ISRAEL JOURNAL OF MATHEMATICS 83 (1993), 321–328

CONNECTIVITY AND DYNAMICS FOR RANDOM SUBGRAPHS OF THE DIRECTED CUBE

BY

BÉLA BOLLOBÁS

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, Cambridge CB2 1SB, U.K.

AND

CRAIG GOTSMAN

Department of Computer Science
Technion, Haifa 32000, Israel

AND

ELI SHAMIR*

Department of Computer Science
The Hebrew University of Jerusalem, Jerusalem 91904, Israel

ABSTRACT

Let $\alpha \in R$, $\epsilon = (\alpha + o(1))/n$ and $p = \frac{1}{2}(1+\epsilon)$. Denote by $\vec{Q}^n_p$ a random subgraph of the directed n-dimensional hypercube $\vec{Q}^n$, where each of the $n2^n$ directed edges is chosen independently with probability p. Then the probability that $\vec{Q}^n_p$ is strong-connected tends to $\exp\{-2\exp\{-\alpha\}\}$. The proof of this main result uses a double-randomization technique. Similar techniques may be employed to yield a simpler proof of the known analogous result for undirected random graphs on the cube.

The main result is applied to the analysis of the dynamic behavior of asynchronous binary networks. It implies that for almost all random binary networks with fixpoints, convergence to a fixpoint is guaranteed.

* Partially supported by grant #438/89 of the Israel Academy of Sciences.
Received March 26, 1992 and in revised form July 27, 1992

1. Introduction

The threshold probability for connectedness of random undirected graphs and behavior in the vicinity of the threshold is well understood, both for the standard random graph model $G_{n,p}$ [5] and for random subgraphs of the cube Q_p^n [3, 6, 1]. The strong-connectedness of random *directed graphs* has been investigated for the standard random graph model $\vec{G}_{n,p}$ [10], but not for random subgraphs of the directed cube $\vec{Q}_p^n$. For $\vec{G}_{n,p}$, Palásti [10] has shown that the threshold probability of strong-connectedness is the same as that for connectedness in $G_{n,p}$, namely $p = \log n/n$, and the hitting time of strong-connectedness coincides with the hitting time of all vertices having positive indegree and outdegree. In this paper, we prove the analogous result for $\vec{Q}_p^n$, namely that the threshold probability of strong-connectedness is $p = 1/2$ and the hitting time of strong-connectedness here also coincides with the hitting time of all vertices having positive indegree and outdegree. Behavior for values of p slightly below the threshold is investigated using a "double randomization" technique, in which the directed edges are selected in two stages. In the first stage, the edges are selected with probability slightly less than p. This causes the appearance of large strong-connected components and vertices with positive indegree or positive outdegree. In the second stage, edges are selected with probability complementing to p. This causes the large strong-connected components to fuse together to one enormous strong-connected component.

Beyond their theoretical value, our results have applications in the field of neural-network-type dynamical systems. In his study of networks of binary gates, Kauffman [9] introduced the following model: A collection of n binary gates are connected in a network such that each gate receives inputs from all others (including itself). Each gate has an internal binary state x_i and computes a specific n-input boolean function f_i. The **state** of the network is the binary vector $\mathbf{x} = (x_1, .., x_n) \in \{0,1\}^n$. Temporal dynamics of these networks can be observed by updating the network state in time by each gate computing its function with the current inputs:

$$x_i \leftarrow f_i(\mathbf{x})$$

Dynamics **regimes** can be either **synchronous** or **asynchronous**. In a synchronous regime (Kauffman's original model), the values of all the gates update simultaneously. In an asynchronous regime, investigated by Glauber [7], each gate updates independently of the others, subject to the single constraint that all processors update at the same average rate. Obviously, any dynamics in system lacking a central clocking mechanism must be asynchronous. When the update rate is not too rapid, a good approximation to the dynamics is one in which, at each time step, *only one* random gate updates. The state then moves to an adjacent vertex of the hypercube, unless (and until) it reaches a state from which there is no exit. In the sequel, we adopt this approximation when investigating asynchronous dynamics.

2. The Strong-Connectedness Threshold

Denote by Q_n the undirected cube and by $\vec{Q}^n$ the directed cube. A vertex $x \in \vec{Q}^n$ is a **source** if it has zero indegree and non-zero outdegree, a **sink** if it has zero outdegree and non-zero indegree, **semi-isolated** if it is a source or a sink and **isolated** if it has zero indegree and outdegree.

We begin with a series of lemmas. The first is a simple extension of the isoperimetric inequality for the vertex boundary in the cube (the case $l = 1$).

