Comp. & Maths. with Appls., Vol. 5, pp. 33–39
Pergamon Press Ltd., 1979. Printed in Great Britain

EVOLUTION OF THE n-CUBE

PAUL ERDŐS and JOEL SPENCER
Nemetvolgyi ut 72c, Budapest

Dedicated to the memory of Yu. D. Burtin

(*Received August* 1978)

Abstract—Let C^n denote the graph with vertices $(\epsilon_1, \ldots, \epsilon_n)$, $\epsilon_i = 0, 1$ and vertices adjacent if they differ in exactly one coordinate. We call C^n the n-cube.

Let $G = G_{n,p}$ denote the random subgraph of C^n defined by letting

$$\text{Prob}\,(\{i, j\} \in G) = p$$

for all $i, j \in C^n$ and letting these probabilities be mutually independent. We wish to understand the "evolution" of G as a function of p. Section 1 consists of speculations, without proofs, involving this evolution. Set

$$f_n(p) = \text{Prof}\,(G_{n,p} \text{ is connected})$$

We show in Section 2:

Theorem

$$\operatorname*{Lim}_n f_n(p) = 0 \text{ if } p < 0.5$$
$$e^{-1} \text{ if } p = 0.5$$
$$1 \text{ if } p > 0.5.$$

The first and last parts were shown by Yu. Burtin[1]. For completeness, we show all three parts.

1. SPECULATIONS

We are guided by the fundamental results of A. Rényi and the senior author[2] on the evolution of random graphs. We think of p increasing (in time, perhaps) from $p = 0$ to $p = 1$ and $G_{n,p}$ evolving from the empty to the complete graph. Of course, G is not a particular graph but a random variable. We say that $p = p(n)$, $G = G_{n,p(n)}$ has a property Γ if

$$\operatorname*{Lim}_n \text{Prob}\,(G \text{ satisfies } \Gamma) = 1$$

and does not have property Γ if the above limit is zero. Erdős and Renyi noted that for many interesting monotone graph theoretical properties (e.g.; connectedness, planarity) there is a threshold function $f(n)$ so that if $p(n) = 0(f(n))$, G does not have Γ and if $f(n) = 0(p(n))$, G does have Γ. We say, informally, that property Γ appears at $p = f(n)$ if $f(n)$ is a threshold function for Γ.

At first, G consists of nonadjacent edges. Threshold functions for the appearance of small subgraphs are relatively easy to compute. For e fixed, connected subgraphs with e edges appear at $p \sim 2^{-n/e+0(n)}$: For such p the largest component has $(e+1)$ points and consists of a path of length e. We are most intrigued by the sizes of the components of G when p reaches $0(n^{-1})$.

Let $p = \lambda/n$, $\lambda < 1$. The degree of a point is approximately Poisson with mean λ. The component containing a fixed point resembles a Galton–Watson process. In each generation, each active member (point) spawns (is adjacent to) X new members where X is Poisson with mean λ. For $\lambda < 1$ the Galton–Watson process "dies" with probability one and the size of the component containing a given point is, in expectation, $(1-\lambda)^{-1}$. The size of the largest component is more difficult as one must consider 2^n not quite independent almost Galton–Watson processes.

With $\lambda > 1$ the nature of G changes dramatically. (This is the "double jump") of [12]). Now with probability $q(\lambda) > 0$ the Galton–Watson process does not stop. Then $(1 - q(\lambda))2^n$ points are in "small" components. What of the remainder? In particular, will there be a component with $(q(\lambda) + 0(1))2^n$ points? What is the size of the second largest component?

As λ increases the number of small components decrease. Perhaps there is a giant component at $\lambda \neq 1+\epsilon$ or perhaps the large components merge later. Somewhere between $p=(1+\epsilon)/n$ and $p=o(1)$ the medium size components disappear.

