Szemerédi's Regularity Lemma and Its Applications to Pairwise Clustering and Segmentation

Anna Sperotto[1] and Marcello Pelillo[2]

[1] Department of Electrical Engineering, Mathematics, and Computer Science
University of Twente, The Netherlands
a.sperotto@utwente.nl
[2] Dipartimento di Informatica
Università Ca' Foscari di Venezia, Italy
pelillo@dsi.unive.it

Abstract. Szemerédi's regularity lemma is a deep result from extremal graph theory which states that every graph can be well-approximated by the union of a constant number of random-like bipartite graphs, called regular pairs. Although the original proof was non-constructive, efficient (i.e., polynomial-time) algorithms have been developed to determine regular partitions for arbitrary graphs. This paper reports a first attempt at applying Szemerédi's result to computer vision and pattern recognition problems. Motivated by a powerful auxiliary result which, given a partitioned graph, allows one to construct a small reduced graph which inherits many properties of the original one, we develop a two-step pairwise clustering strategy in an attempt to reduce computational costs while preserving satisfactory classification accuracy. Specifically, Szemerédi's partitioning process is used as a preclustering step to substantially reduce the size of the input graph in a way which takes advantage of the strong notion of edge-density regularity. Clustering is then performed on the reduced graph using standard algorithms and the solutions obtained are then mapped back into the original graph to create the final groups. Experimental results conducted on standard benchmark datasets from the UCI machine learning repository as well as on image segmentation tasks confirm the effectiveness of the proposed approach.

1 Introduction

Graph-theoretic representations and algorithms have long been an important tool in computer vision and pattern recognition, especially because of their representational power and flexibility. However, there is now a renewed and growing interest toward explicitly formulating computer vision problems within a graph-theoretic setting. This is in fact particularly advantageous because it allows vision problems to be cast in a pure, abstract setting with solid theoretical underpinnings and also permits access to the full arsenal of graph algorithms developed in computer science and operations research. Graph-theoretic problems which have proven to be relevant to computer vision include maximum

A.L. Yuille et al. (Eds.): EMMCVPR 2007, LNCS 4679, pp. 13–27, 2007.

flow, minimum spanning tree, maximum clique, shortest path, maximal common subtree/subgraph, etc. In addition, a number of fundamental techniques that were designed in the graph algorithms community have recently been applied to computer vision problems. Examples include spectral methods and fractional rounding.

In 1941, the Hungarian mathematician P. Turán provided an answer to the following innocent-looking question. What is the maximal number of edges in a graph with n vertices not containing a complete subgraph of order k, for a given k? This graph is now known as a Turán graph and contains no more than $n^2(k-2)/2(k-1)$ edges. Later, in another classical paper, T. S. Motzkin and E. G. Straus [12] provided a novel proof of Turán's theorem using a continuous characterization of the clique number of a graph. Thanks to the contributions of P. Erdös, B. Bollobás, M. Simonovits, E. Szemerédi, and others, Turán study developed soon into one of the richest branches of 20th-century graph theory, known as *extremal graph theory*, which has intriguing connections with Ramsey theory, random graph theory, algebraic constructions, etc. Very roughly, extremal graph theory studies how the intrinsic structure of graphs ensures certain types of properties (e.g., cliques, colorings and spanning subgraphs) under appropriate conditions (e.g., edge density and minimum degree) (see, e.g., [3]).

Among the many achievements of extremal graph theory, Szemerédi's *regularity lemma* is certainly one of the best known [6]. It states, essentially, that every graph can be partitioned into a small number of random-like bipartite graphs, called regular pairs, and a few leftover edges. Szemerédi's result was introduced in the mid-seventies as a tool for his celebrated proof of the Erdös-Turán conjecture on arithmetic progressions in dense sets of integers. Since then, the lemma has emerged as a fundamental tool not only in extremal graph theory but also in theoretical computer science, combinatorial number theory, etc. We refer to [11,10] for a survey of the (theoretical) applications of Szemerédi's lemma and its generalizations.

The regularity lemma is basically an existence predicate. In its original proof, Szemerédi demonstrated the existence of a regular partition under the most general conditions, but he did not provide a constructive proof to obtain such a partition. However, in 1992, Alon et al. [1] succeeded in developing the first algorithm to create a regular partition on arbitrary graphs, and showed that it has polynomial computational complexity. Other polynomial-time algorithms can be found in [8,5,9].

