1 Graph Sparsfication

In this section we discuss the approximation of a graph $G(V, E)$ by a sparse graph $H(V, F)$ on the same vertex set. In particular, we consider any graph with $|E| = \Omega(n^{1+\delta})$ edges to be dense; we wish to find sparse representations with $|F| = O(n)$ edges that have proportionally the same number of edges crossing any cut. That is, if we scale the values of every edge in the sparse graph by $|E|/|F|$, the value of each cut will remain (approximately) the same.

A common application of graph sparsification is internet traffic routing. Consider building a undirected network N on n nodes, and suppose we would like to route 1 unit of flow (directionless) between each pair of nodes in N under some capacity constraints on the edges. A complete graph K_n with $c(e) = 1 \; \forall \; e \in E$ would suffice, but for practical purposes it is undesirable to use so many $(O(n^2))$ edges. If we require the number of edges to be linear in n, we might consider a star graph and scale up the capacities on the edges to $n-1$. The drawback in this case is that there is a single node of degree $n-1$ which has too much $(O(n^2))$ traffic flowing through it – if this node were to fail it would bring down the entire network. A reasonable goal, then, is to produce a graph on n vertices and $O(n)$ edges that maintains the connectivity of the complete graph while also having approximately uniform vertex degree.

The **expansion** $\rho(G)$ of an undirected graph $G(V, E)$ is defined as the minimum cut value weighted by the size of the smaller cut partition:

$$\rho(G) = \min_{S \subseteq V} \frac{c(S, \bar{S})}{\min(|S|, |\bar{S}|)} = \min_{S \subseteq V, |S| \leq \frac{n}{2}} \frac{c(S, \bar{S})}{|S|}.$$

Taking all edge-values to be 1, our goal from above is equivalent to finding an approximately regular graph G with $O(n)$ edges and $\rho(G) = \Omega(1)$. For comparison, note that the complete graph has $O(n^2)$ edges and achieves $\rho(K_n) = n/2$. We give two methods of constructing such G.

1.1 Erdős-Rényi Random Graphs

The Erdős-Rényi $G(n, p)$ model for constructing random graphs denotes a graph on n vertices where each of the $n(n-1)/2$ possible edges are included in the edge set independently with probability p. To get $m = O(n)$, we may choose $p = c/n$ for some constant $c > 0$, then we may compute

$$\mathrm{E}[m] = \frac{c(n-1)}{2},$$
$$\mathrm{E}[d(v)] = c\left(\frac{n-1}{n}\right) \quad \forall \; v \in V.$$

The following claim shows that we may sample from an Erdős-Rényi graph distribution and obtain a suitable G with high probability.

Claim 1 *If we choose $p = \frac{\log n}{n\epsilon^2}$, then $G(n, p)$ will have $O(\frac{n \log n}{n\epsilon^2})$ edges and with high probability, the size of every cut in G will be within $(1 \pm \epsilon)$ of its expected value, so $\rho(G) = \Omega(1)$.*

1.2 Random d-Regular Graphs

Recall that a d-regular graph is one in which all vertices have the same degree d.

Theorem 1 *For all $d \geq 3$, there exists a constant $\alpha > 0$ such that with high probability, a random d-regular graph G has expansion $\rho(G) \geq \alpha$.*

Proof: (To simplify the calculations, we present the proof for d sufficiently large instead of $d \geq 3$. The proof for $d = 3$ is similar.)

First we note that we can generate d-regular graphs on n vertices via the configuration model: we split each vertex into d mini-vertices, and find the edges by generating a random perfect matching on the mini-vertices. Then when we combine each vertex's mini-vertices, every vertex will have degree d. Note that a graph produced in this way may have multiple edges or self-loops.

To prove the theorem we need to show that for each set $S \subset V$, $|S| = k$, the probability that $|c(S, \bar{S})| < \alpha k$ is sufficiently small. Assume that there exists such an S of size k. For a given k, there are $\binom{n}{k}$ possible choices for S. For a given S, there are $\binom{dk}{\alpha k}\binom{dn-dk}{\alpha k}$ ways to choose the minivertices in S and minivertices in $\bar{S}$ involved in a cut of size αk.

