Lecture 4: The effect of faults on expansion II: Random faults

11th, 16th and 17th September 2008

4.1 Introduction: The Random fault model

Until now we were talking mostly about faults which are caused by some *adversary* who knows about the network, so he/she affects the network in the worst way. It has been said for a long time that computer science treats the worst case scenarios, which happen rarely, so these results are not practically much useful. But here we will be studying random faults exploring highly probable events (i.e. more frequently occurring scenarios). Note that, in this lecture *node* and *vertex* are used interchangeably.
The model. Each node of the network can become faulty with probability p independent of all other nodes. More formally let $G = (V, E)$ be a graph corresponding to network under study. For each vertex $v \in V$, let us define a random variable X_v s.t.

$$P(X_v = 1) = 1 - p$$

and

$$P(X_v = 0) = p$$

where $0 \leq p \leq 1$. $X_v = 0$ means node v is faulty. A faulty node is removed from network with all of its edges. Let $W = \{v \in V : X_v = 1\}$ and G_f is the graph induced by W. Although the model is an ideal one in the sense that fault probability i.e. p, for all nodes is equal and independent of all other nodes, but it is a good start to understand the fundamentals.

4.2 A lower bound on the fault probability for graph survival

Intuitively it appears that in general this situation might be easier to handle since there is no malicious adversarial intent behind the distribution of node failures. But, in general this is not true. We begin this section by showing that there are families of graphs for which a fault probability of $\Theta(\alpha)$ causes the graph to disintegrate into sublinear fragments, where α is the node expansion of the graph. In other words, in these graphs $\Theta(\alpha n)$ random node failures can be catastrophic: they don't even allow us to find a linear sized connected component, hence making it impossible to find a linear sized connected component with good expansion. Let's first prove a *lemma* that will be used in the proof of main theorem of this section.

Lemma 4.1 *Let $G = (V, E)$ be a graph with maximum vertex degree δ. Then $n\delta^{2(r-1)}$ is an upper bound on the number of connected subgraphs of G induced by exactly r vertices where $n = |V|$.*

Proof. We know that there is atleast one spanning tree with $r - 1$ edges for a connected subgraph with r vertices. And if the subgraph is an induced one then no two different subgraph can have same spanning tree as they will have different vertex set and one vertex set induces exactly one subgraph. Also for every tree with $r - 1$ edges there is atleast one rooted tree which in turn is characterized by *Eulerian tour* of length $2(r - 1)$, in which each edge is used atmost twice, starting at root of that tree and a *Eulerian tour* can be seen as a special *random walk* which starts and ends at the same vertex.
Let N_1^r = number of connected subgraphs induced by exactly r vertices and containing a given vertex $v \in V$, and N_2^r = number of spanning trees of size r rooted at same vertex v, then

$$N_1^r \leq N_2^r \tag{1}$$

Let N_3^r = number of *random walks* of size r starting at vertex v, then

$$N_2^r \leq N_3^{2(r-1)} \tag{2}$$

Also during *random walk* at any step we have atmost δ choices to select next vertex so