This paper reports perhaps the first practical application of the regularity lemma and related algorithms. Our original motivation was to study how to take advantage of the information provided by Szemerédi's regular partitions in a pairwise clustering context. Indeed, pairwise (or similarity-based) data clustering techniques are gaining increasing popularity over traditional feature-based grouping algorithms. In many application domains, in fact, the objects to be clustered are not naturally representable in terms of a vector of features. On the other hand, quite often it is possible to obtain a measure of the similarity/dissimilarity between objects. Hence, it is natural to map (possibly implicitly) the data to

be clustered to the nodes of a weighted graph, with edge-weights representing similarity or dissimilarity relations. However, a typical problem associated to pairwise grouping algorithms is the scaling behavior with the number of data. On a dataset containing N examples, the number of potential comparisons scales with $O(N^2)$, thereby hindering their applicability to problems involving very large data sets, such as high-resolution imagery and spatio-temporal data. Recent sophisticated attempts to deal with this problem use optimal embeddings [15], the Nystrom method [2,7], and out-of-sample dominant sets [14].

The solution outlined in this paper combines the notion of regularity introduced by Szemerédi with that of graph-based clustering. In this context, the regularity lemma is used as a preclustering strategy, in an attempt to work on a more compact, yet informative, structure. Indeed, this structure is well-known in extremal graph theory and is commonly referred to as the *reduced graph*. An important auxiliary result, the so-called Key Lemma, reveals that this graph does inherit many of the essential structural properties of the original graph. In summary, our approach consists basically in a two-phase procedure. In the first phase, the input graph is decomposed into small pieces using Szemerédi's partitioning process and the corresponding (weighted) reduced graph is constructed, the weights of which reflect edge-densities between class pairs of the original partition. Next, a standard graph-based clustering procedure is run on the reduced graph and the solution found is mapped back into original graph to obtain the final groups. In our simulations we used dominant-set algorithms to perform the clustering on the reduced graph, but any other pairwise algorithm such as Normalized Cut would work equally well. Note that our approach differs from other attempts aimed at reducing the complexity of pairwise grouping processes, such as [2,7,14], as we perform no sampling of the original data but work instead on a derived structure which does retain the important features of the original one.

We performed some experiments both on standard datasets from the UCI machine learning repository and on image segmentation tasks, and the preliminary results obtained confirm the effectiveness of the proposed approach.

2 Szemerédi's Regularity Lemma

Let $G = (V, E)$ be an undirected graph with no self-loops, where V is the set of vertices and E is the set of edges, and let $X, Y \subseteq V$ be two disjoint subsets of vertices of G. We define the *edge density* of the pair (X, Y) as:

$$d(X, Y) = \frac{e(X, Y)}{|X||Y|} \tag{1}$$

where $e(X, Y)$ denotes the number of edges of G with an endpoint in X and an endpoint in Y, and $|\cdot|$ denotes the cardinality of a set. Note that edge densities are real numbers between 0 and 1.

Given a positive constant $\varepsilon > 0$, we say that the pair (A, B) of disjoint vertex sets $A, B \subseteq V$ is *ε-regular* if for every $X \subseteq A$ and $Y \subseteq B$ satisfying

$$|X| > \varepsilon|A| \quad \text{and} \quad |Y| > \varepsilon|B| \tag{2}$$

we have

$$|d(X,Y) - d(A,B)| < \varepsilon. \tag{3}$$

Thus, in an ε-regular pair the edges are distributed fairly uniformly.

A partition of V into pairwise disjoint classes $C_0, C_1, \ldots, C_k$ is said *equitable* if all the classes C_i $(1 \leq i \leq k)$ have the same cardinality. The *exceptional set* C_0 (which may be empty) has only a technical purpose: it makes it possible that all other classes have exactly the same number of vertices. An equitable partition $C_0, C_1, \ldots, C_k$, with C_0 being the exceptional set, is called *ε-regular* if $|C_0| < \varepsilon|V|$ and all but at most εk^2 of the pairs (C_i, C_j) are ε-regular $(1 \leq i < j \leq k)$.

Theorem 1 (Szemerédi's regularity lemma [16]). *For every positive real ε and for every positive integer m, there are positive integers $N = N(\varepsilon, m)$ and $M = M(\varepsilon, m)$ with the following property: for every graph G with $|V| \geq N$ there is an ε-regular partition of G into $k+1$ classes such that $m \leq k \leq M$.*

Given an $r \times r$ symmetric matrix (p_{ij}) with $0 \leq p_{ij} \leq 1$, and positive integers $n_1, n_2, \ldots, n_r$, a *generalized random graph* R_n for $n = n_1 + n_2 + \ldots + n_r$ is obtained by partitioning n vertices into classes C_i of size n_i and joining the vertices $x \in V_i$, $y \in V_j$ with probability p_{ij}, independently for all pairs $\{x, y\}$. Now, as pointed out by Komlós and Simonovits [11], the regularity lemma asserts basically that every graph can be approximated by generalized random graphs. Note that, for the lemma to be useful, the graph has to to be dense. Indeed, for sparse graphs it becomes trivial as all densities of pairs tend to zero [6].

