QUEUES, RANDOM GRAPHS AND BRANCHING PROCESSES*

Lajos Takács
Case Western Reserve University,
Cleveland, Ohio

ABSTRACT

In this paper it is shown that certain basic results of queueing theory can be used successfully in solving various problems of random graphs and branching processes.

Key Words: Queues, Random Graphs, Branching Processes, Limit Theorems.

AMS subject classifications: 60K25, 90B22, 05C80, 60J80.

DR/MS Index subject classification: 683 Queues, 482 Graphs, 572 Stochastic Model Applications.

1. INTRODUCTION

Some basic results of queueing theory turn out to be also useful in various other fields of mathematics. By way of illustration we will focus on two topics, namely, random graphs and branching processes.

In what follows we describe a general queueing model and define a random graph and a branching process related to it. After deriving some specific results for the queueing model considered, we shall demonstrate their applications in the fields mentioned above.

The queueing model. Let us suppose that in the time interval $(0,\infty)$ customers arrive at random at a counter and are served singly by one server in order of arrival. It is assumed that the server starts working at time $t = 0$ and at that time i $(i = 0, 1, 2,...)$ customers are already waiting for service. The initial i customers are numbered $1, 2, ..., i$ and the customers arriving subsequently are numbered $i+1, i + 2,...$ in the order of their arrivals. Denote by ν_r the number of customers arriving during the service time of the r-th customer. This queueing model will be

---

* Invited paper.

Research supported by the National Science Foundation Grant DMS-85-10073.

characterized by the initial queue size i and the sequence of random variables $\nu_1, \nu_2, \ldots, \nu_r, \ldots$. Throughout this paper we use the abbreviation

$$N_r = \nu_1 + \nu_2 + \ldots + \nu_r \tag{1}$$

for r = 1, 2, ..., and write $N_0 = 0$.

We shall now consider the following random variables: $\rho(i)$, the number of customers served in the initial busy period, for i = 1,2,... and $\rho(0) = 0$. We write

$$f_n^{(i)} = P\{\rho(i) = n\} \tag{2}$$

for $1 \le i \le n$ and $f_0^{(0)} = 1, f_n^{(0)} = 0$ for $n \ge 1$.

We define $\zeta_r (r = 1, 2, \ldots)$ as the queue size immediately after the r-th service ends, and write

$$p_i(n, k) = P\{\zeta_r < k \text{ for } 1 \le r \le n \text{ and } \rho(i) = n\} \tag{3}$$

for $1 \le i \le n$ and $k \ge 1$.

Following D. G. Kendall [10] we say that the initial i customers (if $i \ge 1$) in the queue form the 0-th generation. The customers (if any) arriving during the total service time of the initial i customers form the 1-st generation. Generally, the customers (if any) arriving during the total service time of the customers in the (r-1)-th generation form the r-th generation for r = 1, 2,... . Denote by γ_r (r = 0, 1, 2, ...) the number of customers in the r-th generation and define

$$\tau(i) = \sup\{r: \gamma_r > 0, r \ge 0\}. \tag{4}$$

If $\gamma_r > 0$ for all $r \ge 0$, then $\tau(i) = \infty$. Write

$$h_i(n, k) = P\{\tau(i) \le k \text{ and } \rho(i) = n\} \tag{5}$$

for $1 \le i \le n$ and $k \ge 1$.

An associated random graph. We associate a random graph with the queueing process considered. If $i \geq 1$ and if the initial busy period consists of n services, that is, if $\rho(i) = n$, then let us assume that the graph has vertex set $\{1, 2,..., n\}$. Furthermore, in the graph, two vertices r and s are joined by an edge if and only if the r-th customer arrives during the service time of the s-th customer.

If $\rho(i) = n$ in the queueing process and $i = 1$, then the graph is a rooted tree with n labeled vertices, the root being labeled 1. If $\rho(i) = n$ and $i > 1$, then the graph is a forest consisting of i distinct rooted trees whose roots are labeled 1, 2,..., i respectively. The total number of vertices in the forest is n.

We define the height of the forest as the maximal distance between any vertex of the graph and the root of the tree which contains the given vertex. In the notation of (4), $\tau(i)$ is the height of the random forest for $i \geq 1$. If $i = 1$, the forest becomes a tree.

An associated branching process. Now let us assume that in the queueing process $\nu_1, \nu_2,..., \nu_r,...$ is a sequence of independent and identically distributed random variables for which

$$P\{\nu_r = j\} = p_j \tag{6}$$

for $j = 0, 1, 2,...$ and $r = 1, 2,...$. In this case the sequence of random variables $\{\gamma_0, \gamma_1,..., \gamma_r,...\}$ forms a branching process. We can imagine that in a population initially we have i ($i \geq 1$) individuals and in each generation each individual reproduces independently of the others and has probability p_j ($j = 0, 1, 2,...$) of giving rise to j descendants in the following generation. The random variable γ_r is the number of individuals in the r-th generation. Define

$$\rho^*(i) = \gamma_0 + \gamma_1 + ... + \gamma_r + ..., \tag{7}$$

that is, $\rho^*(i)$ is the total number of individuals (total progeny) in the process. Possibly, $\rho^*(i) = \infty$. Also we define

$$\tau^*(i) = \inf\{r: \gamma_r = 0, r \geq 0\}, \tag{8}$$

that is, $\tau^*(i)$ is the time of extinction ($\tau^*(i) = \infty$ if $\gamma_r > 0$ for all $r \geq 0$).

