Szemerédi's Regularity Lemma

Chelsea Edmonds, Angeliki Koutsoukou-Argyraki and Lawrence C. Paulson
Computer Laboratory, University of Cambridge CB3 0FD
{cle47,ak2110,lp15}@cam.ac.uk

March 17, 2025

Abstract

Szemerédi's regularity lemma [2] is a key result in the study of large graphs. It asserts the existence an upper bound on the number of parts the vertices of a graph need to be partitioned into such that the edges between the parts are random in a certain sense. This bound depends only on the desired precision and not on the graph itself, in the spirit of Ramsey's theorem. The formalisation follows online course notes by Tim Gowers[1] and Yufei Zhao[2]. Similar material is found in many textbooks [1].

Contents

Acknowledgements

The authors were supported by the ERC Advanced Grant ALEXANDRIA (Project 742178) funded by the European Research Council.

[1]https://www.dpmms.cam.ac.uk/~par31/notes/tic.pdf
[2]https://yufeizhao.com/gtacbook/ and https://yufeizhao.com/gtac/gtac.pdf are drafts of a textbook in preparation.

References

[1] R. Diestel. *Graph Theory*. Springer, 2017.

[2] E. Szemerédi. Regular partitions of graphs. Technical Report STAN-CS-75-489, Stanford University Computer Science Department, Apr. 1975.

1 Szemerédi's Regularity Lemma

theory *Szemeredi*
imports *HOL−Library.Disjoint-Sets Girth-Chromatic.Ugraphs HOL−Analysis.Convex*

begin

We formalise Szemerédi's Regularity Lemma, which is a major result in the study of large graphs (extremal graph theory). We follow Yufei Zhao's notes "Graph Theory and Additive Combinatorics" (MIT), latest version here: https://yufeizhao.com/gtacbook/ and W.T. Gowers's notes "Topics in Combinatorics" (University of Cambridge, Lent 2004, Chapter 3) https://www.dpmms.cam.ac.uk/~par31/notes/tic.pdf. We also used an earlier version of Zhao's book: https://yufeizhao.com/gtac/gtac.pdf.

1.1 Partitions

1.1.1 Partitions indexed by integers

definition *finite-graph-partition* :: [*uvert set*, *uvert set set*, *nat*] ⇒ *bool*
where *finite-graph-partition V P n* ≡ *partition-on V P* ∧ *finite P* ∧ *card P* = *n*

lemma *finite-graph-partition-0* [*iff*]:
finite-graph-partition V P 0 ⟷ *V* = {} ∧ *P* = {}
⟨*proof*⟩

lemma *finite-graph-partition-empty* [*iff*]:
finite-graph-partition {} *P n* ⟷ *P* = {} ∧ *n* = *0*
⟨*proof*⟩

lemma *finite-graph-partition-equals*:
finite-graph-partition V P n ⟹ (⋃*P*) = *V*
⟨*proof*⟩

lemma *finite-graph-partition-subset*:
⟦*finite-graph-partition V P n*; *X* ∈ *P*⟧ ⟹ *X* ⊆ *V*
⟨*proof*⟩

lemma *trivial-graph-partition-exists*:
assumes *V* ≠ {}
shows *finite-graph-partition V* {*V*} (*Suc 0*)
⟨*proof*⟩

lemma *finite-graph-partition-finite*:
assumes *finite-graph-partition V P k finite V X* ∈ *P*
shows *finite X*
⟨*proof*⟩

lemma *finite-graph-partition-gt0*:

assumes *finite-graph-partition V P k finite V X* $\in$ *P*
shows *card X* $>$ *0*
⟨*proof*⟩

lemma *card-finite-graph-partition*:
assumes *finite-graph-partition V P k finite V*
shows $(\sum X{\in}P.\ card\ X) = card\ V$
⟨*proof*⟩

1.1.2 Tools to combine the refinements of the partition *P i* for each *i*

These are needed to retain the “intuitive” idea of partitions as indexed by integers.

