The Birth of the Giant Component

Dedicated to Paul Erdős on his 80th birthday

Svante Janson, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel

Department of Mathematics, Uppsala University;
Computer Science Department, Stanford University;
Department of Discrete Mathematics, Adam Mickiewicz University; and
Department of Mathematics, The Ohio State University

Abstract. Limiting distributions are derived for the sparse connected components that are present when a random graph on n vertices has approximately $\frac{1}{2}n$ edges. In particular, we show that such a graph consists entirely of trees, unicyclic components, and bicyclic components with probability approaching $\sqrt{\frac{2}{3}}\cosh\sqrt{\frac{5}{18}} \approx 0.9325$ as $n \to \infty$. The limiting probability that it consists of trees, unicyclic components, and at most one other component is approximately 0.9957; the limiting probability that it is planar lies between 0.987 and 0.9998. When a random graph evolves and the number of edges passes $\frac{1}{2}n$, its components grow in cyclic complexity according to an interesting Markov process whose asymptotic structure is derived. The probability that there never is more than a single component with more edges than vertices, throughout the evolution, approaches $5\pi/18 \approx 0.8727$. A "uniform" model of random graphs, which allows self-loops and multiple edges, is shown to lead to formulas that are substantially simpler than the analogous formulas for the classical random graphs of Erdős and Rényi. The notions of "excess" and "deficiency," which are significant characteristics of the generating function as well as of the graphs themselves, lead to a mathematically attractive structural theory for the uniform model. A general approach to the study of stopping configurations makes it possible to sharpen previously obtained estimates in a uniform manner and often to obtain closed forms for the constants of interest. Empirical results are presented to complement the analysis, indicating the typical behavior when n is near 20000.

0. Introduction. When edges are added at random to n initially disconnected points, for large n, a remarkable transition occurs when the number of edges becomes approximately $\frac{1}{2}n$. Erdős and Rényi [13] studied random graphs with n vertices and $\frac{n}{2}(1+\mu)$ edges

---

This research was supported in part by the National Science Foundation under grant CCR-86-10181, and by Office of Naval Research contract N00014-87-K-0502.

as $n \to \infty$, and discovered that such graphs almost surely have the following properties: If $\mu < 0$, only small trees and "unicyclic" components are present, where a unicyclic component is a tree with one additional edge; moreover, the size of the largest tree component is $(\mu - \ln(1+\mu))^{-1} \ln n + O(\log\log n)$. If $\mu = 0$, however, the largest component has size of order $n^{2/3}$. And if $\mu > 0$, there is a unique "giant" component whose size is of order n; in fact, the size of this component is asymptotically αn when $\mu = -\alpha^{-1}\ln(1-\alpha) - 1$. Thus, for example, a random graph with approximately $n \ln 2$ edges will have a giant component containing $\sim \frac{1}{2}n$ vertices.

The research that led to the present paper began in a rather curious way, as a result of a misunderstanding. In 1988, the students in a class taught by Richard M. Karp performed computer experiments in which graphs with a moderately large number of vertices were generated by adding one edge at a time. A rumor spread that these simulations had turned up a surprising fact: As each of the random graphs evolved, the story went, never once was there more than a single "complex" component; i.e., there never were two or more components present simultaneously that were neither trees nor unicyclic. Thus, the first connected component that acquired more edges than vertices was destined to be the giant component. As more edges were added, this component gradually swallowed up all of the others, and none of the others ever became complex before they were swallowed.

Reports of those experiments suggested that a great simplification of the theory of evolving graphs might be possible. Could it be that such behavior occurs almost always, i.e., with probability approaching 1 as $n \to \infty$? If so, we could hope for the existence of a much simpler explanation of the fact that a giant component emerges during the graph process, and we could devise rather simple algorithms for online graph updating that would take advantage of the unique-complex-component phenomenon. At that time the authors who began this investigation (DEK and BP) were unaware of Stepanov's posthumous paper [36]. We were motivated chiefly by the work of Bollobás [5], who had shown that a component of size $\geq n^{2/3}$ is almost always unique once the number of edges exceeds $\frac{1}{2}n + 2(\ln n)^{1/2} n^{2/3}$; moreover, Bollobás proved that such a component gets approximately 4 vertices larger when each new edge is added. His results blended nicely with the unique-complex-component conjecture.

However, we soon found that the conjecture is false: There is nonzero probability that a graph with $\frac{1}{2}n$ edges will contain several pretenders to the giant throne, and this probability increases when the number of edges is slightly more than $\frac{1}{2}n$. We also learned that Stepanov [36] had already obtained similar results. Thus we could not hope for a theory of random graphs that would be as simple as the conjecture promised. On the other hand, we learned that the graph evolution process does satisfy the conjecture with reasonably high probability; hence algorithms whose efficiency rests on the assumption of a unique complex component will not often be inefficient.

Further analysis revealed, in fact, that we must have misunderstood the initial reports of experimental data. The actual probability that an evolving graph never has two complex components approaches the limiting value $5\pi/18 \approx 0.8727$; therefore the rumor that got us started could not have been true. In fact, the computer experiments by Karp's students had simply reported the state of the graph when exactly $\frac{1}{2}n$ edges were present, and at certain other fixed reporting times. A false impression arose because there is high probability that a random graph with $\frac{1}{2}n$ edges has at most one complex component; indeed, the probability is $0.9957 + O(n^{-1/3})$. More complicated configurations sometimes arise momentarily just after $\frac{1}{2}n$ edges are reached. However, the fallacious rumor of 1988 has turned out to have beneficial effects, because it was a significant catalyst for the discovery of some remarkably beautiful patterns.

Sections 1–10 of this paper provide a basic introduction to the theory of evolving graphs and multigraphs, using generating functions as the principal tool. Two models of graph evolution are presented in section 1, the "graph process" and the "multigraph process." Their generating functions are introduced in section 2, and special aspects of those functions related to trees and cycles are discussed in section 3. Section 4 explains how to derive properties of a graph's more complex features by means of differential equations; the equations are solved for multigraphs in section 5 and for graphs in section 6. The resulting decomposition of multigraphs turns out to be surprisingly regular. Section 7 explains the regularities and begins to analyze the algebraic properties of the functions obtained in section 5. Related results for connected graphs are discussed in section 8. Section 9 explains the combinatorial significance of the algebraic structure derived earlier. Finally, section 10 presents a quantitative lemma about the characteristics of random graphs near the critical point $\mu = 0$, making it possible to derive exact values for many relevant statistics.

Readers who cannot wait to get to the "good stuff" should skim sections 1–10 and move on to section 11, which begins a sequence of applications of the basic theory. The first step is to analyze the distribution of bicyclic components; then, in section 12, the same ideas are shown to yield the joint distribution of all kinds of components. The formulas obtained there have a simple structure suggesting that the traditional approach of focussing on connected components is unnecessarily complicated; we obtain a simpler and more symmetrical theory if we first consider the *excess* of edges over vertices, exclusive of tree components, then look at other properties like connectedness after conditioning on the excess. Section 13 motivates this principle, and section 14 derives the probability distribution of a graph's excess as it passes the critical point. These ideas help to nail down the probability that a graph with $\frac{1}{2}n$ edges is planar, as shown in section 15.

Section 16 begins the discussion of what may well be the most important notion in this paper; readers who have time for nothing else are encouraged to look at Figure 1, which shows the initial stages of the "big bang." The evolution of a graph or multigraph passes through discrete transitions as the excess increases, and important aspects of those changes are illustrated in Figure 1; section 17 proves that this illustration represents a Markov process that characterizes almost all graph evolutions. The $\frac{5\pi}{18}$ phenomenon alluded to above is discussed in section 18, which establishes $\frac{5\pi}{18}$ as an upper bound for the probability in question. Section 19 shows that, for small n, the probability of retaining at most one complex component during the critical stage is in fact greater than $\frac{5\pi}{18}$, decreasing monotonically with n.

The excess of a graph is of principal importance at the critical point, but a secondary concept called *deficiency* becomes important shortly thereafter. A graph with deficiency 0 is called "clean"; such graphs are obtained from 3-regular graphs by splitting edges and/or by attaching trees to vertices of cycles. Section 20 explains how deficiency evolves jointly with increasing excess. Figure 2, at the end of that section, illustrates another Markov process that goes on in parallel with Figure 1. Section 21 shows that most graphs stay clean until they have acquired approximately $\frac{1}{2}n + n^{3/4}$ edges. Section 22 looks more closely at the moment a graph first becomes unclean.

Section 23 tracks the growth of excess and deficiency as a multigraph continues to evolve through $\frac{1}{2}n + n^{4/5}$, $\frac{1}{2}n + n^{5/6}$, ... edges. The excess and deficiency are shown to be approximately normally distributed about certain well-defined values. Specifically, when the number of edges is $\frac{n}{2}(1 + \mu)$, with $\mu = o(1)$, the excess will be approximately $\frac{2}{3}\mu^3 n$ and the deficiency will be approximately $\frac{2}{3}\mu^4 n$. These statistics complement the well-known fact that the emerging giant component has almost surely grown to encompass approximately $2\mu n$ vertices.

Sections 24 to 26 develop a theory of "stopping configurations," by which it is possible to study the first occurrences of various events during a multigraph's evolution. In particular, an explicit formula is derived for the asymptotic distribution of the time when the excess first reaches a given value r. A closed formula is derived for the "first cycle constant" of [14].

Section 27 completes the discussion initiated in sections 17 and 18, by proving the $\frac{5\pi}{18}$ phenomenon as a special case of a more general result about the infinite Markov process in Figure 1.

Finally, section 28 presents empirical data, showing to what extent the theory relates to practice when n is not too large. Section 29 discusses a number of open questions raised by this work.

1. Graph evolution models. We shall consider two ways in which a random graph on n vertices might evolve, corresponding to sampling with and without replacement. The first of these, introduced implicitly in [4] and explicitly in [7, proof of Lemma 2.7] and [14], turns out to be simpler to analyze and simpler to simulate by computer, therefore more likely to be of importance in applications to computer science: We generate ordered pairs $\langle x, y\rangle$ repeatedly, where $1 \le x, y \le n$, and add the (undirected) edge x — y to the graph. Each ordered pair $\langle x, y\rangle$ occurs with probability $1/n^2$, so we call this the *uniform model* of random graph generation. It may also be called the *multigraph process*, because it can generate graphs with self-loops x — x, and it can also generate multiple edges. Notice that a self-loop x — x is generated with probability $1/n^2$, while an edge x — y with $x \ne y$ is generated with probability $2/n^2$ because it can occur either as $\langle x, y\rangle$ or $\langle y, x\rangle$.

The second evolution procedure, introduced by Erdős and Rényi [12], is called the *permutation model* or the *graph process*. In this case we consider all $N = \binom{n}{2}$ possible edges x — y with $x < y$ and introduce them in random order, with all $N!$ permutations considered equally likely. In this model there are no self-loops or multiple edges.

A multigraph M on n labeled vertices can be defined by a symmetric $n \times n$ matrix of nonnegative integers m_{xy}, where $m_{xy} = m_{yx}$ is the number of undirected edges x — y in G. For purposes of analysis, we shall assign a *compensation factor*

$$\kappa(M) = 1 \Bigg/ \prod_{x=1}^{n} \left(2^{m_{xx}} \prod_{y=x}^{n} m_{xy}! \right) \tag{1.1}$$

to M; if $m = \sum_{x=1}^{n} \sum_{y=x}^{n} m_{xy}$ is the total number of edges, the number of sequences $\langle x_1, y_1\rangle\langle x_2, y_2\rangle \ldots \langle x_m, y_m\rangle$ that lead to M is then exactly

$$2^m\, m!\, \kappa(M)\,. \tag{1.2}$$

(The factor 2^m accounts for choosing either $\langle x, y\rangle$ or $\langle y, x\rangle$; the $2^{m_{xx}}$ in the denominator of $\kappa(M)$ compensates for the case $x = y$. The other factor $m!$ accounts for permutations of the pairs, with $m_{xy}!$ in $\kappa(M)$ to compensate for permutations between multiple edges.)

Equation (1.2) tells us that $\kappa(M)$ is a natural weighting factor for a multigraph M, because it corresponds to the relative frequency with which M tends to occur in applications. For example, consider multigraphs on three vertices $\{1, 2, 3\}$ having exactly three edges. The edges will form the cycle $M_1 = \{1 — 2,\ 2 — 3,\ 3 — 1\}$ much more often than they will form three identical self-loops $M_2 = \{1 — 1,\ 1 — 1,\ 1 — 1\}$, when the multigraphs are generated in a uniform way. For if we consider the 3^6 possible sequences $\langle x_1, y_1\rangle\langle x_2, y_2\rangle\langle x_3, y_3\rangle$ with $1 \le x, y \le 3$, only one of these generates the latter multigraph, while the cyclic multigraph is obtained in $2^3\,3! = 48$ ways. Therefore it makes sense to assign weights so that $\kappa(M_2) = \frac{1}{48}\kappa(M_1)$, and indeed (1.1) gives $\kappa(M_1) = 1$, $\kappa(M_2) = \frac{1}{48}$.

Notice that a given multigraph M is a graph—i.e., it has no loops and no multiple edges—if and only if $\kappa(M) = 1$. Notice also that if M consists of several disjoint components $M_1, \ldots, M_k$, with no edges between vertices of M_i and M_j for $i \ne j$, we have

$$\kappa(M) = \kappa(M_1)\, \ldots\, \kappa(M_k)\,. \tag{1.3}$$

2. Generating functions. We shall use bivariate generating functions (bgf's) to study labeled graphs and multigraphs and their connected components. If $\mathcal{F}$ is a family of multigraphs with labeled vertices, the associated bgf is the formal power series

$$F(w,z) = \sum_{M\in\mathcal{F}} \kappa(M)\, w^{m(M)} \frac{z^{n(M)}}{n(M)!}\,, \tag{2.1}$$

where $m(M)$ and $n(M)$ denote the number of edges and the number of vertices of M. We can do many operations on such power series without regard to convergence. It follows from (1.2) and (2.1) that m steps of the uniform evolution model on n vertices will produce a multigraph in $\mathcal{F}$ with probability

$$\frac{2^m\, m!\, n!}{n^{2m}}\, [w^m z^n]\; F(w,z)\,, \tag{2.2}$$

where the symbol $[w^m z^n]$ denotes the coefficient of $w^m z^n$ in the formal power series that follows it. Similarly, if $\mathcal{F}$ is a family of graphs with labeled vertices, the probability that m steps of the permutation model will produce a graph in $\mathcal{F}$ is

$$\frac{n!}{\binom{N}{m}}\, [w^m z^n]\, F(w,z)\,, \qquad N = \binom{n}{2}\,. \tag{2.3}$$

Formulas (2.2) and (2.3) are asymptotically related by the formula

$$\binom{N}{m} = \frac{n^{2m}}{2^m\, m!} \exp\left(-\frac{m}{n} - \frac{m^2}{n^2} + O\left(\frac{m}{n^2}\right) + O\left(\frac{m^3}{n^4}\right)\right)\,, \qquad 0 \le m \le N, \tag{2.4}$$

which follows from Stirling's approximation.

Incidentally, the exponential factor in (2.4) is the probability that m steps of the multigraph process will produce no self-loops or multiple edges. When $m = \frac{1}{2}n$, this probability is $e^{-3/4} + O(n^{-1}) \approx 0.472$.

When we say that the n vertices of a multigraph are "labeled," it is often convenient to think of the labeling as an assignment of the numbers 1 to n. But a strict numeric convention would require us to recompute the labels whenever vertices are removed or when multigraphs are combined. The actual value of a label is, in fact, irrelevant; what really counts is the relative order *between* labels. Labeled multigraphs are multigraphs whose vertices have been totally ordered. In this paper all graphs and multigraphs are assumed to be labeled, i.e., totally ordered, even when the adjective "labeled" is not stated.

The bgf (2.1) is an exponential generating function in z, and the factor $\kappa(M)$ is multiplicative according to (1.3). Therefore the product of bgf's

$$F_1(w,z)F_2(w,z)\, \ldots\, F_k(w,z)$$

represents ordered k-tuples of labeled multigraphs $\langle M_1, M_2, \ldots, M_k\rangle$, each M_j being from family $\mathcal{F}_j$. Unordered k-tuples $\{M_1, \ldots, M_k\}$ from a common family $\mathcal{F}$ have the bgf $F(w,z)^k/k!$, if $\mathcal{F}$ does not include the empty multigraph. For example, the bgf for a 3-cycle is $w^3z^3/3!$, and the bgf for two isolated vertices is $z^2/2!$; hence the bgf for a 3-cycle and two isolated vertices is $(w^3z^3/6)(z^2/2) = 10w^3z^5/5!$. (There are 10 such graphs, one for each choice of the isolated points.)

Let $C(w,z)$ be the bgf for all connected multigraphs, and let $G(w,z)$ be the bgf for the set of all multigraphs. Then we have

$$e^{C(w,z)} = \sum_{k\geq 0} \frac{C(w,z)^k}{k!} = G(w,z) \tag{2.5}$$

because the term $C(w,z)^k/k!$ is the bgf for multigraphs having exactly k components. Similarly, if $\widehat{C}(w,z)$ and $\widehat{G}(w,z)$ are the corresponding bgf's for graphs instead of multigraphs, we have

$$e^{\widehat{C}(w,z)} = \widehat{G}(w,z)\,, \tag{2.6}$$

a well-known formula due to Riddell [32]. The bgf for all graphs is obviously

$$\widehat{G}(w,z) = \sum_{n\geq 0} (1+w)^{n(n-1)/2} \frac{z^n}{n!}\,. \tag{2.7}$$

Therefore (2.6) gives us the bgf for connected graphs,

$$\begin{aligned}\widehat{C}(w,z) &= \ln\left(1 + z + (1+w)\frac{z^2}{2} + (1+w)^3\frac{z^3}{6} + \cdots\right)\\ &= z + w\frac{z^2}{2} + (3w^2 + w^3)\frac{z^3}{6} + \cdots\,.\end{aligned} \tag{2.8}$$

The bgf $G(w,z)$ for all multigraphs can be found as follows: The coefficient of $z^n/n!$ is $\sum \kappa(M)w^{m(M)}$, summed over multigraphs M on n vertices. This is

$$\prod_{x=1}^{n}\left(\left(\sum_{m_{xx}\geq 0} \frac{w^{m_{xx}}}{2^{m_{xx}}\, m_{xx}!}\right) \prod_{y=x+1}^{n} \left(\sum_{m_{xy}\geq 0} \frac{w^{m_{xy}}}{m_{xy}!}\right)\right) = \prod_{x=1}^{n} e^{w/2}(e^w)^{n-x} = e^{wn^2/2}\,.$$

Hence the desired formula is slightly simpler than (2.7):

$$G(w,z) = \sum_{n\geq 0} e^{wn^2/2} \frac{z^n}{n!}\,. \tag{2.9}$$

The corresponding bgf for connected multigraphs is therefore

$$\begin{aligned}C(w,z) &= \ln G(w,z)\\ &= \bigl(1+\tfrac12 w+\tfrac18 w^2+\tfrac1{48}w^3+\cdots\bigr)z+\bigl(w+\tfrac32 w^2+\tfrac76 w^3+\cdots\bigr)\frac{z^2}{2}\\ &\qquad+\bigl(3w^2+\tfrac{17}{2}w^3+\cdots\bigr)\frac{z^3}{6}+\cdots\,. \end{aligned}\tag{2.10}$$

In this case the coefficient of w^3z^3 is $\frac{17}{2}/3!$, because the connected multigraphs with three edges on three vertices have total weight $\frac{17}{2}$. (The 3-cycle has weight 1; there are 9 multigraphs obtainable by adding a self-loop to a tree, each of weight $\frac12$; and there are six multigraphs obtainable by doubling one edge of a tree, again weighted by $\frac12$.)

Notice that expression (2.2) is $[w^m z^n]\,F(w,z)\,/\,[w^m z^n]\,G(w,z)$, the ratio of the weight of multigraphs in $\mathcal{F}$ to the weight of all possible multigraphs. Similarly, expression (2.3) is $[w^m z^n]\,\widehat{F}(w,z)\,/\,[w^m z^n]\,\widehat{G}(w,z)$.

It is convenient to group the terms of (2.8) and (2.10) according to the excess of edges over vertices in connected components. Let $\mathcal{C}_r$ and $\widehat{\mathcal{C}}_r$ denote the families of connected multigraphs and graphs in which there are exactly r more edges than vertices; let $C_r(w,z)$ and $\widehat{C}_r(w,z)$ be the corresponding bgf's. Then we have

$$\begin{aligned}C(w,z) &= \sum_r C_r(w,z) = \sum_r w^r C_r(wz)\,,\\ \widehat{C}(w,z) &= \sum_r \widehat{C}_r(w,z) = \sum_r w^r \widehat{C}_r(wz)\,,\end{aligned}\tag{2.11}$$

where $C_r(z)$ and $\widehat{C}_r(z)$ are univariate generating functions for $\mathcal{C}_r$ and $\widehat{\mathcal{C}}_r$. A univariate generating function $F(z)$ is $\sum \kappa(M)z^n/n!$, summed over all graphs or multigraphs in a given family $\mathcal{F}$. We obtain it from a bgf by setting $w=1$, thereby ignoring the number of edges. Univariate generating functions are easier to deal with than bgf's, so we generally try to avoid the need for two independent variables whenever possible.

3. Trees, unicycles, and bicycles. Let us say that a connected component has *excess* r if it belongs to $\mathcal{C}_r$, i.e., if it has r more edges than vertices. A connected graph on n vertices must have at least $n-1$ edges. Hence $C_r=0$ unless $r\ge -1$. In the extreme case $r=-1$, we have $\mathcal{C}_{-1}=\widehat{\mathcal{C}}_{-1}$, the family of all unrooted trees, which are *acyclic components*. In the next case $r=0$, the generating functions C_0 and $\widehat{C}_0$ represent *unicyclic components*, which are trees with an additional edge. Similarly, C_1 and $\widehat{C}_1$ represent *bicyclic components*. In the present paper we shall deal extensively with sparse components of these three kinds, so it will be convenient to use the special abbreviations

$$\begin{aligned}&U(z)=C_{-1}(z)=\widehat{C}_{-1}(z) &&\text{for unrooted trees;}\\ &V(z)=C_0(z)\text{ and }\widehat{V}(z)=\widehat{C}_0(z) &&\text{for unicyclic components;}\\ &W(z)=C_1(z)\text{ and }\widehat{W}(z)=\widehat{C}_1(z) &&\text{for bicyclic components.}\end{aligned}$$

According to a well-known theorem of Sylvester [37] and Borchardt [8], often attributed erroneously to Cayley [10] although Cayley himself credited Borchardt, we have $U(z) = \sum_{n=1}^{\infty} n^{n-2}z^n/n!$. The other four generating functions begin as follows:

$$\begin{aligned}
V(z) &= \tfrac{1}{2}z + \tfrac{3}{4}z^2 + \tfrac{17}{12}z^3 + \tfrac{71}{24}z^4 + \tfrac{523}{80}z^5 + \tfrac{899}{60}z^6 + \cdots\,;\\
\widehat{V}(z) &= \tfrac{1}{6}z^3 + \tfrac{5}{8}z^4 + \tfrac{37}{20}z^5 + \tfrac{61}{12}z^6 + \cdots\,;\\
W(z) &= \tfrac{1}{8}z + \tfrac{7}{12}z^2 + \tfrac{101}{48}z^3 + \tfrac{83}{12}z^4 + \tfrac{12487}{576}z^5 + \tfrac{3961}{60}z^6 + \cdots\,;\\
\widehat{W}(z) &= \tfrac{1}{4}z^4 + \tfrac{41}{24}z^5 + \tfrac{95}{12}z^6 + \cdots\,.
\end{aligned}$$

All of these generating functions can be expressed succinctly in terms of the tree function

$$T(z) = \sum_{n\ge 1} n^{n-1}\frac{z^n}{n!} = z + z^2 + \tfrac{3}{2}z^3 + \cdots\,, \tag{3.1}$$

which generates *rooted* labeled trees and satisfies the functional relation

$$T(z) = ze^{T(z)} \tag{3.2}$$

due to Eisenstein [11]. Indeed, the relation

$$U(z) = T(z) - \tfrac{1}{2}T(z)^2 \tag{3.3}$$

is well known, as are the formulas

$$V(z) = \frac{1}{2}\ln\frac{1}{1-T(z)}\,, \tag{3.4}$$

$$\widehat{V}(z) = \frac{1}{2}\ln\frac{1}{1-T(z)} - \frac{1}{2}T(z) - \frac{1}{4}T(z)^2\,; \tag{3.5}$$

see [14]. We can prove (3.4) and (3.5) by noting that the univariate generating function for connected unicyclic multigraphs whose cycle has length k is

$$\frac{T(z)^k}{2k}\,;$$

summing over $k \ge 1$ gives (3.4), and summing over $k \ge 3$ gives (3.5). (If $k = 1$, the cycle is a self-loop; hence the multigraph is essentially a rooted tree and the compensation factor is $\frac{1}{2}$. If $k = 2$, the cycle is a duplicate edge; hence the multigraph is essentially an unordered pair of rooted trees, and the compensation factor again is $\frac{1}{2}$. If $k \ge 3$, the unicyclic component is essentially a sequence of k rooted trees, divided by $2k$ to account for cyclic order and change of orientation.)

The generating function $\widehat{W}(z)$ was shown by G. N. Bagaev [1] to be

$$\widehat{W}(z) = \frac{T(z)^4\bigl(6 - T(z)\bigr)}{24\bigl(1 - T(z)\bigr)^3}\,. \tag{3.6}$$

Then E. M. Wright made a careful study of all the generating functions $\widehat{C}_k(z)$, which he called W_k, in a series of significant papers [41, 43, 44, 45]. We will show below that the bgf for bicyclic connected multigraphs is

$$W(z) = \frac{T(z)\bigl(3 + 2T(z)\bigr)}{24\bigl(1 - T(z)\bigr)^3}\,. \tag{3.7}$$

The coefficients of powers of $1/\bigl(1-T(z)\bigr)$ arise in numerous applications, so Knuth and Pittel [24] began to catalog some of their interesting properties. For each n the function $t_n(y)$ defined by

$$\frac{1}{\bigl(1 - T(z)\bigr)^y} = \sum_{n\ge 0} t_n(y)\,\frac{z^n}{n!} \tag{3.8}$$

is a polynomial of degree n in y, called the *tree polynomial* of order n. The coefficient of y^k in $t_n(y)$ is the number of mappings from an n-element set into itself having exactly k cycles. For fixed y and $n \to \infty$, we have [24, Lemma 2 and (3.16)]

$$t_n(y) = \frac{\sqrt{2\pi}\, n^{n-1/2+y/2}}{2^{y/2}\Gamma(y/2)} + O(n^{n-1+y/2})\,. \tag{3.9}$$

We can, for example, express the number of connected bicyclic graphs on n vertices in terms of the tree polynomial t_n, namely

$$\tfrac{5}{24}\, t_n(3) - \tfrac{19}{24}\, t_n(2) + \tfrac{13}{12}\, t_n(1) - \tfrac{7}{12}\, t_n(0) + \tfrac{1}{24}\, t_n(-1) + \tfrac{1}{24}\, t_n(-2)\,, \tag{3.10}$$

because (3.6) can be rewritten

$$\widehat{W}(z) = \frac{5}{24\bigl(1 - T(z)\bigr)^3} - \frac{19}{24\bigl(1 - T(z)\bigr)^2} + \frac{13}{12\bigl(1 - T(z)\bigr)} - \frac{7}{12} + \frac{1 - T(z)}{24} + \frac{\bigl(1 - T(z)\bigr)^2}{24}\,.$$

Equation (3.9) tells us that only the first term $\frac{5}{24}\, t_n(3)$ of (3.10) is asymptotically significant. Extensions of (3.9) appear in equations (19.13) and (19.14) below.

We can also express quantities like (3.10) in terms of Ramanujan's function [30]

$$\begin{aligned} Q(n) &= 1 + \frac{n-1}{n} + \frac{n-1}{n}\,\frac{n-2}{n} + \frac{n-1}{n}\,\frac{n-2}{n}\,\frac{n-3}{n} + \cdots \\ &= \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} - \frac{4}{135n} + O(n^{-3/2})\,, \end{aligned} \tag{3.11}$$

which Wright [41] called $1+h(n)/n^n$. For we have

$$t_n(1)=n^n\,;\qquad t_n(2)=n^n(1+Q(n))\,;\qquad t_n(y+2)=n\frac{t_n(y)}{y}+t_n(y+1)\,,\quad y\neq 0. \tag{3.12}$$

(See [24, equations (2.7), (3.14), and (1.9)].) Furthermore, we have

$$[z^n]\,V(z)=\tfrac{1}{2}n^{n-1}Q(n)\,; \tag{3.13}$$

this follows from a well-known formula of Rényi [31].

4. The cyclic components. For theoretical purposes it proves to be important to partition a multigraph into its *acyclic part*, consisting entirely of isolated vertices or trees, and its *cyclic part*, consisting entirely of components that each contain at least one cycle. The cyclic part can in turn be partitioned into the *unicyclic part*, consisting entirely of unicyclic components, and the *complex part*, consisting entirely of components that have more edges than vertices. A multigraph is called cyclic if it equals its cyclic part, complex if it equals its complex part. In this section and the next, we will study the generating functions for cyclic and complex multigraphs. The formulas turn out to be surprisingly simple, and they will be the key to much of what follows.

Let $F(w,z)$ be the bgf for all cyclic multigraphs, i.e., for all multigraphs whose acyclic part is empty. Formulas (2.5) and (2.11) tell us that

$$F(w,z)=e^{C_0(w,z)+C_1(w,z)+\cdots}=e^{C(w,z)-C_{-1}(w,z)}=G(w,z)\,e^{-U(wz)/w};$$

in other words,

$$G(w,z)=e^{U(wz)/w}\,F(w,z)\,. \tag{4.1}$$

Indeed, this makes sense, because $e^{U(wz)/w}$ is the bgf for all acyclic multigraphs. We will analyze $F=F(w,z)$ by studying a linear differential equation satisfied by $G=G(w,z)$, and seeing that a similar equation is satisfied by F.

Let ϑ_w be the differential operator $w\frac{\partial}{\partial w}$, and let ϑ_z be $z\frac{\partial}{\partial z}$. The operator ϑ_w corresponds to marking an edge of a multigraph, i.e., giving some edge a special label, because ϑ_w multiplies the coefficient of $w^m z^n$ by m. Similarly, ϑ_z corresponds to marking a vertex, because it multiplies the coefficient of $w^m z^n$ by n. (For a general discussion of marking, see [16, sections 2.2.24 and following].) We have

$$\vartheta_w\,G(w,z)\;=\;w\sum_{n\geq 0}\frac{n^2}{2}e^{wn^2/2}\frac{z^n}{n!}\;=\;\frac{w}{2}\,\vartheta_z^2\,G(w,z)\,;$$

hence G satisfies the differential equation

$$\frac{2}{w}\,\vartheta_w\,G\;=\;\vartheta_z^2\,G\,. \tag{4.2}$$

Again, this makes sense: The left side represents all multigraphs having a marked edge and an orientation assigned to that edge, and with the edge count decreased by 1. The right side represents all multigraphs with an ordered pair $\langle x, y\rangle$ of marked vertices. Orienting and discounting an edge is the same as marking two vertices.

We can also write (4.2) in the suggestive form

$$G(w,z) \;=\; e^z + \frac{1}{2}\int_0^w \vartheta_z^2\, G(w,z)\, dw\,, \tag{4.3}$$

using the boundary condition $G(0,z) = e^z$. (The generating function for all multigraphs with no edges is, of course, e^z.) The operator ϑ_z^2 corresponds to choosing an ordered pair $\langle x, y\rangle$, and the operator $\frac{1}{2}\int_0^w$ corresponds to disorienting that edge and blending it into the existing multigraph. (Notice that the English words "differentiation" and "integration" are remarkably apt synonyms for the combinatorial operations of marking and blending.)

Most of our work will involve ϑ_z instead of ϑ_w, so we shall often write simply ϑ without a subscript when we mean ϑ_z. The marking operator ϑ has a simple effect on the generating functions $U(z)$ for unrooted trees and $T(z)$ for rooted trees. Indeed, we have

$$\vartheta\, U(z) \;=\; T(z)\,, \tag{4.4}$$

because an unrooted tree with a marked vertex is the same as a rooted tree. Furthermore

$$\vartheta\, T(z) \;=\; \sum_{k\ge 1} T(z)^k \;=\; \frac{T(z)}{1-T(z)}\,, \tag{4.5}$$

because a rooted tree with a marked vertex is combinatorially equivalent to an ordered sequence $\langle T_1, T_2, \ldots, T_k\rangle$ of rooted trees, for some $k \ge 1$. The sequence represents a path of length k from the marked vertex to the root, with rooted subtrees sprouting from each point on that path.

Now let $U = U(wz)/w$ be the function $C_{-1}(w,z)$ that appears in (4.1), and let $T = T(wz) = C_0(w,z)$. We have

$$\begin{aligned}
\vartheta_z U &= z\frac{\partial}{\partial z}\frac{U(wz)}{w} = z\frac{wU'(wz)}{w} = z\frac{T(wz)}{wz} = \frac{T}{w}\,;\\
\vartheta_w U &= w\frac{\partial}{\partial w}\frac{U(wz)}{w} = w\left(\frac{zU'(wz)}{w} - \frac{U(wz)}{w^2}\right) = \frac{T-U}{w} = \frac{T^2}{2w}\,.
\end{aligned}$$

Thus

$$\frac{2}{w}\,\vartheta_w\, U \;=\; (\vartheta_z\, U)^2\,. \tag{4.6}$$

In words: "Orienting and discounting an edge of an unrooted tree is equivalent to constructing an ordered pair of rooted trees."

We are now ready to convert (4.2) into a differential equation satisfied by $F = F(w, z)$:

$$\begin{aligned}
&\vartheta_w G = \vartheta_w(e^U F) = (\vartheta_w e^U)F + e^U(\vartheta_w F) = e^U\bigl((\vartheta_w U)F + \vartheta_w F\bigr)\,;\\
&\vartheta_z G = \vartheta_z(e^U F) = e^U\bigl((\vartheta_z U)F + \vartheta_z F\bigr)\,;\\
&\vartheta_z^2 G = e^U\bigl((\vartheta_z^2 U)F + (\vartheta_z U)^2 F + 2(\vartheta_z U)(\vartheta_z F) + \vartheta_z^2 F\bigr)\,.
\end{aligned}$$

Therefore, using (4.6), we have

$$\frac{2}{w}\vartheta_w F \;=\; (\vartheta_z^2 U)F \;+\; 2(\vartheta_z U)(\vartheta_z F) \;+\; \vartheta_z^2 F\,. \tag{4.7}$$

And like our other formulas, this one makes combinatorial sense as well as algebraic sense: The left side tells us that the right side should yield all ways that the cyclic part of a multigraph can grow, since $\frac{2}{w}\vartheta_w F$ is the number of ways it can go backward one step. The first term on the right corresponds to marking two vertices of an unrooted tree (in the acyclic part of the multigraph); joining them will produce a unicyclic component, thereby increasing the number of components in F. The middle term corresponds to marking a vertex in some tree of the acyclic part and another vertex in the cyclic part; joining them will add new vertices to one of F's existing components. The remaining term corresponds to marking two vertices in the cyclic part. If such marked vertices belong to the same component, say a component of excess r, a new edge between them will change the excess of the component to $r+1$. Otherwise, the marked vertices belong to different components, having respective excesses r and s, possibly with $r = s$; joining them will merge the components into a new component of excess $r + s + 1$.

Similarly, we can proceed to study the bgf $E(w, z)$ for the complex part of a multigraph, the part whose components all have positive excess. (The letter E stands for excess.) We have

$$F(w, z) = e^{V(wz)}\, E(w, z)\,, \tag{4.8}$$

where $V = V(wz)$ generates unicyclic components. It is easy to verify the identity

$$\frac{2}{w}\vartheta_w V \;=\; \vartheta_z^2 U \;+\; 2(\vartheta_z U)(\vartheta_z V)\,, \tag{4.9}$$

which corresponds to a combinatorially evident fact. Indeed,

$$\vartheta_z^2 U = \frac{1}{w}\frac{T}{1-T}\,; \qquad\qquad \vartheta_w V = \vartheta_z V = \frac{T}{2(1-T)^2}\,. \tag{4.10}$$

Therefore we find

$$\frac{2}{w}\vartheta_w E \;=\; (\vartheta_z^2 V)E + (\vartheta_z V)^2 E + 2(\vartheta_z U)(\vartheta_z E) + 2(\vartheta_z V)(\vartheta_z E) + \vartheta_z^2 E\,. \tag{4.11}$$

5. Enumerating complex multigraphs. To solve the differential equation (4.11), we can first write it in the form

$$\frac{1}{w}(\vartheta_w - T\vartheta_z)E \;=\; \frac{1}{2}\,e^{-V}\,\vartheta_z^2\,e^V\,E\,. \tag{5.1}$$

Now we partition $E = E(w,z)$ into terms of equal excess, as we did for $C(w,z)$ in (2.11):

$$E(w,z) = \sum_r E_r(w,z) = \sum_r w^r E_r(wz)\,. \tag{5.2}$$

The univariate generating function $E_r(z)$ represents all complex multigraphs having exactly r more edges than vertices; in particular, $E_0(z) = 1$, since only the empty multigraph is "complex" and has excess 0. Differentiation yields

$$\vartheta_w E(w,z) = \sum_r \bigl(rw^r E_r(wz) + w^r(\vartheta E_r)(wz)\bigr)\,,$$

$$\vartheta_z E(w,z) = \sum_r w^r(\vartheta E_r)(wz)\,,$$

where $(\vartheta E_r)(wz)$ here means $\vartheta_z E_r(z)$ with the argument z subsequently replaced by wz, namely $wzE_r'(wz)$. Therefore, if we equate the coefficients of w^{r-1} on both sides of (5.1) and set $w = 1$, we obtain a differential recurrence for the univariate generating functions $E_r = E_r(z)$:

$$(r + \vartheta - T\vartheta)E_r \;=\; \tfrac{1}{2}\,e^{-V}\,\vartheta^2\,e^V\,E_{r-1}\,. \tag{5.3}$$

It is convenient to introduce a new variable

$$\zeta = \frac{T(z)}{1 - T(z)} \tag{5.4}$$

and to express E_r in terms of ζ instead of z. Note that

$$1 + \zeta = \frac{1}{1 - T(z)}\,; \qquad T(z) = \frac{\zeta}{1+\zeta}\,; \qquad z = \frac{\zeta}{1+\zeta}\exp\left(\frac{-\zeta}{1+\zeta}\right)\,. \tag{5.5}$$

Equation (5.3) now takes the form

$$\bigl(r + (1+\zeta)^{-1}\vartheta\bigr)E_r = \tfrac{1}{2}(1+\zeta)^{-1/2}\vartheta^2(1+\zeta)^{1/2}E_{r-1}\,, \tag{5.6}$$

since $e^V = 1/\bigl(1 - T(z)\bigr)^{1/2} = (1+\zeta)^{1/2}$ by (3.4). We will see later that the variable ζ, which represents an ordered sequence of one or more rooted trees, has important significance in the study of graphs and multigraphs.

In the ζ world, with ϑ still denoting $z\frac{d}{dz}$, we have the operator equation

$$\vartheta \cdot f(\zeta) = f'(\zeta)\zeta(1+\zeta)^2 + f(\zeta)\vartheta\,, \tag{5.7}$$

because

$$z\frac{d\zeta}{dz} = \frac{T(z)}{1-T(z)} zT'(z)\left(\frac{1}{T(z)} + \frac{1}{1-T(z)}\right) = \frac{T}{(1-T)^3} = \zeta(1+\zeta)^2\,.$$

Equation (5.7) allows us to commute ϑ with functions of ζ. For example, we find

$$\begin{aligned}(1+\zeta)^{-1/2}\vartheta(1+\zeta)^{1/2} &= (1+\zeta)^{-1/2}\bigl(\tfrac12(1+\zeta)^{-1/2}\zeta(1+\zeta)^2 + (1+\zeta)^{1/2}\vartheta\bigr)\\ &= \tfrac12\zeta(1+\zeta) + \vartheta\,;\end{aligned}$$

hence (5.6) can be rewritten

$$\bigl(r + (1+\zeta)^{-1}\vartheta\bigr)E_r = \tfrac12\bigl(\tfrac12\zeta(1+\zeta)+\vartheta\bigr)^2 E_{r-1}\,. \tag{5.8}$$

To simplify the equation even further, we seek a function $f_r(\zeta)$ such that

$$\vartheta \cdot f_r(\zeta) = (1+\zeta)f_r(\zeta)\bigl(r + (1+\zeta)^{-1}\vartheta\bigr)\,;$$

then the differential equation (5.8) will become

$$\vartheta\bigl(f_r(\zeta)E_r\bigr) \;=\; \tfrac12(1+\zeta)f_r(\zeta)\bigl(\tfrac12\zeta(1+\zeta)+\vartheta\bigr)^2 E_{r-1}\,,$$

which can be solved by integration. According to (5.7), the desired factor $f_r(\zeta)$ is a solution to

$$\frac{f_r'(\zeta)}{f_r(\zeta)} = \frac{r}{\zeta(1+\zeta)} = \frac{r}{\zeta} - \frac{r}{1+\zeta}\,;$$

so we let $f_r(\zeta) = \zeta^r(1+\zeta)^{-r}$, which incidentally equals $T(z)^r$. We have derived the equation

$$\vartheta\left(\frac{\zeta^r\,E_r}{(1+\zeta)^r}\right) = \frac12\,\frac{\zeta^r}{(1+\zeta)^{r-1}}\left(\frac12\zeta(1+\zeta)+\vartheta\right)^2 E_{r-1}\,. \tag{5.9}$$

This differential equation determines E_r uniquely when $r > 0$, given E_{r-1}, since ζ^r vanishes when $z = 0$.

Now all the preliminary groundwork has been laid, and we are ready to calculate E_r. We know that $E_0 = 1$. A bit of experimentation soon reveals a fairly simple pattern: We can prove by induction on r that the solution to (5.9) has the form

$$E_r(z) = \sum_{d=0}^{2r} e_{rd}(1+\zeta)^r\zeta^{2r-d} = \sum_{d=0}^{2r}\frac{e_{rd}T(z)^{2r-d}}{\bigl(1-T(z)\bigr)^{3r-d}}\,, \tag{5.10}$$

where the coefficients e_{rd} are rational numbers, and where $e_{r(2r)} = 0$ for $r > 0$. Let $e_{rd} = 0$ when $d < 0$ or $d > 2r$. Assuming that (5.10) holds for some r, we use (5.7) and (5.8) to compute

$$\begin{aligned}
A_r &= \bigl(\tfrac{1}{2}\zeta(1+\zeta)+\vartheta\bigr)E_r \\
&= \sum_{d=0}^{2r} e_{rd}(1+\zeta)^r\zeta^{2r-d}\zeta(1+\zeta)^2\left(\frac{\frac{1}{2}}{1+\zeta}+\frac{r}{1+\zeta}+\frac{2r-d}{\zeta}\right) \\
&= \sum_{d=0}^{2r+1} a_{rd}(1+\zeta)^{r+1}\zeta^{2r+1-d}\,, \\
a_{rd} &= (3r+\tfrac{1}{2}-d)e_{rd}+(2r+1-d)e_{r(d-1)}\,;
\end{aligned} \tag{5.11}$$

$$\begin{aligned}
B_r &= \bigl(\tfrac{1}{2}\zeta(1+\zeta)+\vartheta\bigr)A_r \\
&= \sum_{d=0}^{2r+1} a_{rd}(1+\zeta)^{r+1}\zeta^{2r+1-d}\zeta(1+\zeta)^2\left(\frac{\frac{1}{2}}{1+\zeta}+\frac{r+1}{1+\zeta}+\frac{2r+1-d}{\zeta}\right) \\
&= \sum_{d=0}^{2r+2} b_{rd}(1+\zeta)^{r+2}\zeta^{2r+2-d}\,, \\
b_{rd} &= (3r+\tfrac{5}{2}-d)a_{rd}+(2r+2-d)a_{r(d-1)}\,.
\end{aligned} \tag{5.12}$$

Moreover, the left side of Equation (5.9) is a polynomial,

$$\vartheta\bigl(\zeta^r(1+\zeta)^{-r}E_r\bigr) = \vartheta\sum_{d=0}^{2r} e_{rd}\zeta^{3r-d} = \sum_{d=0}^{2r}(3r-d)e_{rd}(1+\zeta)^2\zeta^{3r-d}\,.$$

The corresponding polynomial on the right-hand side is

$$\tfrac{1}{2}\zeta^r(1+\zeta)^{1-r}\bigl(\tfrac{1}{2}\zeta(1+\zeta)+\vartheta\bigr)^2E_{r-1} = \tfrac{1}{2}\textstyle\sum_d b_{(r-1)d}(1+\zeta)^2\zeta^{3r-d}\,;$$

therefore we can complete the induction proof by setting

$$e_{rd} \;=\; \frac{b_{(r-1)d}}{6r-2d}\,, \qquad\qquad 0 \le d \le 2r. \tag{5.13}$$

It is easy to check that $a_{r(2r+1)} = 0$ and $b_{r(2r+2)} = 0$, hence $e_{r(2r)} = 0$ when $r > 0$.

In particular, $a_{00} = \frac{1}{2}$, $b_{00} = \frac{5}{4}$, $b_{01} = \frac{1}{2}$, and we obtain

$$E_1(z) = (1+\zeta)\bigl(\tfrac{5}{24}\zeta^2+\tfrac{1}{8}\zeta\bigr) = \left(\frac{5}{24}\frac{T(z)^2}{\bigl(1-T(z)\bigr)^3}+\frac{1}{8}\frac{T(z)}{\bigl(1-T(z)\bigr)^2}\right). \tag{5.14}$$

A complex multigraph of excess 1 must consist of a single bicyclic component, so $E_1(z)$ is the function we called $W(z)$ in (3.7). If our only goal had been to compute $W(z)$, we could of course have gotten this result easily and directly. The more elaborate machinery above has been developed so that the generating function $E_r(z)$ can readily be computed and analyzed for larger values of r.

6. Enumerating complex graphs. For graphs instead of multigraphs, the calculations are more intricate, but it is instructive to look at them and see how they differ. As in (4.1) and (4.8), we separate off the cyclic and complex parts of the bgf by writing

$$\widehat{G}(w,z) \;=\; e^{U(wz)/w}\,\widehat{F}(w,z)\,; \qquad\qquad \widehat{F}(w,z) \;=\; e^{\widehat{V}(wz)}\widehat{E}(w,z). \tag{6.1}$$

Adding a new edge to a graph means that we want to mark an unordered pair of *distinct* vertices, and the operator corresponding to this is $\frac{1}{2}(\vartheta_z^2 - \vartheta_z)$. We must also avoid duplicating an edge that's already present, so we must also subtract ϑ_w. Therefore the differential equation satisfied by $\widehat{G}$ is not (4.2) but

$$\frac{1}{w}\vartheta_w \widehat{G} = \left(\frac{\vartheta_z^2 - \vartheta_z}{2} - \vartheta_w\right)\widehat{G}\,; \tag{6.2}$$

and the integral equation corresponding to (4.3) is

$$\widehat{G}(w,z) = e^z + \int_0^w \left(\frac{\vartheta_z^2 - \vartheta_z}{2} - \vartheta_w\right)\widehat{G}(w,z)\,dw\,. \tag{6.3}$$

A computation similar to our derivation of (4.7) now leads to a differential equation defining $\widehat{F}$:

$$\frac{1}{w}\vartheta_w \widehat{F} \;=\; \left(\left(\frac{\vartheta_z^2 - \vartheta_z}{2} - \vartheta_w\right)U\right)\widehat{F} \,+\, (\vartheta_z U)(\vartheta_z \widehat{F}) \,+\, \left(\frac{\vartheta_z^2 - \vartheta_z}{2} - \vartheta_w\right)\widehat{F}\,. \tag{6.4}$$

The analog of (5.1) turns out to be

$$\frac{1}{w}\big(\vartheta_w \widehat{E} \,-\, T\vartheta_z \widehat{E}\big) \;=\; e^{-\widehat{V}}\left(\frac{\vartheta_z^2 - \vartheta_z}{2} - \vartheta_w\right)e^{\widehat{V}}\,\widehat{E}\,; \tag{6.5}$$

converting to univariate generating functions $\widehat{E}_r(w,z) = w^r\,\widehat{E}_r(wz)$ yields

$$\big(r + \vartheta - T\vartheta\big)\,\widehat{E}_r \;=\; e^{-\widehat{V}}\left(1 - r + \frac{\vartheta^2 - 3\vartheta}{2}\right)e^{\widehat{V}}\,\widehat{E}_{r-1}\,. \tag{6.6}$$

Again we multiply by the integration factor $\zeta^r/(1+\zeta)^r$, but the differential equation turns out to be rather messy:

$$\vartheta\left(\left(\frac{\zeta}{1+\zeta}\right)^r \widehat{E}_r\right) = \left(\frac{\zeta^r}{(1+\zeta)^{r-1}}\right)\left(1-r+\frac{\zeta^4(10+14\zeta+5\zeta^2)}{8(1+\zeta)^2}+\frac{\zeta^3-3\zeta-3}{2(1+\zeta)}\vartheta+\frac{\vartheta^2}{2}\right)\widehat{E}_{r-1}\,. \tag{6.7}$$

At least it is linear, and it allows us to compute $\widehat{E}_r$ for small r. It turns out that the solution has the form

$$\widehat{E}_r = \sum_{d\ge 0} \hat{e}_{rd}\frac{\zeta^{5r-d}}{(1+\zeta)^{2r}} = \sum_{d\ge 0} \hat{e}_{rd}\frac{T(z)^{5r-d}}{(1-T(z))^{3r-d}}\,, \tag{6.8}$$

for appropriate coefficients $\hat{e}_{rd}$. We have, of course, $\hat{e}_{00} = 1$ and $\hat{e}_{0d} = 0$ for $d \neq 0$. When $r > 0$, the values of $\hat{e}_{rd}$ satisfy the following recurrence, equivalent to (6.7):

$$(3r - d)\hat{e}_{rd} + (6r - d + 1)\hat{e}_{r(d-1)} = \sum_{j=0}^{6} c_j(r-1, d)\hat{e}_{(r-1)(d-j)}\,, \tag{6.9}$$

where

$$\begin{aligned}
c_0(r,d) &= (6r - 2d + 5)(6r - 2d + 1)/8\,,\\
c_1(r,d) &= (132r^2 + (166 - 80d)r + 45 - 50d + 12d^2)/4\,,\\
c_2(r,d) &= (398r^2 + (584 - 220d)r + 205 - 160d + 30d^2)/4\,,\\
c_3(r,d) &= (316r^2 + (515 - 160d)r + 207 - 129d + 20d^2)/2\,,\\
c_4(r,d) &= (279r^2 + (484 - 130d)r + 208 - 112d + 15d^2)/2\,,\\
c_5(r,d) &= (13r - 3d + 10)(5r - d + 5)\,,\\
c_6(r,d) &= (25r^2 + (43 - 10d)r + 18 - 9d + d^2)/2\,.
\end{aligned} \tag{6.10}$$

It is not at all obvious that this recurrence has a solution. We can use it to compute $\hat{e}_{rd}$ for $d = 0, 1, \ldots, 3r - 1$, but then the value of $\hat{e}_{r(3r-1)}$ must satisfy a nontrivial equation when we set $d = 3r$. To get the values of $\hat{e}_{rd}$ when $d \geq 3r$, we can start by assuming that $\hat{e}_{rd} = 0$ for $d \geq 6r$ and work backward. We will prove later that the recurrence always does have a solution, and that the last nonzero coefficient for fixed r can be completely characterized by an almost unbelievable (but true) formula: If $\binom{s-2}{2} \leq r < \binom{s-1}{2}$, then

$$\hat{e}_{r(5r-s)} = \binom{\binom{s}{2}}{s+r}\frac{1}{s!}\,. \tag{6.11}$$

Moreover, $\hat{e}_{rd} = 0$ for all $d > 5r - s$. Here is a table of values for small r, in case the reader would like to check a computer program that is based on the formulas above:

$d =$	0	1	2	3	4	5	6	7	8	9	10
$\hat{e}_{0d} =$	1										
$\hat{e}_{1d} =$	$\frac{5}{24}$	$\frac{1}{4}$									
$\hat{e}_{2d} =$	$\frac{385}{1152}$	$\frac{175}{96}$	$\frac{133}{32}$	$\frac{79}{16}$	$\frac{49}{16}$	$\frac{5}{6}$	$\frac{1}{24}$				
$\hat{e}_{3d} =$	$\frac{85085}{82944}$	$\frac{5005}{512}$	$\frac{97097}{2304}$	$\frac{7777}{72}$	$\frac{43621}{240}$	$\frac{200561}{960}$	$\frac{950569}{5760}$	$\frac{14001}{160}$	$\frac{7021}{240}$	$\frac{773}{144}$	$\frac{3}{8}$

7. A surprising pattern. The numbers $\hat{e}_{rd}$ that characterize cyclic graphs of excess r do not appear to have any nice mathematical properties. But when we calculate the corresponding coefficients e_{rd} for multigraphs, as defined in (5.11)–(5.13), we run into patterns that cry out for explanation. For example, here is a table showing the values for small r:

$d =$	0	1	2	3	4	5	6	7	8	9
$e_{0d} =$	1									
$e_{1d} =$	$\frac{5}{24}$	$\frac{1}{8}$								
$e_{2d} =$	$\frac{385}{1152}$	$\frac{35}{64}$	$\frac{91}{384}$	$\frac{1}{48}$						
$e_{3d} =$	$\frac{85085}{82944}$	$\frac{25025}{9216}$	$\frac{23023}{9216}$	$\frac{2849}{3072}$	$\frac{19}{160}$	$\frac{1}{384}$				
$e_{4d} =$	$\frac{37182145}{7962624}$	$\frac{11316305}{663552}$	$\frac{3556553}{147456}$	$\frac{3658655}{221184}$	$\frac{1656083}{294912}$	$\frac{8723}{10240}$	$\frac{1969}{46080}$	$\frac{1}{3840}$		
$e_{5d} =$	$\frac{5391411025}{191102976}$	$\frac{929553625}{7077888}$	$\frac{7994161175}{31850496}$	$\frac{8068525465}{31850496}$	$\frac{341105765}{2359296}$	$\frac{327803333}{7077888}$	$\frac{1606891}{207360}$	$\frac{140569}{245760}$	$\frac{4043}{322560}$	$\frac{1}{46080}$

Anybody who has played with integers knows that the numerator of e_{32}, 23023, is equal to $7 \cdot 11 \cdot 13 \cdot 23$; moreover, the denominator is $9216 = 2^{10} \cdot 3^2$. Further experiments show that the factorization of, say, e_{55}, is $2^{-18} \cdot 3^{-3} \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 47 \cdot 151$. The occurrence of so many small prime factors cannot be a coincidence!