Now $\rho^*(i)$, defined by (7), has the same distribution as $\rho(i)$ in the queueing process, namely (2). Also $\tau^*(i)$, defined by (8), has the same distribution as $\tau(i)+1$ where $\tau(i)$ is defined by (4).

The following discussion is based mostly on a simple combinatorial theorem which is a generalization of the classical ballot theorem. It can be formulated as follows:

Theorem 1. *Let us suppose that a box contains* n *cards marked with nonnegative integers* $k_1, k_2, \ldots, k_n$ *respectively where* $k_1 + k_2 + \ldots k_n = k \leq n$. *All the* n *cards are drawn without replacement from the box. Denote by* v_r *the number drawn at the* r-*th drawing* (r = 1, 2,...,n). *Then*

$$P\{v_1 + \ldots + v_r < r \text{ for } 1 \leq r \leq n\} = (n - k)/n \tag{9}$$

provided that all the possible results are equally probable.

Proof. Denote by S(n, k) the number of favorable results, that is, the number of sequences $(k_1, k_2, \ldots, k_n)$ in which the sum of the first r numbers drawn is less than r for every r = 1, 2, ... , n. If we take into consideration that the last number drawn may be $k_1, k_2, \ldots, k_n$, then we can write down that

$$S(n, k) = \sum_{i=1}^{n} S(n - 1, k - k_i). \tag{10}$$

Evidently, S(1, 0) = 1 and S(n, n) = 0 for all $n \geq 1$. From (10) it follows by mathematical induction that

$$S(n, k) = (n - k)\,(n - 1)! \tag{11}$$

for $0 \leq k \leq n$ and $n \geq 1$. If n = 1, then (11) is true. Suppose that (11) is true for S(n - 1, k), $0 \leq k \leq n-1$, $n \geq 2$. If k = n, then S(n, k) = 0. If k < n and $n \geq 2$, then by (10) and by the induction hypothesis

$$S(n, k) = \sum_{i=1}^{n} (n - 1 - k + k_i)(n - 2)! = (n - k)\,(n - 1)\,! \;. \tag{12}$$

Consequently, (11) is true for all $n = 1, 2,...$ and $0 \le k \le n$. Finally, $S(n, k)/n!$ yields the desired probability (9).

Theorem 1 immediately implies the following more general result.

Theorem 2. *Let* $\nu_1, \nu_2,..., \nu_n$ *be interchangeable discrete random variables taking on nonnegative integers only, and define* $N_r = \nu_1 + ...+ \nu_r$ *for* $r = 1, 2,...,$ n. *We have*

$$P\{N_r<r \textit{ for } 1\le r\le n \mid N_n = k\} = \begin{cases}(n-k)/n & \textit{if } k = 0,1,...,n,\\ 0 & \textit{if } k >n,\end{cases} \tag{13}$$

provided that $P\{N_n = k\} > 0$.

Proof. If we apply Theorem 1 to every realization of the sequence of the random variables $\nu_1, \nu_2,..,\nu_n$, we get (13).

In a particular case Theorem 1 was formulated by J. Bertrand [2] in 1887 and was proved in the same year by D. André [1]. The above generalizations have been given by L. Takács [26].

2. THE QUEUEING MODEL

In the queueing model defined in the Introduction, the probability (2), that is, the probability that the initial busy period consists of n services, is completely determined by the joint distribution of the random variables $\nu_1, \nu_2,..., \nu_n$. Since $\rho(i) = n$ if and only if $N_r > r - i$ for $i \le r < n$ and $N_n = n - i$, we have

$$f_n^{(i)} = P\{N_r > r - i \text{ for } i \le r < n \text{ and } N_n = n - i\} \tag{14}$$

for $1 \le i \le n$.

If $\nu_1, \nu_2,..., \nu_n$ are interchangeable random variables, then we can express (14) in the following equivalent form:

$$f_n^{(i)} = P\{N_r < r \text{ for } 1 \le r \le n - i \text{ and } N_n = n - i\} \tag{15}$$

for $1 \le i \le n$. For, if the random variables $\nu_1, \nu_2,...,\nu_n$ are interchangeable, (14) remains unchanged if we replace ν_r by ν_{n+1-r} for every $r = 1, 2,..., n$.

Theorem 3. *If* $\nu_1, \nu_2, \ldots, \nu_n$ *are interchangeable random variables, then we have*

$$f_n^{(i)} = \frac{i}{n} P\{N_n = n - i\} \tag{16}$$

for $1 \le i \le n$.

