Analytic Combinatorics of Random Graphs
A direct combinatorial approach for deriving limit laws

APC 527 — Random Graphs and Networks
Prof. Emmanuel Abbe, EunJee Lee
Final Project Report

Maryam Bahrani
Princeton University, Fall 2016
# Contents

1 Introduction 3  
1.1 Motivation 3  
1.2 Prior work 3  
1.3 Acknowledgements 3  

2 Model and notation 3  

3 Connected multigraphs with fixed excess 5  
3.1 Previous work 5  
3.2 Definitions and bijections 5  
3.3 Generating function for connected kernels 6  
3.4 Asymptotic analysis 7  
3.5 Counting cubic multigraphs 7  

4 Limit laws 7  
4.1 Limit laws through generating functions 8  
4.2 Multigraphs with marked degrees 8  
4.3 Limit law of degrees 9  
4.4 Observing sharp threshold for disappearance of isolated vertices 9  
4.5 Back to limit law of degrees 10  

5 Next steps and concluding remarks 10  
5.1 Extension to multiple communities 10  
5.2 From Multigraphs to Simple Graphs 11  
5.3 A note on equivalence of models 11
1 Introduction

1.1 Motivation

My goal in this project was to explore structural properties of random graphs through the lens of generating functions. This was motivated by the following reasons:

- Combinatorial approaches are more direct than classical probabilistic arguments and can improve our understanding of the objects under study.
- Generating functions capture the exact behavior of a combinatorial class. Given a generating function, powerful theorems from complex analysis allow for the derivation of asymptotics, limit laws, etc. They can also be used to build efficient uniform samplers from different random graph models.
- These derivations are nearly automatic, generalizable to different models of random graphs, and receptive of non-trivial constraints. While the same results for simpler models such as Erdős-Rényi might be easier to derive using probabilistic approaches, other classes of graphs might be better studied via a combinatorial perspective (e.g. see [1] for a combinatorial treatment of the configuration model).

1.2 Prior work

The literature on the analytic combinatorics of random graphs is relatively sparse. Many graph questions, including general limit laws for the degrees, at least in the stochastic block model, have not been addressed by analytic combinatorics at all.

1.3 Acknowledgements

I have benefited extensively from the guidance of Élie de Panafieu, who suggested the problem and introduced the methodology and many of the generating functions, as well as Jérémie Lumbroso, who introduced me to Élie and answered my questions along the way.

2 Model and notation

In this note, we will work with an extension of simple graphs to have oriented and allow for loops and multiedges.

Definition 1. A multigraph $G$ is a pair $(V(G), E(G))$, where $V(G)$ denotes the set of labeled vertices and $E(G)$ the set of labeled edges. Each edge is an ordered triplet $(u, v, e)$, where $u$ and $v$ are vertex labels and $e$ is the edge label. Also, define $m(G) = |E(G)|$ and $n(G) = |V(G)|$.

Given a multigraph, we can obtain a simple graph by removing loops and multiedges and ignoring edge orientations. Conversely, any simple graph with $m$ edges can be made into a multigraph in $2^m m!$ ways ($2^m$ ways of assigning edge orientations, and $m!$ ways of picking edge labels).

We work with multigraphs in this note because they have simpler generating functions. We can derive a stronger result for multigraphs for a given number of loops and multiedges, which can in turn be extended to simple graphs by constraining these numbers to be zero, as done in [1].

¹Labels are distinct consecutive integers starting at 1.
Using the symbolic method (developed by Flajolet and Sedgewick [6]), we can translate combinatorial operations (such as set and sequence) into generating function equations, following Table 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>EGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral (element of size 0)</td>
<td>{\epsilon}</td>
<td>1</td>
</tr>
<tr>
<td>Atom (element of size 1)</td>
<td>{Z}</td>
<td>z</td>
</tr>
<tr>
<td>Disjoint union</td>
<td>A + B</td>
<td>A(z) + B(z)</td>
</tr>
<tr>
<td>(Labeled) product</td>
<td>A ⋆ B</td>
<td>A(z) ⋅ B(z)</td>
</tr>
<tr>
<td>Sequence</td>
<td>SEQ (A)</td>
<td>\frac{1}{1-A(z)}</td>
</tr>
<tr>
<td>Set</td>
<td>SET (A)</td>
<td>e^{A(z)}</td>
</tr>
<tr>
<td>Substitution</td>
<td>A ⋄ B</td>
<td>A(B(z))</td>
</tr>
</tbody>
</table>

Table 1: Relevant labeled combinatorial classes and their exponential generating functions (EGFs), based on Lumbroso and Morcrette [7].