It is, in fact, easy to see the pattern in the numbers e_{r0}, which satisfy the recurrence

$$e_{r0} = \frac{(6r-1)(6r-5)}{24r} e_{(r-1)0} \tag{7.1}$$

according to rules (5.11)–(5.13). The numbers $\hat{e}_{r0}$ also satisfy the same recurrence, according to (6.9) and (6.10). Therefore we find

$$e_{r0} = \hat{e}_{r0} = \frac{(6r)!}{2^{5r} 3^{2r} (3r)!\,(2r)!}\,. \tag{7.2}$$

But the recurrence defining e_{rd} for $d > 0$ is much more complex, and we have no a priori reason to expect these numbers to have any mathematical virtues. The following theorem provides an algebraic explanation of what is going on.

Theorem 1. *The numbers e_{rd} defined in (5.10) can be expressed as*

$$e_{rd} = \frac{(6r-2d)!\,P_d(r)}{2^{5r}\,3^{2r-d}\,(3r-d)!\,(2r-d)!}\,, \tag{7.3}$$

where $P_d(r)$ is a polynomial of degree d defined by the formulas

$$P_d(r) = [z^d]\,F(z)^{2r-d}\,, \tag{7.4}$$

$$F(z) = 3! \sum_{n \ge 0} \frac{(4z)^n}{(n+3)!} = \frac{6}{(4z)^3}\left(e^{4z} - \frac{(4z)^2}{2} - 4z - 1\right). \tag{7.5}$$

Proof. By the duplication and triplication formulas for the Gamma function, expression (7.3) can also be written

$$e_{rd} = g_{rd}P_d(r), \qquad g_{rd} = \frac{3^r\,\Gamma(r+\frac{5}{6}-\frac{d}{3})\,\Gamma(r+\frac{1}{2}-\frac{d}{3})\,\Gamma(r+\frac{1}{6}-\frac{d}{3})}{2^{r+d}\,2\pi\,\Gamma(r+1-\frac{d}{2})\,\Gamma(r+\frac{1}{2}-\frac{d}{2})}\,. \tag{7.6}$$

Therefore recurrence equation (5.11) becomes

$$\begin{aligned} a_{rd} &= 3(r+\tfrac{1}{6}-\tfrac{d}{3})g_{rd}P_d(r) + 2(r+\tfrac{1}{2}-\tfrac{d}{2})g_{r(d-1)}P_{d-1}(r) \\ &= 3(r+\tfrac{1}{6}-\tfrac{d}{3})g_{rd}A_d(r)\,, \\ A_d(r) &= P_d(r) + \tfrac{4}{3}P_{d-1}(r)\,. \end{aligned} \tag{7.7}$$

Similarly, but without as much cancellation, (5.12) becomes

$$\begin{aligned} b_{rd} &= 3(r+\tfrac{5}{6}-\tfrac{d}{3})3(r+\tfrac{1}{6}-\tfrac{d}{3})g_{rd}A_d(r) + 2(r+1-\tfrac{d}{2})3(r+\tfrac{1}{2}-\tfrac{d}{3})g_{r(d-1)}A_{d-1}(r) \\ &= \tfrac{9}{2}g_{r(d-1)}B_d(r)\,, \\ B_d(r) &= (r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{2})A_d(r) + \tfrac{4}{3}(r+1-\tfrac{d}{2})(r+\tfrac{1}{2}-\tfrac{d}{3})A_{d-1}(r)\,. \end{aligned} \tag{7.8}$$

Relation (5.13) becomes

$$(3r+3-d)g_{(r+1)d}P_d(r+1) = \frac{9}{4}\frac{(r+1-\frac{d}{3})(r+\frac{5}{6}-\frac{d}{3})(r+\frac{1}{2}-\frac{d}{3})}{(r+1-\frac{d}{2})}g_{r(d-1)}P_d(r+1) = \frac{1}{2}b_{rd}\,;$$

hence the original recurrence takes the following form:

$$(r+1-\tfrac{d}{3})(r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{3})P_d(r+1) = (r+1-\tfrac{d}{2})B_d(r)\,. \tag{7.9}$$

The boundary conditions are

$$P_d(r) = 0 \quad \text{for } d < 0; \qquad P_0(r) = 1; \qquad P_{2d}(d) = 0 \quad \text{for } d > 0. \tag{7.10}$$

It is by no means obvious that a polynomial $P_d(r)$ will satisfy (7.7), (7.8), and (7.9). The key observation that makes everything work is that a solution to the simpler recurrence

$$(r+\tfrac{1}{2}-\tfrac{d}{3})P_d(r+\tfrac{1}{2}) = (r+\tfrac{1}{2}-\tfrac{d}{2})A_d(r) \tag{7.11}$$

suffices to solve the more complex one. This new recurrence is sort of a "half step" between solutions of (7.7), (7.8), and (7.9); it tells us about multigraphs whose excess is an integer plus $\frac{1}{2}$, whatever that may mean.

A solution to (7.11) in the extended domain implies a solution to (7.9). For we will then have

$$(r+1-\tfrac{d}{3})(r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{3})P_d(r+1) = (r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{3})(r+1-\tfrac{d}{2})A_d(r+\tfrac{1}{2})$$

and

$$\begin{aligned}(r+1-\tfrac{d}{2})B_d(r) &= (r+1-\tfrac{d}{2})(r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{2})A_d(r) + \tfrac{4}{3}(r+1-\tfrac{d}{2})^2(r+\tfrac{1}{2}-\tfrac{d}{3})A_{d-1}(r)\\ &= (r+1-\tfrac{d}{2})(r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{3})P_d(r+\tfrac{1}{2})\\ &\qquad + \tfrac{4}{3}(r+1-\tfrac{d}{2})(r+\tfrac{5}{6}-\tfrac{d}{3})(r+\tfrac{1}{2}-\tfrac{d}{3})P_{d-1}(r+\tfrac{1}{2})\,.\end{aligned}$$

Moreover, $P_d(\frac{d}{2}) = 0$ when $d > 0$.

We can solve the simultaneous recurrences (7.7) and (7.11) by constructing solutions to (7.7) that have the desired form (7.4), namely

$$P_d(r) = [z^d]\,F(z)^{2r-d}\,, \qquad A_d(r) = [z^d]\,F(z)^{2r-d}\bigl(1+\tfrac{4}{3}\,z\,F(z)\bigr)\,,$$

and noting that the function $F(z)$ of (7.5) satisfies

$$\vartheta F(z) = 4z\,F(z) + 3 - 3F(z)\,. \tag{7.12}$$

Thus we have

$$\begin{aligned} dP_d(r+\tfrac{1}{2}) &= [z^d]\,\vartheta\bigl(F(z)^{2r+1-d}\bigr)\\ &= [z^d]\,(2r+1-d)F(z)^{2r-d}\bigl(4zF(z)+3-3F(z)\bigr)\\ &= (6r+3-3d)\bigl(A_d(r)-P_d(r+\tfrac{1}{2})\bigr)\,,\end{aligned}$$

and (7.11) holds. $\square$

Incidentally, the theory of confluent hypergeometric functions provides us with alternative expressions for the function $F(z)$ in (7.5). We have, for example,

$$\begin{aligned} F(z) = F(1;\,4;\,4z) &= 3\int_0^1 e^{4zt}(1-t)^2\,dt\\ &= \frac{3e^{4z}}{64z^3}\,\gamma(3,4z) = 3e^{4z}\left(\frac{1}{3\cdot 0!} - \frac{4z}{4\cdot 1!} + \frac{4^2z^2}{5\cdot 2!} - \frac{4^3z^3}{6\cdot 3!} + \cdots\right)\,. \end{aligned} \tag{7.13}$$

The general theory of [23] also allows us to write

$$P_d(r) = \frac{2r-d}{2r}\,[z^d]\,G(z)^{2r}\,, \tag{7.14}$$

where $G(z) = 1 + z - \frac{1}{5}z^2 + \frac{2}{15}z^3 - \frac{19}{175}z^4 + \frac{2}{21}z^5 - \frac{2018}{23625}z^6 + \cdots$ is defined implicitly by the relation

$$G\bigl(z\,F(z)\bigr) = F(z)\,. \tag{7.15}$$

Corollary. *For fixed $d \geq 0$ we have*

$$e_{rd} = \frac{3^r}{2^r}\frac{(r+d-1)!}{2\pi\, d!}\bigl(1 + O(r^{-1})\bigr) \tag{7.16}$$

as $r \to \infty$. Moreover, e_{rd} is a rational number whose numerator has at most

$$d + O\bigl(d(\log d)^2/\log r\bigr) \tag{7.17}$$

prime factors greater than $6r$, and whose denominator has no prime factors greater than $3r$.

Proof. The obvious bounds

$$\binom{2r-d}{d} = [z^d]\,(1+z)^{2r-d} \leq [z^d]\,F(z)^{2r-d}$$

$$\leq [z^d]\left(\frac{1}{1-z}\right)^{2r-d} = \binom{2r-1}{d} \tag{7.18}$$

tell us that $P_d(r) = (2r)^d/d! + O(r^{d-1})$. Formula (7.16) now follows from (7.3) and Stirling's approximation. (We will derive a more precise estimate, suitable when d varies with r, in section 23 below, Lemma 8.)

All prime factors greater than $6r$ must appear as prime factors of $P_d(r)$. We will prove the upper bound (7.17) by showing that $m_d P_d(r)$ is an integer, where

$$m_d = 5^{\lfloor d/2\rfloor}6^{\lfloor d/3\rfloor}7^{\lfloor d/4\rfloor}\ldots = \prod_{k\geq 2}(k+3)^{\lfloor d/k\rfloor}\,. \tag{7.19}$$

It will follow that the denominator of $P_d(r)$ contains no prime factors greater than $2r+1$, and that if the numerator contains k prime factors greater than $6r$, we have $(6r)^k < m_d P_d(r) \leq m_d\binom{2r-1}{d} < m_d(2r)^d$; i.e., $k\log 6r < d\log 2r + \log m_d = d\log 2r + O\bigl(d(\log d)^2\bigr)$.

The coefficient of z^d in any power of $F(z)$ is a sum of terms $f_1^{k_1}f_2^{k_2}f_3^{k_3}\ldots$, where $f_j = [z^j]\,F(z) = \frac{4}{5}\frac{4}{6}\cdots\frac{4}{(j+3)}$ and $k_1 + 2k_2 + 3k_3 + \cdots = d$. Thus, for example, the factor 7 occurs in the denominator of $f_1^{k_1}f_2^{k_2}f_3^{k_3}\ldots$ exactly $k_4 + k_5 + \cdots \leq d/4$ times. It follows that the denominator of P_d is a divisor of m_d. ∎

The estimate (7.17) can be sharpened for small d, because $P_d(r)$ always has $(2r-d)$ as a factor when $d > 0$. For example,

$$P_1(r) = 2r-1, \qquad P_2(r) = \frac{(r-1)(10r-7)}{5}, \qquad P_3(r) = \frac{(2r-3)(10r^2-21r+10)}{15}.$$

There are no prime factors $> 6r$ when $d \le 1$, and there is at most one when $d \le 3$.

Instead of writing

$$E_r(z) = \sum_{d=0}^{2r} e_{rd} \frac{T(z)^{2r-d}}{\bigl(1-T(z)\bigr)^{3r-d}}\,,$$

it is sometimes convenient to use coefficients e'_{rd} such that

$$E_r(z) = \sum_{d=0}^{2r} \frac{e'_{rd}}{\bigl(1-T(z)\bigr)^{3r-d}}\,. \tag{7.20}$$

The following table shows that the numbers e'_{rd} tend to alternate in sign:

$d =$	0	1	2	3	4	5	6	7	8
$e'_{0d} =$	1								
$e'_{1d} =$	$\frac{5}{24}$	$-\frac{7}{24}$	$\frac{1}{12}$						
$e'_{2d} =$	$\frac{385}{1152}$	$-\frac{455}{576}$	$\frac{77}{128}$	$-\frac{43}{288}$	$\frac{1}{288}$				
$e'_{3d} =$	$\frac{85085}{82944}$	$-\frac{95095}{27648}$	$\frac{119119}{27648}$	$-\frac{201355}{82944}$	$\frac{38623}{69120}$	$-\frac{803}{34560}$	$-\frac{139}{51840}$		
$e'_{4d} =$	$\frac{37182145}{7962624}$	$-\frac{40415375}{1990656}$	$\frac{141292151}{3981312}$	$-\frac{62775713}{1990656}$	$\frac{116866321}{7962624}$	$-\frac{15867137}{4976640}$	$\frac{850003}{4976640}$	$\frac{25129}{1244160}$	$-\frac{571}{2488320}$

Again, patterns lurk beneath the surface, and there is a prevalence of small prime factors; for example, $-e'_{55} = \frac{7541601353}{63700992} = 2^{-18} \cdot 3^{-5} \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 31 \cdot 229$. We can in fact prove the existence of a pattern similar to that of the original coefficients e_{rd}:

Corollary. *The numbers e'_{rd} defined in (7.20) can be expressed as*

$$e'_{rd} = \frac{(6r-2d)!\,Q_d(r)}{2^{5r}\,3^{2r-d}\,(3r-d)!\,(2r-d)!}\,, \tag{7.21}$$

where $Q_d(r)$ is a polynomial of degree d for which $Q_d\bigl(\frac{d}{3}-\frac{1}{2}\bigr) = 0$ when $d > 0$.

Proof. By definition, we have

$$e'_{rd} = \sum_{k=0}^{d} \binom{2r-k}{d-k} (-1)^{d-k} e_{rk}\,, \tag{7.22}$$

because the quantity $T^{2r-k} = \bigl(1-(1-T)\bigr)^{2r-k}$ contributes $\binom{2r-k}{d-k}(-1)^{d-k}$ to the coefficient of $(1-T)^{d-3r}$. Now if we plug in equations (7.3) and (7.21), we find that

$$\begin{aligned} Q_d(r) &= \sum_{k=0}^{d} \frac{(-1)^{d-k} P_k(r)}{3^{d-k}(d-k)!} \frac{(6r-2k)!}{(6r-2d)!} \frac{(3r-d)!}{(3r-k)!} \\ &= \sum_{k=0}^{d} \Bigl(-\frac{4}{3}\Bigr)^{d-k} \binom{3r-k-\frac{1}{2}}{d-k} P_k(r) = \sum_{k=0}^{d} \Bigl(-\frac{4}{3}\Bigr)^{d-k} \binom{3r-k+\frac{1}{2}}{d-k} A_k(r)\,, \end{aligned} \tag{7.23}$$

clearly a polynomial in r of degree $\le d$. In fact, the leading term is

$$\sum_k \left(-\frac{4}{3}\right)^{d-k} \frac{(3r)^{d-k}}{(d-k)!}\frac{(2r)^k}{k!} = \frac{(-2)^d r^d}{d!}\,,$$

so the degree is exactly d. If we set $r = \frac{d}{3} - \frac{1}{2}$, the sum reduces to $A_d\bigl(\frac{d}{3} - \frac{1}{2}\bigr)$, which we know is zero for $d > 0$ by (7.11). $\square$

It is interesting to try to compute the coefficients e'_{rd} directly, by proceeding as we did in section 5 but using the variable $\xi = 1+\zeta = \bigl(1-T(z)\bigr)^{-1}$ in place of ζ. The calculations are essentially the same, even slightly simpler, until we get to the analog of equation (5.13); the recurrences that replace (5.11)–(5.13) are

$$a'_{rd} = (3r+\tfrac{1}{2}-d)e'_{rd} - (3r+\tfrac{3}{2} - d)e'_{r(d-1)}\,; \tag{7.24}$$

$$(3r-d)e'_{rd} - (2r+1-d)e'_{r(d-1)} = \tfrac{1}{2}\bigl((3r-\tfrac{1}{2}-d)a'_{(r-1)d} - (3r+\tfrac{1}{2}-d)a'_{(r-1)(d-1)}\bigr)\,. \tag{7.25}$$

It appears to be quite difficult to derive (7.21) directly from these recurrences. The recurrence for $Q_d(r)$, corresponding to equation (7.9) for $P_d(r)$, turns out to be

$$\begin{aligned}(r - \tfrac{d}{3})(r-\tfrac{1}{2}-\tfrac{d}{3})Q_d(r) = (r - \tfrac{d}{2})&(r-\tfrac{1}{2}-\tfrac{d}{2})Q_d(r-1)\\ &+ \tfrac{4}{3}(r+\tfrac{1}{6}-\tfrac{d}{3})(r-\tfrac{1}{2}-\tfrac{d}{3})Q_{d-1}(r)\\ &- 4(r - \tfrac{d}{2})(r-\tfrac{1}{2}-\tfrac{d}{3})(r - \tfrac{d}{3})(r-\tfrac{1}{6}-\tfrac{d}{3})^{-1}Q_{d-1}(r-1)\\ &+ 4(r+\tfrac{1}{6}-\tfrac{d}{3})(r-\tfrac{1}{2}-\tfrac{d}{3})Q_{d-2}(r-1)\,,\end{aligned} \tag{7.26}$$

and we can proceed to solve it for $d = 1, 2, \ldots,$ if we first multiply both sides by the summation factor $\Gamma(r-\frac{d}{3})\Gamma(r-\frac{1}{2}-\frac{d}{3})\Gamma(r+1-\frac{d}{2})^{-1}\Gamma(r+\frac{1}{2}-\frac{d}{2})^{-1}$. The equation for $d > 0$ then takes the form

$$S_d(r) = S_d(r-1) + g_d(r) + g_d(r-\tfrac{1}{2})\,,$$

$$S_d(r) = \frac{\Gamma(r+1-\frac{d}{3})\,\Gamma(r+\frac{1}{2}-\frac{d}{3})}{\Gamma(r+1-\frac{d}{2})\,\Gamma(r+\frac{1}{2}-\frac{d}{2})}Q_d(r)\,,$$

$$g_d(r) = \frac{\Gamma(r+1-\frac{d}{3})\,\Gamma(r+\frac{1}{2}-\frac{d}{3})}{\Gamma(r+1-\frac{d}{2})\,\Gamma(r+\frac{1}{2}-\frac{d}{2})}f_d(r)\,,$$

where $f_d(r) = Q_d(r) - \frac{r-d/2}{r-d/3}Q_d(r-\frac{1}{2})$ is a polynomial of degree $d-1$. For example, $f_1(r) = -\frac{4}{3}$ and $f_2(r) = \frac{8}{3}r - \frac{4}{3}$. There is apparently no analog of the simple relation (7.11) that made everything work nicely in the theorem above.

A generating function for $Q_d(r)$, analogous to (7.4), can be found by analyzing (7.23) more carefully. Let $H(z)$ satisfy

$$H(z) = F\bigl(z\,H(z)^{-1/3}\bigr) = 1 + z + \tfrac{7}{15}\,z^2 + \tfrac{1}{15}\,z^3 + \cdots\,; \tag{7.27}$$

then the elementary theory in [23] proves that

$$\bigl(x - \tfrac{d}{3}\bigr)\,[z^d]\,H(z)^x = x\,[z^d]\,F(z)^{x-d/3}\,. \tag{7.28}$$

Hence, by (7.11) and (7.4),

$$A_d(r) = \frac{r+\frac{1}{2}-\frac{d}{3}}{r+\frac{1}{2}-\frac{d}{2}}\,[z^d]\,F(z)^{2r+1-d} = [z^d]\,H(z)^{2r+1-2d/3}\,. \tag{7.29}$$

And (7.23) can therefore be "summed":

$$\begin{aligned} Q_d(r) &= \sum_{k=0}^{d}\left(-\frac{4}{3}\right)^k \binom{3r-d+k+\frac{1}{2}}{k} A_{d-k}(r)\\ &= \sum_{k=0}^{d}\left(\frac{4}{3}\right)^k \binom{-3r+d-\frac{3}{2}}{k}\,[z^{d-k}]\,H(z)^{(-2/3)(-3r-3/2+d-k)}\\ &= [z^d]\,\left(\frac{4}{3}\,z + H(z)^{-2/3}\right)^{-3r-3/2+d}. \end{aligned} \tag{7.30}$$

In particular,

$$Q_0(r) = 1; \qquad Q_1(r) = -2(r+\tfrac{1}{6}); \qquad Q_2(r) = 2(r-\tfrac{1}{6})(r-\tfrac{1}{5}).$$

Although $Q_1(r) = -A_1(r)$ and $Q_2(r) = A_2(r)$, we have $Q_3(r) = -A_3(r) + \frac{16}{135}(r-\frac{1}{2})$.

8. Sparse components. We can readily compute the univariate generating functions $C_1(z)$, $C_2(z)$, $C_3(z)$, …, $C_r(z)$ for bicyclic, tricyclic, tetracyclic, …, $(r+1)$-cyclic components, now that we know the simple form of $E_1(z)$, $E_2(z)$, $E_3(z)$, …, $E_r(z)$, because of the fact that

$$\sum_{r\ge 0} w^r\,E_r \;=\; \exp\biggl(\sum_{r\ge 1} w^r\,C_r\biggr). \tag{8.1}$$

Differentiating this formula with respect to w and equating coefficients of w^{r-1} leads to the expression

$$r\,E_r \;=\; \sum_{k=1}^{r} k\,C_k\,E_{r-k}\,, \tag{8.2}$$

from which we may find C_r by calculating

$$C_r \;=\; E_r \;-\; \frac{1}{r}\sum_{k=1}^{r-1} k\,C_k\,E_{r-k}\,. \tag{8.3}$$

Since we know that $E_r = (1+\zeta)^r \sum_{d=0}^{2r-1} e_{rd}\zeta^{2r-d}$ for $r>0$, it follows by induction that C_r can be written in the same form,

$$C_r = (1+\zeta)^r \sum_{d=0}^{2r-1} c_{rd}\zeta^{2r-d}\,, \tag{8.4}$$

for appropriate coefficients c_{rd}. (The variable ζ stands for $T(z)/\bigl(1-T(z)\bigr)$, as in section 5.) Indeed, relation (8.3) tells us that we can compute c_{rd} by evaluating a double sum

$$c_{rd} = e_{rd} - \frac{1}{r}\sum_{k=1}^{r-1} k \sum_j c_{kj}e_{(r-k)(d-j)}\,; \tag{8.5}$$

the inner sum here is over the range $\max(0, d+1-2r+2k) \le j \le \min(d, 2k-1)$, which is always nonempty for $0<k<r$ except when $d = 2r-1$. We always have $c_{r(2r-1)} = e_{r(2r-1)} = 1/\bigl(2^{r+1}(r+1)!\bigr)$. Here is a table of the coefficients for for small r:

$d=$	0	1	2	3	4	5	6	7	8	9
$c_{1d}=$	$\frac{5}{24}$	$\frac{1}{8}$								
$c_{2d}=$	$\frac{5}{16}$	$\frac{25}{48}$	$\frac{11}{48}$	$\frac{1}{48}$						
$c_{3d}=$	$\frac{1105}{1152}$	$\frac{985}{384}$	$\frac{1373}{576}$	$\frac{515}{576}$	$\frac{223}{1920}$	$\frac{1}{384}$				
$c_{4d}=$	$\frac{565}{128}$	$\frac{12455}{768}$	$\frac{26581}{1152}$	$\frac{12227}{768}$	$\frac{2089}{384}$	$\frac{9583}{11520}$	$\frac{27}{640}$	$\frac{1}{3840}$		
$c_{5d}=$	$\frac{82825}{3072}$	$\frac{387005}{3072}$	$\frac{371195}{1536}$	$\frac{10154003}{41472}$	$\frac{121207}{864}$	$\frac{519883}{11520}$	$\frac{1573507}{207360}$	$\frac{2597}{4608}$	$\frac{803}{64512}$	$\frac{1}{46080}$

In applications, the leading coefficients c_{r0} of C_r are the most important, as are the leading coefficients e_{r0} of E_r, because these govern the dominant asymptotic behavior of $[z^n]\,C_r(z)$ and $[z^n]\,E_r(z)$. Therefore it is convenient to write

$$c_r = c_{r0}, \qquad\qquad e_r = e_{r0}\,. \tag{8.6}$$

We have seen in (7.2) that there is a simple way to express the numbers e_r in terms of factorials. The values c_r are then easily computed by using relation (8.3), but with c_r and e_r substituted respectively for C_r and E_r.

Asymptotically speaking, the values of c_{rd} and e_{rd} are equivalent when r is large.

Theorem 2. *For fixed $d \geq 0$ we have*

$$c_{rd} = e_{rd}\bigl(1 + O(r^{-1})\bigr) = \frac{3^r}{2^r}\frac{(r+d-1)!}{2\pi\, d!}\bigl(1 + O(r^{-1})\bigr) \tag{8.7}$$

as $r \to \infty$.

Proof. We know the asymptotic value of e_{rd} from (7.16). To complete the proof, we need only show that the double sum in (8.5) is $O_d(e_{rd}/r)$, where O_d implies a bound for fixed d as $r \to \infty$.

Since $c_{rd} \leq e_{rd}$, each term in the double sum is bounded above by an absolute constant (depending on d) times

$$\frac{3^r}{2^r}\frac{k}{r}\frac{(k+j-1)!}{j!}\frac{(r-k+d-j-1)!}{(d-j)!} = \frac{3^r}{2^r}\frac{k}{r}\frac{(r+d-2)!}{d!}\binom{d}{j}\bigg/\binom{r+d-2}{k+j-1}.$$

We have $\binom{r+d-2}{k+j-1} \geq r+d-2$ except when $k=1$ and $j=0$ or $k=r-1$ and $j=d$. Therefore all but one term is $O_d(e_{rd}/r^2)$, and the exceptional term is $O_d(e_{rd}/r)$. There are $O(rd)$ terms altogether, so the overall double sum is $O_d(e_{rd}/r)$. □

The simple form (8.4) of $C_r(z)$, the generating function for $(r+1)$-cyclic multigraphs, makes it possible for us to deduce a formula for the corresponding graph-based function $\widehat{C}_r(z)$, which turns out to be only about 50% more complicated. In fact, we will prove a result that applies to the generating functions for infinitely many models of random graphs, including both $G(w,z)$ and $\widehat{G}(w,z)$ as special cases.

Our starting point for this calculation is the formal power series relation

$$\widehat{G}(w,z) = G(\ln(1+w), z/\sqrt{1+w}\,). \tag{8.8}$$

which is an immediate consequence of (2.7) and (2.9). It follows that

$$\widehat{C}(w,z) = C(\ln(1+w), z/\sqrt{1+w}\,). \tag{8.9}$$

We can therefore obtain a near-polynomial formula for $\widehat{C}_r(z)$ as a special case of the following result.

Theorem 3. *If $f(w) = 1 + f_1 w + f_2 w^2 + \cdots$ and $g(w) = 1 + g_1 w + g_2 w^2 + \cdots$ are arbitrary formal power series with $f(0) = g(0) = 1$, and if*

$$\widetilde{C}(w,z) = C\left(wf(w), z\frac{g(w)}{f(w)}\right) = \sum_r w^r \widetilde{C}_r(wz)\,, \tag{8.10}$$

where C is the bgf (2.10) for connected multigraphs, then there exist coefficients $\tilde{c}_{rd}$ such that

$$\widetilde{C}_r(z) = \sum_{d=0}^{3r+2} \tilde{c}_{rd}\zeta^{3r+2-d}(1+\zeta)^{-2} = \sum_{d=0}^{3r+2} \tilde{c}_{rd}\frac{T(z)^{3r+2-d}}{\bigl(1-T(z)\bigr)^{3r-d}} \tag{8.11}$$

for all $r > 0$.

Proof. Consider Ramanujan's function $Q(n)$ of (3.11), which has the asymptotic value $\sqrt{\frac{\pi n}{2}} + O(1)$ as $n \to \infty$. Following Knuth [22], we shall say that a function $s(n)$ of the form $p(n) + q(n)Q(n)$ is a *semipolynomial* when p and q are polynomials. The *degree* of a semipolynomial is computed by assuming that $Q(n)$ is of degree $\frac{1}{2}$. For example, $3+2n+(1+n)Q(n)$ is a semipolynomial of degree $\frac{3}{2}$. More formally, if d is any nonnegative integer, the semipolynomial $p(n) + q(n)Q(n)$ has degree $\le \frac{1}{2}d$ if and only if p has degree $\le \frac{1}{2}d$ and q has degree $< \frac{1}{2}d$.

The formulas (3.12) of section 3, taken from [24], show that generating functions of the form $F(z) = \sum_{k=1}^{d} a_k \big/ \bigl(1 - T(z)\bigr)^k$ are precisely those whose coefficients satisfy

$$[z^n]\, F(z) = \frac{n^n s(n)}{n!}$$

where $s(n)$ is a semipolynomial of degree $\le \frac{1}{2}(d-1)$.

Consider now the expansion

$$\sum_r w^r f(w)^r C_r\bigl(zwg(w)\bigr) = \sum_r w^r\, \widetilde{C}_r(wz)$$

which follows from (8.10) and (2.11). We will study how each term on the left contributes to terms on the right. First, when $r = -1$ we have

$$\begin{aligned}
\frac{U\bigl(zwg(w)\bigr)}{w\, f(w)} &= \sum_{n\ge1} \frac{n^{n-2} z^n w^{n-1} (1 + g_1 w + \cdots)^n}{n!\,(1 + f_1 w + \cdots)} \\
&= \sum_{n\ge1} \frac{n^{n-2} z^n w^{n-1} \bigl(1 + np_0(n)w + np_1(n)w^2 + \cdots\bigr)}{n!\,(1 + f_1 w + \cdots)}
\end{aligned}$$

where each $p_l(n)$ is a polynomial of degree $\le l$. The effect is to make $\widetilde{C}_{-1}(z) = U(z)$, and to contribute a linear combination of $U(z)$, $T(z)$, and $\bigl(1-T(z)\bigr)^{-1}$, $\ldots$, $\bigl(1-T(z)\bigr)^{-2l+1}$ to $\widetilde{C}_l(z)$ for each $l \ge 0$. Next, when $r = 0$ we have

$$V\bigl(zwg(w)\bigr) = \frac{1}{2} \sum_{n\ge1} n^{n-1} Q(n) z^n w^n \bigl(1 + np_0(n)w + np_1(n)w^2 + \cdots\bigr)\,;$$

this contributes $V(z)$ to $\widetilde{C}_0(z)$ and a linear combination of $\bigl(1-T(z)\bigr)^{-1}, \ldots, \bigl(1-T(z)\bigr)^{-2l}$ to $\widetilde{C}_l(z)$ for each $l > 0$. Finally, when $r > 0$ we have, by (5.11),

$$w^r f(w)^r C_r\bigl(zwg(w)\bigr) = \sum_{n\ge0} \frac{n^n s(n)}{n!} z^n w^{n+r} \bigl(1 + np_0(n)w + np_1(n)w^2 + \cdots\bigr) f(w)^r\,,$$

where $s(n)$ is a semipolynomial of degree $\le \frac{3}{2}r - \frac{1}{2}$. This contributes a linear combination of $\bigl(1-T(z)\bigr)^{-1}, \ldots, \bigl(1-T(z)\bigr)^{-2l-r}$ to $\widetilde{C}_l(z)$ for each $l \ge r$. The proof of (8.11) is complete, because $U(z) = \frac{1}{2}\zeta(2+\zeta)/(1+\zeta)^2$ and $T(z) = \zeta/(1+\zeta)$. $\square$

Incidentally, our proof shows that the only contribution to the coefficient of the "leading term" $T(z)^{3l+2}/\bigl(1-T(z)\bigr)^{3l}$ of $\widetilde{C}_l(z)$ comes from $C_l(z)$ itself. Therefore $\widetilde{C}_r(z)$ and $C_r(z)$ have identical leading coefficients. In particular, $\hat{c}_{r0} = c_{r0} = c_r$. We will see below that this gives the same asymptotic characteristics to the limiting distribution of component types in the uniform and permutation models when $m \approx \frac{1}{2}n$.

Theorem 3 justifies our earlier assertion that the recurrence (6.9)–(6.10) for $\hat{e}_{rd}$ has a solution. The coefficients $\hat{c}_{rd}$ can be computed from those coefficients $\hat{e}_{rd}$ using the relation $\widehat{C}_r = \widehat{E}_r - \frac{1}{r}\sum_{k=1}^{r-1} k\widehat{C}_k\widehat{E}_{r-k}$; but that makes $\widehat{C}_r$ a polynomial of degree $5r$ with denominator $(1+\zeta)^{2r}$, so the numerator and denominator must be divided by $(1+\zeta)^{2r-2}$. A simpler recurrence for $\widehat{C}_r$ was found by Wright [41], who proved Theorem 3 in the special case $\widetilde{C}_r = \widehat{C}_r$ by a different method. Translated into the notation of the present paper, Wright's recurrence is

$$\vartheta\left(\frac{\zeta}{1+\zeta}\right)^r \widehat{C}_r = \frac{\zeta^r}{2(1+\zeta)^{r-1}}\left(\sum_{j=0}^{r-1}(\vartheta\widehat{C}_j)(\vartheta\widehat{C}_{r-1-j}) + \bigl(\vartheta^2 - 3\vartheta - 2(r-1)\bigr)\widehat{C}_{r-1}\right), \quad r > 0, \tag{8.12}$$

with $\vartheta\widehat{C}_0 = \frac{1}{2}\zeta^3(1+\zeta)^{-1}$. As we saw for the related sequence $\widehat{E}_r$ in section 6, it isn't obvious that this recurrence has a solution of the desired form

$$\widehat{C}_r(z) = \sum_{d=0}^{3r+2} \hat{c}_{rd}\zeta^{3r+2-d}(1+\zeta)^{-2} = \sum_{d=0}^{3r+2} \hat{c}_{rd}\frac{T(z)^{3r+2-d}}{\bigl(1-T(z)\bigr)^{3r-d}}, \tag{8.13}$$

when $r > 0$. Theorem 3 provides an algebraic proof, while Wright proved the existence by a combination of algebraic and combinatorial methods that we will consider in the next section. Here is a table of the first few values of the coefficients:

$d =$	0	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{c}_{1d} =$	$\frac{5}{24}$	$\frac{1}{4}$											
$\hat{c}_{2d} =$	$\frac{5}{16}$	$\frac{55}{48}$	$\frac{73}{48}$	$\frac{3}{4}$	$\frac{1}{24}$								
$\hat{c}_{3d} =$	$\frac{1105}{1152}$	$\frac{395}{72}$	$\frac{15131}{1152}$	$\frac{2399}{144}$	$\frac{8303}{720}$	$\frac{557}{144}$	$\frac{3}{8}$						
$\hat{c}_{4d} =$	$\frac{565}{128}$	$\frac{26165}{768}$	$\frac{133651}{1152}$	$\frac{523789}{2304}$	$\frac{80573}{288}$	$\frac{317611}{1440}$	$\frac{77773}{720}$	$\frac{89}{3}$	$\frac{839}{240}$	$\frac{1}{12}$			
$\hat{c}_{5d} =$	$\frac{82825}{3072}$	$\frac{67005}{256}$	$\frac{1770535}{1536}$	$\frac{31448897}{10368}$	$\frac{438258631}{82944}$	$\frac{1146749}{180}$	$\frac{86265}{16}$	$\frac{304411}{96}$	$\frac{25180997}{20160}$	$\frac{109627}{360}$	$\frac{781}{20}$	$\frac{439}{240}$	$\frac{1}{120}$

Notice that $\hat{c}_{rd} = 0$ for sufficiently large values of d; we do not have to go all the way up to $d = 3r+2$. In fact, we will see in the next section that the final nonzero coefficient is $\hat{c}_{r(3r+2-s)}$ when $\binom{s-2}{2} \le r < \binom{s-1}{2}$, and it has the value exhibited in (6.11).

The asymptotic value of the leading coefficients $\hat{c}_{r0} = c_{r0} = c_r$ has an interesting history. Wright [44] gave a complicated argument establishing that $\hat{c}_{r0}$ is asymptotically $\left(\frac{3}{2}\right)^r (r-1)!$ times a certain constant, for which he obtained the numerical value 0.159155. Stepanov [35] independently computed the numerical value '0,46...' for three times the constant; the approximation 0.48 would have been more accurate, but Stepanov was perhaps conjecturing that the true value would be $\frac{1}{3}+\frac{1}{\pi}\bigl(\sqrt{3}+\ln(2-\sqrt{3}\,)\bigr) \approx 0.46546$, which he announced at the same time in connection with another problem concerning the size of the largest component when the centroid is removed from a random tree. Wright's constant was identified as $1/2\pi$ by G. N. Bagaev and E. F. Dmitriev [2], who presented without proof a list of asymptotic expressions for the solution of several related enumeration problems. Lambert Meertens independently found a proof in 1986, but did not publish it; his approach was reported later in [3]. A detailed analysis was also carried out by V. A. Voblyĭ [38], who obtained a number of interesting auxiliary formulas. In particular, if we write $c(z) = c_1 z + c_2 z^2 + c_3 z^3 + \cdots$, Voblyĭ proved the formal power series relation

$$\vartheta c(z) = -\frac{1}{6} + \frac{1}{3z}\left(1 - \frac{I_{-2/3}(1/3z)}{I_{1/3}(1/3z)}\right). \tag{8.14}$$

In other words, he proved that the coefficients c_r show up in the asymptotic series

$$\frac{I_{-2/3}(1/3z)}{I_{1/3}(1/3z)} \sim 1 - \frac{z}{2} - 3c_1 z^2 - 6c_2 z^3 - 9c_3 z^3 - \cdots, \tag{8.15}$$

as $z \to 0$. This is interesting because the left-hand side can also be expressed as a continued fraction

$$2z + \cfrac{1}{8z + \cfrac{1}{14z + \cfrac{1}{20z + \cfrac{1}{26z + \cdots}}}}, \tag{8.16}$$

using the standard recurrence $zI_{\nu+1}(z) = zI_{\nu-1}(z) - 2\nu I_\nu(z)$ for the modified Bessel functions $I_\nu(z)$. In the course of his investigation, Voblyĭ noticed that the coefficients of $e^{c(z)}$ have a simple form, although he did not mention their combinatorial significance; these are the numbers we have called e_r. He gave the formulas

$$\frac{2^r}{3^r} e_r = (-1)^r (1/3, r) = \frac{\Gamma(r+5/6)\,\Gamma(r+1/6)}{2\pi r!}, \tag{8.17}$$

which are equivalent to (7.2). Here (ν, r) denotes Hankel's symbol,

$$(\nu, r) = \frac{1}{r!}\prod_{k=1}^{r}(\nu + k - \tfrac{1}{2})(\nu - k + \tfrac{1}{2}).$$

9. Structure of complex multigraphs. The generating functions E_r, C_r, $(1+\zeta)^{2r}\widehat{E}_r$, and $(1+\zeta)^2\widehat{C}_r$ are polynomials in ζ, and these polynomials have a combinatorial interpretation that provides considerable insight into what is happening as a graph or multigraph evolves. The inner structure in the case of $\widehat{C}_r$ was studied by Wright in his original paper [41]; we will see that his results for graphs become simpler when we consider the analogous results for multigraphs.

Let M be a cyclic multigraph of excess r, i.e., any multigraph with no acyclic components, having r more edges than vertices. We can "prune" M by repeatedly cutting off any vertex of degree 1 and the edge leading to that vertex; this eliminates as many edges as vertices, so the pruned multigraph $\overline{M}$ still has excess r. Each vertex of $\overline{M}$ has degree at least 2. Such multigraphs are called *smooth.*

Conversely, given any smooth multigraph $\overline{M}$, we obtain all multigraphs M that prune down to it by simply sprouting a tree from each vertex of $\overline{M}$ (i.e., identifying that vertex with the root of a rooted tree). Since $T(z)$ is the generating function for rooted trees, it follows that

$$F_r(z) = \overline{F}_r\bigl(T(z)\bigr)\,, \tag{9.1}$$

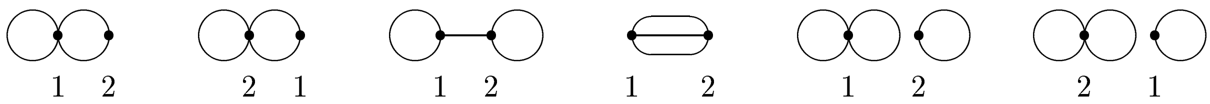
where $F_r(z)$ is the generating function for all cyclic multigraphs of excess r and $\overline{F}_r$ is the generating function for all smooth multigraphs of excess r. Thus, for example, we must have

$$\overline{F}_1(z) = \frac{1}{24}\,z(3+2z)/(1-z)^{7/2}\,, \tag{9.2}$$

because we know from (3.4), (4.8), (5.2), and (5.14) that

$$F_1(z) = e^{V(z)}E_1(z) = \frac{1}{24}\,T(z)\bigl(3+2T(z)\bigr)/\bigl(1-T(z)\bigr)^{7/2}.$$

The coefficient of z^n in $\overline{F}_1(z)$ is the sum of $\kappa(\overline{M})$ over all multigraphs $\overline{M}$ on n labeled vertices having $n+1$ edges and all vertices of degree 2 or more, divided by $n!$. For example, the coefficient of z is $1/8$; this is the compensation factor of the multigraph with a single vertex x and two loops from x to itself. The coefficient of z^2 is $\frac{25}{48} = \frac{25}{24}/2!$; the smooth labeled multigraphs



have compensation factors $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{6}$, $\frac{1}{16}$, and $\frac{1}{16}$, respectively, summing to $\frac{25}{24}$.

The smooth multigraph $\overline{M}$ obtained by repeatedly pruning M is called the *core* of M (see [26]). Let $\overline{F}$ be any family of smooth multigraphs, and let F be the set of all cyclic multigraphs whose core is a member of $\overline{F}$. The argument that proves (9.1) also proves that the univariate and bivariate generating functions for F and $\overline{F}$ are related by the equations

$$F(z) = \overline{F}\bigl(T(z)\bigr)\,; \qquad F(w,z) = \overline{F}\bigl(w,\,T(wz)/w\bigr)\,. \tag{9.3}$$

In particular we have $\widehat{E}_r(z) = \widehat{\widehat{E}}_r\bigl(T(z)\bigr)$, where $\widehat{\widehat{E}}_r$ counts all smooth graphs of excess r having no unicyclic components. This relationship accounts for the curious formula (6.11) about the last nonvanishing coefficient $\hat{e}_{rd}$; we can reason as follows: The minimum number of vertices among all graphs of excess r, when $\binom{s-2}{2} \le r < \binom{s-1}{2}$, is s, because a graph on $s-1$ vertices has at most $\binom{s-1}{2}$ edges and $\binom{s-1}{2} < s-1+r$. The coefficient of the minimum power of ζ in $\widehat{E}_r = \widehat{\widehat{E}}_r\bigl(\zeta/(1+\zeta)\bigr)$ therefore comes entirely from the $\binom{s(s-1)/2}{s+r}$ graphs on s labeled vertices having exactly $s+r$ edges. All such graphs are smooth.

When M has no unicyclic components we can go beyond pruning to another kind of vertexectomy that we will call *cancelling*: If any vertex has degree 2, we can remove it and splice together the two edges that it formerly touched. Repeated application of this process on any smooth multigraph $\overline{M}$ of excess r will lead to a multigraph $\overline{\overline{M}}$ of excess r in which every vertex has degree 3 or more. (A self-loop $\langle x, x\rangle$ is assumed to contribute 2 to the degree of x. A vertex with a self-loop will be connected to at least one other vertex, because there are no unicycles, so we will never cancel it.) The multigraph $\overline{\overline{M}}$ can be called *reduced*. Only the middle two multigraphs of the six pictured above are reduced.

There are only finitely many reduced multigraphs of excess r. For if such a multigraph has n vertices of degrees $d_1, d_2, \ldots, d_n$, it has $n + r = \frac{1}{2}(d_1 + d_2 + \cdots + d_n) \ge \frac{3}{2}n$ edges, hence $n \le 2r$. The extreme case $n = 2r$ occurs if and only if the multigraph is 3-regular, i.e., every vertex has degree exactly 3. We will see later that such regularity is, in fact, normal: The complex components of a random graph or multigraph with $\frac{1}{2}n + o(n^{3/4})$ edges almost always reduce to components that are 3-regular.

The reduced multigraph $\overline{\overline{M}}$ obtained by pruning and cancellation from a given complex multigraph M is called the *kernel* of M (see [26]). Our immediate goal is to find the generating function for all smooth multigraphs $\overline{M}$ without unicyclic components that have a given reduced multigraph $\overline{\overline{M}}$ as their kernel. For this it will be convenient to introduce another representation of a multigraph M: We label both the vertices and the edges, and we assign an arbitrary orientation to each edge, thereby obtaining a *directed edge-labeled* multigraph. Let $V = V(M)$ be the set of vertex labels and $E = E(M)$ the set of edge labels. Each edge $e \in E$ has a *dual edge* $\overline{e}$, and $\overline{E}$ is the set of all dual edges. The multigraph M is then represented as a mapping M from $E \cup \overline{E}$ to V, with the interpretation that each directed edge e runs from $M(e)$ to $M(\overline{e})$. The dual of $\overline{e}$, namely $\overline{\overline{e}}$, is e; thus $\overline{e}$ runs from $M(\overline{e})$ to $M(e)$.

If the vertex labels are $1, \ldots, n$ and if the edge labels are $1, \ldots, m$, the multigraph mapping M takes the set $\{1, \ldots, m, \overline{1}, \ldots, \overline{m}\}$ into the set $\{1, \ldots, n\}$. Any such mapping is equivalent to a sequence $\langle x_1, y_1\rangle \ldots, \langle x_m, y_m\rangle$ of ordered pairs generated by the multigraph process of section 1, where $x_k = M(k)$ and $y_k = M(\overline{k})$.

The number of different mappings M corresponding to a given multigraph M is $2^m m!\,\kappa(M)$, where κ is the compensation factor defined in (1.1). This holds because

$2^m m!$ is the number of ways to orient the edges and to assign edge labels, and κ accounts for duplicate assignments that leave us with the same mapping M.

Duplicate assignments can be treated more formally as follows. A *signed permutation* σ of a set E and its dual $\overline{E}$ is a permutation of $E \cup \overline{E}$ with the property that $\sigma\overline{e} = \overline{\sigma e}$ for all e. (The group of all signed permutations on a set of m elements is conventionally called the hyperoctahedral group $\mathbf{B}_m$; it is the group of all $2^m m!$ symmetries of an m-cube.) Given a multigraph represented as a mapping M from $E \cup \overline{E}$ to V, an *edge automorphism* is a signed permutation σ of $E \cup \overline{E}$ with the property that $M(\sigma e) = M(e)$.

It is easy to see that the number of edge automorphisms of M is $1/\kappa(M)$. Such a mapping σ must be the product of one of the $2^{m_{xx}}(m_{xx})!$ signed permutations of the m_{xx} self-loops from x to x, for each x, times one of the $(m_{xy})!$ signed permutations of the m_{xy} edges from x to y, for each $x < y$. Edge automorphisms are the automorphisms of multigraphs with labeled vertices and unlabeled edges; this explains why $\kappa(M)$ is used as a weighting function for each M in the generating functions we have been discussing.

We are now ready to prove a basic lemma about multigraphs, motivated by but noticeably simpler than the corresponding result for graphs obtained by Wright [41]:

Lemma 1. *If $\overline{\overline{M}}$ is a reduced multigraph having ν vertices, μ edges, and compensation factor κ, the generating function for all smooth, complex multigraphs $\overline{M}$ that reduce to $\overline{\overline{M}}$ under cancellation is*

$$\frac{\kappa\, z^\nu}{(1-z)^\mu\, \nu!}\,. \tag{9.4}$$

Proof. This result is "intuitively obvious," but it requires a formal proof to ensure that everything is counted properly in the presence of compensation factors. We assume that $\overline{\overline{M}}$ is represented by a fixed mapping from edges and dual edges to vertices, where the set of edge labels is $\{[1], \ldots, [\mu]\}$ and the set of vertex labels is $\{(1), \ldots, (\nu)\}$. The dual of edge $[j]$ will be denoted by $\overline{[j]} = [-j]$. The given multigraph mapping can be represented as a function M from $\{-\mu, \ldots, -1, 1, \ldots, \mu\}$ to $\{1, \ldots, \nu\}$, such that edge $[j]$ runs from $\bigl(M(j)\bigr)$ to $\bigl(M(-j)\bigr)$ and edge $[-j]$ runs from $\bigl(M(-j)\bigr)$ to $\bigl(M(j)\bigr)$. Square brackets and round parentheses are used notationally here in order to distinguish edge labels from vertex labels, although M is a function from integers to integers.

Let s_n be the coefficient of z^n in $z^\nu/(1-z)^\mu$. This quantity s_n is the number of solutions $\langle n_1, \ldots, n_\mu \rangle$ to the equation

$$n_1 + \cdots + n_\mu = n - \nu \tag{9.5}$$

in nonnegative integers. Let $m - \mu = n - \nu$; then m is the number of edges in an n-vertex multigraph that cancels to $\overline{\overline{M}}$.

We will construct $2^m m!\, n!\, s_n/\nu!$ sequences of ordered pairs $\langle x_1, y_1 \rangle \ldots \langle x_m, y_m \rangle$ of integers $1 \le x_j, y_j \le n$ such that (a) every constructed sequence defines a smooth multigraph that cancels to $\overline{\overline{M}}$; (b) every sequence that defines such a smooth multigraph is constructed exactly $1/\kappa$ times. This will prove the lemma, because of (2.2). As noted earlier, constructing a sequence $\langle x_1, y_1\rangle \ldots \langle x_m, y_m\rangle$ is equivalent to constructing a map $\overline{M}$ from $\{-m, \ldots, -1, 1, \ldots, m\}$ into $\{1, \ldots, n\}$, if we let $x_j = \overline{M}(j)$ and $y_j = \overline{M}(-j)$.

The construction is as follows. For each ordered solution $\langle n_1, \ldots, n_\mu\rangle$ to (9.5), we effectively insert n_j new vertices into edge $[j]$, thereby undoing the effect of cancellation. Formally, we construct a set of m edge labels

$$E = \{\, [j,k] \mid 1 \le j \le \mu,\ 0 \le k \le n_j \,\} \tag{9.6}$$

and a set of n vertex labels

$$V = \{\, (i) \mid 1 \le i \le \nu \,\} \ \cup\ \{\, (j,k) \mid 1 \le j \le \mu,\ 1 \le k \le n_j \,\}\,. \tag{9.7}$$

Edge $[j,k]$ runs from vertex (j,k) to vertex $(j,k+1)$, where we define for convenience

$$(j,0) = \bigl(M(j)\bigr)\,, \qquad (j, n_j+1) = \bigl(M(-j)\bigr)\,. \tag{9.8}$$

Thus the original edge $[j]$ from $\bigl(M(j)\bigr)$ to $\bigl(M(-j)\bigr)$ has become a sequence of n_j+1 edges $[j,0] \ldots [j,n_j]$ between the same two vertices, with intermediate vertices $(j,1), \ldots, (j,n_j)$.

