A Fast New Algorithm for Weak Graph Regularity

László Miklós Lovász
MIT

Joint work with Jacob Fox and Yufei Zhao

January 11, 2019

Regularity between sets

Given X and Y sets of vertices of G, let

Definition

Given a graph G and two sets of vertices X and Y, we say the pair (X, Y) is ϵ-*regular* if for any $X' \subset X$ with $|X'| \geq \epsilon|X|$, $Y' \subset Y$ with $|Y'| \geq \epsilon|Y|$, we have

$$\Big| d(X', Y') - d(X, Y) \Big| \leq \epsilon.$$

- $d(X, Y) = \frac{e(X,Y)}{|X||Y|}$.
- $e(X, Y) =$ number of pairs of vertices in $X \times Y$ that have an edge between them.

Roughly says graph between X and Y is “random-like”.

Szemerédi's regularity lemma

Given a partition $\mathcal{P}$ of the set of vertices V, we say it is *equitable* if the size of any two parts differs by at most one. Given an equitable partition $\mathcal{P}$ of the set of vertices V, it is ϵ-*regular* if all but $\epsilon|\mathcal{P}|^2$ pairs are ϵ-regular. For every $\epsilon > 0$, there is an $M(\epsilon)$ such that for any graph $G = (V, E)$, there is an equitable, ϵ-regular partition of the vertices into at most $M(\epsilon)$ parts.

Drawback!

The standard proof gives $M(\epsilon) \leq T(\epsilon^{-5})$, where T is the tower function, i.e. $T(5) = 2^{2^{2^{2^{2}}}} = 2^{65536}$.

Unfortunately, Gowers (1997) showed $M(\epsilon) \geq T(\epsilon^{-c})$ for some c.

Frieze-Kannan regularity lemma: weaker regularity property, better bounds.

Frieze-Kannan (weak) regularity lemma

Definition

Given a partition $\mathcal{P} = \{V_1, V_2, ..., V_k\}$ of the set of vertices V, it is *Frieze-Kannan ϵ-regular* (FK-ϵ-regular) if for any pair of sets $S, T \subseteq V$, we have

$$\left| e(S,T) - \sum_{i,j=1}^{k} d(V_i, V_j)|S \cap V_i||T \cap V_j| \right| \leq \epsilon |V|^2$$

Frieze-Kannan regularity lemma

Let $\epsilon > 0$. Every graph has a Frieze-Kannan ϵ-regular partition with at most $2^{2/\epsilon^2}$ parts.

Definition

Given two (weighted) graphs G_1 and G_2 on the same vertex set V, we define their *cut distance*

Weak regularity

Given a partition $\mathcal{P}$, let $G_{\mathcal{P}}$ be the weighted graph obtained by taking, between each $X, Y \in \mathcal{P}$, weighted edges with weight $d(X, Y)$. Partition $\mathcal{P}$ is FK-ϵ-regular if and only if $d_{\square}(G, G_{\mathcal{P}}) \leq \epsilon$. The Frieze Kannan regularity lemma then says Let $\epsilon > 0$. Every graph has a partition with at most $2^{2/\epsilon^2}$ parts such that $d_{\square}(G, G_{\mathcal{P}}) \leq \epsilon$.

We also have the counting lemma. Given two graphs G_1 and G_2 on the same vertex set, for any graph H on k vertices, we have

$$|\hom(H, G_1) - \hom(H, G_2)| \leq e(H) d_{\square}(G_1, G_2) n^k.$$

Algorithmic regularity

Alon-Duke-Lefmann-Rödl-Yuster in 1994 showed that it is NP-hard to test whether a pair (X, Y) is ϵ-regular. :(But! If a pair (X, Y) not ϵ-regular, they can find S, T showing they are not $\epsilon^4/16$-regular, time $O_\epsilon(n^{\omega+o(1)})$. This implies an algorithm with at most $T(\epsilon^{-O(1)})$ parts. ($\omega < 2.373$)
In terms of deterministic algorithms, we have

- Frieze-Kannan (1999) - Spectral approach.
- Kohayakawa-Rödl-Thoma (2003) - $O_\epsilon(n^2)$-time algorithm.
- Alon-Naor (2006) - Polynomial-time algorithm, at most $T(O(\epsilon^{-7}))$ parts.

There is also a folklore result in a Tao blog post (2010): Randomized algorithm in time $O_\epsilon(1)$, ϵ-regular partition.

