RANDOM GRAPHS AND PROBABILITY MODELS

by

T. N. Bhargava and D. L. Fisk

TECHNICAL REPORT NO. 22

TO BE PRESENTED AT THE 36th SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE, SYDNEY, AUSTRALIA, AUGUST 28-SEPTEMBER 8, 1967

PREPARED UNDER RESEARCH GRANT NO. NsG-568
(PRINCIPAL INVESTIGATOR: T. N. BHARGAVA)

FOR

NATIONAL AERONAUTICS and SPACE ADMINISTRATION

Reproduction in whole or in part is permitted for any purpose of the United States Government.

DEPARTMENT OF MATHEMATICS

KENT STATE UNIVERSITY

KENT, OHIO

N68-25430
(ACCESSION NUMBER) (THRU) 1
12 (PAGES) (CODE) 19
NASA CR # 84674 (NASA CR OR TMX OR AD NUMBER) (CATEGORY)
FACILITY FORM 602

[redacted]

May 1967

Random Graphs and Probability Models*

T. N. Bhargava and D. L. Fisk**
Kent State University, U. S. A.

Abstract

In this paper we present some preliminary results from the theory of random graphs and digraphs, particularly in reference to their evolution as a time-dependent process. The main ideas come from the work of Erdös and Rényi, who describe a random graph as an independent trials stochastic process. The number, N, of edges in the graph is kept to be fixed, $o \leq N \leq \binom{n}{2}$. Evolution of a random digraph (directed graph) may similarly be described, e.g. Palasti's recent work on strong connectedness of random digraphs. Another approach is that of Gilbert, in which N itself is a random variable, and which is similarly carried over to digraphs. We study in detail some specific situations for the case when the number of diedges (directed edges) in a random digraph remain fixed by virtue of the fact that either the valencies (number of diedges starting from a point), or the densities (number of edges terminating at a point) or both remain fixed throughout the evolutionary process. The case when both the valencies and the densities are random variables gives rise to the varying number of diedges. A formal description of a multiple Markovian process for varying densities but fixed valencies, is also given with results for some simple cases.

---

*Prepared with the support of a NASA Research Grant No. NsG-568 at Kent State University, U. S. A.

**With Northwestern University until August 15, 196[redacted]

The aim of this report is to indicate some of the work which is being pursued with a view of apply graph-theoretic methods, in particular the notions of random graphs and digraphs, in applied probability situations. For most of our definitions, notations and concepts, we refer to Erdös and Rényi [3], and Bhargava [1]. A _graph_ consists of a set V of n labeled points (called _vertices_), and a set of N lines (called _edges_) joining none, some, or all of the pairs of points, i.e. $0 \leq N \leq \binom{n}{2}$; we denote such a graph by $G_{n,N}$. It is assumed that a graph has no parallel edges or loops. A _digraph_ (directed graph) consists of a set V or vertices, and a set E of _diedges_ (directed edges) such that $\Phi \subseteq E \subseteq V \times V - \{(i,i): \ i \in V\}$, where VxV denotes the cartesian product of set V with itself. Once again, we assume that a digraph has no multiple diedges and no loops, and denote a digraph with n vertices and N diedges by $D_{n,N}$ where N is the number of elements in the set E; clearly $0 \leq N \leq n(n-1)$. We remark that digraphs with loops are also of interest but their consideration is omitted here.

A random evolutionary graph may be defined in many different ways. For example, a very general way consists in choosing a random subset from the set, V, of vertices, and then another random subset which may be quite arbitrary; then we define a random mapping from the first chosen subset to the second. Next, we select a third subset according to a probability distribution which may depend on the first two choices of the subsets, and then define a random mapping between the second and the third subsets. This procedure may be continued as long as we want, each random mapping defining a class of edges so that at any particular stage we have a graph consisting of randomly chosen edges.

A simpler way of defining a random graph is to first pick a point at random from the set, V, of vertices, and then to move from this point to another with a certain probability distribution. This defines an edge. From the point arrived at, we then move to another point according to a probability distribution which depends on the previous points visited; the possibility of repeating points that have been visited before, or returning to the starting point is not excluded. The points may either be chosen independently in succession, or be chosen in a Markov sequence. However, such generalizations are not very fruitful, and sometimes not even necessary. A simple and useful approach is that of Erdös and Rényi [2] who essentially started the study of random graphs: A graph $G_{n,N}$ is said to be a <u>random graph</u> if it is chosen with same probability as any other member from the class of all graphs consisting of n vertices and N edges; this probability is $1/\binom{\binom{n}{2}}{N}$. The evolutionary process is described by looking at a random graph as a stochastic process. At time t = 1, one chooses an edge from the $\binom{n}{2}$ possible edges with probability $1/\binom{n}{2}$; at t = 2, a new edge is chosen with probability $1/\{\binom{n}{2}-1\}$; and so on.