Proof. Formula (16) is an immediate consequence of (13).

If $\nu_1, \nu_2, \ldots, \nu_n$ is a sequence of independent and identically distributed random variables, then they are interchangeable and (16) holds unchangeably. In this case, as an alternative, we can use generating functions for the determination of $f_n^{(i)}$. See L. Takács [24] and [27]. Let us introduce the notations: $P\{\nu_r = j\} = p_j$ $(j = 0,1,2,\ldots)$,

$$g(z) = \sum_{j=0}^{\infty} p_j z^j \tag{17}$$

for $|z| \le 1$,

$$a = \sum_{j=0}^{\infty} j p_j , \tag{18}$$

$f_n = f_n^{(1)}$ for $n \ge 1$, and

$$f(w) = \sum_{n=1}^{\infty} f_n w^n \tag{19}$$

for $|w| \le 1$. Since the number of customers served in the initial busy period can be expressed as the sum of i independent random variables each having the same distribution $\{f_n\}$, we have

$$\sum_{n=i}^{\infty} f_n^{(i)} w^n = [f(w)]^i \tag{20}$$

for $i = 0, 1, 2, \ldots$.

On the other hand, if we take into consideration that the number of customers arriving during the first service time may be $j = 0, 1, 2,\ldots$, then we can write that

$$f_n = \sum_{j=0}^{n-1} p_j f_{n-1}^{(j)} \tag{21}$$

for $n \geq 1$. Hence

$$f(w) = w \sum_{j=0}^{\infty} p_j [f(w)]^j = w\, g(f(w)) \tag{22}$$

for $|w| \leq 1$. Accordingly, $z = f(w)$, $|w| \leq 1$, is a solution of the equation

$$z = w\, g(z). \tag{23}$$

The function $g(z)$ is regular in the unit disk $|z| < 1$, and $|wg(z)| \leq |w| < |z|$ if $|z| = 1$ and $|w| < 1$. Thus by Rouché's theorem, equation (23) has exactly one root $z = \zeta$ in the unit disk whenever $|w| < 1$. Thus necessarily, $\zeta = f(w)$. By Lagrange's expansion we obtain

$$\zeta^i = i \sum_{n=1}^{\infty} \frac{w^n}{n!} \left(\frac{d^{n-1} z^{i-1} [g(z)]^n}{dz^{n-1}} \right)_{z=0} = i \sum_{n=i}^{\infty} \frac{w^n}{n} P\{N_n = n - i\} \tag{24}$$

for $i = 1, 2,\ldots$. This proves (16) once more.

Theorem 4. *For an arbitrary set of random variables* $\nu_1, \nu_2,\ldots, \nu_n$ *the probability* (3) *can be expressed in the following way:*

$$p_i(n,k) = P\{i - k < r - N_r < i \text{ for } 1 \leq r < n \text{ and } n - N_n = i\} \tag{25}$$

if $1 \leq i \leq n$ *and* $k \geq 1$.

Proof. If $\rho(i) = n$, then $N_r > r - i$ for $1 \le r < n$, $N_n = n - i$, and consequently $\zeta_r = N_r - r + 1$ for $1 \le r \le n$.

Now let us suppose again that $\nu_1, \nu_2, \ldots, \nu_r, \ldots$ is a sequence of independent and identically distributed random variables with the distribution $P\{\nu_r = j\} = p_j$ $(j = 0, 1, 2, \ldots)$ and generating function $g(z)$ defined by (17). We are interested in determining the probabilities (5). Let us introduce the generating functions

$$h_k(z) = \sum_{n=1}^{\infty} P\{\tau(1) \le k \text{ and } \rho(1) = n\} z^n \tag{26}$$

for $k = 1, 2, \ldots$ and $|z| \le 1$. In analogy with (20), we have

$$\sum_{n=1}^{\infty} P\{\tau(i) \le k \text{ and } \rho(i) \le n\} z^n = [h_k(z)]^i \tag{27}$$

for $i \ge 1, k \ge 1$ and $|z| \le 1$. Thus the problem of finding (5) can be reduced to the determination of (26).

Theorem 5. *If $\nu_1, \nu_2, \ldots, \nu_r, \ldots$ is a sequence of independent and identically distributed random variables with generating function* (17), *then*

$$h_k(z) = zg(h_{k-1}(z)) \tag{28}$$

for $k = 1, 2, \ldots$ *and* $|z| \le 1$ *where* $h_0(z) = 0$.

Proof. In exactly the same way as we proved (22) we obtain that

$$h_k(z) = z \sum_{j=0}^{\infty} p_j [h_{k-1}(z)]^j \tag{29}$$

for $k \ge 1$ and $|z| \le 1$ where $h_0(z) = 0$. This proves (28).

3. EXAMPLES FOR THE QUEUEING PROCESS

The following examples have some interest in their own, but they are also useful in studying random graphs.