LEMMA 1: ([2] p. 129) *For every set $A \subset Q^n$ with $|A| \geq \sum_{i=0}^{r} \binom{n}{i}$, there are at least $\sum_{i=0}^{r+l} \binom{n}{i}$ vertices of Q^n within distance l of A.*

The next lemma is the standard isoperimetric inequality for the edge boundary in the cube.

LEMMA 2: ([2] p. 125) *For every set $A \subset Q^n$ with $|A| \geq k$, the number of edges connecting A and $Q^n - A$ is at least $k(n - \lceil \log_2 k \rceil)$.*

LEMMA 3: *In Q^n there are at most $(4en)^{k-1}$ trees of order k that contain a given vertex.*

Proof: Let T be a tree of order k in Q^n containing a given vertex v. Denote by m_j, $j = 1, .., k-1$ the number of vertices in T at distance j from v. Then $\sum_{j=1}^{k-1} m_j = k - 1$, where $m_{k-1} \geq 1$. The number of different trees containing precisely m_j vertices at distance j from v is not more than

$$\binom{n}{m_1}\binom{m_1(n-1)}{m_2}\cdots\binom{m_{k-2}(n-1)}{m_{k-1}}.$$

Since, in general,

$$\prod \binom{n_i}{l_i} \leq \binom{\sum n_i}{\sum l_i},$$

and summing over all partitions of $k-1$ into m_j gives a factor of $2^{2(k-1)}$ at most,

$$\begin{aligned}\#(\text{trees of order } k \text{ containing } v) &\le 2^{2(k-1)}\binom{n+(n-1)\sum_{j=1}^{k-2} m_j}{\sum_{j=1}^{k-1} m_j} \\ &\le 2^{2(k-1)}\binom{n+(n-1)(k-2)}{k-1} \\ &\le (4en)^{k-1}. \quad \blacksquare\end{aligned}$$

The next lemma maintains that if A and B are large enough subsets of Q^n, there are many short edge-disjoint paths connecting them.

LEMMA 4: *For every constant $c > 0$ there is a constant $c_0 > 0$ such that if A and B are subsets of Q^n, each containing at least $2^n/n^c$ vertices, and $l = c_0(\log_2 n)^{1/2}n^{1/2}$, then Q^n contains at least $2^{n-2l\log_2 n}$ edge disjoint $A-B$ paths, each having length at most $\lfloor l \rfloor$.*

Proof: In proving the lemma, we may, and shall assume that n is sufficiently large. Let $c_0 > 0$ be such that the resulting l and $r = \lceil n/2 - l/3 \rceil$ satisfy

$$2^n/n^c \ge \sum_{i=0}^{r}\binom{n}{i}.$$

Since, by a generous upper bound on the tail of the binomial distribution,

$$\sum_{i=r+l+1}^{n}\binom{n}{i} \le \frac{1}{2}\sum_{i=0}^{r}\binom{n}{i},$$

then, by Lemma 1, the set D of vertices at distance $\le l$ from A, contains at least

$$\begin{aligned}\sum_{i=0}^{r+l}\binom{n}{i} &= 2^n - \sum_{i=r+l+1}^{n}\binom{n}{i} \\ &\ge 2^n(1-1/2n^c)\end{aligned}$$

vertices. At least $2^n/2n^c$ of these vertices must be in B. Denote $B_0 = B \cap D$. By considering, for each $x \in B_0$, an appropriate shortest path to A, we can find a set of vertex-disjoint trees in Q^n, each tree rooted in A, having height at most l, such that these trees cover B_0. A tree whose root has degree d contains d paths from A to B, sharing only the root. Such a tree covers at most n^l vertices of B, hence there are at least $|B|/n^l$ edge-disjoint $A-B$ paths of length at most l. Finally, because $l \ge c$ for sufficiently large n,

$$|B|/n^l \ge 2^{n-2l\log_2 n}. \quad \blacksquare$$

We are now ready to state the central result of this paper.

THEOREM 1: *Let $p = \frac{1}{2}(1-\epsilon)$, where $0 < \epsilon < \log\log n/n$. Almost every $\vec{Q}_p^n$ is such that if x is not a sink and y is not a source, then $\vec{Q}_p^n$ contains an oriented path from x to y.*

Proof: We shall construct $\vec{Q}_p^n$ in two stages. Let $p_0 = \frac{1}{2}(1-\epsilon_0)$ be such that $(1-p_0)(1-1/n^2) = 1-p$. First we select edges with probability p_0 to obtain $\vec{Q}_{p_0}^n$ and then add edges independently with probability $1/n^2$ to obtain $\vec{Q}_p^n$. Note that, by definition, $(1+\epsilon_0)^n \leq 2\log n$.