1.2 Edges

All edges between two sets of vertices, *X* and *Y*, in a graph, *G*

definition *all-edges-between* :: *nat set* ⇒ *nat set* ⇒ *nat set* × *nat set set* ⇒ (*nat* × *nat*) *set*
where *all-edges-between X Y G* ≡ {(*x*,*y*). *x*∈*X* ∧ *y*∈*Y* ∧ {*x*,*y*} ∈ *uedges G*}

lemma *all-edges-between-subset*: *all-edges-between X Y G* ⊆ *X*×*Y*
⟨*proof*⟩

lemma *max-all-edges-between*:
assumes *finite X finite Y*
shows *card* (*all-edges-between X Y G*) ≤ *card X* ∗ *card Y*
⟨*proof*⟩

lemma *all-edges-between-empty* [*simp*]:
all-edges-between {} *Z G* = {} *all-edges-between Z* {} *G* = {}
⟨*proof*⟩

lemma *all-edges-between-disjnt1*:
assumes *disjnt X Y*
shows *disjnt* (*all-edges-between X Z G*) (*all-edges-between Y Z G*)
⟨*proof*⟩

lemma *all-edges-between-disjnt2*:
assumes *disjnt Y Z*
shows *disjnt* (*all-edges-between X Y G*) (*all-edges-between X Z G*)
⟨*proof*⟩

lemma *all-edges-between-Un1*:
all-edges-between (*X* ∪ *Y*) *Z G* = *all-edges-between X Z G* ∪ *all-edges-between Y Z G*
⟨*proof*⟩

lemma *all-edges-between-Un2*:
$all\text{-}edges\text{-}between\ X\ (Y \cup Z)\ G = all\text{-}edges\text{-}between\ X\ Y\ G \cup all\text{-}edges\text{-}between\ X\ Z\ G$
⟨*proof*⟩

lemma *finite-all-edges-between*:
assumes *finite X finite Y*
shows *finite* (*all-edges-between X Y G*)
⟨*proof*⟩

1.3 Edge Density and Regular Pairs

The edge density between two sets of vertices, X and Y, in G. Authors disagree on whether the sets are assumed to be disjoint!. Quite a few authors assume disjointness, e.g. Malliaris and Shelah https://www.jstor.org/stable/23813167.

definition $edge\text{-}density\ X\ Y\ G \equiv card(all\text{-}edges\text{-}between\ X\ Y\ G)\ /\ (card\ X * card\ Y)$

lemma *edge-density-ge0*: $edge\text{-}density\ X\ Y\ G \geq 0$
⟨*proof*⟩

lemma *edge-density-le1*: $edge\text{-}density\ K\ Y\ G \leq 1$
⟨*proof*⟩

lemma *all-edges-between-swap*:
$all\text{-}edges\text{-}between\ X\ Y\ G = (\lambda(x,y).\ (y,x))$ ‘ $(all\text{-}edges\text{-}between\ Y\ X\ G)$
⟨*proof*⟩

lemma *card-all-edges-between-commute*:
$card\ (all\text{-}edges\text{-}between\ X\ Y\ G) = card\ (all\text{-}edges\text{-}between\ Y\ X\ G)$
⟨*proof*⟩

lemma *edge-density-commute*: $edge\text{-}density\ X\ Y\ G = edge\text{-}density\ Y\ X\ G$
⟨*proof*⟩

ϵ-regular pairs, for two sets of vertices. Again, authors disagree on whether the sets need to be disjoint, though it seems that overlapping sets cause double-counting. Authors also disagree about whether or not to use the strict subset relation here. The proofs below are easier if it is strict but later proofs require the non-strict version. The two definitions can be proved to be equivalent under fairly mild conditions, but even those conditions turn out to be onerous.

definition *regular-pair*:: *real* ⇒ *uvert set* ⇒ *uvert set* ⇒ *ugraph* ⇒ *bool*
(‹-−*regular*′-*pair*› [*999*]*1000*)
where $\varepsilon{-}regular\text{-}pair\ X\ Y\ G \equiv$
$\forall\ A\ B.\ A \subseteq X \wedge B \subseteq Y \wedge (card\ A \geq \varepsilon * card\ X) \wedge (card\ B \geq \varepsilon * card\ Y) \longrightarrow$

