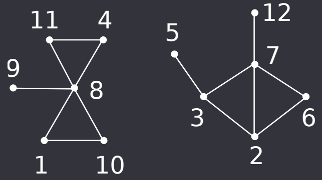
# Analytic combinatorics of connected graphs

Élie de Panafieu
Bell Labs France, Nokia

Séminaire Flajolet, juin 2017

# Introduction




Graph with $n = 12$ vertices, $m = 14$ edges, excess $k = m - n = 2$.

Generating function (GF): $w^m \frac{z^n}{n!}$

**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 95<br>Pittel Wormald 05<br>van der Hofstad Spencer 05 |
| asympt. expansion | Flajolet Salvy Schaeffer 04 | present work |

# I - From connected graphs to 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

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

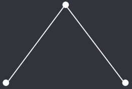
**Problem:** divergent series. One of the few tools we have is Bender's Theorem (1975).

# Bender's Theorem (simplified version)

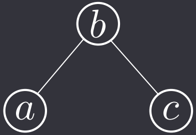
Convergent series $F(z)$ and divergent series $G(z) = \sum_{k \geq 1} g_k z^k$.
If $g_k \to +\infty$ "*fast enought*" (*e.g.* factorial), then

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

**Intuition:** If $F(z)$, $G(z)$ are the gf of the families $F$, $G$, then the objects in $F \circ G$ are typically unbalanced, with one large object from $G$ and the others very small



$\in F$ | $(a, b, c) \in G$ |




$\in FoG$

# Classic application

GF of connected graphs, without considering the number of edges

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

The hypothesis of Bender's Theorem are satisfied

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

Thus almost all graphs with $n$ vertices are connected.

# Generalization?

Flajolet, Salvy, Schaeffer 2004 analyzed around $w = -1$

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

asymptotic expansion of connected graphs with fixed excess.

Typical $(n, m)$-graph with $m = \Theta(n)$ are not connected (Erdős Rényi 1960): they contain trees and unicycles.

Thus, Bender's Theorem cannot be applied. Many "magical" cancelations occur in Equation (1).

# Solution

**Positive graphs:** $\mathrm{SG}^{>0}$ graphs where all components have positive excess, *i.e.* no trees, no unicycles.
The gf of connected graphs of excess $k > 0$ is

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

A variant of Bender's Theorem is applicable, if $\mathrm{SG}_\ell^{>0}(z)$ is known.

$$n![z^n]\,\mathrm{CSG}_k(z) = n! \sum_{r=0}^{d-1} [z^n]\,\mathrm{SG}_{k-r}^{>0}(z)[y^r] \left( 1 + \sum_{\ell \geq 1} \mathrm{SG}_\ell^{>0}(z) y^\ell \right)^{-1} + \mathcal{O}(.).$$

**Erdős and Rényi 1960:** almost all positive $(n, m)$-graphs are connected when $m = \Theta(n)$.
Property used by Pittel and Wormald 2005.
Simplest way to remove the trees: forbid the degrees 0 and 1 (applied by Wright 1980, and Pittel Wormald 2005).

# Positive graphs and Cores

**Core:** graph with min deg $\geq 2$.
Graph $\rightarrow$ Core: remove repeatedly the vertices of deg 0 and 1.

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

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

Factor $e^{\frac{z}{2} + \frac{z^2}{4}}$ to avoid loops and double edges.

# Positive graphs and kernels

**Kernel:** multigraph with min deg $\geq 3$.
Core $\to$ kernel: merge the edges sharing a vertex of deg 2.

If we allow loops and multiple edges, this construction can be reversed, going from kernels to positive multigraphs.

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

A similar formula exists for positive simple graphs, by keeping track of the loops and multiple edges in the kernel.

$\mathrm{Kernel}_k(z)$ is a polynomial of deg $2k$

$$2m = \sum_{v \in G} \deg(v) \geq 3n, \qquad n \leq 2k, \qquad m \leq 3k.$$

# Which formula is best?

Asymptotics for fixed $k$

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

Asymptotics for large $k$

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

Both used in the asymptotic analysis of

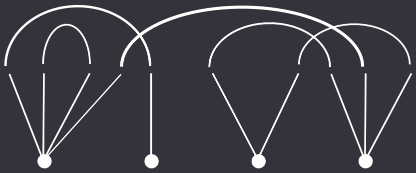
$$n![z^n]\,\mathrm{CSG}_k(z) = n!\sum_{r=0}^{d-1}[z^n]\,\mathrm{SG}_{k-r}^{>0}(z)[y^r]\left(1+\sum_{\ell\geq 1}\mathrm{SG}_\ell^{>0}(z)y^\ell\right)^{-1}+\mathcal{O}(.).$$

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

# II - Multigraphs

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

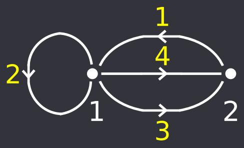
**Configuration model:** Bollobás 1980, Wormald 1978.

