and, for each clause $(u \vee v)$, the two arcs $\bar{u} \rightarrow v$ and $\bar{v} \rightarrow u$, which express the clause as an implication and its contrapositive. A 2-CNF is satisfiable if and only if no variable x is *contradictory*, meaning that x and $\bar{x}$ lie in the same strongly connected component (Aspvall, Plass, and Tarjan 1979). 2-SAT is therefore decidable in linear time, while k -SAT is NP-complete for $k \geq 3$. Counting the solutions of a 2-CNF is nonetheless #P-hard (Valiant 1979) their number was recently analyzed by Chatterjee et al. (2024).

The transition. Let $\mathbb{P}_{\text{SAT}}(n, p)$ be the probability that a 2-CNF with n variables, each of the $2n(n-1)$ possible clauses being present with probability p , is satisfiable. Chvátal and Reed (1992), Goerdt (1996) and Fernandez de la Vega (2001) proved that it tends to 1 for $p = c/n$ with $c < 1/2$ and to 0 for $c > 1/2$; Bollobás et al. (2001) located the critical window at $p = (1 + \mu n^{-1/3})/(2n)$ and proved that the limit is $1 - \Theta(|\mu|^{-3})$ as $\mu \rightarrow -\infty$ and $\exp(-\Theta(\mu^3))$ as $\mu \rightarrow +\infty$, with later refinements by Kim (2008) and Dovgal (2019). Müller, Neininger, and Zhu (2024) described the limit distribution of the solutions. The shape of the transition is the function

$$\mathbb{P}_{\text{SAT},\infty}(\mu) = \lim_n \mathbb{P}_{\text{SAT}}(n, (1 + \mu n^{-1/3})/(2n)). \quad (7.1)$$

Structure. An implication digraph is symmetric: with each arc $u \rightarrow v$ it contains $\bar{v} \rightarrow \bar{u}$. That symmetry makes its enumeration harder than that of digraphs. It also sorts the strongly connected components into two kinds. An *ordinary* component contains no variable together with its negation (Figure 1); its image under the symmetry is a disjoint ordinary component, sink-like exactly when the original is source-like. A *contradictory* component contains a variable together with its negation, hence all its variables with their negations (Figure 2); it is its own image under the symmetry, no path joins it to another contradictory component, and it is isolated as soon as it is source-like or sink-like. A 2-CNF is satisfiable exactly when it has no contradictory component.

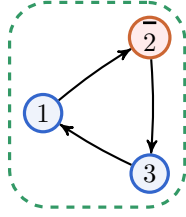


Figure 1: An ordinary component.

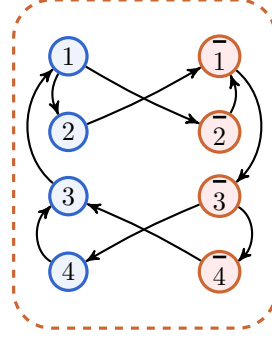


Figure 2: A contradictory component.

A new type of generating function. For a family with $b_{n,m}$ implication digraphs on n variables and m clauses, and $b_n(w) = \sum_m b_{n,m} w^m$, the exponential and graphic generating functions of Section 6.1 are joined by a new type, the *implication generating function* $\hat{B}$:

$$\begin{aligned} B(z, w) &= \sum_n b_n(w) \frac{z^n}{n!}, \\ \hat{B}(z, w) &= \sum_n \frac{b_n(w)}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}, \\ \dot{B}(z, w) &= \sum_n \frac{b_n(w)}{(1+w)^{n(n-1)}} \frac{z^n}{2^n n!}. \end{aligned}$$

The three types are converted into one another by exponential Hadamard products, as in Equation (6.1), and the probability that a 2-CNF of the (n, p) model belongs to a family $\mathcal{H}$ is $2^n n! (1-p)^{n(n-1)} [z^n] \dot{H}(z, p/(1-p))$.

The operation matching the product is the *implication product* of a digraph D with an implication digraph F , illustrated in Figure 3. The vertices of D are relabeled and each vertex j is replaced by the literal x_j or $\bar{x}_j$, giving a digraph $L(D)$, whose mirror image $R(D)$ carries the negated literals with the arcs reversed. An arbitrary set of arcs from $L(D)$ to F is added, then an arbitrary set of arcs from $L(D)$ to $R(D)$

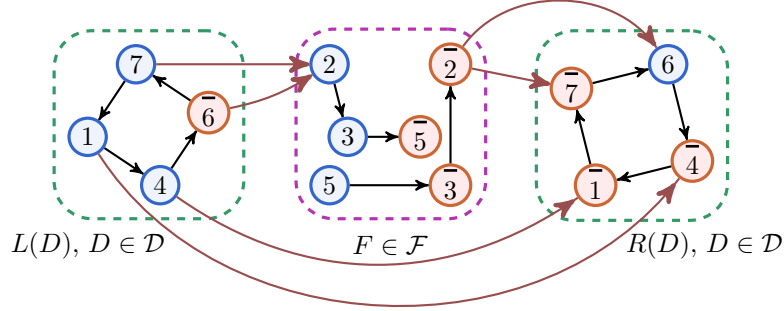


Figure 3: An element of the implication product of a digraph family $\mathcal{D}$ and an implication digraph family $\mathcal{F}$.

joining distinct variables, and the result is completed by symmetry. In the figure, D receives the labels $\{1, 4, 6, 7\}$ with 6 negated and F the labels $\{2, 3, 5\}$; the arcs $7 \rightarrow 2$ and $\bar{6} \rightarrow 2$ enter F and their mirrors $\bar{2} \rightarrow \bar{7}$ and $\bar{2} \rightarrow \bar{6}$ leave it, while $1 \rightarrow \bar{4}$ goes from $L(D)$ to $R(D)$ together with its contrapositive $4 \rightarrow \bar{1}$. If $\mathcal{C}$ is the implication product of a digraph family $\mathcal{D}$ with an implication digraph family $\mathcal{F}$, then

$$\dot{C}(z, w) = \hat{D}(z, w) \dot{F}(z, w).$$

Exact enumeration. The inclusion-exclusion on source-like components that led to Equation (6.4) carries over, with two adjustments forced by the symmetry. The source-like ordinary components are marked together with the isolated pairs of symmetric ordinary components; and, since a source-like contradictory component is isolated, a 2-CNF without non-isolated ordinary source-like component is a set of isolated contradictory components and of pairs of symmetric ordinary ones. Writing $\text{Ord}(z, w)$ for the exponential generating function of the strongly connected digraphs to which ordinary components reduce and $\text{Contr}(z, w)$ for that of the allowed contradictory components, the 2-CNFs whose components are so constrained have implication generating function

$$\text{CNF}_{\text{Ord}, \text{Contr}}(z, w) = \frac{e^{\text{Contr}(z, w) - \text{Ord}(2z, w)/2} \widehat{\square}_z \text{Set}(z, w)}{e^{-\text{Ord}(z, w)} \widehat{\square}_z \text{Set}(z, w)},$$

where, using the generating function $\text{SG}(z, w)$ of simple graphs from Equation (3.1),

$$\begin{aligned} \widehat{\text{Set}}(z, w) &= \sum_{n \geq 0} \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!} = \text{SG}\left(z, -\frac{w}{1+w}\right), \\ \dot{\text{Set}}(z, w) &= \sum_{n \geq 0} \frac{1}{(1+w)^{n(n-1)}} \frac{z^n}{2^n n!} = \text{SG}(z/2, (1+w)^{-2} - 1). \end{aligned}$$

Satisfiable 2-CNFs are the case $\text{Contr} = 0$ and $\text{Ord} = \text{SCC}$, for which Equation (6.4) turns the denominator into $1/\text{SG}(z, w)$ and

$$\text{SAT}(z, w) = \text{SG}(z, w) \left(\sqrt{\text{SG}(z, w) \boxdot_z \frac{1}{\text{SG}(z, w)} \boxdot_z \text{SG}(z, (1+w)^{-2} - 1)} \right). \quad (7.2)$$

The number of satisfiable 2-CNFs with n variables and m clauses is thus a coefficient of an expression built from the generating function of graphs alone. We present the result for the most widely accepted model of 2-CNFs, with distinct clauses on distinct variables; the models with repeated clauses, akin to multigraphs, could naturally be treated as well.

The transition, numerically. Deroulers and Monasson (2006) estimated by simulation, up to $n = 5 \times 10^6$ variables, that $\mathbb{P}_{\text{SAT},\infty}(0) = 0.907 \pm 10^{-3}$. From Equation (7.2), Sergey Dovgal computed the exact probabilities for large n . Extrapolating them predicts, accurately but without proof,

$$\mathbb{P}_{\text{SAT},\infty}(0) = 0.90622396067 \pm 10^{-11},$$

together with the whole curve of Figure 4. Since Equation (7.2) is built from the same Hadamard products as Equation (6.4) and Equation (6.7), we expect the analytic tools of Section 6.3 to extend to 2-CNFs and to yield $\mathbb{P}_{\text{SAT},\infty}(\mu)$ in closed form, presumably as a contour integral of Airy functions. Chapter 11 returns to it.