EXAMPLE 1. Let us suppose that in the queueing model defined in the Introduction $\nu_1, \nu_2, \ldots, \nu_r, \ldots$ is a sequence of independent random variables and the distribution of ν_r is given by

$$P\{\nu_r = j\} = e^{-\lambda p q^{r-1}} (\lambda p q^{r-1})^j / j! \tag{30}$$

for $j = 0, 1, 2, \ldots$ and $r = 1, 2, \ldots$ where $\lambda > 0$, $0 < p < 1$ and $q = 1 - p$. In this case the random variables $\nu_1, \nu_2, \ldots, \nu_n$ are not interchangeable, but by (14) we obtain that

$$f_n^{(i)} = e^{-\lambda(1-q^n)} (\lambda p)^{n-i} \sum_{\substack{j_1+\ldots+j_n=n-i \\ j_1+\ldots+j_r > r-i \ (i \le r \le n)}} \frac{q^{j_1 + 2j_2 + \ldots + nj_n - (n-i)}}{j_1!\, j_2! \ldots j_n!} \tag{31}$$

for $1 \le i \le n$, or equivalently,

$$f_n^{(i)} = \frac{e^{-\lambda(1-q^n)} (\lambda p)^{n-i}}{(n-i)!} \phi_n^{(i)}(q) \tag{32}$$

where

$$\phi_n^{(i)}(q) = \sum_{\substack{j_1+\ldots+j_n=n-i \\ j_1+\ldots+j_r < r \ (1 \le r \le n-i)}} \frac{(n-1)!}{j_1!\, j_2! \ldots j_n!} q^{n(n-1)-j_1-2j_2-\ldots-nj_n} \tag{33}$$

and is a polynomial of degree $(n - i)(n + i - 3)/2$ in q.

If we take into consideration that in (33) $j_n = j$ where $0 \le j \le n - i$, then we obtain the recurrence formula

$$\phi_n^{(i)}(q) = \sum_{j=0}^{n-i} \binom{n-i}{j} \phi_{n-1}^{(i+j-1)}(q)\, q^{n-i-j} \tag{34}$$

for $1 \le i \le n$ where $\phi_0^{(0)}(q) = 1$ and $\phi_n^{(0)}(q) = 0$ for $n \ge 1$.

We note that

$$\phi_n^{(i)}(1) = i\, n^{n-i-1} \tag{35}$$

for $1 \le i \le n$. This can be proved by (34). If we put $q = 1$ in (34), we obtain that

$$\phi_n^{(i)}(1) = \sum_{j=0}^{n-i} \binom{n-i}{j} \phi_{n-1}^{(i+j-1)}(1) \tag{36}$$

and (35) follows from (36) by mathematical induction.

EXAMPLE 2. Let us suppose that in the queueing process customers arrive in the time interval $(0, \infty)$ in accordance with a Poisson process with parameter λ, and the service times are independent random variables each having the same exponential distribution function

$$H(x) = \begin{cases} 1 - e^{-\mu x} & \text{for } x \ge 0, \\ 0 & \text{for } x < 0, \end{cases} \tag{37}$$

and are independent of the arrival times. In this case $v_1, v_2, \ldots, v_r, \ldots$ is a sequence of independent and identically distributed random variables for which

$$P\{v_r = j\} = p_j = qp^j \tag{38}$$

if $j = 0, 1, 2, \ldots$ where $p = \lambda / (\lambda + \mu)$ and $q = \mu / (\lambda + \mu)$.

Now by (16) we obtain that

(39) $$P\{\rho(i) = n\} = \frac{1}{n}\binom{2n-i-1}{n-1} p^{n-i} q^n$$

for $1 \le i \le n$, and (3) can also be expressed in the following way:

(40) $$p_i(n, k) = P\{-i < \eta_r \le k - i \text{ for } 1 \le r < 2n - i \text{ and } \eta_{2n-i} = -i\}$$

for $1 \le i \le n$ and $k \ge 1$ where the random variables $\eta_0, \eta_1, \ldots, \eta_r, \ldots$ describe a random walk on the real axis. A particle starts at $x = 0$ and in each step it moves either a unit distance to the right with probability p or a unit distance to the left with probability q. The successive displacements are independent and η_r denotes the position of the particle after the r - th step ($r = 1, 2, \ldots$); $\eta_0 = 0$. The probabilities (40) can easily be determined explicitly by using the reflection principle. See L. Takács [25]. In particular, for $i = 1$, we obtain that

(41) $$\begin{aligned} p_1(n+1,k) &= p^n q^{n+1} \sum_j \left[\binom{2n}{n+j(k+1)} - \binom{2n}{n+j(k+1)+k}\right] \\ &= \frac{2^{2n+1} p^n q^{n+1}}{(k+1)} \sum_{r=1}^{k} \left(\cos \frac{r\pi}{k+1}\right)^{2n} \left(\sin \frac{r\pi}{k+1}\right)^2. \end{aligned}$$

