The Employee Party Problem

David W. Matula, Washington University, St. Louis

April, 1972

ABSTRACT

A set of n employees have intermingled such that each pair have met with (independent) probability p. The boss wishes to host a party for a maximum number of the n employees such that every pair of the employees invited have met, and wonders how big a party to plan. Let D_{np} be the random variable giving the size of the largest subset of employees every pair of which have met. We show that the density of D_{np} is quite spiked. For example, with $n = 1000$, $p = .5$, we exhibit bounds showing that $\text{Prob}\{D_{1000,.5} = 15\} > .8$. Our main result is the following.

<u>Theorem</u>: For any $0 < p = 1/b < 1$, $\varepsilon > 0$, let

$d = 2\log_b n - 2\log_b \log_b n + 2\log_b(e/2) + 1$. Then

$\lim_{n\to\infty} \text{Prob}\{\lfloor d-\varepsilon \rfloor \le D_{np} \le \lfloor d+\varepsilon \rfloor\} = 1$.

The Main Theorem

The employee party problem is equivalent to the problem of determining the order of the largest complete subgraph of a random graph. A random graph G_{np} is defined to have n vertices where each of the possible n(n-1)/2 edges occurs with probability p. The order of the largest complete subgraph of the random graph G_{np} is denoted by D_{np}. Thus D_{np} is a random variable with a discrete distribution over the integers 1 to n.

Considerable information about D_{np} can be gleaned from analysis of the random variable K_{dnp}, which gives the number of d-membered complete subgraphs of the random graph G_{np}. In the paper "On the Complete Subgraphs of a Random Graph" we provide proofs of the following formulas for the mean and standard deviation of K_{dnp}.

$$E(K_{dnp}) = \binom{n}{d} p^{d(d-1)/2}$$

$$SD(K_{dnp}) = \left\{ \bar{K} \sum_{j=\max\{0,2d-n\}}^{d} \binom{n-d}{d-j} \binom{d}{j} p^{[d(d-1)-j(j-1)]/2} - \bar{K}^2 \right\}^{\frac{1}{2}}$$

$$\text{where } \bar{K} = E(K_{dnp}) = \binom{n}{d} p^{d(d-1)/2}$$

In that paper the following bounds on the distribution of D_{np} are then derived.

$$\text{Prob}\{D_{np} \geq d\} \leq \binom{n}{d} p^{d(d-1)/2} \qquad (1)$$

$$\text{Prob}\{D_{np} \geq d\} \geq 1 / \sum_{j=\max\{0,2d-n\}}^{d} \frac{\binom{n-d}{d-j}\binom{d}{j}}{\binom{n}{d}} p^{-j(j-1)/2} \qquad (2)$$

Inequality (1) is elementary since

$$\mathrm{Prob}\{D_{np} \geq d\} = \mathrm{Prob}\{K_{dnp} \neq 0\} \leq E(K_{dnp}).$$

Inequality (2) follows from the fact that any non-negative valued random variable X with mean μ and standard deviation σ must have

$$\mathrm{Prob}\{X = 0\} \leq \frac{\sigma^2}{\mu^2+\sigma^2} .$$

For $E(K_{dnp}) = \binom{n}{d} p^{d(d-1)/2} << 1$, inequality (1) is readily seen to be quite tight. Numerical experience in the previously referenced paper with inequality (2) suggests that this lower bound is quite poor. Nevertheless it will now be shown that the inequalities (1) and (2) are adequate to prove a surprisingly sharp bound on the density of D_{np} as $n \to \infty$. In fact, as n grows, the density mass of D_{np} will become spiked to the extent that the value of D_{np} will behave almost like a deterministic property of the random graph G_{np}.