The dual of edge $[j,k]$ will be denoted by $-[j,k]$. We also define

$$[-j,k] = -[j, n_{|j|} - k]\,, \qquad (-j,k) = (j, n_{|j|} + 1 - k)\,; \tag{9.9}$$

this means that the original edge $[-j]$ has become the edge sequence $[-j,0] \ldots [-j,n_j]$, which is the reverse of $[j,0] \ldots [j,n_j]$. Edge $[-j,k]$ runs from $(-j,k)$ to $(-j,k+1)$.

To complete the construction, let f be any one-to-one mapping from V to $\{1, \ldots, n\}$ that preserves the order of the original labels $(1), \ldots, (\nu)$; and let g be any *signed bijection* from $\overline{E} \cup E$ to $\{-m, \ldots, -1, 1, \ldots, m\}$. $\bigl($A signed bijection is a one-to-one correspondence such that $g(\overline{e}) = -g(e)$.$\bigr)$ Then we define

$$\overline{M}\bigl(g([j,k])\bigr) = f\bigl((j,k)\bigr)\,, \tag{9.10}$$

for all $[j,k]$ in $\overline{E} \cup E$. This mapping $\overline{M}$ corresponds to a sequence $\langle x_1, y_1\rangle \ldots \langle x_m, y_m\rangle$ that defines a multigraph $\overline{M}$ on $\{1, \ldots, n\}$, as stated above. We have constructed $2^m m!\, n!\, s_n/\nu!$ such sequences, since there are $2^m m!$ choices for g and $n!/\nu!$ for f, given any solution $\langle n_1, \ldots, n_\mu\rangle$ to (9.5).

It is clear that $\overline{M}$ is a smooth multigraph on n vertices that cancels to the given reduced multigraph $\overline{\overline{M}}$, and that every such $\overline{M}$ is constructed at least once. We need to verify that every mapping $\overline{M}$ is obtained exactly $1/\kappa$ times among the $2^m m!\, n!\, s_n/\nu!$ constructed mappings.

Suppose $\overline{M}$ has been constructed from $(\langle n_1, \ldots, n_\mu\rangle, f, g)$, and suppose σ is one of the $1/\kappa$ edge automorphisms of $\overline{\overline{M}}$. We will define a new construction $(\langle n'_1, \ldots, n'_\mu\rangle, f', g')$ that produces the same mapping $\overline{M}$. Our notational conventions allow us to regard σ as a permutation of $\{-\mu, \ldots, -1, 1, \ldots, \mu\}$, where

$$\sigma(-j) = -\sigma j \qquad \text{and} \qquad M(\sigma j) = M(j)\,. \tag{9.11}$$

The new construction is defined by

$$\begin{aligned}
n'_j &= n_{|\sigma(j)|}\,, && 1 \le j \le \mu\,; \\
f'\big((i)\big) &= f\big((i)\big)\,, && 1 \le i \le \nu\,; \\
f'\big((j,k)'\big) &= f\big((\sigma j, k)\big)\,, && 1 \le j \le \mu\,, \quad 1 \le k \le n'_j\,; \\
g'([j,k]') &= g([\sigma j, k])\,, && 1 \le j \le \mu\,, \quad 0 \le k \le n'_j\,; \\
\overline{M}'\big(g'([j,k]')\big) &= f'\big((j,k)'\big)\,, && 1 \le |j| \le \mu\,, \quad 0 \le k \le n'_{|j|}\,.
\end{aligned} \tag{9.12}$$

Here $(j,k)'$ and $[j,k]'$ are the new vertex and edge labels corresponding to $\langle n'_1, \ldots, n'_\mu\rangle$; they are defined in (9.6)–(9.9).

It is easy to verify that the definitions in (9.12) imply validity of the same formulas for the whole range of j and k values:

$$\begin{aligned}
f'\big((j,k)'\big) &= f\big((\sigma j, k)\big)\,, && 1 \le |j| \le \mu\,, \quad 0 \le k \le n'_{|j|} + 1\,; \\
g'([j,k]') &= g([\sigma j, k])\,, && 1 \le |j| \le \mu\,, \quad 0 \le k \le n'_{|j|}\,.
\end{aligned} \tag{9.13}$$

For example, if $j > 0$ we have

$$f'\big((j,0)'\big) = f'\big((M(j))\big) = f\big((M(j))\big) = f\big((M(\sigma j))\big) = f\big((\sigma j, 0)\big)\,;$$

$$\begin{aligned}
f'\big((j, n'_j + 1)'\big) = f'\big((M(-j))\big) &= f\big((M(-j))\big) \\
&= f\big((M\big(\sigma(-j)\big))\big) = f\big((M(-\sigma j))\big) = f\big((\sigma_j, n'_j + 1)\big)\,;
\end{aligned}$$

$$\begin{aligned}
f'\big((-j,k)'\big) = f'\big((j, n'_j{+}1{-}k)'\big) &= f\big((\sigma j, n'_j{+}1{-}k)\big) \\
&= f\big((\sigma j, n_{|\sigma j|}{+}1{-}k)\big) = f\big((-\sigma j, k)\big) = f\big((\sigma(-j), k)\big)\,.
\end{aligned}$$

Therefore if l is any value in $\{-m, \ldots, -1, 1, \ldots, m\}$, we can verify that $\overline{M}'(l) = \overline{M}(l)$, as follows: There are unique j and k such that $l = g([\sigma j, k])$. Hence $l = g'([j,k]')$, and

$$\overline{M}'(l) = f'\big((j,k)'\big) = f\big((\sigma j, k)\big) = \overline{M}(l)\,.$$

Conversely, if $(\langle n'_1, \ldots, n'_\mu\rangle, f', g')$ is another construction that makes $\overline{M}'(l) = \overline{M}(l)$ for all l, we can reverse this process and find a unique edge automorphism σ satisfying all the conditions of (9.12). Exactly ν of the vertices of $\overline{M} = \overline{M}'$ have degree ≥ 3, since $\overline{\overline{M}}$ is reduced; these are the images under f and f' of $(1), \ldots, (\nu)$, and they have the same order in $\overline{M}$. Therefore $f'\bigl((i)\bigr) = f\bigl((i)\bigr)$ for $1 \leq i \leq \nu$.

Let $l = g'([j,0])$. Since $\overline{M}'(l) = f'\bigl((j,0)\bigr) = f'\bigl((M(j))\bigr) = f\bigl((M(j))\bigr)$, we know that $\overline{M}(l)$ must be a vertex of degree ≥ 3, so there must be a value j' (either positive or negative) such that $l = g([j',0])$. This rule defines $\sigma j = j'$. We have $\overline{M}(l) = f\bigl((j',0)\bigr) = f\bigl((M(j'))\bigr)$, hence $M(\sigma j) = M(j)$.

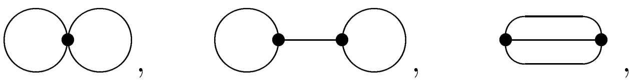
Let us say that the edge $[j,k]'$ of $\overline{M}'$ corresponds to the edge $[j',k']$ of $\overline{M}$ if $g'([j,k]') = g[j',k']$. We have defined σj for $1 \leq j \leq \mu$ in such a way that $[j,0]'$ corresponds to $[\sigma j, 0]$. Suppose we know that $[j,k]'$ corresponds to $[\sigma j, k]$ for some $k < n'_j$; then $-[j,k]'$ also corresponds to $-[\sigma j, k]$. Also $\overline{M}'\bigl(g'(-[j,k]')\bigr) = \overline{M}'\bigl(g'([-j, n'_j - k]')\bigr) = f'\bigl((-j, n'_j - k)'\bigr) = f'\bigl((j, k+1)'\bigr) = \overline{M}'\bigl(g'([j,k+1]')\bigr)$ is a vertex v of degree 2 in $\overline{M}$, which therefore equals $\overline{M}\bigl(g(-[\sigma j, k])\bigr) = f\bigl((-\sigma j, n_{|\sigma j|} - k)\bigr)$. Consequently we have $k < n_{|\sigma j|}$, $f'\bigl((j, k+1)'\bigr) = f\bigl((\sigma j, k+1)\bigr)$, and $v = \overline{M}'\bigl(g'([j,k+1]')\bigr) = \overline{M}\bigl(g([\sigma j, k+1])\bigr)$. Now $[j, k+1]'$ must correspond to $[\sigma j, k+1]$, since there is only one value $l \neq -g'([j,k]')$ such that $\overline{M}(l) = v$. In this way we prove inductively that $[j,k]'$ corresponds to $[\sigma j, k]$ for $0 \leq k \leq n'_j$, and that $n'_j = n'_{|\sigma j|}$. Hence (9.12) holds. $\square$

Let $\overline{\overline{F}}$ be a family of reduced multigraphs, and let $\overline{F}$ be the family of all smooth complex multigraphs that reduce under cancellation to a member of $\overline{\overline{F}}$. The bivariate generating functions of $\overline{F}$ and $\overline{\overline{F}}$ are then related by the equation

$$\overline{F}(w,z) = \overline{\overline{F}}\bigl(w/(1-wz), z\bigr)\,, \tag{9.14}$$

because Lemma 1 establishes this relation in the case that $\overline{\overline{F}}$ has only one member. Equation (9.14) says simply that every edge in $\overline{\overline{F}}$, represented by w, is to be replaced by a sequence of one or more edges, represented by $w/(1-wz) = w + w^2 z + w^3 z^2 + \cdots$; perhaps this means that Lemma 1 is indeed obvious and that the lengthy proof was unnecessary. It is, however, comforting to know that a formal verification is possible, when one is beginning to learn the power of generating function techniques. And somehow, examples of multigraphs with numerous self-loops and repeated edges do seem to mandate a formal proof, because compensation factors change when edges are manipulated.

As an example of Lemma 1, let us derive explicitly the generating function $\overline{E}_1(z) = \overline{C}_1(z)$ for all smooth bicyclic multigraphs. All such multigraphs cancel to a reduced multigraph of excess 1, which can have at most 2 vertices and 3 edges. There are only three possibilities,



(9.15)

having $\kappa = \frac{1}{8}$, $\frac{1}{4}$, and $\frac{1}{6}$, respectively. Therefore

$$\overline{E}_1(z) = \overline{C}_1(z) = \frac{z}{8(1-z)^2} + \frac{z^2}{8(1-z)^3} + \frac{z^2}{12(1-z)^3} = \frac{z(3+2z)}{24(1-z)^3}\,, \tag{9.16}$$

in agreement with (9.2). Wright [41] states that there are 15 connected, unlabeled, reduced multigraphs of excess 2, and 107 of excess 3.

If a reduced multigraph of excess r has exactly $2r-d$ vertices, we will say that it has *deficiency* d. A reduced multigraph of deficiency 0 is 3-regular; we will call such multigraphs *clean*.

Corollary. *The coefficient e_{rd} in (5.10) and (7.3) is $(2r-d)!^{-1}\sum\kappa(\overline{\overline{M}})$, summed over all reduced, labeled multigraphs $\overline{\overline{M}}$ of excess r and deficiency d. The coefficient c_{rd} in (8.4) can be obtained in the same way, but restricting the sum to connected multigraphs.* $\square$

This corollary leads to a completely different proof of Theorem 1, because it allows us to obtain formula (7.3) for e_{rd} by a combinatorial counting argument. Consider a reduced multigraph that has exactly d_k vertices of degree k, for each $k \ge 3$; then $d_3 + d_4 + \cdots = n$ and $3d_3 + 4d_4 + \cdots = 2m$. We can calculate $\sum\kappa(\overline{\overline{M}})$ over all such $\overline{\overline{M}}$ by counting the number of relevant sequences $\langle x_1, y_1\rangle \ldots \langle x_m, y_m\rangle$ and dividing by $2^m m!$; and the number of ways to choose $\langle x_1, y_1\rangle \ldots \langle x_m, y_m\rangle$ is clearly a product of multinomial coefficients,

$$\frac{(2m)!}{3!^{d_3}\ 4!^{d_4}\ \ldots}\quad\frac{n!}{d_3!\ d_4!\ \ldots}\,,$$

since the first factor is the number of ways to partition $2m$ slots into d_k labeled classes of size k for each k, and the second factor counts the assignments of vertex labels to those classes. To obtain all reduced multigraphs of excess r and deficiency d, we sum over all sequences of nonnegative integers $\langle d_3, d_4, \ldots\rangle$ such that $\sum_{k\ge3} d_k = 2r-d$ and $\sum_{k\ge3} kd_k = 6r-2d$, or equivalently

$$\sum_{k\ge3}(k-3)d_k = d \qquad\text{and}\qquad \sum_{k\ge3}(k-2)d_k = 2r\,.$$

Let

$$f_{cd} = \sum\left\{\prod_{k\ge3}\frac{1}{k!^{d_k}\ d_k!}\ \middle|\ \sum_{k\ge3}(k-3)d_k = d \quad\text{and}\quad \sum_{k\ge3}(k-2)d_k = c\right\}. \tag{9.17}$$

We have just proved that

$$e_{rd} = \frac{(6r-2d)!}{2^{3r-d}(3r-d)!}\,f_{(2r)d}\,. \tag{9.18}$$

And we can readily calculate a bivariate generating function for the coefficients f_{rd}:

$$\begin{aligned}
\sum_{r,d\geq 0} f_{rd} w^d z^r &= \sum_{d_3,d_4,\ldots\geq 0} \prod_{k\geq 3} \frac{w^{(k-3)d_k} z^{(k-2)d_k}}{k!^{d_k}\, d_k!} \\
&= \prod_{k\geq 3} \sum_{d_k\geq 0} \left(\frac{w^{k-3}z^{k-2}}{k!}\right)^{d_k} \frac{1}{d_k!} \\
&= \prod_{k\geq 3} \exp(w^{k-3}z^{k-2}/k!) \\
&= \exp\left(w^{-3}z^{-2}\sum_{k\geq 3}\frac{(wz)^k}{k!}\right) = \exp\left(\frac{z}{6}F\left(\frac{wz}{4}\right)\right),
\end{aligned}$$

where F is the function defined in (7.5). Comparing (9.18) to (7.3) now yields the promised proof of (7.4):

$$\begin{aligned}
P_d(r) &= 2^{2r+d}3^{2r-d}(2r-d)!\, f_{(2r)d} \\
&= 2^{2r+d}3^{2r-d}(2r-d)!\, [w^d z^{2r}]\, \exp\bigl(z\,F(wz/4)/6\bigr) \\
&= 2^{2r+d}3^{2r-d}(2r-d)!\, [w^d z^{2r-d}]\exp\bigl(z\,F(w/4)/6\bigr) \\
&= 2^{2d}(2r-d)!\, [w^d z^{2r-d}]\exp\bigl(z\,F(w/4)\bigr) = [w^d]\, F(w)^{2r-d}\,.
\end{aligned}$$

These observations also allow us to express e_{rd} in the suggestive form

$$e_{rd} = \frac{1}{2^{3r-d}(3r-d)!}\left\{ {6r-2d \atop 2r-d} \right\}_{\geq 3}, \qquad (9.19)$$

where $\left\{ {m \atop n} \right\}_{\geq 3}$ denotes the number of ways to partition an m-element set into n subsets, each containing at least 3 elements. The asymptotic behavior of the integers $2^{3r-d}(3r-d)!\, e_{rd}$ will therefore be analogous to the asymptotic behavior of Stirling numbers.

Lemma 1 captures the combinatorial essence of the generating functions for all complex multigraphs. We can obtain a similar generating function for graphs instead of multigraphs, but we must work a bit harder, and the formulas are not as attractive. The following improvement over Wright's original treatment [41] is based on an approach suggested by V. E. Stepanov [36].

Lemma 2. *Let $\overline{\overline{M}}$ be a reduced multigraph having ν vertices, μ edges, compensation factor κ, and μ_{xy} edges between x and y for $1 \leq x \leq y \leq \nu$. The generating function for all smooth, complex graphs $\overline{G}$ that lead to $\overline{\overline{M}}$ under cancellation is*

$$\frac{\kappa\, z^\nu}{(1-z)^\mu\, \nu!}\, P(\overline{\overline{M}}, z)\,, \qquad (9.20)$$

where

$$P(\overline{\overline{M}}, z) = \prod_{x=1}^{\nu} \left(z^{2\mu_{xx}} \prod_{y=x+1}^{\nu} z^{\mu_{xy}-1}\bigl(\mu_{xy} - (\mu_{xy}-1)z\bigr) \right) \tag{9.21}$$

is a polynomial in z such that $P(\overline{\overline{M}}, 1) = 1$.

Proof. We argue as in Lemma 1, but we must restrict the solutions $\langle n_1, \ldots, n_\mu \rangle$ of (9.5) to cases that produce a graph instead of a multigraph. Thus, each n_j that corresponds to a self-loop must be ≥ 2, so we use $z^2/(1-z)$ instead of $1/(1-z)$ in the contribution that n_j makes to the overall generating function. A subsequence $\langle n_j, \ldots, n_{j+k-1} \rangle$ that corresponds to $k = \mu_{xy}$ edges between distinct vertices $x < y$ must have the property that at most one of $\langle n_j, \ldots, n_{j+k-1} \rangle$ is zero; hence we use

$$\frac{z^k}{(1-z)^k} + \frac{kz^{k-1}}{(1-z)^{k-1}} = \frac{z^{k-1}\bigl(k-(k-1)z\bigr)}{(1-z)^k}$$

instead of $1/(1-z)^k$ in its contribution. The net effect is to multiply the previous generating function by $P(\overline{\overline{M}}, z)$. $\square$

Replacing z by $T(z)$ gives the generating function for all graphs that prune and cancel to $\overline{\overline{M}}$. For example, the generating function $\widehat{E}_1(z) = \widehat{C}_1(z) = \widehat{W}(z)$ of (3.6) can be read off from (9.15): It is

$$\frac{T(z)^5}{8\bigl(1-T(z)\bigr)^2} + \frac{T(z)^6}{8\bigl(1-T(z)\bigr)^3} + \frac{T(z)^4\bigl(3-2T(z)\bigr)}{12\bigl(1-T(z)\bigr)^3}\,. \tag{9.22}$$

The degree of the polynomial $P(\overline{\overline{M}}, z)$ is the total number of "penalty points" of $\overline{\overline{M}}$, where each self-loop costs two penalty points, and where each cluster of $\mu_{xy} > 1$ multiple edges between distinct vertices costs $\mu_{xy} - 1$. If $\overline{\overline{M}}$ is a graph, the degree is zero and $P(\overline{\overline{M}}, z) = 1$. At the other extreme, if all edges of $\overline{\overline{M}}$ are self-loops, the degree is 2μ.

The quantity $T(z)^\nu/\bigl(1-T(z)\bigr)^\mu$ becomes $\zeta^\nu(1+\zeta)^{\mu-\nu}$, when we express it in terms of the variable $\zeta = T(z)/\bigl(1-T(z)\bigr)$ introduced in section 5; the quantity $P\bigl(\overline{\overline{M}}, T(z)\bigr)$ becomes $P\bigl(\overline{\overline{M}}, \zeta/(1+\zeta)\bigr)$. If we restrict consideration to connected multigraphs of excess r, we get rational functions of ζ with denominator $(1+\zeta)^{r+2}$; this denominator occurs when there are $(r+1)$ self-loops in $\overline{\overline{M}}$. However, we have seen in Theorem 3 that the denominator of $\widehat{C}_r$ is always a divisor of $(1+\zeta)^2$. There seems to be no easy combinatorial explanation for the cancellation that occurs when the contributions of different $\overline{\overline{M}}$ are added together. Some of the properties of connected graphs are easier to derive by combinatorics, others are easier to derive by algebra.

The actual coefficients of $P\bigl(\overline{\overline{M}}, \zeta/(1+\zeta)\bigr)$ do not make any significant difference asymptotically, when graphs are sparse; we will see later that the asymptotic behavior as

$\zeta \to \infty$ is what counts, hence we only need to know that $P(\overline{\overline{M}}, 1) = 1$. We observed earlier that the leading coefficients $\hat{e}_{r0}$ and e_{r0} of $\widehat{E}$ and E are equal, as are the leading coefficients $\hat{c}_{r0}$ and c_{r0}. Now Lemma 2 shows in fact that each reduced multigraph $\overline{\overline{M}}$ makes the same contribution to the leading coefficient for graphs as it does for multigraphs.

10. A lemma from contour integration. Studies of random graphs that have $m \approx \frac{1}{2}n$ edges are traditionally broken into two cases, the "subcritical" case where $m < \frac{1}{2}n$ and the "supercritical" case where $m > \frac{1}{2}n$. It is desirable, however, to have estimates of probabilities that hold uniformly for all m in the vicinity of $\frac{1}{2}n$, passing smoothly from one side to the other. The following lemma, based on techniques introduced in [14], will be our key tool for the computation of probabilities.

Lemma 3. *If $m = \frac{1}{2}n(1 + \mu n^{-1/3})$ and if y is any real constant, we have*

$$\frac{2^m\, m!\, n!}{(n-m)!\, n^{2m}}\,[z^n]\,\frac{U(z)^{n-m}}{\bigl(1-T(z)\bigr)^y} = \sqrt{2\pi}\,A(y,\mu)\,n^{y/3-1/6} + O\bigl((1+|\mu|^B)n^{y/3-1/2}\bigr) \tag{10.1}$$

uniformly for $|\mu| \le n^{1/12}$, where $B = \max(4,\, \frac{9}{2} - y)$ and

$$A(y,\mu) = \frac{e^{-\mu^3/6}}{3^{(y+1)/3}} \sum_{k\ge 0} \frac{\bigl(\frac{1}{2}3^{2/3}\mu\bigr)^k}{k!\,\Gamma\bigl((y+1-2k)/3\bigr)}\,. \tag{10.2}$$

As $\mu \to -\infty$, we have

$$A(y,\mu) = \frac{1}{\sqrt{2\pi}\,|\mu|^{y-1/2}}\left(1 - \frac{3y^2+3y-1}{6|\mu|^3} + O(\mu^{-6})\right); \tag{10.3}$$

as $\mu \to +\infty$, we have

$$A(y,\mu) = \frac{e^{-\mu^3/6}}{2^{y/2}\mu^{1-y/2}}\left(\frac{1}{\Gamma(y/2)} + \frac{4\mu^{-3/2}}{3\sqrt{2}\,\Gamma(y/2-3/2)} + O(\mu^{-2})\right). \tag{10.4}$$

Moreover, (10.1) can be improved to

$$\frac{2^m\, m!\, n!}{(n-m)!\, n^{2m}}\,[z^n]\,\frac{U(z)^{n-m}}{\bigl(1-T(z)\bigr)^y} = \sqrt{2\pi}\,A(y,\mu)\,n^{y/3-1/6}\bigl(1 + O(\mu^4 n^{-1/3})\bigr) \tag{10.5}$$

if $|\mu|$ goes to infinity with n while remaining $\le n^{1/12}$.

Proof. First we need to derive some auxiliary results about the function A. If α is any positive number, we define a path $\Pi(\alpha)$ in the complex plane that consists of the following three straight line segments:

$$s(t) = \begin{cases} -e^{-\pi i/3}\,t, & \text{for } -\infty < t \le -2\alpha; \\ \alpha + it\sin\pi/3, & \text{for } -2\alpha \le t \le +2\alpha; \\ e^{+\pi i/3}\,t, & \text{for } +2\alpha \le t < +\infty. \end{cases} \tag{10.6}$$

Now we define

$$A(y,\mu) = \frac{1}{2\pi i}\int_{\Pi(1)} s^{1-y} e^{K(\mu,s)}\,ds\,, \tag{10.7}$$

where $K(\mu,s)$ is the polynomial

$$K(\mu,s) = \frac{(s+\mu)^2(2s-\mu)}{6} = \frac{s^3}{3} + \frac{\mu s^2}{2} - \frac{\mu^3}{6}\,. \tag{10.8}$$

Our first goal is to show that $A(y,\mu)$ satisfies (10.2), (10.3), and (10.4).

To get (10.2), we make the substitution $u = s^3/3$. As s traverses $\Pi(1)$, the variable u traverses an interesting contour Γ that begins at $-\infty$ and hugs the lower edge of the negative axis, then circles the origin counterclockwise and returns to $-\infty$ along the upper edge of the axis. On this contour Γ we have Hankel's well-known formula for the reciprocal Gamma function,

$$\frac{1}{\Gamma(z)} = \frac{1}{2\pi i}\int_\Gamma \frac{e^u\,du}{u^z}\,.$$

(See, for example, [18, Theorem 8.4b].) So we can expand (10.7) into an absolutely convergent series, after substituting $3^{1/3}u^{1/3}$ for s:

$$\begin{aligned}\int_{\Pi(1)} s^{1-y}e^{K(\mu,s)}\,ds &= \frac{e^{-\mu^3/6}}{3^{(y+1)/3}}\int_\Gamma \frac{e^u\exp\bigl(\frac12 3^{2/3}\mu u^{2/3}\bigr)\,du}{u^{(y+1)/3}}\\ &= \frac{e^{-\mu^3/6}}{3^{(y+1)/3}}\int_\Gamma \sum_{k\ge0}\frac{\bigl(\frac12 3^{2/3}\mu\bigr)^k e^u\,du}{k!\,u^{(y+1-2k)/3}}\,.\end{aligned}$$

Interchanging summation and integration, and applying Hankel's formula, gives (10.2).

To get (10.3) and (10.4), we note first that the integral (10.7) can be taken over any path $\Pi(\alpha)$, not just $\Pi(1)$, because $e^{K(\mu,s)}$ has no singularities. Moreover, we can "straighten out" the path $\Pi(\alpha)$, changing it to a single straight line from $\alpha - i\infty$ to $\alpha + i\infty$, if α is sufficiently large. For we can readily verify that the integrand is exponentially small on any large circular arc $s = Re^{i\theta}$, as $|\theta|$ increases from $\pi/3$ to the angle where $R\cos\theta = \alpha$: The real part of s^3 is $R^3\cos 3\theta$, which increases from $-R^3$ to $4\alpha^3 - 3R^2\alpha$; and the real part of s^2 lies between $-R^2$ and $-R^2/2$. Hence the real part of $K(\mu,s)$ will be at most $-cR^2$ for some positive $c = c(\alpha)$ on the entire arc, whenever $\alpha > 0$ and $\alpha > -\frac12\mu$; this will make $s^{1-y}e^{K(\mu,s)}$ exponentially small.

If μ is negative, let $\alpha = -\mu$; then

$$A(y,-\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\alpha+it)^{1-y}e^{K(-\alpha,\alpha+it)}\,dt$$

$$= \frac{1}{2\pi\sqrt{\alpha}} \int_{-\infty}^{\infty} (\alpha + it/\sqrt{\alpha}\,)^{1-y} e^{K(-\alpha,\alpha+it/\sqrt{\alpha}\,)}\, dt$$

$$= \frac{1}{2\pi\alpha^{y-1/2}} \int_{-\infty}^{\infty} \left(1 + \frac{it}{\alpha^{3/2}}\right)^{1-y} e^{-t^2/2 - it^3/(3\alpha^{3/2})}\, dt\,, \tag{10.9}$$

and we can find the asymptotic value of the remaining integral by using Laplace's standard technique of "tail-exchange" (see [17, section 9.4]):

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left(1 + \frac{it}{\alpha^{3/2}}\right)^{1-y} e^{-t^2/2 - it^3/(3\alpha^{3/2})}\, dt \\
&\quad = \int_{-\alpha^{\epsilon}}^{\alpha^{\epsilon}} \left(1 + \frac{it}{\alpha^{3/2}}\right)^{1-y} e^{-t^2/2 - it^3/(3\alpha^{3/2})}\, dt \;+\; O\big(e^{-\alpha^{2\epsilon}/3}\big) \\
&\quad = \int_{-\alpha^{\epsilon}}^{\alpha^{\epsilon}} e^{-t^2/2} \left(1 + \frac{(1-y)\,it}{\alpha^{3/2}} - \frac{it^3}{3\alpha^{3/2}} + O(\alpha^{6\epsilon-3})\right) dt \;+\; O\big(e^{-\alpha^{2\epsilon}/3}\big) \\
&\quad = \sqrt{2\pi} \;+\; O\big(\alpha^{6\epsilon-3}\big)\,.
\end{aligned}$$

If we expand the integrand further, to terms that are $O(\alpha^{12\epsilon-6})$, we obtain

$$A(y,-\alpha) = \frac{1}{\sqrt{2\pi}\alpha^{y-1/2}} \left(1 \;-\; \frac{3y^2+3y-1}{6\alpha^3} \;+\; O\big(\alpha^{12\epsilon-6}\big)\right).$$

The method can clearly be extended, in principle, to give a complete asymptotic series in powers of α^{-3}, beginning as shown in (10.3).

We also want to know the asymptotic value of $A(y,\mu)$ as $\mu \to +\infty$, and for this we need to work a bit harder. A combination of the methods we have used to prove (10.2) and (10.3) will establish (10.4). The idea now is to integrate on the path $\mu^{-1} + it/\sqrt{\mu}$:

$$\begin{aligned}
A(y,\mu) &= \frac{e^{K(\mu,\mu^{-1})}}{2\pi\sqrt{\mu}} \int_{-\infty}^{\infty} \left(\mu^{-1} + \frac{it}{\sqrt{\mu}}\right)^{1-y} \exp\big(it(\mu^{-1/2} + \mu^{-5/2}) \\
&\qquad\qquad - t^2\big(\tfrac{1}{2} + \mu^{-2}\big) - \tfrac{1}{3} it^3 \mu^{-3/2}\big)\, dt \\
&= \frac{e^{K(\mu,\mu^{-1})}}{2\pi\mu^{1-y/2}} \int_{-\infty}^{\infty} (\mu^{-1/2} + it)^{1-y} e^{-t^2/2} g(it,\mu)\, dt \\
&= \frac{e^{K(\mu,\mu^{-1})}}{2\pi i\mu^{1-y/2}} \int_{-\infty i}^{\infty i} (v + \mu^{-1/2})^{1-y} e^{v^2/2} g(v,\mu)\, dv\,,
\end{aligned}$$

where the last step replaces it by v. We can distort the path of v so that it crosses the positive real axis, and then replace $v^2/2$ by u to get Hankel's contour Γ again:

$$\begin{aligned} A(y,\mu) &= \frac{e^{K(\mu,\mu^{-1})}}{2\pi i\mu^{1-y/2}}\int_\Gamma (\sqrt{2u}+\mu^{-1/2})^{1-y}e^u g(\sqrt{2u},\mu)\,\frac{du}{\sqrt{2u}} \\ &= \frac{e^{K(\mu,\mu^{-1})}}{2^{1+y/2}\pi i\mu^{1-y/2}}\int_\Gamma (1+(2\mu u)^{-1/2})^{1-y}u^{-y/2}e^u g(\sqrt{2u},\mu)\,du\,. \end{aligned}$$

For definiteness we can stipulate that the contour Γ lies entirely on the negative axis, except for a circular loop about 0 with a radius of 1. When u is on the negative axis, say $u=-t$, the quantity $\sqrt{2u}$ will be $-i\sqrt{2t}$ on the first part of Γ and $+i\sqrt{2t}$ on the last, so we will have

$$g(\sqrt{2u},\mu) = \exp\bigl(\mp i\sqrt{2t}(\mu^{-1/2}+\mu^{-5/2}) - 2t\mu^{-2} \pm \tfrac{1}{3}i(2t)^{3/2}\mu^{-3/2}\bigr)\,.$$

On the portions of Γ for which $|u|\ge\mu^\epsilon$, the integrand is superpolynomially small;* hence

$$\int_\Gamma = \int_{\Gamma[\mu^\epsilon]} + O\bigl(e^{-\mu^\epsilon/2}\bigr)\,,$$

where $\Gamma[\mu^\epsilon]$ is the subcontour that runs along the lower edge of the negative axis from $-\mu^\epsilon$ to the circle $u=e^{i\theta}$ and back to $-\mu^\epsilon$ on the top edge of the axis. On $\Gamma[\mu^\epsilon]$ we have

$$\begin{aligned} &(2\mu u)^{-1/2} = O(\mu^{-1/2})\,, \\ &g(\sqrt{2u},\mu) = 1+\sqrt{2u}\,\mu^{-1/2}+O(\mu^{\epsilon-1})\,; \end{aligned}$$

and $\int_\Gamma |u^{-y/2}e^u|\,du$ exists. Hence

$$\begin{aligned} &\int_{\Gamma[\mu^\epsilon]} \bigl(1+(2\mu u)^{-1/2}\bigr)^{1-y}u^{-y/2}e^u g\bigl(\sqrt{2u},\mu\bigr)\,du \\ &\quad= \int_{\Gamma[\mu^\epsilon]} \Bigl(u^{-y/2}+\frac{1}{\sqrt{2\mu}}\bigl((1-y)u^{-(y+1)/2}+2u^{-(y-1)/2}\bigr)\Bigr)e^u\,du + O(\mu^{\epsilon-1}) \\ &\quad= 2\pi i\biggl(\frac{1}{\Gamma(y/2)}+\frac{1}{\sqrt{2\mu}}\biggl(\frac{1-y}{\Gamma\bigl((y+1)/2\bigr)}+\frac{2}{\Gamma\bigl((y-1)/2\bigr)}\biggr)\biggr) + O(\mu^{\epsilon-1})\,. \end{aligned}$$

The coefficient of $\mu^{-1/2}$ vanishes, because $\Gamma\bigl((y+1)/2\bigr) = \frac{1}{2}(y-1)\Gamma\bigl((y-1)/2\bigr)$. We can use the same method to expand the integrand further, obtaining (10.4).

* "Superpolynomially small" means that it approaches zero faster than any negative power of the argument.

Notice that $1/\Gamma(y/2)$ or $1/\Gamma(y/2-3/2)$ may be zero, but not both. Therefore (10.4) gives the asymptotically leading term of $A(y,\mu)$ in all cases.

Whew—we have worked pretty hard to establish (10.2)–(10.4), and we still haven't begun to tackle the main assertion of the lemma. Fortunately, the work we have done so far will help streamline the rest of the proof. The next step is to analyze the factor at the left of (10.1); a routine application of Stirling's approximation shows that

$$\frac{2^m\, m!\, n!}{(n-m)!\, n^{2m}} \;=\; \sqrt{2\pi n}\, 2^{n-m}\, e^{-\mu^3/6-n}\left(1+O\left(\frac{1+\mu^4}{n^{1/3}}\right)\right), \tag{10.10}$$

uniformly for $|\mu|\le n^{1/12}$ as $n\to\infty$, when $m=\frac{n}{2}(1+\mu n^{-1/3})$.

Now we turn to the other parts of (10.1). Equation (3.2) implies that T has an analytic continuation in which $T(ze^{-z})=z$ for $|z|<1$. Hence, by (3.3) and Cauchy's formula for $[z^n]\,f(z)$, we can substitute $\tau=ze^{-z}$ and get

$$\begin{aligned}[z^n]\,\frac{U(z)^{n-m}}{\bigl(1-T(z)\bigr)^y} &= \frac{1}{2\pi i}\oint \frac{U(\tau)^{n-m}\,d\tau}{\bigl(1-T(\tau)\bigr)^y\,\tau^{n+1}}\\ &= \frac{e^n\,2^{m-n}}{2\pi i}\oint (1-z)^{1-y}e^{nh(z)}\,\frac{dz}{z}\,,\end{aligned} \tag{10.11}$$

where

$$\begin{aligned}h(z) &= z-1-\frac{m}{n}\ln z+\left(1-\frac{m}{n}\right)\ln(2-z)\\ &= z-1-\ln z-\left(1-\frac{m}{n}\right)\ln\frac{1}{1-(z-1)^2}\,.\end{aligned} \tag{10.12}$$

The contour in (10.11) should keep $|z|<1$. Notice that $h(1)=h'(1)=0$; if $m=\frac{1}{2}n$ we also have $h''(1)=0$. This triple zero accounts for the procedure we shall use to investigate the value of (10.11) for large n.

Let $\nu=n^{-1/3}$, and let α be the positive solution to

$$\mu \;=\; \alpha^{-1}\;-\;\alpha\,. \tag{10.13}$$

We will evaluate (10.11) on the path $z=e^{-(\alpha+it)\nu}$, where t runs from $-\pi n^{1/3}$ to $\pi n^{1/3}$:

$$\oint f(z)\,\frac{dz}{z}=i\nu\int_{-\pi n^{1/3}}^{\pi n^{1/3}} f(e^{-(\alpha+it)\nu})\,dt\,. \tag{10.14}$$

It will turn out that the main contribution to the value of this integral comes from the vicinity of $t=0$.

The magnitude of $e^{h(z)}$ depends on $\Re h(z)$.† If $z = \rho e^{i\theta}$, we have

$$\Re h(\rho e^{i\theta}) = \rho\cos\theta - 1 - \frac{m}{n}\ln\rho + \frac{1}{2}\left(1-\frac{m}{n}\right)\ln\left(4 - 4\rho\cos\theta + \rho^2\right). \tag{10.15}$$

The derivative with respect to θ is $-\rho g(\theta)\sin\theta$, where

$$g(\theta) = 1 - \frac{2\left(1-\frac{m}{n}\right)}{4 - 4\rho\cos\theta + \rho^2} \geq \frac{(2-\rho)^2 - 2\left(1-\frac{m}{n}\right)}{4 - 4\rho\cos\theta + \rho^2}; \tag{10.16}$$

and $g(\theta)$ is positive when $\rho = e^{-\alpha\nu}$, because $2(1-\frac{m}{n}) = 1 - \mu\nu < 1 + \alpha\nu < (2 - e^{-\alpha\nu})^2$. (We always have $0 < \alpha\nu < 2$ when $|\mu| \leq n^{1/12}$, and it is not difficult to verify that $(2-e^{-x})^2 > 1 + x$ when $0 < x < 2$.) Hence $\Re h(e^{-(\alpha+it)\nu})$ decreases as $|t|$ increases, and $|e^{nh(z)}|$ has its maximum on the circle $z = e^{-(\alpha+it)\nu}$ when $t = 0$.

Looking further at $nh(e^{-s\nu})$, we have the asymptotic estimate

$$n\,h(e^{-s\nu}) = \tfrac{1}{3}s^3 + \tfrac{1}{2}\mu s^2 + O\big((\mu^2 s^2 + s^4)\nu\big), \tag{10.17}$$

uniformly in any region such that $|s\nu| \leq c$ where $c < \ln 2$. This follows from (10.12), using the expansion

$$\ln\frac{1}{1-(e^u-1)^2} = u^2 + u^3 + O(u^4), \qquad |u| \leq c.$$

We also have

$$(1-e^{-s\nu})^{1-y} = s^{1-y}\nu^{1-y}\big(1 + O(s\nu)\big). \tag{10.18}$$

Therefore if $f(z) = (1-z)^{1-y}e^{nh(z)}$ is the integrand of (10.11) and (10.14), we have

$$e^{-\mu^3/6} f(e^{-s\nu}) = \nu^{1-y}s^{1-y}e^{K(\mu,s)}\big(1 + O(s\nu) + O(\mu^2 s^2\nu) + O(s^4\nu)\big), \tag{10.19}$$

when $s = O(n^{1/12})$. (This restriction on s ensures that $\mu^2 s^2\nu$ and $s^4\nu$ are bounded, hence the O terms of (10.17) can be moved out of the exponent.)

The exponent $K(\mu, s)$ in (10.19), when $s = \alpha + it$, is

$$K(\alpha^{-1}-\alpha, \alpha+it) = \left(\tfrac{1}{2}\alpha^{-1} - \tfrac{1}{6}\alpha^{-3}\right) + it - \tfrac{1}{2}(\alpha+\alpha^{-1})t^2 - \tfrac{1}{3}it^3.$$

The real part is bounded above by $\frac{1}{3} - t^2$, for all $\alpha > 0$, since $3\alpha^{-1} - \alpha^{-3} \leq 2 \leq \alpha + \alpha^{-1}$, with equality iff $\alpha = 1$. Hence the integrand $f(e^{-s\nu})$ becomes superpolynomially small when $|t|$ grows, and we have

$$\begin{aligned}
\frac{e^{-\mu^3/6}}{2\pi i}\oint f(z)\,\frac{dz}{z} &= \frac{\nu e^{-\mu^3/6}}{2\pi}\int_{-n^{1/12}}^{n^{1/12}} f\big(e^{-(\alpha+it)\nu}\big)\,dt + O\big(e^{-(\alpha+\alpha^{-1})n^{1/6}/3}\big) \\
&= \frac{\nu^{2-y}}{2\pi i}\int_{\alpha-n^{1/12}i}^{\alpha+n^{1/12}i} s^{1-y}e^{K(\mu,s)}\,ds + O(\nu^{3-y}R) + O\big(e^{-(\alpha+\alpha^{-1})n^{1/6}/3}\big) \\
&= \nu^{2-y}A(y,\mu) + O(\nu^{3-y}R) + O\big(e^{-\max(2,|\mu|)n^{1/6}/3}\big),
\end{aligned}$$

† $\Re(x+iy) = x$ denotes the real part of the complex number $x + iy$.

where $s = \alpha + it$ and

$$R = \int_{-\infty}^{\infty} \bigl(|s^{2-y}| + \mu^2 |s^{3-y}| + |s^{5-y}| \bigr) \bigl| e^{K(\mu,s)} \bigr| \, dt = R_1 + R_2 + R_3 .$$

The lemma will be proved if we can show that $R = O\bigl(1+\mu^B\bigr)$ and that $R/A(y,\mu) = O(\mu^4)$ as $|\mu| \to \infty$.

To show that each remainder integral R_1, R_2, R_3 is small, we will let $s = \alpha + iu/\beta$, where $u = \beta t$ and

$$\beta = \sqrt{\alpha + \alpha^{-1}}\,. \tag{10.20}$$

Notice that when $\mu \le 0$ we have $\alpha \ge 1$ and $\alpha = |\mu| + O\bigl(|\mu|^{-1}\bigr)$; when $\mu \ge 0$ we have $0 < \alpha \le 1$ and $\alpha^{-1} = \mu + O(\mu^{-1})$. Therefore in both cases

$$\beta = |\mu|^{1/2} + O\bigl(|\mu|^{-1/2}\bigr) \qquad \text{as } |\mu| \to \infty. \tag{10.21}$$

The first remainder, R_1, is

$$\int_{-\infty}^{\infty} |\alpha + it|^{2-y} \, |e^{K(\mu,\alpha+it)}| \, dt = \frac{e^{\alpha^{-1}/2 - \alpha^{-3}/6}}{\beta} \int_{-\infty}^{\infty} \left| \alpha + \frac{iu}{\beta} \right|^{2-y} e^{-u^2/2} \, du\,.$$

If $\mu < 0$, we have $\alpha\beta \ge \sqrt{2}$, hence

$$R_1 \le \frac{O(1)\, \alpha^{2-y}}{\beta} \int_{-\infty}^{\infty} \max\left(1,\, \left|1 + \frac{iu}{\sqrt{2}}\right|^{2-y}\right) e^{-u^2/2} \, du\,;$$

and the integral exists, so this is $O\bigl(|\mu|^{3/2-y}\bigr)$ by (10.21). Similarly, $R_2 = O\bigl(|\mu|^{2+5/2-y}\bigr)$ when $\mu < 0$, and $R_3 = O\bigl(|\mu|^{9/2-y}\bigr)$.

On the other hand, when $\mu > 0$ we have $\alpha\beta \le \sqrt{2}$, and we need to be more cautious. Instead of letting t run from $-\infty$ to $+\infty$ through real values in the derivation above, we will distort the path slightly near the origin, so that t passes through the point $-i/\beta$ and so that $\beta s = \alpha\beta + iu$ never has magnitude less than 1. (We used essentially the same sort of contour when deriving (10.4).) Then u passes through the point $-i$, and we have

$$R_1 \le \frac{O(1)\, e^{-\mu^3/6}}{\beta^{\,3-y}} \int_{-\infty}^{\infty} \max\bigl(1,\, \bigl|\sqrt{2} + iu\bigr|^{2-y}\bigr) e^{-u^2/2} \, du\,.$$

We therefore have $R_1 = O(e^{-\mu^3/6}\mu^{y/2-3/2})$; similarly, $R_2 = O(e^{-\mu^3/6}\mu^{y/2-4/2+2})$ and $R_3 = O(e^{-\mu^3/6}\mu^{y/2-6/2})$. From (10.4) we know that $A(y,\mu)$ grows at least as fast as $e^{-\mu^3/6}\mu^{y/2-5/2}$. So in this case the remainders behave even better than we have claimed in (10.5), although the error term $O(\mu^4/n^{1/3})$ is still necessary because of (10.10). $\square$

If we differentiate the integral (10.7) with respect to s and with respect to μ, we obtain a recurrence relation for $A(y,\mu)$ and a formula for the derivative:

$$(y-2)A(y,\mu) = \mu A(y-2,\mu) + A(y-3,\mu)\,; \tag{10.22}$$

$$A'(y,\mu) = \tfrac{1}{2}A(y-2,\mu) - \tfrac{1}{2}\mu^2 A(y,\mu)\,. \tag{10.23}$$

(The prime here denotes differentiation with respect to the second argument, μ. The derivative with respect to y could also be worked out; but it depends on the derivative of the Gamma function in a rather complicated way, and it is not expressible directly in terms of A itself.)

The derivative is more easily investigated if we define

$$B(y,\mu) = e^{\mu^3/6}\,A(y,\mu)\,. \tag{10.24}$$

Then

$$(y-2)B(y,\mu) = \mu B(y-2,\mu) + B(y-3,\mu)\,; \tag{10.25}$$

$$B'(y,\mu) = \tfrac{1}{2}B(y-2,\mu)\,. \tag{10.26}$$

It is easy to verify that the infinite series of (10.2) satisfies these relations. Repeated application of (10.25) and (10.26) leads to a third-order differential equation for $B = B(y,\mu)$:

$$8B''' - 4\mu^2 B'' + 2\mu(2y-9)B' - (y-2)(y-5)B = 0\,. \tag{10.27}$$

We can see from (10.22) that, for any fixed $\mu \ge 0$, there are infinitely many negative values of y such that $A(y,\mu) = 0$. For if $y < 0$ and there is no root between $y-1$ and y, then $A(y-1,\mu)$ and $A(y,\mu)$ have the same sign; hence $A(y+2,\mu)$ has the opposite sign, and there's a root between y and $y+2$. Therefore we cannot use equation (10.5) until $|\mu|$ is sufficiently large, at least not when $y < 0$ and $\mu \ge 0$.

Lemma 3 implies the nonobvious inequality $A(y,\mu) \ge 0$ for all $y \ge 0$, since $A(y,\mu)$ is proportional to the limiting value of the coefficients of $U(z)^{n-m}/\bigl(1-T(z)\bigr)^y$, and these coefficients are nonnegative. Moreover, $A(y,\mu)$ is strictly positive for $y \ge 2$ and all μ. For if $y \ge 2$ and $A(y,\mu_0) = 0$, we have $B(y,\mu_0) = 0$; but $B'(y,\mu) \ge 0$ by (10.26), hence we must have $B(y,\mu) = 0$ for all $\mu \le \mu_0$, which is impossible because $B(y,\mu)$ is a nonconstant analytic function of μ by (10.2).

When $y = 1$ there is a "closed form" in terms of the Airy function:

$$A(1,\mu) = e^{-\mu^3/12}\mathrm{Ai}(\mu^2/4)\,; \tag{10.28}$$

this is proved in [14, (A.12) and (A.19)]. If we differentiate (10.28) with respect to μ, taking note of the fact that (10.22) gives

$$A(-1,\mu) = -\mu A(0,\mu)\,, \tag{10.29}$$

we find

$$A(0,\mu) = -\tfrac{1}{2}\mu e^{-\mu^3/12}\mathrm{Ai}(\mu^2/4) - e^{-\mu^3/12}\mathrm{Ai}'(\mu^2/4)\,. \tag{10.30}$$

Therefore in particular,

$$e^{\mu^3/12}A(1,\mu) \qquad\qquad \text{and} \qquad\qquad e^{\mu^3/12}\bigl(A(0,\mu)+\tfrac{1}{2}\,\mu\,A(1,\mu)\bigr)$$

are even functions of μ. The well-known relations between Airy functions and Bessel functions,

$$\mathrm{Ai}(z) = \frac{1}{\pi}\sqrt{\frac{z}{3}}\,K_{1/3}\Bigl(\frac{2}{3}z^{3/2}\Bigr), \qquad \mathrm{Ai}'(z) = \frac{-z}{\pi\sqrt{3}}\,K_{2/3}\Bigl(\frac{2}{3}z^{3/2}\Bigr),$$

yield the additional formulas

$$A(1,\mu) = \frac{e^{-\mu^3/12}\mu}{2\pi\sqrt{3}}\,K_{1/3}\Bigl(\frac{\mu^3}{12}\Bigr)\,, \tag{10.31}$$

$$A(0,\mu)+\frac{\mu}{2}A(1,\mu) = A(3,\mu)-\frac{\mu}{2}A(1,\mu) = \frac{e^{-\mu^3/12}\mu^2}{4\pi\sqrt{3}}\,K_{2/3}\Bigl(\frac{\mu^3}{12}\Bigr)\,. \tag{10.32}$$

Since we know $A(y,\mu)$ for $y=-1$, 0, and 1, we can use (10.22) to determine $A(y,\mu)$ for all negative integers y, and for $y=3$ as indicated in (10.32). But a new idea is needed if we hope to have a closed form when $y=2$. It is possible to express $A(2,\mu)$ as an infinite sum of Bessel functions,

$$A(2,\mu) = \frac{1}{3}\biggl(e^{-\mu^3/6} + e^{-\mu^3/12}\Bigl(\sum_{k\ge 0}(-1)^k\Bigl(I_{k+1/3}\Bigl(\frac{\mu^3}{12}\Bigr) - I_{k+2/3}\Bigl(\frac{\mu^3}{12}\Bigr)\Bigr)\Bigr)\biggr)\,, \tag{10.33}$$

but this may be as close to a closed form as possible unless we use general hypergeometric functions. Equation (10.33) follows from (10.2) and the hypergeometric identity

$$\begin{aligned} &F(\tfrac{1}{2}+a,\ 1+2a-b-c;\ 1+2a-b,\ 1+2a-c;\ 2z) \\ &\qquad = \frac{e^z\,\Gamma(a)}{(z/2)^a}\sum_{k\ge 0}\frac{(-1)^k(2a)^{\overline{k}}b^{\overline{k}}c^{\overline{k}}(k+a)I_{k+a}(z)}{(1+2a-b)^{\overline{k}}\,(1+2a-c)^{\overline{k}}k!} \end{aligned} \tag{10.34}$$

[34, equation (2.8)]; here $x^{\overline{k}}$ denotes $\Gamma(x+k)/\Gamma(x)$, and we obtain (10.33) by setting $z=\mu^3/12$, $(a,b,c)=\bigl(\frac{1}{3},\frac{1}{3},1\bigr)$ and $\bigl(\frac{2}{3},\frac{2}{3},1\bigr)$.

The facts that $K_{1/3}(z) = 3^{-1/2}\pi\bigl(I_{-1/3}(z)-I_{1/3}(z)\bigr)$, $K_{2/3}(z) = 3^{-1/2}\pi\bigl(I_{-2/3}(z) - I_{2/3}(z)\bigr)$, and $e^{-z} = I_0(z) + 2\sum_{k\ge1}(-1)^k I_k(z)$ suggest that we look for an identity of the form

$$\begin{aligned} A(y,\mu) &= \Bigl(\frac{\mu}{2}\Bigr)^{2-y} e^{-\mu^3/12}\sum_{k\ge0} a_k(y) I_{(k+y-2)/3}\left(\frac{\mu^3}{12}\right) \\ &= \frac{e^{-\mu^3/6}}{3^{(y-2)/3}}\sum_{k\ge0}\frac{a_k(y)}{\Gamma\bigl(\frac{k+y+1}{3}\bigr)}\Bigl(\frac{\mu}{2\cdot 3^{1/3}}\Bigr)^k F\Bigl(\frac{2k+2y-1}{6};\frac{2k+2y-1}{3};\frac{\mu^3}{6}\Bigr)\,. \end{aligned} \tag{10.35}$$

Any formal power series in μ has such an expansion, for all $y > -1$. But the coefficients $a_k(y)$ do not appear to have a simple form except in the cases already mentioned. We have

$$a_0(y) = \frac{1}{3}\,, \quad a_1(y) = \frac{y-1}{3}\,, \quad a_2(y) = \frac{y(y-3)}{6}\,, \quad a_3(y) = \frac{(y^2-1)(y-6)}{18}$$
$$a_4(y) = \frac{(y-1)(y+2)(y^2-11y+12)}{72}\,, \quad a_5(y) = \frac{(y+3)y(y-1)(y-3)(y-14)}{360}\,.$$

Splitting (10.2) into three sums according to the value of k mod 3 yields a closed form for $A(y,\mu)$ in terms of general hypergeometric series:

$$\begin{aligned} A(y,\mu) = e^{-\mu^3/6}\Biggl(&\frac{1}{3^{(y+1)/3}\,\Gamma\bigl((y{+}1)/3\bigr)}\; F\left(\frac{2{-}y}{6}, \frac{5{-}y}{6}; \frac{1}{3}, \frac{2}{3}; \frac{\mu^3}{6}\right) \\ &+ \frac{1}{3^{(y-1)/3}\,\Gamma\bigl((y{-}1)/3\bigr)}\,\frac{\mu}{2}\, F\left(\frac{4{-}y}{6}, \frac{7{-}y}{6}; \frac{2}{3}, \frac{4}{3}; \frac{\mu^3}{6}\right) \\ &+ \frac{1}{3^{(y-3)/3}\Gamma\bigl((y{-}3)/3\bigr)}\,\frac{\mu^2}{8}\, F\left(\frac{6{-}y}{6}, \frac{9{-}y}{6}; \frac{4}{3}, \frac{5}{3}; \frac{\mu^3}{6}\right)\Biggr)\,. \end{aligned} \tag{10.36}$$

11. Application to bicyclic components. Now we are ready to begin using the basic theoretical results of the preceding sections. We will start by considering the case when the parameter μ of Lemma 3 is very small, say $\mu = O(n^{-1/3})$. Then there are $m = \frac{1}{2}n + O(n^{1/3})$ edges.