Now the probabilities

(42) $$h_i(n,k) = P\{\tau(i) \le k \text{ and } \rho(i) = n\} = a_i(n,k)\, p^{n-i} q^n$$

for $1 \le i \le n$ and $k \ge 1$ can be obtained by Theorem 5. In this case $g(z) = qz / (1-pz)$; and if we define

(43) $$A_k(x) = \sum_{n=1}^{\infty} a_1(n,k)\, x^n$$

for $k \ge 1$ and $|x| < 1$, then by (28) we obtain that

(44) $$A_k(x) = \frac{x}{1 - A_{k-1}(x)}$$

for $k \geq 1$ where $A_0(x) = 0$. By (44) we can prove that

(45) $$A_k(x) = x^{1/2} U_{k-1}(1 / 2\sqrt{x}) / U_k(1 / 2\sqrt{x})$$

for $k = 1, 2,...$ where $U_k(x)$ $(k = 0, 1, 2,...)$ are Chebyshev's polynomials of the second kind. We have $U_0(x) = 1$, $U_1(x) = 2x$, and $U_{k+2}(x) = 2x\, U_{k+1}(x) - U_k(x)$ for $k = 0, 1, 2,...$. Since $U_k(\cos \varphi) = \sin (k + 1)\varphi / \sin \varphi$ for $k = 0, 1, 2,...$, it follows that $x_r = \cos (r \pi / (k + 1))$, $r = 1, 2,..., k$, are the k roots of $U_k(x) = 0$. Thus by Lagrange's formula we obtain

(46) $$\frac{U_{k-1}(x)}{U_k(x)} = \sum_{r=1}^{k} \frac{U_{k-1}(\cos \frac{r\pi}{k+1})}{U'_k(\cos \frac{r\pi}{k+1})(x - \cos \frac{r\pi}{k+1})} = \frac{1}{k+1} \sum_{r=1}^{k} \frac{(\sin \frac{r\pi}{k+1})^2}{x - \cos \frac{r\pi}{k+1}}.$$

If we put $x = y^2 / 4$ in (45), then by (43) and (46) we obtain that

(47) $$A_k(y^2/4) = \sum_{n=1}^{\infty} a_1(n, k) (\frac{y}{2})^{2n} = \frac{y^2}{2(k+1)} \sum_{r=1}^{k} \frac{(\sin \frac{r\pi}{k+1})^2}{1 - y \cos \frac{r\pi}{k+1}}.$$

Expanding the extreme right member into Taylor series and forming the coefficient of y^{2n+2}, we obtain that

(48) $$a_1(n + 1, k) = \frac{2^{2n+1}}{k+1} \sum_{r=1}^{k} (\cos \frac{r\pi}{k+1})^{2n} (\sin \frac{r\pi}{k+1})^2.$$

A comparison of (41), (42) and (48) shows that

(49) $$h_1(n + 1, k) = p_1(n + 1, k) = a_1(n + 1, k)\, p^n q^{n+1}$$

or, in other words, if $i = 1$, the two probabilities (40), and (42) are the same for every $n \geq 1$ and $k \geq 1$. This identity is not evident, but it can also be proved directly.

4. CAYLEY'S FORMULA

A forest is a simple graph that has no cycles. In other words, a forest is a simple graph, all of whose components are trees.

Denote by $F(n, i)$, $1 \leq i \leq n$, the number of forests having vertex set $\{1, 2,..., n\}$ and i components which are trees such that vertices 1, 2,..., i all belong to different trees. We have

$$F(n, i) = i\, n^{n-i-1} \tag{50}$$

for $1 \leq i \leq n$. In particular,

$$F(n, 1) = n^{n-2} \tag{51}$$

is the number of distinct trees with n labeled vertices.

In 1889 A. Cayley [6] discovered formula (51). Since then various proofs have been given for (51) by O. Dziobek [8], H. Prüfer [19], G. Bol [3], L.E. Clarke [7], J.W. Moon [15], [16], A. Rényi [20], [21] and others.

A. Cayley [6] also stated formula (50), but he did not indicate how to prove it. In 1959 A. Rényi [20] gave an analytic proof for (50). For other proofs of (50) we refer to J.W. Moon [17], [18], J. Riordan [22], V.F. Kolchin [13] and V.N. Sachkov [23]. All these proofs of (50) presume that the particular case (51) has already been proved. In what follows we shall provide a simple proof of (50) which does not presume (51).

Let us consider the queueing model defined in the Introduction. Suppose that at time $t = 0$, i ($1 \leq i \leq n$) customers are waiting for service, and in the time interval $(0, n]$ exactly $n - i$ customers arrive in such a way that independently of each other each customer may join the queue in any interval $(r - 1, r]$ ($r = 1, 2,..., n$) with probability $1 / n$. The service times are assumed to be constant of unit length. In this case $\nu_1, \nu_2,..., \nu_n$ are interchangeable random variables with sum

$\nu_1 + \nu_2 + \ldots + \nu_n = n - i$. If $\rho_n(i)$ denotes the total number of customers served during the initial busy period, then by Theorem 3 we have

$$P\{\rho_n(i) = n\} = \frac{i}{n} \tag{52}$$

for $1 \le i \le n$.

