# Supersaturation

Zdeněk Dvořák

December 11, 2020

For graphs $F$ and $G$, let $s_F(G)$ be the number of $|F|$-element sets $X \subseteq V(G)$ such that $F \subseteq G[X]$, i.e., the number of places where $F$ appears as a subgraph of $G$. If $G$ has more than $\mathrm{ex}(|G|; F)$ edges, then clearly $s_G(G) > 0$; can we say something about how large $s_F(G)$ is? It turns out that if the density of $G$ is larger than $\overline{\mathrm{ex}}(|G|; F)$, then $F$ appears in $G$ with positive density.

**Lemma 1.** *For any graph $F$ and any $\varepsilon > 0$, there exist $\beta > 0$ and $n_0$ such that the following holds. If $G$ is a graph with $n \geq n_0$ vertices and at least $(\overline{\mathrm{ex}}(\infty; F) + \varepsilon)\binom{n}{2}$ edges, then $s_F(G) \geq \beta n^{|F|})$.*

*Proof.* Without loss of generality, we can assume $\overline{\mathrm{ex}}(\infty; F) + \varepsilon \leq 1$. Let $n_0$ be the smallest integer greater than $2|F|$ such that $\overline{\mathrm{ex}}(n_0; F) < \overline{\mathrm{ex}}(\infty; F) + \varepsilon/2$. For a set $M \subseteq V(G)$ of size $n_0$ chosen uniformly at random, we have $E[\|G[M]\|/\binom{n_0}{2}] \geq \overline{\mathrm{ex}}(\infty; F) + \varepsilon$. Since $\|G[M]\|/\binom{n_0}{2} \leq 1$, we have

$$\Pr\left[\|G[M]\|/\binom{n_0}{2} \geq \overline{\mathrm{ex}}(\infty; F) + \varepsilon/2\right] \geq \frac{\varepsilon}{2(1 - \overline{\mathrm{ex}}(\infty; F) - \varepsilon/2)}.$$

Let $\gamma = \frac{\varepsilon}{2(1-\overline{\mathrm{ex}}(\infty;F)-\varepsilon/2)}$. Consider $X \subseteq V(G)$ of size $|F|$ chosen uniformly at random. We can imagine that we first choose $M$ at random and then choose $X \subseteq M$ at random. If $G[M]$ has at least $(\overline{\mathrm{ex}}(\infty; F) + \varepsilon/2)\binom{n_0}{2} > \overline{\mathrm{ex}}(n_0; F)$ edges, then $G[M]$ contains $F$ as a subgraph, and with probability at least $\binom{n_0}{|F|}^{-1}$ the set $X$ hits the vertex set of this subgraph. Therefore,

$$\Pr[F \subseteq G[X]] \geq \frac{\Pr[\|G[M]\|/\binom{n_0}{2} \geq \overline{\mathrm{ex}}(\infty; F) + \varepsilon/2]}{\binom{n_0}{|F|}} \geq \frac{\gamma}{\binom{n_0}{|F|}}.$$

In other words, there are at least $\gamma\binom{n_0}{|F|}^{-1}\binom{n}{|F|} \geq \gamma\binom{n_0}{|F|}^{-1}\frac{n^{|F|}}{2^{|F|}|F|!}$ sets $X$ such that $F \subseteq G[X]$. Hence, the claim of the lemma holds with

$$\beta = \frac{\gamma}{\binom{n_0}{|F|}2^{|F|}|F|!}.$$

□

In the rest of this text, we will study in more detain the behavior of the number of triangles in graphs whose density exceeds the bound given by Mantel's theorem. Let us remark that such a detailed study was also performed for cliques in graphs whose density exceeds the bound given by Turán's theorem, using similar ideas (but being technically rather more involved).

It will be useful to work in the induced subgraph setting. Let $i_F(G)$ denote the number of $|F|$-element subsets $X \subseteq V(G)$ such that $G[X]$ is isomorphic to $F$, and let

$$\mathrm{ex}_i(n, m; F) = \min\{i_F(G) : |G| = n, \|G\| = m\}.$$

Hence, we are interested in the behavior of the function $\mathrm{ex}_i(n, m; K_3)$ when $m > n^2/4$. Let us start with a simple result. Let $N_3$ denote the graph consisting of 3 isolated vertices.

