

# Local Structure: Subgraph Counts II

Graph Theory (Fall 2011)  
Rutgers University  
Swastik Kopparty

We now see some examples of interesting relationships between the hom numbers, as well as some of the diverse techniques that go into proving such relationships.

## 1 Triangles versus Edges

We will first focus on the relationship between the quantities  $n = \text{hom}(K_1, G)$ ,  $e = \text{hom}(K_2, G)$  and  $t = \text{hom}(K_3, G)$ .

As an example, note that  $e = n(n-1)$  iff  $t = n(n-1)(n-2)$  (since this happens iff  $G = K_n$ ). Turan's theorem tells us that if  $e > n^2/2$ , then  $t \geq 1$ . Below we will prove some more general inequalities between  $n$ ,  $e$  and  $t$  that capture these relations and much more.

**Theorem 1.**  $t \leq e^{3/2}$ .

*Proof.* This theorem is a corollary of the (much more general) Kruskal-Katona theorem. The Kruskal-Katona theorem has a very hands-on proof, based on iteratively modifying the graph. We will see a linear-algebraic proof.

Let  $A$  be the  $n \times n$  adjacency matrix of  $G$  ( $A_{uv} = 1$  if vertex  $u$  is adjacent to vertex  $v$ , and  $A_{uv} = 0$  otherwise). Note that  $A$  is symmetric.

It turns out that  $e$  and  $t$  are both fundamentally related to  $A$ .

$$\text{Tr}(A^2) = \sum_u \left( \sum_v A_{uv} A_{vu} \right) = \sum_{u,v} A_{uv} = e.$$

$$\text{Tr}(A^3) = \sum_u \sum_v \sum_w A_{uv} A_{vw} A_{wu} = t.$$

Now since  $A$  is symmetric, all its eigenvalues  $\lambda_1, \dots, \lambda_n$  are real, and so  $\text{Tr}(A^2) = \sum \lambda_i^2$  and  $\text{Tr}(A^3) = \sum \lambda_i^3$ .

Thus we have

$$e^{1/2} = \text{Tr}(A^2)^{1/2} = \left( \sum \lambda_i^2 \right)^{1/2} \geq \left( \sum |\lambda_i|^3 \right)^{1/3} \geq \left( \sum \lambda_i^3 \right)^{1/3} = \text{Tr}(A^3)^{1/3} = t^{1/3}.$$

So  $t \leq e^{3/2}$ . □

This theorem is essentially tight when  $G$  is the disjoint union of a clique on  $\sqrt{e}$  vertices and  $n - \sqrt{e}$  isolated vertices.

In the other direction, lots of edges imply the existence of a lot of triangles.

**Theorem 2** (Goodman).

$$t \geq e \cdot \left(2\frac{e}{n} - n\right).$$

*Proof.* This proof is based on direct counting.

We let  $d(u)$  denote the degree of  $u$ , and  $d(u, v)$  denote the number of common neighbors of  $u$  and  $v$ .

Notice that for every  $u, v \in V$ , we have  $d(u) + d(v) - d(u, v) \leq n$  (since this expression counts the number of vertices adjacent to at least one of  $u, v$ ).

Summing this up over all *adjacent*  $(u, v)$ , we get:

$$\sum_{(u,v) \in E} (d(u) + d(v) - d(u, v)) \leq n \cdot e.$$

Simplifying, we get:

$$\sum_{u \in V} d(u)^2 + \sum_{v \in V} d(v)^2 - t \leq n \cdot e.$$

Thus

$$t \geq 2 \sum_{u \in V} d(u)^2 - n \cdot e \geq 2 \frac{(\sum_{u \in V} d(u))^2}{n} - n \cdot e = 2\frac{e^2}{n} - n \cdot e.$$

□

It is instructive to see what would lead to equality in the above proof: (1) for every adjacent  $u, v$ , the union of the neighborhoods of  $u$  and  $v$  equals the set of all vertices, and (2) all vertices have the same degree. Complete  $r$ -partite graphs with parts of equal size have this property.

