Analytic combinatorics of connected graphs

Élie de Panafieu
Bell Labs France, Nokia

Seminar Flajolet of May 2017

Part of this work has been published at FPSAC16.

Model: connected (simple) graphs

Drawing with $n = 5$, $m = 4$, $w^m \frac{z^n}{n!}$

Excess $k = m - n$: always ≥ -1.

Goal: asymptotic expansion of the number of connected (n, m)-graphs when $m \approx (1 + \alpha)n$

$$\mathrm{CSG}_{n,m} = D_{n,m} \left(\sum_{r=0}^{d-1} c_r n^{-r} + \mathcal{O}(n^{-d}) \right).$$

Related work:

	k fixed	$k \to +\infty$
asymptotics	Wright 1980	Bender Canfield McKay 1995 Pittel Wormald 2005 van der Hofstad Spencer 2005
asymptotic expansion	Flajolet Salvy Schaeffer 2004	present work

1 From connected graphs to graphs with degree constraints

gf of graphs

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

A graph is a set of connected graphs

$$\mathrm{SG}(z, w) = e^{\mathrm{CSG}(z,w)},$$

so we obtain the following exact formula

$$\mathrm{CSG}(z, w) = \log \left(1 + \sum_{n \geq 1} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!} \right). \tag{1}$$

Problem: divergent series. One of the few tools we have is Bender's Theorem (1975). Let first look at the easiest problem of enumerating without consideration for the number of edges.

Bender's Theorem (simplified): Consider a convergent series $F(z)$ and a divergent series $G(z) = \sum_{k\geq 1} g_k z^k$. If (g_k) grows "*fast enought*" to infinity (*e.g.* like a factorial), then

$$[z^n]F(G(z)) = \sum_{r=0}^{d-1} g_{n-r}[y^r]F'(G(y)) + \mathcal{O}(g_{n-d}).$$

Intuition: If $F(z)$, $G(z)$ are the gf of the families F, G, then what do the objects in $F \circ G$ look like?

Two drawings: balanced or unbalanced.

Application: If we forget the number of edges, Equation (1) becomes

$$\mathrm{CSG}(z) = \log\left(1 + \sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right).$$

Bender's Theorem is applicable, and gives

$$\mathrm{CSG}_n = 2^{\binom{n}{2}}(1 - 2n2^{-n} + o(2^{-n})).$$

Thus almost all graphs with n vertices are connected.

Generalization: Flajolet, Salvy, Schaeffer 2004 analyzed Equation (1) around $w = -1$ and extracted the asymptotic expansion of connected graphs with fixed excess. It is a technical article. The reason is that a typical (n, m)-graph with $m = \Theta(n)$ is not connected. The main responsible are trees. Thus, Bender's Theorem cannot be applied. Said otherwise: many "magical" cancelations occur in Equation (1). They are necessary to cancel the trees.

Solution: let us define the *positive graphs* as the graphs where all components have positive excess, *i.e.* are neither trees nor unicycles. Then the gf of connected graphs of excess $k > 0$ is

$$\mathrm{CSG}_k(z) = [y^k] \log\left(1 + \sum_{\ell\geq 1} \mathrm{SG}_\ell^{>0}(z) y^\ell\right),$$

and a variant of Bender's Theorem is applicable, provided that the gf of $\mathrm{SG}_\ell^{>0}(z)$ is known. Indeed, according to Erdős and Rényi 1960, almost all positive (n, m)-graphs are connected when $m = \Theta(n)$. This property was used by Pittel and Wormald 2005, we obtained the asymptotic of $\mathrm{CSG}_{n,n+k}$ when $k \to +\infty$. The easiest way to remove the trees is to forbid the degrees 0 and 1 (idea already applied by Wright 1980, and Pittel Wormald 2005).

Positive graphs and cores: A *core* is a graph with minimum degree ≥ 2. Any graph can be reduced to its core by removing repetedly the vertices of degree 0 and 1 (drawing). A core is a positive core with an additional set of isolated cycles, so

$$\mathrm{Core}_k(z) = \mathrm{Core}_k^{>0}(z) e^{\frac{1}{2}\log(\frac{1}{1-z}) - \frac{z}{2} - \frac{z^2}{4}}.$$

A positive graph is a positive core where a rooted tree is attached to each vertex

$$\mathrm{SG}_k^{>0}(z) = \mathrm{Core}_k^{>0}(T(z)) = \sqrt{1 - T(z)} e^{\frac{z}{2} + \frac{z^2}{4}} \mathrm{Core}_k(T(z)). \tag{2}$$

Observe that we would not have the factor $e^{\frac{z}{2} + \frac{z^2}{4}}$ if loops and multiple edges were allowed.

