Regularity, uniformity, and quasirandomness

Jozsef Solymosi[†]

Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2

Graph theory is the appropriate language for discussing binary relations on objects. Results in graph theory have numerous applications in biology, chemistry, computer science, and physics. In cases of multiple relations, instead of binary relations more general structures known as hypergraphs are the right tools. However, it turns out that because of their extremely complex structure, hypergraphs are very difficult to deal with. As with number theory, there are questions about hypergraphs that are easy to state but very difficult to answer. In this issue of PNAS, Rödl *et al.* (1) extend a powerful tool, the regularity lemma, from graphs to hypergraphs.

Contrary to the general terminology, in extremal graph theory regularity is a measure of randomness. Random graphs are easy to work with, especially when one wants to estimate the (expected) number of small subgraphs. In complex structures, like in dense graphs, one can substitute randomness with weaker but still useful properties. The motivation behind graph regularity is to arrange the vertices of a graph in such a way that the graph becomes similar to the union of a few random graphs, and then one can apply standard counting methods from probability theory. In order to define hypergraph regularity, one has to introduce somehow complicated and technical notations. However, even without these notations we can formulate the most important consequence of the so-called hypergraph regularity method. The method, which is the combination of the hypergraph regularity lemma and a counting lemma is described by Rödl *et al.* (1). Similar results with the same consequences have been obtained independently by Gowers (2). Inspired by the methods of refs. 1 and 2, very recently Tao (T. Tao, personal communication) gave another proof of the main results. The road to the hypergraph regularity and counting lemmas was long and challenging.

Graph Regularity

Graph regularity was first introduced by Szemerédi (3), who used it to prove his celebrated theorem that every dense subset of integers contains arbitrary long arithmetic progressions. Today, one of the main tools in extremal graph theory is Szemerédi's regularity lemma (4), which makes arbitrary (usually large and dense) graphs manageable.[‡] It was

widely expected that hypergraph regularity could provide a similarly useful tool to deal with hypergraphs. The problem is that one can easily formulate fake hypergraph regularity lemmas by simply generalizing the original regularity lemma. The question was if one can find the "right" hypergraph lemma that can be used to prove theorems that do not follow from an application of the ordinary regularity lemma. Chung (5) was

Rödl *et al.* extend a powerful tool, the regularity lemma, from graphs to hypergraphs.

the first to come up with generalizations of regularity; however, her result had certain limitations. Her findings were not strong enough for applications to Szemerédi-type theorems, but still they formed a significant precursor to the more modern hypergraph regularity lemmas. After several years of hard work, Rödl and his students (1) have devised a solution providing a right notation of hypergraph regularity and proving the corresponding theorems using purely combinatorial tools. Gowers' approach (2) is somehow different, more analytic. The notations and proofs are related to his earlier proof of Szemerédi's theorem using Fourier analysis (6). One should mention here that the Cauchy–Schwarz-type arguments Gowers uses in his counting lemma were very influential in the recent results of Green and Tao (7) on long arithmetic progressions in the primes.

An Important Corollary

Graphs and hypergraphs are general combinatorial objects. A graph G is given by its vertex set $V(G)$ and the edge set $E(G)$, a list of vertex pairs that are connected by an edge. The notation of a hypergraph is similar. Given a set S as the vertex set, a family of the subsets of S will define the hyperedges. In this paper, we will focus on k -uniform hypergraphs, on hypergraphs where all the edges have the same size, k . With this notation, the two uniform hypergraphs are the ordinary graphs.

Given a k -uniform hypergraph, H_k^n , on an n -element vertex set, $V(H_k^n)$ a clique, K_{k+1} , is a $k+1$ -element subset of $V(H_k^n)$ such that any k -tuple of K_{k+1} is an edge of the hypergraph H_k^n . Two cliques are said to be edge-disjoint if they don't have a common edge. Any set of pairwise edge-disjoint cliques in H_k^n has cardinality at most $\binom{n}{k}/(k+1)$ because every clique has $k+1$ edges. The main result of ref. 1 is that if a hypergraph contains a large set, S , of pairwise edge-disjoint cliques, then it contains many cliques. In particular, the hypergraph contains at least one clique that is not in S . We will refer to the result below as the *Removal Lemma* for k -uniform hypergraphs. The reason why it is called *Removal Lemma* is that one can formulate the statement in the following equivalent way. If a hypergraph contains few cliques, then after removing only few edges from the hypergraph, the remaining hypergraph will not contain cliques at all.

Removal Lemma. For any $c > 0$ real number and $k \geq 2$ integer, there is a $\delta > 0$ that depends on c and k only, such that the following is true. If H_k^n contains a set, S , of pairwise edge-disjoint cliques with cardinality $|S| \geq c \binom{n}{k}$, then H_k^n contains at least $\delta \binom{n}{k+1}$ cliques.

A typical application of the result would be as follows. We want to prove that a given hypergraph contains two cliques sharing an edge. If we can show that there is a large set of pairwise edge-disjoint cliques, then we are done. To illustrate the method, we prove a generalization of Roth's theorem (8) about three-term arithmetic progressions in dense subsets of integers. We will show that if S is a dense subset of a large $N \times N$ integer grid, then S contains an isosceles equilateral triangle, three points with coordinates (x, y) ,

See companion article on page 8109.