Using \( z \) to mark vertices and \( w \) to mark edges, the bivariate generating function for a family \( \mathcal{F} \) of multigraphs is defined as

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

Note that this implies the generating function for a vertex is \( z \) and the generating function for an oriented edge is \( w/2 \).

To derive the generating function for all multigraphs, we observe that multigraph \( G \) with \( n \) vertices and \( m \) edges can be thought of as either

- a set of \( m \) labeled oriented edges, each chosen from \( n^2 \) possibilities, or
- a sequence of \( 2m \) vertices (of all \( n^{2m} \) such sequences), interpreted as \( m \) pairs of vertices, where the order within a pair determines the orientation of the corresponding edge, and the order of the pairs follows the edge labels.

The first interpretation gives the following EGF for multigraphs

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

while the second interpretation gives

\[
MG(z, w) = \sum_{n \geq 0} \sum_{m \geq 0} \frac{w^m n^{2m} z^n}{2^m m! n!}.
\]

As expected, these two generating functions are identical:

\[
\sum_{n \geq 0} \sum_{m \geq 0} \frac{w^m n^{2m} z^n}{2^m m! n!} = \sum_{n \geq 0} \frac{z^n}{n!} \sum_{m \geq 0} \frac{(n^2 w/2)^m}{m!} = \sum_{n \geq 0} e^{(n^2 w/2) z^n} \frac{z^n}{n!}.
\]
3 Connected multigraphs with fixed excess

As a first step in working with generating functions of multigraphs, we will derive from first principles the generating function and asymptotic growth of the number of connected multigraphs with of small fixed excess, i.e. the difference between the number of edges and the number of vertices.

3.1 Previous work

In the original paper by Erdős-Rényi (1959) [4], the authors start out deriving the threshold for connectivity directly using generating functions. However, since the asymptotics of connected graphs were unknown at the time, they proceed to use probabilistic arguments. Wright (1980) [8] first derived the asymptotics of simple graphs with fixed excess, and Flajolet et al. (2004) [5] derived the complete asymptotic expansion. The results in this section are due to de Panafieu (2016) [2].

3.2 Definitions and bijections

The excess of a multigraph $G$ is defined as $m(G) - n(G)$.

**Definition 2.** A kernel is a multigraph with minimum degree at least 3. If all vertices have degree exactly 3, the graph is called a cubic multigraph.

In order to enumerate kernels by their excess, we need to make sure that the number of kernels of fixed excess $k$ is finite. This is indeed true since

$$2m = \sum_{v \in V(G)} \deg(v) \geq 3n \implies n \leq 2k \text{ and } m \leq 3k.$$ 

Furthermore, we observe that each connected multigraph $G$ of excess $k$ can be uniquely converted into a connected kernel of excess $k$ via the following procedure:

- Repeatedly remove all vertices of degree 1;
- Repeatedly replace every vertex of degree 2 with an edge connecting its neighbors. Remove isolated loops.

Both steps preserve excess and connectivity, and the procedure removes all vertices of degree less than 3, so we are left with a kernel of excess $k$ at the end. It can easily be shown (by induction on iterations in every step) that the connected kernel thus obtained is unique. In other words, every connected multigraph of excess $k$ can be built uniquely from a connected kernel of excess $k$, by replacing all vertices of the kernel with rooted trees, and replacing all edges of the kernel with paths of rooted trees (see Figure 1).

**Remark 1.** The smallest possible excess without disconnecting a graph is $k = -1$, corresponding to trees. An excess of $k = 0$ gives rise to multi-unicycles. The kernels of trees and multi-unicycles are empty.

