Conditional Regularity and Efficient testing of bipartite graph properties

Ilan Newman
Haifa University

Based on work with Eldar Fischer and Noga Alon
Let $F$ be a graph property that is defined by a finite collection of forbidden induced graphs $\{F_1, \ldots \}$, each on at most $k$ vertices.

Def: For two (bipartite) graphs $G, H$ on the same set of $n$ vertices, we say that $G$ is $\varepsilon$-far from $H$ if

$|E(G) \Delta E(H)| > \varepsilon n^2$

Def: $G$ is $\varepsilon$-far from $F$ if it is $\varepsilon$-far from any $H$ that has $F$. 
Thm [AFKS99] If $G$ (large enough) is $\varepsilon$-far from $F$ then $\delta$-fraction of its random induced subgraphs of size $k$ are members of $F$.

Caveat: $\delta = \delta(\varepsilon, k) = 1/\text{tower(tower}(1/\varepsilon))$.

Best upper bound on $\delta$ is $(\varepsilon)^{\Omega(\log(1/\varepsilon))}$ [Alon02, Alon Shapira 03].

Our Goal: find a 'more efficient' version for bipartite graphs.
Main Result

Thm 1. If a bipartite $G$ (large enough) is $\varepsilon$-far from $F$ then $\delta$-fraction of its random induced subgraphs of size $k$ are members of $F$.

Here $\delta = \delta(\varepsilon, k) = \text{poly}(1/\varepsilon)$. 
• $A,B$ - disjoint set of vertices, \[ \text{density}(A,B) = \frac{e(A,B)}{|A||B|}. \]

• $(A,B)$ has density $< \delta$ or at least $1-\delta$, then $(A,B)$ is $\delta^{1/3}$-regular (in the Szemerédi sense). We call such a pair $\delta$-homogeneous.

• Regularity Lemma: there exists a partition to $O(1)$ sets in which most pairs are regular - But, can not expect strong regularity as above (e.g for a random graph in $G(n,1/2)$).

• We show that under some condition this is possible for bipartite graphs (and with very efficient partitions).
We move to 0/1, nxn matrices instead of bipartite graphs.

- **Partitions**: An r-partition of $M$ is a partition of its row set into $r' < r$ parts and column set into $r'' < r$ parts.
- **Blocks**: A subset of rows $R'$, and subset of columns $C'$ of $M$ define a block (pair in graph), which will be denoted by $(R',C')$. 
\( M: \)

\[
\begin{array}{|c|c|c|}
\hline
& C' \\
\hline R' & & \\
\hline
\end{array}
\]

r- partition (does not need to be of consecutive rows/columns).

A block \((R', C')\)

Note: the partition is not necessarily into equal size parts!!

Def: The weight of a block \((R', C')\) is \( |R'| |C'| / n^2 \)
Def:

Let $M$ be a $0/1$, $n \times n$ matrix with an $r$-partition of $M$, $P$. $P$ is said to be $(\delta, r)$-partition if the total weight of $\delta$-homogeneous blocks is at least $1 - \delta$.

Note: such a $P$ is a regular partition.
Thm2: For every $k, \delta > 0$ and matrix $M$ (large enough) either:

- $M$ has a $(\delta, r)$-partition with $r < (k/\delta)^{O(k)}$
- OR,
- For every $k \times k$, 0/1 matrix $B$, at least $g(\delta, k) = (k/\delta)^{O(k^2)}$-fraction of the $k \times k$ matrices of $M$ are $B$. 
Proof of the conditional regularity

Definition:

• For two vectors $u,v \in \{0,1\}^n$ (e.g. two rows or two columns) denote
  \[ \mu(u,v) = \text{hamming}(u,v)/n. \]

• An $r$-partition of the rows of $M$, \{\(V_0, V_1, \ldots, V_s\)\} is a \((\delta, r)\)-clustering if $s < r$, $|V_0| < \delta n$ and for every $i=1,\ldots,s$ if $u,v \in V_i$ then $\mu(u,v) < \delta$. 
Claim: A $(\delta, r)$-partition defines a $(4\delta^{1/3}, r)$-clustering of the rows.

Claim: The inverse (with different parameters) is also true.
Proof cont.

Let $T = (10k)^{2k}$. Let $F$ be a fixed $k \times k$ matrix.

Want to show that if the columns of $M$ cannot be $(\delta, T)$-clustered then a random $k \times k$ matrix of $M$ is $F$ with 'high probability.

- Chose a random set of columns $S$ (with repetitions) of size $5T/\delta$. We will chose a random set of $10k$ rows. Show that this submatrix contains (at least one copy of) $F$. 

Claim 1: $S$ contains $T$ columns of pairwise distance at least $0.5 \delta n$.

Simply by sequentially picking the $i$th column if its distance is at least $0.5 \delta n$ form all previously picked columns.

This could be done for $T$ steps as there is no $(\delta, T)$-clustering of col.
Claim 2: Assume that $S'$ is a set of $T$ columns of pairwise distance at least $0.5 \delta n$.

Then if we chose a set $R$ of 10k rows independently, at random. The projections of $S'$ on $R$ are distinct with high prob.
Proof: For a random row and two columns $c_1, c_2$ in $S'$, $\text{Prob}(c_1[r] = c_2[r]) < 1 - \delta/2$.

Thus, $c_1,c_2$ have the same projection on $R$ with probability $< (1 - \delta/2)^{10k}$.

Hence, expected number of 'bad pairs' is at most

$$\binom{T}{2}(1-\frac{\delta}{2})^{10k} < 0.1$$
Lemma: Every \((10k) \times T\) matrix with no two identical columns contains every possible \(k \times k\) submatrix.

Proof: let \(t=10k\), \(T=t^{2k}\)

Sauer-Perles-Shelah
Open Problems and comments

- Can there be an 'efficient' 'conditional regularity' version for general graphs?
  By Gowers this cannot be a 'syntactical' generalization of the result here.

If all forbidden subgraphs are bipartite the result here holds even for general graphs.
• Assume that $G$ is $\varepsilon$-far from being triangle free (that is: need to delete at least $\varepsilon n^2/2$ edges to cancel all triangles). Does $G$ contain $2^{-O(1/\varepsilon)} n^3$ triangles?

• What happens in higher dimensions?