Alcom-FT Technical Report Series
ALCOMFT-TR-01-63

International Journal of Foundations of Computer Science

CONNECTIVITY PROPERTIES IN RANDOM REGULAR GRAPHS WITH EDGE FAULTS

S. NIKOLETSEAS
Computer Technology Institute, Patras University, Greece

K. PALEM
Courant Institute, New York University

P. SPIRAKIS
Computer Technology Institute, Patras University, Greece

M. YUNG
CertCo NYNY, Columbia University

Received (received date)
Revised (revised date)
Communicated by Editor's name

ABSTRACT

We introduce a new model of random graphs, that of random regular graphs with edge faults (which we denote by $G^r_{n,p}$), obtained by selecting the edges of a random member of the set of all regular graphs of degree r independently and with probability p. We can thus represent a communication network in which the links fail independently and with probability $f = 1 - p$.
In order to deal with this new model, we extend the notion of configurations and the translation lemma between configurations and random regular graphs provided by B. Bollobás, by introducing the concept of random configurations, to account for edge faults, and by providing an extended translation lemma between random configurations and $G^r_{n,p}$ graphs.
We investigate important connectivity properties of $G^r_{n,p}$ by estimating the ranges of r, f for which, with high probability, $G^r_{n,p}$ graphs a) are highly connected b) become disconnected and c) admit a giant connected component of small diameter.

Keywords: random graphs, graph connectivity, giant connected component, reliable network computing

1. Introduction

In recent years the development and use of distributed systems and communication networks has increased dramatically. In addition, state-of-the-art multiprocessor architectures compute over structured, regular interconnection networks. In such environments, several applications may share the same network while executing concurrently. This may lead to unavailability of certain network resources (e.g. links) for certain applications. Similarly, faults may cause unavailability of links or nodes. The aspect of *reliable distributed computing* (which means computing with the available resources and resisting faults) adds value to applications developed in such environments.

When computing in the presence of faults, one cannot assume that the actual structure of the computing environment is known. Faults may happen even in execution time. In addition, what is a "faulty" or "unavailable" link for one application may in fact be the de-allocation of that link because it is assigned (e.g. by the network operation system) to another application. The problem of analyzing allocated computation or communication in a network over a *randomly assigned subnetwork* and *in the presence of faults* has a nature different from fault analysis of special, well-structured networks (e.g. hypercube ([15, 27]) or the faulty PRAM models ([17, 18, 19])), which do not deal with network aspects. Our work addresses this interesting issue, i.e. analyzing the average case taken over a set of possible topologies. Such analyses have been suggested in the past (see e.g. [8, 13, 20]) and still motivate further research (see e.g. [12, 23, 24, 25]). Our present work focuses on multiconnectivity and existence of giant component properties, required for reliable distributed computing in such randomly allocated unreliable environments.

2. Definitions, State of the Art and Summary of Results

Let H be an undirected graph. A *random graph of type*$-H$ is obtained by selecting edges of H independently and with probability p. Let $f = 1 - p$ be called the *link failure probability*. The connectivity properties of random graphs of type$-H$ were studied in [2, 3, 11, 21, 26], among others. Most of these works relate the spanning trees of H to the probability that a random graph of type$-H$ is connected.

We study here a simple but fundamental type of H: H being a *random* member of the class of all regular graphs of n nodes and degree r, leading to *random regular graphs with edge faults*, a new class which we call $G^r_{n,p}$. In the model $G^r_{n,p}$, the random selection of a regular graph is capturing the *random allocation of a subnetwork* to a computation (e.g. by the operating system) and the edge failure probability $f = 1 - p$ is capturing the link faults that may happen.

An explicit definition of the new probability space $G^r_{n,p}$ follows.

Definition 1 (the $G^r_{n,p}$ probability space) *Let G^r_n be the probability space of all random regular graphs with n vertices where the degree of each vertex is r. The probability space $G^r_{n,p}$ of random regular graphs with edge faults is constructed by the following two subsequent random experiments: first, we choose a random regular graph from the space G^r_n and, second, we randomly and independently delete each edge from this graph with probability $f = 1 - p$.*

Clearly, the probability of two random regular graphs with edge faults with the same number of edges need not be the same in this case.

We will also frequently use in this paper the terms "almost certainly" (a.c.) and "with high probability" (w.h.p.) with their standard meaning for random graph properties. A property defined in a random graph holds almost certainly when its probability tends to 1 as the independent variable (usually the number of vertices in the graph) tends to infinity. "With high probability" means that the probability of the property of the random graph (or the success probability of a randomized algorithm) is at least $1 - n^{-\alpha}$, where $\alpha > 0$ is a constant and n is the number of vertices in the graph.

The $G_{n,p}$ model has been extensively studied (see e.g. [6]) and continues to motivate research in many fields. The theory of random graphs began with the celebrated works of Erdös and Renyi ([10]) in 1960 and by now researchers know a lot about the probable structure of these objects (see e.g. the paper on the birth of the giant component of [16] and, for an excellent review of various probabilistic techniques employed in the analysis of random graphs, see [1]). We also note that in [7] the threshold for $k-$connectivity of $G_{n,p}$ has been studied, but this result does not address the question of *the length* of the disjoint paths. This question has been recently investigated in [23], where the necessary probability p so that there almost certainly exist (at least) x vertex disjoint paths each of length (at most) $2l$ between any pair of vertices has been discussed. This work also presented a preliminary form of some of the results developed here. We also refer to the recent work of Flajolet et al ([12]) where the existence of edge-disjoint paths of given length between pairs of vertices in $G_{n,p}$ random graphs has been studied.

In order to deal with this new model $G^r_{n,p}$, we extend the notion of configurations and the translation lemma between configurations and random regular graphs provided by B. Bollobás ([4, 6]), by introducing here the concept of *random configurations* to account for edge faults, and by also providing an *extended translation lemma* between random configurations and random regular graphs with edge faults.