$$N_3^r \leq \delta^r \tag{3}$$

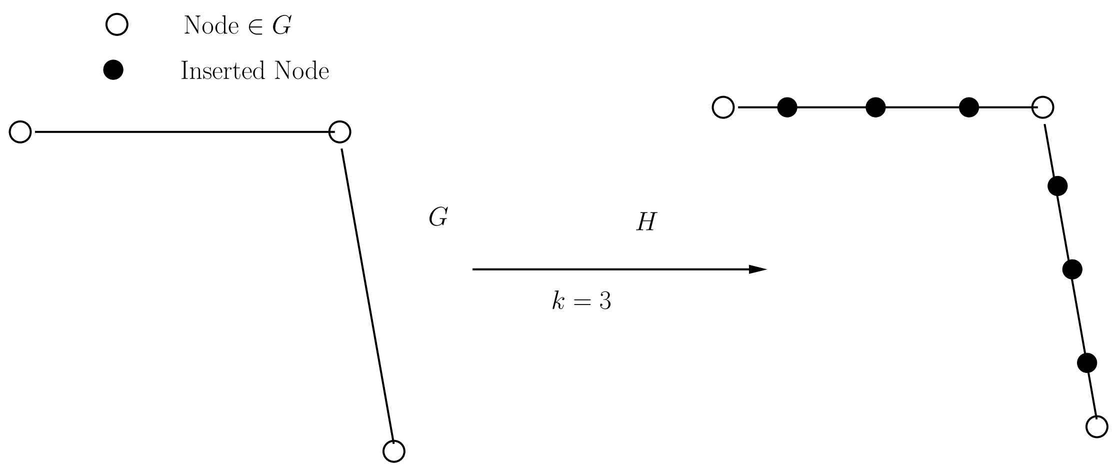
Now from (1), (2) and (3) we have

$$N_1^r \leq \delta^{2(r-1)} \tag{4}$$

Now summing over all $|V|$ vertices, number of connected subgraphs of G induced by exactly r vertices is $\leq n\delta^{2(r-1)}$. ■

Theorem 4.1 *There is a constant γ s.t. for all $\alpha < \gamma$ there is an infinite family of graphs with expansion $\theta(\alpha)$ for which fault probability $c_2\alpha$ causes that graph to break into the components of size $o(n)$ with high probability.*

Proof. We use the family of graphs constructed in the proof of theorem 3.2 of lecture 3, i.e. let $G_n = (V_n, E_n)$ be an infinite family of constant degree expanders with constant expansion γ and degree δ. Construct a graph, H, which is a copy of G(which is a member of family G_n) with each edge of G replaced by a chain of k nodes (between its two endpoints). Figure 1 illustrates the construction. From proof of that theorem in lecture 3 we know that H has expansion $\theta(1/k)$.



H is constructed by replacing each edge of G by a chain of k (=3 here) nodes.

Figure 1: Construction of H from G

Let us take a subgraph H' of H which has following properties:

1. H is *connected* and contains *exactly* r nodes of G.

2. All boundary nodes of H' are of G, i.e. $\Gamma_H(H') \subset V(G)$.

Clearly, to satisfy the second property we should construct H' in a fashion that if a node $v \in V(G)$ is in H' then all k nodes which were inserted during construction of H on any edge (of G) of v should also be in H'(see figure 2). We say that *a subgraph "survives" if none of its nodes became faulty.*

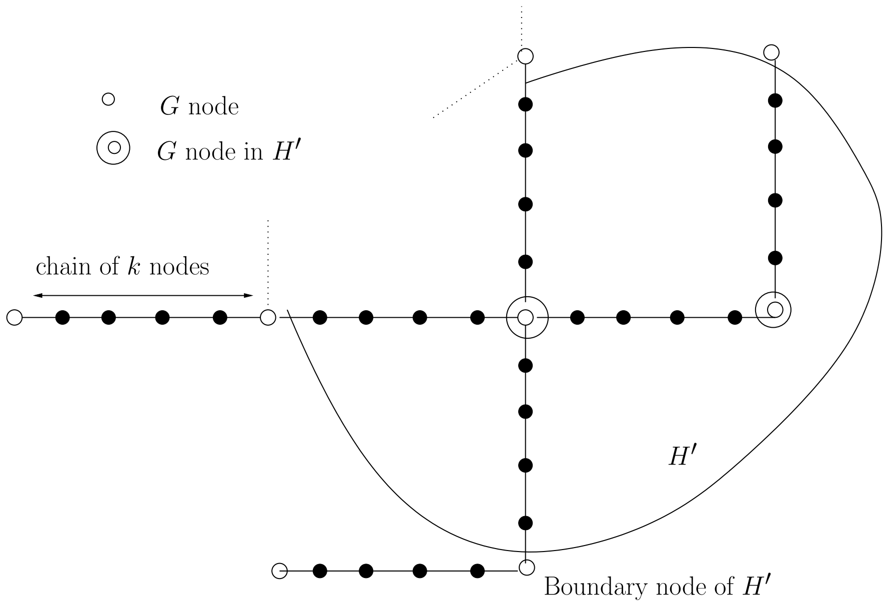


Figure 2: Choosing H' from H

Let failure probability is $p = \frac{(2\ln\delta+2)}{k}$. As we have already proved that expansion(α) of H is $\theta(\frac{1}{k})$ and δ is a constant so p is $c_2\alpha$ for some constant c_2.