<u>Theorem:</u> For any $0 < p = 1/b < 1$, $\varepsilon > 0$, let

$$d = 2 \log_b n - 2 \log_b \log_b n + 2 \log_b (e/2) + 1. \qquad (3)$$

Then

$$\lim_{n\to\infty} \mathrm{Prob}\{\lfloor d-\varepsilon \rfloor \leq D_{np} \leq \lfloor d+\varepsilon \rfloor\} = 1.$$

<u>Proof:</u> For $1 < b < \infty$, $\varepsilon > 0$, let d be given by (3). It will first be shown that $\lim_{n\to\infty} \mathrm{Prob}\{D_{np} > \lfloor d+\varepsilon \rfloor\} = 0$. Let $Z(n) = \mathrm{Prob}\{D_{np} > \lfloor d+\varepsilon \rfloor\}$, and let $\delta = \lfloor d+\varepsilon \rfloor + 1 - d$. Note that $\varepsilon < \delta \leq 1+\varepsilon$ for any n.

For sufficiently large n, clearly $n \geq d + \delta \geq 1$. Thus using inequality (1),

$$Z(n) = \mathrm{Prob}\{D_{np} \geq d+\delta\}$$

$$\leq \binom{n}{d+\delta} p^{(d+\delta)(d+\delta-1)/2}$$

Now $\binom{n}{d+\delta} \leq n^{d+\delta}/(d+\delta)!$, and by Stirling's formula for sufficiently large n, $(d+\delta)! > d^d/e^d$. For sufficiently large n,

$$Z(n) < n^{d+\delta}\ e^d\ p^{[d(d-1)/2+\delta(d-1)+(\delta+\delta^2)/2]}/d^d$$

From (3) recalling that $p = 1/b$, $p^{d-1} = (2\ \log_b n)^2/e^2n^2$

and

$$p^{d(d-1)/2} = \frac{(2\ \log_b n)^d}{e^d\ n^d}\ .$$

Thus for sufficiently large n,

$$Z(n) < \frac{(2\ \log_b n)^{2\delta}}{n^\delta}\left(\frac{2\ \log_b n}{d}\right)^d$$

From (3)

$$\left(\frac{2\ \log_b n}{d}\right)^d < \left(1 + \frac{2\ \log_b \log_b n}{d}\right)^d$$

$$\leq e^{2\ \log_b \log_b n} = (\log_b n)^{2\ \log_b e}$$

Thus for sufficiently large n

$$Z(n) < 2^{2\delta}(\log_b n)^{2\ \log_b e+2\delta}/n^\delta .$$

Now $0 < \varepsilon < \delta < \varepsilon+1$, so for sufficiently large n,

$$Z(n) < 2^{2(\varepsilon+1)}(\log_b n)^{2\ \log_b e+2(\varepsilon+1)}/n^\varepsilon .$$

Thus for fixed ε and b with $1 < b$, $\varepsilon > 0$,

$$\lim_{n\to\infty} Z(n) = 0.$$

To complete the theorem it will now be shown that

$\lim_{n\to\infty} \text{Prob}\{D_{np} \geq \lfloor d-\varepsilon\rfloor\} = 1.$

Let $\alpha = d - \lfloor d - \epsilon \rfloor$, and note that $\epsilon \le \alpha < 1 + \epsilon$.

For sufficiently large n, $n > 2d$ and $d - \alpha \ge 2$, so then

$$\text{Prob}\,\{ D_{np} \ge d - \alpha \} \ge 1 \Big/ \sum_{j=0}^{d-\alpha} \frac{\binom{n-d+\alpha}{d-\alpha-j}\binom{d-\alpha}{j}}{\binom{n}{d-\alpha}}\, b^{j(j-1)/2}.$$

Note that

$$\sum_{j=0}^{d-\alpha} \frac{\binom{n-d+\alpha}{d-\alpha-j}\binom{d-\alpha}{j}}{\binom{n}{d-\alpha}} = 1.$$

Hence to show that $\lim_{n\to\infty} \text{Prob}\,\{ D_{np} \ge d-\alpha \} = 1$, it need only be shown that

$$\lim_{n\to\infty} \sum_{j=2}^{d-\alpha} \frac{\binom{n-d+\alpha}{d-\alpha-j}\binom{d-\alpha}{j}}{\binom{n}{d-\alpha}} \left(b^{j(j-1)/2} - 1 \right) = 0. \qquad (4)$$

For $2 \le j \le d - \alpha$, let

$$a_j = \frac{\binom{n-d+\alpha}{d-\alpha-j}\binom{d-\alpha}{j}}{\binom{n}{d-\alpha}} \left(b^{j(j-1)/2} - 1 \right). \qquad (5)$$

Equation (4) will be verified in three steps:

(i) $\lim_{n\to\infty} \sum_{j=2}^{\lfloor \frac{2}{3} d \rfloor} a_j = 0$,

(ii) for n sufficiently large, $a_{j-1} < a_j$ for $\lfloor \frac{2}{3} d \rfloor < j \le d - \alpha$,

(iii) $\lim_{n\to\infty} d\, a_{d-\alpha} = 0$.