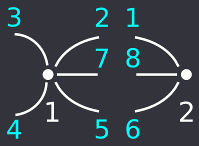


This motivates us to work first on multigraphs.
EdP Lander Analco16.

# Multicores

**Multigraphs:** loops and multiple edges allowed, labelled oriented edges. Replacing edges with half-edges, a multicore becomes a set of sets of size $\geq 2$




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

Change of variable $m \to k + n$, closed form of the sum over $n$

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

# Multicores

**Multigraphs:** loops and multiple edges allowed, labelled oriented edges. Replacing edges with half-edges, a multicore becomes a set of sets of size $\geq 2$


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

Change of variable $m \to k+n$, closed form of the sum over $n$

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

# Multicores

**Multigraphs:** loops and multiple edges allowed, labelled oriented edges. Replacing edges with half-edges, a multicore becomes a set of sets of size $\geq 2$


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

Change of variable $m \to k + n$, closed form of the sum over $n$

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

# Connected multigraphs

The gf of positive multigraphs of excess $k$ is

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

where $B(z,x) = \left(1 - T(z)\frac{e^x-1-x}{x^2/2}\right)^{-1}$.

Bender's Theorem when $k = \alpha n + \mathcal{O}(n^{-d})$

$$\mathrm{CMG}_{n,n+k} \sim n!2^{n+k}(n+k)! \sum_{r=0}^{d-1} (2(k-r)-1)!![z^n x^{2k}] A_r(z,x) B(z,x)^k$$

where

$$A_r(z,x) = \sqrt{(1-T(z))B(z,x)}[y^r]\left(1 + \sum_{\ell \geq 1} \mathrm{MG}_\ell^{>0}(z) y^\ell\right)^{-1}$$

Saddle-point method (Pemantle Wilson 2013) to conclude.

# III- Connected simple graphs

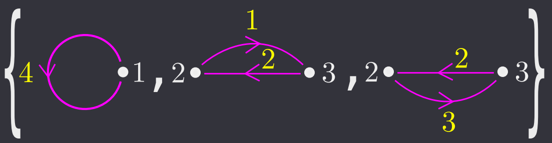
**From multigraphs to simple graphs:** $\mathrm{MG}^{\star}$ denote the multigraphs without loops and double edges

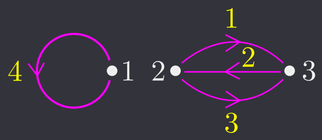
$$\mathrm{CSG}_{n,m} = 2^m m! \, \mathrm{CMG}^{\star}_{n,m}$$

**inclusion-exclusion principle** to remove the loops and double edges (Collet, EdP, Gardy, Gittenberger, Ravelomanana Eurocomb17).

# Patchworks

Patchwork: set of loops and double edges (not necessarily disjoint)




$$P(z,w,u) = \sum_{\text{patchwork } P} u^{\text{nb loops \& double edges}} \frac{w^{m(P)}}{2^{m(P)} m(P)!} \frac{z^{n(P)}}{n(P)!}$$

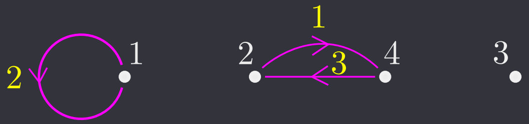
A patchwork of excess $k$ is a set of isolated loops and double edges, and a finite nb of more complex patterns

$$P_k(z,u) := [y^k] P(z/y, y, u) = e^{uz/2 + u^2 z^2/4} P_k^{>0}(z)$$

# Simple Cores

MCore$(z, w, u)$ is the gf of multicores where $u$ marks the loops and double edges, so MCore$(z, w, 0) =$ Core$(z, w)$.

Inclusion-exclusion: compute MCore$(z, w, u + 1)$, gf of multicores where each loop and double edge is either marked or left unmarked

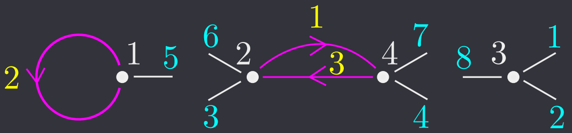



$$P(z \quad , w, \; u \;)e^{z} \qquad .$$

# Simple Cores

MCore$(z, w, u)$ is the gf of multicores where $u$ marks the loops and double edges, so MCore$(z, w, 0) =$ Core$(z, w)$.