8 Spectrum of random graphs and tree walks

This chapter presents our work with Eva Hainzl (Hainzl and Panafieu 2026), begun during her visit to Paris, supported by the RandNET project, on a problem suggested to us by Nicolas Curien and Laurent Ménard. Its contributions are the proof that the generating functions of tree walks graded by excess are rational functions of the Catalan series, a conjecture of Bauer and Golinelli (2001), obtained by transporting Wright’s kernel decomposition to walks; and an expansion of the spectral measure

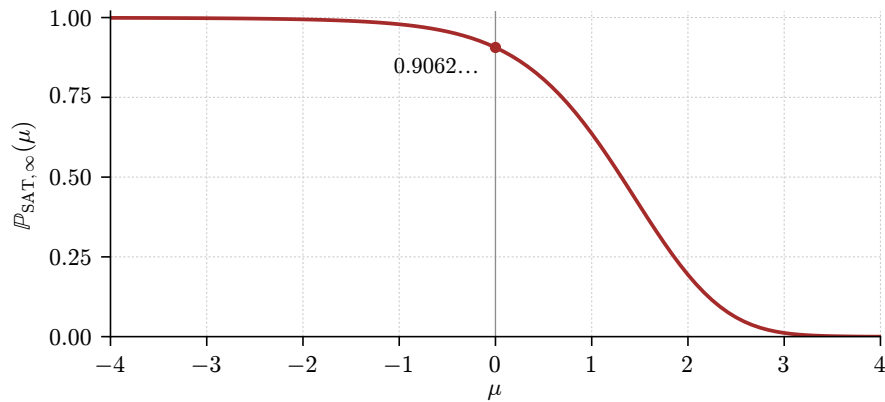


Figure 4: The predicted limit probability $\mathbb{P}_{\text{SAT},\infty}(\mu)$ of satisfiability inside the critical window, extrapolated from the exact probabilities given by Equation (7.2).

of the sparse random graph $G(n, c/n)$ in powers of $1/c$, with explicit densities up to order five, extending a computation of Enriquez and Ménard (2016).

From the spectrum to tree walks. Let μ_n^c be the empirical spectral measure of the adjacency matrix of $G(n, c/n)$, rescaled by $c^{-1/2}$:

$$\mu_n^c = \frac{1}{n} \sum_{\lambda \in \text{Sp}(c^{-1/2}A(G(n, c/n)))} \delta_\lambda,$$

where δ_λ denotes the Dirac measure at λ . The spectral measure of $G(n, p)$, suitably rescaled, converges to the semicircle law for p fixed (Wigner 1958), and more generally as long as $np \rightarrow \infty$ (Tran, Vu, and Wang 2013, Lemma A.1). In the sparse regime the limit μ^c exists and is determined by its moments (Zakharevich 2006), but has a dense set of atoms, and a continuous part exactly when $c > 1$ (Chayes et al. 1986; Bordenave, Sen, and Virág 2017). The moment method makes it a counting problem. The ℓ -th moment is the normalized expected number of closed walks of length ℓ , and a closed walk visiting m vertices and e distinct edges contributes $O(n^{m-e-1})$: only those whose induced graph is a tree survive, the *tree walks*, and the odd moments vanish in the limit. Writing $w_{m, 2\ell}$ for the number of tree walks of length 2ℓ spanning a tree on m labeled vertices, and grading by the *excess* $\xi = \ell - m + 1$, half the length minus the number of edges, the generating function of the even moments is

$$M_{\mu^c}(z) = \sum_{\ell \geq 0} m_{2\ell}(\mu^c) z^{2\ell} = \sum_{\xi \geq 0} W_\xi(z^2) c^{-\xi}, \quad W_\xi(z) = \sum_{\ell \geq 0} \frac{w_{\ell-\xi+1, 2\ell}}{(\ell-\xi+1)!} z^\ell.$$

Approximating μ^c for large c is thus a matter of computing W_ξ for the first few ξ , as approximating a critical random graph was a matter of the first few kernels. The walks of excess 0 traverse each edge exactly twice; they are the contour walks of plane trees, so $W_0 = C$ is the Catalan series $C(z) = (1 - \sqrt{1 - 4z})/(2z) = \sum_n \text{Cat}(n) z^n$, and Wigner's law is recovered at order c^0 .

Why tree walks are not trees. A tree walk is a root with subwalks attached, but the subwalks are not independent: the number of times the walk leaves a vertex toward its parent is constrained by the number of times it enters it, and the subwalks of siblings interleave. The family is not obtained by substituting a class into itself. Bauer and Golinelli (2001) found the way around: record the number of departures from the root in a *catalytic* variable, exponential because the departures are shuffled among subtrees, and discard it once the decomposition closes by the Laplace operator $\mathcal{L}_{t=1}(A(t)) = \sum_k k! [t^k] A(t)$. Their functional equation for all tree walks is well suited to computation, and allowed them to tabulate $w_{m, 2\ell}$ up to half-length 120 (Sloane 2017), but the Laplace operator places it outside singularity analysis. They conjectured that each W_ξ is a rational function of the Catalan generating function C .

Kernel walks. Our contribution is to apply the catalytic variable to a finite family only. The *kernel* of a tree walk is obtained by deleting the leaf and root edges traversed exactly twice and merging, at non-root vertices of degree 2 whose two edges are traversed consecutively, the two edges into one (see Figure 5). A tree walk of excess ξ that is its own kernel has at most $3\xi - 1$ vertices, $2\xi - 2$ edges traversed exactly twice, and half-length $4\xi - 2$, so there are finitely many kernel walks of each

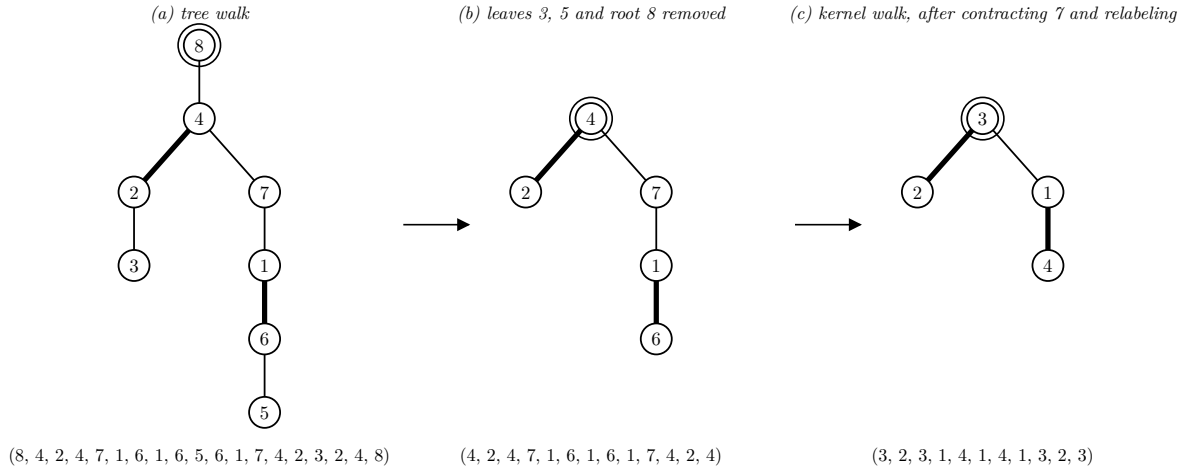


Figure 5: The reduction of a tree walk to its kernel. Thick edges are traversed four times, thin ones twice; the root is circled. The leaves 3 and 5, visited once, are removed, and so is the root 8, visited only at the ends of the walk, which reroots the walk at 4; the vertex 7, of degree 2 with both edges traversed twice, is contracted into a single edge. Relabeling gives the kernel walk on the right, of the same excess 2 as the original.

excess, and their generating function $K_\xi(u, v, z)$, with u marking the edges traversed twice, v the vertices and z the half-length, is a polynomial. Reversing the reduction is a substitution, each edge traversed twice becoming a sequence of back-and-forth steps and an excess-free walk being inserted before and after every step, and gives

$$W_\xi(z) = C(z) \sum_{s=0}^{2\xi-2} \frac{K_{\xi,s}(zC(z)^2)}{(1 - zC(z)^2)^{s+1}}, \quad (8.1)$$

where $K_{\xi,s}$ marks the kernel walks with s edges traversed twice. Each W_ξ is a rational function of the Catalan series, which proves the conjecture of Bauer and Golinelli. The extremal kernels, those with $2\xi - 2$ or $2\xi - 3$ such edges, are counted in closed form, $\text{Cat}(\xi - 1)z^{4\xi-2}$ and $(3\xi - 1) \text{Cat}(\xi - 1)z^{4\xi-3}$, and control the singular behavior, as the cubic kernels did for critical graphs. The polynomials K_ξ themselves are computed with the catalytic variable, through the *superreduced* walks, the kernel walks with no edge traversed twice: splitting the departures from the child of the root into those returning to the root and the others gives their generating function as a fixed point inside a Laplace operator, which determines finitely many coefficients per excess, and kernel walks follow by Lagrange inversion.

Back to the spectrum. The Stieltjes transform of a measure is $S_\mu(z) = \sum_\ell m_\ell(\mu)z^{-\ell-1} = z^{-1}M_\mu(1/z)$, and recovers the measure by the Stieltjes–Perron inversion $\rho(x) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \text{Im } S_\mu(x + i\varepsilon)$. Under the substitution $z \mapsto 1/z^2$, Equa-

tion (8.1) becomes an expansion of S_{μ^c} in powers of $1/c$ whose terms are rational functions of the Stieltjes transform S_σ of the semicircle law,

$$S_{\mu^c}(z) = S_\sigma(z) + \frac{1}{c} \frac{S_\sigma(z)^5}{1 - S_\sigma(z)^2} + \frac{1}{c^2} \frac{S_\sigma(z)^7 + 4S_\sigma(z)^9 - 6S_\sigma(z)^{11} + 2S_\sigma(z)^{13}}{(1 - S_\sigma(z)^2)^3} + \dots,$$

with poles at $S_\sigma(z)^2 = 1$, that is, at the edge of the semicircle. They cannot be inverted term by term into densities, and indeed μ^c has atoms.