$|edge\text{-}density\ A\ B\ G - edge\text{-}density\ X\ Y\ G| \le \varepsilon$ **for** ε::*real*

lemma *regular-pair-commute*: ε−*regular-pair X Y G* $\longleftrightarrow$ ε−*regular-pair Y X G*
⟨*proof*⟩

lemma *edge-density-Un*:
assumes *disjnt X1 X2 finite X1 finite X2*
shows *edge-density* (*X1* ∪ *X2*) *Y G* = (*edge-density X1 Y G* ∗ *card X1* + *edge-density X2 Y G* ∗ *card X2*) / (*card X1* + *card X2*)
⟨*proof*⟩

lemma *edge-density-partition*:
assumes *finite-graph-partition U P n*
shows *edge-density U W G* = ($\sum$ *X*∈*P*. *edge-density X W G* ∗ *card X*) / *card U*
⟨*proof*⟩

Let P, Q be partitions of a set of vertices V. Then P refines Q if for all $A \in P$ there is $B \in Q$ such that $A \subseteq B$.

For the sake of generality, and following Zhao's Online Lecture https://www.youtube.com/watch?v=vcsxCFSLyP8&t=16s we do not impose disjointness: we do not include $i \neq j$ below.

definition *irregular-set*:: [*real*, *ugraph*, *uvert set set*] ⇒ (*uvert set* × *uvert set*) *set*
(‹-−*irregular′-set*› [*999*]*1000*)
where ε−*irregular-set* ≡ λ*G P*. {(*R*,*S*)|*R S*. *R*∈*P* ∧ *S*∈*P* ∧ ¬ ε−*regular-pair R S G*}
for ε::*real*

A regular partition may contain a few irregular pairs as long as their total size is bounded as follows.

definition *regular-partition*:: [*real*, *ugraph*, *uvert set set*] ⇒ *bool*
(‹-−*regular′-partition*› [*999*]*1000*)
where
ε−*regular-partition* ≡ λ*G P* .
partition-on (*uverts G*) *P* ∧
($\sum$ (*R*,*S*) ∈ *irregular-set* ε *G P*. *card R* ∗ *card S*) ≤ ε ∗ (*card* (*uverts G*))^2 **for** ε::*real*

lemma *irregular-set-subset*: ε−*irregular-set G P* ⊆ *P* × *P*
⟨*proof*⟩

lemma *irregular-set-swap*: (*i*,*j*) ∈ ε−*irregular-set G P* $\longleftrightarrow$ (*j*,*i*) ∈ ε−*irregular-set G P*
⟨*proof*⟩

lemma *finite-irregular-set* [*simp*]: *finite P* ⟹ *finite* (ε−*irregular-set G P*)
⟨*proof*⟩

1.4 Energy of a Graph

Definition 3.7 (Energy), written $q\ (U,\ W)$

definition *energy-graph-subsets*:: [*uvert set*, *uvert set*, *ugraph*] ⇒ *real* **where**
energy-graph-subsets U W G ≡
card U ∗ *card W* ∗ (*edge-density U W G*)^2 / (*card* (*uverts G*))^2

Definition for partitions

definition *energy-graph-partitions* :: [*ugraph*, *uvert set set*, *uvert set set*] ⇒ *real*
where *energy-graph-partitions G P Q* ≡ $\sum$ *R*∈*P*.$\sum$ *S*∈*Q*. *energy-graph-subsets R S G*

lemma *energy-graph-subsets-0* [*simp*]:
energy-graph-subsets {} *B G* = *0* *energy-graph-subsets A* {} *G* = *0*
⟨*proof*⟩

lemma *energy-graph-subsets-ge0* [*simp*]:
energy-graph-subsets U W G ≥ *0*
⟨*proof*⟩

lemma *energy-graph-partitions-ge0* [*simp*]:
energy-graph-partitions G U W ≥ *0*
⟨*proof*⟩

lemma *energy-graph-subsets-commute*:
energy-graph-subsets U W G = *energy-graph-subsets W U G*
⟨*proof*⟩

lemma *energy-graph-partitions-commute*:
energy-graph-partitions G W U = *energy-graph-partitions G U W*
⟨*proof*⟩