For this new model of random regular graphs with edge faults we show that:

1. For all failure probabilities $f = 1 - p \leq n^{-\epsilon}$ ($\epsilon \geq \frac{3}{2r}$ fixed) and any $r \geq 3$ the biggest part of $G^r_{n,p}$ (i.e. the whole graph except of $O(1)$ vertices) remains connected and this connected part (excluding the vertices that were originally neighbours of the $O(1)$-sized disconnected set) can not be separated, almost certainly, unless we remove more than r vertices. Note interestingly that the situation for this range of f and r is very similar, despite the faults, to the properties of G^r_n for $r \geq 3$ (which is r-connected).
2. $G^r_{n,p}$ is *disconnected* a.c. for constant f and any $r = o(\log n)$, but is *highly connected*, almost certainly, when $r \geq \alpha \log n$, where $\alpha > 0$ an appropriate constant.
3. Even when $G^r_{n,p}$ becomes disconnected, it still has *a giant component of small diameter*, even when $r = O(1)$. We provide an $O(n \log n)$ algorithm to construct a giant component.

We note the following important application of our work: multitasking in distributed memory multiprocessors is usually performed by assigning an arbitrary subnetwork (of the interconnection network) to each task (called the *computation graph*). Each parallel program may then be expressed as communicating processors over the computation graph. Note that a multiconnectivity value k of the computation graph means also that the execution of the application can tolerate up to $k-1$ *on-line additional faults.*

Recently, Andreas Goerdt ([14]) continued the work presented in a preliminary version of this paper ([23]) and showed the following results: if the degree r is fixed then $p=\frac{1}{r-1}$ is a threshold probability for the existence of a linear sized component in the faulty version of almost all random regular graphs. In fact, he further shows that if each edge of an *arbitrary* graph G with maximum degree bounded above by r is present with probability $p=\frac{\lambda}{r-1}$, when $\lambda<1$, then the faulty version of G has only components whose size is at most logarithmic in the number of nodes, with high probability. His result implies some kind of optimality of random regular graphs with edge faults.

3. Multiconnectivity Properties of $G^r_{n,p}$

Let G^r_n be the class of all random regular graphs on n vertices of degree r. It is known (see e.g. [6]) that G^r_n is a.c. $r-$connected for $r\geq 3$. In the following subsection we show the impressive effect of constant edge faults in connectedness, since we prove that random regular graphs with faults become, in the case of small degrees, disconnected, with high probability.

3.1. $G^r_{n,p}$ becomes disconnected in the case of small degrees and constant edge failure probability

We note that it is not as easy (from the technical point of view) as in the $G_{n,p}$ case to argue about random regular graphs, because of the stochastic dependencies on the existence of the edges due to regularity. The following notion of *configurations* was introduced by B. Bollobás ([4, 6]) to translate statements for random regular graphs to statements for the corresponding configurations which avoid the edge dependencies due to regularity and thus are much easier to deal with:

Definition 2 (Bollobás, [4]) *Let $w=\cup_{j=1}^{n}w_j$ be a fixed set of $2m=\sum_{j=1}^{n}d_j$ labeled vertices where $|w_j|=d_j$. A configuration F is a partition of w into m pairs of vertices, called edges of F.*

Given a configuration F, let $\theta(F)$ be the (multi)graph with vertex set V in which (i,j) is an edge if and only if F has a pair (edge) with one element in w_i and the other in w_j.

Note that every regular graph $G\in G^r_n$ is of the form $\theta(F)$ for exactly $(r!)^n$ configurations. However not every configuration F with $d_j=r$ for all j corresponds to a $G\in G^r_n$ since F may have an edge entirely in some w_j or parallel edges joining w_i and w_j.

Let ϕ be the set of all configurations F and let G^r_n be the set of all regular graphs. Given a property (set) $Q\subseteq G^r_n$ let $Q^*\subseteq\phi$ such that $Q^*\cap\theta^{-1}(G^r_n)=\theta^{-1}(Q)$.

In order to identify the set $\{\theta^{-1}(G):G\in G^r_n\}$, we consider possible cycles of length one (self-loops) and two (loops) among pairs w_i,w_j in $\theta(F)$. Let X_1,X_2 be the numbers of self-loops, loops in $\theta(F)$, respectively.

Clearly, $\theta^{-1}(G_n^r) = \{F : X_1 = X_2 = 0\}$. Thus, we can estimate the probability of a property Q, from the corresponding set of configurations Q^* such that $Q^* \cap \theta^{-1}(G_n^r) = \theta^{-1}(Q)$, by using the fact that:

$$\Pr\{Q\} = \Pr\{Q^* | X_1 = X_2 = 0\} \geq 1 - \frac{1 - \Pr\{Q^*\}}{\Pr\{X_1 = X_2 = 0\}} \qquad (1)$$

The distributions of X_1, X_2 are asymptotically independent Poisson distributions, with mean $E(X_i) = \frac{\lambda^i}{2i}$, $i = 1, 2$ where

$$\lambda = \frac{1}{m} \sum_{i=1}^{n} \binom{d_i}{2} = r - 1$$

thus

$$\Pr\{X_1 = X_2 = 0\} \sim e^{-\frac{\lambda}{2} - \frac{\lambda^2}{4}}$$

So, we get the following important lemma:

Lemma 1 (Bollobás, [6]) *If $r \geq 2$ is fixed and Q^* holds for a.e. configuration, then Q holds for a.e. $r-$regular graph.*

□

The main importance of the above lemma is that when we study random regular graphs whose degree does not grow too fast with n, instead of considering the set of all random regular graphs, we can study the (much more easier to deal with) set of configurations.

In order to deal with edge failures, we introduce here the following extension of the notion of configurations:

Definition 3 (random configurations) *Let $w = \cup_{j=1}^{n} w_j$ be a fixed set of $2m = \sum_{j=1}^{n} d_j$ labeled "vertices" where $|w_j| = d_j$. Let F be any configuration of the set ϕ. For each edge of F, remove it with probability $1 - p$, independently. Let $\hat{\phi}$ be the new set of objects and $\hat{F}$ the outcome of the experiment. $\hat{F}$ is called a* random configuration.

By introducing probability p in every edge, direct translation of the proof of Lemma 1 leads (since in both $\bar{Q}$ and $\hat{Q}$ each edge has the same probability and independence to be deleted, thus the modified spaces follow the properties of Q and Q^*) to the following extension to random configurations.

