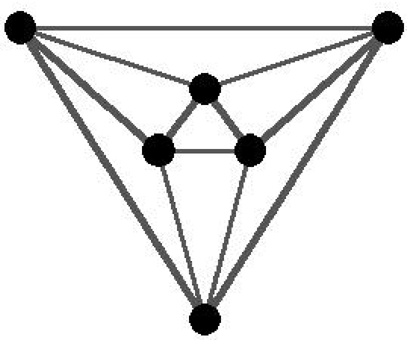
Abel Prize Laureate 2012
Endre Szemerédi

Szemerédi`s Regularity lemma

A main ingredient in Szemerédis theorem about arithmetic progressions in sets of positive density is the Regularity lemma. Szemerédi used a weak form of this lemma, for bipartite graphs, to prove the theorem. Later he also proved a strong version, for more general graphs.

Szemerédi´s Regularity Lemma is a result in graph theory. The lemma states that for every large enough graph, the set of nodes can be dvided into subsets of about the same size so that the edges between different subsets behave almost randomly. In 1975, Szemerédi introduced a weak version of this lemma, restricted to socalled bipartite graphs, in order to prove his famous theorem about arithmetic progressions. In 1978 he proved the full lemma. A graph consists of nodes and edges. The edges are connections between the nodes, and between two nodes there might or might not be an edge.

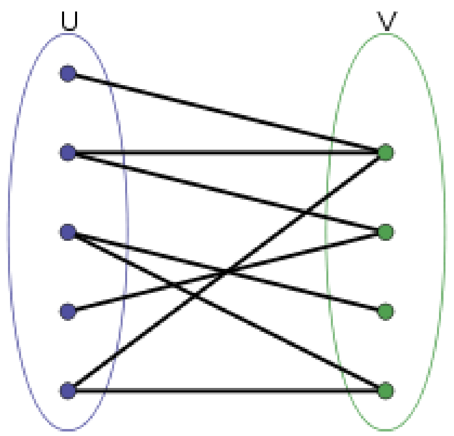


A graph with six nodes and twelve edges

A graph can be viewed as an abstract mathematical object, or as an illustration of some network. A road map is an example of a graph, where the crossings are the nodes and the roads in between are the edges. The friendship graph is another example, where the nodes can be thought of as a collection of people, and the edges as friendship relations. We usually draw a mathematical graph using dots for the nodes and straight lines for the edges. Graphs can have all kinds of complexity, the simplest one being graphs with only nodes and no edges. A graph with edges connecting any pair of nodes is called a complete graph. In general, a randomly chosen graph is something in between these two extremes.

Now, for any graph, consider two disjoint subsets of nodes, denoted by X and Y. If all the nodes in X are connected to all the nodes in Y, the number of edges between the two sets, $e(X,Y)$ is the product of the number of nodes in the two sets. In general, the number of edges between the two sets will be less than this product. The fraction of the number of edges between nodes in X and Y, and the highest possible number of edges, given by the product of the cardinalities of the two sets, is called the density of the pair (X,Y), and is denoted $d(X,Y)$. The density is 1 if all possible edges between the two sets are present in the graph, and 0 if there are no edges. You may think of the density as the average probability of the existence of an edge between two randomly picked nodes of X and Y.

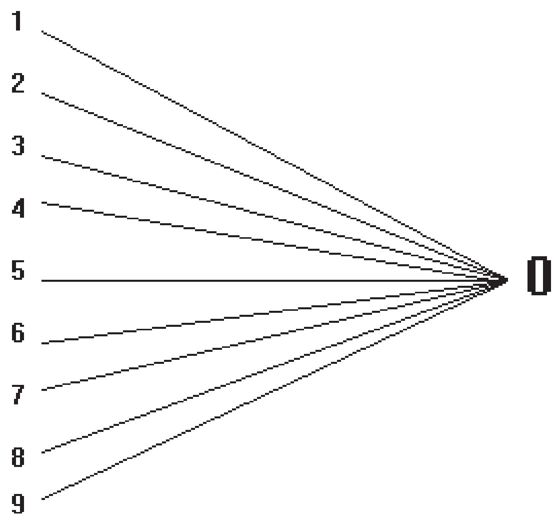
Let (X,Y) be a pair of nodes of density $d(X,Y)$. For different choices of subsets U and V of X and Y, repectively, we can compute the density $d(U,V)$ in the same way as above. The regularity of the the pair (X,Y) measures how the density varies when we examine all pairs of subsets of *X* and *Y* of a certain size. In the complete case, where all nodes in *X* are connected to all nodes in *Y*, we have full regularity. The other extreme, a graph with no edges at all between *X* and *Y*, also gives full regularity. And, mayby a bit surprising, a randomlike edge set for a pair *(X,Y)* gives high regularity. The reason is that for a randomlike edge set, the average probaility of finding an edge between two nodes in *X* and *Y*, is more or less independent of which subsets of *X* and *Y* we consider.



The density of the pair (U,V) is d(U,V)=8/20=0,4

For an arbitrary graph we normally know very little about the edge set. But Szemerédis Regularity lemma makes us look at the edge set in a more manageable way. The strategy is to split the set of nodes of the graph into more or less equally sized subsets, and compute the regularity of each pair of subsets. If we can do this in such a way that each pair has a high degree of regularity, we get a helpful tool to study the graph, even if the edge set of the graph is an inaccessible and apparently patternless collection of lines. This is exactly what Szemerédi`s Regularity lemma gives us. The lemma says that no matter which regularity we require, we can always find a partition of the set of nodes into subsets such that the edge set for any two of the subsets is sufficienly randomlike, i.e. have the required degree of regularity.

We can consider a special case of the Regularity lemma, where the graph has the following form. Let *1,2,3,...,N* and *0* be the nodes and let *A* be some subset of *{1,2,3,...,N}*. Suppose there are no edges between the nodes *1,2,3,...,N*, and that we have an edge between a node *m* and *0* if and only if *m* is in *A*. A partition of the set of nodes will consist of one subset which contains *0*, and the rest will be subsets of *{1,2,3,..., N}* which we assume are intervals. For simplicity we throw out all nodes except *0* from the subset containing *0*. Thus the subsets are either (more or less) equally sized subintervals of *{1,2,3,...,N}*, or *{0}*. For two subintervalls of *{1,2,3,..., N}* the density is *0*, since there are no internal edges in *{1,2,3,..., N}*, and the density of an interval and *{0}* is the fraction $\#(A\cap\{1,2,3,...,N\})/\#A$. The Regularity lemma in this case says that for an arbitrary subset *A* of *{1,2,3,..., N}* we can split *{1,2,3,..., N}* in (almost) equally long subintervals such that for each subinterval *I* the subset $A\cap I$ is randomlike, i.e. for any subinterval *J* of *I* the density of *A* in *J* is almost the same as the density of *A* in *I*.



E. Szemerédi; *On sets of integers containing no k elements in arithmetic progression*.
Acta Arithmetica, 27: pp.199–245, 1975.

E. Szemerédi; *Regular partitions of graphs*.
In Problèmes Combinatoires et Théorie des Graphes (Col loq. Internat. CNRS, Univ. Orsay, Orsay, 1976), pp. 399–401, Paris, 1978. Colloques Internationaux CNRS n. 260.