Definition 3.7 (Energy of a Partition), or following Gowers, mean square density: a version of energy for a single partition of the vertex set.

abbreviation *mean-square-density* :: [*ugraph*, *uvert set set*] ⇒ *real*
where *mean-square-density G P* ≡ *energy-graph-partitions G P P*

lemma *mean-square-density*:
mean-square-density G U ≡
($\sum$ *R*∈*U*. $\sum$ *S*∈*U*. *card R* ∗ *card S* ∗ (*edge-density R S G*)^2) / (*card* (*uverts G*))^2
⟨*proof*⟩

Observation: the energy is between 0 and 1 because the edge density is bounded above by 1.

lemma *sum-partition-le*:
assumes *finite-graph-partition V P k finite V*
shows ($\sum$ *R*∈*P*. $\sum$ *S*∈*P*. *real* (*card R* ∗ *card S*)) ≤ (*real*(*card V*))^2
⟨*proof*⟩

lemma *mean-square-density-bounded*:
  assumes *finite-graph-partition* (*uverts G*) *P k finite* (*uverts G*)
  shows *mean-square-density G P* $\leq$ *1*
$\langle$*proof*$\rangle$

1.5 Partitioning and Energy

See Gowers's remark after Lemma 11. Further partitioning of subsets of the vertex set cannot make the energy decrease. We follow Gowers's proof, which avoids the use of probability.

lemma *sum-products-le*:
  fixes *a* :: *'a* $\Rightarrow$ *real*
  assumes $\bigwedge i.\ i \in I \Longrightarrow a\ i \geq 0$
  shows $(\sum i{\in}I.\ a\ i * b\ i)^2 \leq (\sum i{\in}I.\ a\ i) * (\sum i{\in}I.\ a\ i * (b\ i)^2)$ (**is** *?L* $\leq$ *?R*)
$\langle$*proof*$\rangle$

lemma *energy-graph-partition-half*:
  assumes *P*: *finite-graph-partition U P n*
  shows *card U* $*$ (*edge-density U W G*)^2 $\leq$ ($\sum$ *R*$\in$*P*. *card R* $*$ (*edge-density R W G*)^2)
$\langle$*proof*$\rangle$

proposition *energy-graph-partition-increase*:
  assumes *P*: *finite-graph-partition U P k* **and** *V*: *finite-graph-partition W Q l*
  shows *energy-graph-partitions G P Q* $\geq$ *energy-graph-subsets U W G*
$\langle$*proof*$\rangle$

The following is the fully general version of Gowers's Lemma 11 Further partitioning of subsets of the vertex set cannot make the energy decrease. Note that *V* should be *uverts G* even though this more general version holds.

lemma *energy-graph-partitions-increase-half*:
  assumes *ref*: *refines V Q P* **and** *finite V* **and** *part-VP*: *partition-on V P*
    and *U*: $\{\} \notin U$
  shows *energy-graph-partitions G Q U* $\geq$ *energy-graph-partitions G P U*
      (**is** *?egQ* $\geq$ *?egP*)
$\langle$*proof*$\rangle$

proposition *energy-graph-partitions-increase*:
  assumes *refines V Q P refines V' Q' P'*
    and *finite V finite V'*
  shows *energy-graph-partitions G Q Q'* $\geq$ *energy-graph-partitions G P P'*
$\langle$*proof*$\rangle$

The original version of Gowers's Lemma 11 (also in Zhao) is not general enough to be used for anything.