We can describe this construction of connected multigraphs from connected kernels symbolically using the substitution operator (see Table 1).

---

2In the case of connected multigraphs, there can only be one isolated loop at the very end of the procedure.

3The substitution $A \circ B$ replaces every atom of $A$ with an element of $B$. 

---
Figure 1: Constructing a connected multigraph (right) from a connected kernel (left).
The construction preserves excess — both graphs have excess 1. Vertex and edge labels as well as edge orientations have been omitted for clarity.

3.3 Generating function for connected kernels

Let $CK_k(z, w)$ be the generating function for connected kernels of excess $k$, and let $ck_{n,m}$ be the number of such kernels on $n$ vertices and $m$ edges. We showed in the Subsection 3.2 that $ck_{n,m}$ is finite, and, for small excess $k$, $ck_{n,m}$ is a computable polynomial.

The generating function for connected kernels of excess $k$ is thus

$$CK_k(z, w) = \sum_{m \leq 3k} ck_{m-k,m} \frac{w^m z^{m-k}}{2^m m! (m-k)!}.$$

To build multigraphs from kernels, we will introduce the generating function of rooted trees, characterized by the recursive relation $T = Z \ast \text{SET}(T)$, yielding the generating function $T(z) = ze^{T(z)}$ (also known as the Cayley function, see Flajolet and Sedgewick [6] for a full treatment). The bivariate generating function for rooted trees is $T(zw)/w$, where the division by $w$ ensures that a tree has one fewer edges than vertices.

Let $P$ be an oriented path with rooted trees hanging from its vertices\(^4\). Then $P$ can be thought of as a sequence of (rooted tree, edge) pairs, with an additional edge added to the end, giving the following generating function:

$$P(z, w) = \frac{1}{1 - \left(\frac{T(zw)}{w} \cdot w\right) \cdot w} = \frac{w}{1 - T(zw)}.$$

The generating function for connected multigraphs of excess $k$, $\text{CMG}_k(z, w)$, is then given via substitution,

$$\text{CMG}_k(z, w) = CK_k \left( \frac{T(zw)}{w}, \frac{w}{1 - T(zw)} \right) = \sum_{m \leq 3k} ck_{m-k,m} \frac{T(zw)^m}{2^m m! (m-k)!} \cdot \left(\frac{w}{1 - T(zw)}\right)^{m-k}.$$

For enumeration purposes, we need an expression for $cmg_{n,n+k}$, which can be written as

$$n! [z^n] \text{CMG}_k = cmg_{n,n+k} \frac{w^{n+k}}{2^{n+k} (n+k)!} \Longrightarrow cmg_{n,n+k} = 2^{n+k}(n+k)!n! [z^n] \text{CMG}_k(z, 1),$$

\(^4\)Not counting the two vertices at the ends, since we are replacing edges of the kernel with these paths; the vertices at the ends originally belonged to the kernel and will be replaced by rooted trees.
where \([z^n]f(z)\) denotes the coefficient of \(z^n\) in the expansion of \(f(z)\) and \(\text{cmg}_{n,m}\) denotes the number of connected multigraphs on \(n\) vertices and \(m\) edges. Furthermore,

\[
\text{CMG}_k(z, 1) = \sum_{m \leq 3k} c_{k-m,k,m} \frac{\left(\frac{1}{1-T(z)}\right)^m}{2^m m!} T(z)^{m-k} (m-k)!.
\]

### 3.4 Asymptotic analysis

We will derive the asymptotics of the univariate function \(\text{CMG}_k(z) = \text{CMG}_k(z, 1)\), which is a rational function in \(T(z)\).