Lemma 2 (extended translation lemma) *Let $r \geq 2$ fixed and $\bar{Q}$ be a property for $G_{n,p}^r$ graphs. If $\hat{Q}$ holds for a.e. random configuration, then the corresponding property $\bar{Q}$ holds for a.e. graph in $G_{n,p}^r$.* □

Important Remark: We note that, *for monotone increasing properties*, results proved using the above extended translation lemma for $r \geq 2$ fixed, *obviously extend to greater (unbounded)* r, since (because of the monotonicity) increasing r can only strengthen the probability of such properties. In this work, we use this remark for the monotone properties a) of graph connectivity and b) of the existence of a giant connected component in a graph.

Using the notion of random configurations, we can now prove the following theorem:

Theorem 1 *When* $2 \leq r \leq \frac{\sqrt{\log n}}{2}$ *and* $p = \Theta(1)$ *then* $G^r_{n,p}$ *has at least one isolated node with probability at least* $1 - n^{-k}, k \geq 2$.

Proof. Consider the random configuration $\hat{F}$ corresponding to $G^r_{n,p}$. In the configuration F (out of which $\hat{F}$ was produced) there exist $n/(r+1)$ "vertices" that do not share an "edge". The probability that each particular one of them (say v) is isolated in $\hat{F}$ is:

$$\Pr\left\{v \text{ isolated in } \hat{F}\right\} = (1-p)^r$$

With respect to the above event, the $n/(r+1)$ vertices define $n/(r+1)$ *independent* experiments. Let $\hat{E}$ the event "there exists at least one isolated vertex in $\hat{F}$ out of the $\frac{n}{r+1}$ vertices". Thus we have:

$$\Pr\left\{\hat{E}\right\} = 1 - (1-(1-p)^r)^{\frac{n}{r+1}}$$

Since for all $x \in R, 1 - x \leq e^{-x}$ we get:

$$\Pr\left\{\hat{E}\right\} \geq 1 - \exp\left(-\frac{n}{r+1}(1-p)^r\right)$$

But for $r \leq \frac{\sqrt{\log n}}{2}$ and by setting $f = 1-p$, we get:

$$(1-p)^r \geq (1-p)^{\frac{\sqrt{\log n}}{2}} \geq n^{\frac{\log f}{2\sqrt{\log n}}} = n^{-\frac{\log(1/f)}{2\sqrt{\log n}}}$$

Thus:

$$\Pr\left\{\hat{E}\right\} \geq 1 - \exp\left(-\frac{n^{1-\frac{\log(1/f)}{2\sqrt{\log n}}}}{0.5\sqrt{\log n}+1}\right) \geq 1 - n^{-c}$$

for any constant $c > 0$. Thus, by the remark above, if $\bar{E}$ is the event that $G^r_{n,p}$ has an isolated point, then:

$$\Pr\left\{\bar{E}\right\} \geq 1 - \frac{1-\Pr\{\hat{E}\}}{\Pr\{\hat{X}_1 = \hat{X}_2 = 0\}} \geq 1 - \frac{n^{-c}}{\Pr\{X_1 = X_2 = 0\}}$$

But $\Pr\{X_1 = X_2 = 0\} \sim e^{-\frac{\lambda}{2}-\frac{\lambda^2}{4}}$. Since $\lambda = r-1$, we get $e^{-\frac{\lambda}{2}-\frac{\lambda^2}{4}} \sim n^{-\frac{1}{16}}$. Thus, $\Pr\{\bar{E}\} \geq 1 - n^{-k}$, for $k \geq c - \frac{1}{16}$. The proof is completed by choosing $c \geq 2 + \frac{1}{16}$. $\square$

In the next subsection we study the case of constant link failure probability f, which represents a worst case for connectivity preservation. Still, we are able to show that logarithmic degrees suffice to guarantee that $G^r_{n,p}$ remains w.h.p. highly connected, despite these constant edge failures.

3.2. Large degree random regular graphs are robust even for constant edge failure probability

Let $r \geq \alpha \log n$, for α being a suitably chosen constant, and let the probability of edge failures f be constant (w.l.o.g. let $f = 1/2$).

Then, for any particular vertex v, whose degree after the failures is denoted by $deg(v)$, we have:

$$\Pr\{(1-\beta)rp \leq deg(v) \leq (1+\beta)rp\} \geq 1 - 2e^{-\frac{\beta^2}{2}rp} \geq 1 - 2n^{-\gamma}$$

where $\gamma \geq 2$ a constant, by Chernoff bounds, and by choosing $\alpha \geq 2\gamma/(\beta^2 p)$. Clearly then:

$$\Pr\{\text{for all } v, (1-\beta)rp \leq deg(v) \leq (1+\beta)rp\} \geq 1 - 2n^{-(\gamma-1)}$$

We note that, interestingly, this property is shared by the graphs $G_{n,p'}$ for $p' \geq (\alpha p \log n)/n$, by the same technical argument. Indeed, remark that in $G^r_{n,p}$ graphs, the neighbors of any vertex are determined by first selecting r edges at random among all $n-1$ possible, and by then sampling them randomly and independently with the same probability p. In $G_{n,p'}$ graphs, all $n-1$ possible adjacent edges of a vertex are sampled independently and with the same probability p'. In either case, we end up, with high probability, with almost the *same number* of edges adjacent to any vertex, which are *randomly selected*. We note that it is true that the spaces $G^r_{n,p}$ and $G_{n,p'}$ are different, but the difference includes only the graphs where although p' is small, the degree of a vertex may exceed r, and this event is of negligible probability compared to the rest, from Chernoff bounds. In fact, by choosing (for an $\beta \in (0,1)$) $np' = \frac{rp}{1+\beta}$, the model $G_{n,p'}$ is weaker that the $G^r_{n,p}$. That is, in this case the "randomly truncated" random regular graph $G^r_{n,p}$ and $G_{n,p'}$ *produce almost the same graph outcomes and have the same properties*, with extremely high probability. The minimum necessary probability p_c so that $G_{n,p}$ graphs are a.c. k-connected when $p \geq p_c$, is given in [6] (p. 152, lemma 5). By calculating the maximum value of k such that

$$p' = \frac{\alpha \log n}{2(1+\beta)n} \geq p_c = \frac{\log n + (k-1)\log\log n - \omega(n)}{n}$$

where $\omega(n) \to \infty$, $\omega(n) \leq \log\log\log n$, we get the following theorem.