Let us see what the above inequalities tell us in terms of the density of edges and triangles. Let  $x = \frac{e}{n^2}$  and let  $y = \frac{t}{n^3}$  (note that  $x, y \in [0, 1]$ ). Then we have:

$$\max\{0, x(2x - 1)\} \leq y \leq x^{3/2}.$$

As remarked earlier, the right hand inequality is best possible for every  $x$ . However, the left hand inequality is not always best possible. Whenever  $x$  is of the form  $1 - \frac{1}{r}$  with  $r$  an integer, we have  $y = x(2x - 1)$  for the complete  $r$ -partite graph, but for every other  $x > 1/2$ , it turns out that  $y$  can never be as small as  $x(2x - 1)$ . The true behavior of  $y$  as a function of  $x$  turns out to be rather complex, and was only recently resolved by Razborov using some high-power machinery (“flag algebras”) designed especially for tackling inequalities between graph homomorphism numbers.

## 2 Generalizations

There are many directions in which the above results and methods can get generalized. We now see two direct generalizations.

**Theorem 3** ( $C_k$  vs  $C_\ell$ ). *If  $k \leq \ell$  and  $k$  is even, then*

$$\text{hom}(C_k, G)^{1/k} \geq \text{hom}(C_\ell, G)^{1/\ell}.$$

*Proof.* As in the proof of Theorem 1, we will relate it to the eigenvalues of  $G$ .

Notice that for any integer  $r$ ,

$$\text{hom}(C_r, G)^{1/r} = \text{Tr}(A^r)^{1/r} = \left( \sum_{i=1}^n \lambda_i^r \right)^{1/r}.$$

This implies the desired inequality (by monotonicity of norms).  $\square$

**Theorem 4** (Moon-Moser inequality for  $K_r$ ). *If  $G$  has  $n$  vertices and edge density  $\beta$ , then:*

$$\frac{\text{hom}(K_r, G)}{n^r} \geq \beta(2\beta - 1)(3\beta - 2) \dots ((t-1)\beta - (t-2)),$$

Notice that this implies Turan's theorem, and shows that when the edge density of a graph  $G$  is a constant larger than  $\text{ex}(n, K_r)/\binom{n}{2}$ , there is a constant density of  $K_r$  in  $G$ .

*Proof.* Define  $d(v_1, \dots, v_t)$  to be the number of common neighbors of  $v_1, \dots, v_t$ .

Let  $v_{[t]}$  be the tuple  $(v_1, \dots, v_t)$ . Denote by  $v_{[t]-i}$  the tuple of vertices  $(v_1, v_2, \dots, v_{i-1}, v_{i+1}, \dots, v_t)$ . Observe that for any vertices  $v_1, \dots, v_t$ , we have:

$$\left( \sum_{i=1}^t d(v_{[t]-i}) \right) - (t-1)d(v_{[t]}) \leq n,$$

(since the LHS counts the number of vertices adjacent to at least  $t-1$  of the vertices  $v_1, \dots, v_t$ ).

Let  $C_t$  denote the set of  $(v_1, \dots, v_t)$  such that  $v_1, \dots, v_t$  form a  $t$ -clique in  $G$ . Note that  $|C_t| = \text{hom}(K_t, G)$ .

Summing the above inequality up over  $(v_1, \dots, v_t) \in C_t$  ranging over all  $t$ -tuples which form  $t$ -cliques, we get:

$$\left( \sum_{i=1}^t \left( \sum_{v_{[t]-i} \in C_t} d(v_{[t]-i}) \right) \right) - (t-1) \sum_{v_{[t]} \in C_t} d(v_{[t]}) \leq n \text{hom}(K_t, G).$$

Thus:

$$t \cdot \left( \sum_{v_{[t-1]} \in C_{t-1}} d(v_{[t-1]})^2 \right) - (t-1) \text{hom}(K_{t+1}, G) \leq n \cdot \text{hom}(K_t, G).$$

Applying the Cauchy-Schwarz inequality to the first term, we get:

$$t \cdot \left( \frac{(\sum_{v_{[t-1]} \in C_{t-1}} d(v_{[t-1]})^2)}{\text{hom}(K_{t-1}, G)} \right) - (t-1) \text{hom}(K_{t+1}, G) \leq n \cdot \text{hom}(K_t, G),$$

which simplifies to:

$$\text{hom}(K_{t+1}, G) \geq \frac{1}{t-1} \left( t \cdot \frac{\text{hom}(K_t, G)^2}{\text{hom}(K_{t-1}, G)} - n \cdot \text{hom}(K_t, G) \right).$$