Theorem 4. *The probability that a random graph or multigraph with n vertices and $\frac{1}{2}n + O(n^{1/3})$ edges has exactly r bicyclic components, and no components of higher cyclic order, is*

$$\left(\frac{5}{18}\right)^r \sqrt{\frac{2}{3}}\,\frac{1}{(2r)!} + O(n^{-1/3})\,. \tag{11.1}$$

Proof. (The special case $r = 0$ and $m = \frac{1}{2}n$ of this theorem was Corollary 9 of [14].) Consider first the case of random multigraphs, since this case is simpler. If there are n vertices, m edges, r bicyclic components, and no components with higher cyclic order, there must be exactly $n{-}m{+}r$ acyclic components. The probability of such a configuration, according to (2.2), is therefore

$$\frac{2^m\, m!\, n!}{n^{2m}}\,[z^n]\,\frac{U(z)^{n-m+r}}{(n-m+r)!}\,e^{V(z)}\,\frac{W(z)^r}{r!}\,, \tag{11.2}$$

where $U(z)$, $V(z)$, $W(z)$ are the generating functions (3.3), (3.4), and (3.7). Now

$$W(z) = \frac{5}{24}\,\frac{1}{\bigl(1-T(z)\bigr)^3} - \frac{7}{24}\,\frac{1}{\bigl(1-T(z)\bigr)^2} + \frac{1}{12}\,\frac{1}{\bigl(1-T(z)\bigr)}\,, \tag{11.3}$$

using the coefficients e'_{1d} of (7.20); so we see that $W(z)^r$ is a polynomial of degree $3r$ in $\bigl(1-T(z)\bigr)^{-1}$, with leading coefficient $\bigl(\frac{5}{24}\bigr)^r$. Lemma 3 tells us that the leading term of $W(z)^r$ is the only significant one, asymptotically speaking, because the other terms contribute at most $n^{-1/3}$ times as much as the leading term. We can also write

$$U(z)^r = 2^{-r}\bigl(1-\bigl(1-T(z)\bigr)^2\bigr)^r\,; \tag{11.4}$$

this allows us to replace $U(z)^r$ by 2^{-r} in (11.2). Since $e^{V(z)} = \bigl(1-T(z)\bigr)^{-1/2}$, the value of (11.2) is

$$\frac{(n-m)!}{(n-m+r)!\,r!\,2^r}\,\frac{\sqrt{2\pi}\,n^r}{3^{r+1/2}\Gamma(r+1/2)}\left(\frac{5}{24}\right)^r\bigl(1+O(n^{-1/3})\bigr)\,.$$

This simplifies to (11.1) using the fact that

$$\frac{(n-m)!}{(n-m+r)!} = \frac{2^r}{n^r}\bigl(1+O(rn^{-2/3}+r^2n^{-1})\bigr)\,,$$

and using a special case of the duplication formula for the Gamma function,

$$\Gamma(r+1/2) = \frac{(2r)!\,\sqrt{\pi}}{4^r r!}\,. \tag{11.5}$$

On the other hand if we are dealing with random graphs we must replace (11.2) by

$$\frac{n!}{\binom{n(n-1)/2}{m}}\,[z^n]\,\frac{U(z)^{n-m+r}}{(n-m+r)!}\,e^{\widehat{V}(z)}\,\frac{\widehat{W}(z)^r}{r!}\,, \tag{11.6}$$

where $\widehat{V}(z)$ and $\widehat{W}(z)$ appear in (3.5) and (3.6). Again we have $\widehat{W}(z) = \frac{5}{24}\bigl(1-T(z)\bigr)^{-3}$ plus less significant terms, so $\widehat{W}(z)$ produces an effect similar to $W(z)$. But $\widehat{V}(z) = V(z) - \frac{1}{2}T(z) - \frac{1}{4}T(z)^2$; so we now want the coefficient of $[z^n]$ in an expression proportional to

$$\frac{U(z)^{n-m}}{\bigl(1-T(z)\bigr)^{3r+1/2}}\,e^{-T(z)/2-T(z)^2/4}\,,$$

which has an exponential factor not covered by Lemma 3. The proof of Lemma 3 shows, however, that this exponential factor simply changes the result by a factor of $e^{-3/4} + O(n^{-1/3})$: We multiply (10.18) by $\exp(-e^{-s\nu}/2 - e^{-2s\nu}/4) = e^{-3/4} + O(s\nu)$.

Furthermore, (11.6) contains a factor $e^{+3/4}$ to cancel the $e^{-3/4}$, because of (2.4). Therefore the leading term of the asymptotic probability for graphs is the same as it was for multigraphs. □

Corollary. *The probability that a random graph or multigraph with n vertices and $\frac{1}{2}n$ edges has only acyclic, unicyclic, and bicyclic components is*

$$\sqrt{\frac{2}{3}}\cosh\sqrt{\frac{5}{18}}+O(n^{-1/3})\approx 0.9325\,. \tag{11.7}$$

Proof. The sum over r of the estimate made in Theorem 4 clearly gives a lower bound, so we must prove that it is also an upper bound. That sum can be written

$$\frac{2^m\,m!\,n!}{(n-m)!\,n^{2m}}\,[z^n]\,U(z)^{n-m}f_{n-m}(z)\,,$$

where

$$f_l(z)=\sum_{r\ge 0}\frac{l!}{(l+r)!}\,\frac{\bigl(U(z)W(z)\bigr)^r}{r!}\,e^{V(z)}\,. \tag{11.8}$$

If we look at the proof of Theorem 4, and the proof of Lemma 3 on which it is based, we see that the calculations all depend on $f_l(ze^{-z})$, where $|z|\le e^{-\nu}$ and $\nu=n^{-1/3}$. In this region,

$$|T(ze^{-z})|\le e^{-\nu}\,,\qquad |1-T(ze^{-z})|\ge \nu+O(\nu^2)\,. \tag{11.9}$$

Thus the sum $f_{n-m}(ze^{-z})$ converges *uniformly* for all n and all $|z|\le e^{-\nu}$. Uniform convergence allows us to interchange summation and integration. (Notice that the function $h(z)$ in the proof of Lemma 3, which influences the behavior of the integrand most strongly as $n\to\infty$, is independent of r.) $\square$

Another proof of (11.7) will be given below.

12. Components of higher cyclic order. Now let's consider components that are tricyclic, tetracyclic, etc. (Notice that tricyclic components correspond to $C_2(z)$, not $C_3(z)$, in the notation of section 2; our notation has mathematical advantages, but it is slightly out of phase with the traditional terminology.)

Theorem 5. *The probability that a random graph or multigraph with n vertices and $\frac{1}{2}n+O(n^{1/3})$ edges has exactly r_1 bicyclic components, r_2 tricyclic components, …, r_q $(q+1)$-cyclic components, and no components of higher cyclic order, is*

$$\left(\frac{4}{3}\right)^r\sqrt{\frac{2}{3}}\,\frac{c_1^{r_1}}{r_1!}\,\frac{c_2^{r_2}}{r_2!}\,\cdots\,\frac{c_q^{r_q}}{r_q!}\,\frac{r!}{(2r)!}+O(n^{-1/3})\,, \tag{12.1}$$

where $r=r_1+2r_2+\cdots+qr_q$ and the constants c_j are defined in (8.6).

Proof. If there are n vertices and m edges, there must be exactly $n-m+r$ acyclic components. So we can argue as in Theorem 4 to find

$$\begin{aligned}\frac{2^m\,m!\,n!}{n^{2m}}\,[z^n]\,\frac{U(z)^{n-m+r}}{(n-m+r)!}\,e^{V(z)}\,\frac{C_1(z)^{r_1}}{r_1!}\,\frac{C_2(z)^{r_2}}{r_2!}\,\cdots\,\frac{C_q(z)^{r_q}}{r_q!}&\\ =\frac{c_1^{r_1}}{r_1!}\,\frac{c_2^{r_2}}{r_2!}\,\cdots\,\frac{c_q^{r_q}}{r_q!}\,\frac{\sqrt{2\pi}}{3^{r+1/2}\Gamma(r+1/2)}+O(n^{-1/3})\,.&\end{aligned}$$

Formula (12.1) now follows from (11.5) as before. $\square$

Let's illustrate the consequences of Theorem 5 by computing the limiting probabilities for small values of the parameters $(r_1, r_2, \ldots, r_q)$. Here is a list of all configurations with $r_1 + r_2 + \cdots + r_q > 1$ that occur with limiting probability .000005 or more, showing the probabilities rounded to five decimal places:

$$\begin{array}{rrr} [2] = .00263 & [0,2] = .00008 & [1,0,0,0,0,1] = .00002 \\ [1,1] = .00105 & [1,0,0,0,1] = .00004 & [2,1] = .00001 \\ [1,0,1] = .00031 & [0,1,1] = .00003 & [0,1,0,1] = .00001 \\ [1,0,0,1] = .00010 & [3] = .00002 & [1,0,0,0,0,0,1] = .00001 \end{array}$$

(The notation [2] stands for the case $r_1 = 2$, $r_2 = r_3 = \cdots = 0$; similarly $[r_1, \ldots, r_q]$ implies that there are no complex components of cyclic order greater than $q + 1$.)

The sum of these probabilities, .00431, is nicely balanced by $\sqrt{2/3}$ plus the sum of probabilities when there is only one complex component, i.e., when $r_q = 1$ and all other r's are zero:

$$\begin{aligned} &.81650 + .11340 + .03780 + .01547 + .00678 + .00307 + .00141 \\ &\qquad + .00066 + .00031 + .00015 + .00007 + .00003 + .00002 + .00001\,; \end{aligned}$$

this comes to $.99568 = .99999 - .00431$.

Suppose $\mathcal{R}$ is any countably infinite set of configurations $[r_1, r_2, \ldots, r_q]$, where q might be unbounded. We would like to prove that a random graph or multigraph with approximately $\frac{1}{2}n$ edges lies in $\mathcal{R}$ with limiting probability

$$\sum \{\, P[r_1, r_2, \ldots, r_q] \bigm| [r_1, r_2, \ldots, r_q] \in \mathcal{R} \,\}\,, \tag{12.2}$$

where $P[r_1, r_2, \ldots, r_q]$ is the limiting value stated in Theorem 5. The technique we used to prove (11.7) does not apply, because the infinite sums over which integration takes place might not converge uniformly when q is unbounded.

However, we are obviously justified in claiming that (12.2) is a *lower* bound for the stated probability, because the sum over any finite subset of $\mathcal{R}$ yields a lower bound.

We will prove below that the sum of $P[r_1, r_2, \ldots, r_q]$ over all possible configurations $[r_1, r_2, \ldots, r_q]$ is 1. Consequently, the sum (12.2) must in fact be the limiting probability of a random graph or multigraph being in $\mathcal{R}$, not just a lower bound. If (12.2) were too low, we would not obtain 1 by adding the complementary probabilities $P[r_1, r_2, \ldots, r_q]$ for $[r_1, r_2, \ldots, r_q] \notin \mathcal{R}$. This observation will lead to the promised "second proof" of (11.7), if we also sum less significant terms to obtain the error bound $O(n^{-1/3})$.

13. Excess Edges. The notion of "excess" was used somewhat informally in the introductory sections of this paper. Let us now define it formally, saying that the *excess* of a graph or multigraph is the number of edges plus the number of acyclic components, minus the number of vertices. Thus a $(q+1)$-cyclic component has excess q, when $q \geq 0$. If a graph or multigraph has r_1 bicyclic components, r_2 tricyclic components, etc., then it has excess $r = r_1 + 2r_2 + 3r_3 + \cdots$.

If G and G' are graphs on the same vertices, and if $G \cup G'$ and $G \cap G'$ denote the graphs obtained by taking the union and intersection of their edges, the excesses satisfy

$$r(G) + r(G') \leq r(G \cup G') + r(G \cap G')\,.$$

For we can start with empty graphs and insert the edges of $G \cap G'$, preserving equality. Then if we insert an edge of $G \setminus G'$ or of $G' \setminus G$, each side of the inequality increases by either 0 or 1; and the left side cannot increase by 1 unless the right side does also. For example, if the left side increases by 1 when we add an edge of $G \setminus G'$, the endpoints of that edge are in non-trees of G, so they surely are in non-trees of $G \cup G'$.

We have seen in Theorem 5 that the limiting joint probability distribution of the random variables $(r_1, r_2, \ldots)$ in a large random graph or multigraph with approximately $\frac{1}{2}n$ edges has the form

$$\frac{c_1^{r_1}}{r_1!}\,\frac{c_2^{r_2}}{r_2!}\;\cdots\;\frac{c_q^{r_q}}{r_q!}\,f(r)\,, \tag{13.1}$$

where $r = r_1 + 2r_2 + \cdots + qr_q$ is the excess of the graph and $r_l = 0$ for $l > q$. Indeed, this is not surprising, if we look at the problem in another way.

Let $\mathcal{S}$ be the set of all multigraphs of configuration $[r_1, r_2, \ldots, r_q]$, and let $S(w,z)$ be its bgf. The probability that a given multigraph with m edges and n vertices lies in $\mathcal{S}$ is then

$$\Pr_{mn}(\mathcal{S}) = \frac{[w^m z^n]\,S(w,z)}{[w^m z^n]\,G(w,z)}\,. \tag{13.2}$$

We can also express this as

$$\Pr_{mn}(\mathcal{S}) = \Pr_{mn}(\mathcal{S} \mid r)\,\Pr_{mn}(\mathcal{E}_r)\,, \tag{13.3}$$

where $\Pr_{mn}(\mathcal{S} \mid r)$ means the probability of obtaining an element of $\mathcal{S}$ given that the excess is r, and $\Pr_{mn}(\mathcal{E}_r)$ is the probability that a random multigraph has excess r:

$$\Pr_{mn}(\mathcal{S} \mid r) = \frac{[w^m z^n]\,S(w,z)}{[w^m z^n]\,e^{U(w,z)+V(w,z)}E_r(w,z)}\,, \quad \Pr_{mn}(\mathcal{E}_r) = \frac{[w^m z^n]\,e^{U(w,z)+V(w,z)}E_r(w,z)}{[w^m z^n]\,G(w,z)}\,. \tag{13.4}$$

Since all elements of $\mathcal{S}$ have excess r, we can compute $[w^m z^n]\, S(w,z)$ with univariate generating functions:

$$\begin{aligned} S(w,z) &= e^{U(w,z)+V(w,z)}\frac{C_1(w,z)^{r_1}}{r_1!}\frac{C_2(w,z)^{r_2}}{r_2!}\ldots\frac{C_q(w,z)^{r_q}}{r_q!} \\ &= e^{U(wz)/w+V(wz)}\frac{\bigl(wC_1(wz)\bigr)^{r_1}}{r_1!}\frac{\bigl(w^2C_2(wz)\bigr)^{r_2}}{r_2!}\ldots\frac{\bigl(w^qC_q(wz)\bigr)^{r_q}}{r_q!} \\ &= e^{U(wz)/w+V(wz)}w^r\frac{C_1(wz)^{r_1}}{r_1!}\frac{C_2(wz)^{r_2}}{r_2!}\ldots\frac{C_q(wz)^{r_q}}{r_q!}\,; \end{aligned}$$

hence

$$[w^m z^n]\, S(w,z) = [z^n]\,\frac{U(z)^{n+r-m}}{(n+r-m)!}e^{V(z)}S(z)\,, \tag{13.5}$$

if we let

$$S(z) = \frac{C_1(z)^{r_1}}{r_1!}\frac{C_2(z)^{r_2}}{r_2!}\ldots\frac{C_q(z)^{r_q}}{r_q!}\,.$$

Similarly

$$[w^m z^n]\, e^{U(w,z)+V(w,z)}E_r(w,z) = [z^n]\,\frac{U(z)^{n+r-m}}{(n+r-m)!}e^{V(z)}E_r(z)\,.$$

A multigraph with m edges, n vertices, and excess $r>0$ has $t=n+r-m$ components that are trees (including isolated vertices). Suppose it has n_1 vertices in complex components and n_0 vertices in trees and unicyclic components. Then

$$\begin{aligned} \Pr_{mn}(\mathcal{S}\mid r) = \frac{[z^n]\,\dfrac{U(z)^t}{t!}e^{V(t)}S(z)}{[z^n]\,\dfrac{U(z)^t}{t!}e^{V(t)}E_r(z)} &= \frac{\displaystyle\sum_{n_0+n_1=n}\bigl([z^{n_0}]\,U(z)^t e^{V(z)}\bigr)\bigl([z^{n_1}]\,S(z)\bigr)}{\displaystyle\sum_{n_0+n_1=n}\bigl([z^{n_0}]\,U(z)^t e^{V(z)}\bigr)\bigl([z^{n_1}]\,E_r(z)\bigr)} \\ &= \sum_{n_1}\Pr(\mathcal{S}\mid r,n_1)\Pr_{mn}(n_1\mid\mathcal{E}_r)\,, \end{aligned}$$

where

$$\Pr(\mathcal{S}\mid r,n_1) = \frac{[z^{n_1}]\,S(z)}{[z^{n_1}]\,E_r(z)}\,; \tag{13.6}$$

$$\Pr_{mn}(n_1\mid\mathcal{E}_r) = \frac{\bigl([z^{n-n_1}]\,U(z)^t e^{V(z)}\bigr)\bigl([z^{n_1}]\,E_r(z)\bigr)}{[z^n]\,U(z)^t e^{V(z)}E_r(z)}\,. \tag{13.7}$$

Thus, $\Pr(\mathcal{S})$ has been expressed in terms of a simple ratio (13.6), the number of multigraphs consisting of precisely r_j components of excess j for $1\le j\le q$, divided by the number of complex multigraphs of excess r. We know from section 9 that there are coefficients s_d such that

$$S(z)=\frac{s_0T(z)^{2r}}{\bigl(1-T(z)\bigr)^{3r}}+\frac{s_1T(z)^{2r-1}}{\bigl(1-T(z)\bigr)^{3r-1}}+\cdots+\frac{s_{2r-1}T(z)}{\bigl(1-T(z)\bigr)^{r+1}}\,.$$

Indeed, section 9 tells us that s_d is $\sum\kappa(\overline{\overline{M}})/(2r-d)!$, summed over all reduced multigraphs of configuration $[r_1,r_2,\ldots,r_q]$ having exactly $2r-d$ vertices. We can also write

$$S(z)=\frac{s'_0}{\bigl(1-T(z)\bigr)^{3r}}+\frac{s'_1}{\bigl(1-T(z)\bigr)^{3r-1}}+\cdots+\frac{s'_{2r}}{\bigl(1-T(z)\bigr)^{r}}\,, \tag{13.8}$$

letting $s'_d=\sum_k\binom{2r-k}{d-k}(-1)^{d-k}s_k$ as in (7.22). Therefore,

$$n!\,[z^n]\,S(z)=s'_0t_n(3r)+s'_1t_n(3r-1)+\cdots+s'_{2r}t_n(r)\,,$$

expressing the relevant number of multigraphs in terms of the tree polynomials (3.8); and (3.9) tells us that

$$n!\,[z^n]\,S(z)=s'_0\frac{\sqrt{2\pi}\,n^{n-1/2+3r/2}}{2^{3r/2}\Gamma(3r/2)}\bigl(1+O(n^{-1/2})\bigr)\,.$$

Similarly, we have

$$n!\,[z^n]\,E_r(z)=e'_{r0}\frac{\sqrt{2\pi}\,n^{n-1/2+3r/2}}{2^{3r/2}\Gamma(3r/2)}\bigl(1+O(n^{-1/2})\bigr)\,.$$

Therefore the ratio (13.6) is

$$\Pr(\mathcal{S}\mid r,n_1)=\frac{s'_0}{e'_{r0}}\bigl(1+O(n_1^{-1/2})\bigr)\,;$$

and we can sum over n_1 to get

$$\Pr_{mn}(\mathcal{S})=\left(\frac{s'_0}{e'_{r0}}+O(\epsilon)\right)\Pr_{mn}(\mathcal{E}_r)\,, \tag{13.9}$$

where ϵ is the expected value of $n_1^{-1/2}$ in the probability distribution (13.7).

Moreover, the leading coefficient is

$$s'_0=s_0=\frac{c_1^{r_1}}{r_1!}\frac{c_2^{r_2}}{r_2!}\cdots\frac{c_q^{r_q}}{r_q!}\,; \tag{13.10}$$

and e'_{r0} is just e_r, the sum of (13.10) over all configurations $[r_1, r_2, \ldots, r_q]$ with $r_1 + 2r_2 + \cdots + qr_q = r$. This derivation explains why we obtained a formula of the form (13.1) in Theorem 5.

With graphs instead of multigraphs, the same considerations apply, but we must add more terms to the formulas. For example, (13.8) becomes

$$\widehat{S}(z) = \frac{\hat{s}'_0}{\bigl(1 - T(z)\bigr)^{3r}} + \frac{\hat{s}'_1}{\bigl(1 - T(z)\bigr)^{3r-1}} + \cdots + \hat{s}'_{3r} + \hat{s}'_{3r+1}\bigl(1 - T(z)\bigr) + \hat{s}'_{3r+2}\bigl(1 - T(z)\bigr)^2 . \tag{13.11}$$

The leading coefficient $\hat{s}'_0$ is the same as s_0, so the asymptotic behavior is the same as before, if we assume that m is large enough to make the expected value of $n_1^{-1/2}$ approach zero.

We can estimate the expected value of $n_1^{-1/2}$ by finding the expected value of

$$\frac{[z^{n_1}]\, S(z) - (s_0/e_r)E_r(z)}{[z^{n_1}]\, E_r(z)}\,; \tag{13.12}$$

indeed, this expected value is the true error in the approximation (13.9), so it is even more relevant than the expected value of $n_1^{-1/2}$. Since $S(z) - (s_0/e_r)E_r(z)$ can be expressed as $(s'_1 - e'_{r1}s_0/e_r)/\bigl(1 - T(z)\bigr)^{3r-1}$ plus less significant terms, the desired expected value times $\Pr_{mn}(\mathcal{E}_r)$ is obtained by applying Lemma 3 as we did in the proof of Theorem 4, but with $3r$ replaced by $3r - 1$. The result, when m is near $\frac{1}{2}n$, is proportional to $n^{-1/3}$.

The expected value of n_1^k can be computed if we replace $S(z)$ by $\vartheta^k E_r(z)$ in these formulas, because $[z^n]\,\vartheta^k E_r(z) = n^k [z^n]\, E_r(z)$. This has the effect of changing the leading term from $e_r/\bigl(1 - T(z)\bigr)^{3r}$ to $(3r)(3r+2)\ldots(3r+2k-2)e_r/\bigl(1 - T(z)\bigr)^{3r+2k}$, so the result when m is near $\frac{1}{2}n$ is proportional to $n^{2k/3}$. We have proved

Corollary. *If $m = \frac{1}{2}n(1 + \mu n^{-1/3})$ and $|\mu| \le n^{1/12}$, the kth moment $\mathrm{E}_{mn}(n_1^k \mid r)$ of the number of vertices in complex components, given that the total excess is r, is*

$$\alpha_{kr}\frac{\Gamma(r + \frac{1}{2})}{3^{2k/3}\Gamma(r + \frac{1}{2} + \frac{2}{3}k)} n^{2k/3}\bigl(1 + O(\mu) + O(n^{-1/3})\bigr)\,, \qquad \textit{if } \mu = O(1); \tag{13.13}$$

$$\alpha_{kr}\frac{n^{2k/3}}{\mu^{2k}}\bigl(1 + O(|\mu|^{-3}) + O(\mu^4 n^{-1/3})\bigr)\,, \qquad \textit{if } \mu \to -\infty; \tag{13.14}$$

$$\alpha_{kr}\frac{n^{2k/3}\mu^k}{2^k}\frac{\Gamma(\frac{3}{2}r + \frac{1}{4})}{\Gamma(\frac{3}{2}r + \frac{1}{4} + k)}\bigl(1 + O(\mu^{-1}) + O(\mu^4 n^{-1/3})\bigr)\,, \qquad \textit{if } \mu \to +\infty; \tag{13.15}$$

here $\alpha_{kr} = (3r)(3r+2)\ldots(3r+2k-2)$.

Proof. These expressions are α_{kr} times the ratios of formulas (10.2), (10.3), and (10.4) when $y = 3r + \frac{1}{2} + 2k$ to their values when $y = 3r + \frac{1}{2}$. $\square$

Notice that when m is approximately $\frac{1}{2}n - n^{3/4}$, the probable value of n_1 is proportional to $n^{2/3-2/12} = n^{1/2}$; when $m \approx \frac{1}{2}n + n^{3/4}$, it is proportional to $n^{2/3+1/12} = n^{3/4}$. These are the extreme cases $|\mu| = n^{1/12}$ at the limits of Lemma 3's range.

We can use formula (13.3) whenever $\mathcal{S}$ is a collection of multigraphs whose complex components have total excess r. We can use formula (13.6) whenever $\mathcal{S}$ also places no restriction on its non-complex (acyclic and unicyclic) components. For example, we can determine the conditional probability that a random graph with $\frac{1}{2}n$ edges has a bicyclic component of each of the three types in (9.15), given that it has excess 1. The generating functions $S(z)$ for the three cases are respectively $\frac{1}{8}T/(1-T)^2$, $\frac{1}{8}T^2/(1-T)^3$, $\frac{1}{12}T^2/(1-T)^3$; so the respective conditional probabilities are

$$O(n^{-1/3}), \qquad \frac{3}{5} + O(n^{-1/3}), \qquad \frac{2}{5} + O(n^{-1/3}) . \tag{13.16}$$

All probabilities that are conditional on excess r must, of course, be multiplied by $\Pr_{mn}(\mathcal{E}_r)$, the probability that a random multigraph has excess r. Lemma 3 and the method of Theorem 5 make this easy to compute:

Corollary. *A graph or multigraph with $m = \frac{1}{2}n(1 + \mu n^{-1/3})$ edges and n vertices has excess r with probability*

$$\Pr_{mn}(\mathcal{E}_r) = \sqrt{2\pi}\, e_r\, A(3r + \tfrac{1}{2}, \mu) + O\biggl(\frac{1+\mu^4}{n^{1/3}}\biggr) , \tag{13.17}$$

uniformly for $|\mu| \le n^{1/12}$ as $n \to \infty$, where $e_r = e_{r0}$ is given by (7.2) and $A(y, \mu)$ is given by (10.2). When $\mu \to -\infty$, the probability is $O(|\mu|^{-3r})$; when $\mu \to +\infty$ it is $O(\mu^{3r/2}e^{-\mu^3/6})$. □

(The special case $r = 0$ in (13.17), without the error bound, was found by Britikov [9], who proved that a random graph has excess 0 with probability approaching $\sqrt{2\pi}\,A(\frac{1}{2}, \mu)$, for fixed μ as $n \to \infty$.)

Here is a table that shows how the probabilities of having excess r change as the graph or multigraph evolves past the critical point $m = \frac{1}{2}n$:

	$r=0$	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
$\mu=-3$	.994	.006	.000	.000	.000	.000	.000	.000	.000	.000	.000
$\mu=-2$	.983	.015	.001	.000	.000	.000	.000	.000	.000	.000	.000
$\mu=-1$	.947	.043	.008	.002	.000	.000	.000	.000	.000	.000	.000
$\mu=0$	.816	.113	.040	.017	.007	.003	.001	.001	.000	.000	.000
$\mu=1$	.475	.179	.115	.077	.052	.035	.023	.015	.010	.007	.004
$\mu=2$	.100	.082	.085	.086	.084	.079	.073	.066	.058	.051	.043
$\mu=3$	.003	.004	.007	.010	.013	.017	.020	.024	.028	.031	.034

The mean excess is approximately .308, 1.544, 6.364, 19.009, for $\mu = 0, 1, 2, 3$.

In this paper we are interested mainly in graphs or multigraphs with approximately $\frac{1}{2}n$ edges, but it is instructive to consider also the formulas that arise when m is somewhat smaller. The excess is then almost surely zero. In fact, we can obtain a formula that has a much better error bound than (13.17), in the case $r = 0$ and $\mu < -n^{-\epsilon}$: If we set $\lambda = 2m/n$, and if $m < \frac{1}{2}n - n^{2/3+\epsilon}$, the probability of excess 0 can be shown to be exactly

$$\frac{2^m m!\, n!}{(n-m)!\, n^{2m}} \, [z^n] \, \frac{U(z)^{n-m}}{\bigl(1-T(z)\bigr)^{1/2}}$$

$$= \frac{S(m)S(n)}{\sqrt{2\pi}\, S(n-m)} \oint \left(1 - \frac{it\beta}{1-\lambda}\right)^{1/2} \left(1 + \frac{it\beta}{\lambda}\right)^{-1} e^{h(n,\lambda,t)-t^2/2}\, dt\,, \tag{13.18}$$

where

$$S(n) = \frac{n!\, e^n}{n^n \sqrt{2\pi n}} = 1 + O\left(\frac{1}{n}\right)\,, \tag{13.19}$$

$$\beta = \sqrt{\frac{\lambda(2-\lambda)}{(1-\lambda)n}}\,, \tag{13.20}$$

$$h(n,\lambda,t) = n\,h(\lambda + it\beta) - n\,h(\lambda) + t^2/2 \tag{13.21}$$

$$= \frac{n}{2} \sum_{k\ge 3} \frac{(it\beta)^k}{k} \left(\frac{(-1)^k}{\lambda^{k-1}} - \frac{1}{(2-\lambda)^{k-1}}\right)\,, \tag{13.22}$$

and the contour of integration makes $z = \lambda + it\beta$ traverse the circle $|z| = \lambda$ as t varies. The function $h(z)$ in (13.21) is the function defined in (10.12). We are essentially simplifying the proof of Lemma 3 by choosing a path of integration through the saddle point $z = \lambda$, as in the proof of Theorem 4 in [14]. The proof of that theorem justifies restricting t to a neighborhood of zero, so that the tail-exchange method can be applied as in the derivation following (10.9). It follows that the probability of excess 0 is $1 - O\bigl(n^2/(n-2m)^3\bigr)$ for m in the stated range. We have in fact the estimate

$$\Pr_{mn}(\mathcal{E}_0) = 1 - \frac{5}{24}\,\alpha^{-3}\bigl(1 + O(\alpha^{-3}) + O(\alpha n^{-1/3})\bigr) \tag{13.23}$$

when $m = \frac{1}{2}n(1 - \alpha\, n^{-1/3})$, uniformly for $(\ln n)^2 \le \alpha \le \frac{1}{2}n^{1/3}$.

It is interesting to note that the tail-exchange method can be used to extend (13.23) to an asymptotic series in α^{-1} and $\alpha n^{-1/3}$, although the integral (13.18) actually diverges if we let t run through all real values from $-\infty$ to $+\infty$ instead of describing the stated contour. Indeed, the magnitude of the integrand in (13.18) for large real values of $|t|$ is approximately $|t|^{n-2m-1/2}$.

14. Probability distribution of the excess. One way to check our calculations is to verify that the probabilities in (13.17) sum to 1. Thus we want to prove that

$$\sum_{k\geq 0}\sqrt{\frac{2\pi}{3}}\,\frac{(\frac{1}{2}\,3^{2/3}\mu)^k}{k!}\sum_{r\geq 0}\frac{(6r)!}{2^{5r}3^{3r}(2r)!\,(3r)!\,\Gamma(r+1/2-2k/3)}=e^{\mu^3/6}\,. \tag{14.1}$$

The inner sum is a hypergeometric series whose sum is known;

$$\frac{1}{\Gamma(1/2-2k/3)}\,F\left(\frac{1}{6},\frac{5}{6}\,;\,\frac{1}{2}-\frac{2k}{3}\,;\,\frac{1}{2}\right)=\frac{2^{1/2+2k/3}\sqrt{\pi}}{\Gamma\big((1-k)/3\big)\,\Gamma\big((2-k)/3\big)}\,. \tag{14.2}$$

Indeed, the special hypergeometric

$$f(a,b,z)=F(a,1-a;b;z)\,,$$

which is related to a Legendre function, satisfies

$$\frac{f(a,b,\frac{1}{2})}{\Gamma(b)}=\frac{2^{1-b}\sqrt{\pi}}{\Gamma\big(\frac{1}{2}(a+b)\big)\,\Gamma\big(\frac{1}{2}(1-a+b)\big)}\,. \tag{14.3}$$

This well-known relation can be obtained by applying Euler's identity

$$F(a,b;c;z)=(1-z)^{c-a-b}F(c-a,c-b;c;z)$$

and Gauss's identities

$$F(a,b;c;1)=\big(\Gamma(c-a-b)\Gamma(c)\big)/\big(\Gamma(c-a)\Gamma(c-b)\big)\,,$$
$$F(2a,2b;a+b+\frac{1}{2};z)=F\big(a,b;a+b+\frac{1}{2};4z(1-z)\big)\,,$$

which can be found, for example, in [17, (5.92), (5.111), exercise 5.28]):

$$(1-z)^{1-b}F(a,1-a;b;z)=F(b-a,b+a-1;b;z)=F\big(\tfrac{1}{2}b-\tfrac{1}{2}a,\tfrac{1}{2}b+\tfrac{1}{2}a-\tfrac{1}{2};b;4z(1-z)\big)\,;$$

we obtain (14.3) by letting $z\to\frac{1}{2}$.

The sum (14.2) vanishes except when $k=3m$, and in this case the kth term on the left of (14.1) reduces to simply $(\mu^3/6)^m/m!$ because of the formula

$$\Gamma(\tfrac{1}{3}-m)\,\Gamma(\tfrac{2}{3}-m)=3^{3m-1/2}2\pi\,\frac{m!}{(3m)!}\,. \tag{14.4}$$

Hence (14.1) is true. It is remarkable that so much of nineteenth century mathematics has turned out to be relevant to the study of random graphs.

When $\mu = 0$, the generating function for the limiting probabilities of excess r turns out to have a closed form: It is

$$\sum_{r\ge 0}\left(\frac{4}{3}\right)^r\sqrt{\frac{2}{3}}\,\frac{e_r r!\,z^r}{(2r)!} = \sqrt{\frac{2}{3}}\,F\left(\frac{1}{6},\frac{5}{6};\frac{1}{2};\frac{z}{2}\right) = \frac{2\cos\bigl(\frac{2}{3}\arcsin\sqrt{z/2}\,\bigr)}{\sqrt{6-3z}}\,. \tag{14.5}$$

From this expression it is easy to calculate the limiting value of the mean excess when $m = \frac{1}{2}n$, namely $\frac{1}{2} - 3^{-3/2} \approx 0.308$. The variance, similarly, is $\frac{23}{27} - 3^{-3/2}$.

The limiting mean excess when the number of edges is $\frac{1}{2}n(1+\mu n^{-1/3})$ does not seem to have a simple closed form, although we can express it as a hypergeometric series and find the asymptotic value. Suppose we insert the factor z^r into the left-hand side of (14.1). Then the left-hand side of (14.2) becomes

$$\frac{1}{\Gamma(1/2-2k/3)}\,F\left(\frac{1}{6},\frac{5}{6};\frac{1}{2}-\frac{2k}{3};\frac{z}{2}\right) = \frac{1}{\Gamma(1/2-2k/3)}\,f\left(\frac{1}{6},\frac{1}{2}-\frac{2k}{3},\frac{z}{2}\right). \tag{14.6}$$

To evaluate the derivative of such a function at $\frac{1}{2}$, we can use the identity

$$z(1-z)f'(a,b,z) = \left(az + \frac{1-a-b}{2}\right)f(a,b,z) - \frac{1-a-b}{2}\,f(-a,b,z)\,, \tag{14.7}$$

which is readily verified by checking that the coefficients of z^n agree on both sides. To get the mean value of r, we want to differentiate (14.6) with respect to z and set $z=1$; and according to (14.7), this is equivalent to replacing (14.6) by

$$\frac{1}{\Gamma(1/2-2k/3)}\Bigl(\bigl(\tfrac{1}{2}+\tfrac{2k}{3}\bigr)f\bigl(\tfrac{1}{6},\tfrac{1}{2}-\tfrac{2k}{3},\tfrac{1}{2}\bigr) - \bigl(\tfrac{1}{3}+\tfrac{2k}{3}\bigr)f\bigl(-\tfrac{1}{6},\tfrac{1}{2}-\tfrac{2k}{3},\tfrac{1}{2}\bigr)\Bigr)\,. \tag{14.8}$$

Again, $f(\frac{1}{6},\frac{1}{2}-\frac{2k}{3},\frac{1}{2})$ vanishes unless $k=3m$. The contribution to the mean from this half of (14.8) is just what we had when we were summing the probabilities, but with an additional factor of $(\frac{1}{2}+\frac{2k}{3})$; so it is

$$e^{-\mu^3/6}\sum_{m\ge 0}\bigl(\tfrac{1}{2}+2m\bigr)\frac{(\mu^3/6)^m}{m!} = \frac{1}{2}+\frac{\mu^3}{3}\,. \tag{14.9}$$

The other half of (14.8) is, however, more complicated, since all values of k make a contribution. According to (14.3), we want to evaluate

$$\sum_{k\ge 0}\sqrt{\frac{2\pi}{3}}\,\frac{(\frac{1}{2}3^{2/3}\mu)^k}{k!}\,\frac{2^{1/2+2k/3}\sqrt{\pi}\,(1/3+2k/3)}{\Gamma(1/6-k/3)\,\Gamma(5/6-k/3)} = \Sigma_0+\Sigma_1+\Sigma_2\,,$$

where Σ_j is a hypergeometric series corresponding to $k = 3m + j$:

$$\Sigma_0 = \frac{1}{3\sqrt{3}} F\left(\frac{5}{6}, \frac{7}{6}; \frac{1}{3}, \frac{2}{3}; \frac{\mu^3}{6}\right);$$
$$\Sigma_1 = -\frac{1}{\sqrt{3}} \frac{\mu\sqrt{\pi}}{6^{1/3}\Gamma(5/6)} F\left(\frac{7}{6}, \frac{3}{2}; \frac{2}{3}, \frac{4}{3}; \frac{\mu^3}{6}\right);$$
$$\Sigma_2 = -\frac{5\sqrt{3}}{2} \frac{\mu^2}{6^{2/3}} \frac{\sqrt{\pi}}{\Gamma(1/6)} F\left(\frac{3}{2}, \frac{11}{6}; \frac{4}{3}, \frac{5}{3}; \frac{\mu^3}{6}\right). \tag{14.10}$$

As $z \to +\infty$, such hypergeometric series satisfy the asymptotic formula

$$F(a,b;c,d;z) = \frac{\Gamma(c)\Gamma(d)}{\Gamma(a)\Gamma(b)} z^\delta e^z \left(1 + \frac{\delta(a+b-1) - ab + cd}{z} + O(z^{-2})\right), \tag{14.11}$$

where $\delta = a + b - c - d$; this follows by plugging the right-hand side into the differential equation

$$\vartheta(\vartheta + c - 1)(\vartheta + d - 1)F = z(\vartheta + a)(\vartheta + b)F$$

satisfied by the left. We obtain

$$e^{-\mu^3/6}\Sigma_0 = \frac{1}{3\sqrt{3}} \frac{\Gamma(\frac{1}{3})\Gamma(\frac{2}{3})}{\Gamma(\frac{5}{6})\Gamma(\frac{7}{6})} \left(\frac{\mu^3}{6} + \frac{1}{4} + O(\mu^{-3})\right);$$
$$e^{-\mu^3/6}\Sigma_1 = -\frac{1}{\sqrt{3}} \frac{\sqrt{\pi}}{\Gamma(\frac{5}{6})} \frac{\Gamma(\frac{2}{3})\Gamma(\frac{4}{3})}{\Gamma(\frac{7}{6})\Gamma(\frac{3}{2})} \left(\frac{\mu^3}{6} + \frac{1}{4} + O(\mu^{-3})\right);$$
$$e^{-\mu^3/6}\Sigma_2 = -\frac{5\sqrt{3}}{2} \frac{\sqrt{\pi}}{\Gamma(\frac{1}{6})} \frac{\Gamma(\frac{4}{3})\Gamma(\frac{5}{3})}{\Gamma(\frac{3}{2})\Gamma(\frac{11}{6})} \left(\frac{\mu^3}{6} + \frac{1}{4} + O(\mu^{-3})\right); \tag{14.12}$$

therefore $e^{-\mu^3/6}(\Sigma_0 + \Sigma_1 + \Sigma_2) = (\frac{2}{3} - \frac{4}{3} - \frac{4}{3})(\frac{\mu^3}{6} + \frac{1}{4} + O(\mu^{-3}))$. Subtracting this from (14.9), and using computer algebra to refine the estimate further, gives us the answer we seek:

Theorem 6. *The expected value of the excess, when there are $\frac{1}{2}n(1 + \mu n^{-1/3})$ edges, approaches*

$$\frac{2}{3}\mu^3 + 1 + \frac{5}{24}\mu^{-3} + \frac{15}{16}\mu^{-6} + O(\mu^{-9}), \tag{14.13}$$

for fixed $\mu \ge \delta > 0$ as $n \to \infty$. □

This method of calculation shows also that the variance will be $O(\mu^6)$ and the kth moment will be $O(\mu^{3k})$; each derivative of (14.6) can do no worse than multiply by μ^3, because of (14.7).

Incidentally, the $O(\mu^{-3})$ terms in all three equations of (14.12) turn out to equal $\frac{5}{48}\mu^{-3} + \frac{15}{32}\mu^{-6} + O(\mu^{-9})$, and this is no coincidence. We have, in fact,

$$\frac{\Gamma(\frac{5}{6})\Gamma(\frac{7}{6})}{\Gamma(\frac{1}{3})\Gamma(\frac{2}{3})} F\left(\frac{5}{6}, \frac{7}{6}; \frac{1}{3}, \frac{2}{3}; z^3\right) \sim \frac{\Gamma(\frac{7}{6})\Gamma(\frac{9}{6})}{\Gamma(\frac{2}{3})\Gamma(\frac{4}{3})} z\, F\left(\frac{7}{6}, \frac{3}{2}; \frac{2}{3}, \frac{4}{3}; z^3\right)$$
$$\sim \frac{\Gamma(\frac{3}{2})\Gamma(\frac{11}{6})}{\Gamma(\frac{4}{3})\Gamma(\frac{5}{3})} z^2 F\left(\frac{3}{2}, \frac{11}{6}; \frac{4}{3}, \frac{5}{3}; z^3\right), \tag{14.14}$$

in the sense that all three functions have the same asymptotic series $\sum s_k z^{-3k} e^{z^3}$ as $z \to \infty$. This follows because all three functions satisfy the same differential equation, and because their asymptotic behavior depends only on the differential equation except for a constant of proportionality. It is well known that the general hypergeometric functions $F(a_1, \ldots, a_m;\, b_1, \ldots, b_n;\, z)/\Gamma(b_1) \ldots \Gamma(b_n)$ and $z^{1-b_1} F(a_1 + 1 - b_1, \ldots, a_m + 1 - b_1;\, b_2 + 1 - b_1, \ldots, b_n + 1 - b_1, 2 - b_1;\, z)/\Gamma(b_2 + 1 - b_1) \ldots \Gamma(b_n + 1 - b_1)$ both satisfy the differential equation $\vartheta(\vartheta + b_1 - 1) \ldots (\vartheta + b_n - 1)F = z(\vartheta + a_1) \ldots (\vartheta + a_m)F$. In the case of (14.14), even more is true: The three asymptotically equivalent functions shown there can be written respectively as $\frac{1}{3}\bigl(G(z) + G(\omega z) + G(\omega^2 z)\bigr)$, $\frac{1}{3}\bigl(G(z) + \omega^2 G(\omega z) + \omega G(\omega^2 z)\bigr)$, $\frac{1}{3}\bigl(G(z) + \omega G(\omega z) + \omega^2 G(\omega^2 z)\bigr)$, where

$$G(z) = \frac{1}{\sqrt{3}} \sum_{k \ge 0} \frac{\Gamma(3/2 + k) z^k}{\Gamma(1/2 + k/3)\, k!} \tag{14.15}$$

and $\omega = e^{2\pi i/3}$.

15. Deficiency, planarity, and complexity. The calculations in the preceding section can be combined with the structure theory of section 9 to yield the following general result.

Theorem 7. *Let $\overline{\overline{M}}$ be a reduced multigraph of excess r and deficiency d, i.e., a reduced multigraph having $2r - d$ vertices and $3r - d$ edges. The probability that the complex part of a random graph or multigraph reduces to $\overline{\overline{M}}$ is asymptotically*

$$\frac{\sqrt{2\pi}\, \kappa(\overline{\overline{M}})}{(2r - d)!} A(3r + \tfrac{1}{2} - d, \mu)\, n^{-d/3} \tag{15.1}$$

when there are $\frac{1}{2}n(1 + \mu n^{-1/3})$ edges and n vertices, $|\mu| = o(n^{1/12})$. Here $\kappa(\overline{\overline{M}})$ denotes the compensation factor (1.1), and $A(y, \mu)$ is defined in (10.2). The sum of (15.1) over all $\overline{\overline{M}}$ of deficiency 0 is 1. For each $d \ge 0$, the probability that a random multigraph has deficiency $\ge d$ is $O\bigl((1 + \mu^4)^d n^{-d/3}\bigr)$, uniformly in n and μ.

Proof. When $d = 0$, this theorem is a consequence of the corollary following (9.16), together with (13.17) and (14.1).

When $d > 0$, (15.1) is clear, but we need two auxiliary results of independent interest before we can prove the desired uniform estimate.

Lemma 4. *Let $E_{r(\geq d)}$ denote the generating function for all complex multigraphs of excess r whose deficiency is at least d. Then*

$$\frac{e_{rd}T}{(1-T)^{3r-d}} - \frac{(2r-d-1)e_{rd}T}{(1-T)^{3r-d-1}} \leq E_{r(\geq d)} \leq \frac{e_{rd}T}{(1-T)^{3r-d}}\,, \tag{15.2}$$

where inequality between generating functions means that the coefficients of every power of z obey the stated relation.

Proof. The claim is trivial when $d = 2r-1$; and it is true when $r = 1$, because $E_1 = \frac{5}{24}T(1-T)^{-3} - \frac{1}{12}T(1-T)^{-2}$. The lower bound is easily seen to be a lower bound on $e_{rd}T^{2r-d}/(1-T)^{3r-d}$ itself.

The proof of the upper bound now proceeds by induction on r. Let

$$E'_r = \sum_{k=0}^{d-1} e_{rk}(1+\zeta)^r\zeta^{2r-k} + e_{rd}\zeta(1+\zeta)^{3r-d-1}\,, \tag{15.3}$$

in the notation of section 5. We want to prove that $E_r \leq E'_r$; it suffices to show, by (5.8), that

$$\big(r + (1-T(z))\vartheta\big)E'_r = \big(r + (1+\zeta)^{-1}\vartheta\big)E'_r \geq \tfrac{1}{2}\left(\tfrac{1}{2}\,\zeta(1+\zeta) + \vartheta\right)^2 E'_{r-1}\,, \tag{15.4}$$

considering both sides as generating functions in powers of z. Proceeding as in (5.11) and (5.12) to form

$$A'_r = \big(\tfrac{1}{2}\,\zeta(1+\zeta) + \vartheta\big)E'_r\,, \qquad B'_r = \big(\tfrac{1}{2}\,\zeta(1+\zeta) + \vartheta\big)A'_r\,,$$

a bit of algebra shows that when $0 \leq d \leq 2r-3$ we have

$$\begin{aligned}
&\big(r + (1+\zeta)^{-1}\vartheta\big)E'_r - \tfrac{1}{2}B'_{r-1} \\
&\quad = \frac{1}{2}\zeta(1+\zeta)^{r+1}\bigg(\sum_{k\geq 0}\zeta^{2r-d-2-k}\big((\alpha_k+\beta_k)e_{rd} - (\gamma_k+\delta_k+\epsilon_k)e_{(r-1)d}\big) \\
&\qquad\qquad - (2r-1-d)^2 e_{(r-1)(d-1)}\zeta^{2r-d-2}\bigg)\,,
\end{aligned} \tag{15.5}$$

where

$$\alpha_k = 2(3r-d)\binom{2r-d-2}{k+1}\,, \qquad \beta_k = 2(r+1)\binom{2r-d-2}{k}\,;$$

$$\gamma_k = \big(3r-\tfrac{1}{2}-d\big)\big(3r-\tfrac{5}{2}-d\big)\binom{2r-d-3}{k+1}\,, \qquad \delta_k = \big(9r-\tfrac{13}{2}-3d\big)\binom{2r-d-3}{k}\,,$$

$$\epsilon_k = \binom{2r-d-3}{k-1}\,.$$

Obviously $\beta_k e_{rd} \ge \epsilon_k e_{(r-1)d}$, since $e_{rd} \ge e_{(r-1)d}$. And the inequality $9r - \frac{13}{2} - 3d \le (3r - \frac{1}{2} - d)(3r - \frac{5}{2} - d)$ for $0 \le d \le 2r-3$ yields

$$(\gamma_k + \delta_k) e_{(r-1)d} \le \bigl(3r - \tfrac{1}{2} - d\bigr)\bigl(3r - \tfrac{5}{2} - d\bigr)\binom{2r-d-2}{k+1} e_{(r-1)d} \le \alpha_k e_{rd}\,.$$

In fact, (5.11)–(5.13) imply that

$$2(3r-d)e_{rd} \ge \bigl(3r - \tfrac{1}{2} - d\bigr)\bigl(3r - \tfrac{5}{2} - d\bigr) e_{(r-1)d} + (2r-d)(2r-1-d)e_{(r-1)(d-1)}\,;$$

so (15.5) is a polynomial in ζ with nonnegative coefficients, and thus a power series in z with nonnegative coefficients, proving (15.4). The case $d = 2r-2$ needs to be handled separately, but it offers no difficulty. $\square$

Lemma 5. *There exists a constant $\epsilon > 0$ such that, for every fixed $d \ge 0$, a random multigraph with n vertices and $m = \frac{n}{2}(1 + \mu n^{-1/3})$ edges has excess r and deficiency $\ge d$ with probability*

$$\begin{cases} O(\mu^{4d-3/2} n^{-d/3} e^{-\epsilon(r - \frac{2}{3}\mu^3)^2/\mu^3}), & \text{if } r \le \mu^3, \\ O(n^{-d/3} e^{-\epsilon r}), & \text{if } r \ge \mu^3, \end{cases}$$

uniformly in n, r, and μ when $\mu \le n^{1/12}$.

Proof. Let $p_{rd} = p_{rd}(n, \mu)$ be the stated probability. It suffices to prove the lemma when $\mu \ge 1$ and $r \ge 1$. For if $r = d = 0$, the result follows from Lemma 3; and $p_{0d} = 0$ when $d > 0$. On the other hand, if $\mu < 1$ we have $p_{rd}(n,\mu) \le \sum_{j=r}^{\infty} p_{jd}(n,1)$.

