Amitabha Bagchi
Scribe: Amitabha Bagchi

Lecture 3: The effect of faults on expansion I: Adversarial faults

20th and 26th August and 9th September 2008

In this lecture and a subsequent lecture we will study the effect of faults on the expansion of a network. We will begin in this lecture by studying the effect of adversarial faults i.e. faults placed by a malicious adversary whose goal is to degrade the expansion of the network.

More formally let us consider a graph $G = (V, E)$. An adversary with power f is a process which can delete at most f nodes from V to obtain a vertex set V_f. The subgraph $G_f = (V_f, E_f)$ induced by the vertex set V_f on G is referred to as the *faulty graph*.

We will be concerned with the edge expansion α of G and G_f. Clearly in the process of transforming G to G_f, an adversary intent on ruining the expansion of a graph can create bottlenecks to make the expansion arbitrarily bad. For example in Figure 1 we see that with with something like $\theta(\sqrt{n})$ faults the adversary can reduce the expansion of the $\sqrt{n} \times \sqrt{n}$ mesh to $\theta(\frac{1}{n})$.

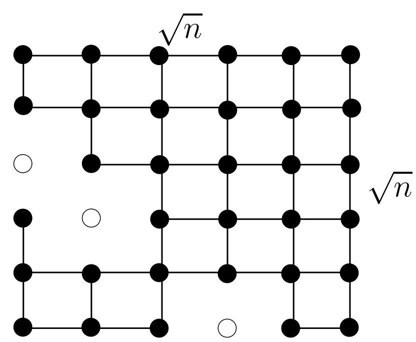

Figure 1: An adversarial selection of faults on the mesh

However, it may still be possible to identify a large portion of the graph whose expansion is relatively unaffected. A simple approach to this might be to simply cut away those parts of the graph whose expansion has fallen below a desired threshold and then see how much we have left. In the rest of this lecture we prove that this intuitive approach actually works. We show in Section 3.1 that if an adversary is given at most $c \cdot \alpha \cdot |V|$ faults to work with, for some suitably small constant c, then it is always possible to obtain a subgraph of G_f that has size $\theta(|V|)$ and expansion at least a constant times α. We will also show, in Section 3.2, that up to a constant factor this is a tight result in the sense that there is a class of graphs such that an instance of this class with n vertices fall into pieces of size $o(n)$ if the adversary has the power to remove $c_1 \cdot \alpha \cdot n$ vertices for a suitably large constant c_1.

3.1 Tolerating a bounded number of adversarial faults

If the number of vertices in G is n, we show that if an adversary is allowed no more than $c \cdot \alpha \cdot n$ faults, there always exists a subgraph of G_f called H which has $\Theta(n)$ nodes and an expansion of $\Omega(\alpha)$. In order to find this H we proceed by pruning away those parts of G_f that have bad expansion. We formalize this process below as an algorithm called *Prune* described in Figure 2. First, we need to introduce some notation. Given a graph G, we define $\Gamma_G(S)$ to be the set of nodes in the neighborhood of a node set S in G. The algorithm generates a sequence of graphs G_0 to G_m, where the final graph G_m is the graph H we are looking for.

Algorithm *Prune(ϵ)*

1. $G_0 \leftarrow G_f;\ i \leftarrow 0$
2. **while** $\exists S_i \subseteq V(G_i)$ such that $|\Gamma_{G_i}(S_i)| \leq \epsilon \cdot \alpha |S_i|$ and $|S_i| \leq |V(G_i)|/2$
   (a) $G_{i+1} \leftarrow G_i \setminus S_i$
   (b) $i \leftarrow i + 1$
3. **end while**
4. $H \leftarrow G_i;\ m \leftarrow i$

Figure 2: The pruning algorithm

Note that we do not claim that *Prune*(ϵ) is a polynomial time algorithm. In fact, any polynomial time implementation of this algorithm would be able to find us the expansion of a graph in polynomial time, and that problem is known to be NP-hard. This algorithm simply helps us prove the following existential result:

Theorem 3.1 *Let G be a graph with n nodes, maximum degree δ and node expansion α. Suppose that the adversary can select at most $f = \frac{\alpha n}{4\delta k^2}$ faulty nodes for some constant $k > 1$. Then,* Prune*$(1 - \frac{1}{k})$ returns a subgraph H of size at least $n - \frac{f \cdot k}{\alpha}$ with expansion at least $(1 - \frac{1}{k}) \cdot \alpha$.*

Proof. Let us argue first that H has the required expansion. This clearly follows from the fact that if it did not, then there would be a set with small expansion which would qualify as a candidate for pruning. Now we move on to proving the bound on the size of H given the restriction on f. Let $\mathcal{S}$ be the union of all the regions culled by *Prune*$(1 - \frac{1}{k})$. We will show that $|\mathcal{S}| \leq \frac{k \cdot f}{\alpha}$ by contradiction.