The lemma allows us to specify a lower bound m on the number of classes. A large value of m ensures that the partition classes C_i are sufficiently small, thereby increasing the proportion of (inter-class) edges subject to the regularity condition and reducing the intra-class ones. The upper bound M on the number of partitions guarantees that for large graphs the partition sets are large too. Finally, it should be noted that a singleton partition is ε-regular for every value of ε and m.

The regularity lemma permits εk^2 pairs to be irregular. Following a path pointed out by Szemerédi [16], many researchers studied if the result could be strengthened, avoiding the presence of such pairs. However, it turned out that forcing the lemma in that way would affect its generality [1].

The problem of checking if a given partition is ε-regular is a quite surprising one from a computational complexity point of view. In fact, as proven in [1], it turns out that constructing an ε-regular partition is easier than checking if a given one responds to the ε-regularity criterion.

Theorem 2. *The following decision problem is co-NP-complete. Given a graph G, an integer $k \geq 1$ and a parameter $\varepsilon > 0$, and a partition of the set of vertex of G into $k+1$ parts. Decide if the given partition is ε-regular.*

Recall that the complexity class *co-NP-complete* collects all problems which complement are NP-complete [4]. In other words, a *co-NP-complete* problem is one for which there are efficiently verifiable proofs of its no-instances, i.e., its counterexamples.

3 Finding Regular Partitions in Polynomial Time

The original proof of Szemerédi's lemma [16] is not constructive, yet this has not narrowed the range of its applications in the fields of extremal graph theory, number theory and combinatorics. In the early 1990's, Alon et al. [1] proposed a new formulation of the lemma which emphasizes the algorithmic nature of the result.

Theorem 3 (A constructive version of the regularity lemma [1]). *For every $\varepsilon > 0$ and every positive integer t there is an integer $Q = Q(\varepsilon, t)$ such that every graph with $n > Q$ vertices has an ε-regular partition into $k + 1$ classes, where $t \leq k \leq Q$. For every fixed $\varepsilon > 0$ and $t \geq 1$ such a partition con be found in $O(M(n))$ sequential time, where $M(n) = O(n^{2.376})$ is the time for multiplying two $n \times n$ matrices with $0, 1$ entries over the integers. The partition can be found in time $O(\log n)$ on a EREW PRAM with a polynomial number of parallel processor.*

In the remaining of this section we shall derive the algorithm alluded to in the previous theorem. We refer the reader to the original paper [1] for more technical details and proofs.

Let H be a bipartite graph with equal color classes $|A| = |B| = n$. We define the *average degree* of H as

$$d = \frac{1}{2n} \sum_{i \in A \cup B} deg(i) \tag{4}$$

where, $deg(i)$ is the degree of the vertex i.

Let y_1 and y_2 be two distinct vertices, such that $y_1, y_2 \in B$. Alon et al. [1] define the *neighborhood deviation* of y_1 and y_2 by

$$\sigma(y_1, y_2) = |N(y_1) \cap N(y_2)| - \frac{d^2}{n} \tag{5}$$

where $N(x)$ is the set of neighbors of the vertex x. For a subset $Y \subseteq B$, the *deviation* of Y is defined as:

$$\sigma(Y) = \frac{\sum_{y_1, y_2 \in Y} \sigma(y_1, y_2)}{|Y|^2} \tag{6}$$

Now, let $0 < \varepsilon < \frac{1}{16}$. It can be shown that if there exists $Y \subseteq B$, $|Y| > \varepsilon n$ such that $\sigma(Y) \geq \frac{\varepsilon^3}{2} n$ then at least one of the following cases occurs [1]:

1. $d < \varepsilon^3 n$ (which amounts to saying that H is ε-regular);
2. there exists in B a set of more than $\frac{1}{8}\varepsilon^4 n$ vertices whose degrees deviate from d by at least $\varepsilon^4 n$;
3. there are subsets $A' \subseteq A$, $B' \subseteq B$, $|A'| \geq \frac{\varepsilon^4}{4} n$, $|B'| \geq \frac{\varepsilon^4}{4} n$ and $|d(A', B') - d(A, B)| \geq \varepsilon^4$.

Note that one can easily check if 1 holds in time $O(n^2)$. Similarly, it is trivial to check if 2 holds in $O(n^2)$ time, and in case it holds to exhibit the required subset of B establishing this fact. If both cases above fail we can proceed as follows. For each $y_0 \in B$ with $|deg(y_0) - d| < \varepsilon^4 n$ we find the set of vertices $B_{y_0} = \{y \in B : \sigma(y_0, y) \geq 2\varepsilon^4 n\}$. It can be shown that there exists at least one such y_0 for which $|B_{y_0}| \geq \frac{\varepsilon^4}{4} n$. The subsets $B' = B_{y_0}$ and $A' = N(y_0)$ are the required ones. Since the computation of the quantities $\sigma(y, y')$, for $y, y' \in B$, can be done by squaring the adjacency matrix of H, the overall complexity of this algorithms is $O(M(n)) = O(n^{2.376})$.

