The Giant Component Threshold for Random Regular Graphs with Edge Faults

Andreas Goerdt

TU Chemnitz, Theoretische Informatik
D-09107 Chemnitz, Germany

Abstract. Let G be a given graph (modelling a communication network) which we assume suffers from static edge faults: That is we let each edge of G be present independently with probability p (or absent with *fault probability* $f = 1 - p$). In particular we are interested in robustness results for the case that the graph G itself is a random member of the class of all regular graphs with given degree.
Our result is: If the degree d is fixed then $p = 1/(d-1)$ is a threshold probability for the existence of a linear-sized component in a faulty version of almost all random regular graphs. We show: If each edge of an *arbitrary* graph G with maximum degree bounded above by d is present with probability $p = \lambda/(d-1)$ where $\lambda < 1$ is fixed then the faulted version of G has only components whose size is at most logarithmic in the number of nodes with high probability. If on the other hand G is a random regular graph with degree d and $p = \lambda/(d-1)$ where $\lambda > 1$ then for almost all G the faulted version of G has a linear-sized component with high probability. Note that these results imply some kind of optimality of random regular graphs among the class of graphs with the same degree bound.
The theme is: Use the known expansion properties of almost all random regular graphs to obtain strong robustness results. This has not been done systematically before.

Introduction

Modern multiprocessor architectures and communication networks compute over structured interconnection graphs like meshes. Here several applications share the same network while executing concurrently. This may of course lead to unavailability of links and nodes in certain cases and we may assume to compute over a subnetwork being randomly assigned by the operating system. Moreover, this subnetwork may suffer from edge or node faults. Our work addresses robustness properties in case the subnetwork is a random regular graph suffering from edge faults. Random regular graphs make an (at least theoretically) popular choice because they combine low degree with high expansion almost always (see [AS 92] for an introduction to expansion and [Bo 85] to random regular graphs). Our study continues the work begun in [Ni et al. 94] [NiSp 95] which are the only papers known to us which investigate random regular graphs with edge faults. In particular we build on the first of the abovementioned papers, where the authors show: Almost all faulty versions (i. e. with probability tending to 1 when the number of nodes goes to infinity) of random regular graphs with degree d where $64 < d < (1/2)\sqrt{\log n}$ and $p \geq 32/d$ have a linear-sized component. The second of the abovementioned papers deals with expansion properties of such a linear-sized component. For the case that $d \geq 3$ is fixed, we strengthen the result of [Ni et al. 94] as far as possible. Precise formulations of our results are theorems 1 and 2.

Our result adds to the list of theorems giving threshold probabilities for the existence of a linear-sized component: If G is the fully connected graph we are in the realm of the classical theory of random graphs. Threshold for a linear size component is $p = 1/n$, n being the number of nodes.In case of the $d \times d$ $2-$dimensional grid we have $p = 1/2$. (At least in [KaNeTa 94] it is assured that results from [Ke 80] can be looked at this way). If G is the hypercube in d dimensions, i. e. we have $n = 2^d$ nodes, the threshold for a linear-sized component is $1/d$ [AjKoSz 82]. Without giving the actual value, the existence of a threshold probability roughly between one third and one half in case of the butterfly network is shown in [KaNeTa 94].

Note that in case of the fully connected G and the hypercube the threshold probability is something like the inverse of the degree. Our threshold probability for random regular graphs fits in nicely.

In spite of the fact that the existence of a linear-sized component is not the only measure of the robustness of a network it is a *necessary* requirement for efficiently simulating computations of the fault free graph in the faulted version: Without a linear-sized component computation on the faulted version must be slower by more than a constant factor than computation on the unfaulted version. In case of the $2-$dimensional grid the threshold probability of a linear-sized component coincides with the threshold probability of other properties related more closely to the efficiency of computation, such as routing or embedding the fault-free version into the faulty one [Ra 89] [Ma 92].

1 Below the threshold probability

Let $H = (V, F)$ be an undirected graph of maximum degree $\leq d$ where d is a fixed constant. According to our fault model we consider the probability space of faulty versions $G = (V, E)$ of H where $E \subseteq F$. The probability of G is defined by: Each edge of F is present, i. e. in E, independently with probability $p < 1/(d-1)$, where p is a fixed constant.

For a fixed node $v \in V$ we consider the following probabilistic algorithm to generate the subgraph reachable from v in a faulty version of H . The algorithm is like a breadth-first search, we describe it as a sequence of sets of nodes and edges: We set $S_0 = \{v\}$, where v is the fixed initial node. The sets of edges E_1 and $\neg E_1$ are obtained probabistically: Each edge of H, $e = (v, w)$, incident with v is either included in E_1 with probability p or in $\neg E_1$ with $1 - p$. We let

$T_1 = \{w \mid (v,w) \in E_1\}$ be the set of nodes which are reached from v by edges in E_1. Finally $S_1 = T_1$ is the new front of our graph generation procedure.