We begin by seeing that:

$$|V(H')| \geq (k+1)(r-1) \tag{5}$$

Let us argue this, as H' has r nodes of G and is connected so it has atleast $r-1$ edges of G (of course, replaced by chain of k nodes) and we can associate with each edge (of G) $k+1$ nodes where k are those inserted nodes and 1 is the corresponding nodes of G. Now using *multiplicative law* of probability:

$$P[H' \text{ survives }] \;=\; (1-p)^{|V(H')|}$$

using (5) and considering $(1-p) \leq 1$,

$$\leq \;(1-p)^{(k+1)(r-1)}$$

Since $e^{-x} \geq 1 - x$, it follows that

$$\leq \quad e^{-p(k+1)(r-1)}$$
$$\leq \quad e^{-pk(r-1)}$$

Substituting $p = \frac{(2\ln\delta+2)}{k}$, we get

$$= \quad e^{-(2\ln\delta+2)(r-1)}$$

Now by definition of H' once we pick up r nodes then other one's are fixed so for every H' there is a connected subgraph of G induced by exactly r nodes and *vice-versa*. So using lemma 4.1 maximum number of such H' having exactly r nodes is $n\delta^{2(r-1)}$ where $n = |V(H)|$. So taking $r = \ln n + 1$, we get

$$\begin{aligned}
P[\exists\ a\ H' \text{ that survives}] &\leq \quad \text{Number of possible } H' \times P[H' \text{ survives }] \\
&\leq n\delta^{2(r-1)}e^{-2(2\ln\delta+2(r-1)} \\
&\leq n\delta^{2\ln n}e^{-2(\ln\delta+2)\ln n} \\
&\leq n\delta^{2\ln n}e^{\ln\delta^{-2\ln n}}e^{\ln n^{-2}} \\
&= \frac{n}{n^2} \\
&= \frac{1}{n}
\end{aligned}$$

Clearly $\theta(k\ln n)$ size components are very less probable. So we can safely say that graph is broken in components of size $o(n)$ with high probability. ■

However, the expansion of a graph is not necessarily the critical point (i.e., the point at which the graph disintegrates into components of size $o(n)$, with high probability) for all graphs. There are several important classes of graphs which can sustain a much higher fault probability and still yield a linear sized connected component with good expansion. One specific case is the mesh. In the following, we describe a general technique to quantify this higher fault probability.

4.3 An upper bound on fault probability for retaining graph expansion

Let us introduce some more notations and definitions which will be needed shortly. Let $G = (V, E)$ be a connected graph and let $U \subseteq V$ then U is called

