

Lecture 4 - A little sparse regularity

In this lecture, we will prove a counting lemma for sparse graphs, concentrating on the case of triangles, since this case contains all the important ideas and is notationally simpler.

For graphs, the counting lemma says that if two weighted graphs are close in cut norm, then they have similar triangle densities. To be more specific, we consider weighted tripartite graphs on the vertex set $X \cup Y \cup Z$, where X , Y and Z are finite sets. Such a weighted graph g is given by three functions $g_{XY}: X \times Y \rightarrow \mathbb{R}$, $g_{XZ}: X \times Z \rightarrow \mathbb{R}$ and $g_{YZ}: Y \times Z \rightarrow \mathbb{R}$, although we often drop the subscripts if they are clear from context. We write $\|g\|_{\square} = \max\{\|g_{XY}\|_{\square}, \|g_{XZ}\|_{\square}, \|g_{YZ}\|_{\square}\}$.

For dense graphs, the counting lemma is as follows. Typically, \tilde{g} will be a low complexity approximation to g given by applying the weak regularity lemma.

Proposition 1 (Triangle counting lemma, dense setting) *Let g and \tilde{g} be weighted tripartite graphs on $X \cup Y \cup Z$ with weights in $[0, 1]$. If $\|g - \tilde{g}\|_{\square} \leq \epsilon$, then*

$$|\mathbb{E}_{x \in X, y \in Y, z \in Z}[g(x, y)g(x, z)g(y, z) - \tilde{g}(x, y)\tilde{g}(x, z)\tilde{g}(y, z)]| \leq 3\epsilon.$$

Proof Unless indicated otherwise, all expectations are taken over $x \in X$, $y \in Y$, $z \in Z$ uniformly and independently. From the definition of the cut norm, we have that

$$|\mathbb{E}_{x \in X, y \in Y}[(g(x, y) - \tilde{g}(x, y))a(x)b(y)]| \leq \epsilon \tag{1}$$

for every function $a: X \rightarrow [0, 1]$ and $b: Y \rightarrow [0, 1]$. It follows that

$$|\mathbb{E}[g(x, y)g(x, z)g(y, z) - \tilde{g}(x, y)g(x, z)g(y, z)]| \leq \epsilon,$$

since the expectation has the form (1) if we fix any value of z . Similarly, we have

$$|\mathbb{E}[\tilde{g}(x, y)g(x, z)g(y, z) - \tilde{g}(x, y)\tilde{g}(x, z)g(y, z)]| \leq \epsilon$$

and

$$|\mathbb{E}[\tilde{g}(x, y)\tilde{g}(x, z)g(y, z) - \tilde{g}(x, y)\tilde{g}(x, z)\tilde{g}(y, z)]| \leq \epsilon.$$

The result then follows from telescoping and the triangle inequality. \square

When considering sparse graphs G , it will be useful to think of G as being a subgraph of a highly pseudorandom graph Γ . If Γ has density p , we write ν for the measure of Γ , given by

$$\nu(x, y) = \begin{cases} 0 & \text{if } xy \notin E(\Gamma); \\ p^{-1} & \text{if } xy \in E(\Gamma). \end{cases}$$

The subgraph G is then represented by a function g which is given by

$$g(x, y) = \begin{cases} 0 & \text{if } xy \notin E(G); \\ p^{-1} & \text{if } xy \in E(G). \end{cases}$$

In what follows, we will not need to assume that G or Γ are graphs. It will be sufficient to assume that $0 \leq g \leq \nu$, where ν is a sufficiently well-behaved function. Here, we will assume that ν satisfies the following pseudorandomness condition, which we call the 3-linear forms condition.

Definition 1 (3-linear forms condition) *A weighted tripartite graph ν with vertex sets X, Y and Z satisfies the 3-linear forms condition if*

$$\mathbb{E}_{x,x' \in X, y,y' \in Y, z,z' \in Z} [\nu(y,z)\nu(y',z)\nu(y',z')\nu(x,z)\nu(x',z)\nu(x,z')\nu(x',z') \cdot \nu(x,y)\nu(x',y)\nu(x,y')\nu(x',y')] = 1 + o(1) \quad (2)$$

and also (2) holds when one or more of the twelve ν factors in the expectation are erased.

This essentially says that the weighted graph ν has asymptotically the expected H -density for any subgraph H of $K_{2,2,2}$. For example, triangle density in ν is given by the expression $\mathbb{E}[\nu(x,y)\nu(x,z)\nu(y,z)]$, where x, y, z vary independently and uniformly over X, Y, Z , respectively. The pseudorandomness assumption requires, amongst other things, that this triangle density be $1 + o(1)$, the normalization having accounted for the other factors.