In an obvious manner, one can define a <u>random digraph</u> (or a <u>directed random graph</u> as called in Palasti [5]). Another useful approach is that of Gilbert [4] in which the evolution is considered by assigning a probability p, $0 < p < 1$, of joining a pair of vertices by means of an edge; or equivalently by erasing with a probability $q = 1-p$, an edge between any pair of vertices from the complete graph (a graph which has all possible edges i.e. $\binom{n}{2}$). A comparison of Erdős-Rényi and Gilbert approaches is of interests in itself.

One may now ask various interesting questions about such graphs and digraphs. For example: What is the probability that a random graph $G_{n,N}$ (a random digraph $D_{n,N}$) is connected (strongly connected) or has a property P at a certain stage of the process? What is the probability distribution of the first occurrence of a connected graph (strongly connected digraph)? What is the probability that a graph (digraph) ever becomes connected (strongly connected)? These and some other questions have been considered by Erdős and Rényi [2], [3], and Palasti [5] in their very interesting papers.

In all of the schemes described above, the number, n, of vertices is kept fixed while the number, N, of edges (diedges) varies in time. A scheme for describing a random evolutionary digraph in which N is also kept fixed has been essentially described in some detail by Bhargava [1]. This, of necessity, changed the character of randomness of the digraph, but such an approach appears to be more useful in applications to various fields; for example, networks. The model in [1] is mainly used to analyze time changes in terms of the aggregate of subdigraphs of the given digraph, and presents techniques for doing so. In this section we present a formalization of the above type of random digraph, for fixed n and N, in terms of a digraph-valued Markov process. The number, N, of diedges may be taken as a random variable using a Gilbert type approach. We also remark that all the evolutionary processes described so far, with the exception of the Bhargava approach, are independent trials processes, and as such not applicable in dynamic studies.

Let $V = \{1,2,\ldots,n\}$ denote the set of n vertices of a given digraph D which is such that a vertex i of the digraph has valency c_i (the number of diedges starting from it), and density d_i (the number of diedges ending at it), $i = 1,2,\ldots,n$. Clearly $\sum_i c_i = \sum_i d_i = N$, say. There are four possible ways of describing the evolution of such a digraph: (a) Both c_i's and d_i's remain fixed through the time, (b) c_i's remain fixed but d_i's change in time, (c) d_i's remain fixed but c_i's change in time, (d) both c_i's and d_i's change in time. In the first three cases (a), (b), and (c) the number of edges remain fixed, whereas in the case (d) the number of edges is a time-dependent random variable. We consider the simple case (b) where the number of edges and the number of vertices remain fixed and the change in digraph takes place because of the change in the probability distribution of the c_i's i.e. of the random vector $(c_1, c_2, \ldots, c_n)$.

It now seems possible to describe, rather simply, the evolution of a random digraph as a formal stochastic process with all its standard problems of limiting stable distributions, statistical inference about the parameters involved, etc. In various practical situations a random digraph is found to be a reasonable probabilistic model under suitable conditions. We give a very general description of a digraph-valued Markov process. Because of the hopelessness of the general situation, we specialize our process to some specific situations.

Let $V = \{1,2,\ldots,n\}$ be a set of vertices, and let $\mathcal{E} = \{e_{ij} = \langle i,j\rangle \mid i \neq j, i,j \in V\}$ be the set of all ordered pairs of distinct points in V, i.e. $\mathcal{E}$ is the set of diedges; the number of elements in $\mathcal{E}$ is $n(n-1)$. Let $\mathcal{E}_i = \{e_{ij} \mid i \neq j, j \in V\}$, $i \in V$, be the set of all diedges starting from vertex i; the number of elements in $\mathcal{E}_i$ is $n-1$. If $2^{\mathcal{E}_i}$ denotes the class of all subsets of E_i, then there are 2^{n-1} elements in the class $2^{\mathcal{E}_i}$.

We denote by $G(V,\mathcal{E}) = \{\langle V,E\rangle \mid E \subset \mathcal{E}\}$ the class of all digraphs on set V; the number of such digraphs is $2^{n(n-1)}$, and any digraph $\langle V,E\rangle$ is completely described by the sequence $(\langle i,E_i\rangle)_{1 \leq i \leq n}$, where $E_i \subset \mathcal{E}_i$ for all $1 \leq i \leq n$.