Using Lemma 4 and arguing as in the proof of Lemma 3, equation (10.11), we obtain

$$\begin{aligned} p_{rd} &= \frac{2^m m!\, n!}{n^{2m}}\,[z^n]\,\frac{U^{n-m+r}}{(n-m+r)!}\,e^V\,E_{r(\ge d)} \\ &\le \frac{2^m m!\, n!}{n^{2m}}\,[z^n]\,\frac{U^{n-m+r}}{(n-m+r)!}\,\frac{e_{rd}\,T}{(1-T)^{3r-d+1/2}} \\ &= \frac{2^m m!\, n!\, e_{rd}\, e^n\, 2^{m-n-r}}{(n-m+r)!\, n^{2m}\, 2\pi i}\oint \left(\frac{z(2-z)}{1-z}\right)^r (1-z)^{d-2r+1/2} e^{nh(z)}\,dz\,, \end{aligned}$$

with $h(z)$ as in (10.12), and where the integral is taken around a circle $z = \rho e^{i\theta}$ with $0 < \rho < 1$. On this circle, both $\bigl|(2-z)/(1-z)\bigr|$ and $|1-z|^{-1}$ attain their maxima at $z = \rho$. Moreover, by (10.16) we have $\frac{d}{d\theta}\Re h = -\rho g(\theta)\sin\theta$, where $g(\theta) > \bigl((2-\rho)^2 - 1\bigr)/9 > \frac{2}{9}(1-\rho)$; therefore

$$\Re h(\rho e^{i\theta}) \le h(\rho) + \frac{2}{9}(1-\rho)\rho(\cos\theta - 1) \le h(\rho) - \frac{4}{9\pi^2}\rho(1-\rho)\theta^2, \qquad \text{for } |\theta| \le \pi.$$

Now $p_{rd} = 0$ if $d \ge 2r$, because we are assuming that $r \ge 1$. Hence $d - 2r + 1/2 < 0$, and the contour integral including the factor $1/(2\pi i)$ is less than

$$\begin{aligned} &\frac{\rho}{2\pi}\left(\frac{\rho(2-\rho)}{1-\rho}\right)^r (1-\rho)^{d-2r+1/2} e^{nh(\rho)} \int_{-\pi}^{\pi} \exp\left(-\frac{4n\rho(1-\rho)}{9\pi^2}\theta^2\right) d\theta \\ &\qquad < \frac{3}{4}\sqrt{\frac{\pi}{n}}\rho^{r+1/2}(2-\rho)^r(1-\rho)^{d-3r}e^{nh(\rho)}\,. \end{aligned}$$

In the following argument, unspecified positive constants will be denoted by ϵ_1, ϵ_2, $\ldots$, while positive numbers that may depend on d will be denoted by C_1, C_2, $\ldots$. Let $\nu = n^{-1/3}$. If we apply (10.10) to the coefficient in front of the contour integral, and if we use the estimate

$$\frac{(n-m+r)!}{(n-m)!} > (n-m)^r = \left(\frac{n}{2}\right)^r (1-\mu\nu)^r > \left(\frac{n}{2}\right)^r e^{-2\mu\nu r},$$

which is valid when $\mu\nu \le \frac{1}{2}$, we obtain the upper bound

$$p_{rd} \le C_1 e_{rd} n^{-r} \rho^r (2-\rho)^r (1-\rho)^{d-3r} e^{nh(\rho)-\mu^3/6+2\mu\nu r}, \tag{15.6}$$

where ρ is any number between 0 and 1.

Suppose now that $r \le 12\mu^3$, and set $\rho = 1 - \xi\mu\nu$, $r = \frac{2}{3}x\mu^3$. If $\xi = O(1)$, we have

$$nh(1-\xi\mu\nu) = \frac{1}{3}\xi^3\mu^3 + \frac{1}{2}\xi^2\mu^3 + O(1)$$

as in (10.17). Therefore, since $\rho(2-\rho) < 1$,

$$\begin{aligned} p_{rd} &\le C_2 e_{rd} n^{-r} (\xi\mu\nu)^{d-3r} e^{\xi^3\mu^3/3+\xi^2\mu^3/2-\mu^3/6} \\ &\le C_3 3^r 2^{-r} r^{r+d-1/2} e^{-r} n^{-d/3} (\xi\mu)^{d-3r} e^{\xi^3\mu^3/3+\xi^2\mu^3/2-\mu^3/6} \\ &= C_3 r^{d-1/2} n^{-d/3} (\xi\mu)^d e^{k(x,\xi)\mu^3/6}, \quad k(x,\xi) = 2\xi^3 + 3\xi^2 - 1 + 4x\ln\left(\frac{x}{e\xi^3}\right), \end{aligned}$$

by (7.16) and Stirling's formula. Given x between 0 and 18, we minimize $k(x,\xi)$ by letting ξ be the positive root of $\xi^3 + \xi^2 = 2x$; notice that this makes $\xi \le 3$, justifying our assumption that $\xi = O(1)$. The minimum $k(x,\xi)$ satisfies

$$\begin{aligned} k(x,\xi) &= \xi^2 - 1 + 2(\xi^3+\xi^2)\ln\left(\frac{1+\xi^{-1}}{2}\right) \\ &\le \xi^2 - 1 + 2(\xi^3+\xi^2)\,\frac{\xi^{-1}-1}{2} = -(\xi+1)(\xi-1)^2. \end{aligned}$$

We also have $|\xi - 1| \geq \epsilon_1 |x-1|$, hence $k(x,\xi) \leq -\epsilon_2(x-1)^2$. Our estimates have shown that

$$p_{rd} \leq C_5 r^{d-1/2} n^{-d/3} \mu^d e^{-\epsilon_2(x-1)^2\mu^3/6},$$

when $r = \frac{2}{3}x\mu^3 \leq 12\mu^3$, so the first half of the lemma has been proved.

When $r \geq 12\mu^3$, let us set $\rho = 1-\eta$ and $r = yn$. In this case we will in fact prove the lemma for a much larger range of μ, assuming only that $\mu\nu \leq \delta$, when δ is a suitably small constant. If $\delta \leq \frac{1}{5}$ we can assume that $0 < y < \frac{3}{5}$, since m is at most $\frac{1+\delta}{2}n$ and since $p_{rd} = 0$ when $r \geq m$. Using (7.16) and (15.6) again, we find

$$p_{rd} \leq C_6 r^{d-1/2} \eta^d e^{nl(y,\eta) - \mu^3/6 + 2r\mu\nu},$$

where

$$\begin{aligned} l(y,\eta) &= y \ln\left(\frac{3y(1-\eta^2)}{2e\eta^3}\right) + h(1-\eta) \\ &= y \ln\left(\frac{3y(1-\eta^2)}{2e\eta^3}\right) - \eta - \ln(1-\eta) + \frac{1-\mu\nu}{2}\ln(1-\eta^2)\,. \end{aligned}$$

Given y, the minimum value of $l(y,\eta)$ occurs when $y = \eta^2(\eta+\mu\nu)/(3-\eta^2)$. However, we do not need to find the exact minimum, in order to achieve the upper bound in the lemma; it will suffice to be close to the minimum when y is small. Therefore we choose η in such a way that the calculations will be relatively simple:

$$y \;=\; \frac{2\eta^3}{3(1-\eta^2)}\,. \tag{15.7}$$

With this choice, we always have $\eta < \frac{3}{4}$; and

$$l(y,\eta) = f(\eta) = -\frac{2\eta^3}{3(1-\eta^2)} - \eta - \ln(1-\eta) + \frac{1-\mu\nu}{2}\ln(1-\eta^2)\,.$$

If we set $\eta = \mu\nu$, this function $f(\eta)$ reduces to

$$\sum_{k=1}^{\infty} \eta^{2k+1}\left(\frac{1}{2k} + \frac{1}{2k+1} - \frac{2}{3}\right) < \frac{\eta^3}{6} = \frac{\mu^3}{6n}\,.$$

On the other hand, the actual value of η must be larger than $2\mu\nu$, because $2\mu\nu$ is too small to satisfy (15.7):

$$\frac{2(2\mu\nu)^3}{3\bigl(1-(2\mu\nu)^2\bigr)} \leq \frac{16(\mu\nu)^3}{3(1-\frac{4}{25})} = \frac{400\mu^3}{63n} < \frac{r}{n} = y\,.$$

When $\eta > \mu\nu$ we have

$$f'(\eta) = -\frac{\eta\bigl(\eta^3 + 3\eta^2\mu\nu + 3(\eta-\mu\nu)\bigr)}{3(1-\eta^2)^2} < -\eta(\eta-\mu\nu) < -(\eta-\mu\nu)^2\,;$$

hence when η satisfies (15.7) we have

$$l(y,\eta) < \frac{\mu^3}{6n} - \frac{(\eta-\mu\nu)^3}{3} \le \frac{\mu^3}{6n} - \epsilon_3\eta^3 \le \frac{\mu^3}{6n} - \epsilon_4 y\,.$$

We have proved that

$$p_{rd} \le C_7 r^d y^{d/3} e^{-\epsilon_4 r + 2r\mu\nu},$$

and this is at most $C_8 n^{-d/3} e^{-\epsilon_5 r}$ if δ is less than $\frac{1}{2}\epsilon_4$. □

Returning to the proof of Theorem 7, its final claim now follows for $\mu \le n^{1/12}$ by summing the upper bounds of Lemma 5 over all values of r. The claim is trivial when $\mu > n^{1/12}$. □

As remarked earlier, the fact that (15.1) sums to 1 allows us to compute asymptotic probabilities of any collection of graphs or multigraphs obtained as a union over an infinite set of reduced multigraphs, as long as at least one multigraph in the set is clean (has deficiency zero). We simply sum the individual probabilities, neglecting unclean cases.

One corollary of Theorem 7 is the fact that a random graph with $\frac{1}{2}(n+\mu n^{2/3})$ edges is clean with probability $1-o(1)$ whenever $\mu = o(n^{1/12})$. Stepanov proved this for $\mu \le 0$ [36, Theorem 3] and conjectured that it would also hold for positive μ. His conjecture was proved for all fixed μ by Łuczak, Pittel, and Wierman [28].

Erdős and Rényi remarked in their pioneering paper [13, §8] that, if x is any real number, the probability that a graph with $\frac{1}{2}n + xn^{1/2}$ edges is nonplanar "has a positive lower limit, but we cannot calculate its value. It may even be 1, though this seems unlikely." They gave no proof that the limiting probability is positive, and their remark was embedded in a section of [13] that contains a technical error (see [27]); but a proof of their assertion was found later by Stepanov [36, Corollary 2 following (10)]. In the other direction, the fact that nonplanarity occurs with probability strictly less than 1 follows from the fact that a graph with $\frac{1}{2}n + o(n^{2/3})$ edges has excess 0 with probability $\sqrt{\frac{2}{3}}$, as observed in [14, Corollary 8].

We are now in a position to make a more precise estimate of the probability in question.

Theorem 8. *The probability that a graph with $\frac{1}{2}n+o(n^{2/3})$ edges is nonplanar approaches a limit ρ as $n\to\infty$, where*

$$0.000229 \le \rho \le 0.012926. \tag{15.8}$$

Proof. The condition $m = \frac{1}{2}n + o(n^{2/3})$ is equivalent to saying that $\mu = o(1)$ when $m = \frac{1}{2}n(1+\mu n^{-1/3})$, so we can let $\mu = 0$ in the asymptotic formulas above. By Theorem 7, the constant ρ is the sum $\sum \sqrt{2\pi} A(3r+\frac{1}{2}, 0)\kappa(\overline{\overline{M}})/(2r)! = \sum \sqrt{\frac{2}{3}}\left(\frac{4}{3}\right)^r r!\,\kappa(\overline{\overline{M}})/(2r)!^2$ over all nonplanar, reduced, labeled, clean multigraphs $\overline{\overline{M}}$, where $r = r(\overline{\overline{M}})$ is the excess of $\overline{\overline{M}}$.

A clean multigraph cannot contain a subgraph that is homeomorphic to the complete graph K_5, i.e., a subgraph that cancels to K_5, because K_5 has deficiency 5. Adding an edge to any multigraph increases the excess by 0 or 1 and increases the deficiency by 0, 1, or 2 (see section 20 below for further discussion); hence all subgraphs of a clean multigraph are clean. Indeed, this argument implies that a random graph with $\frac{1}{2}n+o(n^{2/3})$ edges has probability $O(n^{-5/3})$ of containing a K_5.

Therefore, if a sparse graph or multigraph is nonplanar, its nonplanarity comes almost surely from a subgraph that cancels to the complete bipartite graph $K_{3,3}$, which is clean and has excess 3.

One way to obtain bounds on ρ is to restrict consideration to reduced multigraphs whose components all have excess ≤ 3. If such a multigraph contains a $K_{3,3}$, it corresponds only to nonplanar graphs; if it does not, it corresponds only to planar graphs. The difference between the upper and lower bounds so obtained is the probability that a random graph of $\frac{1}{2}n + o(n^{2/3})$ edges has at least one component of excess ≥ 4, i.e., that at least one component is more than tetracyclic.

The multigraph $K_{3,3}$ has compensation factor 1, because it is a graph, and its vertices can be labeled in $\frac{1}{2}\binom{6}{3} = 10$ different ways. Thus it contributes only $\frac{10}{6!} = \frac{1}{72}$ to the constant $c_3 = \frac{1105}{1152}$ that accounts for all clean connected multigraphs of excess 3.

Let $f_r = [z^r]\ \exp(c_1 z + c_2 z^2 + c_3 z^3)$ and $g_r = [z^r]\ \exp\bigl(c_1 z + c_2 z^2 + (c_3 - \frac{1}{72})z^3\bigr)$. Then the quantities

$$p = \sum_{r\geq 0} \sqrt{\frac{2}{3}} \left(\frac{4}{3}\right)^r f_r \frac{r!}{(2r)!} \qquad \text{and} \qquad q = \sum_{r\geq 0} \sqrt{\frac{2}{3}} \left(\frac{4}{3}\right)^r g_r \frac{r!}{(2r)!}$$

are respectively the probability that a sparse graph has all components of excess ≤ 3 and the probability that, moreover, no component cancels to $K_{3,3}$. These series converge rapidly and lead to the numerical bounds $p - q$ and $1 - q$ in (15.8). $\square$

It is interesting to study the expected number $\mathrm{E}n_1$ of vertices in complex components, as a function of μ, because it will be the expected number of vertices in the giant component when μ increases. We have $\mathrm{E}n_1 = \sum_r \mathrm{E}(n_1|r)\Pr(\mathcal{E}_r)$. By (13.17) and the remarks preceding (13.13), each term in this sum can be approximated, to within relative error $O\bigl((1+\mu^4)n^{-1/3}\bigr)$, by $3r\sqrt{2\pi}e_r A(3r+\frac{5}{2},\mu)n^{2/3}$. Let us, for simplicity, assume that μ is bounded. Then the proof of Lemma 5 is easily modified to show that the rth term of the sum is $O\bigl(n^{2/3}(r+1)e^{-\epsilon r}\bigr)$, uniformly in n and r. Thus, by dominated convergence, $\mathrm{E}n_1 = \bigl(f(\mu)+o(1)\bigr)n^{2/3}$, where

$$f(\mu) = \sum_{r\geq 0} 3r\sqrt{2\pi}\, e_r\, A(3r+\tfrac{5}{2},\mu)\,. \tag{15.9}$$

Equation (10.23) tells us that

$$\tfrac{1}{2}\,\mu^2 f(\mu) + f'(\mu) = \tfrac{1}{2}\,\textstyle\sum_{r\ge 0} 3r\sqrt{2\pi}\,e_r\,A(3r+\tfrac{1}{2},\mu) = \tfrac{3}{2}\,g(\mu)\,, \tag{15.10}$$

where $g(\mu)$ is the expected value of r; we calculated $g(\mu)$ in the discussion leading up to Theorem 6. Thus, we obtain the estimate

$$f(\mu) = 2\mu - \mu^{-2} - \tfrac{27}{8}\,\mu^{-5} - \tfrac{495}{16}\,\mu^{-8} + O(\mu^{-11})\,, \tag{15.11}$$

for $\mu \ge \delta > 0$, by combining (15.9) with the asymptotic formula for $g(\mu)$ in (14.13).

We can express $f(\mu)$ in "closed hypergeometric form" by proceeding as in (14.9) and (14.10). The result is

$$\begin{aligned} f(\mu) = {} & -\frac{2^{-2/3}\pi}{3^{7/6}\,\Gamma\bigl(\frac{2}{3}\bigr)}\,e^{-\mu^3/6} + \mu - \frac{\mu}{4}e^{-\mu^3/6}\,F\left(\frac{1}{3};\,\frac{4}{3};\,\frac{\mu^3}{6}\right) \\ & + e^{-\mu^3/6}\left(\frac{2^{1/3}\sqrt{\pi}}{3^{7/6}\,\Gamma\bigl(\frac{7}{6}\bigr)}\,F\left(\frac{1}{2},\frac{5}{6};\,\frac{1}{3},\frac{2}{3};\,\frac{\mu^3}{6}\right)\right. \\ & \qquad - \frac{\mu}{2\sqrt{3}}\,F\left(\frac{5}{6},\frac{7}{6};\,\frac{2}{3},\frac{4}{3};\,\frac{\mu^3}{6}\right) \\ & \qquad \left. + \frac{3^{1/6}\sqrt{\pi}\,\mu^2}{2^{7/3}\,\Gamma\left(\frac{5}{6}\right)}\,F\left(\frac{7}{6},\frac{3}{2};\,\frac{4}{3},\frac{5}{3};\,\frac{\mu^3}{6}\right)\right). \end{aligned} \tag{15.12}$$

It is instructive to compare this expression with alternative formulas for the same quantity obtained in [28] by a different method:

$$\begin{aligned} f(\mu) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \Bigl(\sum_{r\ge 1} f_r x^{3r/2}\Bigr) e^{G(x,\mu)}\,dx \\ &= \mu + \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{1-e^{G(x,\mu)}}{x^{3/2}}\,dx - \frac{1}{4}\int_0^\infty e^{G(x,\mu)}\,dx\,. \end{aligned} \tag{15.13}$$

Here $G(x,\mu) = \bigl((\mu-x)^3-\mu^3\bigr)/6$, and $f_r n^{n+(3r-1)/2}$ is Wright's asymptotic estimate [44] for the number of connected graphs with excess r.

16. Evolutionary paths. Consider any graph or multigraph that evolves by starting out with isolated vertices and then by acquiring edges one at a time. Initially its excess is 0; then each new edge either preserves the current excess or increases it by 1. We observed in section 4, following (4.7), that a new edge augments the excess if and only if both of its endpoints are currently in the cyclic part. We observed in section 13 that many interesting statistics about random graphs can be usefully represented in terms of probabilities that are conditional on the graph having a given excess. Therefore it is natural to look more closely at the way a graph changes character as its excess grows.

Every evolution of a graph or multigraph traces a path from left to right in the following diagram, which shows the beginning of an infinite partial ordering of all possible configurations $[r_1, r_2, \ldots, r_q]$:

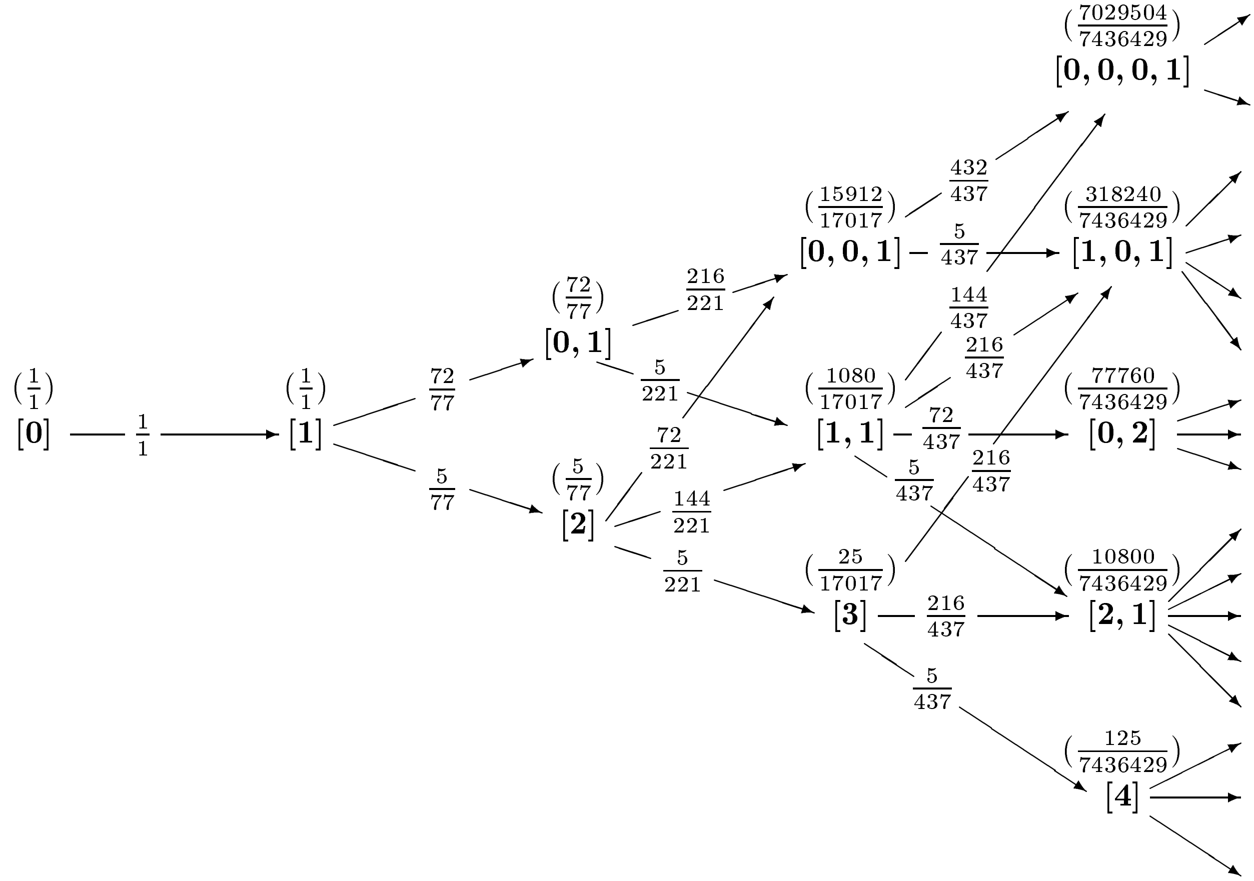


Figure 1. The evolution of complex components. Each configuration $[r_1, r_2, \ldots, r_q]$ stands for a graph or multigraph with r_1 bicyclic components, r_2 tricyclic components, $\ldots$, r_q $(q+1)$-cyclic components. As a graph evolves, its excess $r_1 + 2r_2 + 3r_3 + \cdots$ increases in unit steps, and the configurations follow a path from left to right in this partial ordering.

When complex components begin to form, they follow a path in this diagram, with the indicated transition probabilities. The upper path is followed most frequently; on this path there is a unique complex component that will become the "giant." Parenthesized ratios are the probabilities of reaching a given configuration. At the moment the excess first reaches 2, the configuration must either be $[0, 1]$ (one tricyclic component) or $[2]$ (two bicyclic components). When the excess goes from 2 to 3, we go either from $[0, 1]$

to $[0,0,1]$ or $[1,1]$, or from $[2]$ to $[0,0,1]$, [1,1], or [3]; and so on. Each configuration $[r_1, r_2, \ldots, r_q]$ corresponds to a partition of the excess $r = r_1 + 2r_2 + \cdots + qr_q$. The fraction in parenthesis shown above each configuration in Figure 1 is the limiting probability $c_1^{r_1} c_2^{r_2} \ldots c_q^{r_q}/(r_1!\, r_2! \,\ldots\, r_q!\, e_r)$ that a random graph of excess r has configuration $[r_1, r_2, \ldots, r_q]$. This is the limiting probability that the infinite path traced out in the infinite extension of Figure 1 will pass through $[r_1, r_2, \ldots, r_q]$ during the evolution of a random graph or multigraph on a large number of vertices.

A random graph almost always acquires nearly $\frac{1}{2}n$ edges before taking the first step from [0] to [1] in Figure 1. Indeed, the uniform estimate (13.17), with $\mu = -n^{1/21}$, implies that the probability of excess r when $m = \frac{1}{2}n\exp(-n^{-2/7})$ is of order $n^{-r/7}$.

The fractions shown on arcs leading between configurations are transition probabilities, namely the limiting probabilities that a random graph of configuration $[r_1, r_2, \ldots, r_q]$ will go to another specified configuration when its excess next changes. For example, a random graph in configuration [2], having two bicyclic components and no other complex components, will proceed next to configuration [1,1] with probability $\frac{144}{221}$. These transition probabilities have a fairly simple characterization:

Theorem 9. *Let $r_1 + 2r_2 + \cdots + qr_q = r$ and $\delta_1 + 2\delta_2 + 3\delta_3 + \cdots = 1$. The asymptotic probability that a random graph or multigraph of configuration $[r_1, r_2, \ldots, r_q]$, having no acyclic components, will change to configuration $[r_1+\delta_1, r_2+\delta_2, \ldots, r_q+\delta_q, \delta_{q+1}, \ldots]$ when a random edge is added, can be computed as follows:*

Nonzero δ's	Probability
$\delta_1 = 1$	$\frac{5}{4}/(3r+\frac{1}{2})(3r+\frac{5}{2})$
$\delta_j = -1,\ \delta_{j+1} = 1$	$9j(j+1)r_j/(3r+\frac{1}{2})(3r+\frac{5}{2})$
$\delta_j = -2,\ \delta_{2j+1} = 1$	$9j^2 r_j(r_j-1)/(3r+\frac{1}{2})(3r+\frac{5}{2})$
$\delta_j = -1,\ \delta_k = -1,\ \delta_{j+k+1} = 1,\ j < k$	$18j\,k\,r_j r_k/(3r+\frac{1}{2})(3r+\frac{5}{2})$

In all other cases, the probability is 0. The estimates are correct to within $O(n^{-1/2})$ when there are n vertices.

Proof. As usual, it is easiest to consider first the uniform multigraph process. We know that the generating function for the cyclic multigraphs under consideration is

$$S(z) = e^{V(z)}\,\frac{C_1(z)^{r_1}}{r_1!}\,\frac{C_2(z)^{r_2}}{r_2!}\,\cdots\,\frac{C_q(z)^{r_q}}{r_q!}\,; \tag{16.1}$$

the number of such multigraphs, weighted as usual by the compensation factor (1.1), is $[z^n]\,S(z)$. We also know from (3.4) that $V(z) = -\frac{1}{2}\ln\bigl(1-T(z)\bigr)$, hence

$$e^{V(z)} = \frac{1}{\bigl(1-T(z)\bigr)^{1/2}}\,.$$

We observed in section 4 that the operator $\vartheta = z\frac{d}{dz}$ corresponds to "marking" or singling out a particular vertex. The function $\vartheta^2 S(z)$ can therefore be regarded as the generating function for multigraphs of configuration $[r_1, r_2, \ldots, r_q]$ together with an ordered pair of marked vertices $\langle x, y\rangle$. When $S(z)$ is a product $A(z)B(z)$, the familiar relation

$$\vartheta^2\bigl(A(z)B(z)\bigr) = \bigl(\vartheta^2 A(z)\bigr)B(z) + 2\bigl(\vartheta A(z)\bigr)\bigl(\vartheta B(z)\bigr) + A(z)\bigl(\vartheta^2 B(z)\bigr) \tag{16.2}$$

has a natural combinatorial interpretation: The product $A(z)B(z)$ stands for ordered pairs of graphs, generated respectively by $A(z)$ and $B(z)$, with no edges between them; the first term $\bigl(\vartheta^2 A(z)\bigr)B(z)$ of (16.2) corresponds to cases when both of the marked vertices $\langle x, y\rangle$ are in the graph generated by $A(z)$; the last term corresponds to cases when both x and y belong to the $B(z)$ graph. The middle term $2\bigl(\vartheta A(z)\bigr)\bigl(\vartheta B(z)\bigr)$ corresponds to the cases where x is in A and y is in B or vice versa.

We can use this idea in connection with (16.1) to understand what happens when the graph gains a new edge. The coefficient of z^n in $\vartheta^2 S(z)$ represents all possibilities $\langle x, y\rangle$; we can divide this into cases by writing

$$\vartheta^2 S(z) = S(z)\biggl(\sum_{0\le j\le q} \frac{\vartheta^2 f_j(z)}{f_j(z)} + 2\sum_{0\le j<k<q} \frac{\vartheta f_j(z)}{f_j(k)}\,\frac{\vartheta f_k(z)}{f_k(z)}\biggr) \tag{16.3}$$

where $f_0(z) = e^{V(z)}$ and $f_j(z) = C_j(z)^{r_j}/r_j!$ for $j \ge 1$. A term like $S(z)\bigl(\vartheta^2 f_j(z)\bigr)/f_j(z)$, say, then corresponds to cases where x and y both belong to $(j+1)$-cyclic components.

Each of the factors $f_j(z)$ is a linear combination of powers of the quantity $\xi = 1+\zeta = 1/(1-T(z))$. For example, $f_0(z) = \xi^{1/2}$ and $f_1(z) = \frac{5}{24}\xi^3 - \frac{7}{24}\xi^2 + \frac{1}{12}\xi$, according to (3.4) and (11.3). Hence it is easy to compute ϑf_j and $\vartheta^2 f_j$, using rule (4.5):

$$\begin{aligned}\vartheta(\xi^\alpha) &= \alpha\xi^{\alpha+2} - \alpha\xi^{\alpha+1}\,;\\ \vartheta^2(\xi^\alpha) &= \alpha(\alpha+2)\xi^{\alpha+4} - \alpha(2\alpha+3)\xi^{\alpha+3} + \alpha(\alpha+1)\xi^{\alpha+2}\,.\end{aligned} \tag{16.4}$$

The overall function $S(z)$ has the form $\xi^{3r+1/2}P(\xi^{-1})$ for some polynomial P, with $P(0) \ne 0$; hence the coefficient $[z^n]\,S(z)$ is $t_n(3r+\frac12)P(0)\bigl(1+O(n^{-1/2})\bigr)/n!$ by (3.8) and (3.9). It follows from (16.4) that $\vartheta^2 S(z) = \xi^{3r+9/2}Q(\xi^{-1})$ for some polynomial Q, where $Q(0) = (3r+\frac12)(3r+\frac52)P(0)$. Hence

$$n^2 = \frac{[z^n]\,\vartheta^2 S(z)}{[z^n]\,S(z)} = (3r+\tfrac12)(3r+\tfrac52)\,\frac{t_n(3r+\frac92)}{t_n(3r+\frac12)}\bigl(1+O(n^{-1/2})\bigr)\,. \tag{16.5}$$

The transition probabilities we wish to compute are the fractions of $(3r+\frac12)(3r+\frac52)$ that occur when ϑ^2 operates on individual factors of $S(z)$.

For example, consider first the term $S(z)\bigl(\vartheta^2 f_0(z)\bigr)/f_0(z)$ of (16.3). This corresponds to the case where both x and y belong to a cyclic component (possibly the same one), thereby creating a new bicyclic component; thus it corresponds to having $\delta_1 = 1$ and all other $\delta_j = 0$. In this case $[z^n]\,S(z)\bigl(\vartheta^2 f_0(z)\bigr)/f_0(z) \sim \frac{1}{2}\cdot\frac{5}{2}\,t_n(3r+\frac{9}{2})P(0)/n!$, and the latter is asymptotically $\frac{5}{4}/(3r+\frac{1}{2})(3r+\frac{5}{2})$ of the total $[z^n]\,\vartheta^2 S(z)$.

The term $2S(z)\bigl(\vartheta f_0(z)\bigr)\bigl(\vartheta f_j(z)\bigr)/f_0(z)f_j(z)$, similarly, gives the probability that a vertex from a cyclic component joins with a $(j+1)$-cyclic component; this occurs with probability $2(\frac{1}{2})(3j\,r_j)/(3r+\frac{1}{2})(3r+\frac{5}{2})$. The net effect on components corresponds to $\delta_j = -1$, $\delta_{j+1} = +1$.

There is also another way to get $\delta_j = -1$ and $\delta_{j+1} = +1$, namely if both x and y belong to the same $(j+1)$-cyclic component. The probability of this case works out to be $(3j)(3j+2)r_j/(3r+\frac{1}{2})(3r+\frac{5}{2})$; hence the total transition probability for $\delta_j = -1$ and $\delta_{j+1} = +1$ is $9j(j+1)r_j/(3r+\frac{1}{2})(3r+\frac{5}{2})$ as stated in the theorem.

Notice that

$$\vartheta^2 C_j^{r_j} = r_j C_j^{r_j-1}(\vartheta^2 C_j) + r_j(r_j-1)C_j^{r_j-2}(\vartheta C_j)^2\,. \tag{16.6}$$

We have just taken care of the first term; the second term corresponds to vertices x and y in distinct C_j's, when the new edge makes $\delta_j = -2$ and $\delta_{2j+1} = +1$. The probability is $9j^2 r_j(r_j-1)/(3r+\frac{1}{2})(3r+\frac{5}{2})$.

Finally, the term $2S(z)\bigl(\vartheta f_j(z)\bigr)\bigl(\vartheta f_k(z)\bigr)/f_j(z)f_k(z)$ of (16.3) represents a case that occurs with probability $2(3j\,r_j)(3k\,r_k)/(3r+\frac{1}{2})(3r+\frac{5}{2})$ and corresponds to $\delta_j = \delta_k = -1$, $\delta_{j+k+1} = +1$.

If we are working with the graph process instead of the multigraph process, we must use $\widehat{C}_j(z)$ instead of $C_j(z)$; but $f_0(z)$ is still essentially of degree $-1/2$ in ξ^{-1}, and $f_j(z)$ is still of degree $-3j$, so the asymptotic calculations work out as before.

However, in a random graph we must use the operator $\frac{1}{2}(\vartheta_z^2 - \vartheta_z) - \vartheta_w$ instead of ϑ_z^2, and we must work with bivariate generating functions, as discussed in section 6. The bgf corresponding to (16.1) is almost univariate, however:

$$\hat{F}(w,z) = w^r\, e^{\hat{V}(wz)}\,\frac{C_1(wz)^{r_1}}{r_1!}\,\frac{C_2(wz)^{r_2}}{r_2!}\,\cdots\,\frac{C_q(wz)^{r_q}}{r_q!}\,.$$

It is not difficult to see that the effect of ϑ_z^2 swamps the effects of ϑ_z and ϑ_w, asymptotically, so the multigraph analysis carries through. $\square$

One amusing consequence of Theorem 9 is that we can use it to discover and prove formula (7.2) for the numbers e_r in a completely different way. The probability of reaching the configuration $[r]$, consisting of r bicyclic components and none of higher cyclic order, is $c_1^r/(r!\,e_r)$. The only way to reach this configuration, when $r > 0$, is from $[r-1]$, and the transition probability is

$$\frac{\frac{5}{4}}{(3r-\frac{5}{2})(3r-\frac{1}{2})} = \frac{c_1^r/(r!\,e_r)}{c_1^{r-1}/\bigl((r-1)!\,e_{r-1}\bigr)}\,.$$

Since $c_1 = 5/24$, we have $e_r = (6r-5)(6r-1)e_{r-1}/24r$, and (7.2) follows by induction. This indirect method is probably the simplest way to deduce the fact that Wright's constant is $1/(2\pi)$.

17. A near-Markov process. We proved in Theorem 9 that the transition probabilities shown in Figure 1 are the limiting probabilities, averaged over all multigraphs, that a multigraph reaching a particular state will take a particular step as its excess increases. But we did not prove that those transition probabilities are independent of past history. For all we know, the path taken to a particular configuration during the evolution of a random graph might strongly influence the probability distribution of its next leap forward. The next theorem addresses this question.

Theorem 10. *For any fixed R, an evolving random graph or multigraph almost surely carries out a random walk in the first R levels of the partial ordering shown in Figure 1, with transition probabilities that approach the limiting values derived in Theorem 9.*

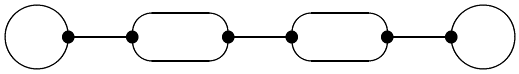
Proof. As in previous proofs, it suffices to consider random multigraphs. We will show that the transition probabilities have the asymptotic behavior of Theorem 9 for all random multigraphs that remain clean—i.e., for all multigraphs that reduce, under the pruning and cancelling algorithms of Section 9, to 3-regular multigraphs $\overline{\overline{M}}$ having $2r$ vertices and $3r$ edges, when the excess is $r \le R$. We know from Theorem 7 that the multigraph will be clean with probability $1-O\bigl((1+\mu^4)n^{-1/3}\bigr)$; and we know from (13.17) that the probability of excess r becomes superpolynomially small as the number of edges passes $\frac{n}{2}$. So the excess almost surely increases past any given value before a large multigraph becomes unclean. For example, if $\mu \to \infty$ with $\mu = o(n^{1/12})$, the probability of excess $\le R$ approaches zero while the probability of remaining clean is $1 - o(1)$.

The proof for clean multigraphs is not as trivial as might be expected: Multigraphs that follow a given path to $[r_1, r_2, \ldots, r_q]$ in the partial ordering are *not* uniformly distributed, among all multigraphs whose complex parts are enumerated by the generating function

$$e^V (C_1^{r_1}/r_1!)(C_2^{r_2}/r_2!) \ldots (C_q^{r_q}/r_q!)$$

assumed in the proof of Theorem 9. Past history does affect the frequency of certain types of components. For example, a tricyclic component that prunes and cancels to $K_{3,3}$ cannot evolve along the path $[1] \to [2] \to [0,0,1]$; removing any edge of $K_{3,3}$ leaves a connected graph.

Let's try to clarify the situation by working an example. Consider the reduced multigraph



; (17.1)

suppose we wish to compute the transition probabilities for multigraphs of excess 3 that prune and cancel to (17.1) after following the path $[1] \to [0,1] \to [0,0,1]$. The generating function for all such multigraphs, assuming that there are no acyclic components, would be $\frac{1}{32}\, e^V T^6/(1-T)^9$, if we did not specify the past history $[1] \to [0,1] \to [0,0,1]$; but it turns out to be only $\frac{8}{9}$ as much when we prescribe the history. The reason is, intuitively, that (17.1) has 9 edges, and a multigraph with history $[1] \to [0,1] \to [0,0,1]$ can reduce to it only if the "middle" edge is not the last to be completed. The latter event happens with probability $\frac{8}{9}$.

A formal proof of the $\frac{8}{9}$ phenomenon can be given as follows. The generating function $e^V T^6 (1-T)^9$ expands to $e^V T^6 \sum_{n_1,n_2,\ldots,n_9 \ge 0} T^{n_1} T^{n_2} \ldots T^{n_9}$; the individual terms represent the insertion of $\langle n_1, \ldots, n_9 \rangle$ vertices into the nine edges of (17.1), after which a tree is sprouted at each vertex. The resulting multigraph will have n vertices and $m = n + 3$ edges; there will be $6 + n_1 + \cdots + n_9$ root vertices and $9 + n_1 + \cdots + n_9$ "critical" edges on paths between root vertices. Suppose we color each critical edge with one of 9 colors, corresponding to the original edge of (17.1) from which it was subdivided. Then among the $m!$ permutations of edges that could generate any such multigraph, exactly $\frac{8}{9}$ have the property that the last critical edge has some color besides the "middle" color. (This follows by symmetry between $n_1, n_2, \ldots, n_9$.) Such permutations are precisely those for which the history will be $[1] \to [0,1] \to [0,0,1]$; hence we obtain (17.1) with exactly $\frac{8}{9}$ times its overall probability, given that history.

It turns out that there are 17 unlabeled clean, connected, reduced multigraphs of excess 3; and exactly 6 of them occur with weight $\frac{8}{9}$ when the past history is $[1] \to [0,1] \to [0,0,1]$. Those 6 occur with weight $\frac{1}{9}$ when the past history is $[1] \to [2] \to [0,0,1]$, and the other 11 do not occur at all in that case.

In general, given any $\overline{\overline{\overline{M}}}$ that can arise for a given past history, there will be a fraction $\beta > 0$ such that each multigraph reducing to $\overline{\overline{\overline{M}}}$ arises β times as often with the given history as it does overall. The reason is a slight generalization of the method by which we proved the $\frac{8}{9}$ phenomenon: Each permutation of colors of critical edges is equally likely to be the sequence of last appearances in a random permutation of $n_1 + n_2 + \cdots + n_{3r}$ critical edges, and such permutations determine the past history. The generating function for $\overline{\overline{\overline{M}}}$ will then be a constant multiple of $e^V \bigl(T^{2r_1}/(1-T)^{3r_1}\bigr)\bigl(T^{2r_2}/(1-T)^{3r_2}\bigr) \ldots \bigl(T^{2r_q}/(1-T)^{3r_q}\bigr)$. Hence the asymptotic transition probabilities will be the same for every feasible $\overline{\overline{\overline{M}}}$, exactly as calculated in Theorem 9. $\square$

18. An emerging giant. The classic papers of Erdős and Rényi [12, 13] tell us that an evolving graph almost surely develops a single giant component, which eventually is surrounded by only a few trees and later by only isolated vertices, until the entire graph becomes connected. Thus there will be a time when the graph reaches some configuration $[0, 0, \ldots, 0, 1]$ on the top line of Figure 1 and stays on that top line ever afterward.

Indeed, the most probable path in Figure 1 is the one that goes directly from $[1]$ to $[0, 1]$ to $[0, 0, 1]$ and so on, never leaving the top line. The first transition probability is $\frac{72}{77}$, the next is $\frac{216}{221}$, and subsequent steps are ever more likely to stay in line. In such cases we can see the "seed" around which the giant component is forming, before that component has become in any way gigantic. (The complex components of any given finite excess almost always have only $O(n^{2/3})$ vertices, a vanishingly small percentage of the total; each step at the beginning of Figure 1 occurs after adding about $n^{2/3}$ more edges.)

If we assume that the transition probabilities in Figure 1 are exact, the overall probability that an evolving graph adheres strictly to the top line—never having more than one complex component throughout its entire evolution—is

$$\prod_{r=1}^{\infty} \frac{r(r+1)}{(r+\frac{1}{6})(r+\frac{5}{6})} = \frac{\Gamma(\frac{7}{6})\,\Gamma(\frac{11}{6})}{\Gamma(1)\,\Gamma(2)} = \frac{5}{36}\,\Gamma\Bigl(\frac{1}{6}\Bigr)\,\Gamma\Bigl(\frac{5}{6}\Bigr) = \frac{5\pi}{18}\,. \tag{18.1}$$

Numerically, this comes to 0.8726646, roughly 7 times out of every 8.

Is $\frac{5\pi}{18}$ the true limiting probability that an evolving graph or multigraph never acquires two simultaneous components of positive excess, throughout its evolution? We can at least prove that $\frac{5\pi}{18}$ is an upper bound. For we know from Theorem 10 that an evolving graph will hug the top line of Figure 1 for at least R steps with probability

$$\prod_{r=1}^{R} \frac{r(r+1)}{(r+\frac{1}{6})(r+\frac{5}{6})} = \frac{5\pi}{18} + O(R^{-1}) + O(n^{-1/3}) \tag{18.2}$$

for any fixed R, as $n \to \infty$.

It is natural to conjecture that $\frac{5\pi}{18}$ is also a lower bound, because a large component tends to propagate itself as soon as it becomes large enough. Still, it is conceivable that a random graph might have a tendency to leave the top line briefly when it first becomes unclean. The transition probability for remaining on the top line becomes strictly less than $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$ when the graph has a positive deficiency. For example, suppose the initial bicyclic component is already unclean; it will then correspond to the double self-loop of (9.15). We know from (13.16) that this case arises with probability $O(n^{-1/3})$. But if it does occur, the generating function for the complex part will be a constant multiple of $T/(1-T)^2$ instead of $T^2/(1-T)^3$, so the proof technique of Theorem 9 will yield a transition probability from $[1]$ to $[0, 1]$ of only $\frac{8}{9}$ instead of $\frac{72}{77}$. In general, when the deficiency is d, the asymptotic transition probability drops to

$$(r-\tfrac{d}{3})(r-\tfrac{d}{3}+1)/(r-\tfrac{d}{3}+\tfrac{1}{6})(r-\tfrac{d}{3}+\tfrac{5}{6})\,.$$

This probability estimate is, moreover, valid only when the excess is reasonably small as a function of n; otherwise the trees that sprout from the pruned multigraph $\overline{M}$ will not be large enough to assert their asymptotic behavior.

19. A monotonicity property. During the time when an evolving graph or multigraph stays clean, we can show that the asymptotic top-line transition probabilities $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$ are in fact *lower* bounds for the correct (non-asymptotic) probabilities. More precisely, the proof of Theorem 9 shows that the true transition probability is a ratio of expressions involving the tree polynomials $t_n(y)$, when there are n vertices in the cyclic part of the multigraph. We will prove that this ratio decreases monotonically to $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$ as $n \to \infty$.

First we need to prove an auxiliary result about tree polynomials that is interesting in its own right. Let us generalize the definition of $t_n(y)$ in (3.8) by introducing a new parameter $m \geq 0$:

$$\frac{T(z)^m}{\bigl(1-T(z)\bigr)^y} = \sum_{n=0}^{\infty} t_{m,n}(y)\,\frac{z^n}{n!}\,. \tag{19.1}$$

Thus

$$t_{m,n}(y) = \sum_{j=0}^{m} \binom{m}{j}(-1)^j t_n(y-j) \tag{19.2}$$

is the mth backward difference of $t_n(y)$.

Lemma 6. *Let m be a nonnegative integer. For any fixed integer $n > m$ and arbitrary real $y > 0$, the ratio $t_{m,n+1}(y)/t_{m,n}(y)$ is an increasing function of y. Equivalently, for fixed $y > 0$ and any integer $n > m$, the ratio $t'_{m,n}(y)/t_{m,n}(y)$ is an increasing function of n.*

Proof. The two statements of the lemma are clearly equivalent, because $t_{m,n}(y)$ is positive when $y > 0$ and $n > m$.

Equation (2.12) of [24] states that

$$t_n(y) = n^{n-1}\sum_{k\geq 0} \frac{y^{\overline{k+1}}}{k!}\;\frac{(n-1)^{\underline{k}}}{n^k}\,, \tag{19.3}$$

where $x^{\overline{k}}$ means $x(x+1)\ldots(x+k-1)$ and $x^{\underline{k}}$ means $x(x-1)\ldots(x-k+1)$. Therefore, by (19.2),

$$\begin{aligned} t_{m,n}(y) &= n^{n-1}\sum_{k\geq m-1} (k+1)\,\frac{y^{\overline{k+1-m}}}{(k+1-m)!}\;\frac{(n-1)^{\underline{k}}}{n^k} \\ &= n^{n-m}\sum_{k=0}^{n-m} (k+m)\,\frac{y^{\overline{k}}}{k!}\;\frac{(n-1)^{\underline{k+m-1}}}{n^k}\,. \end{aligned} \tag{19.4}$$

It follows that the inequality $t'_{m,n}(y)/t_{m,n}(y) < t'_{m,n+1}(y)/t_{m,n+1}(y)$ is equivalent to

$$\frac{\sum_{k=0}^{N} a_k\alpha_k}{\sum_{k=0}^{N} b_k\alpha_k} > \frac{\sum_{k=0}^{N} a_k\beta_k}{\sum_{k=0}^{N} b_k\beta_k}\,, \tag{19.5}$$

where $N = n+1-m$ and

$$\begin{aligned} a_k &= (k+m)\,\frac{y^{\overline{k}}}{k!}\,, & b_k &= (k+m)\,\frac{d}{dy}\,\frac{y^{\overline{k}}}{k!}\,, \\ \alpha_k &= (n-1)^{\underline{k+m-1}}\,n^{n-m-k}\,, & \beta_k &= n^{\underline{k+m-1}}\,(n+1)^{n+1-m-k}\,. \end{aligned} \tag{19.6}$$

The following condition is sufficient to prove (19.5), assuming positive denominators:

$$\frac{a_0}{b_0} > \frac{a_1}{b_1} > \cdots > \frac{a_N}{b_N} \qquad \text{and} \qquad \frac{\alpha_0}{\beta_0} > \frac{\alpha_1}{\beta_1} > \cdots > \frac{\alpha_N}{\beta_N}\,. \tag{19.7}$$

For we have

$$\begin{aligned} &\sum_{k=0}^{N} b_k\beta_k \sum_{j=0}^{N} a_j\alpha_j - \sum_{j=0}^{N} b_j\alpha_j \sum_{k=0}^{N} a_k\beta_k \\ &\qquad = \sum_{0\le j<k\le n} (b_k a_j - b_j a_k)(\beta_k\alpha_j - \beta_j\alpha_k) > 0\,. \end{aligned} \tag{19.8}$$

(Historical note: Inequality (19.5) under condition (19.7) goes back at least to Seitz in 1936 [33]; see [29, Section 2.5, Theorem 4], where a supplementary condition is needed: The product of the denominators must be positive. In linearly ordered discrete probability space, the inequality is equivalent to saying that $E\bigl(f(X)g(X)\bigr) \ge E\bigl(f(X)\bigr)E\bigl(g(X)\bigr)$ whenever f and g are increasing functions of the random variable X. This inequality is, in turn, a special case of the celebrated FKG inequality [15], which applies to certain partially ordered probability spaces. The equality in (19.8), which reduces to Lagrange's identity when we set $a_k = \alpha_k$ and $b_k = \beta_k$, is the Binet-Cauchy identity for det AB when A is a matrix of size $2\times n$ and B is $n\times 2$.)

And (19.7) is not difficult to verify, under the substitutions (19.6). We have

$$\begin{aligned} \frac{b_{k+1}}{a_{k+1}} &= \frac{1}{y} + \frac{1}{y+1} + \cdots + \frac{1}{y+k} = \frac{b_k}{a_k} + \frac{1}{y+k}\,; \\ \frac{\alpha_{k+1}}{\alpha_k} &= \frac{n-k-m}{n} < \frac{n-k-m+1}{n+1} = \frac{\beta_{k+1}}{\beta_k}\,. \end{aligned}$$

(When $m = 0$ we omit the terms for $k = 0$.) $\square$

Assume now that the cyclic part of a random multigraph contains n vertices. The "top line" transition probability from a single clean component of excess r to a single component of excess $r+1$ is $1-p_{nr}$, where p_{nr} is the probability that a new bicyclic component will be formed. By the argument of Theorem 9,

$$p_{nr} = \frac{[z^n]\,\bigl(\vartheta^2 V(z)\bigr)S(z)}{[z^n]\,\vartheta^2\bigl(V(z)S(z)\bigr)}\,, \tag{19.9}$$

where $V(z) = 1/\bigl(1-T(z)\bigr)^{1/2}$ is the generating function for unicyclic components and $S(z) = T(z)^{2r}/\bigl(1-T(z)\bigr)^{3r}$ is a prototypical generating function for clean components of excess r. We want to show that p_{nr} is an increasing function of n, since we want $1-p_{nr}$ to be decreasing.

Let's work on a simpler problem first, showing that

$$q_{nr} = \frac{[z^n]\,\bigl(\vartheta\, A(z)\bigr)S(z)}{[z^n]\,\vartheta\bigl(A(z)S(z)\bigr)} \tag{19.10}$$

is an increasing function of n whenever

$$A(z) = \frac{T(z)^a}{\bigl(1-T(z)\bigr)^b}\,, \qquad b > \frac{3}{2}\,a\,. \tag{19.11}$$

Here a is a nonnegative integer; we will assume that $n \ge 2r+a$, so that the denominator of (19.10) is nonzero. We have

$$\begin{aligned}
\vartheta A(z) &= \frac{b\,T(z)^a}{\bigl(1-T(z)\bigr)^{b+2}} - \frac{(b-a)\,T(z)^a}{\bigl(1-T(z)\bigr)^{b+1}}\,,\\
\vartheta\bigl(A(z)S(z)\bigr) &= \frac{(3r+b)T(z)^{2r+a}}{\bigl(1-T(z)\bigr)^{3r+b+2}} - \frac{(r+b-a)T(z)^{2r+a}}{\bigl(1-T(z)\bigr)^{3r+b+1}}\,;
\end{aligned}$$

hence

$$\begin{aligned}
q_{nr} &= \frac{b\,t_{2r+a,n}(3r+b+2) - (b-a)\,t_{2r+a,n}(3r+b+1)}{(3r+b)\,t_{2r+a,n}(3r+b+2) - (r+b-a)\,t_{2r+a,n}(3r+b+1)}\\
&= \frac{b}{3r+b}\left(1 - \frac{r(2b-3a)/(3rb+b^2)}{\left(\dfrac{t_{2r+a,n}(3r+b+2)}{t_{2r+a,n}(3r+b+1)} - \dfrac{r+b-a}{3r+b}\right)}\right).
\end{aligned}$$

Since the coefficients of $t_{2r+a,n}(y)$ are nonnegative, we have

$$t_{2r+a,n}(3r+b+2)/t_{2r+a,n}(3r+b+1) \ge 1 > (r+b-a)/(3r+b)\,.$$

It follows that q_{nr} is increasing iff

$$\frac{t_{2r+a,n}(3r+b+2)}{t_{2r+a,n}(3r+b+1)} < \frac{t_{2r+a,n+1}(3r+b+2)}{t_{2r+a,n+1}(3r+b+1)} \,. \tag{19.12}$$

And (19.12) does hold, because $t_{2r+a,n+1}(y)/t_{2r+a,n}(y)$ is an increasing function of y by Lemma 6.

Incidentally, this argument also shows that q_{nr} is constant when $b = \frac{3}{2}a$ and decreasing when $0 < b < \frac{3}{2}a$.