In order to understand the final partitioning algorithm we need the following two lemmas.

Lemma 1 (Alon et al. [1]). *Let H be a bipartite graph with equal classes $|A| = |B| = n$. Let $2n^{-1/4} < \varepsilon < \frac{1}{16}$. There is an $O(n^{2.376})$ algorithm that verifies that H is ε-regular or finds two subsets $A' \subseteq A$, $B' \subseteq B$, $|A'| \geq \frac{\varepsilon^4}{4} n$, $|B'| \geq \frac{\varepsilon^4}{4} n$ and $|d(A', B') - d(A, B)| \geq \varepsilon^4$.*

It is quite easy to check that the regularity condition can be rephrased in terms of the average degree of H. Indeed, it can be seen that if $d < \varepsilon^3 n$, then H is ε-regular, and this can be tested in $O(n^2)$ time. Next, it is necessary to count the number of vertices in B whose degrees deviate from d by at least $\varepsilon^4 n$. Again, this operation takes $O(n^2)$ time. If the number of deviating vertices is more than $\frac{\varepsilon^4}{8} n$, then the degrees of at least half of them deviate in the same direction and if we let B' be such a set of vertices and $A' = A$ we are done. Otherwise, it can be shown that there must exist $Y \subseteq B$ such that $|Y| \geq \varepsilon n$ and $\sigma(Y) \geq \frac{\varepsilon^3}{2} n$. Hence, our previous discussion shows that the required subsets A' and B' can be found in $O(n^{2.376})$ time.

Given an equitable partition P of a graph $G = (V, E)$ into classes $C_0, C_1 \ldots C_k$, Szemerédi [16] defines a measure called *index of partition*:

$$ind(P) = \frac{1}{k^2} \sum_{s=1}^{k} \sum_{t=s+1}^{k} d(C_s, C_t)^2. \tag{7}$$

Since $0 \leq d(C_s, C_t) \leq 1$, $1 \leq s, t, \leq k$, it can be seen that $ind(P) \leq \frac{1}{2}$.

The following lemma is the core of Szemerédi's original proof.

Lemma 2 (Szemerédi [16]). *Let $G = (V, E)$ be a graph with n vertices. Let P be an equitable partition of V into classes $C_0, C_1 \ldots C_k$, with C_0 being the exceptional class. Let $\gamma > 0$. Let k be the least positive integer such that $4^k > 600\gamma^{-5}$. If more than γk^2 pairs (C_s, C_t) $(1 \leq s < t \leq k)$ are γ-irregular, then there is an equitable partition Q of V into $1 + k4^k$ classes, the cardinality of the exceptional class being at most*

$$|C_0| + \frac{n}{4^k} \tag{8}$$

and such that

$$ind(Q) > ind(P) + \frac{\gamma^5}{20}. \tag{9}$$

The idea formalized in the previous lemma is that, if a partition violates the regularity condition, then it can be refined by a new partition and, in this case, the $ind(P)$ measure can be improved. On the other hand, the new partition adds only "few" elements to the current exceptional set so that, in the end, its cardinality will respect the definition of equitable partition.

We are now in a position to sketch the complete partitioning algorithm. The procedure is divided into two main steps: in the first step all the constants needed during the next computation are set; in the second one, the partition is iteratively created. An iteration is called *refinement step*, because, at each iteration, the current partition is closer to a regular one.

Given any $\varepsilon > 0$ and a positive integer t, the constants $N = N(\varepsilon, t)$ and $T = T(\varepsilon, t)$ are defined as follows. Let b be the least positive integer such that

$$4^b > 600\left(\frac{\varepsilon^4}{16}\right)^{-5}, \quad b \geq t. \tag{10}$$

Let f be the integer-valued function defined as:

$$f(0) = b, \quad f(i+1) = f(i)4^{f(i)}. \tag{11}$$

Put $T = f\left(\lceil 10(\frac{\varepsilon^4}{16})^{-5}\rceil\right)$ and $N = \max\{T4^{2T}, \frac{32T}{\varepsilon^5}\}$. Given a graph $G = (V, E)$ with $n \geq N$ vertices, an ε-regular partition of G into $k+1$ classes, where $t \leq k \leq T$, can be constructed using the following $O(M(n)) = O(n^{2.376})$ algorithm.[1]