On the other hand, we have

$$P\{\rho_n(i) = n\} = \frac{F(n, i)}{n^{n-i}} \tag{53}$$

for $1 \le i \le n$. To prove (53) let us consider the graph associated with the queueing process. The graph has vertex set $\{1, 2, \ldots, n\}$. There are n^{n-i} possible graphs, and they are equally probable. If $\rho_n(i) = n$, then the graph consists of i tree components such that vertices $1, 2, \ldots, i$ all belong to different trees. Conversely, to every such graph there corresponds a queueing process for which $\rho_n(i) = n$. The number of favorable graphs (i.e. graphs corresponding to queueing processes for which $\rho_n(i) = n$) is $F(n, i)$. This proves (53). A comparison of (52) and (53) implies (50).

5. RANDOM GRAPHS

Here we are concerned with a random graph connected with the structure of polymers in chemistry. We define a random graph $\Gamma_n(p)$ on the vertex set $V_n = \{1, 2, \ldots, n\}$ in the following way. First, we form a complete graph K_n on the vertex set V_n. The complete graph is a simple graph which is undirected and has no loops or multiple edges. It contains all the possible $\binom{n}{2}$ edges. If in K_n, each edge, independently of the others, is either retained with probability p or removed with probability q where $p + q = 1$ and $0<p<1$, then we obtain $\Gamma_n(p)$.

Now let us fix a set of i vertices ($1 \le i \le n$) in V_n, say $\{1, 2, \ldots, i\}$, and let us define $\rho_n(i)$ as the number of vertices in the union of all those components of $\Gamma_n(p)$ which contain at least one vertex of the set of vertices $\{1, 2, \ldots, i\}$. Thus $\rho_n(i)$ is i plus the number of vertices in the set $\{i + 1, \ldots, n\}$ which are connected

by a path (an uninterrupted sequence of edges) with one of the vertices in the set $\{1, 2,..., i\}$.

J.W. Kennedy [11] derived a recurrence formula for the determination of the distribution of $\rho_n(1)$ and calculated the probabilities $P\{\rho_n(1) = k\}$ for $k \le 5$.

In what follows we shall demonstrate that $\rho_n(i)$ has the same distribution as the number of customers served in the initial busy period in an appropriate queueing process.

To describe the relevant queueing process let us suppose that n customers, numbered 1, 2,..., n, are served singly by one server. The service times are assumed to be constant of unit length. The server starts working at time $t = 0$ and at that time i $(1 \le i \le n)$ customers, 1, 2,..., i, are already waiting for service. The server attends to these i customers and all the new arriving customers as long as they come. If there are no more customers in the system, the server leaves the system unattended for a time interval of unit length before resuming his duty again. It is assumed that no customer joins the queue more than once, and that any customer who has not yet joined the queue until time $t = r - 1$ $(r = 1, 2,...)$ may join the queue in the time interval $(r - 1, r]$ with probability p $(0 < p < 1)$ independently of the other customers.

The distribution of $\rho_n(i)$ in the random graph $\Gamma_n(p)$ is the same as the distribution of the number of customers served in the initial busy period in the queueing process defined above.

Denote by ν_r $(r = 1, 2,...)$ the number of customers arriving at the counter in the time interval $(r - 1, r]$.

We have

$$P\{\nu_1 = j_1, \nu_2 = j_2,..., \nu_k = j_k\}$$

(54)

$$= \frac{(n - i)!}{j_1!\, j_2!... j_k!\, (n - i - j_1 -...- j_k)!}\, p^{j_1 +...+j_k}\, q^{k(n-i)-kj_1-(k-1)j_2-...-j_k}$$

for $j_1 +...+ j_k \le n - i$ and $k \ge 1$.

Theorem 6 . *The distribution of* $\rho_n(i)$ *is given by*

(55)
$$P\{\rho_n(i) = k\} = \binom{n-i}{k-i} p^{k-i} q^{k(n-k)} \phi_k^{(i)}(q)$$

for $1 \le i \le k \le n$ *where* $\phi_k^{(i)}(q)$ *is defined by* (33).

Proof. By (14) and (54) we have

$$P\{\rho_n(i) = k\} = \sum_{\substack{j_1+\dots+j_k = k-i \\ j_1+\dots+j_r > r-i (i \le r \le k)}} \frac{(n-i)!p^{j_1+\dots+j_k} q^{k(n-i) - kj_1 - (k-1) j_2 - \dots - j_k}}{j_1!\, j_2! \dots j_k!(n-i-j_1-\dots-j_k)!} = \tag{56}$$

$$= \sum_{\substack{j_1+\dots+j_k = k-i \\ j_1+\dots+j_r < r (1 \le r \le k-i)}} \frac{(n-i)!p^{k-i} q^{k(n-i) - j_1 - 2j_2 - \dots - kj_k}}{j_1!\, j_2! \dots j_k!(n-k)!}$$

for $1 \le i \le k \le n$. This proves (55).