Enriquez and Ménard (2016) showed that rescaling μ^c by a correction to the classical factor $c^{-1/2}$ removes the pole at first order. Pushing their computation with Equation (8.1), we found that rescaling by $p_5(1/c)^{-1/2}$ with

$$p_5(x) = 1 + x + x^2 + 4x^3 + 33x^4 + 386x^5,$$

each coefficient being determined by the requirement that the corresponding pole cancel, removes the poles up to order 5: the terms of the rescaled expansion become $C(z)$ times polynomials in $zC(z)^2$, and the inversion formula applies to each. Writing f_i for the resulting densities, the first three are

$$f_0(x) = \frac{\sqrt{4-x^2}}{2\pi}, \quad f_1(x) = (1-x^2) \frac{\sqrt{4-x^2}}{2\pi}, \quad f_2(x) = (1-6x^2+5x^4-x^6) \frac{\sqrt{4-x^2}}{2\pi}$$

on $(-2, 2)$, the first being the semicircle and the others the densities of signed measures of mass zero, so that the moments of the rescaled measure are those of $f_0 + f_1/c + \dots + f_5/c^5$ up to $O(c^{-6})$.

The computation does not explain why admissible coefficients of p exist five times in a row, and again at orders 6 and 7. We conjecture that a unique power series $P(x)$ with nonnegative integer coefficients does this at every order, which amounts to a complete asymptotic expansion of μ^c as $c \rightarrow \infty$. We prove that part of the cancellation is structural: the two leading singular terms of every W_ξ are controlled by the extremal kernels, so any rescaling with the right first two coefficients lowers the order of the pole by two. For the rest we have no explanation, and no combinatorial meaning for the series involved. The expansion seems divergent, in the manner of Stirling's series (Olver 1997): the polynomial factor of f_5 has degree 18 and coefficients up to 904447, the f_i oscillate, and for a given c there is an optimal truncation order, beyond which the partial sums cease to be nonnegative. That order is also the one that matches the eigenvalue histograms of sampled matrices best; Figure 6 compares the first three truncations with such histograms for $c = 5$.

Extension. Hainzl (2026) has since obtained the asymptotics of the k -tours, the tree walks traversing every edge exactly $2k$ times, whose numbers are the hypergraph Catalan numbers of the hypergraph matrix model of Gunnells (2021) and Gunnells (2024). She identified the subfamily carrying the dominant contribution and gave it a combinatorial description amenable to asymptotic extraction, which proves the conjectured growth rate. Whether the same approach can extract the asymptotics of general tree walks is a question we find appealing.

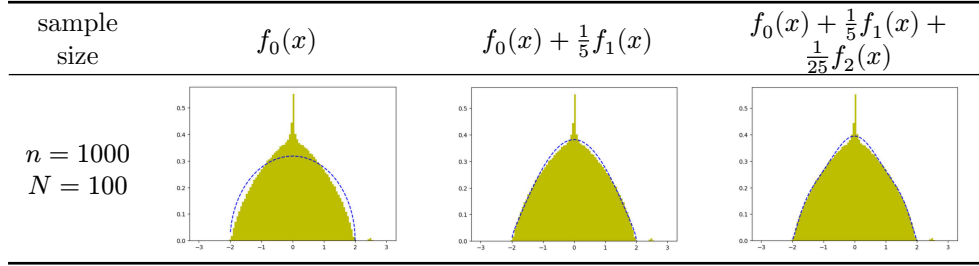


Figure 6: Histograms (100 bins) of eigenvalues of N random adjacency matrices of $G(n, 5/n)$ compared with the first three truncations of the expansion.

9 Cluster graphs

This chapter presents our work with Martijn Gösgens, Lukas Lühtrath, Elena Magnanini and Marc Noy (Gösgens et al. 2024), begun in August 2022 at a workshop of the RandNET project in Eindhoven, where Gösgens presented the problem at the open problem session. Its contribution is the asymptotic behavior of the random cluster graph $G(n, p)_C$, the Erdős–Rényi graph conditioned on having all its components complete. That model is the prior implicit in modularity maximization, so these statistics say how to set the resolution parameter of a community detection algorithm. The chapter is an example of the two approaches, probabilistic and enumerative, being combined: a probabilistic argument identifies the random model, and analytic combinatorics studies it.

Motivation. Community detection aims at partitioning the vertices of a graph into well-connected groups, and one of the most used methods is to maximize the *modularity* of the partition (Newman and Girvan 2004), for instance by the Louvain algorithm (Blondel et al. 2008). Modularity measures the excess of edges inside the communities over their expected number under a *null model*, a random graph without community structure. In its original form it has a *resolution limit* (Fortunato and Barthelemy 2007): in a network with m edges, two communities joined by a single edge are merged by modularity maximization as soon as their numbers of internal edges are below a threshold of order $\sqrt{m}$, so that small communities cannot be resolved in large networks. The *resolution parameter* γ of Reichardt and Bornholdt (2006)

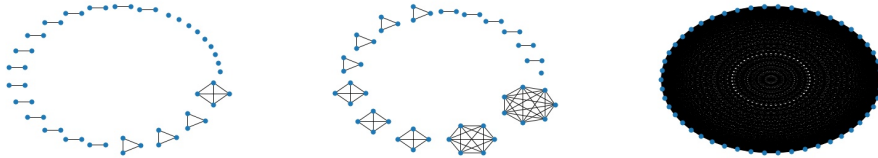


Figure 7: Samples of three random cluster graphs on 50 vertices. The one on the left is sampled with $p = 0.25$, the middle one with $p = 0.51$ and the right one with $p = 0.53$, resulting in the complete graph.

remedies this by fixing the granularity of the communities sought. A partition of the vertex set is represented as a *cluster graph* (Figure 7), a graph whose components are cliques, and for the Erdős–Rényi null model the modularity of the partition G_C of a graph G is

$$\text{ERM}(G_C, G, \gamma) = \frac{1}{m(G)}(m(G \cap G_C) - \gamma m(G_C)), \quad (9.1)$$

where $m(G \cap G_C)$ counts the edges of G inside communities and $m(G_C)$ the pairs of vertices in a common community. Extending an interpretation of Newman (2016) for the unparametrized case, Martijn Gösgens showed that maximizing this modularity is a Bayesian inference, for a natural pair of prior and likelihood. The prior on partitions is the *random cluster graph* $G(n, p)_C$, obtained by drawing $G(n, p)$ and conditioning on the rare event that it is a cluster graph. Given a cluster graph G_C , the observed graph is drawn from the *planted partition model* (Holland, Laskey, and Leinhardt 1983): each pair of vertices is joined by an edge independently, with probability p_{in} if the two vertices belong to the same clique of G_C and p_{out} otherwise, so that for $p_{\text{in}} > p_{\text{out}}$ the communities are denser than their surroundings. The partition maximizing the posterior probability given the observed graph G is then the one maximizing the modularity $\text{ERM}(G_C, G, \gamma)$ for the explicit value

$$\gamma = \frac{\log\left(\frac{1-p_{\text{out}}}{1-p_{\text{in}}} \frac{1-p}{p}\right)}{\log\left(\frac{p_{\text{in}}}{p_{\text{out}}} \frac{1-p_{\text{out}}}{1-p_{\text{in}}}\right)}. \quad (9.2)$$

Maximizing modularity thus assumes a prior on partitions, namely $G(n, p)_C$. Choosing γ well means translating prior knowledge on the communities, their number or their sizes, into a value of p , which calls for the statistics of $G(n, p)_C$.

Enumeration. For p not extremely close to 0 or 1, the conditioning event has vanishing probability and distorts $G(n, p)$ beyond the reach of the usual probabilistic tools. On the other hand, cluster graphs, being sets of cliques, are amenable to the symbolic method. The formula of Chapter 3 for $G(n, p)$ applies with $w = p/(1-p)$. Writing $C(z, w) = \sum_{n \geq 1} w^{\binom{n}{2}} z^n / n!$ for the generating function of cliques and $C_1 = z \partial_z C$ for that of cliques with a distinguished vertex, the numbers $\mathbf{C}$ of cliques and $\mathbf{M}$ of edges of $G(n, p)_C$, and the degree $\mathbf{D}$ of a uniform vertex, have probability generating functions

$$\begin{aligned} \mathbb{E}(u^{\mathbf{C}}) &= \frac{[z^n] e^{uC(z, w)}}{[z^n] e^{C(z, w)}}, \quad \mathbb{E}(u^{\mathbf{M}}) = \frac{[z^n] e^{C(z, uw)}}{[z^n] e^{C(z, w)}}, \\ \mathbb{E}(u^{\mathbf{D}}) &= \frac{[z^n] C_1(uz, w) e^{C(z, w)}}{u [z^n] C_1(z, w) e^{C(z, w)}}. \end{aligned}$$

At $w = 1$, that is $p = 1/2$, these are the statistics of a uniform set partition, whose number of blocks is asymptotically normal with mean $\sim n/\log n$ (Harper 1967). The interest lies on either side, where the coefficients $w^{\binom{n}{2}}$ grow or decay quadratically in the exponent and the analysis changes nature.

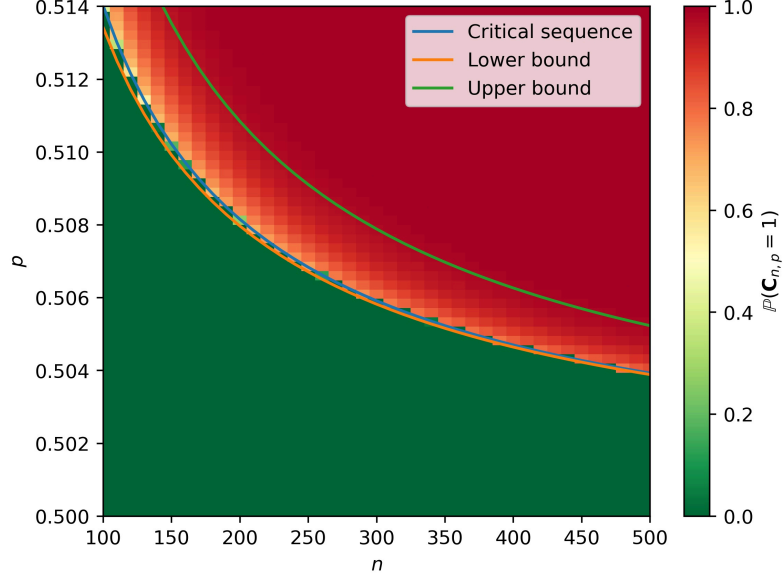


Figure 8: The critical sequence $p_n(1/2)$ with the upper and lower bounds we derived. The background shades the probability that a single clique remains, showing how narrow the window is.