Now assume S_i,E_i, and$\neg E_i$ are defined. If $S_i = \emptyset$ we are finished, the succeeding sets are not any more defined. If $i \geq 1$ and $S_i \neq \emptyset$ and j is the smallest index, for which T_j is undefined, we pick a node $w \in S_i$ according to some deterministic rule (e. g. w could be the node with the smallest number in S_i).The edge sets $E_{i+1}, \neg E_{i+1}$ are constructed as follows: We start with $E_{i+1} = E_i$ and $\neg E_{i+1} = \neg E_i$. Then each edge (w,u) where $u \notin S_0 \cup ... \cup S_i$ is added to E_{i+1} with probability p and to $\neg E_{i+1}$ with $1-p$. We let $T_j = \{u \mid (w,u) \in E_{i+1} \setminus E_i\}$, i. e. T_j collects the nodes discovered via w. We delete w from S_i, and add T_j to S_{i+1}. If S_i now becomes empty, we continue with the current S_{i+1}, if however S_i is not yet empty we repeat the procedure using T_{j+1} instead of T_j. We continue this way until S_i is empty.

First observe that each pair of sets of edges $E, \neg E$ is generated in at most one way by the algorithm: Let $E_1, \neg E_1, E_2, \neg E_2, \ldots,, E_i = E, \neg E_i = \neg E$ and $E'_1, \neg E'_1 \ldots, E'_i = E, \neg E'_i = \neg E$ be sequences of edge sets generated by our algorithm, having the same sets of edges at the end. As the edges incident with v are determined by the first 2 pairs of edge sets we must have $E_1 = E'_1$ and $\neg E_1 = \neg E'_1$. The node w picked deterministically next must be the same in both cases. All edges (w,u) with $u \notin S_1 \cup S_0 = S'_0 \cup S'_1$ are determined in the second round of the algorithm. Hence, $E_2 = E'_2$ and $\neg E_2 = \neg E'_2$. The claim follows inductively. Let $E = E_i$, $\neg E = \neg E_i$ be a pair of edge sets which can be generated by our algorithm, then we have:

The probability to generate $E, \neg E$
$= p^{|E|}(1-p)^{|\neg E|}$
$=$ the probability of those faulty versions $G = (V, E')$ of H with $E' \supseteq E$ and $F \setminus E' \supseteq \neg E$,

where F is the set of edges of the original H.
Let $\mathrm{Comp}_H(v)$ be the maximal connected component containing v and let $\mid \mathrm{Comp}_G(v) \mid$ be the number of nodes in $\mathrm{Comp}_G(v)$. Let in the sequel $t \geq 2$. We bound the probability of the set of faulty versions of our original graph H in which the component of v has more than t nodes:

$\Pr\{G = (V,E) \mid \mid \mathrm{Comp}_G(v) \mid > t\}$
$= \Pr[$The algorithm generates sets $T_1, T_2, ..., T_t$ with
$\mid T_1 \mid \geq 1, \mid T_1 \mid + \mid T_2 \mid \geq 2, \ldots,$ and $\mid T_1 \mid + \cdots + \mid T_t \mid \geq t]$.

Let $s \leq d$ be the number of neighbours of v, then $\mid T_1 \mid = \mathrm{Bin}(s,p)$ in distribution, and for $i > 1$ we have $\mid T_i \mid = \mathrm{Bin}(s,p)$ in distribution for a suitable $s \leq d-1$. (Here $\mathrm{Bin}(s,p)$ is the binomial distribution with parameters s and p). Hence, we get:

$\Pr[$The algorithm generates sets$T_1, T_2, ..., T_t$ with
$\mid T_1 \mid \geq 1, \mid T_1 \mid + \mid T_2 \mid \geq 2, \ldots,$ and $\mid T_1 \mid + \cdots + \mid T_t \mid \geq t]$
$\leq \Pr[1 + Y_1 + \cdots + Y_t \geq t]$

$= \Pr[Y \geq t-1]$,

where $Y_i = \mathrm{Bin}(d-1, p)$ and $Y = \mathrm{Bin}(t(d-1), p)$ in distribution. For the expectation of Y we have

$$\mathrm{EY} = t(d-1)p = t(1-\varepsilon)$$

where ε is a constant depending on p. We will apply the Chernoff bound

$$\Pr[Y > (1+\beta)\mathrm{EY}] < \exp(-\frac{\beta^2}{3}\mathrm{EY})$$

valid for any $\beta > 0$. We have

$$t - 1 = (1+\beta)\mathrm{EY} \geq (1+\beta')\mathrm{EY}$$

with

$$\beta = \frac{t-1}{t(1-\varepsilon)} - 1 \geq \frac{1}{1-\varepsilon'} - 1 = \beta' > 0$$

where $\varepsilon' = \varepsilon/2$ (and t is sufficiently large). Finally, let

$$t = t(n, \varepsilon) = \frac{3}{(\beta')^2}\frac{1}{1-\varepsilon}2\ln n,$$

then

$$\Pr[\mathrm{Y} \geq t-1] \leq \Pr[\mathrm{Y} \geq (1+\beta')\mathrm{EY}] \leq \exp(-\frac{\beta'^2}{3}\mathrm{EY}) \leq \frac{1}{n^2}$$

If $H = (V, F)$ has n nodes, we get theorem 1:

Theorem 1. The probability of those faulty versions $G = (V, E)$ of H which have a component with $\geq t(n, \varepsilon)$ nodes is $O(1/n)$. That is *almost all* faulty versions G are such that *all* components have $O(\ln n)$ nodes, the constant factor in the O-notation depending on the probability p.