1. **Create the initial partition:** Arbitrarily divide the set V into an equitable partition P_1 with classes $C_0, C_1, \ldots, C_b$ where $|C_i| = \lfloor n/b \rfloor$, $i = 1 \ldots b$ and $|C_0| < b$. Let $k_1 = b$.
2. **Check regularity:** For every pair C_r, C_s of P_i verify if it is ε-regular or find $X \subseteq C_r$, $Y \subseteq C_s$, $|X| \geq \frac{\varepsilon^4}{16}|C_r|$, $|Y| \geq \frac{\varepsilon^4}{16}|C_s|$ such that

$$|d(X, Y) - d(C_s, C_r)| \geq \varepsilon^4. \tag{12}$$

3. **Count regular pairs:** If there are at most $\varepsilon\binom{k_i}{2}$ pairs that are not verified as ε-regular, then halt. P_i is an ε-regular partition.
4. **Refine:** Apply Lemma 2 where $P = P_i$, $k = k_i$, $\gamma = \frac{\varepsilon^4}{16}$ and obtain a partition P' with $1 + k_i 4^{k_i}$ classes.
5. Let $k_{i+1} = k_i 4^{k_i}$, $P_{i+1} = P'$, $i = i + 1$ and go to step 2.

Before concluding this section we mention that after Alon et al.'s contribution, other algorithms have been proposed for finding Szemerédi's partitions. In particular, we mention Frieze and Kannan's approach [8], which is based on an intriguing relation between the regularity conditions and the singular values of matrices, and Czygrinow and Rödl's [5], who proposed a new algorithmic version of Szemerédi's lemma for hypergraphs.

[1] Note that Alon et al. [1] proved Theorem 3 with $Q = N$ ($\geq T$).

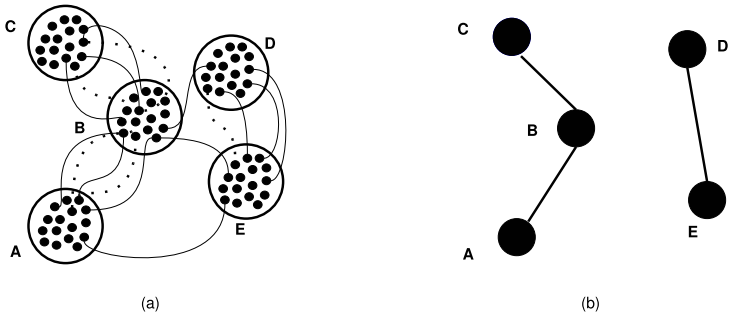


Fig. 1. Example of reduced graph creation. Left: original regular partition. Here, only the pairs (A, B), (B, C) and (D, E) are regular. Right: the graph obtained after the reduction process.

4 The Regularity Lemma and Pairwise Clustering

The regularity lemma postulates the existence of a partition of any graph into disjoint subsets satisfying a uniformity criterion, and the algorithm described in the previous section is able to find such a partition in polynomial time. On the other hand, pairwise (or graph-based) clustering refers to the process of partitioning an edge-weighted similarity graph in such a way as to produce maximally homogeneous classes. It is far too easy to see that, although both processes aim at producing a partition of a given graph, the way in which they attempt to do it is substantially different. Indeed, Szemerédi's partitioning process is very strict as it imposes the creation of same-size classes (this condition being relaxed only for the exceptional set), and this would clearly be too restrictive within a clustering context. Also, the notion of a regular partition, while emphasizing the role of inter-class relations, pays no explicit attention to the relations among vertices in the same class. Indeed, the very nature of the regularity concept indicates that the less intra-class edges, the "better" the partition.

A regular partition reveals the existence of a hidden structure in a graph. Hence, despite their dissimilarities, it could be interesting to try to combine the two partitioning approaches in order to obtain a novel and efficient clustering strategy. In fact, extremal graph theory provides us with some interesting results and abstract structures that can be conveniently employed for our purpose: these are the notion of the *reduced graph* and the so-called *Key Lemma* [11].

Given a graph $G = (V, E)$, a partition P of the vertex-set V into the sets $C_1, C_2, \ldots C_k$ and two parameters ε and d, the reduced graph R is defined as follows. The vertices of R are the clusters $C_1, C_2, \ldots C_k$, and C_i is adjacent to C_j if (C_i, C_j) is ε-regular with density more than d. Figure 1 shows an example of transformation from a partitioned graph to its reduced graph.

Consider now a graph R and an integer t. Let $R(t)$ be the graph obtained form R by replacing each vertex $x \in V(R)$ by a set V_x of t independent vertices, and joining $u \in V_x$ to $v \in V_y$ if and only if (x, y) is an edge in R. $R(t)$ is a graph

in which every edge of R is replaced by a copy of the complete bipartite graph K_{tt}.

The following Lemma shows how to use the reduced graph R and its modification $R(t)$ to infer properties of a more complex graph.