Theorem 2 *Let G be an instance of $G^r_{n,p}$ where $p = \Theta(1)$ and $r \geq \alpha \log n$, where $\alpha > 0$ an appropriate constant. Then G is almost certainly k-connected, where*

$$k = O\left(\frac{\log n}{\log\log n}\right)$$

□

3.3. $G^r_{n,p}$ is a.c. disconnected for any $r = o(\log n)$ and constant f

We remark here that the constant link failure probability dramatically alters the connectivity structure of the regular graph. While G^r_n is a.c. $r-$connected for any $r \geq 3$, we get that $G^r_{n,p}$ is *disconnected* w.h.p. for all $r \leq \frac{\sqrt{\log n}}{2}$ and that large degrees $r \geq \alpha \log n$ are needed for $G^r_{n,p}$ to be a.c. connected. The gap $\left(\frac{1}{2}\sqrt{\log n}, \alpha \log n\right)$ is due to the translation theorem from random configurations to $G^r_{n,p}$ (we recall that the extended translation lemma is valid for $r \geq 2$ fixed). In order to handle this gap, we must avoid the translation lemma. In fact, first consider a graph $G \in G^r_n$. By Turán's celebrated theorem ([28]), it contains an independent set S with cardinality $|S| \geq \frac{n}{r+1}$. Now delete edges from the graph with probability p, and let X_v denote the event that in the new graph G_p obtained in this way a vertex v is isolated. Then, for each pair v, w of vertices of S the events X_v and X_w are independent. Now simple calculations and Chebyshev's inequality reveal that w.h.p. the graph G_p contains an isolated vertex. Then we can repeat the calculations that we did in the proof of Theorem 1 (exactly the same argument but not in the configuration graph, so we avoid the translation lemma in this case) and get that $G^r_{n,p}$ is disconnected a.c. for any $r = o(\log n)$ and constant f.

3.4. $G^r_{n,p}$ is connected (except for $O(1)$ vertices) and admits no small separators for $f = O(n^{-\epsilon})$, for any fixed $\epsilon \geq 3/(2r)$

We now consider the (more practical) case in which $f = 1 - p = o(1)$ and we prove that the desired connectivity properties of random regular graphs are almost preserved despite the link failures.

The technical lemmas related to random configurations are still valid and the following also holds:

Lemma 3 *There is a constant $c = c(r)$ such that $G^r_{n,p}$ graphs do not have a subgraph with $k < c \log n$ vertices and at least $k+1$ edges; this holds with probability at least $1 - e^{-n}$.*

Proof. The proof is by a similar result in [6] (p. 174, Theorem 32) on G^r_n and by the fact that a $G^r_{n,p}$ graph can not have more edges than the corresponding G^r_n graph. □

Consider now in $G^r_{n,p}$ a partition of $V : V = A \cup S \cup B$, where $s_1 = |S| \geq (r-2)\alpha_0, \alpha_0 \geq 3$. Let $\alpha_1 = 4(r-2)\alpha_0$ and consider all partitions in which $|A| = \alpha \geq \alpha_1$. Consider the event: $\varepsilon_1(\alpha) =$ "There is a partition $V = A \cup S \cup B$ as above with no $A - B$ edges, i.e. with S a separating set and $|A| = \alpha$ ".

Theorem 3 *Almost every $G^r_{n,p}$ graph is such that if a smaller than cardinality r set separates it, then only $O(1)$ vertices break away and the rest of the graph (of $O(n)$ size) is $r-$connected.*

Proof. In terms of the terminology above, it suffices to show that $\sum_\alpha Pr\{\varepsilon_1(\alpha)\} \to 0$. By Lemma 3, $G^r_{n,p}$ is such that w.h.p. every set of $k < c(r) \log n$ vertices spans at most k edges.

We have $\binom{n}{s_1}$ choices for S and $\binom{n-s_1}{\alpha}$ choices for A. By using random configurations we see that a configuration edge incident with an element of $\hat{A}$ joins this element to $\hat{B}$ with probability at least $(\frac{n-s_1-\alpha}{n})p$.

Definition 4 *Let $S(\alpha) = Pr\{\varepsilon_1(\alpha)\}$.*

By pairing off elements of $\hat{A}$ one by one and by conditioning on the events of Lemma 3 we get:

$$S(\alpha) \leq \binom{n}{s_1} \binom{n-s_1}{\alpha} \left(1 - \frac{n-s_1-\alpha}{n}\, p\right)^{\frac{\alpha r}{2}}$$

Now, we use $f = 1 - p$ to get :

$$S(\alpha) \le S'(\alpha) = n^{s_1} \binom{n - s_1}{\alpha} \left(\frac{s_1 + \alpha}{n} (1 - f) + f \right)^{\frac{\alpha r}{2}}$$

We will study only the worst, with respect to connectivity issues, case where $r = O\left(\sqrt{\log n}\right)$. For bigger r, things are even better. We note that:

1. $s_1 \ge (r-2)\alpha_0 \Rightarrow n - s_1 \le n - (r-2)\alpha_0$ (and $r = O(\sqrt{\log n})$). Thus $n - s_1 < n$ and $\binom{n-s_1}{\alpha} < n^{\alpha}$.
2. $\alpha \ge \alpha_1 = 4(r-2)\alpha_0 = O\left(\sqrt{\log n}\right)$. Thus $\frac{s_1+\alpha}{n} = O\left(\frac{\sqrt{\log n}}{n}\right) \to 0$ as $n \to +\infty$.
3. Since $f = o(1) \Rightarrow 1 - f \to 1$, thus $(\frac{s_1+\alpha}{n})(1-f) \to 0$ (by the preceding fact).

Clearly, the hardest case for connectivity is when f is large. So, let us assume that $f = \Omega\left(\frac{\sqrt{\log n}}{n}\right)$. Then $S'(\alpha) \le n^{s_1+\alpha}((1+\gamma)f)^{\alpha r}$ for some $\gamma > 0$. We will now need the following Corollary:

Corollary 1 $f \le \frac{1}{1+\gamma}\, n^{-\frac{2s_1+\alpha}{\alpha r}} \Rightarrow S'(\alpha) \le n^{-s_1}$.

Proof. Let $f_1 = (1+\gamma)f$. Clearly, $f_1 \le n^{-\frac{2s_1+\alpha}{\alpha r}} \Rightarrow \log f_1 \le -\frac{2s_1+\alpha}{\alpha r}\log n \Rightarrow \alpha r \log f_1 \le -(2s_1+\alpha)\log n \Rightarrow f_1^{\alpha r} \le \frac{1}{n^{2s_1+\alpha}} \Rightarrow S'(\alpha) \le n^{-s_1}$ □

We now return to the proof of the theorem.