Positive graphs and kernels: a *kernel* is a multigraph with minimum degree ≥ 3. Any graph can be reduced to its kernel by first reducing to its core, then merging the edges sharing a vertex of degree 2 (drawing). If we allow loops and multiple edges, this construction can be reversed, going from kernels to positive multigraphs.

$$\mathrm{MG}_k^{>0}(z) = \frac{\mathrm{Kernel}_k\left(\frac{T(z)}{1-T(z)}\right)}{(1-T(z))^k}. \tag{3}$$

A similar formula exists for positive simple graphs, by keeping track of the loops and multiple edges in the kernel. Furtermore, $\mathrm{Kernel}_k(z)$ is a polynomial of degree $2k$. Indeed, the sum of the degree is twice the number of edges, so

$$2m \geq 3n, \qquad n \leq 2k, \qquad m \leq 3k.$$

When to use what? Equation (2) turns out to be better suited when $k \to +\infty$, while Equation (3) is useful for fixed k.

2 Connected multigraphs

Degree constraints are easier to handle on multigraphs than simple graphs, because loops and multiple edges appear naturally.

Configuration model: introduced by Bollobás 1980 and Wormald 1978.

Drawing explaining the configuration model and the half-edges principle.
Loops and multiple edges are hard to track.

This motivates us to work first on multigraphs, following EdP Lander Analco 2016.

Multigraphs: Loops and multiple edges allowed, labelled oriented edges

Drawing of a multigraph, $n = 2$, $m = 3$, $\frac{w^m}{2^m m!} \frac{z^n}{n!}$.
Representation as a set of labelled half-edges.

$$\begin{aligned}
\mathrm{MCore}(z,w) &= \sum_{m \geq 0} (2m)! [x^{2m}] e^{z(e^x - 1 - x)} \frac{w^m}{2^m m!} \\
&= \sum_{n,m \geq 0} (2m-1)!! [x^{2m}] (e^x - 1 - x)^n \frac{z^n}{n!} w^m \\
&= \sum_{n,k \geq 0} (2(n+k)-1)!! [x^{2k}] \left(\frac{e^x - 1 - x}{x^2}\right)^n \frac{z^n}{n!} w^{n+k} \\
&= \sum_{k \geq 0} (2k-1)!! [x^{2k}] \frac{1}{\left(1 - zw \frac{e^x - 1 - x}{x^2/2}\right)^{k+1/2}} w^k.
\end{aligned}$$

Thus, the gf of multicores of excess k is

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

Thus, the number of positive multigraphs is

$$\mathrm{MG}_k^{>0}(z) = (2k-1)!! [x^{2k}] \sqrt{1 - T(z)}\, B(z,x)^{k+1/2},$$

where $B(z,x) = \left(1 - T(z)\frac{e^x-1-x}{x^2/2}\right)^{-1}$. Its asymptotic, equal to the asymptotic number of connected multigraphs, is obtained applying a classic bivariate saddle-point method (large powers Theorem, Pemantle Wilson 2013)

$$\begin{aligned}\mathrm{CMG}_{n,n+k} &= n!2^{n+k}(n+k)![z^n]\,\mathrm{CMG}_k(z)\\ &\sim n!2^{n+k}(n+k)!(2k-1)!!\frac{\sqrt{1-T(\zeta)}}{2\pi k\sqrt{\det(H)}}\frac{B(\zeta,\lambda)^{k+1/2}}{\zeta^n\lambda^{2k}},\end{aligned}$$

and an asymptotic expansion is computable using a variant of Bender's Theorem

$$\begin{aligned}\mathrm{CMG}_{n,n+k} &= n!2^{n+k}(n+k)!\left(\sum_{r=0}^{d-1}[z^n]\,\mathrm{MG}^{>0}_{k-r}(z)[y^r]\left(1+\sum_{\ell\geq 1}\mathrm{MG}^{>0}_{\ell}(z)y^\ell\right)^{-1} + \mathcal{O}\left([z^n]\,\mathrm{MG}^{>0}_{k-d}(z)\right)\right)\\ &= n!2^{n+k}(n+k)!(2k-1)!!\left(\sum_{r=0}^{d-1}\frac{(2(k-r)-1)!!}{(2k-1)!!}[z^n x^{2k}]A_r(z,x)B(z,x)^k + \mathcal{O}\left(n^{-d}\frac{B(\zeta,\lambda)}{k\zeta^n\lambda^{2k}}\right)\right).\end{aligned}$$