The considerations above apply also if ε is fixed,

$$d = d(n) \to \infty \text{ and } p = p(n) \leq \frac{1-\varepsilon}{d(n)}$$

because if

$$p(n) \leq \frac{1-\varepsilon}{d(n)}, \text{ then } p(n) \leq \frac{1-\varepsilon'}{d(n)-1}$$

for a suitable $\varepsilon > \varepsilon' > 0$ and our proof holds. If $H = (V, E)$ is the complete graph on n nodes, this is one half of the giant component theorem (cf. [AS 92], p. 154).

2 Above the threshold probability

For the whole section we fix the following parameters.

- The degree d of our graphs. We assume $d \geq 3$.
- The probability that a given edge is *not* faulty:

$$p = \frac{\lambda}{d-1} > \frac{1}{d-1}, \text{ hence } \lambda > 1$$

Following [Ni et al. 94] we let G_n^d be the probability space of all regular graphs with n nodes such that the degree of each node is d. Each graph is equally likely. The probability space of faulty random regular graphs is constructed by 2 subsequent probabilistic experiments: First, choose a graph from G_n^d, second, delete each edge from this graph with probability $f = 1 - p$. We call this probability space $G_{n,p}^d$. It is important to understand that the probability of 2 faulty graphs with the same number of edges need not be the same in this case. It is helpful to think of the space of faulty graphs as consisting of pairs (G, H) where H is a subgraph of G. Then 2 pairs where the H's have the same number of edges are equally likely.

In the sequel we prove the following theorem

Theorem 2. Let $d \geq 3$ be fixed. Almost all faulty random regular graphs have a connected component with at least $\varepsilon \cdot n$ nodes. The constant $\varepsilon = \varepsilon(\lambda)$ depends only on λ, not on d (and of course not on n).

To prove the theorem, it helps to consider the probability space of *configurations with egde faults* as introduced in [Ni et al. 94]. Let $W_1, W_2, \ldots, W_n$ be disjoint d-element sets, we call them *classes*. A *configuration* F with vertex set $W = \bigcup_{i=1}^{n} W_i$ is a *partition* of W into 2-element sets called edges (cf. [Bo 85]). Note that dn must be even. The configuration F induces a multigraph (i. e. a graph with loops and multiple edges) $\varphi(F) = (V, E)$ by considering each class as 1 element of V: $V = \{W_1, \ldots W_n\}$. The edges are given by: $(W_i, W_j) \in E$ iff there is an edge $(w, u) \in F$ where $w \in W_i,\ v \in W_j$.

We consider configurations as drawn from a probability space of configurations with each configuration being equally likely. There are $(dn-1)(dn-3) \cdot \ldots \cdot 3 \cdot 1$ configurations. From [Bo 85] we know that the probability space of those configurations F for which $\varphi(F)$ is a d-regular graph (i. e. *without* loops and multiple edges) is $\geq \delta > 0$, where δ depends on d. Moreover, each $G \in G_n^d$ is hit via φ by the same number of configurations: $(d!)^n$. Hence, if a property holds for almost every configuration, the corresponding property holds for almost all d-regular graphs, too.

Definition 3. Let $W_1, \ldots W_n$ be as above n disjoint d-element classes. A random configuration with edge fault probability $f = 1 - p$ is obtained by the following stochastic experiment: First, choose a configuration F from the probability space of all configurations. Second, delete each edge of F independently with the fault probability $f = 1 - p$.

Note that 1 faulty configuration can be obtained from seneral $F's$. A faulty configuration obtained by the second step from a given F is called a faulty version of F. Moreover each faulty configuration with k edges has the same probability. Nevertheless it may sometimes be helpful to look at the space of faulty configurations as a space of pairs as in the case of faulty random regular graphs described above.

Each configuration with edge faults induces a multigraph with edge faults. In particular we get [Ni et al. 94]:

Lemma 4. If a property holds for almost every configuration with edge faults the corresponding property holds for almost every graph from $G^d_{n,p}$.

Accordingly, to prove theorem 2 we prove the analogous result for random configurations with edge faults: For d and p as fixed almost every faulty configuration has a connected component consisting of at least εn classes.

Our proof makes use of the basic technique as introduced in [AjKoSz 82] for the faulty d-dimensional cube. This technique has been applied to the faulty butterfly network in [KaNeTa 94]. We need to fix 2 additional parameters:

– A probability p' with

$$p > p' = \frac{\lambda'}{d-1} > \frac{1}{d-1}\text{ , hence } \lambda > \lambda' > 0$$

.

– And

$$p'' = 1 - \frac{1-p}{1-p'} > 0.\text{ We let } \delta = \delta(\lambda) = \lambda - \lambda'.$$

Note that $(1-p')(1-p'') = 1-p$ and $p'' = p''(\lambda, d) \geq \frac{\delta}{d}$.

We consider a random configuration with fault probability $f = 1-p$ as being obtained by 2 independent stochastic experiments:

– Experiment 1: We choose a configuration F. Then we construct a configuration F' with fault probabilty $f' = 1-p' > 1-p$ by throwing each edge of F independently with probability p'.
– Experiment 2: We throw each edge of F with probability p'' into F' to get F''. In F'' each edge of F is absent with probability $(1-p')(1-p'') = 1-p$, hence F'' is a random configuration with fault probability $f = 1-p$.