corollary *mean-square-density-increase*:
  assumes *refines V Q P finite V*

shows *mean-square-density* G $Q \geq$ *mean-square-density* G P
⟨*proof*⟩

The Energy Boost Lemma says that an irregular partition increases the energy substantially. We assume that $\mathcal{U} \subseteq$ *uverts* G and $\mathcal{W} \subseteq$ *uverts* G are not irregular, as witnessed by their subsets $U1 \subseteq \mathcal{U}$ and $W1 \subseteq \mathcal{W}$. The proof follows Lemma 12 of Gowers.

definition *part2* X $Y \equiv$ *if* $X \subset Y$ *then* $\{X, Y-X\}$ *else* $\{Y\}$

lemma *card-part2*: *card* (*part2* X Y) ≤ 2
⟨*proof*⟩

lemma *sum-part2*: $[\![X \subseteq Y; f\{\} = 0]\!] \Longrightarrow$ *sum* f (*part2* X Y) $= f\,X + f\,(Y-X)$
⟨*proof*⟩

lemma *partition-part2*:
assumes $A \subseteq B$ $A \neq \{\}$
shows *partition-on* B (*part2* A B)
⟨*proof*⟩

proposition *energy-boost*:
fixes ε::*real* **and** U W G
defines *alpha* $\equiv$ *edge-density* U W G
defines $u \equiv \lambda X\ Y.$ *edge-density* X Y G $-$ *alpha*
assumes *finite* U *finite* W
and $U' \subseteq U$ $W' \subseteq W$ $\varepsilon > 0$
and U': *card* $U' \geq \varepsilon *$ *card* U **and** W': *card* $W' \geq \varepsilon *$ *card* W
and *gt*: $|u\ U'\ W'| > \varepsilon$
shows $(\sum A \in$ *part2* U' U. $\sum B \in$ *part2* W' W. *energy-graph-subsets* A B G)
$\geq$ *energy-graph-subsets* U W G $+ \varepsilon\hat{\ }4 *$ (*card* $U *$ *card* W) / (*card* (*uverts* G))^2
(**is** *?lhs* $\geq$ *?rhs*)
⟨*proof*⟩

1.6 Energy boost for partitions

We can always find a refinement that increases the energy by a certain amount.

A necessary lemma for the tower of exponentials in the result. Angeliki's proof

lemma *le-tower-2*: $k * (2\ \hat{}\ Suc\ k) \leq 2\hat{\ }(2\hat{\ }k)$
⟨*proof*⟩

The bound 2^{k+1} comes from a different source by Zhao: "Graph Theory and Additive Combinatorics", https://yufeizhao.com/gtacbook/. It's needed because our *regular-partition* includes the diagonal; otherwise, $k2^k$ would work. Gowers' version has a flatly incorrect bound.

proposition *exists-refinement*:
  assumes *fgp*: *finite-graph-partition* (*uverts G*) *P k* **and** *finite* (*uverts G*)
    and *irreg*: $\neg$ ε*−regular-partition G P* **and** $\varepsilon > 0$
  obtains *Q* **where** *refines* (*uverts G*) *Q P*
    mean-square-density G Q $\geq$ *mean-square-density G P* + εˆ*5*
    $\bigwedge R.\ R \in P \Longrightarrow$ *card* $\{S \in Q.\ S \subseteq R\} \leq 2$ ˆ *Suc k*
    card Q $\leq k * 2$ ˆ *Suc k*
⟨*proof*⟩

1.7 The Regularity Proof Itself

We start with a trivial partition (one part). If it is already ϵ-regular, we are done. If not, we refine it by applying lemma *exists-refinement* above, which increases the energy. We can repeat this step, but it cannot increase forever: by *mean-square-density-bounded* it cannot exceed 1. This defines an algorithm that must stop after at most ϵ^{-5} steps, resulting in an ϵ-regular partition.

theorem *Szemeredi-Regularity-Lemma*:
  assumes $\varepsilon > 0$
  obtains *M* **where** $\bigwedge G.$ *card* (*uverts G*) $> 0 \Longrightarrow \exists P.$ ε*−regular-partition G P* $\wedge$ *card P* $\leq M$
⟨*proof*⟩

The actual value of the bound is visible above: a tower of exponentials of height $2(1 + \epsilon^{-5})$.

end