Results. We present the number of cliques; edges and degrees are treated in Gösgens et al. (2024). For $p < 1/2$, $\mathbf{C}$ satisfies a central limit theorem, with

$$\mathbb{E}(\mathbf{C}) \sim \sqrt{\frac{\log(1-p) - \log p}{2}} \frac{n}{\sqrt{\log n}}, \quad \mathbb{V}(\mathbf{C}) = \Theta\left(\frac{n}{(\log n)^{3/2}}\right),$$

the variance containing a term oscillating with n that we hide in the Θ ; the proof is a saddle point analysis guided by Hayman admissibility, more intricate than we expected. For $p > 1/2$ we have $w > 1$, so $C(z, w)$ has zero radius of convergence: the framework of Dovgal and Nurligareev (2026) for divergent series of graph-like structures applies and shows that $G(n, p)_C$ is the complete graph with probability tending to 1. The mean number of cliques thus drops from $\Theta(n/\sqrt{\log n})$ to 1 across $p = 1/2$, and the transition is sharp, as illustrated in Figure 8: writing $p_n(q)$ for the value at which the probability of a single clique equals $q \in (0, 1)$,

$$p_n(q) = \frac{1}{2} + \frac{\log n}{2n} + O\left(\frac{\log \log n}{n}\right),$$

independently of q at this order.

Numerical experiments by Martijn Gösgens confirm that choosing p by this rule improves the partitions detected when the number of communities is bounded, and

that perturbing p by an amount of order $\log n/n$ visibly changes them, as Figure 9 illustrates. The sharpness of the transition is what makes so small a perturbation matter.

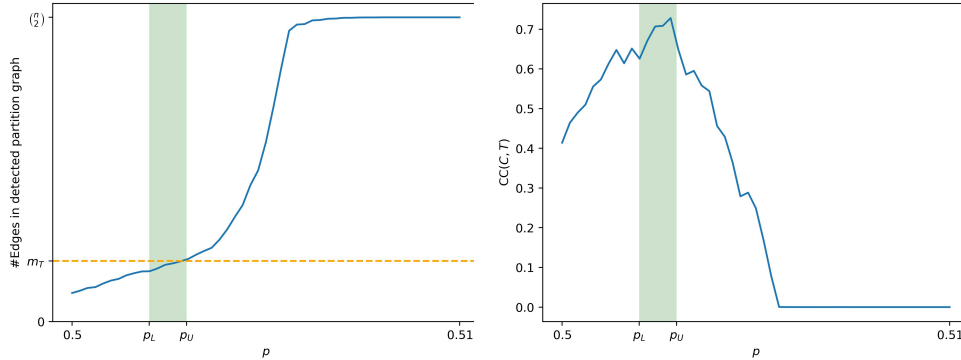


Figure 9: Twenty planted partition models with $n = 1000$ vertices in 5 communities of size 200, with p_{in} and p_{out} chosen so that each vertex expects 10 neighbors inside its community and 10 outside. Communities are detected by maximizing ERM (Equation (9.1)) with resolution parameter $\gamma(p_{\text{in}}, p_{\text{out}}, p)$ (Equation (9.2)), for various p . Left: the number of edges in the detected cluster graph. Right: the correlation coefficient between the detected and the true partition.

Open problems. Two directions remain. Since p should be chosen inside the window above when the number of communities stays bounded, the first term of $p_n(q)$ that depends on q would allow tuning p to prior knowledge on that number; it lies below the $\log \log n/n$ we control. The analysis should also be extended to modularity with the Chung–Lu null model, which accounts for the degrees and is the one used in practice, and which is close to the configuration model of Section 4.1. There the Bayesian interpretation has no known counterpart, and that is what currently stops us.

10 Active clustering

This chapter presents our work with Quentin Lutz, Alex Scott and Maya Stein (Lutz et al. 2021). It answers three questions about the algorithms that recover a hidden set partition from pairwise queries: the algorithms of minimal average complexity are exactly the *chordal* ones, all of them share the same distribution of the number of queries, and that distribution is a q -analog of Dobiński’s formula, with a Gaussian limit law.

The question came from Maria Laura Maag, researcher and engineer at Nokia Bell Labs, about the building of a training set for supervised learning. We explored it with Quentin Lutz, a PhD student co-supervised with Thomas Bonald, and Alex Scott and Maya Stein joined through the RandNET project. We are fond of this work for its industrial motivation, its elegant theoretical framing, and its resolution combining probabilistic, graph-theoretic and analytic arguments, one for each question.

Motivation. Customer service for Nokia products relies on a ticketing system in which clients describe the issues they face. Most tickets fall into categories with a known resolution procedure. An automatic response system reads each ticket, decides whether it belongs to a known category, and either sends the appropriate answer or forwards the ticket to human experts. Supervised learning is the method of choice for such a system, and requires a training set of tickets labeled with their category: about ten thousand tickets across some seventy categories, in the case brought to us. Building it is tedious for the experts, since assigning a category to a ticket means scanning a large part of the list of seventy categories each time. Maria Laura Maag suggested an alternative. Instead of assigning categories, the experts are shown pairs of tickets and asked whether the two belong to the same category, a binary comparison which is markedly easier. The problem is then to choose which pairs to present, so as to recover the category of every ticket with as few questions as possible.

The problem. An *active clustering algorithm* receives a finite set S of n items and recovers a hidden partition σ of S by submitting queries to an oracle, each query being a pair of items, each answer telling whether the two lie in the same block of σ . Its *complexity* is the number of queries. The worst case is uninformative: when σ has only singletons, every pair must be queried, and the complexity is $\binom{n}{2}$ for every algorithm. The relevant measure is the average complexity for a random hidden partition, and we take σ uniform among the set partitions of S ; Lutz et al. (2021) also treats a model closer to the application, discussed at the end.

The problem is a sibling of sorting, where the hidden object is a linear order and the oracle compares two items. The algorithms of minimal average complexity for sorting are not known for all n (Peczarski 2004) only algorithms of optimal *asymptotic* complexity are known. Partitions are simpler combinatorial objects than permutations, and we obtain a complete answer: a characterization of all algorithms of *exactly* optimal complexity.

In machine learning, pairwise queries are known to be preferred by human annotators (Mazumdar and Saha 2017; Chien, Zhou, and Li 2019) and appear as must-link and cannot-link constraints in semi-supervised clustering (Wagstaff et al. 2001; Basu, Bilenko, and Mooney 2004; Gribel, Gendreau, and Vidal 2022), with adaptive choice of the pairs in its active variant (Basu, Banerjee, and Mooney 2004; Mazumdar and Saha 2017) those settings assume a similarity measure between items, whereas ours assumes nothing but the oracle. The problem also belongs to combinatorial search (Aigner 1988; Grebinski and Kucherov 1999; Bouvel, Grebinski, and Kucherov 2005; Alon and Asodi 2005) and to entity resolution (Fellegi and Sunter 1969; Wang et al. 2013; Vesdapunt, Bellare, and Dalvi 2014).

Query graphs and query trees. If u and v belong to the same block, and v and w belong to the same block, then so do u and w . A query is *trivial* when its

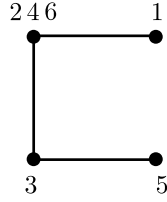


Figure 10: The query graph on $\{1, \dots, 6\}$ after the queries $\{2, 4\}$, $\{1, 2\}$, $\{4, 6\}$, $\{3, 4\}$, $\{3, 5\}$, answered yes, no, yes, no, no.

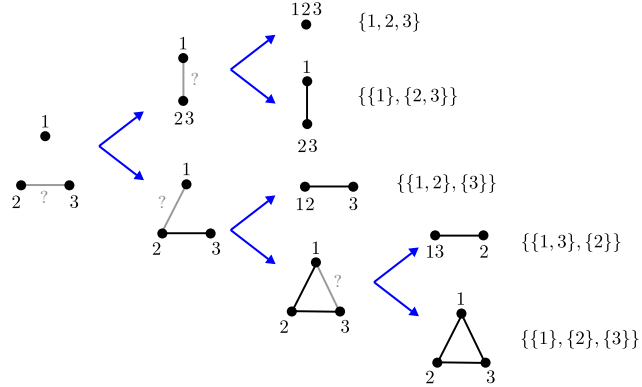


Figure 11: A query tree on $S = \{1, 2, 3\}$, with the query graph at each node and the hidden partition at each leaf.

answer follows from the previous ones by transitivity, and an algorithm never needs to ask one. The information gathered during an execution is held in the *query graph*: its vertices are sets of items known to lie in the same block, and its edges join vertices known to lie in different blocks. A positive answer merges two vertices, a negative one adds an edge, so all queries between two given vertices have the same answer and can be identified; there is no non-trivial query left exactly when the query graph is complete, and its vertices are then the blocks of σ (Figure 10). A deterministic algorithm is then a rooted binary tree, its *query tree*: each internal node holds a query, its two children the two answers, and each leaf the partition recovered (Figure 11). The average complexity is the average depth of the leaves, each weighted by the probability of its partition, which in the uniform case is the same for all.

Optimal algorithms. Our first conjecture was that all algorithms asking no trivial query have the same average complexity; Figure 12 disproved it. Exhaustive search over all query trees on up to 13 items, up to symmetries, suggested the right statement, which we presented at a RandNET meeting, after which Alex Scott joined us and proved it.

A graph is *chordal* if every cycle on four or more vertices has a chord, an edge joining two of its vertices but not in the cycle; equivalently, its induced cycles are triangles (Diestel 2025). Call an algorithm chordal when every query graph in its query tree is chordal. Since chordality is preserved by the contraction of an edge, a query keeps the graph chordal whatever the answer as soon as adding the corresponding edge does, and this happens if and only if the common neighborhood of its two vertices separates them.