Inclusion-exclusion: compute MCore$(z, w, u+1)$, gf of multicores where each loop and double edge is either marked or left unmarked

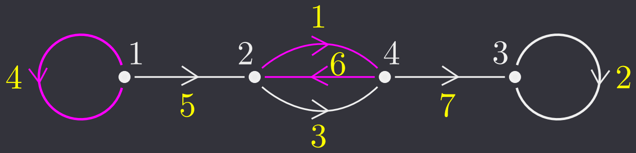



$$P(ze^x, w, \quad u \quad ) e^{z(e^x - 1 - x)} \qquad\qquad .$$

# Simple Cores

MCore$(z, w, u)$ is the gf of multicores where $u$ marks the loops and double edges, so MCore$(z, w, 0) = $ Core$(z, w)$.

Inclusion-exclusion: compute MCore$(z, w, u+1)$, gf of multicores where each loop and double edge is either marked or left unmarked

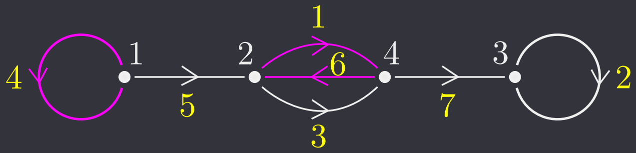



$$(2m)![x^{2m}]P(ze^x, w, \quad u \quad)e^{z(e^x-1-x)}\frac{w^m}{2^m m!} \qquad .$$

# Simple Cores

$\mathrm{MCore}(z, w, u)$ is the gf of multicores where $u$ marks the loops and double edges, so $\mathrm{MCore}(z, w, 0) = \mathrm{Core}(z, w)$.

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

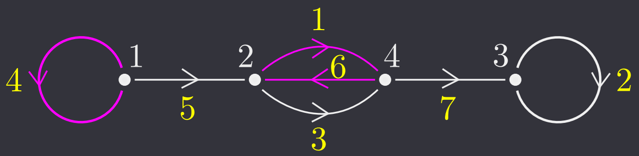



$$\mathrm{MCore}(z, w, u+1) = \sum_{m \geq 0} (2m)! [x^{2m}] P(ze^x, w, \ u \ ) e^{z(e^x - 1 - x)} \frac{w^m}{2^m m!} \quad .$$

# Simple Cores

MCore($z, w, u$) is the gf of multicores where $u$ marks the loops and double edges, so MCore($z, w, 0$) $=$ Core($z, w$).

Inclusion-exclusion: compute MCore($z, w, u+1$), gf of multicores where each loop and double edge is either marked or left unmarked

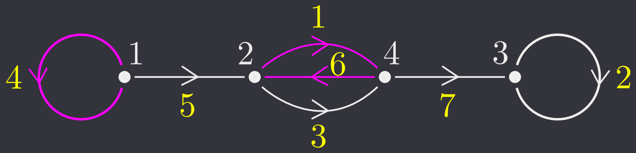



$$\mathrm{Core}(z,w) = \sum_{m\geq 0} (2m)! [x^{2m}] P(ze^x, w, -1) e^{z(e^x - 1 - x)} \frac{w^m}{2^m m!} \quad .$$

# Simple Cores

MCore$(z, w, u)$ is the gf of multicores where $u$ marks the loops and double edges, so MCore$(z, w, 0) =$ Core$(z, w)$.

Inclusion-exclusion: compute MCore$(z, w, u + 1)$, gf of multicores where each loop and double edge is either marked or left unmarked




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

# 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 $k \to +\infty$.

# Conclusion

**But more important** than the precision: 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).

# Conclusion

**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).

**Thank you!**

# Bonus: variant of Bender's Theorem

If $F(z)$ has a positive radius of convergence, and

$$G(z,y) = \sum_{\ell \geq 1} G_\ell(z) y^\ell$$

with

$$G_\ell(\zeta) \leq C E^\ell \Gamma(\ell + \beta),$$

then

$$[z^n y^k] F(G(z,y)) = [z^n] \sum_{r=0}^{d-1} G_{k-r}(z) [y^r] F'(G(z,y)) + \mathcal{O}\left( \frac{E^k}{\zeta^n} \Gamma(k - d + \beta$$