Our proof follows these 2 experiments above:

– Step 1: With high probability the following holds in F': We have at least $\varepsilon' n$ classes which are distributed over at most $\varepsilon' n/\omega(n)$ connected components. (Note that this is weaker than saying $\varepsilon' n$ classes belong to components of size at least $\omega(n)$ each).We call these components *atoms*. The funtion $\omega(n)$ is fixed just before 8, $\varepsilon' = \varepsilon'(\lambda)$ is fixed only after lemma 11. At the moment we only note that $\omega(n)$ is a function independent of d and going (slowly) to infinity. The constant ε' is a constant depending on λ only.

– Step 2: With high probability the edges thrown in by the second experiment induce paths such that the following holds: The atoms which are connected by these paths yield a component having at least εn classes. Again $\varepsilon = \varepsilon(\lambda)$ is fixed further below.

We prove step 2 first. Not surprisingly the proof relies on expansion properties of almost all random configurations [Bo 88] when $d \geq 3$:

Fact 5. There is a constant (independent of d and n) $c > 0$ such that for a random configuration F the following holds with high probability:
(a) For all subsets X of classes we have:

$$\mid \partial X \mid \geq c \cdot d \cdot \mathrm{Min}\{\mid X \mid, \mid \mathrm{Cpl}\ X \mid\}$$

where ∂X is the set of edges of F connecting a class from the set of classes X with a class from Cpl X, the complement of X.
(b)For all X we have

$$\mid \mathcal{N}(X) \mid \geq c \cdot \mathrm{Min}\{\mid X \mid, \mid \mathrm{Cpl}\ X \mid\},$$

where $\mathcal{N}(X)$ is the set of classes adjacent to X in F but not belonging to X.

Conditioning on the set of those configurations F which satisfy the following lemma yields step 2:

Lemma 6. For almost every configuration F holds: Let F' be a faulty version of F such that $\varepsilon' n$ classes belong to $\varepsilon' n/\omega(n)$ atoms. In the probability space of faulty versions of F obtained by throwing each edge of F with probability p'' into F' the following holds with high probability: The newly thrown in edges induce paths which merge at least one third of all classes in atoms (i. e. $(1/3) \cdot \varepsilon' n$ classes) into 1 connected component.

Proof. If the claim of the lemma does not hold, we can partition the set of atoms into 2 disjoint subsets of atoms A and B each containing $\geq \frac{1}{3} \cdot \varepsilon' n$ classes such that no path (of new edges) goes from A to B. We assume for the moment, that for each partition $A,\ B$ as above the following holds: In the *original* configuration F there exist at least $a \cdot n$ edge-disjoint paths between A and B. Moreover each of these paths is of length $\leq b$. Here $a,\ b$ are constants independent of n. For $A,\ B$ fixed we get:

Pr[A and B are not connected by new paths]
$\leq$ Pr[Each of the $a \cdot n$ edge-disjoint paths is not present]
$\leq (1 - p''^b)^{an}$.

Let the random variable X be given by: X = the number of partitions A, B such that none of these $a \cdot n$ paths is present. For the expectation of X we get:

$$\mathrm{EX}\ \leq 2^{\frac{\varepsilon' n}{\omega(n)}} \cdot (1 - p''^b)^{an} \leq 2^{\frac{\varepsilon' n}{\omega(n)}} \cdot \exp(-p''^b an) = o(1)$$

as the number of partitions of the atoms is $\leq 2^{\frac{n}{\omega(n)}}$, $\omega(n)$ is unbounded and $p''^b a$ depends on d, λ but not on n.

We condition on the almost-sure event, that F satisfies the expansion properties from 5. This implies that the required edge disjoint paths exist: For A, B as above, the number of classes in A is $\geq \frac{1}{3} \cdot \varepsilon' n$. The same applies to B. Now, let $A_0 = A$ and $A_{i+1} = A_i \cup \mathcal{N}(A_i)$, with $\mathcal{N}(A_i)$ denoting the set of neighbours of A_i in F (cf. 5). The expansion properties of F imply that there is a constant $j = j(d, \lambda)$, independent of n such that

$$\mid A_j \cap B \mid \geq \frac{1}{2} \mid B \mid \geq \frac{1}{6} \varepsilon' \cdot n$$

(details follow further below). As the number of classes whose distance from a given class is $\leq j$ is $\leq d^{j+1}$ we must have a linear number of edges (w, u) over which B is entered in the construction of A_j. That is for these edges we have a path $(u_0, u_1)(u_1, u_2), \ldots (u_i, w)(w, u)$ with $u_0 \in A$, $u_1 \in A_1 \setminus A, \ldots u_i \in A_i \setminus A_{i-1}$, $w \in A_{i+1} \setminus A_i$, $u \in B$, and $i + 1 \leq j$ such that u is the first vertex on this path belonging to B. Accordingly we have a linear number of such paths. These need not be edge-disjoint. Now consider an arbitrary edge $e = (x, y)$ from such a path. Then e can take part in at most $j \cdot d^{j-1}$ such paths. As each path consists of at most j edges it can have edges in common with at most $j^2 \cdot d^{j-1}$ paths. As j, d are constants independent of n we must have a linear number of edge-disjoint paths.