Let us start by calculating the number of d regular graphs on n vertices, i.e., the number of perfect matchings of K_{dn}.

$$\begin{aligned} f(nd) &= \binom{nd}{2}\binom{nd-2}{2}\cdots\binom{2}{2}\frac{1}{(nd/2)!} \\ &= \frac{(nd)!}{2^{nd/2}(nd/2)!}. \end{aligned}$$

We will use the following Stirling approximation for factorials:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1 + O\left(\frac{1}{n}\right)\right).$$

So

$$f(nd) = c(nd)^{nd/2}e^{-nd/2}\left(1 + O\left(\frac{1}{n}\right)\right),$$

for some constant c. Then the probability of the event that only a certain αk of minivertices match outside of their proper subset is at most

$$\frac{f(dk-\alpha k)f(dn-dk-\alpha k)f(2\alpha k)}{f(dn)}.$$

Therefore, the probability P_k that there is a subset S, with $|S| = k$ and expansion less than α may be bounded as

$$P_k \leq \alpha k\binom{n}{k}\binom{dk}{\alpha k}\binom{dn-dk}{\alpha k}\frac{f(dk-\alpha k)f(dn-dk-\alpha k)f(2\alpha k)}{f(dn)}.$$

Using the simple but useful inequality,

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$$

we have

$$\begin{aligned} P_k &\leq c\alpha k\left(\frac{en}{k}\right)^k\left(\frac{edn}{2\alpha k}\right)^{2\alpha k}\frac{(dk-\alpha k)^{\frac{dk-\alpha k}{2}}(dn-dk-\alpha k)^{(dn-dk-\alpha k)/2}(2\alpha k)^{\alpha k}}{(nd)^{nd/2}} \\ &\leq c\alpha k e^k\left(\frac{ed}{2\alpha}\right)^{2\alpha k}\left(\frac{n}{k}\right)^{k+2\alpha k}\left(\frac{k}{n}\right)^{(dk-\alpha k)/2} \\ &\leq c\alpha k\left(e\left(\frac{ed}{2\alpha}\right)^{2\alpha}\right)^k\left(\frac{k}{n}\right)^{((d-2)k-5\alpha k)/2} \\ &\leq c\alpha k(ce)^k(k/n)^{3k} \end{aligned}$$

for $d \geq 15$, $\alpha < 1/100$. Therefore,

$$\sum_{k=1}^{n/2} P_k = O(1/n^2)$$

■

2 The Probabilistic Method

Theorem 1 gives an affirmative answer to the question of whether there exists a graph with $m = O(n)$ edges and expansion $\rho(G) = \Omega(1)$. It is interesting to note, however, that the proof is non-constructive. We only give a distribution of graphs from which a random sampled instance is likely to have the properties we want. This simple idea is the premise of a combinatorial analysis technique known as the *probabilistic method*: in order to prove the existence of a structure, we merely need to show that there is a positive probability that the structure exists.

2.1 A Simple Example: Monochromatic Coloring

We let $S_1, \ldots, S_m$ be subsets of a larger set S such that each subset S_i contains exactly l elements from S. Is it possible to color the elements of S with two colors — say, red and blue — such that no set S_i is monochromatic?

Lemma 1 *If the number of subsets $m < 2^{l-1}$, then such a coloring is always possible*

Proof: We use the probabilistic method. Toss a coin for each vertex and color the vertex red if the coin lands heads, blue for tails, so the probability that a vertex is red is $1/2$, independent of the color of any other vertex. Then the probability that a given set S_i is entirely red or entirely blue is 2^{-l}, so the probability p^i_{mono} that a S_i is monochromatic is $p^i_{mono} = 2 \cdot 2^{-l} = 2^{-l+1}$.

Recall from basic probability the "union bound" or "subadditivity property" of probabilities. That is, for any (arbitrary, not necessarily disjoint) events $E_1, E_2, \ldots, E_j$,

$$Pr(\cup_{i=1}^{j} E_i) \leq \sum_{i=1}^{j} Pr(E_i).$$

Using the union bound, the probability p_{mono} that some set is monochromatic obeys

$$p_{mono} \leq \sum_{i=1}^{m} p^i_{mono} = \sum_{i=1}^{m} 2^{-l+1} = m \cdot 2^{-l+1} < 1$$

for $m < 2^{l-1}$.

Therefore there is a positive probability that no set is monochromatic, and so there must exist some assignment of colors to vertices such that no set is monochromatic. ■

Note again that this proof is nonconstructive. We've created a distribution over all possible color assignments (namely, the uniform distribution) and used this to show a positive probability that a graph with the desired property exists. By explicitly giving a distribution over the space of all possible colorings, we turn an exhaustive search algorithm into a simple probability calculation.