In order to do this we follow the simple intuition that in order to degrade the expansion of a particular subset of nodes, it is necessary for the adversary to make a large number of nodes in the neighbourhood of that subset faulty where by large we mean a quantity linear in $\alpha \cdot n$. But while it is easy to see that this holds for individual sets, it is not clear that this holds for groups of sets pruned by the algorithm, especially the ones that have edges into each other. How can we say that a set which was a *bad* (in the sense that it's expansion had dropped below $(1 - \frac{1}{k}) \cdot \alpha$) in the round that it was pruned remained bad at the end of m rounds? Can we claim that entire connected components of $\mathcal{S}$ have the same property in G_f that a particular set S_i had in G_i? The following lemma answers these questions in the affirmative:

Lemma 3.1 *For all j with $0 \leq j < m$,*

$$\left| \Gamma_{G_f} \left(\bigcup_{0 \leq i \leq j} S_i \right) \right| \leq \sum_{0 \leq i \leq j} |\Gamma_{G_i}(S_i)| \leq \alpha \cdot \left(1 - \frac{1}{k}\right) \cdot \left| \bigcup_{0 \leq i \leq j} S_i \right|.$$

Proof of Lemma 3.1. Consider the first inequality. Any node v that lies in the neighborhood of $\bigcup_i S_i$ in G_f must lie in the neighborhood of some S_i in G_f. Thus, because v is outside of $\bigcup_i S_i$ and therefore belongs to H, there must be an S_i such that v lies in the neighbourhood of S_i in G_i. Therefore, $\Gamma_{G_f}(\bigcup_i S_i) \subseteq \bigcup_i \Gamma_{G_i}(S_i)$. This proves the first inequality. Each set S_i that is culled by *Prune*$(1 - \frac{1}{k})$ has the property that $|\Gamma_{G_i}(S_i)| \leq \alpha \cdot (1 - \frac{1}{k}) \cdot |S_i|$.

Since the sets S_i are disjoint, $\sum_i |S_i| = |\bigcup_i S_i|$. Hence the second inequality follows. ■

Let us now assume towards contradiction that $|\mathcal{S}| > \frac{f \cdot k}{\alpha}$. Then there must be a j s.t.$\frac{k \cdot f}{\alpha} < | \cup_{0 \le i \le j} S_i|$. Now let us consider two cases.

Case 1.

$$\frac{k \cdot f}{\alpha} < | \cup_{0 \le i \le j} S_i| \le \frac{n}{2}.$$

This is the case in which we assume the set S_j which takes the sum$|\cup_{0 \le i \le j} S_i|$ beyond $\frac{kf}{\alpha}$ doesn't take it beyond $\frac{n}{2}$. So we can use the fact that the expansion of the set $\cup_{0 \le i \le j} S_i$ was atleast α in G. Also it follows from Lemma 3.1 that

$$|\Gamma_{G_f}(\bigcup_{0 \le i \le j} S_i)| \le \alpha \cdot \left(1 - \frac{1}{k}\right) \cdot |\bigcup_{0 \le i \le j} S_i|.$$

Hence, the number of faulty nodes in the neighbourhood of $\bigcup_{0 \le i \le j} S_i$ must be at least $\alpha(1 - (1 - \frac{1}{k})) \cdot |\bigcup_{0 \le i \le j} S_i|$, which is greater than $\alpha \cdot \frac{1}{k} \cdot \frac{k \cdot f}{\alpha} = f$. Since the total number of faults the adversary is allowed to create is at most f, we have a contradiction.

Case 2. In this case the value of j mentioned above has the following property:

$$\left| \bigcup_{0 \le i < j} S_i \right| \le \frac{k \cdot f}{\alpha}$$

and

$$\frac{n}{2} - \frac{k \cdot f}{\alpha} < |S_j| \le \frac{n}{2}.$$

In other words this is the case in which we assume the set S_j which takes the sum$| \cup_{0 \le i \le j} S_i|$ beyond $\frac{kf}{\alpha}$ takes it beyond $\frac{n}{2}$.