When p becomes constant, independent of n, there is one giant component and many small components of bounded size. As p increases the small components merge into the giant component until only isolated points remain unmerged. Total connectedness is achieved at $p=0.5$, as shown in the next section. There is a precise result:

Set $p=0.5+\epsilon/2n$

$$\lim_n \text{Prob}\,(G_{n,p} \text{ is connected}) = e^{-e^{-\epsilon}}.$$

2. CONNECTEDNESS

In this section we prove the Theorem stated in the introduction. Let $g_n(p)$ be probability that G contains isolated points. For $i \in C^n$ we define a random variable

$$X_i = 1 \text{ if } i \text{ is an isolated point of } G$$
$$0 \text{ if not}$$

$$\text{and set } X = \sum_{i \in C^n} X_i,$$

the number of isolated point of G. As each $i \in C^n$ has degree n in C^n

$$E(X_i) = (1-p)^n.$$

We set

$$\mu = 2^n(1-p)^n$$

so that, by linearity of expected value, $E(X)=\mu$. We calculate the second moment applying the formula

$$\text{Var}\,(X) = \sum_i \text{Var}\,(X_i) + \sum_{i \neq j} \text{Cov}\,(X_i, X_J)$$

with values

$$\text{Cov}\,(X_i, X_j) = 0 \text{ if } i,j \text{ not adjacent}$$
$$= \mu^2 p/(1-p) \text{ if } i,j \text{ adjacent}$$

so that

$$\text{Var}\,(X) = \mu + \mu(1-p)^n[(np/(1-p))-1].$$

For $p<0.5$ we apply Kolmogoroff's Inequality:

$$1-g_n(p) = \text{Prob}\,[X=0] \leq \text{Prob}\,[|X-\mu| \geq \mu]$$
$$\leq \text{Var}\,(X)/\mu^2.$$

From our second moment calculation we use only

$$\lim_n \text{Var}\,(X)/\mu^2 = 0.$$

As $f_n(p) \leq 1 - g_n(p)$

$$\operatorname*{Lim}_n f_n(p) = 0.$$

For $p > 0.5$

$$g_n(p) = \text{Prob}\,[X > 0] < E(X) = \mu$$

so

$$\operatorname*{Lim}_n g_n(p) = 0.$$

For $p = 0.5$ more care is required. Set

$$s_k(n) = \Sigma E(X_{i_1} \cdots X_{i_k})$$

summed over all sets $\{i, \ldots, i_k\} \subseteq C^n$. For fixed k the above sum has $\binom{2n}{k} \sim 2^{nk}/k!$ terms. When none of the $i_1, \ldots, i_k$ are the summand is precisely 2^{-nk}. There are at most $\binom{2n}{k-1} n(k-1)$ terms where some i_s, i_t are adjacent. There the summand lies beween 2^{-nd} and $2^{-nk+(k/2)}$ (actually less, as K_k is not a subgraph of C^n). Thus

$$\binom{2^n}{k} 2^{-nk} \leq s_k(n) \leq \binom{2^n}{k} 2^{-nk} + \binom{2^n}{k-1} n(k-1) 2^{-nk+(k/2)}$$

so

$$\operatorname*{Lim}_n s_k(n) = 1/k!$$

For any t, by Inclusion–Exclusion,

$$\text{Prob}\,[X = t] = s_t(n) - s_{t+1}(n) + \cdots$$

and, critically, the sum alternates about Prob $[X = t]$. Hence

$$\operatorname*{Lim}_n \text{Prob}\,[X = t] = e^{-1}/t!$$

(that is, X approaches a Poisson distribution with mean 1—as is to be expected as the X_i are nearly independent) so, in particular

$$\operatorname*{Lim}_n (1 - g_n(p)) = \operatorname*{Lim}_n \text{Prob}\,[X = 0] = e^{-1}.$$

Let $\mathscr{C}_s$ denote the family of connected sets $S \subsetneq C^n$, $|S| = s$ and

$$\mathscr{C} = \bigcup_{s=1}^{2^{n-1}} \mathscr{C}_s$$

For $s \in \mathscr{C}$ set

$$P(S) = \text{Prob}\,[S \text{ is a connected component of } G].$$

Set

$$b(S) = |\{u, v\} \in C^n \colon u \in S,\ v \notin S\}|,$$

the cardinality of the edge boundary of S. Clearly

$$P(S) \leq (1-p)^{b(S)} \leq 2^{-b(S)}$$

for $p \geq 0.5$. Our objective shall be to show

$$\operatorname*{Lim}_{n} \sum_{s \in \mathscr{C}} 2^{-b(S)} = 0. \tag{1}$$

Disconnected G without isolated points must contain a component $S \in \mathscr{C}$. Thus

$$0 \leq 1 - f_n(p) - g_n(p) \leq \sum_{S \in \mathscr{C}} P(S)$$

and hence (1) shall imply our Theorem. Set

$$g(s) = \sum_{S \in \mathscr{C}_s} 2^{-b(S)}. \tag{2}$$

We shall bound $g(s)$.