The counting lemma for sparse graphs is now as follows. Again, we think of \tilde{g} as being a low complexity approximation to g coming from a weak regularity lemma. While we have not proved that a regularity lemma holds in the sparse context, it is not particularly difficult to prove that any graph satisfying the 3-linear forms condition is also upper regular and it is well known that subgraphs of upper regular graphs satisfy analogues of the regularity lemma. This shows that for any $\epsilon > 0$ and any g with $0 \leq g \leq \nu$, there will be \tilde{g} with $0 \leq \tilde{g} \leq 1$ such that $\|g - \tilde{g}\|_{\square} \leq \epsilon$.

Theorem 1 (Triangle counting lemma, sparse setting) *Let ν, g, \tilde{g} be weighted tripartite graphs on $X \cup Y \cup Z$. Assume that ν satisfies the 3-linear forms condition, $0 \leq g \leq \nu$, and $0 \leq \tilde{g} \leq 1$. If $\|g - \tilde{g}\|_{\square} = o(1)$, then*

$$|\mathbb{E}_{x \in X, y \in Y, z \in Z} [g(x,y)g(x,z)g(y,z) - \tilde{g}(x,y)\tilde{g}(x,z)\tilde{g}(y,z)]| = o(1).$$

The proof uses repeated application of the Cauchy–Schwarz inequality, a standard technique in this area, popularized by Gowers. The key additional idea is what we call *densification*. After several applications of the Cauchy–Schwarz inequality, it becomes necessary to analyze the 4-cycle density: $\mathbb{E}_{x,y,z,z'} [g(x,z)g(x,z')g(y,z)g(y,z')]$. To do this, one introduces an auxiliary weighted graph $g': X \times Y \rightarrow [0, \infty)$ defined by $g'(x,y) := \mathbb{E}_z [g(x,z)g(y,z)]$ (this is basically the codegree function). The expression for the 4-cycle density now becomes $\mathbb{E}_{x,y,z} [g'(x,y)g(x,z)g(y,z)]$.

At first glance, it seems that our reasoning is circular. Our aim was to estimate a certain triangle density expression but we have now returned to a triangle density expression. However, g' behaves much more like a dense weighted graph with bounded edge weights, so what we have accomplished is to replace one of the “sparse” g_{XY} by a “dense” g'_{XY} . If we do this two more times, replacing g_{YZ} and g_{XZ} with dense counterparts, the problem reduces to the dense case, which we already know how to handle.

The next lemma is crucial to what follows. It shows that in certain expressions a factor ν can be deleted from an expectation while incurring only a $o(1)$ loss.

Lemma 1 (Strong linear forms) *Let ν, g, \tilde{g} be weighted tripartite graphs on $X \cup Y \cup Z$. Assume that ν satisfies the 3-linear forms condition, $0 \leq g \leq \nu$, and $0 \leq \tilde{g} \leq 1$. Then*

$$\mathbb{E}_{x \in X, y \in Y, z, z' \in Z} [(\nu(x,y) - 1)g(x,z)g(x,z')g(y,z)g(y,z')] = o(1)$$

and the same statement holds if any subset of the four g factors are replaced by \tilde{g} .

Proof We give the proof when none of the g factors are replaced. The other cases require only a simple modification. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \left| \mathbb{E}_{x,y,z,z'} [(\nu(x,y) - 1)g(x,z)g(x,z')g(y,z)g(y,z')] \right|^2 \\ & \leq \mathbb{E}_{y,z,z'} [(\mathbb{E}_x [(\nu(x,y) - 1)g(x,z)g(x,z')])^2 g(y,z)g(y,z')] \mathbb{E}_{y,z,z'} [g(y,z)g(y,z')] \\ & \leq \mathbb{E}_{y,z,z'} [(\mathbb{E}_x [(\nu(x,y) - 1)g(x,z)g(x,z')])^2 \nu(y,z)\nu(y,z')] \mathbb{E}_{y,z,z'} [\nu(y,z)\nu(y,z')]. \end{aligned}$$