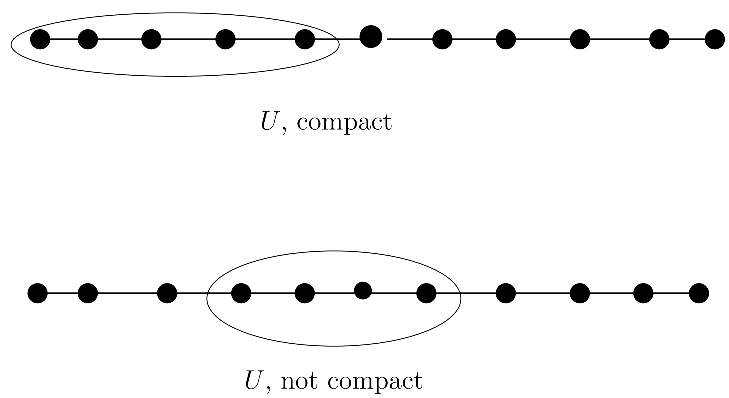


Figure 3: Illustrating the concept of compact sets

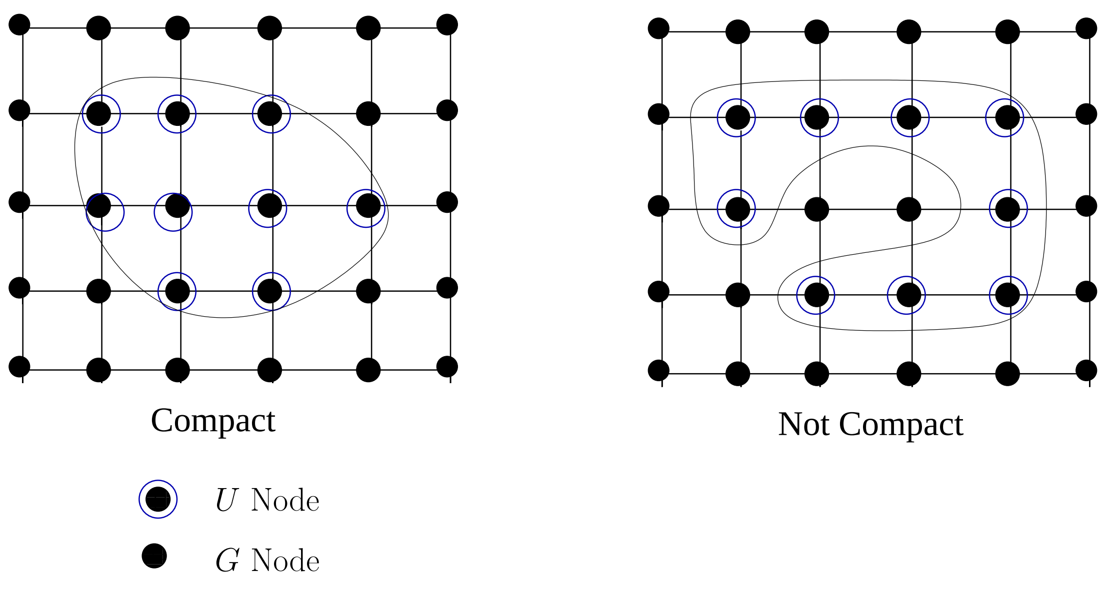


Figure 4: Illustrating the concept of compact sets

compact iff both U and $G \setminus U$ are *connected*. For illustration see figure (3) and figure (4).

Let $P(U)$ is the smallest tree which *spans* boundary of U in G, $\Gamma_G(U)$. Let $\mathcal{U}$ is the set of all *compact* subsets of $V(G)$. Then we define *span*(σ) of graph G as

$$\sigma = \max_{U \in \mathcal{U}} \left\{ \frac{|P(U)|}{|\Gamma_G(U)|} \right\}$$

We need to develop some concepts and proofs which will be used in our main theorem. Let us start with this claim.

Claim 4.2 *Let U_1 and U_2 are two compact sets which share the same boundary, i.e. $\Gamma_G(U_1) = \Gamma_G(U_2)$ then either they are same or disjoint, i.e. either $U_1 = U_2$ or $U_1 \bigcap U_2 = \emptyset$.*

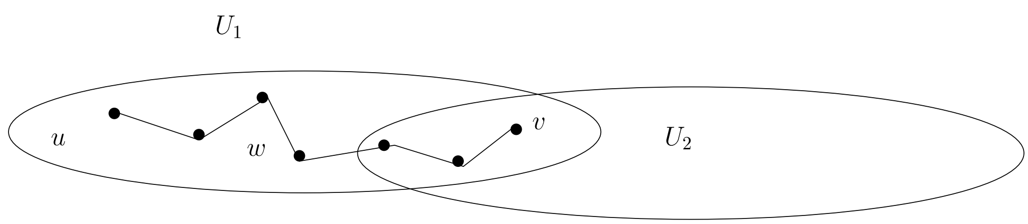


Figure 5: A path from U_2 to U_1

Proof. Let us prove by contradiction. Let $v \in U_1 \cap U_2$ and $u \in U_1 \setminus U_2$. If $U_1 \subseteq U_2$ then such a u wouldn't exist but then we can take $u \in U_2 \setminus U_1$. Notice that here $U_2 \not\subset U_1$ otherwise $U_1 = U_2$ which is contradiction so we are done already. Now if such a u exists then there is a path from v to u lying within U_1(as U_1 is compact) and starting in U_2 but ending in $U_1 \setminus U_2$ which imply there is a node $w \in U_1 \setminus U_2$ (w may be u or some other node on path $v \to u$) s.t. $w \in \Gamma_G(U_2)$. But $w \in U_1$ so $w \notin \Gamma_G(U_1)$. This shows $\Gamma_G(U_1) \neq \Gamma_G(U_2)$, which is *contradiction.* ■