We still have to show the existence of the constant $j = j(d, \lambda)$ used above: From fact 5 we get: As long as

$$\frac{1}{6} \varepsilon' n \leq \mid A_i \mid \leq (1 - \frac{1}{6} \varepsilon') n$$

we get $\mathcal{N}(A_i) \geq c \frac{1}{6} \varepsilon' n$. Hence the smallest j such that

$$\frac{1}{3} \varepsilon' n + c \frac{1}{6} \varepsilon' n \cdot j(d, \lambda) \geq (1 - \frac{1}{6} \varepsilon') n$$

is certainly a constant satisfying our needs. $\square$

Hence step 2 of the proof of theorem 2 is achieved by letting $\varepsilon = \frac{1}{3} \varepsilon'$.

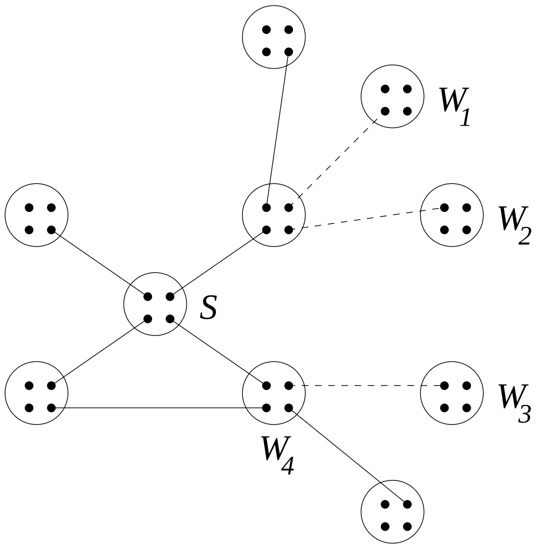
Outline of step 1 of the proof of theorem 2: The existence of the $\varepsilon' \cdot n$ classes and the corresponding atoms is shown by running sufficiently many breadth-first searches on disjoint parts of a faulty configuration. We restrict the number of search steps of each search to $\omega(n)$, a slowly growing function defined below. The choice of $\omega(n)$ implies that the following holds with high probability for *all* breadth-first searches run on 1 faulty configuration: Each search hits at most 1 class 2-times. This allows us to restrict attention to those parts of each search, where no overlapping occurs. These parts have a tree-like structure and therefore behave like branching processes.

The following fact 7 (a) is proved in [AS 92] [Fe 68], (b) can be found as claim 3 in [AjKoSz 81].

Fact 7. We consider a branching process with offspring distribution $\mathrm{Bin}(m, p)$, the binomial distribution where m is the number of trials and p the probability of success.
(a) If the expectation $m \cdot p$ of the offspring distribution is > 1, then the probability that the process dies after a finite number of generations is < 1.
(b) We even have the following uniform version of (a): If $\lambda_0 > 1$ is fixed, there exists a $\mu = \mu_0(\lambda_0) < 1$ such that *for all* m, p with $m \cdot p = \lambda_0$ the probability that the process dies is $\leq \mu_0(\lambda_0) < 1$

So the idea is to run disjoint breadth-first searches each performing at most $\omega(n)$ search steps until altogether a linear number of classes has been detected. Fact 7 (and some tail bounds) ensure that a constant fraction of these classes belong to sufficiently few components.

But here, an additional point arises, as can be seen from the following figure. It shows a typical substructure discovered by a breadth- first search starting at the class S.(We assume $d = 4$, $\omega(n) = 9$). Faulty edges are drawn as dotted lines, class W_4 is hit 2-times:



Interior classes are connected to S via non-faulty paths. Some technical problem is caused by the classes like W_1, W_2, W_3, which are detected by a faulty edge. The problem is that they do not belong to the connected component of S, nevertheless we do not want them to be considered by subsequent searches. Otherwise we would destroy the independence (and hence analyzability) of the various searches. Moreover, the number of classes found by a faulty edge can only be bounded (in the very worst-case) as being d-times the number of interior classes. If we would only use interior classes for our atoms, their number could be a fraction as low as $1/d$ of all classes looked at in our searches: The size of the component to be constructed would decrease in d.

Fortunately, we can allow each class to be hit by a constant fraction of d faulty edges, before it becomes a *bordering* class not to be considered any more by subsequent searches. This still ensures that with high probability a constant fraction of n classes are interior classes of sufficiently few components.