An algorithm has minimal average complexity if and only if it is chordal. The proof is an elegant probabilistic argument. Positive answers merge vertices, so their number, n minus the number of blocks, does not depend on the algorithm, and only the queries with negative answers have to be minimized. An algorithm minimizing the expected number of negative queries between any two blocks, if one exists, therefore reaches the minimal possible complexity, which reduces

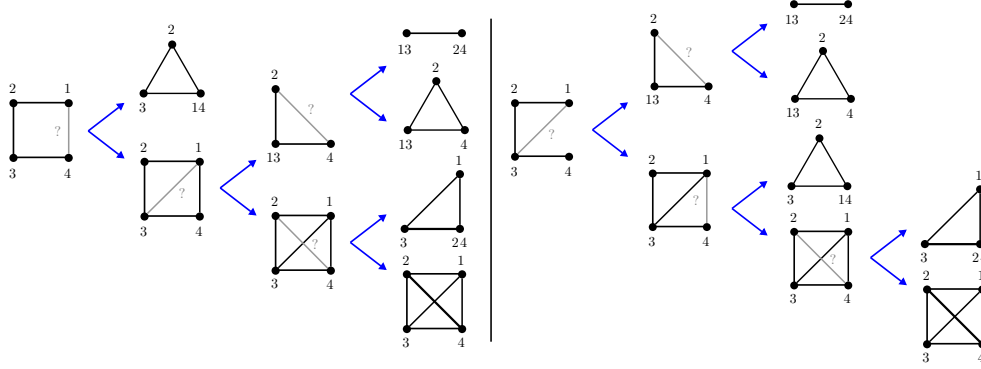


Figure 12: From a partially filled query graph, two possible next queries lead to average complexities $13/5$ and $12/5$: some queries are better than others.

the problem to two blocks. There, queries between two vertices linked by an odd path are wasteful, since the answer is already determined, and a probabilistic argument shows that the expected number of the other negative queries does not depend on the algorithm. That the chordal algorithms are exactly those minimizing the wasteful queries follows from the fact that the only bipartite induced subgraphs of a chordal graph are forests.

All optimal algorithms are alike. The optimal algorithms being identified, the next question is which of them has the smallest variance. We proved that *all chordal algorithms have the same distribution of the number of queries*, on every starting chordal query graph. The proof is graph-theoretic, by induction on the number of non-edges, and its idea is that two chordal algorithms differing in their first query can be compared by exchanging the order of two queries.

The distribution. Since all chordal algorithms share their distribution, it can be computed on the simplest one, the *universal algorithm*: take an item, compare it with every other to find its block, remove the block, and iterate. Let

$$P(z, q) = \sum_{\sigma} q^{\text{queries}(\sigma)} \frac{z^{|\sigma|}}{|\sigma|!}$$

be the generating function of set partitions, with an exponential variable z marking their sizes and an ordinary variable q the number of queries of the universal algorithm. The block of the first item, together with b further items, costs $|r| + b$ queries, where r is the partition of the remaining items, and the whole is a relabeled pair, so that

$$\partial_z P(z, q) = P(qz, q) e^{qz}.$$

Its solution is $P(z, q) = A(z, q) e^{qz/(1-q)}$ with A explicit in terms of the q -Pochhammer symbol $(a; q)_k = \prod_{j=0}^{k-1} (1 - aq^j)$. Writing B_n for the number of set partitions of n items and introducing the classical q -analogs of integers, factorials and the exponential function

$$[m]_q = 1 + q + \dots + q^{m-1}, \quad [m]_q! = \prod_{\ell=1}^m [\ell]_q, \quad \exp_q(z) = \sum_{m \geq 0} \frac{z^m}{[m]_q!},$$

the probability generating function of the number of queries of a chordal algorithm on n items is

$$\frac{1}{B_n} \left(\frac{q}{1-q} \right)^n \sum_{k=0}^n \binom{n}{k} (-1)^k \left(\frac{1-q}{q}; q \right)_k = \frac{1}{B_n} \frac{1}{\exp_q(1/q)} \sum_{m \geq 0} \frac{[m]_q^n}{[m]_q!} q^{n-m}. \quad (10.1)$$

The first form computes the distribution for any given n . We derived the second, which converges for $q > 1/2$, to extract the limit law; at $q = 1$ it reduces to Dobiński's formula $B_n = e^{-1} \sum_{m \geq 0} m^n / m!$, of which it is a q -analog.

The limit law. Let X_n be the number of queries of a chordal algorithm on n items, and W the Lambert function, so that $W(x)e^{W(x)} = x$ for $x \geq 0$. Then $(X_n - E_n)/\sigma_n$ converges in distribution to a standard Gaussian, with

$$E_n = \frac{1}{4}(2W(n) - 1)e^{2W(n)}, \quad \sigma_n = \frac{1}{3} \sqrt{\frac{3W(n)^2 - 4W(n) + 2}{W(n) + 1}} e^{3W(n)}.$$

The proof evaluates the Laplace transform of the normalized variable from the second form of Equation (10.1): a q -analog of Stirling's formula (Moak 1984) approximates the summands, and the Laplace method for sums, a variant of the method of Section 2.2, extracts the asymptotics of the sum, whose dominant terms have index m growing with n . Since $W(n) = \log n - \log \log n + o(1)$, the optimal average complexity is $\binom{n}{2} / \log n$, a factor $\log n$ below the trivial upper bound $\binom{n}{2}$ and a factor $\log n$ above the lower bound $\binom{n/\log n}{2}$ obtained by requiring a query between each pair of the $\sim n/\log n$ blocks of a typical partition.

Beyond the model. These results identify all active clustering algorithms reaching the optimal complexity, settling the analog of the sorting problem for random set partitions. The uniform partition model is not realistic for the machine learning application, however, since its $n/\log n$ blocks of size about $\log n$ are too small for a classifier to learn from. In Lutz et al. (2021) we also analyzed a model in which each item independently chooses one of a fixed number of blocks following some unknown distribution, and found pairwise queries to be more effective there. We studied two models of oracle errors as well, a bounded number of them to be corrected exactly and independent errors of small probability to be corrected with high probability given a budget of extra queries, and gave algorithms for both. Finally, in practice a similarity heuristic between items avoids many negative queries, so that for the industrial application our results are a baseline rather than a recommendation.

11 Perspectives

The supercritical regime and the missing black box. Although the statistics of the giant component of supercritical random graphs are well understood, we would like to derive them again from the exact expression Equation (5.7), through contour

integrals, as explained in Section 5.4. The same obstruction blocks the asymptotic enumeration of DAGs in the (n, m) model above the transition, since Equation (6.11) has the same shape. A black box theorem for these extractions would settle both problems at once. Packaging the analysis of Section 6.3, carried out by hand, into a black box theorem would also open the way to finer statistics of $\text{MD}(n, p)$, starting with those of the giant strong component.

Directed acyclic graphs with constrained degrees. The work in progress of Section 6.2 raises three questions that we intend to settle in the near term. The simple DAG model, where the sources choose distinct targets, should follow by an inversion in the manner of Section 4.3. The asymptotics, for which preliminary work shows at least three regimes according to the growth of the out-degree weights, require singularity analysis in some regimes and large deviations in others, and the picture is not yet settled. The coincidence between DAGs with out-degree constraints and parking functions calls for a bijection, which in the case of out-degrees at most 1 should specialize to the classical correspondences between forests and parking functions.

Constraint satisfaction problems. The first target is a closed form for the limit probability of satisfiability in the 2-SAT window, defined in Equation (7.1), for the reasons given at the end of Chapter 7. The broader ambition is to treat families of constraint satisfaction problems at once. Their worst-case complexity is completely classified by the dichotomy theorem, whereas the probabilistic counterpart, which families have a satisfiability phase transition and where, is far less understood. An entry point combines the two models of our PhD: the hypergraphs of Panafieu (2015), whose edges have arbitrary sizes and carry weights depending on their size, and the inhomogeneous graphs of Panafieu (2015); Panafieu and Ravelomanana (2015), whose vertices carry types. In an *inhomogeneous hypergraph* each vertex has a type in $\{1, \dots, q\}$, and each edge, a sequence of vertices, carries a weight ω_t depending on its type composition $t \in \mathbb{N}^q$. With $\Omega(x) = \sum_t \omega_t x^t / t!$, all of them have generating function

$$\text{IHG}(z, w, x) = \sum_{n \in \mathbb{N}^q} e^{w\Omega(xn)} \frac{z^{|n|}}{n!},$$

the counterpart of $\text{MG}(z, w)$, to which it reduces for $q = 1$ and $\Omega(x) = x^2/2$. Reading a value as a type and a constraint applied to a sequence of variables as an edge, the instances of a constraint satisfaction problem together with one of their solutions are exactly the inhomogeneous hypergraphs whose weight ω_t counts the constraints satisfied by the arrangements of composition t . The generating function above is then the first moment of the number of solutions, and its version with q^2 types, for pairs of solutions, the second moment: the two quantities on which the location of satisfiability thresholds rests (Achlioptas and Peres 2004; Achlioptas and Coja-Oghlan 2008). The kernel decomposition and the critical window analysis of Chapter 5 carry over, with a determinant built from the Hessian of Ω in place of $1 - T$, as we checked in two special cases during our PhD.

Book. The long version of this manuscript, from which it was cut, is closer to a book than to a habilitation thesis, and we intend to make it one. Its aim is to show that graph enumeration by analytic combinatorics rests on a few transferable

ideas and to lower the entry cost enough that more researchers, doctoral students in particular, take up the field.