Now to prove that p_{nr} is increasing, we can write

$$\begin{aligned} p_{nr} &= \frac{[z^n]\,\big(\vartheta^2 V(z)\big)S(z)}{[z^n]\,\vartheta\big(\big(\vartheta V(z)\big)S(z)\big)}\;\frac{[z^n]\,\vartheta\big(\big(\vartheta V(z)\big)S(z)\big)}{[z^n]\,\vartheta^2\big(V(z)S(z)\big)} \\ &= \frac{[z^n]\,\big(\vartheta^2 V(z)\big)S(z)}{[z^n]\,\vartheta\big(\big(\vartheta V(z)\big)S(z)\big)}\;\frac{[z^n]\,\big(\vartheta V(z)\big)S(z)}{[z^n]\,\vartheta\big(V(z)S(z)\big)} \,. \end{aligned}$$

The first factor is of type q_{nr} if we put $A(z) = \vartheta V(z) = \frac{1}{2}\,T(z)/\big(1-T(z)\big)^{5/2}$; here $a = 1$, $b = \frac{5}{2}$, so q_{nr} is increasing. The second factor is of type q_{nr} if we put $A(z) = V(z)$; here $a = 0$, $b = \frac{1}{2}$, and again q_{nr} is increasing. We have proved

Theorem 11. *The probability that a clean random multigraph of excess $r > 0$ will not acquire a new bicyclic component when its excess next changes is strictly greater than the limiting value $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$.* □

Theorem 11 gives further support to the $\frac{5\pi}{18}$ conjecture of Section 18, because $\frac{5\pi}{18}$ was shown there to be an upper bound. If the top-line transition probability were always strictly greater than $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$, we could establish $\frac{5\pi}{18}$ as a lower bound. However, Theorem 11 does not prove the conjecture, because the probability becomes smaller than $r(r+1)/(r+\frac{1}{6})(r+\frac{5}{6})$ when a graph becomes unclean.

Incidentally, when the number of edges gets large, we may need asymptotic formulas for $t_n(y)$ that are valid when y goes to infinity with n. Formula (3.9) can be extended to

$$t_n(y) = \frac{\sqrt{2\pi}\,n^{n-1/2+y/2}}{2^{y/2}\,\Gamma(y/2)}\,\big(1 + O(y^{3/2}n^{-1/2})\big)\,, \tag{19.13}$$

uniformly for $1 \le y \le n^{1/3}$, using the proof technique of Lemma 3. Still larger values of y can be handled by using the saddle point method to derive the following general estimate:

$$t_{a\lambda n,n}(\lambda n + b) = \frac{n!\,e^{n\rho}\rho^{(a\lambda-1)n}\lambda^{(1-b)/2}}{2\sqrt{\pi n}\,(1-\rho)^{\lambda n}}\,\big(1 + O(\sqrt{\lambda}\,) + O(1/\sqrt{\lambda n}\,)\big)\,, \tag{19.14}$$

for fixed a and b as $\lambda \to 0$ and $\lambda n/(\log n)^2 \to \infty$, where

$$\rho = 1 + c\lambda - \sqrt{\lambda(1+c^2\lambda)} = 1 - \sqrt{\lambda} + c\lambda - \frac{c^2}{2}\lambda^{3/2} + O(\lambda^{5/2}), \quad c = \frac{1-a}{2}. \qquad (19.15)$$

For example, to estimate $t_{2r,n}(3r)$ when $r = n^{1/2}$, we can use (19.14) with $a = \frac{2}{3}$, $b = 0$, and $\lambda = 3n^{-1/2}$. The complicated dependence on ρ can also be expressed as

$$\frac{e^{n\rho}\rho^{(a\lambda-1)n}}{(1-\rho)^{\lambda n}} = \exp\bigl(n\bigl(1 - \tfrac{1}{2}\lambda\ln\lambda + \tfrac{1}{2}\lambda + (\tfrac{1}{3} - a)\lambda^{3/2} - \tfrac{1}{4}a^2\lambda^2 + O(\lambda^{5/2})\bigr)\bigr), \qquad (19.16)$$

which is sufficiently accurate if $\lambda \le n^{-1/4}$.

20. The evolution of uncleanness. We get further insight into the behavior of an evolving multigraph by studying how its reduced multigraph $\overline{\overline{M}}$ changes as the excess increases. Let's review the theory of Section 9 in light of what we have learned since then. The generating function for the cyclic part of all multigraphs having excess r and deficiency d is

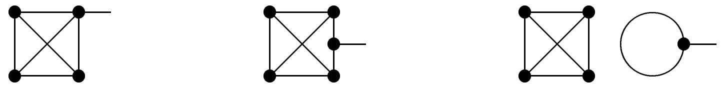
$$E_{rd}(z) = e_{rd}\,\frac{T(z)^{2r-d}}{\bigl(1-T(z)\bigr)^{3r-d+1/2}}. \qquad (20.1)$$

We can interpret it as follows, ignoring the constant factor e_{rd} for a moment: There is a reduced multigraph $\overline{\overline{M}}$ having $\nu = 2r - d$ vertices and $\mu = 3r - d$ edges; each vertex has degree ≥ 3, where a self-loop is considered to add 2 to the degree. We can obtain all cyclic multigraphs M that reduce to $\overline{\overline{M}}$ by a two-step process. First we insert 0 or more vertices of degree 2 on each edge; and we also construct any desired number of cycles, as separate components. All of the newly constructed vertices, including the vertices in the cycles, have degree 2. This first step creates a set of multigraphs with the univariate generating function $z^\nu(1-z)^{-\mu}(1-z)^{-1/2}$, because each edge subdivision corresponds to $(1-z)^{-1}$, and because the cycles are generated by $\exp(\frac{1}{2}z + \frac{1}{4}z^2 + \frac{1}{6}z^3 + \cdots) = (1-z)^{-1/2}$. Now we proceed to step two, which sprouts a rooted tree from every vertex; this changes z to $T(z)$ in the generating function.

The excess increases by 1 when we add a new edge $\langle x, y\rangle$ to M. How does the new edge change $\overline{\overline{M}}$? A moment's thought shows that $\overline{\overline{M}}$ will gain 2, 1, or 0 vertices; this means the deficiency will either stay the same or it will increase by 1 or 2.

In fact there is a nice algebraic and quantitative way to understand what happens, in terms of the generating function. Again we consider a two-step process: First we choose a vertex x of M; this means we apply the marking operator ϑ to the generating function. There are three cases: The marked vertex either belongs to a tree attached to one of the ν special vertices of $\overline{\overline{M}}$, or it belongs to a tree attached to a vertex within one of the μ edges, or it belongs to a tree attached to a vertex in some cycle. We represent Case 1 by attaching a "half-edge" to the existing vertex; we represent Case 2 by introducing a new vertex into the split edge and attaching a half-edge to it; we represent Case 3 by introducing a new vertex with a self-loop and attaching a half-edge to it.

A half-edge is like an edge but it touches only one vertex. For example, if $\overline{\overline{M}}$ is the multigraph K_4, the symbolic representations of the three possible outcomes of step 1 are



Case 1 Case 2 Case 3

Let's call this augmented multigraph $\overline{\overline{M}}'$.

A cyclic multigraph M' with a marked vertex can be reduced by attaching a half-edge to the marked vertex, then pruning all vertices of degree 1 and cancelling all vertices of degree 2. Conversely, the marked cyclic multigraphs that reduce to a given $\overline{\overline{M}}'$ are obtained by adding zero or more vertices to each edge (*including* the half edge), also adding cycles, then sprouting trees from each vertex. Thus the generating function for M' in Case 1 is

$$e_{rd}\,\frac{\nu\,T(z)^{\nu}}{\bigl(1-T(z)\bigr)^{\mu+3/2}}\,; \tag{20.2}$$

the ν in the numerator accounts for the number of vertices that can be chosen, and the extra $\bigl(1-T(z)\bigr)$ in the denominator accounts for the new half-edge. The generating function for M' in Case 2 is

$$e_{rd}\,\frac{\mu\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+5/2}}\,; \tag{20.3}$$

now we have μ edges that can be split, and we include an additional $T(z)$ in the numerator for the new vertex and an additional $\bigl(1-T(z)\bigr)^2$ in the denominator for the new half-edge and the additional split edge. Finally, the generating function for M' in Case 3 is

$$e_{rd}\,\frac{\frac{1}{2}\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+5/2}}\,; \tag{20.4}$$

as in Case 2, the diagram has gained one vertex and two edges. The factor $\frac{1}{2}$ is due to the compensation factor κ of a self-loop.

If our calculations are correct, the sum $(20.2)+(20.3)+(20.4)$ should be the result of applying ϑ to the overall generating function (20.1). And sure enough,

$$\vartheta\,\frac{T(z)^{\nu}}{\bigl(1-T(z)\bigr)^{\mu+1/2}}=\frac{\nu\,T(z)^{\nu}}{\bigl(1-T(z)\bigr)^{\mu+3/2}}+\frac{(\mu+\frac{1}{2})\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+5/2}}\,; \tag{20.5}$$

everything checks out fine.

The next step, choosing y, is the same, except that now we mark a vertex of M' and obtain M''. The transition from $\overline{\overline{M}}'$ to $\overline{\overline{M}}''$ again leads to three cases; we attach another half-edge and possibly split an existing edge or add a new self-loop. In particular, we might split the half-edge of $\overline{\overline{M}}'$. The change in the generating function is once again represented by (20.5), but this time ν and μ have to be adjusted to equal the number of vertices and edges of $\overline{\overline{M}}'$. The left term of (20.5) therefore becomes

$$\frac{\nu^2 T(z)^\nu}{\bigl(1-T(z)\bigr)^{\mu+5/2}} + \frac{\nu(\mu+\frac{3}{2})\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+7/2}}\,, \tag{20.6}$$

and the right term becomes

$$\frac{(\mu+\frac{1}{2})(\nu+1)\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+7/2}} + \frac{(\mu+\frac{1}{2})(\mu+\frac{5}{2})\,T(z)^{\nu+2}}{\bigl(1-T(z)\bigr)^{\mu+9/2}}\,. \tag{20.7}$$

Notice that the first term of (20.5) corresponds to the case that the deficiency increases by 1 when x is chosen, while the second term corresponds to the case where the deficiency stays the same. Similarly, the first terms of (20.6) and (20.7) correspond to an increase in deficiency when y is chosen, after x has already been marked.

By looking at the coefficients of these generating functions we can see why the deficiency rarely increases unless the total number of vertices in the cyclic part is not much larger than ν. Suppose we change the generating function to

$$F(z,s) = \frac{T(z)^\nu}{\bigl(1-s\,T(z)\bigr)^{\mu+1/2}}\,;$$

then

$$\frac{[z^n]\,\frac{\partial}{\partial s}\,F(z,s)\big|_{s=1}}{[z^n]\,F(z,s)|_{s=1}}$$

will be the average number of tree-root vertices that appear within the edges of $\overline{\overline{M}}$. For fixed ν and μ as $n\to\infty$ this number is

$$\frac{[z^n]\,(\mu+\frac{1}{2})\,T(z)^{\nu+1}\bigl(1-T(z)\bigr)^{-\mu-3/2}}{[z^n]\,T(z)^\nu\bigl(1-T(z)\bigr)^{-\mu-1/2}} = \frac{(\mu+\frac{1}{2})t_n(\mu+\frac{3}{2})}{t_n(\mu+\frac{1}{2})}\bigl(1+O(n^{-1/2})\bigr)\,, \tag{20.8}$$

which is approximately $\sqrt{\mu n}$ by (3.9), when μ is large. Thus, there are about $\sqrt{\mu n}$ tree roots, only ν of which will increase the deficiency when chosen; almost all choices of x and y will fall in trees that add new vertices to $\overline{\overline{M}}$ and $\overline{\overline{M}}'$.

If we replace one of the factors $T(z)$ in the numerator of the generating function by $\vartheta T(z) = T(z)/\bigl(1 - T(z)\bigr)$, we multiply the coefficient of z^n by the average size of a rooted tree; we find that each rooted tree contains about $\sqrt{n/\mu}$ vertices.

The number n in these calculations has been the number of vertices in the cyclic part of a multigraph, and the number μ is $3r$. Let's return to our other notational convention, where n is the total number of vertices in the evolving multigraph and $m = \frac{n}{2}(1 + \mu n^{-1/3})$ is the total number of edges. Recall that the average excess r grows as $\frac{2}{3}\mu^3$, for $\mu \le n^{1/12}$; the size of the cyclic part, similarly, has order $\mu n^{2/3}$. The probability that a random new edge falls in the cyclic part (and therefore increases the excess) is therefore of order $(\mu n^{2/3}/n)^2 = \mu^2 n^{-2/3}$; we must add about $n^{2/3}/\mu^2$ more edges before the excess increases. And when it does, the probability of choosing a "bad" x or y, making the new multigraph unclean, is the ratio of $2r$ to the *total* number of tree roots, which is of order

$$\frac{2r}{\sqrt{3r(\mu n^{2/3})}} \approx \frac{\mu^3}{\sqrt{\mu^4 n^{2/3}}} = \mu n^{-1/3}\,.$$

We will probably have to do $n^{1/3}/\mu$ augmentations of excess, adding $(n^{2/3}/\mu^2)(n^{1/3}/\mu) = n/\mu^3$ more edges, before we reach an unclean multigraph. That is why the multigraph tends to stay clean until $\mu = n^{1/12}$, as asserted in Theorem 7.

After x and y are chosen to form the endpoints of a new edge, a third step takes place: This new edge is merged or integrated with the other edges. Symbolically, the two half-edges for x and y are now spliced together. We can complete our study of how the generating function changes at the time of excess augmentation by considering this third and final step.

It is easiest to consider the *inverse* of the final step, namely the operation of marking an edge whose removal would *decrease* the excess. Such an edge must be in the complex part, not the acyclic or unicyclic part. The operator that corresponds to marking an arbitrary edge in a complex multigraph of excess r is $r + \vartheta$, because this multiplies the coefficient of z^n by $r + n$, the total number of edges. However, we also need to figure out the generating function for "insignificant" edges, edges whose removal would leave the excess unchanged. Such edges can be described by an ordered pair consisting of a rooted tree and a multigraph of excess r with a marked vertex; one end of the edge is attached to the marked vertex and the other end is attached to the root of the tree. Thus the appropriate operator for insignificant edges is $T(z)\vartheta$. Altogether we find that the generating function that corresponds to marking a significant edge, given a family of complex multigraphs of excess r, is $r + \vartheta - T\vartheta$. We also should multiply this by two, because we assign an orientation to the edge with the ordered pair $\langle x, y\rangle$.

When the operator $2\bigl(r + \vartheta - T(z)\vartheta\bigr)$ is applied to a generating function of the form

$T(z)^\nu/\bigl(1-T(z)\bigr)^\mu$, with $\mu=\nu+r$, we get

$$\begin{aligned}2\bigl(r+\bigl(1-T(z)\bigr)\vartheta\bigr)\,\frac{T(z)^\nu}{\bigl(1-T(z)\bigr)^\mu} &= \frac{2\bigl((r+\nu)T(z)^\nu\bigr)}{\bigl(1-T(z)\bigr)^\mu}+\frac{2\mu\,T(z)^{\nu+1}}{\bigl(1-T(z)\bigr)^{\mu+1}}\\ &= \frac{2\mu\,T(z)^\nu}{\bigl(1-T(z)\bigr)^{\mu+1}}\,. \end{aligned} \tag{20.9}$$

Therefore the inverse operation we seek, which merges an ordered $\langle x,y\rangle$ into the set of existing edges, takes

$$\frac{T(z)^\nu}{\bigl(1-T(z)\bigr)^{\mu+1}} \;\longmapsto\; \frac{1}{2\mu}\,\frac{T(z)^\nu}{\bigl(1-T(z)\bigr)^\mu}\,. \tag{20.10}$$

For example, the first term of (20.6) will go into

$$\frac{\nu^2\,T(z)^\nu}{2(\mu+1)\bigl(1-T(z)\bigr)^{\mu+3/2}}\,.$$

(First we multiply by $\bigl(1-T(z)\bigr)^{1/2}$ to get rid of the unicyclic components, then we apply the inverse operation (20.10), then we put the unicyclic components back.)

Altogether we find that the generating function $T(z)^{2r-d}/\bigl(1-T(z)\bigr)^{3r-d+1/2}$ for cyclic multigraphs of excess r and deficiency d makes the following contributions to the generating functions for cyclic multigraphs of excess $r+1$ and deficiencies d, $d+1$, and $d+2$, according to (20.6), (20.7), and (20.10):

$$\begin{aligned}&\frac{(6r-2d+5)(6r-2d+1)}{8(3r-d+3)}\,\frac{T(z)^{2r+2-d}}{\bigl(1-T(z)\bigr)^{3r+3-d+1/2}}\\ &\quad+\frac{\bigl((2r-d)(6r-2d+3)+(2r-d+1)(6r-2d+1)\bigr)}{4(3r-d+2)}\,\frac{T(z)^{2r+1-d}}{\bigl(1-T(z)\bigr)^{3r+2-d+1/2}}\\ &\quad+\frac{(2r-d)^2}{2(3r-d+1)}\,\frac{T(z)^{2r-d}}{\bigl(1-T(z)\bigr)^{3r+1-d+1/2}}\,. \end{aligned} \tag{20.11}$$

This is essentially the same as the recurrence relation for e_{rd} in (5.11)–(5.13).

We can illustrate the observations of this section by introducing another partial ordering analogous to Figure 1. Every evolving graph or multigraph traces a path in Figure 2, just as it does in Figure 1; but in Figure 2 the state (r,d) represents excess r and deficiency d. Fractions in brackets above each state are the coefficients e_{rd} of the generating function (5.10). Fractions on the arrows are not transition probabilities but rather the amounts by which each generating function coefficient affects the coefficients at the next level; these fractions are the coefficients in (20.11).

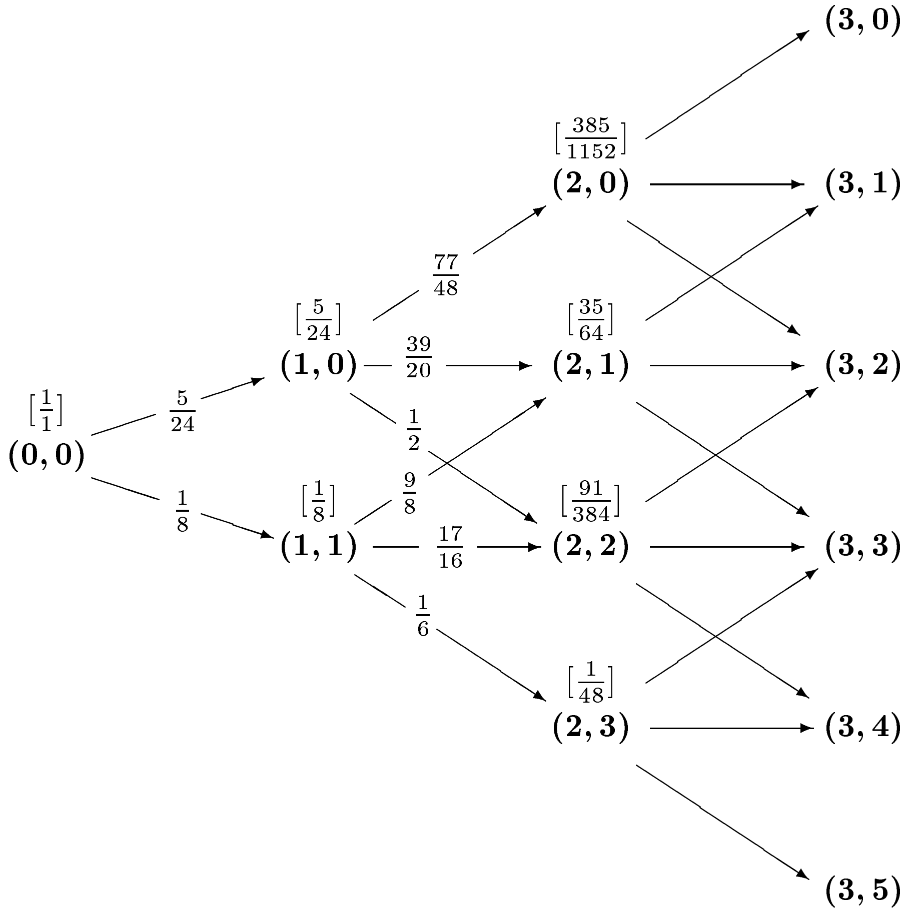


Figure 2. The evolution of deficiency. Each configuration (r,d) stands for a graph or multigraph whose complex part reduces to a multigraph with $2r-d$ vertices and $3r-d$ edges, when vertices of degrees 1 and 2 are eliminated. A graph or multigraph with deficiency 0 is called "clean"; the reduced multigraphs in such cases are 3-regular. When r is small, each unit increase in deficiency occurs with probability of order $n^{-1/3}$; therefore most random graphs stay clean until r is quite large.

21. Waiting for uncleanness. We have seen that a graph almost surely stays clean while it has $\frac{1}{2}(n+\mu n^{2/3})$ edges, as long as μ is $o(n^{1/12})$. What happens when μ gets a bit larger? Another contour integral provides the answer; in this one, we rescale μ in preparation for the appearance of the giant component, but we allow μ to be small enough that there is a substantial overlap with the estimate (10.1) of Lemma 3.

Lemma 7. *If $m = \frac{1}{2}(n+\mu n)$ and $r = \frac{2}{3}\mu^3 n + \rho\sqrt{\mu^3 n}$, we have*

$$\frac{2^m m!\, n!\, e_r}{n^{2m}(n-m+r)!}\,[z^n]\,\frac{U(z)^{n-m+r}T(z)^{2r}}{\bigl(1-T(z)\bigr)^{3r+y}}$$
$$= B(y,\mu,\rho,n)\exp\Bigl(O\bigl((1+|\rho|^3)\mu^{-3/2}n^{-1/2}+(1+|\rho|)\mu^{5/2}n^{1/2}\bigr)\Bigr)\,, \qquad (21.1)$$

where

$$B(y,\mu,\rho,n)=\sqrt{\frac{3}{20\pi n}}\,\mu^{-1-y}\exp\left(-\frac{2}{3}\,\mu^4 n-\frac{3}{20}\,\rho^2\right)\,, \tag{21.2}$$

uniformly for $n^{-1/3}\log n\le\mu\le n^{-1/5}$, $|\rho|\le\frac{2}{3}\mu^{3/2}n^{1/2}$, *and fixed* y *as* $n\to\infty$.

Proof. This is the sort of lemma for which computer algebra really pays off. We can begin by using Stirling's approximation to show that

$$\begin{aligned}\log\left(\frac{2^m m!\,n!\,e_r}{n^{2m}(n-m+r)!\,2^{n-m+r}}\right)&=-n+3r\ln\mu-\tfrac{5}{6}\mu^3 n\\ &\quad-\tfrac{3}{2}\ln\mu+\tfrac{1}{2}\ln\tfrac{3}{2}+\tfrac{2}{3}\mu^4 n+\tfrac{3}{4}\rho^2\\ &\quad+O\big((1+|\rho|^3)\mu^{-3/2}n^{-1/2}+(1+|\rho|)\mu^{5/2}n^{1/2}\big)\,.\end{aligned} \tag{21.3}$$

Now we express the remaining factor by using the trick of (10.11):

$$[z^n]\,\frac{(2U(z))^{n-m+r}T(z)^{2r}}{\big(1-T(z)\big)^{3r+y}}=\frac{1}{2\pi i}\oint(1-z)^{1-y}e^{g(z)}\,\frac{dz}{z}\,, \tag{21.4}$$

where

$$g(z)=nz+(3r-m)\ln z-3r\ln(1-z)+(n-m+r)\ln(2-z)\,. \tag{21.5}$$

As before we can show that the asymptotic value of the integral depends only on the behavior of the integrand near $z=1$. This time we need not worry about a three-legged saddle point, because we are sufficiently far from the critical region near $\mu=0$. A good path of integration turns out to be $z=1-\alpha+it\mu^{-1/2}n^{-1/2}$, where $\alpha=\mu-\frac{2}{3}\mu^2+\frac{3}{5}\rho\mu^{-1/2}n^{-1/2}$. Indeed, some beautiful cancellation occurs in the most significant terms:

$$\begin{aligned}g(1-\alpha+it\mu^{-1/2}n^{-1/2})&=g(1-\alpha)-\tfrac{5}{2}t^2+O\big((\mu^{5/2}n^{1/2}+\mu^{-3/2}n^{-1/2}\rho^2)t\big)\\ &\quad+O\big(((1+|\rho|)\mu^{-3/2}n^{-1/2}+\mu)t^2\big)\,,\end{aligned} \tag{21.6}$$

when $|t|\le\log n$. The O bounds follow from the fact that the power series for $\log z$, $\log(1-z)$, and $\log(2-z)$ converge in the stated ranges.

The other factors of the integrand, besides $e^{g(z)}$, are

$$(1-z)^{1-y}\,\frac{dz}{z}=\mu^{1-y}i\mu^{-1/2}n^{-1/2}\,dt\sum_{k=0}^{\infty}\mu^k\beta^{k+1-y}\,,$$

where $\beta=(\alpha-it\mu^{-1/2}n^{-1/2})/\mu=1-\frac{2}{3}\mu+(\frac{3}{5}\rho-it)\mu^{-3/2}n^{-1/2}$. We can now write the integral as a factor independent of t times

$$\int e^{-5t^2/2}(1+\gamma_1 t+\gamma_2 t^2+\cdots)\,dt\,, \tag{21.7}$$

where the γ's are functions of μ and ρ, and the series is convergent for $|t| \le \log n$. The integrand is superpolynomially small when $t = \pm\log n$; hence we can bound the error terms for $|t| \le \log n$, then integrate from $-\infty$ to ∞, showing that (21.7) is

$$\sqrt{\frac{2\pi}{5}}\,\bigl(1 + O(\mu^{5/2}n^{1/2} + (1+\rho^2)\mu^{-3/2}n^{-1/2})\bigr)\,. \tag{21.8}$$

Finally we observe that the other factors nicely cancel the leading terms of (21.3); only (21.1) and (21.2) are left. The overall formula (21.1) has a weaker estimate than (21.8) because Stirling's approximation (21.3) is more sensitive to the value of ρ and because of the term $g(1-\alpha)$. $\square$

Notice that Lemma 7 matches the first estimate of Lemma 5, which says that the asymptotic probability of excess r is like that for a normal distribution with mean $\frac{2}{3}\mu^3$ and variance of order μ^3, as long as $r = O(\mu^3)$. On the other hand, the extreme tails for larger values of r are not as small as they would be in a normal distribution; they decrease only as shown in the second estimate of Lemma 5. For example, with probability 100^{-m} all edges will join vertices in the first $n/10$ vertices; so there will be at least $0.9n$ isolated vertices, and the excess will be at least $m - n + 0.9n > 0.4n$.

Theorem 12. *The probability that a random multigraph with n vertices and $m = \frac{1}{2}(n + \mu n)$ edges is clean, when $0 \le \mu \le n^{-1/5}$, is*

$$\exp\bigl(-\tfrac{2}{3}\mu^4 n + O\bigl((\mu^{5/2}n^{1/2}\log n + \mu^{-3/2}n^{-1/2}(\log n)^3\bigr)\bigr)\,. \tag{21.9}$$

Proof. The probability decreases as μ increases. Therefore we need to verify the result only for μ greater than $n^{-3/11}$ or so, when the error estimate $\mu^{-3/2}n^{-1/2}(\log n)^3$ does not swamp the main term $\exp(-\frac{2}{3}\mu^4 n) = 1 - \frac{2}{3}\mu^4 n + O(\mu^8 n^2)$.

Formula (21.1) is the probability that a random graph or multigraph with m edges is clean and has excess r, if we set $y = \frac{1}{2}$. That probability is superpolynomially small unless $|\rho| \le \log n$, because of the term $-\rho^2$ in the exponent. Extremely large values of ρ, not covered by the hypotheses of Lemma 7, are also negligible. Therefore we can sum over r by integrating over ρ from $-\log n$ to $+\log n$; and we can then extend the integral from $-\infty$ to ∞ without changing its asymptotic value. Hence the probability of cleanliness is

$$n^{-1/2}\sqrt{\frac{3}{20\pi}}\,\mu^{-3/2}\exp\bigl(-\tfrac{2}{3}\mu^4 n\bigr)\int_{-\infty}^{\infty} e^{-3\rho^2/20}\sqrt{\mu^3 n}\,d\rho = \exp\bigl(-\tfrac{2}{3}\mu^4 n\bigr)\,,$$

plus the error term. Another nice bit of cancellation. $\square$

Corollary. *The average number of edges added to an evolving multigraph until it first becomes unclean is*

$$\tfrac{1}{2}n + \frac{3^{1/4}\Gamma\bigl(\frac{1}{4}\bigr)}{2^{13/4}}\,n^{3/4} + O(n^{8/11+\epsilon})\,, \tag{21.10}$$

and the standard deviation is of order $n^{3/4}$.

Proof. The stated average number is $\sum_{m\ge 0} p_m$, where p_m is the probability in the theorem. When $\mu \le 0$, the probability of uncleanliness is $O(n^{-1/3})$ by Theorem 7, so the sum for $0 \le m < \frac{1}{2}n$ is $\frac{1}{2}n - O(n^{2/3})$. When $0 \le \mu \le n^{-3/11}(\log n)^{6/11}$, the probability of uncleanliness is $O(n^{-1/11}(\log n)^{24/11})$ by (21.9); after that the error is negligible in comparison with the integral

$$\tfrac{1}{2}n \int_0^\infty e^{-(2/3)\mu^4 n}\, d\mu = \tfrac{1}{2}n^{3/4}\,\tfrac{1}{4}\left(\tfrac{3}{2}\right)^{1/4} \int_0^\infty e^{-u}u^{-3/4}\, du = cn^{3/4}\,,$$

where c is the coefficient of $n^{3/4}$ in (21.10). This proves (21.10).

The expected value of m^2 at the stopping time is $\sum_{m\ge 0}(2m+1)p_m$, and we need to be especially careful when evaluating this sum; the simple estimate $p_m = 1 - O(n^{-1/3})$ for $m \le \frac{1}{2}n$ will not do, because it will obliterate significant terms by adding $O(n^{5/3})$. Appropriate accuracy is maintained by computing the expected value of $(m - \frac{1}{2}n)^2$, which is

$$\frac{n^2}{4} + \sum_{m\ge 0}(2m+1-n)p_m = \sum_{m=0}^{n/2}(n-2m)(1-p_m) + \sum_{m=n/2}^{\infty}(2m-n)p_m + O(n)\,.$$

We can show that the terms for $m \le \frac{1}{2}n$ are now negligible, because the cleanliness probability p_m is bounded below by the probability that a multigraph with m edges has excess 0. Therefore $1 - p_m = O\bigl(n^2/(n-2m)^3\bigr)$ when $m \le m_0 = \frac{1}{2}n - n^{2/3+\epsilon}$, by the remarks preceding (13.23); and

$$\sum_{m=0}^{n/2}(n-2m)(1-p_m) = \sum_{m=0}^{m_0} O\left(\frac{n^2}{(n-2m)^2}\right) + \sum_{m=m_0}^{n/2} O(n-2m) = O(n^{4/3-\epsilon}) + O(n^{4/3+2\epsilon})\,.$$

The other terms can be approximated by

$$\sum_{m=n/2}^{\infty}(2m-n)p_m = \int_0^\infty \frac{n^2\mu}{2} e^{-(2/3)\mu^4 n}\, d\mu + O(n^{16/11+\epsilon})\,,$$

with an error estimate coming from the range $0 \le \mu \le n^{-3/11+\epsilon}$ as before. It follows that the variance is asymptotic to this integral minus the square of $\bigl((21.10) - \frac{1}{2}n\bigr)$, namely $\bigl(3^{1/2}\Gamma(\frac{1}{2})2^{-7/2} - c^2\bigr)n^{3/2}$.

Incidentally, the value of c is approximately 0.50155, and the standard deviation is approximately $0.1407n^{3/4}$. $\square$

Once a graph begins to get dirty, its deficiency rises rapidly. For fixed d we can estimate the probability of excess r and deficiency d by taking $y = \frac{1}{2} - d$ and multiplying (21.1) by $r^d/d!$, because of (7.16). The fact that (21.1) has $T(z)^{2r}$ in the numerator instead of $T(z)^{2r-d}$ is unimportant, since $T(z)^{2r} = T(z)^{2r-d}\sum\binom{d}{k}\bigl(T(z)-1\bigr)^k$. We obtain a probability about $(\frac{2}{3}\mu^4 n)^d/d!$ times as large as before, but this is damped rapidly by the factor $\exp(-\frac{2}{3}\mu^4 n)$ when μ becomes greater than $n^{-1/4}$. We will look further at the growth of deficiency in section 23.

22. A closer look. The structure theory of section 20 gives us more detailed information about what happens when an evolving multigraph first changes from clean to unclean. We learned in that section that the process of adding a new edge $\langle x, y\rangle$ can be broken into three parts, namely the introduction of half-edges at x and y followed by the joining of those two edges. The deficiency can increase by 1 during each of the first two stages.

The probability that a clean graph becomes potentially deficient when a half-edge is attached to x is the probability that the image $\overline{\overline{\overline{M}}}'$ of the half-edge after pruning and cancellation does not create a new vertex not in $\overline{\overline{\overline{M}}}$. According to the analysis of section 20, the expected number of times this happens is

$$p_1(n) = \sum_m \frac{2^m m!\, n!}{n^{2m+1}} \, [w^m z^n] \, G_1(w, z) \,, \tag{22.1}$$

$$\begin{aligned} G_1(w, z) &= e^{U(wz)/w} \sum_{r\ge 0} e_r w^r \, \frac{2r\, T(wz)^{2r}}{\bigl(1 - T(wz)\bigr)^{3r+3/2}} \\ &= \tfrac{5}{12} w^3 z^2 + \tfrac{5}{24} \, (2w^3 + 13w^4) \, z^3 + \cdots \,. \end{aligned} \tag{22.2}$$

The factor $2r$ covers the deficient choices of x, as in the first term of (20.5).

Actually (22.2) is an overestimate, because some apparently bad choices of x are "false alarms." If the half-edge of x does not add a vertex to $\overline{\overline{\overline{M}}}$, there's still a possibility that y will be chosen in the acyclic part; then the new edge $\langle x, y\rangle$ will not increase the excess and the multigraph will still be clean. The expected number of false alarms is

$$p_1'(n) = \sum_m \frac{2^m m!\, n!}{n^{2m+2}} \, [w^m z^n] \, \frac{T(wz)}{w} \, G_1(w, z) \,. \tag{22.3}$$

The multigraph becomes unclean when y is chosen if the half-edge for y prunes and cancels to a reduced multigraph $\overline{\overline{\overline{M}}}''$ having the same $2r+1$ vertices as $\overline{\overline{\overline{M}}}'$. This occurs with probability

$$p_2(n) = \sum_m \frac{2^m m!\, n!}{n^{2m+2}} \, [w^m z^n] \, G_2(w, z) \,, \tag{22.4}$$

$$\begin{aligned} G_2(w, z) &= e^{U(wz)/w} \sum_{r\ge 0} e_r w^r \, \frac{(3r + \frac12)(2r+1)\, T(wz)^{2r+1}}{\bigl(1 - T(wz)\bigr)^{3r+7/2}} \\ &= \tfrac12 wz + \tfrac14 \, (2w + 9w^2) z^2 + \cdots \,. \end{aligned} \tag{22.5}$$

Consequently we must have

$$p_1(n) - p_1'(n) + p_2(n) = 1 \tag{22.6}$$

for all n; this identity is a nontrivial property of the bivariate generating functions $G_1(w,z)$ and $G_2(w,z)$. When $n=6$, for example, computer calculations show that

$$p_1(6) = \frac{10288260775}{22039921152} \approx 0.4668\,; \qquad p_1'(6) = \frac{38865625}{612220032} \approx 0.0635\,;$$
$$p_2(6) = \frac{13150822877}{22039921152} \approx 0.5967\,.$$

We can use Lemma 7 to calculate the approximate values of these quantities when n is large, ignoring extreme terms not covered by that lemma:

$$\begin{aligned}
p_1(n) &\sim \frac{1}{n}\int_0^\infty \int_{-\infty}^\infty 2rB(\tfrac32,\mu,\rho,n)(\mu^{3/2}n^{1/2}\,d\rho)(\tfrac12 n\,d\mu) \\
&= 2^{-7/4}3^{-1/4}\Gamma(\tfrac34)\,n^{1/4}\,; \qquad (22.7)\\
p_1'(n) &= \sum_m \frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\,T(wz)G_1(w,z) \\
&\sim \frac12 \int_0^\infty m^{-1}\int_{-\infty}^\infty 2rB(\tfrac32,\mu,\rho,n)(\mu^{3/2}n^{1/2}\,d\rho)(\tfrac12 n\,d\mu) \\
&\sim 2^{-7/4}3^{-1/4}\Gamma(\tfrac34)n^{1/4}\,; \qquad (22.8)\\
p_2(n) &\sim \frac{1}{n^2}\int_{n^{-1/3}}^\infty \int_{-\infty}^\infty (3r+\tfrac12)(2r+1)B(\tfrac72,\mu,\rho,n)(\mu^{3/2}n^{1/2}\,d\rho)(\tfrac12 n\,d\mu) \\
&\sim \frac12\,. \qquad (22.9)
\end{aligned}$$

Notice that $p_1(n)$ and $p_1'(n)$ are unbounded, so they must be regarded as expected values (not probabilities). But $p_1(n)-p_1'(n)$ is the probability of a "true alarm." As we might have guessed, the transition from clean to unclean occurs about half the time when x is chosen, half the time when y is chosen.

23. Giant growth. We know from the classical theory [13] that a giant component will emerge when the number of edges is $\frac{n}{2}(1+\mu)$ for a positive constant μ. The classical theory deals with graphs, but the same phenomenon will occur with multigraphs, because random graphs are generated by the multigraph process if we discard self-loops and duplicate edges; discarded edges do not affect the size of components, and comparatively few edges are discarded until the graph has gotten rather dense (see [4]).

Instead of relying on the classical theory, we can also deduce the existence of a giant component by studying the generating function $G(w,z)$. The proof is indirect: First we count the vertices that lie in trees and unicyclic components, showing that there probably aren't too many of those. Then we show that it is improbable to have two distinct complex components.

The first part is easy, because there is a simple closed form for the expected number of vertices in trees. If we mark just the vertices in trees of size k, by differentiating the generating function

$$G(w,z)\,\exp\bigl(-k^{k-2}w^{k-1}z^k/k! + k^{k-2}w^{k-1}z^k s^k/k!\bigr)$$

with respect to s and setting $s=1$, we see that the expected number of such vertices is just

$$\begin{aligned}\frac{2^m m!\,n!}{n^{2m}}\,[w^m z^n]\,\frac{k^{k-1}}{k!}\,w^{k-1}z^k G(w,z) &= \frac{2^m m!\,n!\,k^{k-1}}{n^{2m}\,k!}\,[w^{m-k+1}z^{n-k}]\,G(w,z)\\ &= \frac{2^m m!\,n!\,k^{k-1}}{n^{2m}\,k!}\,\frac{(n-k)^{2(m-k+1)}}{2^{m-k+1}(m-k+1)!\,(n-k)!}\,;\end{aligned}$$

this can be written

$$\frac{k^{k-1}}{k!}\,\frac{2^{k-1}m^{\underline{k-1}}n^{\underline{k}}}{(n-k)^{2k-2}}\left(\frac{n-k}{n}\right)^{2m} \tag{23.1}$$

in terms of falling factorial powers $x^{\underline{k}} = x(x-1)\ldots(x-k+1)$.

Asymptotically, we have $n^{\underline{k}} = n^k\bigl(1+O(k^2/n)\bigr)$ and $(n-k)^k = n^k\bigl(1+O(k^2/n)\bigr)$ for all k; also $(1-k/n)^n = e^{-k}\bigl(1+O(k^2/n)\bigr)$ for $k \le \sqrt{n}$ and $(1-k/n)^n \le e^{-k}$ for $k \le n$. If μ is a nonzero constant, $\mu > -1$, and if $m = \frac{n}{2}(1+\mu)$, expression (23.1) is

$$\frac{n}{1+\mu}\,\frac{k^{k-1}}{k!}\,(1+\mu)^k e^{-k(1+\mu)}\bigl(1+O(\tfrac{k^2}{n})\bigr) \tag{23.2}$$

for $k \le \sqrt{n}$; and it is superpolynomially small when $k=\sqrt{n}$, because it is $O\bigl(\bigl((1+\mu)e^{-\mu}\bigr)^k k^{1/2}\bigr)$ and $(1+\mu)e^{-\mu} < 1$. It is also superpolynomially small when $k > \sqrt{n}$, because we will prove in section 27 below that a continuous approximation of the quantity

$$e^k\sqrt{\frac{m-k}{m}}\sqrt{\frac{n-k}{n}}\,\frac{2^k\,m^{\underline{k}}\,n^{\underline{k}}}{(n-k)^{2k}}\left(\frac{n-k}{n}\right)^{2m} \tag{23.3}$$

decreases when k increases.

Let σ be defined by the formula

$$(1+\mu)e^{-\mu} = (1-\sigma)e^{\sigma}\,, \qquad \sigma = \mu + O(\mu^2)\,. \tag{23.4}$$

Then σ is the quantity called $1 - x\bigl(\frac{1}{2}(1+\mu)\bigr)$ in [13], and we have

$$\begin{aligned}\sum_{k\ge1}\frac{k^{k-1}}{k!}\bigl((1+\mu)e^{-(1+\mu)}\bigr)^k &= \sum_{k\ge1}\frac{k^{k-1}}{k!}\bigl((1-\sigma)e^{-(1-\sigma)}\bigr)^k\\ &= T\bigl((1-\sigma)e^{-(1-\sigma)}\bigr) = 1-\sigma\,,\end{aligned}$$

when μ is positive. By summing (23.2) over all k, we conclude that the expected total number of vertices in trees is

$$\frac{1-\sigma}{1+\mu}\,n + O(\sigma^{-3})\,; \tag{23.5}$$

the error term $O(\sigma^{-3})$ here comes from summing $\vartheta^2 T\bigl((1-\sigma)e^{-(1-\sigma)}\bigr)$, which brings a factor of k^2 into each term.

For example, if $1+\mu = \ln 4$, we have $1-\sigma = \ln 2$, because $\frac14 \ln 4 = \frac12 \ln 2$. When the number of edges reaches $n \ln 2$ the expected number of vertices in trees will be $\frac12 n$. And in general when the number of edges reaches $\frac{n}{2x} \ln \frac{1}{1-x}$, the expected number of vertices in trees will be $(1-x)n$, for $0 < x < 1$.

The expected number of vertices in unicyclic components can be found in a similar way, by differentiating

$$G(w,z)\, e^{-V(wz)+V(wzs)}$$

with respect to s and setting $s=1$. The generating function is

$$\frac12 \,\frac{T(wz)}{\bigl(1-T(wz)\bigr)^2}\, G(w,z) = \bigl(\vartheta V(wz)\bigr)\, G(w,z)\,, \tag{23.6}$$

and we have

$$\frac{T(z)}{\bigl(1-T(z)\bigr)^2} = \sum_{k\ge1} \frac{k^k Q(k)}{k!}\, z^k \tag{23.7}$$

by (3.12). The expected number of vertices belonging to unicyclic components of size k therefore can be expressed in closed form, analogous to (23.1):

$$\begin{aligned} &\frac12\, \frac{k^k Q(k)}{k!}\, \frac{2^m m!\, n!}{n^{2m}}\, [w^{m-k} z^{n-k}]\, G(w,z) \\ &\qquad = \frac12\, \frac{k^k Q(k)}{k!}\, \frac{2^k m^{\underline{k}} n^{\underline{k}}}{(n-k)^{2k}} \left(\frac{n-k}{n}\right)^{2m}. \end{aligned} \tag{23.8}$$

Summing over k, and breaking the sum into two parts $k \le \sqrt{n}$ and $k > \sqrt{n}$ as above, now yields

$$\frac12 \sum_{k\ge1} \frac{k^k Q(k)}{k!} \bigl((1+\mu)e^{-(1+\mu)}\bigr)^k \bigl(1 + O(\frac{k^2}{n})\bigr) = \frac{1-\sigma}{2\sigma^2} + O(\sigma^{-6} n^{-1})\,. \tag{23.9}$$

(We will obtain sharper bounds in section 27.)

We have assumed in this discussion that μ is a constant. But our relatively coarse asymptotic arguments are in fact valid if μ varies with n, provided that it is not too small. Relation (23.4) defines σ as an analytic function of μ,

$$\begin{aligned} \sigma = \mu &- \frac23 \mu^2 + \frac49 \mu^3 - \frac{44}{135}\mu^4 + \frac{104}{405}\mu^5 - \frac{40}{189}\mu^6 \\ &+ \frac{7648}{42525}\mu^7 - \frac{2848}{18225}\mu^8 + \frac{31712}{229635}\mu^9 - \frac{23429344}{189448875}\mu^{10} + \cdots\,, \end{aligned} \tag{23.10}$$

where the power series converges for $|\mu| < 1$. The quantity $\bigl((1+\mu)e^{-\mu}\bigr)^k$ is superpolynomially small for $k = \sqrt{n}$ if μ is at least, say, $n^{-1/4}\log n$. We are therefore justified in using (23.5)+(23.9) as the expected number of vertices in non-complex components whenever $\mu \geq n^{-1/4}\log n$.

Suppose $\mu = n^{-1/4}\log n$. Then the expected number of vertices in unicyclic components is approximately $\frac{1}{2}\sigma^{-2} \sim \frac{1}{2}\mu^{-2} = \frac{1}{2}n^{1/2}(\log n)^{-2}$, and a similar argument proves that the expected value of the square of this number is approximately $\frac{5}{4}\sigma^{-4} \sim \frac{5}{4}n(\log n)^{-4}$. So the probability of choosing two vertices in unicyclic components is approximately $\frac{5}{4}n^{-1}(\log n)^{-4}$. This probability decreases steadily as m increases, but even if it stayed fixed we would have to add about $\frac{4}{5}n(\log n)^4$ more edges before hitting two unicyclic vertices, i.e., before creating a new bicyclic component. By that time the expected number of vertices in trees and unicyclic components will be nearly zero, so the multigraph will almost surely contain no such vertices. Therefore, if there is only one complex component present when $\mu = n^{-1/4}\log n$, there will almost surely be only one complex component from that time on; it will become gigantic. (We will obtain sharper results in section 27; see Lemma 9 and its corollary.)

Let's look more closely at what happens as the giant component develops. According to (23.5), it will have approximately

$$\left(1 - \frac{1-\sigma}{1+\mu}\right)n = \frac{\mu+\sigma}{1+\mu}\,n = 2\mu n + O(\mu^2 n) \tag{23.11}$$

vertices when $m = \frac{n}{2}(1+\mu)$; this is substantially larger than the number $\frac{1}{2}\mu^{-2} + O(\mu^{-1})$ of unicyclic vertices. When m increases by 1, the value of μn increases by 2, so (23.11) increases by 4. Notice that (23.11) agrees with the leading term of (15.11).

We saw in section 21 that the expected excess r is approximately $\frac{2}{3}\mu^3 n$ when $m = \frac{n}{2}(1+\mu)$, at least for $0 \leq \mu \leq n^{-1/4}$. We will prove momentarily that this relationship continues to hold as long as μ remains $o(1)$; but before giving the proof, let's look at the situation heuristically. The probability that a new edge increases the excess is the probability that both of its endpoints lie in the cyclic part, namely $(2\mu)^2$. The change in r with respect to m is $(dr/d\mu)(d\mu/dm) = (2\mu^2 n)(2/n)$, and this too is $(2\mu)^2$. So the relation $r = \frac{2}{3}\mu^3 n$ is consistent with (23.11) when μ is not too large.

The expected value of the deficiency d turns out to be approximately $\frac{2}{3}\mu^4 n$, about μ times r. Heuristic justification comes from the considerations of section 20: When a new edge $\langle x, y\rangle$ falls in the cyclic part, the probability that x is "bad" (in the sense that it increases the deficiency) will be the number of reduced vertices $2r - d$ divided by the square root of $3r - d$ times the size of the complex part (see the remarks following (20.8)). So it will be approximately $\frac{4}{3}\mu^3 n$ divided by $\bigl((2\mu^3 n)(2\mu n)\bigr)^{1/2}$, namely $\frac{2}{3}\mu$. The same holds for y. Hence the expected increase in d, given that r increases, is $\frac{4}{3}\mu$. And the derivative of $\frac{2}{3}\mu^4 n$ with respect to μ is indeed $\frac{4}{3}\mu$ times the derivative of $\frac{2}{3}\mu^3 n$.

In order to carry out a rigorous proof as μ increases from $n^{-1/4}$ to $n^{-1/5}$ to $n^{-1/6}$ and so on, we need to track the full asymptotic spectrum of the behavior of r and d, not using just the leading terms. It turns out that r and d are approximately given by the following joint functions of μ and σ, whose asymptotic series can be computed from (23.10):

$$r_\mu = \frac{\mu^2 - \sigma^2}{2(1+\mu)}\, n\,; \tag{23.12}$$

$$d_\mu = \frac{3\mu^2 - 3\sigma^2 - \sigma(\mu+\sigma)^2}{2(1+\mu)}\, n\,. \tag{23.13}$$

Notice that the numerators of both r_μ and d_μ are divisible by $(\mu+\sigma)n$, so r_μ and d_μ are multiples of the formula $(\mu+\sigma)n/(1+\mu)$ for giant component size (23.11). The quantity $\mu+\sigma$ can also, incidentally, be expressed as $\ln(1+\mu) - \ln(1-\sigma)$.

These values r_μ and d_μ also have a surprising relation to the confluent hypergeometric series $F(z) = F(1;\, 4;\, 4z)$ of (7.5). It is not difficult to check that

$$F\bigl((\mu+\sigma)/4\bigr) = \frac{6e^{\mu+\sigma}}{(\mu+\sigma)^3}\left(\frac{2r_\mu - d_\mu}{n}\right) = \frac{6(1+\mu)}{(1-\sigma)(\mu+\sigma)^3}\left(\frac{2r_\mu - d_\mu}{n}\right)\,; \tag{23.14}$$

$$\frac{\vartheta F\bigl((\mu+\sigma)/4\bigr)}{F\bigl((\mu+\sigma)/4\bigr)} = \frac{d_\mu}{2r_\mu - d_\mu}\,. \tag{23.15}$$

The quantities r_μ and d_μ are not the exact expected values of r and d. Indeed, the exact expected values are rational numbers, when m and n are integers, while σ is always irrational when μ is rational. But we will prove that the distributions of r and d are approximately normal with expectations r_μ and d_μ.

Before we can prove such a claim, we need to improve the estimate of e_{rd} in (7.16), because that estimate was derived only for fixed d.

Lemma 8. *Let $F(z)$ be the function defined in (7.5). If $r \to \infty$ and if d varies in such a way that $d/r \to 0$, the polynomial $P_d(r) = [z^d]\, F(z)^{2r-d}$ satisfies*

$$P_d(r) = \frac{F(s)^{2r-d}}{s^d}\,\frac{(d/e)^d}{d!}\left(1 + O\left(\frac{d}{r}\right)\right)\,, \tag{23.16}$$

where s is the solution to $\vartheta F(s)/F(s) = d/(2r-d)$.

Proof. We have

$$P_d(r) = \frac{1}{2\pi i}\oint \frac{F(z)^{2r-d}}{z^d}\,\frac{dz}{z} = \frac{1}{2\pi i}\oint e^{(2r-d)f(z)}\,\frac{dz}{z}\,,$$

where $f(z) = \ln F(z) - \bigl(d/(2r-d)\bigr)\ln z$, integrated on the circle $|z| = s$. By hypothesis, $\vartheta f(s) = 0$. Using the expansion formula

$$f(se^t) = \sum_{k=0}^{n} \frac{t^k}{k!}\,\vartheta^k f(s) + \int_0^t \frac{x^n}{n!}\,\vartheta^{n+1} f(se^{t-x})\,dx \tag{23.17}$$

with $n = 2$ and $t = i\theta$, we obtain

$$f(se^{i\theta}) = f(s) - \tfrac{1}{2}\theta^2\vartheta^2 f(s) + O(\theta^3 s)$$

because $|\vartheta^3 f(se^{i\theta})| = O(s)$. If $d \to \infty$, the contour integral is

$$\begin{aligned}
\frac{1}{2\pi}&\int_{-\pi}^{\pi} \exp\bigl((2r-d)\bigl(f(s) - \tfrac{1}{2}\theta^2\vartheta^2 f(s) + O(\theta^3 s)\bigr)\bigr)\,d\theta \\
&= \frac{1}{2\pi\sqrt{d}}\int_{-\pi\sqrt{d}}^{\pi\sqrt{d}} \exp\bigl((2r-d)f(s) - t^2/2 + O(t^2 d/r) + O(t^3 d^{-1/2})\bigr)\,dt \\
&= \frac{F(s)^{2r-d}}{s^d\sqrt{2\pi d}}\bigl(1 + O(d/r) + O(d^{-1/2})\bigr)\,, \qquad (23.18)
\end{aligned}$$

because $\vartheta^2 f(s) = s + O(s^2) = d/(2r-d) + O(d^2/r^2)$. The terms $O(t^2 d/r)$ and $O(t^3 d^{-1/2})$ can safely be moved out of the exponent because they are bounded when $|t| \le d^{1/6}$ and $|t| \le \sqrt{r/d}$. Larger values of $|t|$ are unimportant in the integral because of the factor $e^{-t^2/2}$, and because the relation

$$F(z) = 3\int_0^1 (1-u)^2 e^{4zu}\,du$$

implies that $|F(z)| \le F(\Re z)$; once the real part is sufficiently small, we can neglect the remaining part of the path.