Algorithmic Frieze-Kannan

Frieze-Kannan (1996)

Constant time *probabilistic* algorithm.

Dellamonica-Kalyanasundaram-Martin-Rödl-Shapira

Deterministic algorithm which finds a Frieze-Kannan ϵ-regular partition

- in time $\epsilon^{-6} n^{\omega+o(1)}$ into at most $2^{O(\epsilon^{-7})}$ parts (2012)
- in time $O(2^{2^{\epsilon^{-O(1)}}} n^2)$ into at most $2^{\epsilon^{-O(1)}}$ parts (2015)

Fox-L.-Zhao

Deterministic algorithm which finds a Frieze-Kannan ϵ-regular partition in time $\epsilon^{-O(1)} n^2$.

Algorithmic Frieze-Kannan

We can prove something slightly stronger.

Fox-L.-Zhao

There is an $\epsilon^{-O(1)} n^2$-time algorithm which, given $\epsilon > 0$, an n-vertex graph G, outputs $r \leq \epsilon^{-O(1)}$, subsets $S_1, S_2, \ldots, S_r, T_1, T_2, \ldots, T_r \subset V(G)$ and numbers $c_1, c_2, \ldots, c_r = \pm\epsilon^8/300$ such that

$$d_\square(G, c_1 K_{S_1,T_1} + c_2 K_{S_2,T_2} + \ldots + c_r K_{S_r,T_r}) \leq \epsilon.$$

Main lemma

Main step

There is a $(C/\epsilon)^{O(1)}n^2$-time algorithm which does the following. Suppose that an $n \times n$ matrix A has $\|A\|_{max} \leq C$, and each row $\|\mathbf{a}_i\|_2^2 \leq n$ and column $\|\mathbf{a}^j\|_2^2 \leq n$. Then the algorithm outputs either

- that each singular value of A is at most ϵn, i.e. its operator norm $\|A\| \leq \epsilon n$, or
- sets $S, T \subseteq [n]$ such that

$$\left| \sum_{i \in S, k \in T} a_{i,k} \right| \geq \frac{\epsilon^8}{100} n^2.$$

Main idea

Let A be a "bounded" matrix. The following are "roughly equivalent" (Chung-Graham-Wilson, Thomason):

- $\|A\|_{\square} = o(n^2)$.
- Singular values of A are all $o(n)$.
- $trAA^TAA^T = o(n^4)$.
- Most pairs of rows have inner product $o(n^2)$.

Key idea: Can check fourth condition for pairs along a constant degree *expander graph*.

If off, can find a pair of sets with non-small density.

Idea of testing along expanders first appeared in Kohayakawa-Rödl-Thoma paper.

Iterative proof

Sequence of iterative steps, $A = A_0, A_1, \cdots, A_s$.
In each step, apply main lemma to A_l. If $\|A_l\| \leq \epsilon n$, done.
Else find sets S and T. WLOG sum across $S \times T$ is $\gtrsim \epsilon^8 n^2$.
By slightly decreasing sum across sets, can assume each row and column has positive, not too small sum.
Set $A_{l+1} = A_l - t 1_S 1_T^{\mathsf{T}}$, $t = c\epsilon^8$. L^2-norm of each row and column cannot increase, and $\|A\|_{Fr}^2$ decreases by $\gtrsim \epsilon^{16} n^2$.
Must end after $O(\epsilon^{-16})$ steps.

Counting subgraphs

Suppose we want to count how many times H appears in G?
What if we just want to count is there a single copy?
Even for K_k, Zuckerman showed NP-hard to approximate the size of the largest clique within a factor $n^{1-\epsilon}$, building on an earlier result of Håstad.
How fast can we approximate the count within an additive $\epsilon n^{|V(H)|}$?
A simple randomized algorithm gives 99% certainty:
Sample $10/\epsilon^2$ random k-sets of vertices.
What about deterministic algorithms?

Counting subgraphs

Algorithmic problem

Count the number of copies of a graph H on k vertices in a graph G on n vertices, up to an error of at most ϵn^k.

Duke-Lefmann-Rödl (1996)

Can be done in time $2^{(k/\epsilon)^{O(1)}} n^{\omega+o(1)}$.

Fox-L.-Zhao (2017)

Can be done in time $\epsilon^{-O_H(1)} n^2$.