Let $\mathcal{S}' = \bigcup_{0 \le i < j} S_i$. It follows from the description of the pruning algorithm that $|\Gamma_{G_j}(S_j)| \le (1 - 1/k)\alpha|S_j|$. However, in order to upper bound $|\Gamma_{G_f}(S_j)|$, we also have to consider the neighbors S_j might have in $\mathcal{S}'$. According to Lemma 3.1, $|\Gamma_{G_f}(\mathcal{S}')| \le \alpha|\mathcal{S}'| \le k \cdot f$, and therefore, there can be at most $k \cdot f$ nodes in S_j that have neighbors in $\mathcal{S}'$. Since the maximum degree of G is δ, it follows that

$$|\Gamma_{G_f}(S_j)| \le \alpha \cdot \left(1 - \frac{1}{k}\right) \cdot |S_j| + \delta \cdot k \cdot f \ .$$

On the other hand, we know that $|\Gamma_G(S_j)| \geq \alpha|S_j|$. Hence, the number of faults in G_f must be at least $\alpha|S_j|/k - \delta \cdot k \cdot f$. Since we are in Case 2 and from the definition of f it follows that

$$|S_j| \geq \frac{n}{2} - \frac{n}{4\delta k} \geq \frac{3n}{8}$$

because $\delta \geq 2$. Furthermore, $\delta \cdot k \cdot f = \alpha n/(4k)$. Hence,

$$\frac{\alpha|S_j|}{k} - \delta \cdot k \cdot f \geq \frac{3\alpha \cdot n}{8k} - \frac{\alpha \cdot n}{4k} \geq \frac{\alpha \cdot n}{8k} .$$

But from $k > 1$ and $\delta \geq 2$ it follows that $f = \alpha n/(4\delta k^2) < \alpha n/(8k)$, a contradiction. Hence, H is at least $n - \frac{k \cdot f}{\alpha}$ in size and has an expansion of at least $(1 - \frac{1}{k}) \cdot \alpha$. ■

3.2 A lower bound on adversarial faults

The result given in Theorem 3.1 is the best possible up to constant factors in the sense that for every $\alpha > 0$ smaller than some constant there is an infinite family of graphs with expansion α which disintegrate into components of size $o(n)$ if $f \geq c \cdot \alpha n$ for some sufficiently large constant c.

Theorem 3.2 *There exists a constant γ such that, given any $\alpha < \gamma$, there is an infinite family of graphs with expansion α for which there is an adversarial selection of $c \cdot \alpha \cdot n$ faulty nodes causing the graph to break into components of size $o(n)$, where n is the number of nodes in the graph and c is an appropriately chosen constant.*

Proof. Consider an infinite family $\mathcal{G}$ of δ-regular expander graphs with constant degree δ, i.e., δ-regular graphs with the property that every subset of nodes containing at most half of the nodes in the graph has a constant expansion. It is well-known that random δ-regular graphs with $\delta \geq 3$ almost surely have this property.

For any fixed $G \in \mathcal{G}$ of size n and any k, let graph H be a copy of G with each edge being replaced by a chain of k nodes (between its two endpoints), where k is even. Then H has $k \cdot (\delta n)/2 + n = \Theta(k \cdot n)$ nodes. In Figure 3, we show some edges of such a graph transformation with $k = 6$. Note that since every node in G has δ neighbours the total number of edges in G are $\frac{\delta n}{2}$, and so the total number of vertices in H are $\frac{n\delta k}{2} + n$.

Claim 3.3 *Graph H has expansion $\Theta(\frac{1}{k})$.*

Let us assume for a moment that this claim holds and see how we would prove Theorem 3.2. If we make the middle most node on each edge chain faulty, the graph H will break into such star-like components of size $\frac{k\delta}{2}+1$, each having a vertex of G as its centre. So, on a graph $H$of $\frac{n\delta k}{2}nk$ vertices, we can make an adversarial selection of $\frac{n\delta}{2}$ vertices and get components of size $o(n)$.

Now let us prove the claim.

Proof of Claim 3.3. It is clear that the expansion of H is $O(\frac{1}{k})$. This can be seen by looking at the k vertices on any chain between two vertices of G in H. Proving a lower bound, i.e., every subset U of the node set of H such that $|U| < |H|/2$ has expansion at least $\Omega(|U|/k)$, is a little trickier.

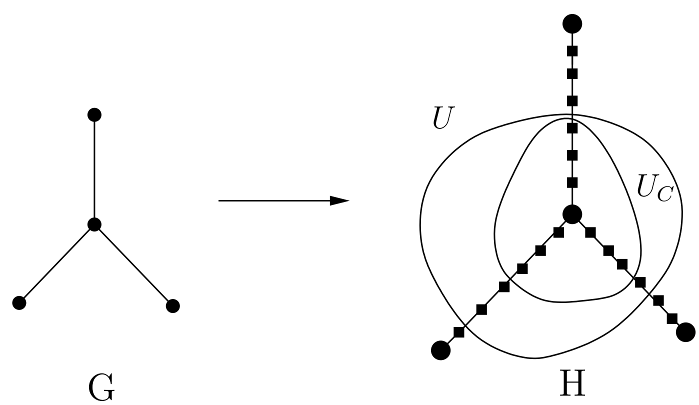

Figure 3: Distinguishing between two kinds of vertices for the lower bound.