Hart[3] has found the minimal $b(S)$, $S \in \mathscr{C}_s$. It is achieved by letting

$$S = \{(\epsilon_1, \ldots, \epsilon_n);\ \sum_{i=1}^{n} \epsilon_i 2^{i-1} < s\}$$

In particular, if $s = 2^k$, S is a k-cube. In general

$$b(S) \geq s[n - \{\lg(s)\}] \tag{3}$$

($\lg$ = log base 2, $\{x\}$ = min integer $y \geq x$). (In [3] the problem stated is to find S with the maximal number of edges. By (5) the problems are equivalent.) We bound

$$|\mathscr{C}_s| \leq 2^n(n)(2n) \cdots ((s-1)n) \leq 2^n(ns)^s$$

as we may count ordered $(x_1, \ldots, x_s)$ each x_i adjacent to some previous x_j. Hence

$$g(s) \leq |\mathscr{C}_s|(\max 2^{-b(S)}) \leq 2^n(ns)^s 2^{-s(n-\{\lg s\})}$$

which is small for $2 \leq s \leq 2^{0.49n}$. (We may assume n is sufficiently large as our theorem concerns a limit in n.) For larger s set

$$s = 2^{n(1-\beta)}$$

and bound

$$|\mathscr{C}_s| \leq \binom{2^n}{s} \leq 2^{ns}/s! < (e2^{\beta n})^s, \tag{4}$$

bounding $s!$ by $(s/e)^s$. Equations (2), (3), (4) do not quite yield a small bound on $g(s)$ (if $p > 0.5$ they do and the proof is considerably simpler) so we require more detailed refinements.

Call $S \in \mathscr{C}_s$, $s = 2^{n(1-\beta)}$, *dense* if $b(S) \leq \beta sn + 10s$

Let $v(s)$ be the number of dense S. We shall bound $v(s)$. We assume $\beta \leq 0.51$ throughout. Fix $S \in \mathscr{C}_s$, dense. For $x \in S$ we define the degree of x,

$$d(x) = |\{y \in S: \{x, y\} \in C^n\}|$$

We call $n-d(x)$ the outdegree of x. Then $b(S)$ is (for any S) the sum of the outdegrees. That is

$$\sum_{x\in S} d(x)+b(S)=|S|n \tag{5}$$

so that, as S is dense,

$$\sum_{x\in S} d(x)\geq sn(1-\beta)-10s\geq 0.48sn.$$

As the average degree is $\geq 0.48n$ and the maximal degree is n, at least (0.48–0.1)/(1–0.1) of the points have degree $\geq 0.1n$. Set

$$T=\{x\in S\colon d(x)\geq 0.1n\} \text{ so } |T|>0.4s$$

(i.e.: a positive proportion of points have high degree.) For $U\subseteq S$ set

$$a(U)=\{x\in S\colon \{u,x\}\in \mathscr{C}^n \text{ for some } u\in U\},$$

the neighborhood of U in S. We now use the probabilistic method to find a small set U with a large number of neighbors. Let U be a random subset of S defined by

$$\text{Prob}\,[s\in U]=\alpha=(\ln n)/n$$

and requiring the events $s\in U$ to be mutually independent. For each $x\in T$

$\text{Prob}\,[x\notin a(U)]=(1-\alpha)^{d(x)}\leq(1-\alpha)^{0.1n}=o(1).$

Then

$$E(|a(U)|\geq E(|a(U)nT|)=\sum_{x\in T}\text{Prob}\,[x\in a(U)]\geq |T|(1-0(1))\geq 0.19s.$$

As $a(U)\leq s$ always, $|a(U)|\geq 0.1s$ with probability at least 0.0. As $|U|$ has binomial distribution $B(s,\alpha)$, $|U|\leq 2s\alpha$ with probability $1-0(1)$. Hence the above two events occur simultaneously with positive probability. That is, there exists a specific $U\subseteq S$ such that

$$\text{(i) } |U|\leq 2s\alpha$$
$$\text{(ii) } |s(U)|\geq 0.1s.$$

(Note the above statement is *not* a probability result. For all S such a U exists.) We set $u=2s\alpha=2s(\ln n)/n$ for convenience.