Theorem 4 (Key Lemma [11]). *Given $d > \varepsilon > 0$, a graph R and a positive integer m, construct a graph G following these steps:*

1. *replace every vertex of R by m vertices*
2. *replace the edges of R with regular pairs of density at least d.*

Let H be a subgraph of $R(t)$ with h vertices and maximum degree $\Delta > 0$, and let $\delta = d - \varepsilon$ and $\varepsilon_0 = \delta^{\Delta}/(2 + \Delta)$. If $\varepsilon \leq \varepsilon_0$ and $t - 1 \leq \varepsilon_0 m$, then H is embeddable into G (i.e., G contains a subgraph isomorphic to H). In fact, we have

$$||H \rightarrow G|| > (\varepsilon_0 m)^h \tag{13}$$

where $||H \rightarrow G||$ denotes the number of labeled copies of H in G.

Given a graph R, the Key Lemma furnishes rules to expand R to a more complex partitioned graph G which respects edge-density bounds. On the other hand, we have another expanded graph, $R(t)$. Because of their construction, $R(t)$ and G are very similar, but they can have a different vertex cardinality. In addition, note that the only densities allowed between vertex subsets in $R(t)$ are 0 and 1. The Key Lemma establishes constraints to the edge density d and the subset size t in order to assure the existence of a fixed graph H embeddable into $R(t)$, which is also a subgraph of G. Let H be a subgraph of $R(t)$. If t is sufficiently small with respect to m, and d is sufficiently high with respect to ε and Δ, it is possible to find small subsets of vertices such that they are connected with a sufficiently high number of edges. The copies of H are constructed vertex by vertex by picking up elements from the previous identified subsets.

As described in [11], a common and helpful combined use of the reduced graph and Key Lemma is as follows (see Figure 2):

- Start with a graph $G = (V, E)$ and apply the regularity lemma, finding a regular partition P
- Construct the reduced graph R of G, w.r.t. the partition P.
- Analyze the properties of R, in particular its subgraphs.
- As it is assured by Theorem 4, every small subgraph of R is also a subgraph of G.

In summary, a direct consequence of the Key Lemma is that it is possible to search for significant substructures in a reduced graph R in order to find common subgraphs of R and the original graph.

Now, let us put ourselves in a pairwise clustering context. We are given a large weighted (similarity) graph $G = (V, E, \omega)$ where the vertices correspond to data points, edges represent neighborhood relationships, and edge-weights reflect similarity between pairs of linked vertices. Motivated by the previous discussion,

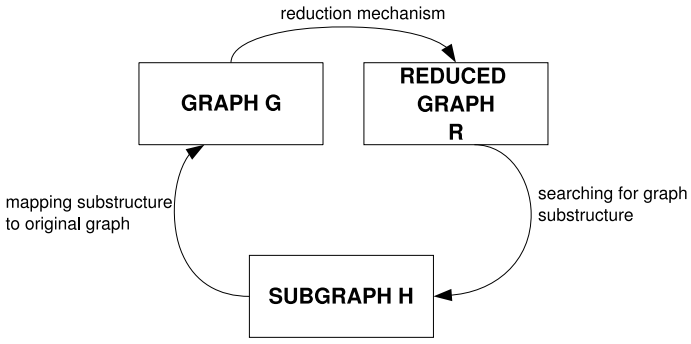


Fig. 2. Reduction strategy to find significant substructures in a graph

here we propose a *two-phase clustering strategy*. In the first phase, we apply the regularity lemma in order to build a compact, yet informative representation of the original data set, in the form of a reduced graph. In the second phase, a pairwise clustering method is used to find meaningful structures in the reduced graph. In the experiments described in the next section, for example, we used the dominant-set approach described in [13], but other pairwise algorithms could be used equally well.

Note that an algorithmic formulation of the regularity lemma for weighted graphs was explicitly introduced in [5], together with an extension of the result concerning hypergraphs. Nevertheless, simple verifications on the previous unweighted definitions show that the algorithms are not influenced by edge-weights. In this case, the *density* between the sets of a pair (A, B) in a becomes:

$$d = d_\omega(A, B) = \frac{\sum_{i=1}^{|A|} \sum_{j=1}^{|B|} \omega(a_i, b_j)}{|A||B|} \tag{14}$$

which is the weighted counterpart of (1).

Since the notion of regularity emphasizes only inter-class relations, there must be a criterion for taking into account intra-class similarities as well. Our approach consists of introducing in the partioning algorithm a rule to select elements to be dispatched to new subsets. Specifically, the criterion used is the average weighted degree

$$\mathrm{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} \omega(i, j) \quad S \subseteq V. \tag{15}$$

All elements in the current subset S are listed in decreasing order by average weighted degree. In so doing, the partition of S takes place simply by subdividing the ordered sequence of elements into the desired number of subsets. The decreasing order has been preferred because of the presence of the exceptional set: in this way we assume that only the less connected vertices join the exceptional set. Hence, the obtained regular partition contains classes the elements of which can already be considered similar to each other.