The second factor is at most $1 + o(1)$ by the linear forms condition. So it remains to analyze the first factor. We have, by another application of the Cauchy-Schwarz inequality,

$$\begin{aligned} & \left| \mathbb{E}_{y,z,z'} [(\mathbb{E}_x [(\nu(x,y) - 1)g(x,z)g(x,z')])^2 \nu(y,z)\nu(y,z')] \right|^2 \\ & = \left| \mathbb{E}_{x,x',y,z,z'} [(\nu(x,y) - 1)(\nu(x',y) - 1)g(x,z)g(x,z')g(x',z)g(x',z')\nu(y,z)\nu(y,z')] \right|^2 \\ & = \left| \mathbb{E}_{x,x',z,z'} [\mathbb{E}_y [(\nu(x,y) - 1)(\nu(x',y) - 1)\nu(y,z)\nu(y,z')] g(x,z)g(x,z')g(x',z)g(x',z')] \right|^2 \\ & \leq \mathbb{E}_{x,x',z,z'} [(\mathbb{E}_y [(\nu(x,y) - 1)(\nu(x',y) - 1)\nu(y,z)\nu(y,z')])^2 g(x,z)g(x,z')g(x',z)g(x',z')] \\ & \quad \cdot \mathbb{E}_{x,x',z,z'} [g(x,z)g(x,z')g(x',z)g(x',z')] \\ & \leq \mathbb{E}_{x,x',z,z'} [(\mathbb{E}_y [(\nu(x,y) - 1)(\nu(x',y) - 1)\nu(y,z)\nu(y,z')])^2 \nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')] \\ & \quad \cdot \mathbb{E}_{x,x',z,z'} [\nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')]. \end{aligned}$$

Using the 3-linear forms condition, the second factor is $1 + o(1)$ and the first factor is $o(1)$ (expand everything and notice that all the terms are $1 + o(1)$ and the signs make all the 1's cancel). \square

Proof of Theorem 1: If ν is identically 1, we are in the dense setting, in which case the theorem follows from Proposition 1. Now we apply induction on the number of $\nu_{XY}, \nu_{XZ}, \nu_{YZ}$ which are identically 1. By relabeling if necessary, we may assume without loss of generality that ν_{XY} is not identically 1. We define auxiliary weighted graphs $\nu', g', \tilde{g}': X \times Y \rightarrow [0, \infty)$ by

$$\begin{aligned} \nu'(x,y) &:= \mathbb{E}_z [\nu(x,z)\nu(y,z)], \\ g'(x,y) &:= \mathbb{E}_z [g(x,z)g(y,z)], \\ \tilde{g}'(x,y) &:= \mathbb{E}_z [\tilde{g}(x,z)\tilde{g}(y,z)]. \end{aligned}$$

We refer to this step as *densification*. The idea is that even though ν and g are possibly unbounded, the new weighted graphs ν' and g' behave like dense graphs. The weights on ν' and g' are not necessarily bounded by 1, but they almost are. To correct this, we will cap the weights by setting $g'_{\wedge 1} := \max\{g', 1\}$ and $\nu'_{\wedge 1} := \max\{\nu', 1\}$ and show that this capping operation has negligible effect.

We have

$$\mathbb{E}[g(x,y)g(x,z)g(y,z) - \tilde{g}(x,y)\tilde{g}(x,z)\tilde{g}(y,z)] = \mathbb{E}[gg' - \tilde{g}\tilde{g}'] = \mathbb{E}[g(g' - \tilde{g}')] + \mathbb{E}[(g - \tilde{g})\tilde{g}'], \quad (3)$$

where the first expectation is taken over $x \in X, y \in Y, z \in Z$ and the other expectations are taken over $X \times Y$ (we will use these conventions unless otherwise specified). The second term on the right-hand side of (3) equals $\mathbb{E}[(g(x,y) - \tilde{g}(x,y))\tilde{g}(x,z)\tilde{g}(y,z)]$ and its absolute value is at most $\|g - \tilde{g}\|_{\square} = o(1)$ (here we use $0 \leq \tilde{g} \leq 1$ as in the proof of Proposition 1). So it remains to bound the first term on the right-hand side of (3). By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} (\mathbb{E}[g(g' - \tilde{g}')])^2 &\leq \mathbb{E}[g(g' - \tilde{g}')^2] \mathbb{E}[g] \leq \mathbb{E}[\nu(g' - \tilde{g}')^2] \mathbb{E}[\nu] \\ &= \mathbb{E}_{x,y} [\nu(x,y) (\mathbb{E}_z [g(x,z)g(y,z) - \tilde{g}(x,z)\tilde{g}(y,z)])^2] \mathbb{E}_{x,y} [\nu(x,y)]. \end{aligned}$$

The second factor is $1 + o(1)$ by the linear forms condition. By Lemma 1, the first factor differs from

$$\mathbb{E}_{x,y}[(\mathbb{E}_z[g(x,z)g(y,z) - \tilde{g}(x,z)\tilde{g}(y,z)])^2] = \mathbb{E}[(g' - \tilde{g}')^2] \quad (4)$$

by $o(1)$ (take the difference, expand the square, and then apply the lemma term-by-term).