Let $\alpha_2 = \lfloor \frac{n-\alpha_1}{2} \rfloor$. Clearly

$$\sum_{\alpha} \Pr\{\varepsilon_1(\alpha)\} \le \sum_{\alpha=\alpha_1}^{\alpha_2} S'(\alpha) \tag{2}$$

Since we assumed $f = \Omega\left(\frac{\sqrt{\log n}}{n}\right)$, we have that

$$\frac{\sqrt{\log n}}{n} < f \le \frac{1}{1+\gamma}\, n^{-\frac{2s_1+\alpha}{\alpha r}}$$

and thus we can apply Corollary 1. Note that: $\frac{S'(\alpha+1)}{S'(\alpha)} \sim n f^r$ i.e. $\frac{S'(\alpha+1)}{S'(\alpha)} \le n(\frac{1}{1+\gamma})^r n^{-\frac{2s_1+\alpha}{\alpha}}$. But $4s_1 \ge \alpha$. Hence $\frac{2s_1+\alpha}{\alpha} \ge \frac{3}{2}$. Thus $\frac{S'(\alpha+1)}{S'(\alpha)} \le n(\frac{1}{1+\gamma})^r n^{-\frac{3}{2}} = \frac{1}{\sqrt{n}}(\frac{1}{1+\gamma})^r < \frac{1}{\sqrt{n}}$. Then, we easily get

$$\sum_{\alpha=\alpha_1}^{\alpha_2} S'(\alpha) \to 0$$

as $n \to +\infty$ and the result follows from inequality (2). □

From the above theorem (by using the extended translation lemma for $r \ge 2$ fixed, and for greater unbounded r by the monotonicity of the property of a graph being connected and admitting no small separators), we get:

Theorem 4 *Let $r \ge 3$ and $f = 1 - p = O(n^{-\epsilon})$ for $\epsilon \ge \frac{3}{2r}$. Then the biggest part of $G^r_{n,p}$ (i.e. the whole graph except of $O(1)$ vertices) remains connected and this* *connected part (excluding the vertices that were originally neighbours of the $O(1)$-sized disconnected set) can not be separated unless we remove more than r vertices, with probability tending to 1 as n tends to $+\infty$.*

$\square$

4. Existence of a Giant Component in $G^r_{n,p}$

Since $G^r_{n,p}$ is a.c. disconnected for $r = o(\log n)$ and $1 - p = f = \Theta(1)$, we would like to know whether at least a large part of the network represented by $G^r_{n,p}$ is still connected, i.e. whether the biggest connected component of $G^r_{n,p}$ is large. Perhaps the most celebrated theorem in Random Graph Theory is that of the Giant Component: $G_{n,p}$ has a giant (i.e. $\Theta(n)$-sized) connected component when, roughly, $p > \frac{1}{n}$. This was first remarked by Erdös and Renyi ([10]). We will here show that:

Theorem 5 *When $f < 1 - \frac{32}{r}$ then $G^r_{n,p}$ admits a giant (i.e. $\Theta(n)$-sized) connected component for any $r \geq 64$ with probability at least $1 - O\left(\frac{\log^2 n}{n^{\alpha/3}}\right)$, where $\alpha > 0$ a constant that can be selected.*

Proof. We will prove the Theorem for r fixed by using the extended translation lemma. The result then obviously extends to greater (unbounded) r by the monotonicity of the giant connected component property.

We will consider the random configuration $\hat{F}$ corresponding to $G^r_{n,p}$. We first select each existing edge of $\hat{F}$ with probability $1/2$, independently. This partitions the edges of $\hat{F}$ into two sets of edges (E_1 and E_2).

Since $f < 1 - \frac{32}{r}$ we have $p > \frac{32}{r}$ i.e. $pr > 32$. In the graph G^r_n we have a total number of edges $\frac{rn}{2}$ and after the experiment of *independently* deleting each edge with probability f and the experiment of the partition, the remaining number of edges in E_1, E_2 satisfies the Bernoulli $B\left(\frac{nr}{2}, \frac{p}{2}\right)$ i.e. by Chernoff bounds we get that $|E_1|, |E_2| \geq (1-\beta)\frac{nrp}{4}$, for any $\beta \in (0,1)$ w.h.p. and since $pr > 32$ we get $|E_1|, |E_2| \geq (1-\beta)8n$ with probability $\geq 1 - \exp\left(-\frac{\beta^2}{2}8n\right)$. We now condition, for E_1, on the event $A_1 =$ "$|E_1| \geq (1-\beta)8n$".

Let $\hat{F}(E_i)$ the random configuration corresponding to the edge set E_i, $i = 1, 2$. The following technical claim will be needed here:

Claim 1 *$\hat{F}(E_1)$ contains a "path" (i.e. a sequence of sets w_i, of r vertices, each connected to the next by an edge) of length at least $\alpha \log n$ ($\alpha > 0$ an appropriate constant that can be selected) with probability at least $1 - n^{-\lambda}, \lambda \geq 2$.*

Proof. Consider now the set E_1 partitioned into g groups $g_1, g_2, \ldots$ of $\alpha \log n$ ($\alpha > 0$ an appropriate constant) edges in each group. For each group g_i let q_i be the probability that it will form a path in $\hat{F}(E_1)$ if thrown alone in the $\binom{rn}{2}$ possible "places". Let a random indicator variable $X_i = 1$ with probability q_i and $X_i = 0$ else. Clearly the random variables X_i are *not independent.* The number of ways for g_i to form a path is calculated as follows:

(a1) Choose the supervertices of the path in $\binom{n}{\alpha \log n}$ ways

(a2) Order them in the path in $(\alpha \log n)!$ ways (a3) In this order $(r^2)^{\alpha \log n - 1}$ paths are defined, because any two specific supervertices can be joined by an edge in r^2 ways.