We obtain the asymptotic expansion when $k = \alpha n + \mathcal{O}(n^{-d})$ if we can expand $A_r(z,x)$ arount (ζ,λ).

Kernel: multigraph with minimum degree at least 3. The gf of kernels of excess k

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

is a polynomial of degree $2k$. This is not a surprise, since the sum of the degrees is twice the number of edges

$$2m \geq 3n, \quad n \leq 2k, \quad m \leq 3k.$$

Drawing of the construction of a kernel from a graph,
and going back to a multigraph.

Since

$$\mathrm{MG}^{>0}_k(z) = \frac{\mathrm{Kernel}_k\left(\frac{T(z)}{1-T(z)}\right)}{(1-T(z))^k},$$

there is a polynomial $Q_k(z)$ of degree $2k$ such that

$$\mathrm{MG}^{>0}_k(z) = \frac{Q_k(T(z))}{(1-T(z))^{3k}}.$$

Its expansion near ζ is computable, thus the expansion of $A_r(z,x)$ near (ζ,λ) is computable.

The same can be done with simple graphs, but care is needed to ensure the loops and multiple edges of the kernels disappear when transforming it into a positive graph. Wright obtained a differential recurrence on the gfs of connected kernels of excess k where loops and multiple edges are marked. This was enought to compute the asymptotics of connected graphs with fixed excess. Bender Canfield McKay 1995 used this recurrence for the asymptotics of connected graphs with large excess. Their article is (very) technical.

3 Connected simple graphs

From multigraphs to simple graphs: Let $\mathrm{MG}^\star$ denote the multigraphs without loops and double edges (hence with no multiple edges either), then

$$\mathrm{CSG}_{n,n+k} = 2^{n+k}(n+k)!\,\mathrm{CMG}^\star_{n,n+k}$$

because each edge has two possible orientations and there are $m = n+k$ labels to be distributed. To remove the loops and double edges, we apply the *inclusion-exclusion principle*. Cite Collet, EdP, Gardy, Gittenberger, Ravelomanana Eurocomb 2017.

Patchwork: set of loops and double edges, that might share vertices and labels, and glued together form a multigraph

Drawing of a patchwork.
Notice that distinct patchwork can lead to the same multigraph.

$$\mathrm{Patch}_k(z,u) = e^{u\frac{z}{2}+u^2\frac{z^2}{4}}\,\mathrm{Patch}_k^{>0}(z,u)$$

where $\mathrm{Patch}_k^{>0}(z,u)$ is a polynomial for which we have a formula (given if we have time).

Simple cores: Let $\mathrm{MCore}(z,w,u)$ denote the gf of multicores where u mark the loops and double edges. Then

$$\mathrm{Core}(z,w) = \mathrm{MCore}(z,w,0).$$

The gf of multicores where each loop and double edge is either marked, or left unmarked, is $\mathrm{MCore}(z,w,u+1)$.

Drawing showing how to build a multicore from this family
using the half-edges idea.

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

Applying the same technique as before, we obtain

$$\mathrm{MCore}_k(z,u+1) = \sum_{\ell=0}^{k}(2(k-\ell)-1)!![x^{2(k-\ell)}]\frac{\mathrm{Patch}_\ell(ze^x,u)}{\left(1-z\frac{e^x-1-x}{x^2/2}\right)^{k-\ell+1/2}},$$

so

$$\mathrm{Core}_k(z) = \mathrm{MCore}_k(z,0) = \sum_{\ell=0}^{k}(2(k-\ell)-1)!![x^{2(k-\ell)}]\frac{\mathrm{Patch}_\ell(ze^x,-1)}{\left(1-z\frac{e^x-1-x}{x^2/2}\right)^{k-\ell+1/2}}.$$