From (5) for sufficiently large n,

$$a_j < \frac{(n-d+\alpha)^{d-\alpha-j}\,[(d-\alpha)!]^2}{(n-d+\alpha)^{d-\alpha}\,[(d-\alpha-j)!]^2\,j!}\,\ell^{j(j-1)/2}$$

$$< \frac{d^{2j}}{(n-d)^j}\,\ell^{j(j-1)/2}$$

$$< \left[\frac{2d^2\,\ell^{(j-1)/2}}{n}\right]^j$$

Now for $j \le \lfloor \frac{2}{3}d \rfloor$,

$$\ell^{(j-1)/2} < \left(\frac{en}{d}\right)^{\frac{j-1}{d-1}} < \left(\frac{en}{d}\right)^{\frac{2}{3}}.$$

Note that from (3)

$$\lim_{n\to\infty} d^k/n = 0 \qquad \text{for any } k. \qquad (6)$$

Hence for $2 \le j \le \lfloor \frac{2}{3}d \rfloor$, and n sufficiently large,

$$a_j < \left[\frac{2ed^2}{n^{1/3}}\right]^j < \frac{2ed^2}{n^{1/3}}$$

so

$$\sum_{j=2}^{\lfloor \frac{2}{3}d \rfloor} a_j < \frac{2ed^2}{n^{1/3}}$$

and

$$\lim_{n\to\infty} \sum_{j=2}^{\lfloor \frac{2}{3}d \rfloor} a_j = 0,$$

proving (i).

From (5) with $2 \le j \le d - \alpha$

$$\frac{a_{j-1}}{a_j} = \frac{(n - 2d + 2\alpha + j)(j)}{(d - \alpha - j + 1)^2} \frac{\left(b^{(j-2)(j-1)/2} - 1\right)}{\left(b^{j(j-1)/2} - 1\right)}$$

$$\le \frac{n\, j}{b^{j-1}}$$

For $j > \lfloor \frac{2}{3} d \rfloor$ and n sufficiently large,

$$b^{j} = \left(\frac{bn}{2 \log_b n}\right)^{\frac{2j}{d-1}} > \left(\frac{n}{d}\right)^{4/3}$$

and

$$\frac{a_{j-1}}{a_j} < \frac{j\, b\, d^{4/3}}{n^{1/3}} < 1,$$

thus verifying (ii).

Now from (5) with n sufficiently large

$$a_{d-\alpha} \le \frac{b^{(d-\alpha)(d-\alpha-1)/2}}{\binom{n}{d-\alpha}}$$

$$\le \frac{d^{d+1} \left[b^{(d-1)/2}\right]^{d-2\alpha} b^{(\alpha^2-\alpha)/2}}{e^d (n - d + \alpha)^{d-\alpha}}$$

$$\le \frac{d^{1+2\alpha}\, n^{d-2\alpha}\, b^{(\alpha^2-\alpha)/2}}{(n-d)^{d-\alpha}}$$

$$\le \frac{d^{1+2\alpha}\, b^{(\alpha^2-\alpha)/2}}{\left(1 - \frac{d}{n}\right)^{d-\alpha} n^{\alpha}} \le \frac{d^{1+2\alpha}\, b^{(\alpha^2-\alpha)/2}}{\left(1 - \frac{d^2}{n}\right) n^{\alpha}}$$

$$\le \frac{2\, d^{3+2\epsilon}\, b^{(1+\epsilon+\epsilon^2)/2}}{n^{\epsilon}}$$

Hence noting (6),

$$\lim_{n \to \infty} d\, a_{d-\alpha} = 0,$$

and the theorem is proved.