Equation (23.18) does not match (23.16) perfectly, although it would be sufficient for the applications considered below. To derive the sharper estimate claimed in (23.16) when d is small, we can apply (23.17) to $f(z) - z$ instead of to $f(z)$, obtaining

$$\begin{aligned}
f(se^{i\theta}) - se^{i\theta} &= f(s) - s - i\theta s + O(\theta^2 s^2)\,; \\
f(se^{i\theta}) &= f(s) + s(e^{i\theta} - i\theta - 1) + O(\theta^2 s^2) \\
&= f(s) + \frac{d}{2r-d}(e^{i\theta} - i\theta - 1) + O\left(\frac{\theta^2 d^2}{r^2}\right).
\end{aligned}$$

The contour integral without the O term can be evaluated exactly,

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi}&\exp\bigl((2r-d)\bigl(f(s)+(e^{i\theta}-i\theta-1)d/(2r-d)\bigr)\bigr)\,d\theta\\&=\frac{F(s)^{2r-d}}{s^d}\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{(e^{i\theta}-1)d}d\theta/e^{i\theta d}\\&=\frac{F(s)^{2r-d}}{s^d}\,[z^d]\,e^{(z-1)d}=\frac{F(s)^{2r-d}}{s^d}\,\frac{(d/e)^d}{d!}\,.\end{aligned}$$

The O term contributes a relative error of $O(d/r)$, because we have

$$\begin{aligned}\int_{-\pi}^{\pi}\bigl|\exp\bigl((e^{i\theta}-1-i\theta)d\bigr)\bigr|\theta^2\,d\theta&=\int_{-\pi}^{\pi}e^{(\cos\theta-1)d}\theta^2\,d\theta\\&\le\int_{-\pi}^{\pi}e^{-c\theta^2 d}\theta^2\,d\theta=O(d^{-3/2})\,,\end{aligned}$$

where $c=2/\pi^2$. ☐

Theorem 13. *The joint distribution of the excess r and deficiency d of a random multigraph with $m=\frac{n}{2}(1+\mu)$ edges is approximately normal about the expected values r_μ and d_μ in (23.12) and (23.13), with zero covariance. More precisely, there exists $\epsilon>0$ such that if*

$$r=r_\mu+\rho\sqrt{\mu^3 n}\,,\qquad d=d_\mu+\delta\sqrt{\mu^4 n}\,,\tag{23.19}$$

the probability that a random multigraph has excess r and deficiency d is

$$\frac{3}{4\pi\sqrt{5}\,\mu^{7/2}n}\exp\left(-\frac{3}{20}\rho^2-\frac{3}{4}\delta^2+O\left((1+|\rho|+|\delta|)^2\,\mu^{1/2}+\frac{1+|\rho|^3}{(\mu^3 n)^{1/2}}+\frac{1+|\delta|^3}{(\mu^4 n)^{1/2}}\right)\right),\tag{23.20}$$

when $n^{-1/4}\le\mu\le\epsilon$ and $n\to\infty$, uniformly for $|\rho|\le\frac12\sqrt{n\mu^3}$ and $|\delta|\le\frac12\sqrt{n\mu^4}$.

Proof. Before proving formula (23.20), we can verify that its leading factor yields total probability 1 when integrated over all values of r and d near r_μ and d_μ: The integral over d gives a factor of $\sqrt{4\pi\mu^4 n/3}$, and the integral over r gives a factor of $\sqrt{20\pi\mu^3 n/3}$.

Let r and d be given by (23.19); the probability of excess r and deficiency d is then

$$\frac{2^m m!\,n!\,e_{rd}}{n^{2m}(n-m+r)!\,2^{n-m+r}}\,[z^n]\,\frac{\bigl(2-T(z)\bigr)^{n-m+r}T(z)^{n-m+3r-d}}{\bigl(1-T(z)\bigr)^{3r-d+1/2}}\,.\tag{23.21}$$

We find the coefficient of z^n by evaluating a contour integral as in (10.11) and (21.4); it is

$$\frac{1}{2\pi i}\oint e^{g(z)}(1-z)^{1/2}\,\frac{dz}{z}\,,\tag{23.22}$$

$$g(z) = nz + (3r-d)\bigl(\ln z - \ln(1-z)\bigr) + r\ln(2-z) - m\ln z + (n-m)\ln(2-z)\,. \tag{23.23}$$

The key to this theorem is the fact that, when $\rho = \delta = 0$, there is a saddle point at $z = 1-\sigma$:

$$\begin{aligned}\frac{g'(1-\sigma)}{n} = 1 + \frac{\mu+\sigma}{2(1+\mu)}\left(\frac{\sigma(\mu+\sigma)}{1-\sigma} + \frac{\sigma(\mu+\sigma)}{\sigma} - \frac{\mu-\sigma}{1+\sigma}\right)\\ - \frac{1+\mu}{2(1-\sigma)} - \frac{1-\mu}{2(1+\sigma)} = 0\,.\end{aligned} \tag{23.24}$$

Moreover, $g''(1-\sigma) = 5\mu n + O(\mu^2 n)$ in that case. If we integrate on the path $z = 1-\sigma+it/\sqrt{\mu n}$, as we did in Lemma 7 (section 21), the logarithm of the result will be

$$g(1-\sigma) + \ln\frac{2^m m!\,n!\,e_{rd}}{n^{2m}(n-m+r)!\,2^{n-m+r}}\sqrt{\frac{1}{10\pi n}} + O(\mu) + O(\mu^{-3/2}n^{-1/2})\,,$$

where $r = r_\mu$ and $d = d_\mu$. The relevant quantity s needed in Lemma 8 is

$$s = \frac{\mu+\sigma}{4} \tag{23.25}$$

because of (23.15). The evaluation of the stated logarithm is tedious, but it can be done in a reasonable amount of time with computer assistance, using some simplifications such as

$$3r-d = \frac{\sigma(\mu+\sigma)^2}{2(1+\mu)}\,n\,, \qquad n-m+r = \frac{1-\sigma^2}{2(1+\mu)}\,n\,.$$

The term $\ln\bigl((6r-2d)!/(3r-d)!\bigr)$ from (7.3) can be evaluated as $(3r-d)\ln(3r-d) - 3r+d+(6r-2d)\ln 2 + \frac12\ln 2 + O(\mu)$. It is not difficult to verify that the terms involving $n\ln n$ cancel. There are three terms involving $n\ln\mu$, namely $(3r-d)\ln\mu^2$, $-d\ln\mu$, and $-(2r-d)\ln\mu^3$, coming respectively from within expansions of $(3r-d)\ln(3r-d)$, $-d\ln s$, and $-(2r-d)\ln(2r-d)$; there are two other terms, $-(3r-d)\ln\sigma$ from within $g(1-\sigma)$ and $+(3r-d)\ln\sigma$ from within $(3r-d)\ln(3r-d)$, which also cancel. The most difficult part of the computation is the sum of about 16 terms that are rational functions in μ and σ, times n; these too sum to zero, using relations (23.14). The net result is that the complicated logarithm sums to $\ln 3 - 2\ln 2 - \ln\pi - \frac12\ln 5 - \frac72\ln\mu - \ln n + O(\mu) + O(\mu^{-4}n^{-1})$; this proves the theorem when $\rho = \delta = 0$.

For the case of general ρ and δ the calculations are similar but even worse. We now choose the integration path

$$z = 1-\sigma-\tfrac35\,\rho/\sqrt{\mu n} + it/\sqrt{\mu n}\,; \tag{23.26}$$

the first-order effects of ρ and δ then cancel out, and the second-order effects contribute $-\frac{3}{20}\rho^2 - \frac34\delta^2$ to the logarithm of the result. □

24. A waiting game. Now let's consider a little game. Start with an empty multigraph and add edges repeatedly at random until either (1) two different complex components are present; or (2) the multigraph is unclean. Case 1 represents the event "we have left the top line of Figure 1 before leaving the top line of Figure 2."

Let $G_0(w,z)$ be the bgf for all multigraphs such that the game has not yet stopped. Then

$$\sum_m \frac{2^m m!\, n!}{n^{2m}}\,[w^m z^n]\,G_0(w,z) \tag{24.1}$$

is the expected running time of the game. We have

$$G_0(w,z) = e^{U(wz)/w} \sum_r w^r K_r(wz)\,, \tag{24.2}$$

where $K_r(z)$ generates all clean cyclic multigraphs, weighted by the probability that they will arise as the cyclic part of a multigraph occurring during the game.

We learned in section 17 how to compute weighting factors that account for the history of transitions in Figure 1 among clean multigraphs; and we learned more specifically in section 20 how these coefficients arise as a multigraph gains random edges. In consequence, we can conclude that $K_r(z) = k_r T(z)^{2r}/\bigl(1-T(z)\bigr)^{3r+1/2}$, where $k_1 = e_1 = \frac{5}{24}$ and the later coefficients obey the rule

$$k_{r+1} = \frac{3}{2} r k_r\,. \tag{24.3}$$

Here's why: Given $k_r T^{2r}/(1-T)^{3r+1/2}$, the generating function for a clean vertex x is

$$\tfrac{1}{2}k_r T^{2r+1}/(1-T)^{3r+5/2} + 3r k_r T^{2r+1}/(1-T)^{3r+5/2}\,, \tag{24.4}$$

where the first term corresponds to cases where x is in the unicyclic part. Similarly, given the generating function $\frac{1}{2}k_r T^{2r+1}/(1-T)^{3r+5/2}$ after x is chosen to be unicyclic, the generating function for a clean unicyclic y is

$$\tfrac{5}{2}\,\tfrac{1}{2}k_r T^{2r+2}/(1-T)^{3r+9/2}\,; \tag{24.5}$$

here $\frac{5}{2} = 1 + 1 + \frac{1}{2}$, for choosing y on the half-edge to x, or on the self-loop attached to that half-edge, or in a different unicyclic component. We obtain a new bicyclic component if and only if both x and y were unicyclic. Therefore the generating function for cases where the game continues is

$$\bigl((3r+\tfrac{5}{2})(3r+\tfrac{1}{2}) - \tfrac{5}{4}\bigr)\,k_r T^{2r+2}/(1-T)^{3r+9/2}\,.$$

As in (20.9) and (20.10), we multiply by $(1-T)/(6r+6)$ to account for merging $\langle x,y\rangle$ with the existing edges. This proves (24.3).

Equation (24.3) implies, of course, that

$$k_r = \frac{5}{36}\left(\frac{3}{2}\right)^r (r-1)!\,. \tag{24.6}$$

Comparing this to the case $d = 0$ of (7.16), we have

$$k_r = \frac{5\pi}{18}\, e_r\bigl(1 + O(r^{-1})\bigr)\,. \tag{24.7}$$

Therefore the similar calculations of section 22, where we found that $p_1(n) - p_1'(n) = 1 - p_2(n) \sim \frac{1}{2}$, tell us that *the game will stop in Case (2) with probability* $\frac{5\pi}{18}$. This provides further evidence in support of the top-line conjecture that was made in section 18.

We can now try to compute the expected time for the game to be completed, but it appears to be quite complicated. The contribution to (24.1) from a given m and r can be obtained by changing e_r to k_r in (13.17) when m and r are not too large; this means we want to evaluate

$$\sum_{k\ge 0}\sqrt{\frac{2\pi}{3}}\,\frac{(\frac{1}{2}\,3^{2/3}\mu)^k}{k!}\left(\frac{1}{\Gamma(1/2-2k/3)} + \sum_{r\ge 1}\frac{5}{36}\left(\frac{1}{2}\right)^r\frac{(r-1)!}{\Gamma(r+1/2-2k/3)}\right) \tag{24.8}$$

in place of (14.1), representing $e^{\mu^3/6}$ times the probability that the game is still alive after m edges. The inner sum is known to be $\frac{5}{72}$ times

$$\begin{aligned}\sum_{r\ge 0}\left(\frac{1}{2}\right)^r\frac{r!}{\Gamma(r+3/2-2k/3)} &= \frac{1}{\Gamma(3/2-2k/3)}\,F\left(1,1;\frac{3}{2}-\frac{2k}{3};\frac{1}{2}\right)\\ &= \frac{1}{\Gamma(1/2-2k/3)}\left(\psi\left(\frac{3}{4}-\frac{k}{3}\right)-\psi\left(\frac{1}{4}-\frac{k}{3}\right)\right)\,,\end{aligned} \tag{24.9}$$

so it has the value $\sqrt{\pi}$ when $k = 0$. (Here, as usual, $\psi(z) = \Gamma'(z)/\Gamma(z)$.) Further study of (24.8) should prove to be interesting.

25. Waiting time in general. Bivariate generating functions provide a useful tool for studying the "first occurrences" of particular graphs or multigraphs, as shown in [14]. The special problems considered in that paper can be put into the following general framework.

Let $\mathcal{S}$ be any collection of multigraphs, with bgf $S(w,z)$. Suppose we wish to study the first time that an evolving multigraph on n vertices does not lie in $\mathcal{S}$. If $[z^n]\,S(0,z) = 0$, the empty graph on n vertices is not in $\mathcal{S}$, so the process never gets started. Otherwise, the probability that an evolving multigraph lies in $\mathcal{S}$ when it has $m-1$ edges but not when it has m is

$$\frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\,(w\vartheta_z^2 - 2\vartheta_w)S(w,z)\,. \tag{25.1}$$

The proof is simple, by definition of the operators ϑ_z and ϑ_w, because the probability in question is

$$\frac{2^{m-1}(m-1)!\,n!}{n^{2(m-1)}}\,[w^{m-1}z^n]\,S(w,z)-\frac{2^m m!\,n!}{n^{2m}}\,[w^m z^n]\,S(w,z)\,.$$

For convenience we shall write

$$\nabla\, S(w,z)=(w\vartheta_z^2-2\vartheta_w)\,S(w,z)\,; \tag{25.2}$$

we call ∇S the bgf for "stopping configurations," while S itself is the bgf for "going configurations."

The operator Φ_n, introduced in [14], is

$$\Phi_n F(w,z)=\sum_{m=1}^{\infty}\frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\,F(w,z)\,. \tag{25.3}$$

Equations (25.1)–(25.3) imply that $\Phi_n\nabla S(w,z)$ is the probability that a stopping configuration will be encountered when some edge is added to an initially empty multigraph. A similar operator

$$\widehat{\Phi}_n\,\widehat{F}(w,z)=\sum_{m=1}^{\infty}\frac{n!}{2m\binom{n(n-1)/2}{m}}\,[w^m z^n]\,\widehat{F}(w,z) \tag{25.4}$$

for graphs instead of multigraphs is considered in [14], but we will restrict consideration to multigraphs for simplicity. (As one might expect from section 6, we should use the operator

$$\widehat{\nabla}=w(\vartheta_z^2-\vartheta_z-2\vartheta_w)-2\vartheta_w \tag{25.5}$$

in place of ∇ when defining stopping configurations for the graph process.)

Several examples will help clarify these definitions and demonstrate their usefulness. Since the bgf $G(w,z)$ for all multigraphs satisfies $\vartheta_z^2 G=2w^{-1}\vartheta_w G$, equation (4.2), we have $\nabla G=0$; this, of course, is obvious, because there are no stopping configurations when all multigraphs are permitted.

Example 1. Let $S(w,z)$ be the bgf for all multigraphs having nothing but self-loops. Clearly $S(w,z)=e^{ze^{w/2}}$, because $ze^{w/2}$ is the bgf for a single vertex with nothing but self-loops. Formula (25.2) now tells us that

$$\nabla S(w,z)=wz^2e^w e^{ze^{w/2}}\,, \tag{25.6}$$

because $\vartheta_z^2 S = z^2 e^w S + z e^{w/2} S$ and $\vartheta_w S = \frac{w}{2} z e^{w/2} S$. Thus, by (25.1), the probability that an evolving multigraph first fails to lie in S when it acquires the mth edge is

$$\begin{aligned}
&\frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\,wz^2 e^w e^{ze^{w/2}}\\
&\qquad = \frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m]\,\frac{we^w e^{(n-2)w/2}}{(n-2)!}\\
&\qquad = \frac{2^{m-1}(m-1)!\,n(n-1)}{n^{2m}}\,[w^{m-1}]\,e^{nw/2} = \frac{n(n-1)}{n^{m+1}}\,.
\end{aligned}$$

And sure enough, $n^{1-m} - n^{-m}$ is obviously the probability that a sequence of edges $\langle x_1, y_1\rangle \ldots \langle x_m, y_m\rangle$ will have $x_1 = y_1, \ldots, x_{m-1} = y_{m-1}, x_m \neq y_m$.

Example 2. Let $S(w,z)$ be the bgf for all acyclic multigraphs, namely $e^{U(w,z)} = e^{U(wz)/w}$. The formulas

$$\vartheta_z U = w^{-1}T\,, \qquad \vartheta_z^2 U = w^{-1}T/(1-T)\,, \qquad \vartheta_w U = \tfrac{1}{2}w^{-1}T^2 \tag{25.7}$$

were derived in section 4, and we have

$$\vartheta_z^2 e^F = (\vartheta_z^2 F)e^F + (\vartheta_z F)^2 e^F \tag{25.8}$$

for any $F = F(w,z)$; hence

$$\nabla e^U = \frac{T}{1-T}\,e^U\,. \tag{25.9}$$

These are the stopping configurations that define the appearance of the first cycle in an evolving multigraph. The term $T^k e^U$ corresponds to a first cycle of length k; therefore if we replace T^k by kT^k and sum over all stopping times, we get an expression for the expected length of the first cycle,

$$\Phi_n\,\frac{T}{(1-T)^2}\,e^U\,. \tag{25.10}$$

This was one of the main problems studied in [14], where it was shown that the expected length is proportional to $n^{1/6}$ although the standard deviation is proportional to $n^{1/4}$.

Example 3. Let $S(w,z) = U(w,z)$ be the bgf for unrooted trees. This is a perverse example, thrown in primarily because (25.7) gives us the information we need to calculate

$$\begin{aligned}
\nabla U &= \frac{T}{1-T} - \frac{T^2}{w}\\
&= wz + (-w+2w^2)z^2 + (-2w^2 + \tfrac{9}{2}\,w^3)z^3 + \cdots\,.
\end{aligned} \tag{25.11}$$

What is the meaning of these negative coefficients?

The example does make sense, if we rephrase our interpretation of (25.1). The exact meaning of

$$\frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\,\nabla S(w,z)$$

is, "the probability that an evolving multigraph leaves $\mathcal{S}$ when the mth edge is added, minus the probability that it enters $\mathcal{S}$ when the mth edge is added." In our example, $U(0,z) = z$; when there are two or more vertices, the empty multigraph is not a tree, but it can become one later. The bgf for becoming a tree is $w^{-1}T^2$, corresponding to an ordered pair of rooted trees with $m-1$ edges. The bgf for adding a new edge $\langle x, y\rangle$ to a tree is $\sum_{k\ge 1} T^k$, where the term T^k corresponds to cases where x and y are at distance k. (Each appearance of $T = T(wz)$ includes an implicit edge touching the tree root, because w and z appear with equal powers in every term.)

Example 3 cautions us to interpret the operators ∇ and Φ_n a bit more carefully. In general, we have the identity

$$\Phi_n \nabla S(w,z) = n!\,[z^n]\,S(0,z) - \lim_{m\to\infty} \frac{2^m m!\,n!}{n^{2m}}\,[w^m z^n]\,S(w,z)\,, \tag{25.12}$$

for any bgf $S(w,z)$ such that the limit exists, because

$$\Phi_n \nabla S(w,z) = \sum_{m=1}^{\infty}\left(\frac{2^{m-1}(m-1)!\,n!}{n^{2(m-1)}}\,[w^{m-1}z^n]\,S(w,z) - \frac{2^m m!\,n!}{n^{2m}}\,[w^m z^n]\,S(w,z)\right).$$

A sufficient condition for the limit to exist is that the coefficients of $\nabla S(w,z)$ are nonnegative. A sufficient condition for the coefficients to be nonnegative is that $S(w,z)$ should represent a family of multigraphs $\mathcal{S}$ with the property that the deletion of any edge preserves membership in $\mathcal{S}$.

Example 4. Let $S(w,z) = G(w,z) - C(w,z)$ be the bgf for all disconnected multigraphs. The stopping configurations now represent the first time an evolving multigraph becomes connected. Since $G(w,z) = e^{C(w,z)}$, we have

$$\begin{aligned}
\vartheta_w C &= \vartheta_w \ln G = (\vartheta_w G)/G\,;\\
\vartheta_z C &= \vartheta_z \ln G = (\vartheta_z G)/G\,;\\
\vartheta_z^2 C &= (\vartheta_z^2 G)/G - (\vartheta_z G)^2/G^2\,;
\end{aligned}$$

hence

$$\nabla S = \nabla G - \nabla C = w(\vartheta_z C)^2\,. \tag{25.13}$$

Of course! This is an edge that joins an ordered pair of vertices marked in distinct components.

Example 5. Let $S(w,z)$ be any bgf of the form

$$S(w,z) = e^{U(w,z)+V(w,z)} H(w,z)\,. \tag{25.14}$$

Then we can use (25.7) and (4.9) to compute

$$\nabla S = e^{U+V} \bigl((2T\vartheta_z - 2\vartheta_w + we^{-V}\vartheta_z^2 e^V)\, H\bigr)\,. \tag{25.15}$$

For example, when $S(w,z) = G(w,z)$, the left side of (25.15) is zero, and $H(w,z)$ is the bgf we have called $E(w,z)$. Equating the right side of (25.15) to zero gives the differential equation (5.1) that we originally used to compute $E(w,z)$.

In the special case $H(w,z) = 1$, the stopping configurations correspond to the first time an evolving multigraph acquires a bicyclic component, i.e., the time when its excess changes from 0 to 1. This is another problem that was considered in [14], where it was shown that the expected number of unicyclic components present at the time is $\frac{1}{6}\ln n + O(1)$. If we express H in terms of univariate generating functions,

$$H(w,z) = \sum_{r\ge 0} w^r H_r(wz)\,, \tag{25.16}$$

then (25.15) can be written

$$\nabla S = e^{U+V} \sum_{r\ge 1} w^r \nabla H_r(wz)\,, \tag{25.17}$$

where the univariate function $H_r(z)$ is related to (5.3):

$$\nabla H_r(z) = e^{-V}\vartheta^2 e^V H_{r-1}(z) - 2\bigl(r + (1-T)\vartheta\bigr)H_r(z)\,. \tag{25.18}$$

Example 6. Specializing Example 5 further, let

$$S(w,z) = e^{U(w,z)+V(w,z)} \sum_{r=0}^{R} w^r E_r(wz)\,, \tag{25.19}$$

where R is any nonnegative integer. Then the stopping configurations ∇S represent the time when an evolving multigraph first acquires excess $R+1$. Expression (25.18) becomes almost trivial because ∇H_r is zero for all $r \ne R+1$; we have

$$\nabla S(w,z) = w^{R+1} e^{U(w,z)}\, \vartheta_z^2 e^{V(wz)} E_R(wz)\,. \tag{25.20}$$

This family $\mathcal{S}$ has the property that $\Phi_n \nabla S = 1$, by (25.12), because a multigraph surely acquires excess $R+1$ at some time $m \le n+R+1$. We can write the identity $\Phi_n \nabla S = 1$ more explicitly, using our known formula for E_R, and using r in place of R:

$$\Phi_n \left(w^{r+1} e^{U(wz)/w}\, \vartheta_z^2 \sum_{d=0}^{2r} e_{rd}\, \frac{T(wz)^{2r-d}}{\bigl(1-T(wz)\bigr)^{3r-d+1/2}} \right) = 1\,, \tag{25.21}$$

for all $n \ge 1$ and $r \ge 0$. Moreover, we can write (25.20) in the form

$$\nabla S(w,z) = 2w^{R+1} e^{U(w,z)+V(wz)} \bigl(R+1+(1-T)\vartheta_z\bigr) E_{R+1}(wz)\,,$$

using (5.3). Setting $r = R+1$ and applying (20.9) gives us another way to express (25.21),

$$\Phi_n \left(w^r e^{U(wz)/w} \sum_{d=0}^{2r} (6r-2d) e_{rd} \, \frac{T(wz)^{2r-d}}{\bigl(1-T(wz)\bigr)^{3r-d+3/2}} \right) = 1\,, \tag{25.22}$$

for all $n \ge 1$ and $r \ge 1$.

For example, the case $r = 1$ of (25.22) is

$$\Phi_n \left(we^U \left(\frac{5}{4}\,\frac{T^2}{(1-T)^{9/2}} + \frac{1}{2}\,\frac{T}{(1-T)^{7/2}} \right) \right) = 1\,. \tag{25.23}$$

The operator Φ_n is defined in (25.3) to be a sum over m, and the mth term of (25.23) is

$$\begin{aligned} &\frac{2^{m-1}(m-1)!\,n!}{n^{2m}}\,[w^m z^n]\, we^{U(w,z)} f\bigl(T(wz)\bigr) \\ &\qquad = \frac{1}{2m(n-m+1)}\,\frac{2^m m!\,n!}{n^{2m}(n-m)!}\,[z^n]\,U(z)^{n-m+1} f\bigl(T(z)\bigr) \\ &\qquad = \frac{1}{4m(n-m+1)}\,\frac{2^m m!\,n!}{n^{2m}(n-m)!}\,[z^n]\,U(z)^{n-m} g\bigl(T(z)\bigr)\,, \end{aligned} \tag{25.24}$$

where $f(T) = \frac{5}{4}T^2/(1-T)^{9/2} + \frac{1}{2}T/(1-T)^{7/2}$ and $g(T) = (2-T)Tf(T)$. We can write

$$g(T) = \frac{5/4}{(1-T)^{9/2}} - \frac{2}{(1-T)^{7/2}} - \frac{1/2}{(1-T)^{5/2}} + \frac{2}{(1-T)^{3/2}} - \frac{3/4}{(1-T)^{1/2}}\,,$$

so we can evaluate (25.24) by summing five applications of formula (10.1). The value is negligibly small unless m is $\frac{1}{2}n + O(n^{2/3})$, hence the factor $4m(n-m+1)$ can be assumed to equal $n^2 + O(n^{5/3})$. The five terms of g yield values of order $n^{4/3}$, n, $n^{2/3}$, $n^{1/3}$, and 1 respectively, according to (10.1); thus the leading term $\frac{5}{4}/(1-T)^{9/2}$ must be responsible for the major contribution to (25.23), and the mth term of the sum when $m = \frac{1}{2}n + \frac{1}{2}\mu n^{2/3}$ will be

$$\tfrac{5}{4} n^{-2/3} \sqrt{2\pi}\, A(\tfrac{9}{2}, \mu) + O(n^{-1})\,.$$

Summing over m yields 1. Therefore it must be true that

$$\int_{-\infty}^{\infty} A\left(\frac{9}{2}, \mu\right) d\mu = \frac{8/5}{\sqrt{2\pi}}\,.$$

This integral formula is not at all obvious from the definition of $A(y,\mu)$ in (10.2), and it would be interesting to find a direct proof.

The argument we have just given can be extended to arbitrary r, starting with (25.22), and it implies the following remarkable result:

$$\int_{-\infty}^{\infty} A(3r+\tfrac{3}{2}\,,\mu)\,d\mu = \frac{1}{3re_r\sqrt{2\pi}}\,, \qquad \text{integer } r \ge 1\,. \tag{25.25}$$

By (8.17) we can also write

$$\int_{-\infty}^{\infty} A(3r+\tfrac{3}{2}\,,\mu)\,d\mu = \frac{1}{3}\left(\frac{2}{3}\right)^r \frac{\Gamma(r)\sqrt{2\pi}}{\Gamma(r+\frac{5}{6})\,\Gamma(r+\frac{1}{6})}\,, \quad \text{integer } r \ge 1\,. \tag{25.26}$$

We have just proved that, if $M_{r,n} = \frac{1}{2}n + \frac{1}{2}U_{r,n}n^{2/3}$ is the number of edges when the excess first reaches r, then

$$\Pr(M_{r,n} = m) \;=\; 6re_r\sqrt{2\pi}\,A\bigl(3r+\tfrac{3}{2}\,,\mu\bigr)n^{-2/3} + O(n^{-1})\,; \tag{25.27}$$

hence $U_{r,n} \to U_r$ in distribution, where U_r has the density function

$$f_r(\mu) = 3re_r\sqrt{2\pi}\,A\bigl(3r+\tfrac{3}{2}\,,\mu\bigr)\,, \qquad -\infty < \mu < \infty\,. \tag{25.28}$$

Combining this formula with (13.17), we have

$$\begin{aligned}\sqrt{2\pi}\,e_r A(3r+\tfrac{1}{2}\,,\mu) = \lim_{n\to\infty}\Pr(\mathcal{E}_r) &= \lim_{n\to\infty}\Pr(M_{r,n} \le m < M_{r+1,n})\\ &= \int_{-\infty}^{\mu}\bigl(f_r(u) - f_{r+1}(u)\bigr)\,du\,,\end{aligned}$$

whence

$$\sqrt{2\pi}\,e_r\,A'(3r+\tfrac{1}{2}\,,\mu) = f_r(\mu) - f_{r+1}(\mu)\,. \tag{25.29}$$

In fact, (25.29) can be derived also by setting $y = 3r + \frac{1}{2}$ in the formula

$$A'(y,\mu) = (y-\tfrac{1}{2})A(y+1,\mu) - \tfrac{1}{2}\,y(y+2)A(y+4,\mu)\,, \tag{25.30}$$

which is a consequence of (10.22) and (10.23).

26. Continuous excess. Let $I(y)$ be the integral in (25.25) when the parameter r is not necessarily an integer:

$$I(y) = \int_{-\infty}^{\infty} A(y,\mu)\,d\mu\,. \tag{26.1}$$

It is natural to conjecture that formula (25.26) holds for y in general:

$$I(y) = \frac{2^{y/3}\sqrt{\pi}\;\Gamma(y/3-1/2)}{3^{y/3+1/2}\,\Gamma(y/3+1/3)\,\Gamma(y/3-1/3)}\,, \qquad y > \frac{3}{2}\,. \tag{26.2}$$

$\bigl($The condition $y > \frac{3}{2}$ is necessary and sufficient for convergence of the integral, because of (10.3) and (10.4).$\bigr)$ And indeed, this conjecture is true.

Theorem 14. *The integral (26.1) has the closed form (26.2).*

Proof. Let $I_0(y)$ be the right-hand side of (26.2); we wish to show that $I(y) = I_0(y)$. Clearly

$$I_0(y+3) = \frac{2y-3}{(y+1)(y-1)}\, I_0(y)\,, \qquad y > \frac{3}{2}\,. \tag{26.3}$$

Since $\int_{-\infty}^{\infty} A'(y,\mu)\,d\mu = 0$ for $y > \frac{1}{2}$, by (10.3) and (10.4), we can integrate (25.30) and replace y by $y-1$ to get the same recurrence for $I(y)$:

$$I(y+3) = \frac{2y-3}{(y+1)(y-1)}\, I(y)\,, \qquad y > \frac{3}{2}\,. \tag{26.4}$$

Therefore $I(y)/I_0(y)$ is a periodic function, and we need only prove asymptotic equivalence $I(y) \sim I_0(y)$ as $y \to \infty$ in order to verify strict equality $I(y) = I_0(y)$ for all $y > \frac{3}{2}$.

The duplication and triplication formulas for the Gamma function provide us with an alternate expression for $I_0(y)$:

$$I_0(3y) = \left(\frac{9}{2}\right)^{y-1} \frac{\Gamma(2y-1)}{\Gamma(3y-1)} \sim \frac{1}{\sqrt{6}} \left(\frac{2e}{3y}\right)^{y}\,. \tag{26.5}$$

To show that $I(y)$ has the same asymptotic behavior, we break the integral into two parts,

$$I(y) = \int_{-\infty}^{0} A(y,\mu)\,d\mu + \int_{0}^{\infty} A(y,\mu)\,d\mu = I_-(y) + I_+(y)\,. \tag{26.6}$$

By definition (10.2) we have

$$I_+(y) = \frac{1}{3^{(y+1)/3}} \int_0^{\infty} \sum_{k\ge 0} \frac{e^{-\mu^3/6}\bigl(\frac{1}{2}3^{2/3}\mu\bigr)^k\,d\mu}{k!\,\Gamma\bigl((y+1-2k)/3\bigr)}\,; \tag{26.7}$$

we will show that the asymptotic value of $I_+(y)$ can be obtained by interchanging summation and integration, then estimating the resulting sum.

Let a_k be the kth term after integration,

$$a_k = \int_0^{\infty} \frac{e^{-\mu^3/6}\bigl(\frac{1}{2}3^{2/3}\mu\bigr)^k\,d\mu}{k!\;\Gamma\bigl((y+1-2k)/3\bigr)} = \frac{2^{(1-2k)/3}3^{k-2/3}\,\Gamma\bigl((k+1)/3\bigr)}{k!\;\Gamma\bigl((y+1-2k)/3\bigr)}\,. \tag{26.8}$$

If $a_k = 0$ then $a_{k+3} = 0$; otherwise we have

$$\frac{a_{k+3}}{a_k} = \frac{(2k+5-y)(2k+2-y)}{4(k+2)(k+3)}\,, \tag{26.9}$$

which is greater than 1 when $k < \frac{1}{4}y - \frac{17}{8} - \frac{5}{4}(y+\frac{3}{2})^{-1}$, less than 1 when k exceeds that value, and nonnegative except for one or two values of k near $\frac{1}{2}y$. So the largest terms a_k occur when k is near $\frac{1}{4}y$. If $y > 5$ and $k > y/2$, we have

$$\left|\frac{a_{k+3}}{a_k}\right| \;\le\; \left(1 - \frac{y-5}{2k}\right)^2 \le \left(e^{3/k}\right)^{(5-y)/3} \le \left(\frac{k+3}{k}\right)^{(5-y)/3} ,$$

and it follows that $a_k = O(k^{(5-y)/3})$ as $k \to \infty$. Therefore $\sum |a_k|$ exists, and the interchange of summation and integration is justified, at least for large y.

Let $k = \frac{1}{4}y + x$, where $|x| \le y^{1/2+\epsilon}$. Then Stirling's formula tells us that

$$\ln a_k = \frac{y+3}{3}\,\ln 2 + \frac{y-2}{3}\,\ln 3 - \frac{2y+3}{6}\,\ln y + \frac{y}{3} - \ln\sqrt{\pi} - \frac{8x^2}{3y} + O(y^{3\epsilon-1/2})\,. \quad (26.10)$$

If $0 < \epsilon < \frac{1}{6}$, this implies that the sum of all terms for $|x| > y^{1/2+\epsilon}$ is superpolynomially small in relation to the sum of terms for $|x| \le y^{1/2+\epsilon}$; hence

$$\sum_{k=0}^{\infty} a_k \;\sim\; \frac{2^{(y+3)/3}3^{(y-2)/3}e^{y/3}}{y^{(2y+3)/6}\,\sqrt{\pi}} \;\int_{-\infty}^{\infty} e^{-8x^2/(3y)}\,dx$$

and we have

$$I_+(y) = \frac{1}{3^{(y+1)/3}}\sum_{k=0}^{\infty} a_k \;\sim\; \frac{1}{\sqrt{6}}\left(\frac{2e}{y}\right)^{y/3} \;\sim\; I_0(y)\,. \quad (26.11)$$

The proof of (26.2) will therefore be complete if we can show that $I_-(y)/I_+(y) \to 0$ as $y \to \infty$. For this we can use (10.9) to show that

$$A(y,-\alpha) \le \frac{1}{2\pi}\,\alpha^{1/2-y}\int_{-\infty}^{\infty}\left|1 + \frac{it}{\alpha^{3/2}}\right|^{1-y} e^{-t^2/2}\,dt \le \frac{\alpha^{1/2-y}}{\sqrt{2\pi}}\,;$$

therefore the first portion of $I_-(y)$ is quite small,

$$\int_{-\infty}^{-y^{1/3}} A(y,\mu)\,d\mu \le \frac{1}{\sqrt{2\pi}}\int_{y^{1/3}}^{\infty}\alpha^{1/2-y}\,d\alpha = O(y^{-1/2-y/3})\,.$$

On the other hand when $-y^{1/3} \le \mu \le 0$ we can integrate (10.7) from $y^{1/3}-i\infty$ to $y^{1/3}+i\infty$, obtaining

$$\begin{aligned} A(y,\mu) &\le \frac{1}{2\pi}\int_{-\infty}^{\infty} |y^{1/3}+it|^{1-y}\, e^{\Re K(\mu, y^{1/3}+it)}\,dt \\ &\le \frac{1}{2\pi}\,y^{(1-y)/3}\,e^{K(\mu,y^{1/3})}\int_{-\infty}^{\infty}\exp\left(-\left(y^{1/3}+\frac{\mu}{2}\right)t^2\right)\,dt \\ &= \frac{y^{(1-y)/3}\exp\bigl(y/3+\mu(3y^{2/3}-\mu^2)/6\bigr)}{\sqrt{2\pi\,(2y^{1/3}+\mu)}} \le \frac{y^{1/6}}{\sqrt{2\pi}}\left(\frac{e}{y}\right)^{y/3}\,; \end{aligned}$$

hence

$$\int_{-y^{1/3}}^{0} A(y,\mu)\,d\mu = O\left(y^{1/2}\left(\frac{e}{y}\right)^{y/3}\right)$$

and $I_-(y) \ll I_+(y)$ as desired. $\square$

Theorem 14 sheds further light on the results of [14], where the first cycle of a random multigraph was shown to have average length asymptotic to $\sqrt{\pi/2}\,I(2)\,n^{1/6}$. According to a lengthy numerical calculation sketched there, this coefficient was determined to be 2.0337, correct to four decimal places. Sure enough, equation (26.2) now confirms that the exact value is

$$\frac{\pi^{1/2}\,\Gamma(1/3)}{2^{1/6}\,3^{2/3}} = 2.03369\,20140\,63898\,89186\,17247\,01028\,49830\,16693-\,. \tag{26.12}$$

Section 7 of [14] also proves implicitly that, if the random variables L and S are respectively the length of the first cycle and the size of the component containing that cycle, we have

$$\mathrm{E}_n\,L^k \;\sim\; \sqrt{\frac{\pi}{2}}\,k!\,I(k+1)\,n^{k/3-1/6}\,; \tag{26.13}$$

$$\mathrm{E}_n\,S^k \;\sim\; 2^{k-1/2}\,\Gamma\bigl(k+\tfrac{1}{2}\bigr)\,I(2k+1)\,n^{2k/3-1/6} \tag{26.14}$$

In particular, the variance of L is asymptotically $\sqrt{2\pi n}$; the asymptotic mean and variance of S are $\sqrt{\pi n/2}$ and $Kn^{7/6}$, where K is the constant in (26.12). For graphs instead of multigraphs, these coefficients should all be multiplied by $e^{3/4}$.

Notice that $I(3)=1$. Hence the function $A(3,\mu)$, which is expressible in terms of Airy series or Bessel functions (see (10.32)), defines a probability density.

Let V_y be a random variable with density function $A(y,\mu)/I(y)$, when $y>\frac{3}{2}$. Then, by (10.22),

$$\mathrm{E}\,V_y = \int_{-\infty}^{\infty} \frac{\mu\,A(y,\mu)\,d\mu}{I(y)} = \frac{y\,I(y+2)-I(y-1)}{I(y)} = \frac{(y-3)I(y-1)}{(y-2)I(y)}\,, \tag{26.15}$$

if $y>\frac{5}{2}$. In particular, the variable U_r of (25.28), which is $V_{3r+3/2}$, has the mean value

$$\frac{(3r-3/2)I(3r+1/2)}{(3r-1/2)I(3r+3/2)} = \left(\frac{3}{4}\right)^{1/3}\frac{\Gamma(2r-2/3)}{\Gamma(2r-1)}\,. \tag{26.16}$$

This is the limit as $n\to\infty$ of $\mathrm{E}\,U_{r,n}$, which represents the mean waiting time for a graph or multigraph to reach excess r. The values are 0.8113, 1.2621, 1.5191, 1.7104, 1.8666, 2.0002, 2.1181, 2.2241, 2.3209, 2.4102 when $1\le r\le 10$.

Similarly, (10.23) implies that

$$\mathrm{E}\, V_y^2 = \frac{I(y-2)}{I(y)} = \frac{(y-2)}{6^{1/3}}\, \frac{\Gamma\big((2y-7)/3\big)}{\Gamma(2y/3-2)}\,, \quad y > \tfrac{7}{2}\,. \tag{26.17}$$

Hence $\mathrm{E}\, V_y = (y/2)^{1/3}\big(1-\frac{7}{6}y^{-1}+O(y^{-2})\big)$, $\mathrm{E}\, V_y^2 = (y/2)^{2/3}\big(1-\frac{2}{3}y^{-1}+O(y^{-2})\big)$, and we have

$$\operatorname{Var} V_y = \frac{5}{2^{2/3}\,3}\, y^{-1/3} + O(y^{-4/3})\,. \tag{26.18}$$

Let us now set $\mu = (y/2)^{1/3} + \sigma z$, where

$$\sigma^2 = \frac{5}{2^{2/3}\,3}\, y^{-1/3}\,. \tag{26.19}$$

An argument similar to the derivation of (26.11) proves that

$$\frac{A(y,\mu)}{I(y)} \sim \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-z^2/2}\,, \qquad z = O(1)\,, \quad y \to \infty\,. \tag{26.20}$$

Therefore $(\operatorname{Var} V_y)^{-1/2}(V_y - \mathrm{E}\, V_y)$ approaches the normal distribution $N(0,1)$ as $y \to \infty$. In particular, this establishes a kind of asymptotic normality of $U_{r,n}$ (and $M_{r,n}$), if we first let $n \to \infty$ and then $r \to \infty$.

27. Proof of the top-line conjecture. We are almost ready to settle the conjecture that was made in section 18, but first we should carry out the promised refinement of our estimates (23.5) and (23.9) for the sizes of the acyclic and unicyclic parts of a random multigraph.

The first step is to consider the quantity (23.3), when $m = \frac{1}{2}n(1+\mu)$ and $k = \kappa n$. If $k \ge m$ or $k \ge n$, expression (23.3) is zero; otherwise $0 \le \kappa < \min\big(\frac{1+\mu}{2}, 1\big)$, and Stirling's approximation yields

$$e^k \sqrt{\frac{m-k}{m}}\, \sqrt{\frac{n-k}{n}}\, \frac{2^k\, m^{\underline{k}}\, n^{\underline{k}}}{(n-k)^{2k}} \left(\frac{n-k}{n}\right)^{2m} = \exp\left(nf(\kappa,\mu) + O\left(\frac{1}{m-k}\right) + O\left(\frac{1}{n-k}\right)\right), \tag{27.1}$$

where

$$f(\kappa,\mu) = \frac{1+\mu}{2}\ln(1+\mu) - \frac{(1+\mu-2k)}{2}\ln(1+\mu-2\kappa) + (\mu-\kappa)\ln(1-\kappa) - \kappa\,. \tag{27.2}$$

Notice that

$$\frac{\partial f(\kappa,\mu)}{\partial\kappa} = \ln(1+\mu-2\kappa) - \ln(1-\kappa) + \frac{\kappa-\mu}{1-\kappa}\,,$$

$$\frac{\partial^2 f(\kappa,\mu)}{\partial\kappa^2} = \frac{(1-\mu)(\mu-\kappa)}{(1+\mu-2\kappa)(1-\kappa)^2}\,,$$

so both first and second derivatives vanish when $\kappa = \mu$. The first derivative is ≤ 0 when $\kappa = 0$; if $0 < \mu < 1$ it increases to zero when $\kappa = \mu$, then becomes negative; if $\mu \le 0$ or $\mu \ge 1$ it decreases steadily. Thus $f(\kappa, \mu)$ is a decreasing function of κ, as claimed in section 23.

We also have

$$\frac{\partial f(\kappa,\mu)}{\partial \mu} = \frac{\ln(1+\mu)}{2} - \frac{\ln(1+\mu-2\kappa)}{2} + \ln(1-\kappa)\,;$$

this derivative decreases steadily, passing through zero when $\mu = \kappa/(2-\kappa)$. Therefore we have

$$f(\kappa,\mu) \le f\left(\kappa, \frac{\kappa}{2-\kappa}\right) = -\left(\frac{\kappa^3}{24} + \frac{\kappa^4}{24} + \frac{11\kappa^5}{320} + \cdots + \frac{1-2j/2^j}{j(j-1)}\,\kappa^j + \cdots\right), \qquad (27.3)$$

for all $\mu > 2\kappa - 1$. In particular, we can conclude that terms like (23.1) and (23.8) are superpolynomially small for all $k \ge n^{2/3+\epsilon}$, since they are $O\bigl(\exp(-n^{3\epsilon}/24)\bigr)$ when $k = n^{2/3+\epsilon}$.

Our next goal is to estimate the sum of (23.8) for $k \ge 1$ when $\mu \ge n^{-1/3}$. This sum $V(m,n)$ is the expected number of vertices in unicyclic components after m steps of the multigraph process. The formulas above allow us to write

$$\begin{aligned} V(m,n) &= \sum_{k \le n^{2/3+\epsilon}} \frac{1}{2}\,\frac{k^k Q(k)}{k!\,e^k} \sqrt{\frac{m}{m-k}}\sqrt{\frac{n}{n-k}}\, e^{nf(k/n,\mu)+O(n^{-1})} + O(e^{-n^\epsilon}) \\ &= \sum_{k \le n^{2/3+\epsilon}} \frac{1}{2}\,\frac{k^k Q(k)}{k!\,e^k} \exp\left(k\bigl(\ln(1+\mu)-\mu\bigr) + \frac{\mu k^2}{2n} - \frac{k^3}{6n^2}\right. \\ &\qquad\qquad \left. +O\left(\frac{\mu^2 k^2}{n} + \frac{k^4}{n^3} + \frac{k}{n}\right)\right) + O(e^{-n^\epsilon})\,. \end{aligned} \qquad (27.4)$$

Let $\mu = \alpha n^{-1/3}$, so that α is the quantity we called μ in sections 10–20 above. We will assume that $\alpha \ge 1$, and also that $\alpha \le cn^{1/3}$ (hence $\mu \le c$), where c is a sufficiently small constant. The terms of $V(m,n)$ are negligible for $k \ge n^{2/3+\epsilon}$, regardless of the value of μ; and when $n^{-1/3} \le \mu \le c$ we can in fact ignore all terms for $k > \alpha^\epsilon \mu^{-2}$. The reason is that

$$\begin{aligned} k\bigl(\ln(1+\mu)-\mu\bigr) + \frac{\mu\,k^2}{2n} - \frac{k^3}{6n^2} &= \frac{-\mu^2 k}{2}\left(\frac{1}{4} + \frac{1}{3}\left(\frac{3}{2} - \frac{k}{\mu n}\right)^2\right) + O(k\mu^3) \\ &\le \frac{-\mu^2 k}{8}\bigl(1+O(\mu)\bigr) \le \frac{-\mu^2 k}{100} \end{aligned}$$

if we choose c small enough. The sum of $O\bigl(e^{-\mu^2 k/100}\bigr)$ for $\alpha^\epsilon/\mu^2 < k < \infty$ is then $O\bigl(\mu^{-2} e^{-\alpha^\epsilon/100}\bigr)$, which is dominated by the error bounds we will encounter below.

When $k \le \alpha^\epsilon \mu^{-2}$, we have $\mu k^2/2n \le \alpha^{2\epsilon-3}/2 \le 1/2$. Therefore we are justified in moving terms out of the exponent in (27.4):

$$\begin{aligned}V(m,n) = \sum_{k\ge1} \frac{1}{2}\frac{k^k Q(k)}{k!}\,\bigl((1+\mu)e^{-(1+\mu)}\bigr)^k \biggl(1 + \frac{\mu k^2}{2n} - \frac{k^3}{6n^2}\\ + O\left(\frac{\mu k^2}{2n} - \frac{k^3}{6n^2}\right)^2 + O\left(\frac{\mu^2 k^2}{n} + \frac{k^4}{n^3} + \frac{k}{n}\right)\biggr)\\ = \sum_{k\ge1} \frac{1}{2}\frac{k^k Q(k)}{k!}\,\bigl((1-\sigma)e^{-(1-\sigma)}\bigr)^k \left(1 + \frac{\mu k^2}{2n} + O(\alpha^{4\epsilon-6}) + O(\frac{\alpha^{2\epsilon-2}}{n^{1/3}})\right). \quad (27.5)\end{aligned}$$

Here σ is the "shadow" of μ as in (23.4) and (23.10), and the error bounds are computed under the assumption $k \le \alpha^\epsilon/\mu^2$. The trick of (23.5) and (23.9) now applies, using (23.7), and we have

$$\begin{aligned}V(m,n) &= \frac{1}{2}\left(1 + O(\alpha^{4\epsilon-6}) + O(\frac{\alpha^{2\epsilon-2}}{n^{1/3}}) + \frac{\mu}{2n}\vartheta^2\right) \frac{T\bigl((1-\sigma)e^{-(1-\sigma)}\bigr)}{\Bigl(1-T\bigl((1-\sigma)e^{-(1-\sigma)}\bigr)\Bigr)^2}\\ &= \frac{1}{2}\left((1 + O(\alpha^{4\epsilon-6}) + O(\frac{\alpha^{2\epsilon-2}}{n^{1/3}}))\frac{1-\sigma}{\sigma^2} + \frac{\mu}{2n}\left(\frac{8-17\sigma+11\sigma^2-2\sigma^3}{\sigma^6}\right)\right). \quad (27.6)\end{aligned}$$

If we had expanded the summand further, we would have obtained still more accuracy; therefore we are allowed to set $\epsilon = 0$ in (27.6). The term $O(\alpha^{-2}n^{-1/3})$ dominates $O(\alpha^{-6})$ when $\alpha \ge n^{1/12}$; it comes from both $O(\mu^2k^2/n)$ and $O(k/n)$ in (27.5).

We are assuming that μ is small, hence $\sigma = \mu\bigl(1+O(\mu)\bigr)$. Thus (27.6) can be simplified to

$$V(m,n) = \left(\frac{1}{2\alpha^2} + \frac{2}{\alpha^5} + O\left(\frac{1}{\alpha^8}\right)\right) n^{2/3}\bigl(1+O(\mu)\bigr),$$

and with an extension of the same approach we obtain an asymptotic expansion that begins

$$V(m,n) = \left(\frac{1}{2\alpha^2} + \frac{2}{\alpha^5} + \frac{20}{\alpha^8} + \frac{320}{\alpha^{11}} + \frac{7040}{\alpha^{14}} + O\left(\frac{1}{\alpha^{17}}\right)\right) n^{2/3}\bigl(1+O(\mu)\bigr). \quad (27.7)$$

This expansion is readily computed if we note that

$$\vartheta^k \frac{T(z)}{\bigl(1-T(z)\bigr)^2} = \frac{2^k\, k!}{\bigl(1-T(z)\bigr)^{2k+2}} + \cdots, \quad (27.8)$$

where the remaining terms $a_{k1}/\bigl(1-T(z)\bigr)^{2k+1} + a_{k2}/\bigl(1-T(z)\bigr)^{2k} + \cdots$ are negligible when we replace $T(z)$ by $1-\mu-O(\mu^2)$. The asymptotic series in (27.7) is obtained also from the integral

$$\tfrac14\int_0^\infty e^{-\alpha^2t/2+\alpha t^2/2-t^3/6}\,dt = \tfrac14\int_0^\infty e^{(\alpha-t)^3/6-\alpha^3/6}\,dt, \quad (27.9)$$

because we can expand $e^{\alpha t^2/2-t^3/6}$ into powers of t and use the formula

$$\int_0^\infty e^{-\alpha^2 t/2} t^k \, dt = \frac{2^{k+1}\, k!}{\alpha^{2k+2}}\,, \tag{27.10}$$

which matches (27.8). The coefficients of (27.7) follow a simple pattern; for example, $7040 = 22 \cdot 16 \cdot 10 \cdot 4 \,/\, 2$. Thus we are led to conjecture the asymptotic series

$$\int_0^\infty e^{(\alpha-t)^3/6-\alpha^3/6}\, dt \;\sim\; 2F\bigl(\tfrac{2}{3}, 1; ; 6/\alpha^3\bigr)/\alpha^2 \quad \text{as} \quad \alpha \to \infty\,; \tag{27.11}$$

the right-hand side here is a formal power series that diverges for all finite α. And indeed, this conjecture is true, as we will see momentarily.