Let $\Gamma(V,N) = \{<V,E> \varepsilon\ G(V,\mathcal{E}) | \#E = N\}$, where $\#E$ denotes the number of elements in the set E. Let $c_i = \#E_i$ be the valency of vertex i; then $\Gamma(V, \sum_{i=1}^{n} c_i) \supset \Gamma(V;c_1,\ldots,c_n)$ $= \{(<i,E_i>)_{1 \leq i \leq n} | \#E_i = c_i,\ 1 \leq i \leq n\}$.

One can now form a very general stochastic process where the outcome on the "t-th trial" is a digraph g_t in the class $G(V_t,\mathcal{E}_t)$. In this paper we formally describe a digraph-valued Markov process for a rather restrictive case, viz. when $g_t\ \varepsilon\ \Gamma(V;c_1(t),\ldots,c_n(t))$ where $\#V = n$ is fixed and the valencies $c_i(t)$ are random variables.

For each $t \geq 0$, $c_i(t)\ \varepsilon\ \{0,1,\ldots,n-1\}$, $1 \leq i \leq n$. Let $\mathcal{C}$ denote the set of n-tuples $\{\bar{i} = <i_1,\ldots,i_n>|i_j$ $\varepsilon\ \{0,1,\ldots,n-1\}\}$, so that $\#\mathcal{C} = n^n$. Let $\|\bar{i}\| = \sum_{j=1}^{n} i_j$, and $\mathcal{C}_N = \{\bar{i}\ \varepsilon\ C \big| \ \|\bar{i}\| = N\}$. Let $\bar{c}(t) = \{<c_1(t),\ldots,c_n(t)\}$; if $\|\bar{c}(t)\| = N$. We have a digraph on n vertices with N diedges, i.e. $C(t)\ \varepsilon\ \mathcal{C}_N$. We formally write

$$P(C(0) = \bar{j}) = p_{\bar{j}},\ \bar{j}\ \varepsilon\ \mathcal{C}.$$

$$P(C(t) = \bar{j} | C(t-1) = \bar{i}) = p_{\bar{i}\bar{j}}(t),\ \bar{i},\bar{j}\ \varepsilon\ \mathcal{C}$$

Thus $(C(t))_{t \geq 0}$ is a Markov chain with initial probability distribution $(p_{\bar{j}})_{\bar{j}\ \varepsilon\ \mathcal{C}}$, and transition probabilities $[p_{\bar{i},\bar{j}}(t)]_{\bar{i},\bar{j}\ \varepsilon\ \mathcal{C}}$ at time $t > 0$.

The process may now be described in the following way:

(i) Once we have $C(0) = <c_1(0),\ldots,c_n(0)>$ we select $c_i(0)$ diedges from $\mathcal{E}_i$ in an appropriate manner; for example, we may select $c_i(0)$ at random from $\mathcal{E}_i$ to form $E_i(0)$. We get the digraph $g_0 = (\langle i, E_i(0)\rangle)_{1 \le i \le n}$

(ii) Given $g_{t-1} = (\langle i, E_{i,t-1}\rangle)_{1 \le i \le n}$, we construct $g_t = (\langle i, E_{i,t}\rangle)_{1 \le i \le n}$ by first obtaining $C(t)$ according to its Markov transition law; then we construct the set $E_i(t)$, where $C(t) = \langle c_1(t), \ldots, c_n(t)\rangle$, and $\# E_i(t) = c_i(t)$. We now have random variables taking values in $\{0, 1, \ldots, c_i(t-1) \wedge c_i(t)\}$, and we choose $x_i(t)$ diedges from $E_i(t-1)$ and $c_i(t) - x_i(t)$ diedges from $\tilde{E}_i(t-1)$ where $\tilde{E}$ denotes complement of E. This is a simple device to weigh the diedges to be chosen for $E_i(t)$ according to whether or not they appeared in $E_i(t-1)$. Then

$$P(E_i(t) = E_i | E_i(t-1), c_i(t), X_i(t))$$

$$= \frac{\binom{c_i(t-1)}{X_i(t)} \binom{n-1-c_i(t-1)}{c_i(t)-X_i(t)}}{\binom{n-1}{c_i(t)}}$$

Each $\langle i, E_i(t)\rangle$, $1 \le i \le n$ is chosen independently according to the procedure outlined above.