Using Theorem 4 from Lumbroso and Morcrette [7] (itself based on Theorem VI.6 on page 404 of Flajolet and Sedgewick [5]),

\[
T(z) \sim z^{-1} (1 - \sqrt{2} \sqrt{1 - ez})
\]

Next, applying Corollary VI.1 (sim-transfer) on page 404 of Flajolet and Sedgewick [5] to every term of the sum in \(\text{CMG}_k(z)\) as \(z \to e^{-1}\), the dominant singularity of \(T(z)\),

\[
[z^n] \text{CMG}_k(z) \sim \sum_{m \leq 3k} \frac{c_{k-m,k,m}}{2^m m! (m-k)!} \frac{1}{2^{m/2}} \frac{n^{m/2-1} e^n}{\Gamma(m/2)}.
\]

The term corresponding to \(m = 3k\) is asymptotically dominant, giving the following asymptotic expression for the number of connected multigraphs of excess \(k\):

\[
\text{cmg}_{n,n+k} = n! 2^{n+k} (n+k)! [z^n] \text{CMG}(z)
\]

\[
\sim n! 2^{n+k} (n+k)! \frac{c_{2k,3k}}{2^{3k} (3k)! (2k)!} \frac{1}{2^{3k/2}} \frac{n^{3k/2-1} e^n}{\Gamma(3k/2)}.
\]

### 3.5 Counting cubic multigraphs

As we saw the dominating term for the number of connected multigraphs of excess \(k\) come from kernels that are cubic multigraphs. The enumeration of such kernels \(c_{2k,3k}\) in fact has a closed form: \(\frac{(6k)!}{(3!)^{2k}}\).

### 4 Limit laws

Generating functions can be used to derive limit laws for different parameters. For example, we can consider a multigraph \(G\) drawn uniformly at random from all multigraphs on \(n\) vertices and \(m\) edges. Let \(X_d\) be a random variable denoting the number of vertices of \(G\) with degree \(d\). We might be interested in the complete distribution of \(X_d\) (with access to all the moments) as \(n\) goes to infinity. Alternatively, we could consider the joint distribution of the vector of random variables \(X = (X_0, X_1, \cdots)\).

In this section, we will discuss how to derive such results using generating functions, and we will use those techniques to derive the Gaussian limit law for the degrees in uniformly random multigraphs.

---

5 This can be seen by representing the cubic multigraph as a sequence of 6\(k\) vertices, in the order of edge labels, and pairwise ordered corresponding to the orientation of edges. Forcing the degrees to equal 3 is equivalent to forcing every vertex to appear exactly 3 times in this sequence. The number of ways to do this is \(\binom{6k}{3,3,\ldots,3} = \frac{(6k)!}{(3!)^{2k}}\).
4.1 Limit laws through generating functions

Once we are equipped with a generating function, we can study combinatorial parameters by marking the parameter using a variable. The resulting marked generating function can be used to derive probability generating functions as well as moment generating functions for the value of the parameter. The moments can be used to show convergence to a distribution in the limit.

More formally, let \( F \) be a combinatorial family, with a generating function \( F(z,u) \), where \( z \) is the size-marking variable and \( u \) is the parameter-marking variable. Additionally, let \( f_{n,t} = [z^n u^t] F(z,u) \) denote the number of members of \( F \) with size \( n \) and parameter value \( t \). Finally, let \( X_n \) be the random variable denoting the value of the parameter in an element of size \( n \) from \( F \) drawn uniformly at random from all size-\( n \) elements.

The probability generating function for \( X_n \) is given by

\[
P_n(u) = \sum_t P[X_n = t] u^t = \frac{f_{n,t}}{\sum_s f_{n,s}} u^t = \frac{[z^n] F(z,u)}{[z^n] F(z,1)}.
\]

Repeated differentiation then gives access to the mean, variance, etc.:

\[
E[X_n] = \frac{d}{du} P_n(1), \quad \text{and} \quad E[X_n(X_n - 1)] = \frac{d^2}{du^2} P_n(1), \ldots.
\]

Furthermore, the moment-generating functions are given by

\[
M_{X_n}(t) = E[e^{tX_n}] = P_n(e^t) = \frac{[z^n] F(z,e^t)}{[z^n] F(z,1)}.
\]

4.2 Multigraphs with marked degrees

In order to derive limit laws for the degrees in a multigraph, we will introduce the infinite-dimensional vector of marker variables \( \delta \), where \( \delta_d \) marks the vertices of degree \( d \). The generating function for these marker variables is

\[
\Delta(x) = \sum_{d \geq 0} \delta_d x^d.
\]