Claim 4.3 *Let G be a graph with span σ, max degree $\delta \geq 3$ and let $|V(G)| = n$ then $n\delta^{3\sigma k}$ is an upper bound on number of sets compact in G with boundary size exactly k.*

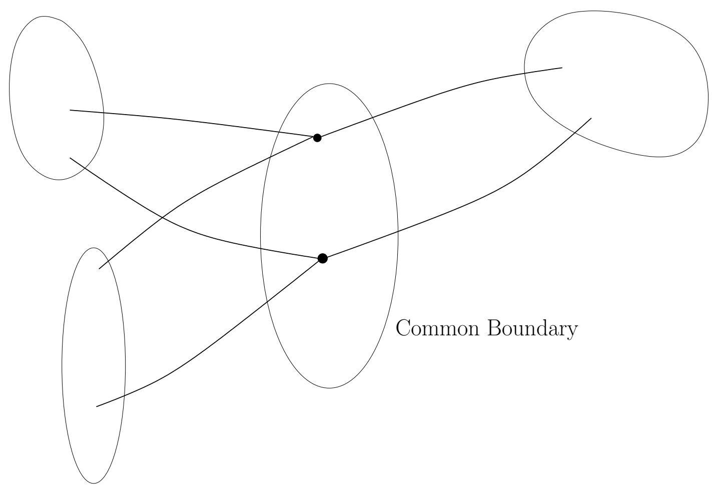


Figure 6: Compact sets sharing a common boundary

Proof. Let N_k is the number of sets compact in G with boundary size exactly k. From definition of span there exists a tree of size $\leq k\sigma$ which spans its boundary. Also let this boundary is also shared by some other compact sets. Then all those compact sets have to be disjoint (using claim 4.2) so each vertex of boundary must have at least one edge to each of those compact sets and max vertex degree is δ so atmost δ compact sets can share a common boundary(see figure 6). Lastly any spanning tree of size $\leq \sigma k$ can be extented to a tree of size σk as $G \setminus U$ is compact and a tree of size σk can cover atmost $\binom{\sigma k}{k}$ boundary sets of size k. Using all these facts,

$$\begin{aligned} N_k &\leq \binom{\sigma k}{k} \times \delta \times (\text{ Number of trees of size } \sigma k) \\ &= \delta \times n\delta^{2(\sigma k-1)} \times \binom{\sigma k}{k} \end{aligned}$$

Since $\binom{n}{k} \leq (\frac{en}{k})^k$, it follows that,

$$\begin{aligned} N_k &\leq \delta \times n\delta^{2(\sigma k-1)} \times (e\sigma)^k \\ &\leq n\delta^{2\sigma k} \times (e\sigma)^k \end{aligned}$$

Since $e\sigma \leq e^{\sigma}$ (as σ, e ≥ 1) and $e^{\sigma} \leq \delta^{\sigma}$ (as $\delta \geq 3 > e$) we have $(e\sigma)^k \leq e^{\sigma k} \leq \delta^{\sigma k}$. Thus it follows that

$$\begin{aligned} N_k &\leq n\delta^{2\sigma k}\delta^{\sigma k} \\ &= n\delta^{3\sigma k} \end{aligned}$$

■

We will also need this lemma which has not been proved here but this result can be drawn from theorem 2.10 of lecture 2. A more rigorous proof by using same concept can be found here[1].

Lemma 4.2 *Let $G = (V, E)$ be a connected graph with maximum degree δ. For any subset $S \subset V$ with $|S| \leq |V|/2$ there exist a compact set $K_G(S)$ in G whose expansion is atmost $\delta\alpha(S)$.*

Lastly, we will use this algorithm to obtain linear size components of good expansion. Let $G_f = (V_f, E_f)$ be the faulty version of G where each node is made faulty independently with probability p. An edge $(u, v) \in E$ remains in E_f if and only if both u and v are non-faulty.