But there are
$$\binom{\binom{rn}{2}}{\alpha \log n}$$
places for the edges of g_i. Thus
$$q_i = \frac{\binom{n}{\alpha \log n}(\alpha \log n)!\ r^{2\alpha \log n - 2}}{\binom{\binom{rn}{2}}{\alpha \log n}}$$

The denominator is at most $(rn)^{2\alpha \log n} = r^{2\alpha \log n} n^{2\alpha \log n}$, thus
$$q_i \geq \frac{\frac{n!}{(\alpha \log n)!(n-\alpha \log n)!}(a \log n)!}{r^2\ n^{2\alpha \log n}}$$
i.e.
$$q_i \geq \frac{n(n-1)\cdots(n-\alpha \log n + 1)}{r^2 n^{2\alpha \log n}}$$
and since $r \leq n-1$ we get
$$q_i \geq \frac{(n-2)\cdots(n-\alpha \log n + 1)}{n^{2\alpha \log n}} \geq \frac{(\frac{n}{2})^{\alpha \log n - 2}}{n^{2\alpha \log n}} \geq \frac{4}{n^{\alpha \log n + 2 + \alpha}}$$
But there are $N \geq \binom{4n}{\alpha \log n}$ groups g_i i.e.
$$N \geq \frac{(4n)!}{(\alpha \log n)!(4n - \alpha \log n)!}$$
i.e. (crudely, and for sufficiently large n)
$$N \geq \frac{(3.9n)^{\alpha \log n}}{(\alpha \log n)!} \geq (3n)^{\alpha \log n}$$
Notice that forall g_i
$$N \cdot q_i \geq 4(1.5)^{\alpha \log n}$$
i.e.
$$N \cdot q_i \geq 4\left(\frac{3}{2}\right)^{\alpha \log n} = 4n^{\alpha \log\left(\frac{3}{2}\right)}$$

As remarked in [9], "a frequently occurring scenario underlying the analysis of many randomized algorithms and processes involves random variables that are, intuitively, dependent in the following *negative* way: if one subset of the variables is "high", then a disjoint subset of the variables is "low". One specific paradigm involving negative dependency is the classical ball and bins experiment".

Definition 5 (Negative Association, [9]) *Let $X := (X_1, \ldots, X_n)$ be a vector of random variables. The random variables X_i are negatively associated if for every two disjoint index sets $I, J \subseteq [n]$,*
$$E[f(X_i, i \in I)g(X_j, j \in J)] \leq E[f(X_i, i \in I)]E[g(X_j, j \in J)]$$
for all functions $f : R^{|I|} \rightarrow R$ and $g : R^{|J|} \rightarrow R$ that are both non-decreasing or both non-increasing.

Now we notice that in our case the random indicator variables X_i and X_j (showing whether a path is formed or not in group g_i, g_j) are negatively associated for all i, j, because if $X_j = 1$ (high) then X_i is lower since some places are already occupied. That is

$$E(X_i|X_j) \leq E(X_i)$$

i.e.

$$E(X_i X_j) \leq E(X_i)E(X_j)$$

(negative covariance).

We are interested in the sum $X = \sum_i X_i$ i.e. in the number of paths of length $\alpha \log n$ in $\hat{F}(E_1)$. From [9] we have Theorem 6. For its proof we will use the following Lemma.

Lemma 4 (Dubhashi and Ranjan, [9]) *Let $X_1, \ldots, X_n$ be random variables satisfying the negative association condition. Then for any non-decreasing functions $f_i, i \in [n]$,*

$$E\left(\prod_{i\in[n]} f_i(X_i)\right) \leq \prod_{i\in[n]} E(f_i(X_i))$$

Proof. Take the non-decreasing functions $f(X_i, i < n) := \prod_{i<n} f_i(x_i)$ and $g(x_n) := f_n(x_n)$ to deduce that $E\left(\prod_{i\in[n]} f_i(X_i)\right) \leq E\left(\prod_{i<n} f_i(X_i) E(f_n(X_n))\right)$ and now use induction.

(end of proof of Lemma 4). □

Theorem 6 (Dubhashi and Ranjan, [9]) *The Chernoff-Hoeffding bounds are applicable to sums of random variables that satisfy the negative association condition, as in definition 5.*

Proof. Let $X_1, \ldots, X_n$ be negatively associated (and bounded) variables. To show that the Chernoff-Hoeffding bounds apply to the sum $X = X_1 + \cdots + X_n$, we use the standard proof of the Chernoff-Hoeffding bound, see for example [22]. The only change needed is in a crucial step, where one uses the fact for *independent* variables, $E(e^{tX}) = E(\prod_i e^{tX_i}) = \prod_i E(e^{tX_i})$. For negatively associated variables, we have, for $t > 0$, $E(e^{tX}) = E(\prod_i e^{tX_i}) \leq \prod_i E(e^{tX_i})$, by Lemma 4 applied with each $f_i(x) := e^{tx}$. The rest of the proof is unchanged, and gives the upper tail bound. For the lower tail, we apply the same argument to the variables $b_i - X_i$, where b_i is an upper bound on the variable X_i. Note that if the X_i variables are negatively associated, then so are the variables $b_i - X_i$. (In fact, because of the monotonicity of the exponential function, the negative covariance relation among the variables suffices to prove the theorem).

(end of proof of Theorem 6) □

Thus, the Chernoff-Hoeffding bounds are applicable in this case and we have that, for any $\gamma \in (0, 1)$,

$$\Pr\{X \geq (1-\gamma)Nq_i\} \geq 1 - e^{-\frac{\gamma^2}{2}Nq_i}$$

i.e. the number of paths of length $\alpha \log n$ in $\hat{F}(E_1)$ is at least $4n^{\alpha \log\left(\frac{3}{2}\right)}$ with probability at least

$$1 - e^{-\frac{\gamma^2}{2} 4n^{\alpha \log\left(\frac{3}{2}\right)}}$$

As an easy conclusion, $\hat{F}(E_1)$ contains at least one such path of length $\alpha \log n$ with probability at least

$$\left(1 - e^{-\frac{\gamma^2}{2} 4n^{\alpha \log\left(\frac{3}{2}\right)}}\right)\left(1 - e^{-\frac{\beta^2}{2} 8n}\right) \geq 1 - n^{-\lambda}$$

for any $\lambda \geq 1$.