The preceding discussion refers to the probability space of configurations with fault probability $f' = 1 - p'$ where $p' = \frac{\lambda'}{d-1} > \frac{1}{d-1}$. We fix some more parameters:

- We let $\gamma = \lambda' - 1$, w.l.o.g. $\gamma < 1/2$ and

  $$\beta = \beta(\lambda, d) = \lfloor \frac{\gamma}{2} \cdot (d-1) \rfloor.$$

  (A class becomes a bordering class only after being hit by $\beta + 1$ faulty edges and no non-faulty one.)
- The maximum number of bordering or interior classes discovered by our searches is bounded above by $\hat{\varepsilon} n$ where $\hat{\varepsilon}$ is specified by

  $$\hat{\varepsilon} = \hat{\varepsilon}(\lambda) = \frac{1}{10} \frac{\gamma}{2} (1 - \gamma)$$
- The offspring distribution induced by expanding a class inside one of our branching processes turns out to be stochastically at least as large as

  $$\mathrm{Bin}(d - 1 - \beta\,,\, p' \cdot (1 - \frac{\hat{\varepsilon}}{1 - \hat{\varepsilon} - \frac{\gamma}{2}})).$$

  The expectation of this random variable can be calculated to be $\geq \lambda_0(\lambda) > 1$. We let

  $$\mu \geq \mu_0(\lambda) > 0$$

  be the probability that a branching process with the offspring distribution as above does not go extinct (cf. fact 7). To simplify notation we let

  $$\alpha = \alpha(\lambda) = 1 - \hat{\varepsilon} - \frac{\gamma}{2}.$$
- The number of interior or bordering classes identified by each search is bounded above by $\omega(n) = \log n$.

Algorithm 8. As in the proof of theorem 1 we formulate our breadth-first searches as breadth-first generation algorithms. We use the following global variables:

- E = the set of non-faulty edges generated
  $\neg E$ = the set of faulty edges generated.
- Free = the set of *vertices* which have not at all been looked at , that is which do not occur in edges from $E \cup \neg E$.
  $\neg$Free = the set of vertices which have already been looked at.
- Dis = the set of bordering or interior *classes* which have already been discovered.

The next 3 variables are local to each single breadth-first search:

- Atom = the set of classes discovered during the current breadth-first search. A class is discovered during a breadth-first search if it is hit by a non-faulty edge (interior class) or by the $\beta+1$'th faulty edge (including the edges from previous searches) and has not yet been hit by a non-faulty edge (bordering class).
- Open = the set of classes in the current search front.
- Open' = the subset of classes already generated from the next search front.

A single breadth-first generation starting from the class S:

Procedure Bfg(S)

1. Open := $\{S\}$; Atom := $\{S\}$; Open' := $\emptyset$
2. while Open $\neq \emptyset$ do
3. Pick a class T from Open according to some deterministic rule. (The class T will depend on the history of the computation.)
4. while $T \cap$ Free $\neq \emptyset$ do
5. Pick a vertex x from T deterministically.
6. Choose the vertex y from Free$\setminus\{x\}$ according to the uniform distribution.
7. Add the edge (x, y) either to E or to $\neg E$. To E with probability p', to $\neg E$ with probability $1-p'$.
8. If the class of y, we call it R, has been discovered before (i. e. $R \in$ Atom or $R \in$ Dis , see below) we do nothing with R. If however R has not been discovered before, we add R to the set Atom if $(x, y) \in E$. If however $(x, y) \notin E$ we add R to Atom only if the edge (x, y) is the $\beta+1$'th faulty edge to hit R. We add R only to Open' if $(x, y) \in E$.
9. if $|$ Atom $| = \omega(n)$ then exit (from the current call to Bfg(S)).
10. if Open $= \emptyset$ then Open := Open'.

Note that Atom includes the bordering classes found. The procedure Gen (for generate) directs the iterated execution of Bfg(S):

Procedure Gen

1. for $i = 1$ to $\hat{\varepsilon}n/\omega(n)$ do
   (The maximal number of interior or bordering classes generated by 1 Bfg(S) is $\omega(n)$.)
2. Pick a class S not from Dis deterministically.
3. Bfg(S)
4. Dis := Dis $\cup$ Atom.

Of course, initially Dis $= \emptyset$.

It is helpful, to visualize the possible computations of Gen as a probability tree: Each node corresponds to a probabilistic choice, the sons of a node represent the possible outcomes of a choice. A possible computation is represented by a path from the root to a leaf. Subsequent choices are independent and the probability of a computation is the product of the probabilities of all the choices made.

We relate the generation procedure to the space of random configurations with fault probability $f' = 1 - p'$: The probability of the algorithm to generate a given set E is equal to the probability of the set of faulty configurations with edge set E. This can be seen from the following observations:

– If E and $\neg E$ are generated by Gen, they can be generated in exactly 1 way by Gen. (Proof: inductively on $\mid E \cup \neg E \mid$.)
– If E and $\neg E$ can be generated by Gen, then

  Pr[Gen generates E and $\neg E$]
  $=$ Pr$\{\hat{E} \mid$ There exists a configuration $F \supseteq E \cup \neg E$ such that $\hat{E}$ is a faulty version of F which has all edges from E and none from $\neg E\}$,

  where the preceding probability refers to the probability space with fault probability $f' = 1 - p'$.

For the subsequent discussion we collect the obvious:

– The number of classes in the set Dis $\cup$ Atom is always $\leq \hat{\varepsilon} \cdot n$.Each class not in Dis $\cup$ Atom can be hit by up to β faulty edges. Accordingly, the number of nodes x which have been touched by the algorithm (i. e. $(x, x') \in E \cup \neg E$ for any x') is always

$$\leq \hat{\varepsilon} \cdot n \cdot d + (1 - \hat{\varepsilon})n\beta \ < \ nd(\hat{\varepsilon} + \frac{\gamma}{2})$$

  Accordingly always

$$\mid \text{Free} \mid \geq nd - (\hat{\varepsilon} \cdot n \cdot d + (1 - \hat{\varepsilon})n\beta) \ > \ nd\alpha$$

– Whenever a choosing step in line 6 of Bfg(S) is executed, we have $\mid Atom \mid < \omega(n)$. Hence, whenever a y is chosen in line 6, the number of free nodes belonging to classes from Atom is certainly $< \omega(n) \cdot d$.