Now we bound $v(s)$. We count triples $(U, a(U), S-U-a(U))$ satisfying (i), (ii). There are at most $\left(\binom{2^n}{u}\right)$ choice for U. (Notation; $\left(\binom{m}{i}\right)=\sum_{j\leq i}\binom{m}{j}$.) There are (and this is the critical saving) at most 2^{nu} choices of $a(U)$ for, having chosen U, we select for each $x\in U$ the points of $a(U)$ adjacent to x in at most 2^n ways. Finally, there are at most $\left(\binom{2n}{0.9s}\right)$ choices of $S-U-a(U)$. Thus,

$$v(s)\leq\left(\binom{2^n}{u}\right)2^{nu}\left(\binom{2^n}{0.9s}\right)\leq 2^{2nu}\left(\binom{2^n}{0.9s}\right) \tag{6}$$

We split the sum (2) into dense and nondense S.

$$g(s)\leq v(s)2^{-s(n-\lceil\lg s\rceil)}+(|\mathscr{C}_s|-v(s))2^{-\beta sn-10s}. \tag{7}$$

By (4)

$$|\mathscr{C}_s|2^{-\beta sn-10s}<(e2^{-10})^s$$

is negligible. (This was why $\beta sn+10s$ was chosen as the cut off point for denseness.) The first summand of (7) is very small if $s\le c2^n/n$. (We omit the calculations.)

For $c2^n/n\le s\le 2^{n-1}$ we must further refine our methods. (Here we are considering the possibility that G consists of several large components.) Set $s=2^{n-\gamma}$, $1\le\gamma\le k\lg n$. $(\gamma=n\beta)$. As before $S\in\mathscr{C}_s$ is dense if $b(S)\le(\gamma+10)s$. Fix a dense S. The average outdegree is $\le\gamma+10$ so all but $0(s)$ points have outdegree $\le(\ln n)^2$. We set

$$R=\{x\in S\colon n-d(x)\le(\ln n)^2\}\text{ so }|S-R|=o(s)$$

and for $x\in S$ define a restricted degree

$$d'(x)=|\{y\in R;\{x,y\}\in C^n\}|.$$

Now

$$\sum_{x\in S}d'(x)=\sum_{y\in R}d(y)\ge|R|(n-(\ln n)^2)=sn(1-0(1))$$

so the average $d'(x)$ is $n(1-0(1))$, the maximum $d'(x)$ is n. Set

$$T'=\{x\in S\colon d'(x)\ge 0.1n\}.$$

Then

$$|S-T'|=o(s).$$

Let U be a random subset of R with independent probabilities

$$\text{Prob}\,[x\in U]=\alpha=(\ln n)/n.$$

On average, all but $o(s)$ points of S are adjacent to U. Thus there exists a triple $(U,a(U),S-U-a(U))$ where

(i) $|U|\le 2\alpha s=o(s)$.
(ii) all $x\in a(U)$ are adjacent to some $y\in U$.
(iii) $|S-U-a(U)|=o(s)$

and critically

(iv) $U\subseteq R$.

In counting triples there is now a critical savings with $a(U)$. For each $u\in U$ there are at most $n^{(\ln n)^2}$ choices (vs a factor of 2^n before) of the $x\in S$ adjacent to u—as there will be all but at most $(\ln n)^2$ of the neighbors of u in C^n. Thus (with $u=2s\alpha$ as before)

$$v(s)\le\left(\binom{2^n}{u}\right)n^{(\ln n)^2u}\left(\binom{2n}{o(s)}\right).\tag{8}$$

With this bound, $g(s)$ is small, $c2^n/n\le s\le 2^{n-1}$. Finally, one requires not only that all $g(s)$ are small but also their sum. This follows immediately from examining the arguments which yield exponentially small bounds on $g(s)$. Given that:

$$\operatorname*{Lim}_n \sum_{S\in\mathscr{C}} 2^{-b(S)} = \operatorname*{Lim}_n \sum_{s=2}^{2^{n-1}} g(s) = 0$$

completing our theorem.

REFERENCES

1. Yu. D. Burtin, On the probability of connectedness of a random subgraph of the n-cube, *Problemy pered. inf.* **13**, (Russian–English summary) (1977).
2. P. Erdös, A. Rényi, On the evolution of random graphs, *Mat. Kutato Int. Kozl* **5**, 17–60 (1960).
3. S. Hart, A note on the edges of the n-cube, *Discrete Math.* **14**, 157–163 (1976).