Letting  $\beta_t$  denote  $\frac{\text{hom}(K_t, G)}{\text{hom}(K_{t-1}, G) \cdot n}$ , we get:

$$\beta_{t+1} \geq \frac{1}{t-1}(t\beta_t - 1).$$

By induction, we can verify that this gives:

$$\beta_t \geq ((t-1)\beta_2 - (t-2)),$$

Thus the quantity of interest,  $\text{hom}(K_r, G)/n^r$  equals:

$$\beta_r \beta_{r-1} \dots \beta_2 \geq \beta_2(2\beta_2 - 1)(3\beta_2 - 2) \dots ((t-1)\beta_2 - (t-2)).$$

This completes the proof (note that  $\beta_2 = \beta$ ). □

### 3 Sidorenko's Conjecture

Sidorenko's conjecture states that for every bipartite graph  $F$  with  $k$ -vertices and  $t$  edges, and for every graph  $G$  with  $n$  vertices and edge density  $\beta$ :

$$\frac{\text{hom}(F, G)}{n^k} \geq \beta^t.$$

In words, the probability that a random function from  $V_F$  to  $V_G$  is a homomorphism is at least the  $t^{\text{th}}$  power of the edge density of  $G$ . The  $t^{\text{th}}$  power of the edge density is what the probability would have been if each edge of  $F$  created an *independent* constraint on the function. Sidorenko's conjecture basically says that these constraints are positively correlated.

This conjecture is open, although it is known for some special kinds of bipartite graphs. We now quickly verify Sidorenko's conjecture for stars and complete bipartite graphs. We will see a proof of Sidorenko's conjecture for general trees in the homework.

**Theorem 5.** *Let  $S_t$  be the star graph with  $t$  edges. Let  $G$  be a graph with  $n$  vertices and edge density  $\beta$ . Then*

$$\frac{\text{hom}(S_t, G)}{n^{t+1}} \geq \beta^t.$$

*Proof.* Note that  $\text{hom}(S_t, G) = \sum_{v \in V_G} d(v)^t$ . Thus

$$\frac{\text{hom}(S_t, G)}{n^{t+1}} = \mathbb{E}_{v \in V_G} \left[ \left( \frac{d(v)}{n} \right)^t \right].$$

Note also that

$$\beta = \frac{1}{n^2} \sum_{v \in V_G} d(v) = \mathbb{E}_{v \in V_G} \left[ \frac{d(v)}{n} \right].$$

The inequality now follows from Holder's inequality. □

**Theorem 6.** *Let  $r, s$  be positive integers. Let  $G$  be a graph with  $n$  vertices and edge density  $\beta$ .*

$$\frac{\text{hom}(K_{r,s}, G)}{n^{r+s}} \geq \beta^{rs}.$$

*Proof.* We will express everything analytically and use convexity. Below  $u_1, \dots, u_r, v_1, \dots, v_s$  and  $u, v$  are all uniformly distributed over  $V_G$ .  $E(u, v)$  is the function which equals 1 if  $u$  is adjacent to  $v$  in  $G$ , and 0 otherwise.

$$\begin{aligned} \frac{\text{hom}(K_{r,s}, G)}{n^{r+s}} &= \mathbb{E}_{u_1, \dots, u_r} \mathbb{E}_{v_1, \dots, v_s} \left[ \prod_{i=1}^r \prod_{j=1}^s A(u_i, v_j) \right] \\ &= \mathbb{E}_{u_1, \dots, u_r} \left( \mathbb{E}_v \left[ \prod_{i=1}^r A(u_i, v) \right] \right)^s \\ &\leq \left( \mathbb{E}_{u_1, \dots, u_r} \mathbb{E}_v \left[ \prod_{i=1}^r A(u_i, v) \right] \right)^s \\ &= \left( \mathbb{E}_v \left[ \mathbb{E}_{u_1, \dots, u_r} \left[ \prod_{i=1}^r A(u_i, v) \right] \right] \right)^s \\ &= (\mathbb{E}_v (\mathbb{E}_u [A(u, v)]^r))^s \\ &\leq (\mathbb{E}_{u,v} [A(u, v)]^{rs}) \\ &= \beta^{rs}. \end{aligned}$$

□

## 4 Algorithmically recognizing graph inequalities

At this stage it is natural to wonder what are all the graph inequalities that are true. We saw a wide range of techniques for proving all kinds of strong graph inequalities. Perhaps every graph inequality can be derived from these techniques?