– The number of choosing steps in 1 execution of Bfg(S) is certainly $\leq \omega(n)d$, as each class can be hit at most d-times.

Now conditional on any arbitrary previous history $\mathcal{H}$ the probability that a node belonging to a class from Atom is chosen in line 6 is equal to

$$\frac{\text{number of untouched nodes belonging to Atom}}{\mid \text{Free} \mid} \leq \frac{\omega(n) \cdot d}{n d \alpha}$$

As each execution of Bfg(S) makes at most $\omega(n)d$ choosing steps the probability that in 1 given call of Bfg(S) 2 or more nodes belonging already to Atom are hit is at most

$$\binom{\omega(n)d}{2} \cdot (\frac{\omega(n)d}{nd\alpha})^2 = o(n^{-1})$$

As Bfg(S) is certainly called *not* more that $nd-$times in a single computation, we get that in almost all computations *no* call of Bfg(S) has 2 or more overlaps. In the following we restrict (i. e. condition) every probability to the set of these computations.

Lemma 9. The probability that a given call of Bfg(S) ends with $\omega(n)$ classes in Atom is at least $\mu_0(\lambda) \cdot \nu(\lambda) > 0$.

Proof. First, we look at the beginning of an execution of Bfg(S). If after finishing S (i. e. after finishing the while-loop from line 4 of Bfg(S) with the class S) $k \geq 3$ non-faulty edges go from S to new classes (i. e. classes not already in Dis), we know that at least $k-2$ of these classes C have the following property: The interior or bordering classes discovered in the branch starting with C do not take part in any kind of overlapping, as far as the present search is concerned. It is not difficult to see that the probability that $k \geq 3$ non-faulty edges leave the class S is bounded away from 0 uniformly in λ (i. e. the lower bound does not depend on the degree d): The probability that a fixed choice of an y in line 6 hits the class S itself or a class from Dis is at most $\frac{\hat{\varepsilon}}{\alpha}$. Hence the probability that a fixed choice yields a new (i. e. not from Dis) interior class (i. e. class reachable by a non-faulty edge) is at least $p' \cdot (1 - \frac{\hat{\varepsilon}}{\alpha})$. As S may have been hit by up to β faulty edges from previous searches, the distribution of the number of non-faulty edges from S to new classes is stochastically at least as large as $\text{Bin}(d - \beta, p'(1 - \frac{\hat{\varepsilon}}{\alpha}))$. The probability that a random variable with this distribution is 0 is bounded above by $\exp(-\lambda'(1 - \frac{\hat{\varepsilon}}{\alpha})$ depending only on λ. An analogous result holds for the probability of the distribution being 1 of 2 and we get: The probability that the call Bfg(S) finishes the class S with $k \geq 3$ non-faulty edges going from S to new classes is $\geq \nu = \nu(\lambda) > 0$. As each Bfg($S$) has only 1 (or 0) overlap we have with probability ν at least 1 *tree* connected to S by a non-faulty edge. Let T be a class from this tree which gets expanded by Bfg(S) (i. e. with which the loop in line 4 is entered). Provided $\mid$ Atom $\mid$ does not surpass $\omega(n)$, the number of sons of T contributing to the next search front (the "offspring" of T) has a distribution stochastically at least as large as $\text{Bin}(d - 1 - \beta, p'(1 - \frac{\hat{\varepsilon}}{\alpha}))$. The expectation of this random variable is at least $\lambda_0(\lambda) > 1$. Hence the interior classes of this tree behave like a branching process with offspring distribution stochastically at least as large as $\mathrm{Bin}(d-1-\beta, p'(1-\frac{\hat{\varepsilon}}{\alpha}))$. Fact 7 implies that $\mathrm{Bfg}(S)$ does not come to an end of the tree with probability $\geq \mu_0 = \mu_0(\lambda) > 0$. Hence $\mathrm{Bfg}(S)$ discovers $\omega(n)$ classes with probability at least μ_0. Altogether we get 2.8. □

Lemma 10. The number of classes discovered in a set Atom of size $\omega(n)$ is at least $\frac{1}{4}\mu_0 \cdot \nu \cdot \hat{\varepsilon} \cdot n$. with high probability.

Proof. Each possible computation of Gen has exactly $t = t(n) = \hat{\varepsilon} \cdot n/\omega(n)$ calls of $\mathrm{Bfg}(S)$. For $1 \leq i \leq t$ let X_i be the indicator random variable of the event "the i'th call of $\mathrm{Bfg}(S)$ ends with $\mid Atom \mid = \omega(n)$". By lemma 9 $\mathrm{X} = \mathrm{X}_1 + \ldots + \mathrm{X}_t$ is at least as large as $\mathrm{Bin}(t, \mu_0 \cdot \nu)$, the binomial random variable with t trials and probability $\mu_0 \cdot \nu$ of success. Let $\sigma = \mu_0 \cdot \nu$.