[†]E-mail: solymosi@math.ubc.ca.

[‡]For a graph $G = (V, E)$ and two disjoint sets $V_1, V_2 \subset V$, we denote by $E(V_1, V_2)$ the set of edges with one endpoint in V_1 and one endpoint in V_2 . The density $d(V_1, V_2)$ is given by $d(V_1, V_2) = |E(V_1, V_2)| / (|V_1||V_2|)$. We say that the graph induced by V_1, V_2 is ϵ -regular if for all $V_1' \subset V_1$ and $V_2' \subset V_2$ with $|V_1'| \geq \epsilon|V_1|$ and $|V_2'| \geq \epsilon|V_2|$, $|d(V_1', V_2') - d(V_1, V_2)| \leq \epsilon$. Szemerédi's regularity lemma claims that for any $\epsilon > 0$ there is a number, $t = t(\epsilon)$, such that any graph's vertex set can be partitioned into t almost equal vertex classes such that with only ϵt^2 exemptions the bipartite graphs between the classes are ϵ -regular.

© 2005 by The National Academy of Sciences of the USA

Fig. 1. Take a tripartite graph in which the vertices of the graph are the red, yellow, and green lines and the edges are defined by the set S . Two vertices are connected by an edge if the crossing point of the corresponding lines is a point of S . A triangle in the graph corresponds to three lines such that any two intersect in a point of S . If there are two triangles sharing an edge, then at least one triangle is not degenerate; thus, we have an isosceles equilateral triangle in S . If S is a dense subset of a large grid, then by the *Triangle Removal Lemma* there are many triangles in the graph. Therefore, there is an edge that is the edge of two triangles, so S contains an isosceles equilateral triangle.

$(x + d, y)$, and $(x, y + d)$, where d is a non-zero integer. It is easy to see that the statement implies Roth's theorem (Fig. 1).

The very same trick can be applied for higher dimensional grids, hyperplanes, and hypergraphs. This calculation leads us to a combinatorial proof of the so-called multidimensional Szemerédi theorem, which was proved by Furstenberg and Katznelson (9) using ergodic theory.

It is not known how δ depends on c . Even in the simplest case, $k = 2$, the gap between the best known upper and lower bounds is huge. When n is large enough, $\delta(\frac{n}{k+1})$ is larger than $\binom{n}{k}/(k+1)$, so

there is at least one clique in H_k^n that is not in S . It is surprising that this seemingly weak statement needs such heavy machinery. In most of the applications, all we need is to show that in a hypergraph there are two cliques that have a common edge. Random hypergraphs almost surely have such a pair of cliques. Therefore, if one can show that a given hypergraph is somehow similar to the random hypergraph, then this could lead to the proof. What we want from a hypergraph regularity lemma is to find for a given hypergraph, H_k^n a partition of the one-, two-, three-, ..., $(k - 1)$ -element subsets of $V(H_k^n)$ into few classes such that the subgraphs,

spanned by the classes, behave in a random-like way with only few exceptions. Also, one should come up with the right definition of "random-like." This plan is nice, but unfortunately for $k > 2$ the solution is quite complicated. In 1978, for $k = 2$, Ruzsa and Szemerédi (10) proved that graph regularity implies the *Removal Lemma* for graphs. What Ruzsa and Szemerédi proved by using the regularity lemma for graphs is the following.

Triangle Removal Lemma. *If a graph on n vertices contains at least cn^2 edge disjoint triangles, then it contains at least δn^3 triangles.*

It was 25 years later when Frankl and Rödl (11) published the $k = 3$ case. This shows how difficult it was to find the right generalization of graph regularity to hypergraphs. There is a test to decide whether a hypergraph regularity is useful or not. Does it imply the *Removal Lemma*? If the answer is yes, then it is a correct concept of regularity indeed. On the contrary, applications of the hypergraph regularity could go beyond the *Removal Lemma*. There are already examples for which the hypergraph regularity method, combined with ergodic theory, analysis, and number theory, are used efficiently to solve difficult problems in mathematics.

1. Rödl, V., Nagle, B., Skokan, J., Schacht, M. & Kohayakawa, Y. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 8109–8113.
2. Gowers, W. T. (2005) *Comb. Probab. Comput.*, in press.
3. Szemerédi, E. (1975) *Acta Arith.* **27**, 199–245.
4. Szemerédi, E. (1978) in *Problèmes Combinatoires et Théorie des Graphes* (Centre Natl. Rech. Sci., Paris), pp. 399–401.
5. Chung, F. R. K. (1990) *Random Struct. Algorithms* **1**, 363–382.
6. Gowers, W. T. (2001) *Geom. Funct. Anal.* **11**, 465–588.
7. Green, B. & Tao, T. (2005) *Ann. Math.*, in press.
8. Roth, K. F. (1953) *J. London Math. Soc.* **28**, 104–109.
9. Furstenberg, H. & Katznelson, Y. (1978) *J. Anal. Math.* **34**, 275–291.
10. Ruzsa, I. & Szemerédi, E. (1976) *Comb. Coll. Math. Soc. J. Bolyai* **18**, 939–945.
11. Frankl, P. & V. Rödl, V. (2002) *Random Struct. Algorithms* **20**, 131–164.