The vertex set cardinality in the reduced graph is a fundamental quantity during the clustering process because it affects the precision of the solution found. In particular, a large cardinality implies a greater precision, as the clusters have to be found in a more articulated context. A low cardinality, instead, makes the clustering phase less accurate because the inclusion or the exclusion of a vertex in a cluster may change dramatically the solution at the finer level of the original graph. Our experience shows that working with medium-size subsets produces the best results in terms of accuracy and speed.

5 Experimental Results

We performed some preliminary experiments aimed at assessing the potential of the approach described in the previous sections. Before presenting the experimental setting and the results obtained, a few technical observations concerning our implementation of Alon et al's algorithm are in order. First, note that the original algorithm stops when a regular partition has been found. In practice, the number of iterations and the vertex set cardinality required is simply too big to be considered. Hence, in our experiments we decide to stop the process either when a regular partition has been found or when the subset size becomes smaller than a predetermined threshold. Further, the next-iteration number of subsets of the original procedure is also intractable, and we therefore decided to split every subset, from an iteration to the next one, using a (typically small) user-defined parameter. Finally, note that in the original formulation of the regularity lemma, the exceptional set is only a technicality to ensure that all other classes have the same cardinality. In a clustering context, the exceptional set is a nuisance as it cannot be reasonably considered as a coherent class. It is therefore necessary to somehow assign its elements to the groups found by the clustering procedure. A naive solution, adopted in our experiments, is to assign them to the closest cluster according to a predefined distance measure.

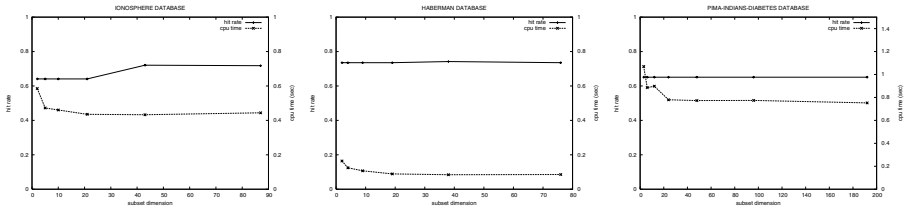
We conducted a first series of experiments on standard datasets from the UCI machine learning repository.[2] Specifically, we selected the following datasets: the Johns Hopkins University Ionosphere database (352 elements, 34 attributes, two classes), the Haberman's Survival database (306 elements, 3 attributes, two classes), and the Pima Indians Diabetes database (768 elements, 8 attributes, two classes). The similarity between data items was computed as a decreasing function of the Euclidean distance between corresponding attribute vectors, i.e., $w(i,j) = \exp(-||\mathbf{v}_i - \mathbf{v}_j||^2/\sigma^2)$, where $\mathbf{v}_i$ is the i-th vector of the dataset and σ is a positive real number which affects the decreasing rate of w.

Table 1 summarizes the result obtained on these data by showing the classification accuracy obtained by our two-phase strategy for each database considered. Recall that in the second phase we used the dominant-set algorithm [13] for clustering the reduced graph. Further, for the sake of comparison, we present the results produced by a direct application of the dominant-set algorithm to the original similarity graph without any pre-clustering step. As can be seen,

[2] `http://www.ics.uci.edu/~mlearn/MLRepository.html`

Table 1. Results obtained on the UCI benchmark datasets. Each row represents a dataset, while the columns represent (left to right): the number of elements in the corresponding dataset, the classification accuracy obtained by the proposed two-phase strategy and that obtained using the plain dominant-set algorithm, respectively, the speedup achieved using our approach w.r.t. plain dominant-set, the size of the reduced graphs, and the compression rate.

Dataset	**Size**	**Classif. Accuracy**		**Speedup**	**R.G. size**	**Compression rate**
		Two-phase	**Plain DS**			
Ionosphere	351	72%	67%	1.13	4	98.9%
Haberman	306	74%	73%	2.16	128	58.1%
Pima	768	65%	65%	2.45	256	66.7%



(a) (b) (c)

Fig. 3. Behavior of classification accuracy (left y-axis) and CPU time (right y-axis) as a function of the size of regularity classes for the three UCI datasets: Ionosphere (a), Haberman (b), and Pima (c)

our combined approach substantially outperforms the plain algorithm. Table 1 shows also the speedup achieved using our approach with respect to a direct application of the dominant-set algorithm. Note that, while the construction of a regular partition depends only on the number of vertices in the original graph and the subset dimension, the dominant set detection is affected by the edge weights. So it can converge faster or slower depending on the dataset we are analyzing. As can be seen, with our pre-clustering strategy we are able to achieve a speedup up to 2.45 on these data.