Let $X(t) = \langle X_1(t), \ldots, X_n(T)\rangle$. Then

$$P(g_t = g | g_{t-1}, C(t), X(t))$$

$$= \prod_{i=1}^{n} \left(\frac{\binom{c_i(t-1)}{X_i(t)} \binom{(n-1)-c_i(t-1)}{c_i(t)-X_i(t)}}{\binom{n-1}{c_i(t)}} \right)$$

where $g_t = \langle V, E(t)\rangle = (\langle i, E_i(t)\rangle)_{1 \le i \le n}$

$g_{t-1} = \langle V, E(t-1)\rangle = (\langle i, E_i(t-1)\rangle)_{1 \le i \le n}$

If $C(t) \in C_N$, then $\|C(t)\| = N$, and $\#E(t) = N$, $g_t \in \Gamma(V,N)$ for every $t \geq 0$. This requires possible representations of N as a sum of n of the integers $\{0,1,\ldots,n-1\}$. For let $C(t) = \langle i_1,\ldots,i_n \rangle$, $\|C(t)\| = \sum_{j=1}^{n} i_j = N$, then the number of elements in C_N, which is the same as the possible values for $C(t)$, is determined by the number of representation of N as a sum of n integers from $\{0,1,\ldots,n-1\}$.

Finally, we consider a simpler case where we assume $X_i(t)$ to be same for each $i = 1,2,\ldots,n$, and that each $X(t)$ has a binomial distribution $b(c_i(t-1) \wedge c_i(t),\lambda)$. By further assuming that $C(t) \equiv \langle c,c,\ldots,c \rangle$ so that $g_t \in \Gamma(V;c,\ldots,c)$, and $X_i(t)$ is $b(c,\lambda)$ for $1 \leq i \leq n$, $t \geq 0$, we get

$$P(g_t|g_{t-1}) = P((E_i(t))_{1 \leq i \leq n}|(E_i(t-1))_{1 \leq i \leq n})$$

$$= \prod_{i=1}^{n} \frac{\binom{c}{k_i}\binom{n-1-c}{d-k_i}}{\binom{n-1}{c}} \binom{c}{k_i} \lambda^{k_i}(1-\lambda)^{k_i}$$

where $k_i = \#(E_i(t) \cap E_i(t-1))$.

For $c = 1$, this reduces to

$$P(g_t|g_{t-1}) = \frac{(n-2)^k}{(n-1)^n} \lambda^k (1-\lambda)^{n-k}$$

where k is the number of i's for which $E_i(t) = E_i(t-1)$.

References

[1] T. N. Bhargava, "On time changes in a digraph.", Jr. of Appl. Prob., 2 (1965), 101-118.

[2] P. Erdös and A. Rényi, "On random graphs, I", Publ. Math. Debrecen, 6 (1959), 290-297.

[3] P. Erdös and A. Rényi, "On the evolution of random graphs", Publ. Math. Inst. Hung. Acad. Sci., 5 (1960), 17-61.

[4] E. N. Gilbert, "Random graphs", Ann. Math. Statist., 30 (1959), 1141-1144.

[5] I. Palasti, "On the strong connectedness of directed random graphs", Stu. Sci. Math. Hung., 1 (1966), 205-214.

Abstrait

Dans ce papier nous présentons des résultats préliminaires de la théorie de"graphes" et de "digraphes," particulièrment à l'égard de leur évolution comme procédés "temps-dépendant." Les idées principales viennent du travail d'Erdos et de Renyi, qui décrivent un graphe choisi au hasard comme un procédé épreuve-stochastique. Le chiffre n aux bords du graph reste fixé $0 \leq N \leq \binom{n}{2}$. L'évolution d'un digraphe (courbe dirigée) peut être décrit pareillement, par exemple, l'ouvrage récent de Palasti sur l'affinité forte entre des graphes choisis au hasard.

Autre approche est celle de Gilbert où "N" même, est un variable choisi au hasard et qui est pareillement rapporté aux digraphes. Nous étudions en détail des situations spécifiques pour le cas où le nombre de "diedges"(bords dirigés) dans un digraph reste fixe à force du fait que ou les valences (nombre de (diedges) portant d'un point), ou les densités (nombre de diedges qui se terminent à un point), ou les deux restent fixés a travers le procédé évolutionnaire. Le cas où tous les deux valences et les densités sont des variables donne naissance au nombre variable de "diedges." Une description formelle d'un multiple procédé Markovien pour des densités variables mais des valences fixés est aussi donnée avec les résultats de quelques cas simples.