The 3-linear forms condition implies that $\mathbb{E}[\nu'] = 1 + o(1)$ and $\mathbb{E}[\nu'^2] = 1 + o(1)$. Therefore, by the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[|\nu' - 1|] \leq \mathbb{E}[(\nu' - 1)^2] = o(1). \quad (5)$$

We want to show that (4) is $o(1)$. We have

$$\mathbb{E}[(g' - \tilde{g}')^2] = \mathbb{E}[(g' - \tilde{g}')(g' - g'_{\wedge 1})] + \mathbb{E}[(g' - \tilde{g}')(g'_{\wedge 1} - \tilde{g}')]. \quad (6)$$

Since $0 \leq g' \leq \nu'$, we have

$$0 \leq g' - g'_{\wedge 1} = \max\{g' - 1, 0\} \leq \max\{\nu' - 1, 0\} \leq |\nu' - 1|. \quad (7)$$

Using (5) and (7), the absolute value of the first term on the right-hand side of (6) is at most

$$\mathbb{E}[(\nu' + 1)|\nu' - 1|] = \mathbb{E}[(\nu' - 1)|\nu' - 1|] + 2\mathbb{E}[|\nu' - 1|] = o(1).$$

Next, we claim that

$$\|g'_{\wedge 1} - \tilde{g}'\|_{\square} = o(1). \quad (8)$$

Indeed, for any $A \subseteq X$ and $B \subseteq Y$, we have

$$\mathbb{E}_{x,y}[(g'_{\wedge 1} - \tilde{g}')(x,y)1_A(x)1_B(y)] = \mathbb{E}[(g'_{\wedge 1} - \tilde{g}')1_{A \times B}] = \mathbb{E}[(g'_{\wedge 1} - g')1_{A \times B}] + \mathbb{E}[(g' - \tilde{g}')1_{A \times B}].$$

By (7) and (5), the absolute value of the first term is at most $\mathbb{E}[|\nu' - 1|] = o(1)$. The second term can be rewritten as

$$\mathbb{E}_{x,y,z}[1_{A \times B}(x,y)g(x,z)g(y,z) - 1_{A \times B}(x,y)\tilde{g}(x,z)\tilde{g}(y,z)],$$

which is $o(1)$ by the induction hypothesis (replace $\nu_{XY}, g_{XY}, \tilde{g}_{XY}$ by $1, 1_{A \times B}, 1_{A \times B}$, respectively, and note that this increases the number of $\{\nu_{XY}, \nu_{XZ}, \nu_{YZ}\}$ which are identically 1). This proves (8).

We now expand the second term on the right-hand side of (6) as

$$\mathbb{E}[(g' - \tilde{g}')(g'_{\wedge 1} - \tilde{g}')] = \mathbb{E}[g'g'_{\wedge 1}] - \mathbb{E}[g'\tilde{g}'] - \mathbb{E}[\tilde{g}'g'_{\wedge 1}] + \mathbb{E}[\tilde{g}'^2]. \quad (9)$$

We claim that each of the expectations on the right-hand side is $\mathbb{E}[(\tilde{g}')^2] + o(1)$. Indeed, we have

$$\mathbb{E}[g'g'_{\wedge 1}] - \mathbb{E}[(\tilde{g}')^2] = \mathbb{E}_{x,y,z}[g'_{\wedge 1}(x,y)g(x,z)g(y,z) - \tilde{g}'(x,y)\tilde{g}(x,z)\tilde{g}(y,z)],$$

which is $o(1)$ by the induction hypothesis (replace $\nu_{XY}, g_{XY}, \tilde{g}_{XY}$ by $1, g'_{\wedge 1}, \tilde{g}'$, respectively, which by (8) satisfies $\|g'_{\wedge 1} - \tilde{g}'\|_{\square} = o(1)$, and note that this increases the number of $\{\nu_{XY}, \nu_{XZ}, \nu_{YZ}\}$ which are identically 1). One can similarly show that the other expectations on the right-hand side of (9) are also $\mathbb{E}[(\tilde{g}')^2] + o(1)$. Thus (9) is $o(1)$ and the theorem follows. \square