We have $\mathrm{EX} \geq t \cdot \mu_0 \cdot \nu$ and by tail bounds

$$\Pr[\mathrm{X} < \frac{1}{2} \cdot \sigma \cdot t] \leq \exp(-\Omega(\sigma t)) = \exp(-c \cdot \frac{\hat{\varepsilon n}}{\omega(n)})$$

with $c = c(\hat{\lambda})$. Hence, with high probability the number of calls of $\mathrm{Bfg}(S)$ at the end of which $\mid \mathrm{Atom} \mid = \omega(n)$ is at least $\frac{1}{4} \cdot \sigma \frac{\hat{\varepsilon} n}{\omega(n)}$. As the different sets Atom are disjoint, at least $\frac{1}{4} \cdot \sigma \hat{\varepsilon} n$ classes belong to atoms with at least $\omega(n)$ classes (with high probability). □

Due to the existence of the bordering classes, the different sets Atom are *not* necessarily connected subgraphs. We condition the remaining probabilities on the almost sure event that at least $s(n) = s(n, \lambda) = \frac{1}{4}\mu_0 \nu \hat{\varepsilon} n$ classes belong to sets Atom with $\omega(n)$ classes. As each bordering class must be hit by a constant fraction of d faulty edges, we have not too many of those:

Lemma 11. There is a constant $\kappa = \kappa(\lambda)$ such that with high probability at least $(1 - \kappa) \cdot s(n)$ classes are distributed over at most $s(n)/\omega(n)$ connected subgraphs.

Proof. Outline of the argument: For any i with $1 \leq i \leq \hat{\varepsilon} n$ the following holds: Conditional on the set of those computations of the algorithm Gen for which the i'th choosing step (i. e. line 6 of $\mathrm{Bfg}(S)$) takes place, the probability that in this step a bordering class is discovered is at most

$$(1 - \frac{\lambda}{d-1})^{\beta+1} \leq (\frac{1}{e})^{\gamma/2}$$

the upper bound being independent of d. For subsequent choosing steps the respective events are not independent, but even negatively correlated. Hence, considering the $s(n)$ choosing steps which put a class into a set Atom which will be filled with $\omega(n)$ classes, the distribution of the number of bordering classes found is majorized by $\mathrm{Bin}(s(n), (\frac{1}{e})^{\gamma/2})$. Let $0 < \tau = \tau(\lambda)$ be small enough such that $\kappa = \frac{1}{e}^{\gamma/2} + \tau < 1$ and let $\mathrm{X} = \mathrm{Bin}(s(n), (\frac{1}{e})^{\gamma/2})$ in distribution. Then

$$\Pr[\mathrm{X} > ((\frac{1}{e})^{\gamma/2} + \tau) \cdot s(n)] \leq e^{-\Omega(n)}.$$

Hence with high probability at least $(1-\kappa) \cdot s(n)$ classes of our sets Atom in question are discovered by non-faulty edges and the lemma follows. □

Finally setting

$$\varepsilon' n \;=\; (1-\kappa) \cdot \frac{1}{4} \cdot \mu_0 \cdot \nu \cdot \hat{\varepsilon} \cdot n$$

yields step 1 of the proof of theorem 2.

References

[AS 92] Noga Alon, Joel H. Spencer. *The Probabilistic Method.* Wiley, 1992.

[AjKoSz 81] M. Ajtai, J. Komlos, E. Szemeredi. The Longest Path in a Random Graph. *Combinatorica*, 1, 1981, 1-12.

[AjKoSz 82] M. Ajtai, J. Komlos, E. Szemeredi.Largest Random Component of a k-Cube. *Combinatorica*,2(1), 1982, 1-7.

[Bo 85] Bela Bollobas. *Random Graphs.* Academic Press, 1985.

[Bo 88] Bela Bollobas. The Isoperimetric Number of Random Regular Graphs. *European Journal of Combinatorics.* 9, 1988, 241-244.

[KaNeTa 94] A. Karlin, G. Nelson, H. Tamaki. On the Fault Tolerance of the Butterfly. In *Proceedings STOC 1994*, 125 -133.

[Ni et al. 94] S. Nikoletseas, K. Palem, P. Spirakis, M. Yung. Short Vertex Disjoint Paths and Multiconnectivity in Random Graphs: Reliable Network Computing. In *Proceedings ICALP 1994.* 508-519.

[NiSp 95] S. Nikoletseas, P. Spirakis. Expander Properties in Random Regular Graphs with Edge Faults. In *Proceedings STACS 1995.* 422-423.

[Ke 80] H. Kesten. The Critical Probability of Bond Percolation on the Square Lattice Equals 1/2. *Communication in Mathematical Physics.* 74, 1980, 41-59.

[Ma 92] T. R. Mathies. Percolation Theory and Computing with Faulty Arrays of Processors. In *3rd SODA 1992.* 100-103.

[Ra 89] P. Raghavan. Robust Algorithms for Packet Routing in a Mesh. In *Proceedings 1st SPAA 1989.* 344-350.

[Fe 68] W. Feller. *An Introduction to Probability Theory an its Applications*, vol. 1. Wiley 1968.