Beyond this manuscript. The other half of our research concerns computational social choice, with François Durand and Jérôme Lang, within the project PROSOC, which we coordinate with Frédérique Bassino and Olivier Spanjaard: the probability of manipulability of the classical voting rules under various cultures, the identification of the least manipulable reasonable rule, the probability of a Condorcet winner when voters and candidates both grow in number, and the multiwinner rules, where approval ballots turn the profile into a random hypergraph and bring the problem back into the world of this manuscript.

Bibliography

- Achlioptas, Dimitris, and Amin Coja-Oghlan. 2008. “Algorithmic Barriers from Phase Transitions.” *arXiv:0803.2122*.
- Achlioptas, Dimitris, and Yuval Peres. 2004. “The Threshold for Random k -SAT Is $2^k \log 2 - O(k)$.” *Journal of the American Mathematical Society* 17(4):947–973.
- Aigner, Martin. 1988. *Combinatorial Search*. John Wiley & Sons, Inc.
- Alon, Noga, and Vera Asodi. 2005. “Learning a Hidden Subgraph.” *SIAM Journal on Discrete Mathematics* 18(4):697–712.
- Aspvall, Bengt, Michael F Plass, and Robert Endre Tarjan. 1979. “A Linear-Time Algorithm for Testing the Truth of Certain Quantified Boolean Formulas.” *Information processing letters* 8(3):121–123.
- Banderier, Cyril, Philippe Flajolet, Gilles Schaeffer, and Michele Soria. 2001. “Random Maps, Coalescing Saddles, Singularity Analysis, And Airy Phenomena.” *Random Structures & Algorithms* 19(3–4):194–246.
- Bassino, F., J. Clément, J. Fayolle, and P. Nicodème. 2008. “Constructions for Clumps Statistics.” In *Fifth Colloquium on Mathematics and Computer Science*, 179–194. Nancy: The Association. Discrete Mathematics & Theoretical Computer Science (DMTCS).
- Basu, Sugato, Arindam Banerjee, and Raymond J Mooney. 2004. “Active Semi-Supervision for Pairwise Constrained Clustering.” In *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM)*, 333–344.
- Basu, Sugato, Mikhail Bilenko, and Raymond J Mooney. 2004. “A Probabilistic Framework for Semi-Supervised Clustering.” In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 59–68.
- Bauer, Michel, and Olivier Golinelli. 2001. “Random Incidence Matrices: Moments of the Spectral Density.” *Journal of Statistical Physics* 103:301–337.
- Bender, E. A., L. B. Richmond, R. W. Robinson, and N. C. Wormald. 1986. “The Asymptotic Number of Acyclic Digraphs. I.” *Combinatorica* 6:15–22.
- Bender, Edward A. 1975. “An Asymptotic Expansion for the Coefficients of Some Formal Power Series.” *Journal of the London Mathematical Society* 2(3):451–458.

- Bender, Edward A., and E. Rodney Canfield. 1978. "The Asymptotic Number of Labeled Graphs with Given Degree Sequences." *Journal of Combinatorial Theory, Series A* 24(3):296–307.
- Bender, Edward A., E. Rodney Canfield, and Brendan D. McKay. 1990. "The Asymptotic Number of Labeled Connected Graphs with a Given Number of Vertices and Edges." *Random Structures & Algorithm* 1:129–169.
- Bergeron, François, Gilbert Labelle, and Pierre Leroux. 1997. *Combinatorial Species and Tree-Like Structures*. Cambridge University Press.
- Bessis, Daniel, Claude Itzykson, and Jean-Bernard Zuber. 1980. "Quantum Field Theory Techniques in Graphical Enumeration." *Adv. Appl. Math.* 1(2):109–157.
- Bianconi, Ginestra, and Matteo Marsili. 2005. "Loops of Any Size and Hamilton Cycles in Random Scale-Free Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2005(6):14.
- Birkhoff, George D, and Waldemar J Trjitzinsky. 1932. "Analytic Theory of Singular Difference Equations." *Acta Mathematica* 60:1–89.
- Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of statistical mechanics: theory and experiment* 2008(10):P10008.
- Bollobás, Béla. 1980. "A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs." *European Journal of Combinatorics* 1:311–316.
- Bollobás, Béla. 1981a. "Threshold Functions for Small Subgraphs." *Mathematical Proceedings of the Cambridge Philosophical Society* 90:197–206.
- Bollobás, Béla. 1981b. *Random Graphs*.
- Bollobás, Béla. 1984. "The Evolution of Random Graphs." *Transactions of the American Mathematical Society* 286(1):257–274.
- Bollobás, Béla, Christian Borgs, Jennifer T Chayes, Jeong Han Kim, and David B Wilson. 2001. "The Scaling Window of the 2-SAT Transition." *Random Structures & Algorithms* 18(3):201–256.
- Bordenave, Charles, Arnab Sen, and Bálint Virág. 2017. "Mean Quantum Percolation." *Journal of the European Mathematical Society (JEMS)* 19(12):3679–3707.
- Borinsky, Michael. 2017. "Renormalized Asymptotic Enumeration of Feynman Diagrams." *Annals of Physics* 385:95–135.
- Borinsky, Michael. 2018a. "Generating Asymptotics for Factorially Divergent Sequences." *The Electronic Journal of Combinatorics* 25(4):32.
- Borinsky, Michael. 2018b. *Graphs in Perturbation Theory: Algebraic Structure and Asymptotics*. Springer.
- Bouillard, Anne, Céline Comte, Élie de Panafieu, and Fabien Mathieu. 2018. "Of Kernels and Queues: When Network Calculus Meets Analytic Combinatorics." In *Proceedings of the 30th International Teletraffic Congress (ITC)*, 49–54. IEEE.
- Bouvel, Mathilde, Vladimir Grebinski, and Gregory Kucherov. 2005. "Combinatorial Search on Graphs Motivated by Bioinformatics Applications: A Brief Survey." In *International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, 16–27.
- Bruijn, N. G. de. 1981. *Asymptotic Methods in Analysis. Corr. Reprint of the 3rd Ed.*

- Caizergues, Emma, and Élie de Panafieu. 2023. “Exact Enumeration of Graphs and Bipartite Graphs with Degree Constraints.” In *European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB)*.
- Caizergues, Emma, François Durand, Marc Noy, Élie de Panafieu, and Vladyslav Ravelomanana. 2026. “Probability of a Condorcet Winner for Large Electorates: An Analytic Combinatorics Approach.” *Journal of Artificial Intelligence Research*.
- Chatterjee, Arnab, Amin Coja-Oghlan, Noela Müller, Connor Riddlesden, Maurice Rolvien, Pavel Zakharov, and Haodong Zhu. 2024. “The Number of Random 2-SAT Solutions Is Asymptotically Log-Normal.” *arXiv:2405.03302*.
- Chayes, JT, L Chayes, Judy R Franz, James P Sethna, and SA Trugman. 1986. “On the Density of State for the Quantum Percolation Problem.” *Journal of Physics A: Mathematical and General* 19(18):L1173.
- Chen, Chung Shue, Peter Keevash, Sean Kennedy, Élie de Panafieu, and Adrian Vetta. 2024. “Robot Positioning Using Torus Packing for Multisets.” In *Proceedings of the 51st International Colloquium on Automata, Languages and Programming (ICALP)*, 43:1–43:18. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik.
- Chien, I Eli, Huozhi Zhou, and Pan Li. 2019. “ HS^2 : Active Learning over Hypergraphs with Pointwise and Pairwise Queries”. In *The 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2466–2475.
- Chvátal, Vášek, and Bruce Reed. 1992. “Mick Gets Some (The Odds Are on His Side) (Satisfiability).” In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science (FOCS)*, 620–627.
- Chyzak, Frédéric, and Marni Mishna. 2026. “Differential Equations Satisfied by Generating Functions of 5-, 6-, and 7-Regular Labelled Graphs: A Reduction-Based Approach.” *Combinatorial Theory* 6(1).
- Chyzak, Frédéric, Marni Mishna, and Bruno Salvy. 2005. “Effective Scalar Products of D-Finite Symmetric Functions.” *Journal of Combinatorial Theory. Series A* 112(1):1–43.
- Collet, Gwendal, Élie de Panafieu, Danièle Gardy, Bernhard Gittenberger, and Vladyslav Ravelomanana. 2020. “Threshold Functions for Small Subgraphs in Simple Graphs and Multigraphs.” *European Journal of Combinatorics* 88:43.
- Curtiss, J. H. 1942. “A Note on the Theory of Moment Generating Functions.” *Annals of Mathematical Statistics* 13:430–433.
- Deroulers, Christophe, and Rémi Monasson. 2006. “Criticality and Universality in the Unit-Propagation Search Rule.” *The European Physical Journal B-Condensed Matter and Complex Systems* 49(3):339–369.
- Diestel, Reinhard. 2025. *Graph Theory*. Springer Nature.
- Doubilet, Peter, Gian-Carlo Rota, and Richard Stanley. 1972. “On the Foundations of Combinatorial Theory. VI: The Idea of Generating Function.”
- Dougherty-Bliss, Robert, Christoph Koutschan, Natalya Ter-Saakov, and Doron Zeilberger. 2025. “The (Symbolic and Numeric) Computational Challenges of Counting 0-1 Balanced Matrices.” *Enumerative Combinatorics and Applications* 5(2):8.