2.2 Chromatic Number of a Graph

A **proper vertex k-coloring** of a graph $G(V, E)$ is an assignment of k colors to vertices such that no two vertices of the same color share an edge. The **chromatic number** $\chi(G)$ is equal to the smallest number of colors needed to have a proper vertex coloring.

Chromatic number might appear to be based mostly on the local structure of a graph. For example, it is simple to see that if a graph G contains K_k as a subgraph, then $\chi(G) \geq k$. In general, very tightly connected subregions of graphs need many colors for proper coloring. A reasonable question to ask is whether there exist graphs with high chromatic number that do not have any particularly dense subregions; their global structure is what makes them require many colors.

As a measure of local connectivity, we define the **girth** of a graph G, $g(G)$, to be the length of the smallest cycle in G. If $g(G) > 3$, we say that G is triangle-free.

In 1954, B. Descartes was the first to show that triangle-free graphs can have arbitrarily high chromatic numbers, but this construction was complicated and contained many short cycles. In 1959, Paul Erdős used the probabilistic method to prove the existence of graphs with arbitrarily high girth and chromatic number.

Theorem 2 (Erdős, 1959) *For every $g, k > 0$, there exists a graph G with $\chi(G) \geq k$ and $g(G) \geq g$.*

Proof: An **independent set** in a graph G is $U \subset V$ such that no two vertices in U are connected by an edge. If a graph has chromatic number k, then there must exist at least one independent set of size at least $\frac{n}{k}$, since each color in a proper coloring corresponds to an independent set.

We consider Erdős-Rényi random graphs $G(n, p)$. In order to show that $\chi(G) \geq k$, it suffices to prove that with high probability the size of any independent set in G is at most $\frac{n}{k}$. We will prove that with high probability for a suitable selection of p, the graph doesn't have any independent set of size $\frac{n}{2k}$.

We use the union bound. The probability that any set of $n/2k$ vertices is an independent set is $(1-p)^{\binom{n/2k}{2}}$. There are $\binom{n}{n/2k}$ possibilities for vertex sets of size $n/2k$. By the union bound, the probability of $G(n, p)$ having such an independent set is therefore at most

$$\begin{aligned} Pr\left[G_{n,p} \text{ has an independent set of size } \frac{n}{2k}\right] &\leq \binom{n}{n/2k}(1-p)^{\binom{n/2k}{2}} \\ &\leq 2^n e^{-pn^2/8k} \\ &\leq e^{\log 2n - pn^2/8k} \\ &\leq e^{n \log 2 - n^{\epsilon+1}/8k}. \end{aligned}$$

where we set $p = n^{\epsilon - 1}$ for some $\epsilon < 1/g$. The above expression tends to zero as $n \to \infty$, so is therefore smaller than $1/4$ for n sufficiently large.

Let X be the random variable counting the number of cycles of length g and smaller. By linearity of expectation,

$$\begin{aligned} \mathrm{E}[X] &= \sum_{i=1}^{g} \binom{n}{i} \frac{(i-1)!}{2} p^i \\ &\leq g(np)^g = gn^{\epsilon g}. \end{aligned}$$

For $0 < \epsilon < 1/g$, the above expression is $o(n)$. Thus, for sufficiently large n, $\mathrm{E}(X) < n/4$. By Markov's inequality,

$$Pr(X > n/2) \leq Pr(X > 2\mathrm{E}(X)) < 1/2.$$

Therefore, if we choose n sufficiently large, and $p = n^{\epsilon - 1}$ for $0 < \epsilon < 1/g$, the probability that $G(n,p)$ has a independent set of size $\frac{n}{2k}$ or that the number of cycles of length at most g is $\frac{n}{2}$ is less than 1 by the union bound. Considering the complement of that event, we see that there must exist a graph G with no independent set of size $\frac{n}{2k}$ and with at most $\frac{n}{2}$ cycles of length at most g.

Now, we can construct a graph G' by removing a vertex from each short cycle of G. The number of vertices in G' is at least $\frac{n}{2}$, the size of the maximum independent set in G' is no more that $\frac{n}{2k}$, and there are no cycles of length less than g. This implies that $\chi(G') > k$ and $g(G') > g$. ■