The multivariate generating function of a family \( F \) of multigraphs with marked degrees is then

\[
F(z, w, \delta) = \sum_{G \in F} \left( \prod_{v \in V(G)} \delta_{\deg(v)} \right) \frac{z^{n(G)}}{n(G)!} \frac{w^{m(G)}}{m(G)!}.
\]

Next, as noted by de Panafieu (2016)[2], we can bijectively cut every multiedge \((v, w, l)\) into a pair of half-edges with new labels, \((v, 2l)\) and \((w, 2l+1)\) (Figure 2). A multigraph is a set of labelled half-edges, or alternatively a set of vertices, each having a set of half-edges incident to them. The generating function for the (vertex, set of half-edges) pair is \( e^{z \Delta(x)} \), giving the following reformulation of the generating function for multigraphs:

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

Therefore,

\[
n! 2^m m! [z^n w^m] MG(z, w, \delta) = (2m)! [x^{2m}] \Delta(x)^n.
\]
4.3 Limit law of degrees

Suppose the number of edges is $m(n)$, a given function of the number of vertices. Using the technique outlined in Subsection 4.1, define $X_n$ as the number of vertices of some fixed degree $d$ in a uniformly random multigraph with $n$ vertices and $m(n)$ edges. We set $\delta_i = 1$ for all $i \neq d$ and show that $X_n$ is normally distributed when $m = \Theta(n)$.

The probability generating function of $X_n$ is

$$P_n(\delta) = P_n(\delta_d) = \frac{[z^n w^m] \text{MG}(z, w, \delta_d)}{[z^n w^m] \text{MG}(z, w, 1)}$$

$$= \frac{[x^{2m}] \left( \delta_d \frac{x^d}{d!} + \sum_{i > 0, i \neq d} \frac{x^i}{i!} \right)}{[x^{2m}] \sum_{i > 0} \frac{x^i}{i!}}$$

$$= \frac{(2m)!}{n^{2m}} [x^{2m}] \left( e^x + (\delta_d - 1) \frac{x^d}{d!} \right)^n.$$

4.4 Observing sharp threshold for disappearance of isolated vertices

Before we continue to the limit law derivation, as a sanity check, let us use this probability generating function to compute the expected number of isolated vertices (degree 0) near its threshold. This is given by

$$E[X_n] = \frac{d}{d \delta_d} P_n(1) = \frac{(2m)!}{n^{2m}} [x^{2m}] ne^{x(n-1)} \frac{x^d}{d!}$$

$$= \frac{(2m)!}{d! n^{2m-1}} \left[ x^{2m-d} \right] e^{x(n-1)}$$

$$= \frac{(2m)!}{d! n^{2m-1}} (n-1)^{2m-d}.$$

In the case of isolated vertices, $d = 0$, and we have

$$E[X_n] = \frac{(n-1)^{2n}}{n^{2m-1}} \sim n - 2m + \frac{(2m)}{n^2} - \frac{(2m^3)}{n^4} + \cdots.$$

We know that the threshold for the disappearance of isolated vertices in a random graph is a sharp one at $p = \log(n)/n$. This corresponds to $m(n) \sim \binom{n}{2} \frac{\log(n)}{n} \sim n \log(n)$. Indeed we observe that

$$\lim_{n \to \infty} \frac{(n-1)^{2n \log(n)}}{n^{2n \log(n-1)}} = 0,$$

i.e. around threshold, there are no isolated vertices, whereas

$$\lim_{n \to \infty} \frac{(n-1)^{2n \log \log(n)}}{n^{2n \log \log(n-1)}} = \infty,$$

implying sharpness of the threshold.
4.5 Back to limit law of degrees

Back to the limit law, using Theorem VIII.8 (Saddle-point estimates of large powers) on page 588 of Flajolet and Sedgewick [6], we can transform the coefficient extraction part \( [x^{2m}] \left( e^x + (\delta_d - 1) \frac{x^d}{d!} \right)^n \) to the form

\[
\frac{B(\xi)^n}{\xi^{2m+1} \sqrt{2\pi n\xi}} \cdot (1 + o(1)),
\]

where \( B(x) = e^x + (\delta_d - 1) \frac{x^d}{d!} \), and \( \xi \) is the unique root of \( \xi B'(\xi)/B(\xi) = 2m/n \). Note that the assumption that \( m = \Theta(n) \) is used here to ensure a reasonable \( \xi \).