- Dovgal, Sergey. 2019. “The Birth of the Contradictory Component in Random 2-Sat.” *arXiv:1904.10266*.
- Dovgal, Sergey, and Khaydar Nurligareev. 2026. “Asymptotics for Graphically Divergent Series: Dense Digraphs and 2-SAT Formulae.” *The Electronic Journal of Combinatorics* 33(3):63.
- Dovgal, Sergey, Élie de Panafieu, and Vlady Ravelomanana. 2023. “Exact Enumeration of Satisfiable 2-SAT Formulae.” *Combinatorial Theory* 3(2):38.
- Dovgal, Sergey, Élie de Panafieu, Dimbinaina Ralaivaosaona, Vonjy Rasendrasahina, and Stephan Wagner. 2024. “The Birth of the Strong Components.” *Random Structures & Algorithms* 64(2):170–266.
- Duchon, Philippe, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. 2004. “Boltzmann Samplers for the Random Generation of Combinatorial Structures.” *Combinatorics, Probability and Computing* 13(4–5):577–625.
- Durand, François, Élie de Panafieu, and Guillem Perarnau. 2026. “Super Condorcet Winners and Limit Coalitional Manipulability of IRV.” In *Proceedings of the 35th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Enriquez, Nathanaël, and Laurent Ménard. 2016. “Spectra of Large Diluted but Bushy Random Graphs.” *Random Structures & Algorithms* 49(1):160–184.
- Erdős, Paul, and Alfréd Rényi. 1960. “On the Evolution of Random Graphs.” *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*.
- Etingof, Pavel. 2002. “Mathematical Ideas and Notions of Quantum Field Theory.” Lecture notes, MIT.
- Evnin, Oleg, and Weerawit Horinouchi. 2024. “A Gaussian Integral That Counts Regular Graphs.” *Journal of Mathematical Physics* 65(9).
- Fellegi, Ivan P, and Alan B Sunter. 1969. “A Theory for Record Linkage.” *Journal of the American statistical association* 64(328):1183–1210.
- Fernandez de la Vega, W. 2001. “Random 2-SAT: Results and Problems.” *Theor. Comput. Sci.* 265(1–2):131–146.
- Flajolet, Philippe, and Andrew Odlyzko. 1990. “Singularity Analysis of Generating Functions.” *SIAM Journal on discrete mathematics* 3(2):216–240.
- Flajolet, Philippe, and Robert Sedgewick. 2009. *Analytic Combinatorics*. Cambridge: Cambridge University Press.
- Flajolet, Philippe, Donald E. Knuth, and Boris Pittel. 1989. “The First Cycles in an Evolving Graph.” *Discrete Mathematics* 75(1–3):167–215.
- Flajolet, Philippe, Bruno Salvy, and Gilles Schaeffer. 2004. “Airy Phenomena and Analytic Combinatorics of Connected Graphs.” *Electronic Journal of Combinatorics* 11(1).
- Foata, Dominique, and John Riordan. 1974. “Mappings of Acyclic and Parking Functions.” *Aequationes Mathematicae* 10:10–22.
- Fortunato, Santo, and Marc Barthelemy. 2007. “Resolution Limit in Community Detection.” *Proceedings of the national academy of sciences* 104(1):36–41.
- Gardy, Danièle. 1995. “Some Results on the Asymptotic Behaviour of Coefficients of Large Powers of Functions.” *Discrete Math.* 139:189–217.
- Gessel, Ira M. 1995. “Enumerative Applications of a Decomposition for Graphs and Digraphs.” *Discrete mathematics* 139(1–3):257–271.

- Gessel, Ira M. 1990. "Symmetric Functions and P-Recursiveness." *Journal of Combinatorial Theory, Series A* 53:257–285.
- Goerdt, Andreas. 1996. "A Threshold for Unsatisfiability." *Journal of Computer and System Sciences* 53(3):469–486.
- Goldschmidt, Christina, and Robin Stephenson. 2023. "The Scaling Limit of a Critical Random Directed Graph." *The Annals of Applied Probability* 33(3):2024–2065.
- Gösgens, Martijn, Lukas Lühtrath, Elena Magnanini, Marc Noy, and Élie de Panafieu. 2024. "The Erdős-Rényi Random Graph Conditioned on Every Component Being a Clique." *arXiv:2405.13454*.
- Goulden, I. P., and D. M. Jackson. 1979. "An Inversion Theorem for Cluster Decompositions of Sequences with Distinguished Subsequences." *Journal of the London Mathematical Society. Second Series* 20:567–576.
- Goulden, I. P., D. M. Jackson, and J. W. Reilly. 1983. "The Hammond Series of a Symmetric Function and Its Application to P-Recursiveness." *SIAM Journal on Algebraic Discrete Methods* 4(2):179–193.
- Grebinski, Vladimir, and Gregory Kucherov. 1999. "Reconstructing Set Partitions." In *Tenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2.
- Gribel, Daniel, Michel Gendreau, and Thibaut Vidal. 2022. "Semi-Supervised Clustering with Inaccurate Pairwise Annotations." *Information Sciences* 607:441–457.
- Guibas, L. J., and A. M. Odlyzko. 1981. "String Overlaps, Pattern Matching, And Nontransitive Games." *Journal of Combinatorial Theory. Series A* 30:183–208.
- Gunnells, Paul E. 2021. "Generalized Catalan Numbers from Hypergraphs." *The Electronic Journal of Combinatorics* 28(1):29.
- Gunnells, Paul E. 2024. "Hypergraph Matrix Models and Generating Functions." *Combinatorics and Number Theory* 13(2):149–178.
- Hainzl, Eva-Maria. 2026. "Formulas and Asymptotics of Hypergraph Catalan Numbers." *arXiv:2508.08865*.
- Hainzl, Eva-Maria, and Élie de Panafieu. 2026. "Tree Walks and the Spectrum of Random Graphs." *European Journal of Combinatorics* 136:26.
- Harary, Frank. 1970. "Covering and Packing in Graphs, I." *Annals of the New York Academy of Sciences* 175(1):198–205.
- Harper, LH. 1967. "Stirling Behavior Is Asymptotically Normal." *The Annals of Mathematical Statistics* 38(2):410–414.
- Hayman, Walter K. 1956. "A Generalisation of Stirling's Formula." *J. Reine Angew. Math.* 196:67–95.
- Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. "Stochastic Blockmodels: First Steps." *Social networks* 5(2):109–137.
- Hwang, Hsien-Kuei. 1998. "On Convergence Rates in the Central Limit Theorems for Combinatorial Structures." *European Journal of Combinatorics* 19(3):329–343.
- Isaev, Mikhail, and Brendan D McKay. 2025. "Asymptotic Enumeration of Graph Factors by Cumulant Expansion." *arXiv:2508.18731*.
- Jacquet, Philippe, and Wojciech Szpankowski. 1998. "Analytical Depoissonization and Its Applications." *Theoretical Computer Science* 201(1–2):1–62.
- Janson, Svante, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel. 1993. "The Birth of the Giant Component." *Random Structures & Algorithms* 4(3):233–358.

- Jenssen, Matthew, Will Perkins, and Aditya Potukuchi. 2025. “On the Evolution of Structure in Triangle-Free Graphs.” *Advances in Mathematics* 480:127.
- Karoński, Michał, and Andrzej Ruciński. 1983. “On the Number of Strictly Balanced Subgraphs of a Random Graph.”
- Karp, Richard M. 1990. “The Transitive Closure of a Random Digraph.” *Random Structures & Algorithms* 1(1):73–93.
- Kauers, Manuel. 2011. “A Mathematica Package for Computing Asymptotic Expansions of Solutions of P-Finite Recurrence Equations.” *RISC Report Series* 11 4.
- Kim, Jeong Han. 2008. “Finding Cores of Random 2-SAT Formulae via Poisson Cloning.” *arXiv:0808.1599*.
- Kreweras, G. 1980. “Une Famille De Polynômes Ayant Plusieurs Propriétés Enumeratives.” *Periodica Mathematica Hungarica* 11:309–320.
- Kung, Joseph P. S., and Catherine Yan. 2003. “Gončarov Polynomials and Parking Functions.” *Journal of Combinatorial Theory. Series A* 102(1):16–37.
- Lipshitz, Leonard. 1988. “The Diagonal of a D-Finite Power Series Is D-Finite.” *Journal of algebra* 113(2):373–378.
- Liskovets, V. A. 1969. “A Recurrent Method for the Enumeration of Graphs with Labelled Vertices.” *Doklady Akademii Nauk SSSR* 184(6):1284–1287.
- Liskovets, V. A. 1970. “The Number of Strongly Connected Directed Graphs.” *Mathematical notes of the Academy of Sciences of the USSR* 8(6):877–882.
- Łuczak, Tomasz. 1990. “Component Behavior near the Critical Point of the Random Graph Process.” *Random Structures & Algorithms* 1(3):287–310.
- Łuczak, Tomasz. 1990. “The Phase Transition in the Evolution of Random Digraphs.” *Journal of Graph Theory* 14(2):217–223.
- Łuczak, Tomasz. 1992. “Sparse Random Graphs with a Given Degree Sequence.” In *Random Graphs, Volume 2*, 165–182. Chichester: Wiley.
- Łuczak, Tomasz, and Taral Guldahl Seierstad. 2009. “The Critical Behavior of Random Digraphs.” *Random Structures & Algorithms* 35(3):271–293.
- Lutz, Quentin, Élie de Panafieu, Alex Scott, and Maya Stein. 2021. “Active Clustering for Labeling Training Data.” *Advances in Neural Information Processing Systems (NeurIPS)* 34:8469–8480.
- Mahler, Kurt. 1940. “On a Special Functional Equation.” *Journal of the London Mathematical Society* 15:115–123.
- Mazumdar, Arya, and Barna Saha. 2017. “Clustering with Noisy Queries.” *Advances in Neural Information Processing Systems (NIPS)* 30.
- Mazumdar, Arya, and Barna Saha. 2017. “Query Complexity of Clustering with Side Information.” *Advances in Neural Information Processing Systems (NIPS)* 30.
- McKay, Brendan D, and Nicholas C Wormald. 1991. “Asymptotic Enumeration by Degree Sequence of Graphs with Degrees $o(n^{1/2})$.” *Combinatorica* 11(4):369–382.
- McKay, Brendan D. 1983. “Applications of a Technique for Labelled Enumeration.”
- Moak, Daniel S. 1984. “The Q-Analogue of Stirling's Formula.” *The Rocky Mountain Journal of Mathematics* :403–413.
- Monteil, Thierry, and Khaydar Nurligareev. 2021. “Asymptotics for Connected Graphs and Irreducible Tournaments.” In *Extended Abstracts EuroComb 2021: European Conference on Combinatorics, Graph Theory and Applications*, 823–828.