(end of proof of Claim 1) $\square$

Let us now condition the rest of the proof of the theorem on the existence of such a path $\Pi = (P_1, P_2, \ldots, P_{\alpha \log n})$, where each "supervertex" P_i (of r vertices) is connected to P_{i+1} by an edge of E_1. We now define a Breadth-First-Search Expansion Process in $\hat{F}(E_2)$ as follows:

BFS Expansion Process

1. Let $B_0 = \{P_1, P_2, \ldots, P_{\alpha \log n}\}; i = 0$.
2. Let $i \leftarrow i + 1$.
3. Let B_i be the set of supervertices not in S_{i-1} that are connected to at least one supervertex in S_{i-1} by an edge.
4. If $B_i \neq \emptyset$ then go to 2.

For each set of supervertices B_i, define:

Definition 6 *Let $S_i = \cup_{j \leq i} B_j$.*

Definition 7 *Let Active(B_i) be the set of those supervertices in B_i each of which has only at most $r/2$ edges connected to S_i.*

Definition 8 *Let Out(B_i) be the set of edges of B_i connecting to vertices which do not belong to S_i.*

Definition 9 *Let R_i be the event $|Active(B_i)| \geq \frac{|B_i|}{2}$.*

Note that the edges we use in the expansion process are independent of the edges we used to get the existence of the path Π. Let $i \leq \log n - o(\log n)$. Choose $\alpha \geq \frac{6 \cdot 32}{rp}$. Assume inductively that

$$\Pr\{R_i\} \geq \left(1 - \frac{1}{n^{\alpha/3}}\right)^i$$

For $i = 0$, $\Pr\{R_i\} = 1$ (basis). By the induction hypothesis and conditioning on R_i, we get that at least $\frac{r}{2}\frac{|Bi|}{2}$ edges of S_i go out of S_i. Each of them will exist with probability p. Thus, by the Bernoulli $B\left(\frac{r}{2}\frac{|Bi|}{2}, p\right)$, and because $rp > 32$, the event $\epsilon_1(i)$ = "at least $(1 - \beta_1)8|B_i|$ edges exist and go out of S_i departing from

S_i" has (by Chernoff bounds) probability at least $1 - \exp\left(-\frac{\beta_1^2}{2} 8|B_i|\right)$ and since $|B_i| \geq \alpha \log n$ we get that

$$\Pr\{\epsilon_1(i)\} \geq 1 - e^{-\frac{\beta_1^2}{2} 8\alpha \log n} \geq 1 - n^{-\beta_1^2 4\alpha}$$

for any $\beta_1 \in (0, 1)$.

Now, condition also on $\epsilon_1(i)$. Then let us see how many new supervertices these edges hit. We can safely assume that during the process, $|S_i|$ (i.e. the number of supervertices in the component in the i-th stage of the BFS) is less than $\frac{n}{8}$ (else we have a giant component already). But then consider the edges out of S_i in the order $e_1, e_2, \ldots, e_x$ $(x \geq (1-\beta_1)8|B_i|)$. Let the set $\{e_1, e_2, \ldots, e_\kappa\}$ have hit m_1 new supervertices. We may again assume $m_1 < \frac{n}{8}$ (since else we are done). Consider $e_{\kappa+1}$. The probability that it hits another *new* supervertex is at least $\frac{n - \frac{n}{8} - \frac{n}{8}}{\frac{7n}{8}} \geq \frac{6}{7}$ and this holds for all $\kappa \leq x$. Consider the Bernoulli $B\left(x, \frac{6}{7}\right)$. By Chernoff bounds, the number of edges that each hits a supervertex separate from those hit before is $(1-\beta_2)x\,\frac{6}{7}$ thus at least $(1-\beta_2)(1-\beta_1)8|B_i|\frac{6}{7}$ i.e. at least $(1-\beta_3)6|B_i|$ with probability at least

$$1 - e^{-\frac{\beta_2^2}{2} 6|B_i|} \geq 1 - n^{-3\alpha\beta_2^2}$$

since $|B_i| \geq \alpha \log n$.

Note that by appropriately choosing β_1, β_2 we get that the number of edges out of S_i that each hits a new supervertex separate from before, is at least $6|B_i|$ with probability at least $1 - n^{-3\alpha\frac{\beta_2^2}{2}}$. But each of these new supervertices (they belong to B_{i+1}) is *active by construction*, because each has only one edge into S_i, thus $r-1$ edges free to expand. So, given that R_i holds, the conditional probability $\Pr\{R_{i+1}|R_i\}$ is at least

$$\Pr\{R_{i+1}|R_i\} \geq \Pr\{\epsilon_1(i)\}\left(1 - n^{-3\alpha\beta_2^2}\right) \geq 1 - n^{-\frac{\alpha}{3}}$$

Thus

$$\Pr\{R_{i+1}\} \geq \Pr\{R_i\}\Pr\{R_{i+1}|R_i\} \geq \left(1 - n^{-\frac{\alpha}{3}}\right)^{i+1}$$

(end of proof of induction)

So, in each stage we get an expansion of the number of connected new supervertices by 6, provided $i \leq \gamma \log n - o(\log n)$ $(\gamma \geq 1)$. Thus, at the end we get a giant component of size $\Theta(n)$, *provided all events* R_i *hold.*

The probability of the event "$\exists i : \bar{R}_i$" is the probability of failure of the process. This is at most

$$\Pr\{\exists i : \bar{R}_i\} \leq \sum_{i=0}^{\log n} \Pr\{\bar{R}_i\}$$

But

$$\Pr\{R_i\} \geq \left(1 - \frac{1}{n^{\frac{\alpha}{3}}}\right)^i \geq 1 - \frac{i}{n^{\frac{\alpha}{3}}}$$

thus

$$\Pr\{\bar{R}_i\} \leq \frac{i}{n^{\frac{\alpha}{3}}} \leq \frac{\log n}{n^{\frac{\alpha}{3}}}$$

and finally

$$\sum_{i=0}^{\log n} \Pr\{\bar{R}_i\} \leq \frac{\log^2 n}{n^{\frac{\alpha}{3}}}$$

This ends the proof of Theorem 5. □

As an immediate consequence of Theorem 5, we get:

Corollary 2 *A giant component of $G^r_{n,p}$ can be constructed in polynomial time, with probability at least $1 - O\left(\frac{\log^2 n}{n^{\alpha/3}}\right)$, where $\alpha > 0$ a constant that can be selected.*

Proof. The described construction of Π and the BFS process can be viewed as an algorithm that (in case of no failures) actually reveals a giant connected component. It is easy to see that the expected running time of this algorithm is $O(n \log n)$ with probability at least $1 - \Theta\left(\frac{\log^2 n}{n^{\frac{\alpha}{3}}}\right)$, where $\alpha > 0$ a constant that can be selected. □

Acknowledgements

We wish to thank D. Knuth for his insightful remarks when this work was presented in a preliminary form and Z. Kedem for his constructive comments in an earlier form of this work. We also wish to thank the anonymous referee of a previous version of our paper for his suggestion that extended the power of Theorem 1 up to $r = o(\log n)$.