Positive graphs:

$$\begin{aligned}\mathrm{SG}_k^{>0}(z) &= \sqrt{1-T(z)}e^{\frac{T(z)}{2}+\frac{T(z)^2}{4}}\,\mathrm{Core}_k(T(z))\\ &= \sum_{\ell=0}^{k}(2(k-\ell)-1)!![x^{2(k-\ell)}]\sqrt{1-T(z)}e^{\frac{T(z)}{2}+\frac{T(z)^2}{4}}\,\mathrm{Patch}_\ell(T(z)e^x,-1)B(z,x)^{k-\ell+1/2}.\end{aligned}$$

Connected graphs: Again, by application of Bender's Theorem variant, we obtain

$$\begin{aligned}\mathrm{CSG}_{n,n+k} &= n!2^{n+k}(n+k)![z^n]\,\mathrm{CSG}_k(z)\\ &= n!2^{n+k}(n+k)![z^n y^k]\log\left(1+\sum_{\ell\geq 1}\mathrm{SG}_\ell^{>0}(z)y^\ell\right)\\ &= n!2^{n+k}(n+k)!\left(\sum_{r=0}^{d-1}[z^n]\,\mathrm{SG}_{k-r}^{>0}(z)[y^r]\left(1+\sum_{\ell\geq 1}\mathrm{SG}_\ell^{>0}(z)y^\ell\right)^{-1}+\mathcal{O}\left(\frac{B(\zeta,\lambda)}{k\zeta^n\lambda^{2k}}\right)\right)\end{aligned}$$

when $k=\alpha n+\mathcal{O}(n^{-d})$.

4 Conclusion

Some related work:

- Wright 1980: asymptotic of connected graphs with fixed excess
- Bender Canfield McKay 1995: $k \to +\infty$ (differential recurrence on the gf of connected kernels)
- Flajolet Savly Schaeffer 2004: asymptotic expansion, fixed excess (Airy connection)
- Pittel Wormald 2005: simpler proof for the asymptotic when $k \to +\infty$ (cores)
- Spencer, van der Hofstad 2005: asymptotic when $k \to +\infty$ (random walks)
- present work: asymptotic **expansion** when $k \to +\infty$.

But more important than the prevision: new techniques

- multigraphs instead of simple graphs, improving the model of Flajolet, Janson, Knuth, Łuczak, Pittel,
- graphs with degree constraints (with Ramos),
- graphs with marked subgraphs (with Collet, Gardy, Gittenberger, Ravelomanana).

Future work:

- structure of random graphs containing a giant component,
- hypergraphs (constraints on the degrees and sizes of the hyperedges, connected hypergraph ...)
- inhomogeneous graphs (stochastic block model).

5 Appendix: patchworks

The generating function of patchworks is equal to

$$P(z, w, u) = \mathrm{SG}\left(ze^{uw/2}, \mathrm{SG}(w, u)e^{-w} - 1\right)e^{-z}.$$

Let $\mathcal{D}$ denote the family of patchworks on two vertices $\{1, 2\}$ that contains only double edges (*i.e.* no loop). Two parts of such a patchwork P may share at most one edge (since all parts are distinct). We now describe a bijection between $\mathcal{D}$ and the non-empty graphs without isolated vertices. Let P denote a patchwork from $\mathcal{D}$, and G the corresponding graph:

- each edge of $\mathrm{MG}(P)$ is represented by a vertex of G,
- each part of P is a double edge, and corresponds to an edge of G.

There are no loops in G because each double edge of P contains two distinct edges. There are no multiple edges in G because the parts of P are distinct. No vertex of G can be isolated since all edges of P belong to at least one part. The generating function of non-empty graphs without isolated vertices is

$$\mathrm{SG}(z, w)e^{-z} - 1,$$

so the generating function of $\mathcal{D}$ is (without taking into account the two vertices)

$$\mathrm{SG}(w, u)e^{-w} - 1.$$

Any patchwork where a set of isolated vertices has been added can be uniquely described as a graph G, where each edge is replaced with a patchwork from $\mathcal{D}$, and a set of loops is added to each vertex. Therefore, the generating function of patchworks satisfies

$$P(z, w, u)e^{z} = \mathrm{SG}\left(ze^{uw/2}, \mathrm{SG}(w, u)e^{-w} - 1\right).$$