Let $X_k = X_k(\vec{Q}_{p_0}^n)$ be the number of vertices from which exactly k vertices can be reached, and let $Y_k = Y_k(\vec{Q}_{p_0}^n)$ be the number of vertices that can be reached from exactly k vertices (through oriented paths in $\vec{Q}_{p_0}^n$). Using Lemmas 2 and 3, with $q_0 = 1 - p_0$ we have

$$\begin{aligned} E_k = \mathbf{E}(X_k) = \mathbf{E}(Y_k) &\leq 2^n(4en)^{k-1}p_0^{k-1}q_0^{k(n-\log_2 k)} \\ &\leq \left[\frac{4enk}{2}2^{n/k}q_0^n\right]^k . \end{aligned}$$

Denote $k_0 = \lfloor 2^n/n^3 \rfloor$. Since $k2^{n/k} \leq 2^n/n^2$ for $2 \leq k \leq k_0$ and $(2q_0)^n = (1+\epsilon_0)^n \leq 2\log n$, we see that, very crudely,

$$\sum_{k=2}^{k_0} E_k \leq 4e\sum_{k=2}^{k_0}(\log n/n)^k = o(1) .$$

In particular, by the Markov inequality, the event $[X_k = Y_k = 0 \text{ for all } 2 \leq k \leq k_0]$ occurs with probability at least $1 - o(1)$. Also, $E_1 = (1+\epsilon_0)^n \leq 2\log n$, so $X_1 \leq \log^2 n$ and $Y_1 \leq \log^2 n$ with probability at least $1 - 2/\log n$ (in fact, with considerably greater probability).

To complete the proof, it suffices to show that if $\vec{Q}_{p_0}^n$ is such that $X_1(\vec{Q}_{p_0}^n) \leq \log^2 n$, $Y_1(\vec{Q}_{p_0}^n) \leq \log^2 n$ and $X_k(\vec{Q}_{p_0}^n) = Y_k(\vec{Q}_{p_0}^n) = 0$ for all $2 \leq k \leq k_0$, then the graph $\vec{Q}_p^n$ obtained from $\vec{Q}_{p_0}^n$ by adding edges with probability $1/n^2$ has the property required by the theorem with probability $1 - o(1)$. Indeed, suppose that $\vec{Q}_{p_0}^n$ is such. Given $x, y \in \vec{Q}_{p_0}^n$, a source and sink respectively, what is the probability that the new edges creating $\vec{Q}_p^n$, added with probability $1/n^2$, do not ensure that $\vec{Q}_p^n$ contains a directed path from x to y ? Let A be the set of vertices that can be reached from x and let B be the set of vertices from which y can be reached. By the preceding discussion, $|A| \geq 2^n/n^3$ and $|B| \geq 2^n/n^3$, both with probability $1 - o(1)$, so Lemma 4 implies that, rather crudely, there are at least $m = 2^{n-n^{1/2}\log^2 n}$ edge-disjoint $A - B$ paths, each of length at most $l = \frac{1}{2}n^{1/2}\log n$. The probability that we do not select all edges of at least one of these paths is at most

$$\begin{aligned}(1 - n^{-2l})^m &\leq \exp\{-mn^{-2l}\} \\ &\leq \exp\{-2^{n-2n^{1/2}\log^2 n}\} \ .\end{aligned}$$

The probability that *some* such pair (x, y) will not be joined by a directed path in $\vec{Q}_p^n$ is at most

$$2^{2n}\exp\{-2^{n-2n^{1/2}\log n}\} < \exp\{-2^{n/2}\} \ .$$

Finally, what is the probability that there is an semi-isolated vertex $x \in \vec{Q}_{p_0}^n$ for which we add an edge $\vec{xy}$ or $\vec{yx}$ during the second randomization ? It is at most $\log^2 n[1 - (1 - 1/n^2)^{2n}] = O(\log^2 n/n) = o(1)$. ∎

Let us state two immediate corollaries of Theorem 1:

THEOREM 2: *Almost every random directed cube process is such that the hitting time of strong connectedness is precisely the hitting time of all vertices having positive indegree and outdegree.*

THEOREM 3: *Let $\alpha \in R$, $\epsilon = (\alpha + o(1))/n$ and $p = \frac{1}{2}(1+\epsilon)$. Then the probability that $\vec{Q}_p^n$ is strong connected tends to* $\exp\{-2\exp\{-\alpha\}\}$.