This work has been supported in part by the NSF grant CCR-89-6949 and by the ESPRIT IV Basic Research Programme of the EC under contract No. 20244 (Project ALCOM-IT).

References

1. N. Alon and J. Spencer, *The Probabilistic Method*, John Wiley and Sons, 1992.
2. D. Bauer, F. Boesch, C. Suffel and R. Tindell, "Connectivity extremal problems and the design of reliable probabilistic networks", *The Theory of Applications of Graphs*, John Wiley and Sons, 1981.
3. F. T. Boesch, F. Harary and J. A. Cabell, "Graphs and models of communication networks vulnerability: connectivity and persistence", *Networks*, vol. 11, pp. 57–63, 1981.
4. B. Bollobás, "A probabilistic proof of an asymptotic formula for the number of labeled regular graphs", *European Journal of Combinatorics*, vol. 1, pp. 311–316, 1980.
5. B. Bollobás, "The distribution of the maximum degree of a random graph", *Discrete Mathematics* 32, pp. 201–203, 1980.
6. B. Bollobás, *Random Graphs*, Academic Press, 1985.
7. B. Bollobás and A. G. Thomason , "Random graphs of small order", *Random Graphs*, *Annals of Discrete Mathematics*, pp. 47–97, 1985.

8. C. Chang and R. Nelson, "Bounds on the Speedup and Efficiency of Partial Synchronization in Parallel Processing Systems", Technical Report RC 16474, IBM T. J. Watson Research Center, pp. 1–41, January 17, 1991.

9. D. Dubhashi and D. Ranjan, "Balls and Bins: A Study in Negative Dependence", BRICS Report Series, RS-96-25, July 1996.

10. P. Erdös and A. Renyi, "On the evolution of random graphs", *Magyar Tud. Akad. Math. Kut. Int. Kozl.* 5, pp. 17–61, 1960.

11. S. Even and R. E. Tarjan, "Network flow and testing graph connectivity", *SIAM Journal of Computing*, vol. 4, pp. 507–518, 1975.

12. P. Flajolet, K. Hatzis, S. Nikoletseas and P. Spirakis, "Trade-offs between Density and Robustness in Random Interconnection Graphs", to appear *in the Proceedings of the IFIP International Conference on Theoretical Computer Science (TCS2000)*, Sendai, Japan, 2000.

13. E. Gelenbe, R. Nelson, R. Phillips and A. Tantawi, "An Approximation for the Processing Time for a Random Graph Model of Parallel Computation", *Proceedings of the ACM FJCC*, 1986.

14. A. Goerdt, "The giant component threshold for random regular graphs with edge faults", *Proceedings of Mathematical Foundations of Computer Science '97 (MFCS'97)*, pp. 279-288, 1997.

15. J. Hastad, T. Leighton and M. Newman, "Fast Computation Using Faulty Hypercubes", *Proceedings of the 21st ACM Symposium on Theory of Computing (STOC)*, pp. 251–263, 1989.

16. S. Janson, D. Knuth, T. Luczak and B. Pittel, "The birth of the giant component", *Journal of Random Structures and Algorithms*, vol. 4, pp. 232–355, 1993.

17. P. Kanellakis and A. Shvartsman, "Efficient Parallel Algorithms Can be Made Robust", Technical Report CS-89-35, Brown University, pp. 1–28, October 24, 1989. (Initial version appeared in the Proceedings of the *8th ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 211–222, 1989).

18. Z. Kedem, K. Palem and P. Spirakis, "Efficient Robust Parallel Computations", *Proceedings of the 22nd ACM Symposium on Theory of Computing (STOC)*, pp. 138–148, 1990.

19. Z. Kedem, K. Palem, A. Raghunathan and P. Spirakis, "Combining Tentative and Definite Executions for Very Fast Dependable Parallel Computing", *Proceedings of the 23rd ACM Symposium on Theory of Computing (STOC)*, 1991.

20. Z. Kedem, K. Palem, P. Spirakis and M. Yung, "Faulty Random Graphs: reliable efficient-on-the-average network computing", Computer Technology Institute, Patras, Greece, Technical Report, 1993.

21. G. A. Margulis, "Probabilistic characteristics of graphs with large connectivity", Problems on Information Transmission, Plenum Press, New York, 1977.

22. R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.

23. S. Nikoletseas, K. Palem, P. Spirakis and M. Yung, "Short Vertex Disjoint Paths and Multiconnectivity in Random Graphs: Reliable Network Computing", *21st International Colloquium on Automata, Languages and Programming (ICALP)*, Jerusalem, pp. 508 – 515, 1994.

24. S. Nikoletseas, G. Pantziou, P. Psycharis and P. Spirakis, "On the reliability of fat-trees", *3rd International European Conference on Parallel Processing (Euro-Par)*, pp. 208 – 217, Passau, Germany, 1997.

25. S. Nikoletseas and P. Spirakis, "Expander Properties in Random Regular Graphs with Edge Faults", *12th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, München, pp. 421 – 432, 1995.

26. J. S. Provan and M. O. Bull, "The complexity of counting cuts and computing the probability that a graph is connected", Working Paper MS/S 81-002, College of Business and Science, University of Maryland at College Park, 1981.

27. M. Rabin, "Efficient Dispersal of Information for Security, Load Balancing and Fault Tolerance", *Journal of the ACM*, vol. 36, no. 2, pp. 335–348, 1989.

28. P. Turán, "On an extremal problem in graph theory", *Matematicko Fizicki Lapok*, 48, pp. 436–452, 1941.

29. N. Wormald, "The asymptotic distribution of short cycles in random regular graphs", *Journal of Combinatorial Theory* (B) 31, pp. 168 – 182, 1981.