We note that the probability that $\Gamma_n(p)$ is connected is obviously

$$P\{\rho_n(1) = n\} = p^{n-1} \phi_n^{(1)}(q) \tag{57}$$

for $n \ge 1$ where $\phi_n^{(1)}(q)$ is determined by (33). By (57) and (33) we have a convenient explicit formula for the probability that a random graph $\Gamma_n(p)$ is connected.

Theorem 7. *If* $n \to \infty$ *and* $p \to 0$ *in such a way that* $np \to a$ *where* $0 < a < \infty$, *then*

$$\lim_{np \to a} P\{\rho_n(i) = k\} = \frac{i}{k} \frac{e^{-ak} (ak)^{k-i}}{(k-i)!} \tag{58}$$

for $1 \le i \le k$.

Proof. *If* $p \to 0$, *then* $q \to 1$, *and by* (35)

$$\lim_{q \to 1} \phi_k^{(i)}(q) = \phi_k^{(i)}(1) = i\, k^{k-i-1}. \tag{59}$$

Furthermore,

$$\lim_{np \to a} q^{(k-1)(n-k)} = \lim_{n \to \infty} (1 - \frac{a}{n})^{(k-1)(n-k)} = e^{-a(k-1)}, \tag{60}$$

and by the Poisson limit theorem

$$\lim_{np \to a} \binom{n-i}{k-i} p^{k-i} q^{n-k} = e^{-a} a^{k-i}/(k-i)! \,. \tag{61}$$

By forming the product of (59), (60) and (61), we obtain the limit (58).

Theorem 7 confirms a conjecture of J.W. Kennedy [11] concerning the limit distribution of $\rho_n(1)$ as $np \to a$. Kennedy calculated the limit (58) for $i = 1$ and $k \leq 5$, and conjectured that (58) is true for $i = 1$ and all $k \geq 1$.

6. THE HEIGHT OF A RANDOM TREE

The number of unlabeled, rooted, plane (ordered) trees with $n + 1$ vertices is

$$C_n = \binom{2n}{n} \frac{1}{(n+1)} \tag{62}$$

for $n = 0, 1, 2, \ldots$. See F. Harary, G. Prins and W. Tutte [9], N.G. de Bruijn and B.J.M. Morselt [5], and D.A. Klarner [12].

Let us suppose that all the possible C_n trees are equally probable and choose a tree at random. Denote by $\sigma(n+1)$ the height of the tree. We are interested in finding the distribution of $\sigma(n+1)$. To find this distribution let us consider the queueing model discussed in Example 2 in Section 3. If the initial queue size is $i = 1$ and if the number of customers served in the initial busy period is $\rho(1) = n + 1$, then the associated random graph can be described as an unlabeled, rooted, plane (ordered) tree with $n + 1$ vertices. The height of the tree $\sigma(n+1) < k$ if and only if in the associated queueing process $\tau(1) \leq k$. Accordingly,

(63) $$P\{\sigma(n+1) < k\} = P\{\tau(1) \le k \text{ and } \rho(1) = n+1\}/P\{\rho(1) = n+1\}.$$

By (39) we have

(64) $$P\{\rho(1) = n+1\} = C_n\, p^n\, q^{n+1},$$

and by (42) and (48) we have

(65) $$P\{\tau(1) \le k \text{ and } \rho(1) = n+1\} = a_1(n+1, k)\, p^n\, q^{n+1} = p_1(n+1, k)$$

where $p_1(n+1, k)$ is given by (41). Consequently,

$$P\{\sigma(n+1) < k\} = \frac{(n+1)}{\binom{2n}{n}} \sum_j \left[\binom{2n}{n+j\,(k+1)} - \binom{2n}{n+j\,(k+1)+k}\right]$$

(66) $$= \frac{(n+1)\,2^{2n+1}}{(k+1)\binom{2n}{n}} \sum_{r=1}^{k} \left(\cos\frac{r\pi}{k+1}\right)^{2n} \left(\sin\frac{r\pi}{k+1}\right)^2$$

for $k = 1, 2, \ldots$.

From (66) we obtain that the limit distribution

(67) $$\lim_{n\to\infty} P\{\frac{\sigma(n)}{\sqrt{2n}} \le x\} = K^*(x)$$

exists. We have

(68) $$K^*(x) = \sum_{k=-\infty}^{\infty} (1 - 4k^2 x^2)\, e^{-2k^2x^2} = \frac{\sqrt{2}\,\pi^{5/2}}{x^3} \sum_{j=1}^{\infty} j^2\, e^{-j^2\pi^2/2x^2}$$

for $x > 0$ and $K^*(x) = 0$ for $x \le 0$. See also L. Takács [25].

The distribution of $\sigma(n)$ and the limit distribution of $\sigma(n) / (n)^{1/2}$ have been determined by I.V. Konovaltsev and E.P. Lipatov [14]. See also N.G. de Bruijn, D.E. Knuth and S.O. Rice [4].