To put this thought on a more concrete footing, let us consider the following computational problem:

### GRAPH-INEQUALITY:

- Input: Integers  $a_1, \dots, a_k, b$  and graphs  $F_1, \dots, F_k$ .
- Output: Is it true that for all graphs  $G$ , we have

$$\sum_{i=1}^k a_i \text{hom}(F_i, G) + b \geq 0.$$

Can we come up with an algorithm for GRAPH-INEQUALITY? Perhaps by combining some of the methods from above? Alternatively, one could imagine an algorithm which just tries out the inequality on all graphs  $G$  of some bounded size, and looks for a violation. Perhaps it is true that

if a graph inequality of the above type holds for all graphs  $G$  on  $\leq \exp(\exp(\exp(\sum |a_i| + \sum |F_i|)))$  vertices, then it automatically holds for all graphs  $G$ ?

The astonishing answer is NO! This problem is undecidable.

**Theorem 7. GRAPH-INEQUALITY** *is undecidable.*

This striking theorem is due to Ioannidis and Ramakrishnan, who proved it in the context of theoretical databases. Recently Hatami and Norine showed (the much stronger result) that linear inequalities between homomorphism *densities* are undecidable.

The main tool here is the undecidability of the solvability of diophantine equations in integers (Hilbert's 10th problem). This extremely fundamental theorem states that solving diophantine equations in integers is undecidable; this problem was posed by Hilbert (asking for an algorithm, well before Godel, Turing and the word "undecidable"), and was solved by Matiyasevic-Davis-Putnam-Robinson.

Formally, the result states that: Given a polynomial  $P(\mathbf{X}) \in \mathbb{Z}[\mathbf{X}]$ , the problem of determining whether there exists  $\mathbf{x} \in \mathbb{N}^t$  such that  $P(\mathbf{x}) = 0$  is undecidable.

*Proof.* We use the following form of the unsolvability of diophantine equations: Given a polynomial  $P(\mathbf{X}) \in \mathbb{Z}[\mathbf{X}]$ , the problem of determining whether there exists  $\mathbf{x} \in \mathbb{N}^t$  such that  $P(\mathbf{x}) \geq 0$  is undecidable. (This version follows easily from the above undecidability, since  $P(\mathbf{x}) = 0$  has a solution if and only if  $-P(\mathbf{x})^2 \geq 0$  has a solution.)

Given a polynomial  $P(\mathbf{X}) \in \mathbb{Z}[\mathbf{X}]$ , we construct  $a_1, \dots, a_k, b$  and  $F_1, \dots, F_k$  such that the two sets:

$$S = \left\{ \sum_{i=1}^k a_i \text{hom}(F_i, G) + b \mid G \text{ a finite graph} \right\},$$

$$T = \{P(\mathbf{x}) \mid \mathbf{x} \in \mathbb{N}^t\},$$

are equal. This implies the undecidability of GRAPH-INEQUALITY.

First, we consider some graphs  $H_1, \dots, H_t$  of a special kind.

**Fact 8.** *For every integer  $t > 0$ , there exist graphs  $H_1, \dots, H_t$  such that for every  $\mathbf{x} \in \mathbb{N}^t$ , there exists a graph  $G$  such that  $\text{hom}(H_i, G) = x_i$ .*

Such graphs are easy to construct (do it!).

Now we use these graphs to come up with the  $F_i$ . Let  $P(\mathbf{X}) = b + \sum_i a_i \prod_{j=1}^t X_j^{c_{i,j}}$ . Let  $F_i = \sum_{j=1}^t c_{i,j} H_j$ .

Then,  $P(\text{hom}(H_1, G), \text{hom}(H_2, G), \dots, \text{hom}(H_t, G)) = b + \sum a_i \text{hom}(F_i, G)$ . Thus,  $S \subseteq T$ .

Furthermore, for every  $\mathbf{x} \in \mathbb{N}^t$ , consider a graph  $G$  with  $\text{hom}(H_i, G) = x_i$  for each  $i$  (we know that such  $G$  exists because of the Fact above). Then  $P(\mathbf{x}) = b + \sum a_i \text{hom}(F_i, G)$ . Thus  $T \subseteq S$ .

So  $S = T$ , and we are done. □