For the lower bound, we have to show that every subset of nodes in H of size at most $|H|/2$ has an expansion of $\Omega(1/k)$. Consider any subset U of H-nodes with $|U| \leq |H|/2$. We differentiate between two sets of nodes within U. To do this we start by defining a set C as the set of all G-nodes with the property that all nodes within a distance of $k/2$ from them in H are in U. We name as U_C the set consisting of all H-nodes within a distance of $k/2$ of the nodes in C. Note that $U_C \subseteq U$. The other set of nodes we will consider is $U' = U \setminus U_C$. See Figure 3 for an illustration of this division of U into two subsets. We will proceed by lowerbounding the size of the boundary of U in two different ways corresponding to these two sets. Then we will take the better of these two lowerbounds to give us the result.

Let us first consider U_C. Since every G-node has exactly $\delta \cdot k/2 + 1$ many nodes within a distance of $k/2$ in H and $|U| \leq |H|/2$, it follows that $|C| \leq |G|/2$. Hence, C has an expansion of at least γ in G for some constant $\gamma > 0$. For every node $v \in \Gamma_G(C)$ there must be an H-node within a distance of $k/2$ from v that is not in U. Hence there must be an H-node w which is in the boundary of U for each $v \in \Gamma_G(C)$. In other words all such nodes w are in $\Gamma_H(U)$. Note that this w belongs exclusively to v since it is within $k/2$ distance of it i.e. this w cannot be claimed by more than one node of $\Gamma_G(C)$. Since $|U_C| = (\delta \cdot k/2 + 1)|C|$, it holds that

$$|\Gamma_H(U)| \geq |\Gamma_G(C)| \geq \gamma|C| = \frac{\gamma}{\delta \cdot k/2 + 1} \cdot |U_C| \, . \tag{1}$$

Next, consider the set $U' = U \setminus U_C$ and let C' be the set of all G-nodes that have at least one U'-node within a distance of $k/2$ in H. From the definition of U' and C' it follows that for each $v \in C'$ there is at least one H-node within a distance of $k/2$ from v that is not in U and hence there must be a H-node w in the boundary of U i.e. all such nodes w belong to $\Gamma_H(U)$. Since $|U'| \leq (\delta \cdot k/2)|C'|$, we get

$$|\Gamma_H(U)| \geq |C'| \geq \frac{1}{\delta \cdot k/2} \cdot |U'| \, . \tag{2}$$

Combining inequalities (1) and (2), it follows that

$$|\Gamma_H(U)| \geq \max\left\{\frac{\gamma}{\delta \cdot k/2 + 1} \cdot |U_C|, \; \frac{1}{\delta \cdot k/2} \cdot |U'|\right\}$$

and since $U = U_C \cup U'$, we immediately get that $|\Gamma_H(U)| = \Omega(|U|/k)$. Hence, the lower bound holds. ■

Notes

The results in this lecture are taken from [2]. Theorem 2.5 of [2] gives another, more far-reaching, lower bound for the power of an adverary. The authors show that even graphs with good "uniform expansion" (a version of isoperimetric inequalities) fall into sublinear components on the removal of $c \cdot \alpha \cdot n$ faults for a suitable value of c. See also [4] for an earlier work in a similar flavour addressing the problem of finding a large component in a graph made faulty by an adversary with bounded power.

The pruning algorithm, as mentioned earlier, is not a polynomial time algorithm since it involves solving the sparsest cut problem, known to be NP-hard. The best known approximation to the sparsest cut problem was improved to $\sqrt{\log n}$ by Arora, Rao and Vazirani [1]. They had initially conjectured that the sparsest cut problem should be approximable in polynomial time to within a constant factor. This would have had the effect of turning the existential results in this paper into algorithmic ones, making it possible to actually find large components with good expansion (perhaps by sacrficing additional constants) whose existence is proved here in polynomial time. However, recently obtained lower bounds indicate that this might not be possible [3].

References

[1] S. Arora, S.B. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proc. of 36th Annual ACM Symp. on Theory of Computing*, pages 222–231, 2004.

[2] 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.

[3] S. Khot and N. Vishnoi. The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into l_1. In *Proc. 46th Annu. IEEE Symp. on Foundations of Computer Science (FOCS '05)*, pages 53–62, 2005.

[4] E. Upfal. Tolerating a linear number of faults in networks of bounded degree. *Information and Computation*, 115(2):312–320, 1994.