Algorithm PruneCompact(ϵ):

1. $G_0 \leftarrow G_f$; $\quad i \leftarrow 0$
2. **while** $|V(G_i)| > |V(G_f)|/4$ and $\exists S_i \in V(G_i)$ s.t. $(|\Gamma_{G_i}(S_i)| \leq (\epsilon\,\delta).\alpha|S_i|$ and $|S_i| \leq |G_i|/2)$
3. $K_i \leftarrow K_{G_i}(S_i)$
4. $G_{i+1} \leftarrow G_i \setminus K_i$
5. $i \leftarrow i + 1$
6. **end while**
7. $H \leftarrow G_i$

Theorem 4.4 *Consider a graph G with max. degree δ, span σ, expansion $\alpha > (r\delta \ln^3 n)/n$ for some sufficiently large constant r and $|\Gamma_G(U)| \geq \log_\delta |U|$ for every $U \subset V(G)$ and $|U| \leq |V(G)|/2$. Then, with high probability, PruneCompact(ϵ) returns H of size $> n/4$ with expansion atleast $(\epsilon/\delta).\alpha$ provided $p \leq 1/(16e\delta^{8\sigma})$.*

Proof. Let $\mathcal{T} = G \setminus H$. Hence $\mathcal{T}$ is the union of all the culled regions. Let $T_1, T_2, \ldots, T_k$ be maximal connected components of $\mathcal{T}$.

Claim 4.5 *$\forall T_i \in \mathcal{T},\ T_i$ is compact in G_f.*

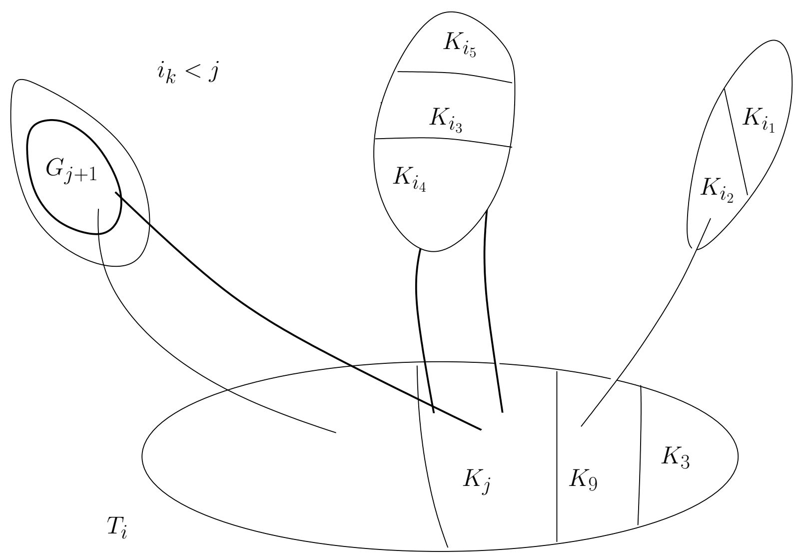


Figure 7: Various component connected through T_i

Proof. Let $K_i's$ be same as appearing in *PruneCompact* algorithm. Let a K_i share its nodes with two (or more) $T's$, say T_j and T_s, but as K_i is compact (so connected), it connects T_j and T_s. This contradicts the maximality of $T_i's$. So each K_i is fully contained within one T_j. So we can write

$$T_i = \bigcup_{j \in I_i} K_j$$

where I_i is some set of integers corresponding to T_i. Now let us prove the claim by contradiction. We remove $K_i's$ from a T_i one by one in increasing order of i (i vary on K_i) and let j is the smallest i s.t. removal of K_j breaks the remaining graph into multiple components. Also G_{j+1} was a single component after removing K_j in *PruneCompact* algorithm, so it must be contained in one of the components of $G_f \setminus \cup_{k \in I_i, k \leq j} K_k$ as we haven't removed any K_i for $i \geq j$. Also G_{j+1} was the only component left after removing K_j during pruning so all the components other than G_{j+1} must have been removed by $K_i' s$ where $i < j$, so all these remaining components consist entirely of some $K_i' s$ where $i < j$. But as just before removing K_j (here, not in pruning) these were connected through K_j(and hence through T_i) to G_{j+1} so we can extend T_i to include those K_i's, but this contradicts the maximality of T_i, which shows that T_i is compact. ■

Now let us consider the case that we have a T_i s.t. $|T_i| \leq |V(G)|/2$, so.