**Lemma 2.** *If $G$ is a graph with degree sequence $d_1, \ldots, d_n$, then*

$$i_{K_3}(G) + i_{N_3}(G) = \binom{n}{3} - (n-2)\|G\| + \sum_{i=1}^{n} \binom{d_i}{2}.$$

*Proof.* Let $p$ denote the number of pairs $(x, \{y, z\})$, where $x$, $y$ and $z$ are distinct vertices and either $xy, xz \in E(G)$ or $xy, xz \notin E(G)$. On one hand,

$$p = 3i_{K_3}(G) + 3i_{N_3}(G) + i_{K_{1,2}}(G) + i_{\overline{K_{1,2}}}(G) = \binom{n}{3} + 2(i_{K_3}(G) + i_{N_3}(G)).$$

On the other hand,

$$\begin{aligned} p &= \sum_{i=1}^{n} \Big[ \binom{d_i}{2} + \binom{n - d_i - 1}{2} \Big] \\ &= \sum_{i=1}^{n} \Big[ \frac{(n-2)(n-1)}{2} - (n-2)d_i + 2\binom{d_i}{2} \Big] \\ &= 3\binom{n}{3} - 2(n-2)\|G\| + 2\sum_{i=1}^{n} \binom{d_i}{2}. \end{aligned}$$

Comparing the two expressions, we get the desired result. □

**Corollary 3.** *If $G$ is a graph with degree sequence $d_1, \ldots, d_n$, then*

$$i_{K_3}(G) \geq \frac{1}{3}\Big[ -(n-2)\|G\| + 2\sum_{i=1}^{n} \binom{d_i}{2} \Big],$$

*with equality iff $G$ is a complete multipartite graph.*

*Proof.* We have

$$\begin{aligned} i_{N_3}(G) &\leq \frac{1}{3}\sum_{i=1}^{n}\binom{n-1-d_i}{2} \\ &= \binom{n}{3} - \frac{2}{3}(n-2)\|G\| + \frac{1}{3}\sum_{i=1}^{n}\binom{d_i}{2}, \end{aligned}$$

with equality iff in $\overline{G}$, the neighborhood of each vertex induces a clique. This is the case iff $\overline{G}$ is a disjoint union of cliques, and thus $G$ is a complete multipartite graph. The desired inequality then follows from Lemma 2. □

**Corollary 4.**

$$\mathrm{ex}_i(n, m; K_3) \geq \frac{m(4m - n^2)}{3n},$$

*with equality iff* $m = t_r(n)$ *for some divisor* $r$ *of* $n$.

*Proof.* Let $G$ be a graph with $n$ vertices, $m$ edges and the degree sequence $d_1, \ldots, d_n$. Cauchy-Schwarz inequality gives

$$-(n-2)m + 2\sum_{i=1}^{n}\binom{d_i}{2} = -nm + \sum_{i=1}^{n} d_i^2 \geq -nm + 4\frac{m^2}{n},$$

with equality iff $d_1 = \ldots = d_n = \frac{2m}{n}$, i.e., $G$ is regular. Corollary 3 gives $i_{K_3}(G) \geq \frac{m(4m-n^2)}{3n}$, with equality iff $G$ is a complete multipartite graph. Hence, the equality holds iff $G$ is a regular multipartite graph, and thus $m = t_r(n)$ for some divisor $r$ of $n$. □

Hence, we know $\mathrm{ex}_i(n, m; K_3)$ exactly for certain isolated points. Next, we show that we can linearly interpolate between these points.

**Lemma 5.** *Let* $c$ *be an arbitrary real number and let* $\varepsilon$ *be a positive real number. For any integer* $n$*, the function* $f(G) = \|G\| + c \cdot i_{K_3}(G) + \varepsilon i_{N_3}(G)$ *is among all* $n$*-vertex graphs maximized only on complete multipartite graphs.*

*Proof.* Suppose $G$ is an $n$-vertex graph maximizing $f$ and consider any non-adjacent vertices $x$ and $y$ of $G$. Let $k_x = \|G[N(x)]\|$, $k_y = \|G[N(y)]\|$, $e_x = i_{N_2}(G - N[x] - y)$, and $e_y = i_{N_2}(G - N[y] - x)$. Let $G_x$ be the graph obtained from $G - y$ by adding a clone of $x$, and $G_y$ the graph obtained from $G - x$ by adding a clone of $y$. Letting $\delta = (\deg x + c \cdot k_x + \varepsilon e_x) - (\deg y + c \cdot k_y + \varepsilon e_y)$, we have

$$\begin{aligned} f(G_x) &= f(G) + \delta + \varepsilon(|N(x) \cup N(y)| - |N(x)|) \\ f(G_y) &= f(G) - \delta + \varepsilon(|N(x) \cup N(y)| - |N(y)|) \end{aligned}$$

Since $G$ maximizes $f$ among the $n$-vertex graphs and $\varepsilon > 0$, it follows that $\delta = 0$ and $|N(x)| = |N(x) \cup N(y)| = |N(y)|$, and thus $N(x) = N(y)$.