3. Applications to Binary Networks

Let $\{f_i : \{0,1\}^n \longrightarrow \{0,1\} : i = 1,..,n\}$ be n-input boolean functions. Denote $\mathbf{x} = (x_1,..,x_n)$ and $\mathbf{F} = (f_1,..,f_n)$. $\mathbf{F}$ induces a Markov chain on $\{0,1\}^n$ with the following transition probabilities:

$$\text{Prob}[(x_1,..,x_i,..,x_n) \longrightarrow (x_1,..,\bar{x}_i,..,x_n)] = 1/n \quad \text{iff } f_i(x_1,..,x_i,..,x_n) = \bar{x}_i \tag{1}$$

($\bar{x}_i$ is the binary complement of x_i). In general, a vertex $\mathbf{x}$ will have $d(\mathbf{x}, \mathbf{F}(\mathbf{x}))$ transitions out of it, where $d(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$. There is a probability $(1 - \frac{d(\mathbf{x},\mathbf{F}(\mathbf{x}))}{n})$ of staying in place. The **transition graph** of the Markov chain is the subgraph of $\vec{Q}^n$ where an edge appears if it has a non-zero transition probability. **Fixpoints** are vertices $\mathbf{x}$ for which $\mathbf{x} = \mathbf{F}(\mathbf{x})$, or equivalently, Prob[$\mathbf{x} \longrightarrow \mathbf{x}$] = 1. These fixpoints are also **attractors** of the network dynamics, in the sense that many dynamic trajectories terminate in these states. Denote by N_n a random element in the space of all possible binary networks of n gates (all different possibilities of gate functions). The number of fixpoints of N_n is asymptotically Poisson distributed with $\lambda = 1$, hence

$$\lim_{n \longrightarrow \infty} \text{Prob} \, [\, N_n \textit{ has at least one fixpoint }] = 1 - \exp\{-1\} \tag{2}$$

In his much-quoted paper on recursive *neural networks*, a special case of binary networks, where the gate functions are restricted to be *linear threshold* functions, Hopfield [8] showed that if the threshold weight matrix is symmetric, rapid convergence to fixpoints is guaranteed. Simulation results of Crisanti and Sompolinsky [4] indicate that convergence is guaranteed for *most* neural networks, even when the weight matrix is not symmetric. We now show that almost all N_n have this property.

THEOREM 4: *Say that N_n is* **convergent** *if there is a trajectory from any vertex of $\{0,1\}^n$ to a fixpoint. Then*

$$\lim_{n \longrightarrow \infty} \text{Prob} \, [\, N_n \textit{ is convergent} \mid N_n \textit{ has at least one fixpoint }] = 1 \tag{3}$$

Proof: The transition graph of N_n is a subgraph of $\vec{Q}^n$. It is easily seen that there is a one-to-one correspondence between $\vec{Q}^n_{\frac{1}{2}}$ and the class of all possible transition graphs of N_n. In the language of random graphs, a fixpoint is a sink. By Theorems 2 and 3, a typical transition graph of N_n consists of an enormous strong-connected component and a few sources and sinks. There are no isolated vertices. Theorem 1 shows explicitly that there exists a series of transitions between any vertex and fixpoint. ∎

4. Extensions

Very little has to be changed to obtain the analogues of Theorems 1–3 for undirected random subgraphs of the cube. The proofs obtained in that way are considerably simpler than the original proofs of these results [3, 6, 1]. Indeed, the analogue of Theorem 1 will state: Almost every Q^n_p is such that if $\deg(x_i) \geq 1, \quad i = 1, 2$, then Q^n_p contains a path from x_1 to x_2. The proof will state that the components of x_i in $Q^n_{p_0}$ are of size $2^n/n^3$, while a sprinkling of edges, with probability $1/n^2$, will fuse together these components (all claims with probability $1 - o(1)$).

References

[1] B. Bollobás, *The evolution of the cube*, in *Combinatorial Mathematics* (C. Berge, D. Bresson, P. Camion, J.F. Maurras and F. Sterboul, eds.), North-Holland, Amsterdam, 1983, pp. 91–97.

[2] B. Bollobás, *Combinatorics*, Cambridge University Press, 1987.

[3] Y.D. Burtin, *On the probability of connectedness of a random subgraph of the n-cube*, Problemy Pered. Inf. (In Russian) **13** (1977), 90–95.

[4] A. Crisanti and H. Sompolinsky, *Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics*, Physical Review **A37** (1988), 4865.

[5] P. ErdHos and A. Rényi, *On random graphs I*, Publ. Math. Debrecen **6** (1959), 290–297.

[6] P. ErdHos and J. Spencer, *Evolution of the n-cube*, Computers and Maths. with Applications **5** (1979), 33–39.

[7] R.J. Glauber, *Time-dependent statistics of the Ising model*, Journal of Math. Physics **4** (1963), 294.

[8] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA **79** (1982), 2554–2558.

[9] S. Kauffman, *Boolean systems, adaptive automata, evolution*, in *Disordered Systems and Biological Organization* (E. Bienenstock, ed.), Springer-Verlag, Berlin, 1986.

[10] E. Palásti, *On the strong-connectedness of directed random graphs*, Studia Scientiarum Mathematicarum Hungarica **1** (1966), 205–214.