A similar calculation allows us to estimate $U(m,n)$, the number of vertices in trees. The analog of (27.5) is

$$U(m,n) = \frac{n}{1+\mu} \sum_{k\ge 1} \frac{k^{k-1}}{k!} \bigl((1-\sigma)e^{-(1-\sigma)}\bigr)^k \left(1 + \frac{\mu k^2}{2n} + O(k\alpha^{3\epsilon-4} n^{-2/3}) + O(k\alpha^{\epsilon} n^{-1})\right); \tag{27.5$'$}$$

we leave a factor of k in the O terms because it will lead to a better final estimate. Then the analogs of (27.6)–(27.10) are

$$U(m,n) = \frac{n}{1+\mu}\left(1 - \sigma + O(\alpha^{3\epsilon-5} n^{-1/3}) + O(\alpha^{\epsilon-1} n^{-2/3}) + \frac{\mu}{2n}\left(\frac{1-\sigma}{\sigma^3}\right)\right); \tag{27.6$'$}$$

$$\begin{aligned} U(m,n) = n + \biggl(-2\alpha + \frac{1}{2\alpha^2} + \frac{11}{8\alpha^5} + \frac{175}{16\alpha^8} + \frac{19005}{128\alpha^{11}} + \frac{735735}{256\alpha^{14}} \\ + O\left(\frac{1}{\alpha^{17}}\right)\biggr) n^{2/3}\bigl(1 + O(\mu)\bigr)\,; \end{aligned} \tag{27.7$'$}$$

$$\vartheta^k T(z) = \frac{2^{k-1}\,\Gamma(k-1/2)}{\Gamma(1/2)} \frac{1}{\bigl(1-T(z)\bigr)^{2k-1}} + \cdots\,, \qquad k \ge 1\,; \tag{27.8$'$}$$

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-\alpha^2 t/2}(e^{\alpha t^2/2-t^3/6}-1)\, dt}{t^{3/2}} = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{(\alpha-t)^3/6-\alpha^3/6}-1}{t^{3/2}}\, dt + \alpha\,; \tag{27.9$'$}$$

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\alpha^2 t/2} t^{k-3/2}\, dt = \frac{2^{k-1}\,\Gamma(k-1/2)}{\sqrt{\pi}\;\alpha^{2k-1}}\,, \qquad k \ge 1\,. \tag{27.10$'$}$$

The asymptotic series (27.7) and (27.7$'$) for $\alpha \to \infty$ blend perfectly with the results obtained in [28] when α is any constant (positive, negative, or zero):

$$V(m,n) = \frac{1}{4}\left(\int_0^\infty e^{(\alpha-t)^3/6-\alpha^3/6}\, dt\right) n^{2/3} + O(n^{1/3})\,; \tag{27.12}$$

$$U(m,n) = n + \left(-\alpha + \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{(\alpha-t)^3/6-\alpha^3/6} - 1}{t^{3/2}}\, dt\right) n^{2/3} + O(n^{1/3})\,. \tag{27.12$'$}$$

These integrals are entire functions of α,

$$\int_0^\infty e^{(\alpha-t)^3/6-\alpha^3/6}\,dt = \frac{6^{1/3}\,\Gamma(1/3)}{3}\,e^{-\alpha^3/6} + \alpha\,F(1;\,\tfrac{4}{3}\,;-\alpha^3/6)\,; \tag{27.13}$$

$$\begin{aligned}\int_0^\infty \frac{e^{(\alpha-t)^3/6-\alpha^3/6}-1}{t^{3/2}}\,dt = -e^{-\alpha^3/6}\Bigl(&\bigl(6^{5/6}\,\Gamma(\tfrac{5}{6})/3\bigr)\,F(\tfrac12,\,\tfrac56;\,\tfrac13,\,\tfrac23;\,\alpha^3/6)\\ &-\bigl(6^{1/2}\,\Gamma(\tfrac12)/6\bigr)\,\alpha\,F(\tfrac56,\,\tfrac76;\,\tfrac23,\,\tfrac43;\,\alpha^3/6)\\ &+\bigl(6^{1/6}\,\Gamma(\tfrac16)/8\bigr)\,\alpha^2\,F(\tfrac76,\,\tfrac32;\,\tfrac43,\,\tfrac53;\,\alpha^3/6)\Bigr)\,.\end{aligned} \tag{27.13$'$}$$

Equation (27.13) is proved by observing that if $g(\alpha) = \int_0^\infty e^{(\alpha-t)^3/6}\,dt$, then $g'(\alpha) = \int_0^\infty \frac{(\alpha-t)^2}{2}\,e^{(\alpha-t)^3/6}\,dt = e^{\alpha^3/6}$. It implies (27.11) by well-known properties of confluent hypergeometric series. Equation (27.13$'$) is proved by setting $h(\alpha) = \int_0^\infty (e^{(\alpha-t)^3/6} - e^{\alpha^3/6})\,t^{-3/2}\,dt = -\int_0^\infty (\alpha-t)^2\,e^{(\alpha-t)^3/6}t^{-1/2}\,dt$ and proving that $h'''(\alpha) = \frac12\,\alpha^2 h''(\alpha) + \frac52\,\alpha h'(\alpha) + \frac{15}{8}\,h(\alpha)$, hence $[\alpha^{k+3}]\,h(\alpha) = [\alpha^k]\,h(\alpha)(k+\frac32)(k+\frac52)/\bigl(2(k+1)(k+2)(k+3)\bigr)$. Recall that we enumerated $n - U(m,n) - V(m,n)$, the expected number of vertices in complex components, using a complementary approach in (15.13), by summing over the excess r.

Lemma 9. *Let V_{mn} be the number of vertices in unicyclic components of a random multigraph with m edges and n vertices. If $m = \frac12 n(1+\mu)$ and $\mu \ge n^{-1/3}$, the expected value of V_{mn}^l is $O(\mu^{-2l})$, for every fixed integer $l \ge 1$.*

Proof. Equation (27.7) proves this for $l = 1$ and $n^{-1/3} \le \mu \le c$, where c is some positive constant. A similar argument applies for arbitrary l, because the generating function $\vartheta^l e^{V(z)}$ is $e^{V(z)}/\bigl(1-T(z)\bigr)^{2l}$ times a polynomial in $T(z)$; this means we are summing terms like (27.5), but with $Q(k)$ replaced by a semipolynomial in k of degree $l - \frac12$. (See the proof of Theorem 3 in section 8.) The analog of (27.6) will then be $O(\sigma^{-2l})$, which is $O(\mu^{-2l})$ if $\mu \le c$. Incidentally, for this range of μ we will have

$$\mathrm{E}\,V_{mn}^l \;=\; \frac{\Gamma(l+1/4)}{\Gamma(1/4)}\,\left(\frac{2}{\sigma^2}\right)^l\bigl(1+O(\mu)+O(\mu^{-3}n^{-1})\bigr)\,. \tag{27.14}$$

If $c \le \mu \le n^\epsilon$, with $\epsilon < \frac14$, let $0 < \delta < 1 - \ln(1+c)/c$. Then each term in the analog of (27.4) with $k \le n^{3/4}$ is $O\bigl(k^{l-1}\exp\bigl(k(\ln(1+\mu)-\mu)+O(\mu^2k^2/n)\bigr)\bigr) = O\bigl(k^{l-1}\exp(-k\delta\mu)\bigr) = O(k^{l-1}e^{-\delta\mu-\delta ck})$. Hence $\mathrm{E}\,V_{mn}^l = O(e^{-\delta\mu})$.

Finally, if $\mu \ge n^\epsilon$ the value of $\mathrm{E}\,V_{mn}^l$ is superpolynomially small, for it is a sum of n terms each of which is bounded by a polynomial in m and n times $(1-1/n)^{2m}$, which is $O(\mu^d e^{-\mu})$ for some finite degree d. ☐

Corollary. *The probability that a random multigraph never acquires a new complex component after it has gained* $m = \frac{1}{2}(n + \alpha n^{2/3}) > \frac{1}{2}n$ *edges is* $1 - O(\alpha^{-3})$.

Proof. We may assume that $\alpha \geq 1$. A new complex component must be bicyclic. A multigraph gains a new bicyclic component if and only if the endpoints of a new edge both fall in unicyclic components. The probability that this occurs at time $m = \frac{1}{2}(n + \mu n)$ is $\mathrm{E}\, V_{mn}^2/n^2 = O(\mu^{-4}n^{-2})$, by the lemma. Summing for $m \geq \frac{1}{2}(n + \alpha n^{2/3})$ gives $O(\alpha^{-3})$ as an upper bound on the probability that at least one new bicyclic component appears after time $\frac{1}{2}(n + \alpha n^{2/3})$. $\square$

Theorem 15. *The probability that an evolving graph or multigraph on* n *vertices never has more than one complex component throughout its evolution approaches* $\frac{5\pi}{18}$ *as* $n \to \infty$.

Proof. Let $\epsilon > 0$ be fixed. By the corollary just proved, there exists a number α, independent of n, such that the probability of a random multigraph obtaining a new complex component after time $m = \frac{1}{2}(n + \alpha n^{2/3})$ is less than ϵ.

By section 14 and the corollary of section 13, there is a number R, independent of n, such that the probability of having excess $> R$ at this time m is less than ϵ. So the probability that a random multigraph leaves the top line after excess R is $< 2\epsilon$. (Either it reaches excess R before time m, or it leaves the top line after time m.)

But the probability that a random multigraph leaves the top line before excess R is $1 - \frac{5\pi}{18} + O(R^{-1}) + O(n^{-1/3})$, by (18.2). We may choose R sufficiently large that this $O(R^{-1})$ is less than ϵ; then we may choose n sufficiently large that the $O(n^{-1/3})$ is less than ϵ. The probability that a random multigraph leaves the top line for such n is therefore between $1 - \frac{5\pi}{18} - 2\epsilon$ and $1 - \frac{5\pi}{18} + 4\epsilon$.

For graphs, we note that an evolving graph may be constructed from an evolving multigraph by ignoring all new edges that would be loops or parallel to an existing edge. Since this reduction preserves or decreases both the excess and the number of complex components, it follows that if the graph leaves the top line after excess R, then the multigraph does too. Hence this event likewise has probability $< 2\epsilon$, and the proof is completed as for multigraphs. $\square$

Theorem 16. *Given any set* S *of infinite paths in Figure 1, the probability that the evolution of a random multigraph follows a path in* S *converges as* $n \to \infty$ *to the corresponding probability for the Markov chain with the transition probabilities given in Theorem 9. Similarly, if the evolution of a random graph, which stops at excess* $\binom{n}{2} - n$ *when the complete graph is reached, is continued along the top line to an infinite path in Figure 1, then the probability that this path lies in* S *converges to the same limit.*

Proof. Given $\epsilon > 0$, let R be as in the preceding proof so that a random graph or multigraph leaves the top line after excess R with probability $< 2\epsilon$. We can also choose R large enough that $c_R > e_R(1-\epsilon)$, by (8.7). Since c_R/e_R is the sum of all Markov transition probabilities for paths that intersect the top line at excess R, if we cut Figure 1 at excess R, the Markov probabilities for paths in S that do not have this property must sum to less than ϵ. When R is large enough, the sum of Markov probabilities for all paths that diverge from the top line after excess R is likewise less than ϵ, because it is $O\bigl(\sum_R^\infty r^{-2}\bigr) = O(R^{-1})$.

Let $P_n(S)$ be the probability that the evolution of a random graph or multigraph on n vertices follows a path in S, and let $P_\infty(S)$ denote the corresponding Markov probability. If S_R is the subset of S having all paths on the top line when the excess is $\geq R$, then $0 \leq P_n(S) - P_n(S_R) < 2\epsilon$ for all $n \leq \infty$. Similarly, if S'_R is the set of all paths that follow a path in S_R up to excess R, but afterwards are arbitrary, then $0 \leq P_n(S'_R) - P_n(S_R) < 2\epsilon$, for $n \leq \infty$. Finally, by Theorem 10, $\bigl|P_n(S'_R) - P_\infty(S'_R)\bigr| < \epsilon$ if n is large enough, and we have $\bigl|P_n(S) - P_\infty(S)\bigr| < 5\epsilon$. $\square$

Theorem 16 says that the evolutionary path, regarded as a random element of the set of all paths in Figure 1, converges in distribution to the Markov process. There are uncountably many paths, but the theorem needs no measurability restriction since the distributions for finite n and for the limit are concentrated on the countable set of paths that eventually follow the top line. Note that we cannot strengthen the statement for random graphs to deduce the limiting probability that the evolution follows a path in S until it stops at excess $\binom{n}{2} - n$; for example, if S is the set of all paths that do *not* eventually follow the top line, the Markov probability $P_\infty(S)$ is zero, while $P_n(S) = 1$ for all finite n.

Corollary. *The probability that an evolving graph or multigraph never has more than l complex components converges to a limit P_l.* $\square$

Closed form expressions for P_l might not exist when $l \geq 2$, but the values can be estimated from below using the following related probabilities:

Corollary. *The probability that an evolving graph or multigraph acquires exactly $l \geq 1$ new complex components during the evolution converges to*

$$p'_l = \Pr\left(\sum_{r=0}^{\infty} I_r = l\right) = \Pr\left(\sum_{r=1}^{\infty} I_r = l-1\right), \qquad (27.15)$$

where $I_0,\ I_1,\ I_2,\ I_3,\ \ldots$ are independent Bernoulli distributed random variables with $\Pr(I_r = 1) = 1 - \Pr(I_r = 0) = 5/(6r+1)(6r+5)$.

In other words, the number of new complex components converges in distribution to $\sum_{r=0}^{\infty} I_r$.

Proof. Let $I_r = 1$ if the Markov process acquires a new bicyclic component when the excess goes from r to $r+1$, and $I_r = 0$ otherwise; in particular $I_0 = 1$ always. By

Theorem 9, $\Pr(I_r = 1) = 5/(6r+1)(6r+5)$ independently of the previous history, and thus the variables are independent. $\square$

The probabilities p'_l have a surprisingly simple generating function: We have

$$\begin{aligned}
p'_l &= [z^l] \prod_{r=0}^{\infty} \left(1 + (z-1)\frac{5}{(6r+1)(6r+5)}\right) \\
&= [z^l] \prod_{r=0}^{\infty} \frac{(r + \frac{1}{2} + \frac{1}{6}\sqrt{9-5z}\,)(r + \frac{1}{2} - \frac{1}{6}\sqrt{9-5z}\,)}{(r+\frac{1}{6})(r+\frac{5}{6})} \\
&= [z^l] \frac{\Gamma(\frac{1}{6})\,\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{2} + \frac{1}{6}\sqrt{9-5z}\,)\,\Gamma(\frac{1}{2} - \frac{1}{6}\sqrt{9-5z}\,)} \\
&= [z^l] \cos\left(\frac{\pi}{6}\sqrt{9-5z}\right) \Big/ \cos\frac{\pi}{3}\,.
\end{aligned} \tag{27.16}$$

Computing the coefficients of the Taylor series for $\cos\bigl(\frac{\pi}{6}\sqrt{9-5z}\bigr)$, we find that the numbers p'_l are rational polynomials in π:

$$\begin{aligned}
p'_1 &= \frac{5\pi}{18} \approx 0.87266\,; \\
p'_2 &= \frac{50\pi}{6^4} \approx 0.12120\,; \\
p'_3 &= \frac{500\pi}{6^6}\left(1 - \frac{\pi^2}{12}\right) \approx 0.00598\,; \\
p'_4 &= \frac{6250\pi}{6^8}\left(1 - \frac{\pi^2}{10}\right) \approx 0.00015\,.
\end{aligned}$$

Let $P'_l = \sum_{j=1}^{l} p'_j$; numerically we have $P_2 > P'_2 \approx 0.99387$, $P_3 > P'_3 \approx 0.99985$, $P_4 > P'_4 \approx 0.999998$.

The number of new complex components is also studied in [19], where further results are given. The methods of [19] do not, however, seem to yield the sharp results obtainable with generating functions.

28. Empirical data. Computer simulations of random multigraphs tend to confirm the theoretical results derived above, although there are a few surprises apparently due to the slow convergence of some asymptotic formulas. In this section we will discuss some of the statistics computed during 1000 trials of the multigraph process on 20,000 vertices, so that readers can obtain a feel for the way in which random multigraphs actually evolve in practice. The data was divided into two groups of 500 runs each, and both groups exhibited essentially the same behavior; therefore the full set of 1000 runs is being treated as a unit here.

When a statistic is given in the form '$x \pm y$' below, x is the sample mean and y is the sample standard deviation divided by $\sqrt{1000}$. The sample standard deviation has been computed by taking the square root of an unbiased estimate of the variance. The "time" of an event is the number of edges present when that event occurred.

The first cycle was formed at time 6769 ± 96; this agrees reasonably well with the asymptotic formula $n/3$ found in [14, Corollary 3]. The size of the first unicyclic component was 188 ± 14. According to (26.14), the mean should be approximately $\sqrt{\pi n/2} \approx 177$.

The length of the first cycle was 3.9 ± 0.1; in fact, the histogram was

length =	1	2	3	4	5	6	7	≥ 8
actual =	321	132	89	88	78	86	60	146
theoretical =	333	133	76	51	37	28	23	318

The distribution has infinite mean, approximately $2.03n^{1/6} + O(n^{3/22})$, and its standard deviation is of order $n^{1/4}$ by (26.13), so the length of the first cycle should not be expected to be a robust statistic. However, the marked deviation in the histogram for cycle lengths ≥ 4 was unexpected. Apparently n must become quite large before the asymptotic probability of first cycle length k will assert itself.

Several people have suggested in conversation that the "last cycle" ought to have the same statistical characteristics as the first. The last cycle is the last unicyclic component that is present during a multigraph's evolution: After it is absorbed into a component of higher complexity, no further unicycles exist, and no further unicycles are formed. (If two cycles disappear simultaneously when the edge $\langle x, y \rangle$ is added, we say that the cycle containing y was the last to go.) The manner in which the giant component swallows other structures is rather like the initial stages of evolution but in reverse: First the unicycles tend to go, then the larger trees, and finally only isolated vertices are left (see Bollobás [6, sections VI.3 and VII.1]). A strong formulation of this symmetry principle was proved by Łuczak [25]; the phenomenon can be explained by the symmetry between $T(z)$ and $2 - T(z)$ in $U(z)$. However, the length of the last cycle has a distinctly different distribution from the length of the first cycle (see [20]). In these computer runs it had the following histogram:

length =	1	2	3	4	5	6	7	≥ 8
observed =	423	144	107	79	63	62	40	82

with mean 3.1 ± 0.1.

The total number of unicyclic components formed during the entire evolution was

number =	1	2	3	4	5	6	7	≥ 8
observed =	53	148	221	219	178	98	44	39

with mean 4.0 ± 0.1.

The excess of the multigraph changed from 0 to 1 at time 10331 ± 13. The number of unicyclic components present was about 2.7 just before this event, and about 1.5 just after. As soon as the excess became positive it began a steady rise:

excess	time	unicyclic size just before	unicyclic size just after	complex size just before	complex size just after
1	10331 ± 13	1606 ± 22	163 ± 9	0	1442 ± 21
2	10501 ± 10	265 ± 14	132 ± 7	1779 ± 22	1912 ± 20
3	10603 ± 8	168 ± 9	111 ± 7	2166 ± 19	2222 ± 19
4	10675 ± 8	132 ± 8	90 ± 5	2433 ± 18	2475 ± 17
5	10738 ± 8	105 ± 6	85 ± 5	2659 ± 17	2680 ± 17
6	10789 ± 7	95 ± 6	76 ± 5	2825 ± 17	2844 ± 16
7	10835 ± 7	83 ± 5	69 ± 4	2980 ± 16	2994 ± 16
8	10880 ± 7	77 ± 5	66 ± 4	3126 ± 16	3137 ± 16
9	10920 ± 7	72 ± 5	62 ± 4	3253 ± 15	3263 ± 15
10	10955 ± 7	66 ± 4	58 ± 4	3371 ± 15	3379 ± 15

The value of $n^{2/3}$ is approximately 737 when $n = 20000$, so each additional edge increases the parameter μ of Lemma 3 by approximately 0.0027. The value of μ when $m = 10955$ is approximately 2.59; then $\frac{2}{3}\mu^3 + 1 + \frac{5}{24}\mu^{-3} + \frac{15}{16}\mu^{-6} \approx 12.6$, so the excess is not quite keeping up with the expected value in Theorem 6. Similarly, formula (26.16) predicts that the excess will reach 1 when $m \approx 10299$, and 10 when $m \approx 10888$; random multigraphs for finite n seem to become complex a bit "late." It is interesting to note that the observed standard deviations kept decreasing as the excess increased, while the discrepancy from (26.16) kept increasing.

The random multigraphs followed paths in Figure 1 with the frequencies shown in Figure 3. When the excess changed from 9 to 10, the transition was from a single C_9 to C_{10} in 977 cases, from C_9 to (C_1, C_9) in 2 cases, from (C_1, C_8) to C_{10} in 8 cases, and from (C_1, C_8) to (C_1, C_9) in the remaining 13 cases. Altogether 897 of the 1000 random multigraphs remained on the top line of Figure 1 throughout their evolution.

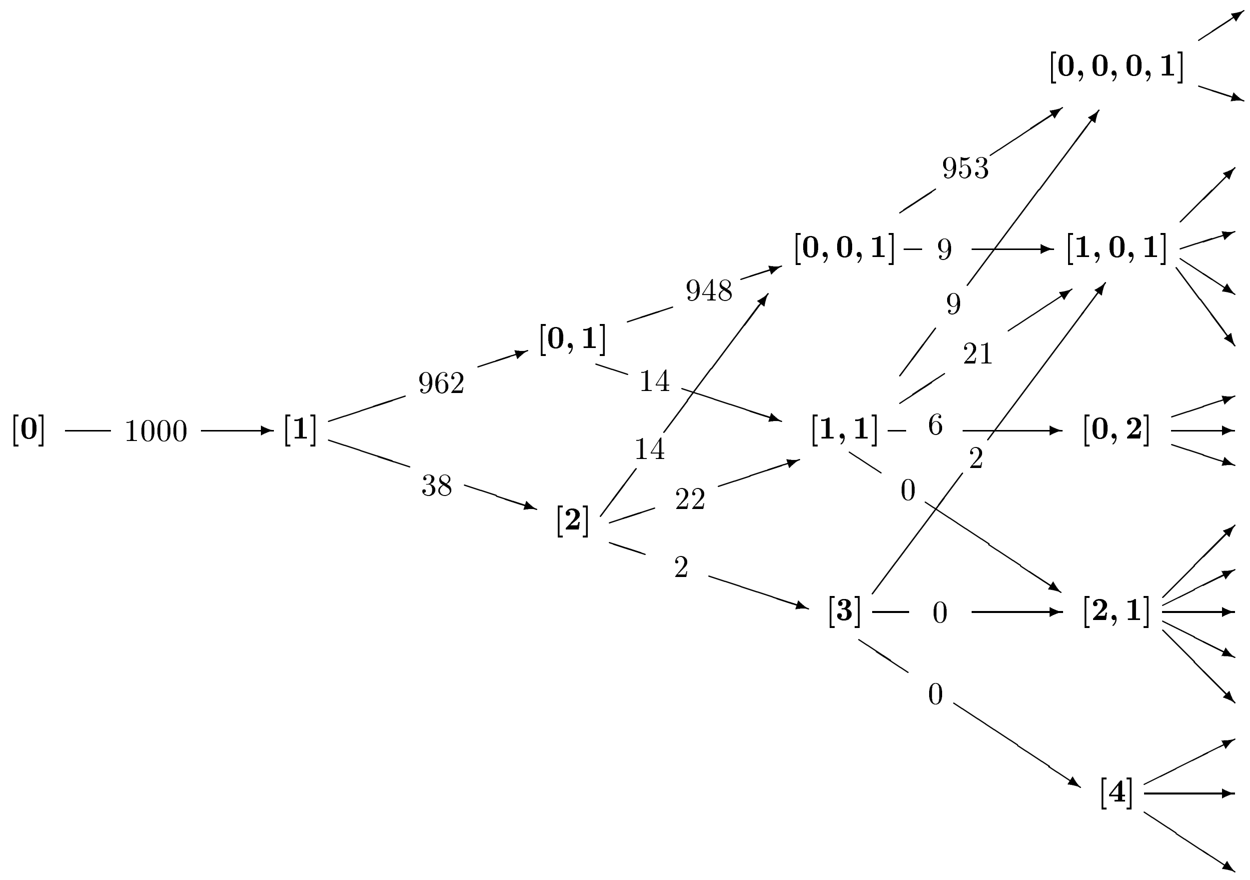


Figure 3. The number of times the paths in Figure 1 were actually traced, when 1000 random multigraphs on 20000 vertices were generated in experimental tests.

There comes a time when the giant component first succeeds in annihilating everything except isolated vertices, after which it remains the only component with edges. In these runs that time was 58352 ± 224. The number of isolated vertices still remaining was then 71 ± 1.

The multigraph finally became connected at time 105294 ± 404. The expected time for an evolving multigraph to have no isolated vertices is $\frac{1}{2}nH_n = \frac{1}{2}n \ln n + \frac{1}{2}\gamma n + \frac{1}{4} + O(n^{-1})$, which is approximately 104807 when $n = 20000$.

29. Open problems. The topics discussed in this paper raise a host of interesting questions, and the answers to those questions will no doubt bring additional striking patterns to light.

But the reader may have noticed that this paper is already rather long. Therefore it seems wise to stop at this point, with the hope that researchers all over the world will enjoy exploring the tantalizing questions that remain.

For example, it would be interesting to find a basis for as many linear combinations of terms $w^r T^a/(1-T)^b$ as possible such that

$$\Phi_n \, w^r \, e^U \, T^a/(1-T)^b$$

has a known value, as in (25.22). We can find many linear combinations of such functions for which Φ_n gives 0, because $\Phi_n \nabla S$ is usually 0 or 1. Notice that

$$\frac{T^a}{(1-T)^{b+1}} = \frac{T^a}{(1-T)^b} + \frac{T^{a+1}}{(1-T)^{b+1}}\,; \tag{29.1}$$

hence terms of excess $r+1$ can be expressed as combinations of terms of excess r. Conversely, we can go from excess r to excess $r+1$, because

$$\frac{T^a}{(1-T)^b} = \frac{T^a}{(1-T)^{b+1}} - \frac{T^{a+1}}{(1-T)^{b+2}} + \frac{T^{a+2}}{(1-T)^{b+3}} - \cdots \tag{29.2}$$

is an infinite series that always "converges" under application of Φ_n; all terms after a certain point are multiples of T^{n+1}, so they do not change the coefficient of z^n.

The stopping configuration machinery suggests many further problems of interest. For example, we should be able to deduce more about the nature of a random multigraph when its deficiency first exceeds a given number d.

The discussion in section 23 characterizes the stochastic behavior of r and d when $\mu = o(1)$; what happens thereafter? Relations (23.12) and (23.13) may well continue to describe the approximate mean values of r and d as $\mu \to \infty$. The shadow point σ defined in (23.2) will approach 0, but it remains an analytic function of μ, and $1-\sigma$ remains a saddle point of the contour integral for $[z^n]\, U^{n-m+r} T^{2r-d}/(1-T)^{3r-d+1/2}$.

The analytic function $T(z)$ has an interesting Riemann surface: There is a quadratic singularity at $z = e^{-1}$, and if we travel around that point we get to a second sheet in which there is a logarithmic singularity at $z = 0$. Winding around that logarithmic singularity takes us to infinitely many other sheets having no finite singularities besides 0. It may be possible to work out a theory under which contour integrals of importance in the study of random graphs could be evaluated by paths that pass through the point $1+\mu$, which lies on the "wrong side" of the quadratic singularity of $T(z)$; $1+\mu$ turns out to be a saddle point for several important generating functions.

Identity (8.15)–(8.16) suggests that the generating functions for random multigraphs might have interesting continued fraction forms. Such expressions could well be of special importance, because they often converge when power series do not.

The fact that the recurrence for the coefficients e_{rd} can be "solved" to yield (7.3)–(7.5) should prove to be a good challenge for computer systems that are now being constructed to solve recurrence relations automatically. The similar recurrence for the coefficients e'_{rd}, discussed in (7.24) and (7.25), will probably be an even greater challenge; at least, no simple derivation of (7.21) from (7.26) is known.

The solution to the recurrence for e_{rd} in section 7 relies on the introduction of a "half excess" stage, in which the polynomials must be evaluated at integers plus $\frac{1}{2}$ although the recurrence in which they are used involves integers only. In section 20 we found, similarly, that it was fruitful to break the process of adding an edge into stages in which "half-edges" were added. Perhaps the theory of fractional differentiation will be of value in future investigations. However, the operators $D^{1/2}$ and $\vartheta^{1/2}$ do not seem to transform the basic functions $T^a/(1-T)^b$ very nicely.

Is there an equation (27.11′) analogous to (27.11)? There must be a reason why the coefficients of (27.7′) tend to have small prime factors.

We have seen numerous examples in which the multigraph process leads to formulas that are mathematically cleaner than the analogous formulas for the graph process. This suggests that an analogous theory be introduced in place of the alternative "$\mathbf{G}_{n,p}$" model of random graphs: Instead of saying that each edge is present with probability p, the multiplicity of each edge should be allowed to have a Poisson distribution with mean p. Readers are encouraged to experiment with such an approach.

Convergence to limiting distributions often appears to be monotonic. For example, the probability that an evolving multigraph on n vertices stays on the top line appears to be strictly decreasing as n increases. How could this be proved?

Our proof of the top-line probability in Theorem 15 was independent of the difficult analyses in Lemma 7 and Theorem 13 about the behavior of random multigraphs with more than $\frac{1}{2}(n + n^{2/3+\epsilon})$ edges; moreover, it did not use the stopping-configuration machinery of sections 24–26, although that theory was in fact motivated by attempts to prove Theorem 15 in a sharper form via generating functions. The top-line phenomenon may perhaps be understood more deeply if we use a generating-function-based approach, and the following ideas may therefore prove to be useful. Let $S(w,z)$ be the bgf for all multigraphs that never leave the top line of Figure 1, where each multigraph is weighted by the probability of having a purely top-line history as discussed in section 17. The discussion of sections 19 and 20 shows that

$$S(w,z) = e^{U(w,z)+V(w,z)}H(w,z)\,, \tag{29.3}$$

where $H(w,z)$ satisfies a differential equation almost like the equation (5.1) that defines $E(w,z)$:

$$\tfrac{1}{w}\,(\vartheta_w - T\vartheta_z)H = \tfrac{1}{2}\,e^{-V}\vartheta_z^2 e^V H - \tfrac{1}{2}\,e^{-V}(\vartheta_z^2 e^V)(H-1)\,. \qquad (29.4)$$

The subtracted term $\frac{1}{2}e^{-V}(\vartheta_z^2 e^V)(H-1)$ accounts for the forbidden case that a new edge marked by ϑ_z^2 lies entirely in the unicyclic part generated by e^V; a second complex component arises if and only if this happens. The correction applies to $H-1$, not H, because the very first complex component does not violate the top-line condition.

Expressing $H(w,z)$ in the form (25.16), we have $H_1 = E_1$, but H_2 is smaller than E_2:

$$H_2 = \frac{5}{16}\,\frac{T^4}{(1-T)^6} + \frac{25}{48}\,\frac{T^3}{(1-T)^5} + \frac{11}{48}\,\frac{T^2}{(1-T)^4} + \frac{1}{48}\,\frac{T}{(1-T)^3}\,.$$

In general we can write

$$H_r = \sum h_{rd}\,\frac{T^{2r-d}}{(1-T)^{3r-d}} \qquad (29.5)$$

for appropriate coefficients h_{rd}. The special case $\mu=\nu=0$ of (20.7) tells us that

$$\vartheta^2 e^V = \vartheta^2(1-T)^{-1/2} = \tfrac{1}{2}\,T(1-T)^{-7/2} + \tfrac{5}{4}\,T^2(1-T)^{-9/2}\,; \qquad (29.6)$$

therefore we can compute the coefficients h_{rd} by making a slight change to the rule for computing e_{rd} that is expressed in (20.11): Subtract 5 from the numerator of the first coefficient term in (20.11), and subtract 1 from the numerator of the second coefficient. The first coefficient now simplifies to

$$\frac{(6r-2d+5)(6r-2d+1)-5}{8(3r-d+3)} = \frac{3r-d}{2}\,.$$

In particular, when $d=0$ we have $h_{(r+1)0} = \frac{3}{2}\,rh_{r0}$; hence h_{r0} is the number we called k_r in (24.3).

Equation (25.17) now gives us a useful expression for the stopping configurations,

$$\begin{aligned}\nabla S &= e^{U(w,z)}\sum_{r\ge 2} w^r(\vartheta_z^2 e^V)H_{r-1}(wz)\\ &= e^{U(w,z)}\sum_{r\ge 2} w^r\left(\frac{1}{2}\,\frac{T(wz)}{\bigl(1-T(wz)\bigr)^{7/2}} + \frac{5}{4}\,\frac{T(wz)^2}{\bigl(1-T(wz)\bigr)^{9/2}}\right)H_{r-1}(wz)\,. \qquad (29.7)\end{aligned}$$

The probability that an evolving multigraph on n vertices leaves the top line of Figure 1 is $\Phi_n\nabla S$.

For fixed r we can evaluate the contribution made to $\Phi_n \nabla S$ by the rth term of (29.7), to within $O(n^{-1/3})$, because the leading coefficient $h_{(r-1)0}$ controls the asymptotic behavior. Indeed, we know from (25.22) and the subsequent discussion that

$$\Phi_n \left(w^r \frac{e^{U(w,z)} T(wz)^{2r}}{\bigl(1 - T(wz)\bigr)^{3r+3/2}} \right) = \frac{1}{6re_r} + O(n^{-1/3}) \tag{29.8}$$

for all fixed r. Therefore when Φ_n is applied to the rth term of (29.7) we get

$$\Phi_n e^U w^r \left(\frac{1}{2} \frac{T}{(1-T)^{7/2}} + \frac{5}{4} \frac{T^2}{(1-T)^{9/2}} \right) H_{r-1} = \frac{5k_{r-1}}{24re_r} + O(n^{-1/3}) \,. \tag{29.9}$$

When $r = 2$, the limit is $\frac{5}{77}$; when $r > 2$, (7.1) and (24.3) imply that

$$\frac{5k_{r-1}}{24re_r} = \left(\frac{5k_{r-2}}{24(r-1)e_{r-1}} \right) \left(\frac{36(r-1)(r-2)}{(6r-1)(6r-5)} \right) .$$

It follows by induction that

$$\frac{5k_{r-1}}{24re_r} = \frac{5}{36(r-1)r} \prod_{k=1}^{r-1} \frac{k(k+1)}{(k+\frac{1}{6})(k+\frac{5}{6})} = \prod_{k=1}^{r-2} \frac{k(k+1)}{(k+\frac{1}{6})(k+\frac{5}{6})} - \prod_{k=1}^{r-1} \frac{k(k+1)}{(k+\frac{1}{6})(k+\frac{5}{6})} .$$

So the sum over r is a telescoping series,

$$\sum_{r \ge 2} \frac{5k_{r-1}}{24re_r} = 1 - \prod_{k=1}^{\infty} \frac{k(k+1)}{(k+\frac{1}{6})(k+\frac{5}{6})} = 1 - \frac{5\pi}{18} \,. \tag{29.10}$$

In other words, convergence to the top-line probability depends entirely on the sum over r of the error term in (29.9).

The number of challenging and potentially fruitful questions that remain unanswered seems to be almost endless. But we shall close this list of research problems by stating what seems to be the single most important related area ripe for investigation at the present time. Wright [42] gave a procedure for computing the number of *strongly connected labeled digraphs* of excess r, analogous to his formulas for connected labeled undirected graphs. Random directed multigraphs are of great importance in computer applications, and it is shocking that so little attention has been given to their study so far. Karp [21] carried Wright's investigations further and discovered a beautiful theorem: A random digraph with $n(1+\mu)$ directed arcs almost surely has a giant strong component of size $\sim \Theta(\mu)^2 n$, when $\Theta(\mu)$ is the factor such that an undirected graph with $\frac{1}{2}n(1+\mu)$ edges almost surely has a giant component of size $\sim \Theta(\mu)n$. (The function $\Theta(\mu)$ is $(\mu+\sigma)/(1+\mu)$, according to (23.11). Karp's investigation was based on $\mathbf{D}_{n,p}$, in which every directed arc is present with probability p, but a similar result surely holds for other models of random digraphs.) A complete analysis of the random *directed* multigraph process is clearly called for, preferably based on generating functions so that extensive quantitative information can be derived without difficulty.

Here is a sketch of how such an investigation might begin. The *directed multigraph process* consists of adding directed arcs $x \to y$ repeatedly to an initially empty multiset of arcs on the vertices $\{1, 2, \ldots, n\}$, where x and y are independently and uniformly distributed between 1 and n. The *compensation factor* $\kappa(M)$ of a multidigraph M with m_{xy} arcs from x to y is $1/\prod_{x=1}^{n}\prod_{y=1}^{n} m_{xy}!\,$; we can use it to compute bivariate generating functions as in (2.1). The bgf for all possible multidigraphs is $\sum_{n\ge 0} e^{n^2 w} z^n/n! = G(2w, z)$.

Let $\mathcal{A}$ be the family of all multidigraphs such that all vertices are reachable from vertex 1 via a directed path, and let $A(w,z)$ be the corresponding bgf. There is a nice relation between $A(w,z)$ and the bgf $C(w,z)$ for connected undirected multigraphs, (2.10): If $A(w,z) = \sum_{n\ge 1} a_n(w) z^n/n!$, we have

$$\sum_{n\ge 1} a_n(w)\, e^{-n^2 w/2}\, \frac{z^n}{n!} \;=\; C(w,z)\,. \tag{29.11}$$

This can be proved by replacing z by $ze^{-w/2}$ and noting that $C(w, ze^{-w/2})$ is the bgf for connected multigraphs without self-loops, and by showing that all members of $\mathcal{A}$ are obtainable from such connected multigraphs M by the following reversible construction: Define a linear ordering $\prec$ on the vertices $\{1, 2, \ldots, n\}$ by saying that $x \prec y$ if $d(x) < d(y)$ or $d(x) = d(y)$ and $x < y$, where $d(x)$ is the distance from 1 to x in M. Then define a multidigraph $D \in \mathcal{A}$ by arcs $x \to y$ whenever x — y in M and $x \prec y$; include arbitrary additional arcs $x \to y$ for all pairs of vertices with $x \succeq y$. The construction is reversible because $d(x)$ is easily seen to be the distance from 1 to x in D, regardless of the choice of additional arcs. The additional arcs correspond to a multiplicative factor $e^{\binom{n+1}{2}w} = e^{n^2 w/2}(e^{w/2})^n$ in an n-vertex multigraph, with one factor e^w for each of the $\binom{n+1}{2}$ vertex pairs $x \succeq y$.

Let $\mathcal{S}$ be the family of all strongly connected multidigraphs, and let $S(w,z) = s_1(w)z + s_2(w)z^2/2! + s_3(w)z^3/3! + \cdots$ be the corresponding bgf. A nontrivial identity discovered by Wright [40] implies that we can calculate the coefficients $s_n(w)$ by using the formula

$$\sum_{n\ge 1} s_n(w)\, e^{-n^2 w/2}\, \frac{z^{n-1}}{(n-1)!}\, \frac{G(w,z)}{G(w, ze^{-nw})} \;=\; C'(w,z)\,, \tag{29.12}$$

where the prime in $C'(w,z)$ denotes differentiation with respect to z. Notice that our generating function $G(w,z)$ satisfies

$$\begin{aligned} &G'(w,z) = e^{w/2} G(w, ze^{w})\,, \quad G''(w,z) = e^{2w} G(w, ze^{2w})\,, \\ &\qquad \ldots\,, \quad G^{(n)}(w,z) = e^{n^2 w/2} G(w, ze^{nw})\,, \quad \ldots\,; \end{aligned} \tag{29.13}$$

thus the denominator $G(w, ze^{-nw})$ in (29.12) is essentially an n-fold integral of $G(w, z)$.

Wright [42] proved that the number of strongly connected digraphs with $n + r$ arcs on n vertices, disallowing self-loops and multiple arcs, is $n!$ times a polynomial in n of degree $3r - 1$, when $n > r > 0$. His proof can be adapted to multidigraphs, and everything becomes much simpler, just as formula (9.4) for multigraphs is simpler than formula (9.20) for graphs. The analogs of (2.11) and (3.4) are

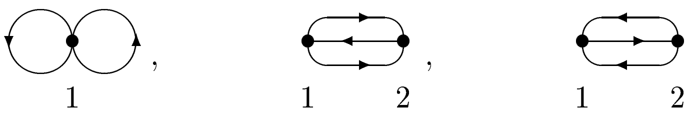
$$S(w, z) = w^{-1} S_{-1}(wz) + S_0(wz) + wS_1(wz) + w^2 S_2(wz) + \cdots, \tag{29.14}$$

where

$$S_{-1}(z) = z\,, \tag{29.15}$$

$$S_0(z) = -\ln(1 - z)\,, \tag{29.16}$$

and $S_r(z)$ for $r \geq 1$ can easily be shown to be $(1 - z)^{-3r}$ times a polynomial in z of degree $< 3r$. For example, the multidigraphs enumerated by $wS_1(wz)$ all arise by inserting ("uncancelling") vertices in the arcs of the reduced multidigraphs



whose generating functions are respectively $\frac{1}{2}w^2 z$, $\frac{1}{4}w^3 z^2$, $\frac{1}{4}w^3 z^2$. The operation of uncancelling corresponds to replacing w by $w/(1 - wz)$, as in Lemma 1; so $wS_1(wz) = \frac{1}{2}w^2 z/(1 - wz)^2 + \frac{1}{2}w^3 z^2/(1 - wz)^3 = \frac{1}{2}w^2 z/(1 - wz)^3$, and $S_1(z) = \frac{1}{2}z/(1 - z)^3$.

In fact, the numerator of $S_r(z)$ turns out to have a surprisingly small degree. Computer calculations indicate that we can write

$$S_r(z) = \frac{s_{r0} z^{2r-1}}{(1 - z)^{3r}} + \frac{s_{r1} z^{2r-2}}{(1 - z)^{3r-1}} + \cdots + \frac{s_{r(2r-2)} z}{(1 - z)^{r+2}}\,, \tag{29.17}$$

a formula analogous to (8.4), at least when $r \leq 5$. The coefficients are

$d =$	0	1	2	3	4	5	6	7	8
$s_{1d} =$	$\frac{1}{2}$								
$s_{2d} =$	$\frac{17}{8}$	$\frac{13}{8}$	$\frac{1}{6}$						
$s_{3d} =$	$\frac{275}{12}$	$\frac{427}{12}$	$\frac{391}{24}$	$\frac{13}{6}$	$\frac{1}{24}$				
$s_{4d} =$	$\frac{26141}{64}$	$\frac{61231}{64}$	$\frac{51299}{64}$	$\frac{18473}{64}$	$\frac{6047}{144}$	$\frac{263}{144}$	$\frac{1}{120}$		
$s_{5d} =$	$\frac{1630711}{160}$	$\frac{1276481}{40}$	$\frac{3125933}{80}$	$\frac{2840093}{120}$	$\frac{3546283}{480}$	$\frac{6743}{6}$	$\frac{25307}{360}$	$\frac{43}{36}$	$\frac{1}{720}$

No reason why $S_r(z)$ should have the simple form (29.17) is apparent; this phenomenon cries out for explanation, if it is indeed true for all $r > 0$, and the explanation will probably lead to new theorems of interest. It can be shown that this conjecture is equivalent to the assertion that the sum of $(-1)^\nu \kappa/\nu!$, over all labelled, reduced, strongly connected multidigraphs of excess r, is zero; or in other words, if we choose a labelled, reduced, strongly connected multidigraph of excess r at random, with probabilities weighted in the natural way by the compensation factor κ, then the probability is $\frac{1}{2}$ that there will be an even number of vertices.

Is there a simple recurrence governing the leading coefficients s_{10}, s_{20}, s_{30}, $\ldots$, perhaps analogous to the relation we observed for ordinary connected components in (8.5)?

Acknowledgments. The authors wish to thank Prof. Richard Askey for helpful correspondence relating to this research.

Appendix. Here is a list of corrections to the related paper [14].

Page 175, line 10: $(1+t)^N$ should be $(1+t)^{-N}$
Page 175, line 11: (3.5) should be (3.6)
Page 182, (4.21): $\sqrt{3t}$ should be $\sqrt{3}\,t$
Page 183, line 18: $\frac{1}{2}\sqrt{3t}$ should be $\frac{i}{2}\sqrt{3t}$
Page 183, line 24: (4.27) should be (4.25)
Page 184, (5.6): $l = 1$ should be $l - 1$
Page 185, line 17: $l = 2$ should be $l = 3$
Page 189, lines 4 and 9: $\frac{1}{2}l(l-1)$ should be $\frac{1}{2}l(l+1)$
Page 192, (7.13): $\frac{31}{45}$ should be $\frac{1}{45}$; $2 + 3\hat{p}_3$ should be $\hat{p}_3$
Page 194, line 15: 'than $\Re h(\lambda) - \lambda - (1 - \frac{1}{2}\lambda)(\ln(1 - \frac{1}{2}\lambda) - \ln(1 + \frac{1}{2}\lambda))$
$< \Re h(\lambda) - \frac{1}{3}\lambda^2$ when'
Page 205, line 7: delete 'number of'
Page 207, (11.9): delete commas in denominator
Page 209, first line of (A.6): $ixt - it^3/3$ should be $ixt + it^3/3$
Page 213, the argument for enveloping series is incomplete
Page 215, (11.12) and (11.14): delete commas in denominators

Bibliography

[1] G. N. Bagaev, "Sluchaĭnye grafy so stepen'i͡u svi͡aznosti 2," *Diskretnyĭ Analiz* **22** (1973), 3–14.

[2] G. N. Bagaev and E. F. Dmitriev, "Perechislenie svi͡aznykh otmechennykh dvudol'nykh grafov," *Doklady Akademii͡a Nauk BSSR* **28** (1984), 1061–1063.

[3] Edward A. Bender, E. Rodney Canfield, and Brendan D. McKay, "The asymptotic number of labeled connected graphs with a given number of vertices and edges," *Random Structures and Algorithms* **1** (1990), 127–169.

[4] Béla Bollobás, "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs," *European Journal of Combinatorics* **1** (1980), 311–316.

[5] Béla Bollobás, "The evolution of random graphs," *Transactions of the American Mathematical Society* **286** (1984), 257–274.

[6] Béla Bollobás, *Random Graphs* (London: Academic Press, 1985).

[7] B. Bollobás and A. Frieze, "On matchings and Hamiltonian cycles in random graphs," in *Random Graphs '83*, edited by Michał Karoński and Andrzej Ruciński, *Annals of Discrete Mathematics* **28** (1985), 23–46.

[8] C. W. Borchardt, "Ueber eine der Interpolation entsprechende Darstellung der Eliminations-Resultante," *Journal für die reine und angewandte Mathematik* **57** (1860), 111–121.

[9] V. E. Britikov, "O strukture sluchaĭnogo grafa vblizi kriticheskoĭ tochki," *Diskretnai͡a Matematika* **1**,3 (1989), 121–128. English translation, "On the random graph structure near the critical point," *Discrete Mathematics and Applications* **1**,3 (1991), 301–309.

[10] A. Cayley, "A theorem on trees," *Quarterly Journal of Pure and Applied Mathematics* **23** (1889), 376–378. Reprinted in his *Mathematical Papers* **13**, 26–28.

[11] Gotthold Eisenstein, "Entwicklung von $\alpha^{\alpha^{\alpha^{\cdot^{\cdot^{\cdot}}}}}$," *Journal für die reine und angewandte Mathematik* **28** (1844), 49–52.

[12] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae (Debrecen)* **6** (1959), 290–297. Reprinted in *Paul Erdős: The Art of Counting* (MIT Press, 1973), 561–568; and in *Selected Papers of Alfréd Rényi* (Akadémiai Kiadó, 1976), 308–315.

[13] P. Erdős and A. Rényi, "On the evolution of random graphs," *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **5** (1960), 17–61. Reprinted in *Paul Erdős: The Art of Counting* (MIT Press, 1973), 574–618; and in *Selected Papers of Alfréd Rényi* (Akadémiai Kiadó, 1976), 482–525.

[14] Philippe Flajolet, Donald E. Knuth, and Boris Pittel, "The first cycles in an evolving graph," *Discrete Mathematics* **75** (1989), 167–215.

[15] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, "Correlation inequalities on some partially ordered sets," *Communications in Mathematical Physics* **22** (1971), 89–103.

[16] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration* (New York: Wiley, 1983).

[17] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics* (Reading, Massachusetts: Addison-Wesley, 1989).

[18] Peter Henrici, *Applied and Computational Complex Analysis*, volume 2, Wiley, 1977.

[19] Svante Janson, "Multicyclic components in a random graph process," *Random Structures and Algorithms* **4** (1993), 71–84.

[20] Svante Janson and Tomasz Łuczak, "The size of the last cycle in the random graph process," *Abstracts of Papers Presented to the American Mathematical Society* **13** (1992), 354, abstract 875-05-131.

[21] Richard M. Karp, "The transitive closure of a random digraph," *Random Structures and Algorithms* **1** (1990), 73–93.

[22] Donald E. Knuth, "An analysis of optimum caching," *Journal of Algorithms* **6** (1985), 181–199.

[23] Donald E. Knuth, "Convolution polynomials," *The Mathematica Journal* **2**,4 (Fall 1992), 67–78.

[24] Donald E. Knuth and Boris Pittel, "A recurrence related to trees," *Proceedings of the American Mathematical Society* **105** (1989), 335–349.

[25] Tomasz Łuczak, "Component behavior near the critical point of the random graph process," *Random Structures and Algorithms* **1** (1990), 287–310.

[26] Tomasz Łuczak, "Cycles in a random graph near the critical point," *Random Structures and Algorithms* **2** (1991), 421–439.

[27] Tomasz Łuczak and John C. Wierman, "The chromatic number of random graphs at the double-jump threshold," *Combinatorica* **9** (1989), 39–49.

[28] Tomasz Łuczak, Boris Pittel, and John C. Wierman, "The structure of a random graph at the point of phase transition," to appear in *Transactions of the American Mathematical Society*.

[29] D. S. Mitrinović, *Analytic Inequalities* (Springer-Verlag, 1970).

[30] S. Ramanujan, "Questions for solution, number 294," *Journal of the Indian Mathematical Society* **3** (1911), 128; **4** (1912), 151–152.

[31] Alfred Rényi, "Some remarks on the theory of trees," *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **4** (1959), 73–85. Reprinted in *Selected Papers of Alfréd Rényi* **2**, 363–374.

[32] Robert James Riddell, Jr., *Contributions to the Theory of Condensation.* Dissertation, University of Michigan, 1951. (The main results of this dissertation were published as R. J. Riddell, Jr., and G. E. Uhlenbeck, "On the theory of the virial development of the equation of state of monoatomic gases," *Journal of Chemical Physics* **21** (1953), 2056–2064.)

[33] G. Seitz, "Une remarque aux inégalités," *Aktuarské Vědy* **6** (1936/37), 167–171.

[34] L. J. Slater, "Expansions of generalized Whittaker functions," *Proceedings of the Cambridge Philosophical Society* **50** (1954), 628–630.

[35] V. E. Stepanov, "Neskol'ko teorem otnositel'no sluchaĭnykh grafov," *Veroi͡atnostnye metody v diskretnoĭ matematike* (Karel'skiĭ filial Akademii͡a Nauk SSSR, Petrozavodsk, 1983), 90–92.

[36] V. E. Stepanov, "O nekotorykh osobennosti͡akh stroenii͡a sluchaĭnogo grafa vblizi kriticheskoĭ tochki," *Teoriya Veroyatnostei i ee Primenenii͡a* **32** (1988), 633–657. English translation, "On some features of the structure of a random graph near a critical point," *Theory of Probability and Its Applications* **32** (1988), 573–594.

[37] J. J. Sylvester, "On the change of systems of independent variables," *Quarterly Journal of Pure and Applied Mathematics* **1** (1857), 42–56. Reprinted in his *Mathematical Papers* **2**, 65–85.

[38] V. A. Voblyĭ, "O koeffitsientakh Raĭta i Stepanova-Raĭta," *Matematicheskie Zametki* 42 (1987), 854–862. English translation, V. A. Voblyi, "Wright and Stepanov-Wright coefficients," *Mathematical Notes* **42** (1987), 969–974.

[39] E. M. Wright, "A relationship between two sequences," *Proceedings of the London Mathematic Society* **17** (1967), 296–304, 547–552.

[40] E. M. Wright, "The number of strong digraphs," *Bulletin of the London Mathematical Society* **3** (1971), 348–350.

[41] E. M. Wright, "The number of connected sparsely edged graphs," *Journal of Graph Theory* **1** (1977), 317–330.

[42] E. M. Wright, "Formulae for the number of sparsely-edged strong labelled digraphs," *Quarterly Journal of Mathematics*, Oxford (2), **28** (1977), 363–368.

[43] E. M. Wright, "The number of connected sparsely edged graphs. II. Smooth graphs and blocks," *Journal of Graph Theory* **2** (1978), 299–305.

[44] E. M. Wright, "The number of connected sparsely edged graphs. III. Asymptotic results," *Journal of Graph Theory* **4** (1980), 393–407.

[45] E. M. Wright, "The number of connected sparsely edged graphs. IV. Large nonseparable graphs," *Journal of Graph Theory* **7** (1983), 219–229.