Therefore, any two non-adjacent vertices of $G$ have the same neighbors, and thus $G$ is a complete multipartite graph. □

**Corollary 6.** *Let $c$ be an arbitary irrational number. For any integer $n$, the function $f(G) = \|G\| + c \cdot i_{K_3}(G)$ is among all $n$-vertex graphs maximized on some Turán graph.*

*Proof.* For $\varepsilon > 0$, let $f_\varepsilon(G) = \|G\| + c \cdot i_{K_3}(G) + \varepsilon i_{N_3}(G)$. Since there are only finitely many $n$-vertex graphs, Lemma 5 implies that there exists a complete multipartite graph $G_0$ satisfying the following condition: For every $\varepsilon_0 > 0$, there exists a positive $\varepsilon < \varepsilon_0$ such that $f_\varepsilon(G) \leq f_\varepsilon(G_0)$ for every $n$-vertex graph $G$. Since $\lim_{\varepsilon \to 0} f_\varepsilon(G) = f(G)$, it follows that $f(G) \leq f(G_0)$ for every $n$-vertex graph $G$.

Let $a_1, \ldots, a_r$ be the sizes of the parts of $G_0$. If $r = 1$, then $G_0$ is the edgeless Turán graph $T_1(n)$; hence, we can assume $r \geq 2$. Let $\alpha = \sum_{i=3}^{r} a_i$, $\beta = \sum_{i<j} a_i a_j$, and $\gamma = \sum_{3 \leq i < j} a_i a_j + c \sum_{3 \leq i < j < k} a_i a_j a_k$. Consider the complete multipartite graph $G$ with parts of sizes $x, y, a_3, \ldots, a_r$, where $x + y = a_1 + a_2$. Then $f(G) = (1 + c\alpha)xy + (\alpha + c\beta)(a_1 + a_2) + \gamma$. Since $c$ is irrational we have $1 + c\alpha \neq 0$. If $1 + c\alpha$ were negative, then we could set $x = 0$ a $y = a_1 + a_2$ (i.e., let $G$ be the graph obtained from $G_0$ by merging two of its parts) and obtain $f(G) > f(G_0)$, which is a contradiction. Therefore, $1 + c\alpha > 0$. Since $G_0$ maximizes $f$, it follows that $xy \leq a_1 a_2$ for all nonnegative integers $x$ and $y$ such that $x + y = a_1 + a_2$, and thus $|a_1 - a_2| \leq 1$.

Symmetrically, we have $|a_i - a_j| \leq 1$ for all $i, j \in \{1, \ldots, r\}$, and thus $G_0 = T_r(n)$. □

For a positive integer $n$, let $\psi_n : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ be the maximum convex function such that $\psi_n(0) = 0$ and $\psi_n(t_r(n)) = i_{K_3}(T_r(n))$ holds for $r \in \{1, \ldots, n\}$.

**Theorem 7.**

$$\mathrm{ex}_i(n, m; K_3) \geq \psi_n(m).$$

*Proof.* Otherwise, there would exist an $n$-vertex graph $G_0$ such that $i_{K_3}(G_0) < \psi_n(\|G_0\|)$. The definition of $\psi_n$ implies that there exists a real (and without loss of generality irrational) $c$ such that the function $f(G) = \|G\| + c \cdot i_{K_3}(G)$ satisfies $f(G_0) > f(t_r(n))$ for every $r \in \{1, \ldots, n\}$. This contradicts Corollary 6. □

Let us remark that the function $\mathrm{ex}(n, m; K_3)$ is actually strictly concave between the points $\{t_r(n) : r \in \{1, \ldots, n\}\}$; this was proven by Razborov, who gave the exact formula for $\mathrm{ex}(n, m; K_3)$ using the flag algebras method.