- Mousset, Frank, Andreas Noever, Konstantinos Panagiotou, and Wojciech Samotij. 2020. “On the Probability of Nonexistence in Binomial Subsets.” *The Annals of Probability* 48(1):493–525.
- Müller, Noela, Ralph Neininger, and Haodong Zhu. 2024. “Random 2-SAT: The Set of Atoms of the Limiting Empirical Marginal Distribution”. *arXiv:2410.17749*.
- Newman, Mark EJ. 2016. “Equivalence between Modularity Optimization and Maximum Likelihood Methods for Community Detection.” *Physical Review E* 94(5):52315.
- Newman, Mark EJ, and Michelle Girvan. 2004. “Finding and Evaluating Community Structure in Networks.” *Physical review E* 69(2):26113.
- Olver, Frank. 1997. *Asymptotics and Special Functions*. CRC Press.
- Panafieu, Élie de. 2014. *Analytic Combinatorics of Graphs, Hypergraphs and Inhomogeneous Graphs*. Doctoral dissertation, Université Paris-Diderot, Sorbonne Paris-Cité.
- Panafieu, Élie de. 2015. “Phase Transition of Random Non-Uniform Hypergraphs.” *Journal of Discrete Algorithms* 31(0):26–39.
- Panafieu, Élie de. 2015. “Enumeration and Structure of Inhomogeneous Graphs.” *Proceedings of the International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC), poster presentation* :12.
- Panafieu, Élie de. 2019. “Analytic Combinatorics of Connected Graphs.” *Random Structures & Algorithms* 55(2):427–495.
- Panafieu, Élie de. 2024. “Asymptotic Expansion of Regular and Connected Regular Graphs.” *arXiv:2408.12459*.
- Panafieu, Élie de, and Sergey Dovgal. 2020. “Counting Directed Acyclic and Elementary Digraphs.” *Proceedings of the 32nd Conference on Formal Power Series and Algebraic Combinatorics (FPSAC), Séminaire Lotharingien de Combinatoire* 84B:12.
- Panafieu, Élie de, and Vlady Ravelomanana. 2015. “Analytic Description of the Phase Transition of Inhomogeneous Multigraphs.” *European Journal of Combinatorics* 48:186–197.
- Panafieu, Élie de, and Michael Wallner. 2026. “Combinatorics of Nondeterministic Walks.” *Combinatorial Theory*.
- Panafieu, Élie de, François Durand, and Jérôme Lang. 2026. “The Communication Complexity of Instant-Runoff Voting.” In *Proceedings of the 35th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Panafieu, Élie de, Danièle Gardy, Bernhard Gittenberger, and Markus Kuba. 2014. “Probabilities of 2-Xor Functions.” In *Proceedings of the 11th Latin American Symposium on Theoretical Informatics (LATIN)*, 454–465. Berlin: Springer.
- Panafieu, Élie de, Danièle Gardy, Bernhard Gittenberger, and Markus Kuba. 2016. “2-Xor Revisited: Satisfiability and Probabilities of Functions.” *Algorithmica* 76(4):1035–1076.
- Panafieu, Élie de, Mohamed Lamine Lamali, and Michael Wallner. 2019. “Combinatorics of Nondeterministic Walks of the Dyck and Motzkin Type.” In *Proceedings of the 16th Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 1–12. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

- Peczarski, Marcin. 2004. “New Results in Minimum-Comparison Sorting.” *Algorithmica* 40(2):133–145.
- Pemantle, Robin, Mark C. Wilson, and Stephen Melczer. 2024. *Analytic Combinatorics in Several Variables*. Cambridge: Cambridge University Press.
- Pittel, Boris. 1990. “On Tree Census and the Giant Component in Sparse Random Graphs.” *Random Structures & Algorithms* 1(3):311–342.
- Pittel, Boris, and Nicholas C. Wormald. 2005. “Counting Connected Graphs inside-Out.” *Journal of Combinatorial Theory, Series B* 93(2):127–172.
- Poincaré, Henri. 1886. “Sur Les Intégrales Irrégulières Des équations Linéaires.” *Acta Math.* 8:295–344.
- Ralaivaosana, Dimbinaina, Vonjy Rasendrasahina, and Stephan Wagner. 2020. “On the Probability That a Random Digraph Is Acyclic.” In *Proceedings of the 31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA)*, 18. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik.
- Read, Ronald C. 1959. *Some Enumeration Problems in Graph Theory*. Doctoral dissertation, University of London.
- Read, Ronald C., and Nicholas C. Wormald. 1980. “The Number of Labelled 4-Regular Graphs.” *Journal of Graph Theory* 4(2):203–212.
- Reichardt, Jörg, and Stefan Bornholdt. 2006. “Statistical Mechanics of Community Detection.” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 74(1):16110.
- Robinson, Robert W. 1973. “Counting Labeled Acyclic Digraphs.”
- Shin, Heesung. 2008. “A New Bijection between Forests and Parking Functions.” *arXiv:0810.0427*.
- Sloane, N.J.A. 2017. “Sequence A294439.” *The on-line encyclopedia of integer sequences*.
- Stanley, Richard P. 1980. “Differentiably Finite Power Series.” *European journal of combinatorics* 1(2):175–188.
- Stark, Dudley, and Nick Wormald. 2018. “The Probability of Non-Existence of a Subgraph in a Moderately Sparse Random Graph.” *Combinatorics, Probability and Computing* 27(4):672–715.
- Stepanov, V. E. 1987. “On Some Features of the Structure of a Random Graph near a Critical Point.” *Theory of Probability and its Applications* 32(4):573–594.
- Tran, Linh V, Van H Vu, and Ke Wang. 2013. “Sparse Random Graphs: Eigenvalues and Eigenvectors.” *Random Structures & Algorithms* 42(1):110–134.
- Valiant, Leslie G. 1979. “The Complexity of Enumeration and Reliability Problems.” *siam Journal on Computing* 8(3):410–421.
- van der Hofstad, Remco. 2017a. “Counting Graphs and Null Models of Complex Networks: Configuration Model and Extensions.” In *Proceedings of the 43rd International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, 1–17. Cham: Springer.
- van der Hofstad, Remco. 2017b. *Random Graphs and Complex Networks*. Cambridge university press Cambridge.

- van der Hofstad, Remco, and Joel Spencer. 2006. "Counting Connected Graphs Asymptotically." *European Journal of Combinatorics* 27(8):1294–1320.
- van der Hofstad, Remco, Johan S. H. van Leeuwen, and Clara Stegehuis. 2018. "Triadic Closure in Configuration Models with Unbounded Degree Fluctuations." *Journal of Statistical Physics* 173(3–4):746–774.
- van der Hofstad, Remco, Johan S. H. van Leeuwen, and Clara Stegehuis. 2021. "Optimal Subgraph Structures in Scale-Free Configuration Models." *The Annals of Applied Probability* 31(2):501–537.
- Vesdapunt, Norases, Kedar Bellare, and Nilesh Dalvi. 2014. "Crowdsourcing Algorithms for Entity Resolution." *Proceedings of the VLDB Endowment* 7(12):1071–1082.
- Wagstaff, Kiri, Claire Cardie, Seth Rogers, and Stefan Schrödl. 2001. "Constrained K-Means Clustering with Background Knowledge." In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, 577–584.
- Wang, Jiannan, Guoliang Li, Tim Kraska, Michael J Franklin, and Jianhua Feng. 2013. "Leveraging Transitive Relations for Crowdsourced Joins." In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 229–240.
- Wigner, Eugene P. 1958. "On the Distribution of the Roots of Certain Symmetric Matrices." *Annals of Mathematics* 67(2):325–327.
- Wong, R, and H Li. 1992. "Asymptotic Expansions for Second-Order Linear Difference Equations." *Journal of computational and applied mathematics* 41(1–2):65–94.
- Wormald, Nicholas. 2018. "Asymptotic Enumeration of Graphs with Given Degree Sequence." In *Proceedings of the International Congress of Mathematicians, Rio de Janeiro 2018 (ICM)*, 3245–3264.
- Wormald, Nicholas C. 1981. "The Asymptotic Connectivity of Labelled Regular Graphs." *Journal of Combinatorial Theory, Series B* 31(2):156–167.
- Wormald, Nicholas C. 1979. "Enumeration of Labelled Graphs II: Cubic Graphs with a Given Connectivity." *Journal of the London Mathematical Society* s2-20(1):1–7.
- Wormald, Nick. 1999. "Models of random regular graphs." *Surveys in Combinatorics* :239–298.
- Wright, E. M. 1971. "The Number of Strong Digraphs." *Bulletin of the London Mathematical Society* 3:348–350.
- Wright, E. M. 1977. "The Number of Connected Sparsely-Edged Graphs." *Journal of Graph Theory* 1:317–330.
- Wright, E. M. 1978. "The Number of Connected Sparsely-Edged Graphs. II: Smooth Graphs and Blocks." *Journal of Graph Theory* 2:299–305.
- Wright, E. M. 1980. "The Number of Connected Sparsely-Edged Graphs. III: Asymptotic Results." *Journal of Graph Theory* 4:393–407.
- Wright, EM. 1970. "Asymptotic Relations between Enumerative Functions in Graph Theory." *Proceedings of the London Mathematical Society* 3(3):558–572.
- Zakharevich, Inna. 2006. "A Generalization of Wigner's Law." *Communications in Mathematical Physics* 268:403–414.

- Zeilberger, Doron. 1990. "A Holonomic Systems Approach to Special Functions Identities." *Journal of computational and applied mathematics* 32(3):321–368.
- Zeilberger, Doron. 2008. "Asyrec: A Maple Package for Computing the Asymptotics of Solutions of Linear Recurrence Equations with Polynomial Coefficients." *The Personal Journal of Shalsho B. Ekhad and Doron Zeilberger*.