$$|\Gamma_G(T_i)| \geq \alpha |T_i| \tag{6}$$

also

$$\begin{aligned}
|\Gamma_{G_f}(T_i)| &\leq \sum_{K_j \in T_i} |\Gamma_{G_j}(K_j)| && (7)\\
&\leq \sum_{K_j \in T_i} \epsilon\alpha |K_j| && (8)\\
&= \epsilon\alpha |T_i| && (9)\\
&\leq \epsilon |\Gamma_{G_f}(T_i)| && (10)
\end{aligned}$$

Here same concept is applied as in lemma 3.1 of lecture 3. In last step (6) is used. Also probability of $k-l$ nodes being faulty among k nodes

$$\begin{aligned}
&= \binom{k}{k-l} p^{k-l} (1-p)^l \\
&\leq \binom{k}{k-l} p^{k-l} \\
&\leq \left(\frac{ek}{l}\right)^{k-l} p^{k-l}
\end{aligned}$$

Here we have used the fact that $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, also in our case number of faulty nodes is $(|\Gamma_G(T_i)| - |\Gamma_{G_f}(T_i)|) \geq (1-\epsilon)|\Gamma_G(T_i)|$, using (10). So this probability is

$$\binom{|\Gamma_G(T_i)|}{|\Gamma_G(T_i)| - |\Gamma_{G_f}(T_i)|} p^{|\Gamma_G(T_i)| - |\Gamma_{G_f}(T_i)|} \leq \left(\frac{ep}{1-\epsilon}\right)^{(1-\epsilon)|\Gamma_G(T_i)|}$$

if we set $\epsilon \leq \frac{1}{4}$ and use $ep \leq \frac{1}{16\delta^{8\sigma}}$

$$\begin{aligned} P[T_i \textit{ is pruned}] &\leq \left(\frac{1}{12\delta^{8\sigma}}\right)^{3/4|\Gamma_G(T_i)|} & (11) \\ &= \frac{1}{12}.\delta^{-6\sigma|\Gamma_G(T_i)|} & (12) \end{aligned}$$

Now let us see upper bound on probability of pruning a T_i of size $\geq \lceil \log_\delta n \rceil$,

$$\begin{aligned} P[\exists \ a \ T_i \ s.t.|\Gamma_G(T_i)| \geq \lceil \log_\delta n \rceil] &\leq \sum_{t=\lceil log_\delta n \rceil}^{n} n\delta^{3\sigma t}\delta^{-6\sigma t} \\ &= \sum_{t=\lceil log_\delta n \rceil}^{n} n\delta^{-3\sigma t} \\ &\leq \delta^{-3\sigma \log_\delta n} \sum_{t=\lceil log_\delta n \rceil}^{n} n \\ &\leq \delta^{-3\sigma \log_\delta n} n^2 \\ &= n^2 \delta^{(\log_\delta n)(-3\sigma)} \\ &= n^2 n^{(-3\sigma)} \\ &\leq \frac{1}{n} \end{aligned}$$

For last step notice that $\sigma > 1$(from definition). Hence we see that it is very unlikely that even order $\theta(\log_\delta n)$ sized components are pruned away, implying that it is highly probable that what we are left with after pruning is a linear size component.

The other case we have to deal with is when all the T_is have boundary smaller than $\lceil \log_\delta n \rceil$ but they add up to $> 2n/3$ nodes. We omit the proof of that case, referring the reader to [1].

■

Here we see that there is an infinite family of graphs having large probability of retaining linear size components under random faults, of course given some bounds on various parameters. But are there any familiar graphs having this kind of robustness? Yes, the d-dimensional mesh has span 2 [1]. The d-dimensional mesh can sustain a fault probability inversely polynomial in d and still have a large component whose expansion Among other things, this result can provide useful insights into the robustness of peer-to-peer networks like CAN [2], which behaves like a d-dimensional mesh in its steady state. Basically, we have shown that CAN can tolerate a fault probability which is inversely polynomial in its dimension without losing too much in its expansion properties is no more than a factor of d worse than the original.

References

[1] A Bagchi, A Bhargava abd A Chaudhary, D Eppstein, and C Scheideler. The effect of faults on network expansion. *Theor. Comput. Syst.*, 39(6):903–928, November 2006.

[2] S. Ratnasamy, P. Francis, M. Handley, R.M. Karp, and S. Schenker. A scalable content-addressable network. In *Proc. of SIGCOMM 2001*, pages 161–172, 2001.