This expression satisfies the format required by Theorem IX.8 (Quasi-powers Theorem) on page 645 of Flajolet and Sedgewick [6] (simply plug in \( \beta_n = n \) and check variability condition), which guarantees a Gaussian distribution for \( X_n \) up to rescaling.

5 Next steps and concluding remarks

5.1 Extension to multiple communities

Our techniques can be extended to cover the limit laws in the case of multiple communities, e.g. the stochastic block model.

We can define the \( k \)-community multigraph \( G \) as a \((2k + \binom{k}{2})\)-tuple:

\[
\left( (V_i(G))_{1 \leq i \leq k}, (E_{ii}(G))_{1 \leq i \leq k}, (E_{ij}(G))_{1 \leq i < j \leq k} \right).
\]

where,

- \( V_i(G) \) denotes the vertices of \( G \) in community \( i \), and \( n_i(G) = |V_i(G)| \);
- \( E_{ii}(G) \) refers to oriented edges of \( G \) within community \( i \), and \( m_{ii}(G) = |E_{ii}(G)| \);
- \( E_{ij}(G) \) refers to unoriented edges of \( G \) across communities \( i \) and \( j \), and \( m_{ij}(G) = |E_{ij}(G)| \).

Additionally, we can define

- vertex markers \( z = (z_1, \ldots, z_k) \),
- intra-community edge-markers \( w = (w_1, \ldots, w_k) \), and
- inter-community edge-markers \( u = (u_1, \ldots, u_{\binom{k}{2}}) \).

One can assign corresponding probabilities from the stochastic block model to \( w \)s and \( u \)s to weigh each graph appropriately.

With these variables, the generating function of a family of \( k \)-community multigraphs is given by

\[
F(z, w, u) = \sum_{G \in \mathcal{F}} \left( \prod_{1 \leq i \leq k} \frac{z_i^{n_i(G)}}{n_i(G)!} \right) \left( \prod_{1 \leq i \leq k} \frac{w_i^{m_{ii}(G)}}{m_{ii}(G)!} \right) \left( \prod_{1 \leq i < j \leq k} \frac{w_{ij}^{m_{ij}(G)}}{m_{ij}(G)!} \right).
\]

Similarly, we can cut edges into half-edges, mark the degrees, etc., to derive parametrized generating functions suitable for deriving limit laws. This could be the subject of future work.

\footnote{A simpler representation would be \( \left( (V_i(G))_{1 \leq i \leq k}, (E_{ii}(G))_{1 \leq i \leq k} \right) \), but would make the contrast between inter- and intra-community edges less apparent.}
5.2 From Multigraphs to Simple Graphs

See Lemma 1 of de Panafieu [2] for an expression of the generating function of (simple) graphs in terms of the generating function of multigraphs.

5.3 A note on equivalence of models

Recall that Erdős-Rényi random graphs can be studied through the uniform model (draw uniformly at random from all graphs on $n$ vertices and $m$ edges) and the binomial model (draw each edge with probability $p$). Recall also that for $m = \lfloor \binom{n}{2} p \rfloor$, standard concentration inequalities ensure that the two models behave similarly for simple graphs. While probabilistic approaches deal with the binomial method, combinatorial approaches work with the uniform model.

The equivalent of the binomial model in the case of multigraphs can be drawn by assigning a random permutation as edge labels to an Erdős-Rényi random graph, where multiedges and loops are allowed. Allowing the number of edges to be a random variable (as opposed to a fixed value as in the uniform model) can skew the distribution away from uniform. Thus, in order to generalize our results from the uniform model to the binomial model, we must ensure that the introduced skewing does not affect the limit laws, presumably using similar concentration equalities. This could be a potential subject for future work.

References