Finally, we note that

$$k^*(x) = \frac{dK^*(x)}{dx} \tag{69}$$

exists for $x > 0$ and $k^*(x)$ satisfies the following equation

$$x^3 k^*(x) = \left(\frac{\pi}{2}\right)^{3/2} k^*\left(\frac{\pi}{2x}\right) \tag{70}$$

for $x > 0$.

REFERENCES

[1] André, D., Solution Directe du Problème Résolu par M. Bertrand, *C.R. Acad. Sci. Paris* 105 (1887) 436-437.

[2] Bertrand, J., Solution d'un Problème, *C.R. Acad. Sci. Paris* 105 (1887) 369.

[3] Bol, G., Über eine kombinatorische Frage, *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität* 12 (1938) 242-245.

[4] De Bruijn, N.G., D.E. Knuth and S.O. Rice, The Average Height of Planted Plane Trees, *Graph Theory and Computing*, Edited by R.C. Read, Academic Press, New York, 1972, 15-22.

[5] De Bruijn, N.G. and B.J.M. Morselt, A Note on Plane Trees, *Journal of Combinatorial Theory* 2 (1967) 27-43.

[6] Cayley, A., A Theorem on Trees, *Quarterly Journal of Pure and Applied Mathematics* 23 (1889) 376-378. [The Collected Mathematical Papers of Arthur Cayley, Vol. XIII, Cambridge University Press, 1897, 26-28.]

[7] Clarke, L.E., On Cayley's Formula for Counting Trees, *Journal of the London Mathematical Society* 33 (1958) 471-474.

[8] Dziobek, O., Eine Formel der Substitutionstheorie, *Sitzungsberichte der Berliner Mathematischen Gesellschaft* 17 (1917) 64-67.

[9] Harary, F., G. Prins and W. Tutte, The Number of Plane Trees, *Nederl. Akad. Wetensch. Proc.* Ser. A. 67 (1964) 319-329. [Indag. Math. 26 (1964) 319-329.]

[10] Kendall, D.G., Some Problems in the Theory of Queues, *Jour. Roy. Stat. Soc. Ser.* B 13 (1951) 151-185.

[11] Kennedy, J.W., Random Clumps, Graphs, and Polymer Solutions, *Theory and Applications of Graphs*, Edited by Y. Alavi and D.R. Lick, Lecture Notes in Mathematics, N. 642, Springer-Verlag, Berlin, 1978, 314-329.

[12] Klarner, D.A., A Correspondence Between Two Sets of Trees, *Nederl. Akad. Wetensch. Proc.* Ser. A 72 (1969) 292-296. [Indag. Math. 31 (1969) 292-296.]

[13] Kolchin,V.F., Random Mappings, Optimization Software Inc., New York, 1986, 150-151.

[14] Konovaltsev, I.V. and E.P. Lipatov, Some Properties of Plane Rooted Trees, *Cybernetics* 6 (1970) 660-667.

[15] Moon, J.W., Various Proofs of Cayley's Formula for Counting Trees, *A Seminar on Graph Theory*, Edited by F. Harary and L. Beinke. Holt, Rinehart and Winsten 1967, 70-78.

[16] Moon, J.W., Another Proof of Cayley's Formula for Counting Trees, *American Mathematical Monthly* 70 (1968) 846-847.

[17] Moon, J.W., A Tree Counting Problem, *The Annals of Mathematical Statistics* 39 (1968) 242-245.

[18] Moon, J.W., Counting Labelled Trees, *Canadian Mathematical Congress*, 1970, 26-27.

[19] Prüfer, H., Nuere Beweis eines Satzes über Permutationen, *Archiv der Mathematik und Physik* 27 (1918) 142-144.

[20] Rényi, A., Some Remarks on the Theory of Trees, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 4 (1959) 73-85.

[21] Rényi, A., On the Enumeration of Trees, *Combinatorial Structures and their Applications*, Edited by R. Guy, H. Hanani, N. Sauer and J. Schonheim, Gordon and Breach, New York, 1970, 355-360.

[22] Riordan,J., Forests of Labeled Trees, *Journal of Combinatorial Theory* 5 (1968) 90-103.

[23] Sachkov, V.N., *Combinatorial Methods of Discrete Mathematics*, (In Russian) Nauka, Moscow, 1977, 181-183.

[24] Takács, L., Investigation of Waiting Time Problems by Reduction to Markov Processes, *Acta Mathematica Acad. Sci. Hungaricae* 6 (1955) 101-129.

[25] Takács, L., Remarks on Random Walk Problems, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 2 (1957) 175-182.

[26] Takács, L., A Generalization of the Ballot Problem and its Application in the Theory of Queues, *Journal of the American Statistical Association* 57 (1962) 327-337.

[27] Takács, L., *Combinatorial Methods in the Theory of Stochastic Processes*, John Wiley and Sons, New York, 1967.