As observed at the end of the previous section, the cardinality of the reduced graph (or, alternatively, the size of the regularity subsets which in our implementation is a user-defined parameter) is of crucial importance as introduces a trade-off between the second-phase precision and the overall performance. The last two columns of Table 1 show the size of the reduced graphs and the corresponding compression rates used in the experiments. These values were chosen manually so as to optimize both accuracy and speed. Note how, using the regularity partitioning process, we were able to achieve compression rates from 66.7% to 98.9% while improving classification accuracy. It is of interest to analyze the behavior of our approach as this parameter is varied. Figure 3 plots the classification accuracy and the CPU time as a function of the size of the regularity classes (which is inversely related to the cardinality of the reduced graph) for

Table 2. The table summarizes information concerning the image segmentation experiments. Each row represents an image, while the columns represent (left to right): the number of pixels in the original image, the number of vertices in the reduced graph (with the subset dimension in parenthesis), the compression rate, and the speedup achieved using our approach w.r.t. plain dominant-set.

Image	**Pixels**	**R.G. size**	**Compression rate**	**Speedup**
Airplane	9600	128(75)	98.7%	4.23
Elephants	9600	32(300)	99.7%	17.79
Lake	9600	128(75)	98.7%	20.94
Tree	9600	64(150)	99.3%	16.63

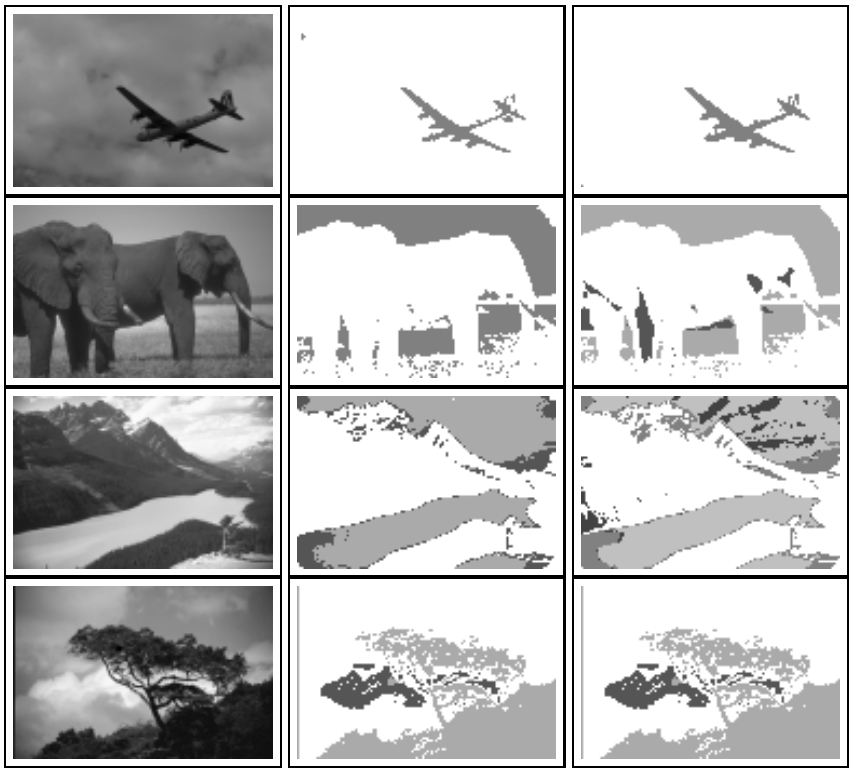


Fig. 4. Segmentation results for 120×80 grayscale images. Left column: original images. Central column: results with our two-phase approach. Right column: results with plain dominant sets.

all datasets considered. The CPU time curve highlights how a small subset dimension requires a higher computational time compared to larger sizes. At the same time, note how the classification accuracy is not substantially affected by

the class size: this means that the preclustering method is resistant to parameter variation and we are therefore allowed to use bigger subset dimensions in order to achieve better performance.

Our two-phase strategy was also applied to the problem of segmenting brightness images. Here, the image is represented as an edge-weighted graph where vertices represent pixels and edge-weights reflect the similarity between pixels. The measure adopted here to assign edge-weights is based on brightness proximity, as in [13]. Specifically, similarity between pixels i and j was measured by $w(i,j) = \exp(-(I(i) - I(j))^2/\sigma^2)$, where $I(i)$ is the normalized intensity value at node i.

Table 2 summarizes the main technical details of our experiments, in particular the compression rate and the speedup, while Figure 4 shows the segmentation results. Overall, both algorithms produced comparable segmentation results, despite the fact the with our approach we are looking for dominant sets in graphs which are at least 98% smaller than the original graph images and with a speedup of up to 20.94.

6 Conclusions

With this paper we have tried to import into the computer vision and pattern recognition fields a profound result from extremal graph theory which asserts essentially that all graphs can be decomposed in such a way as to satisfy a certain uniformity criterion. Since its introduction in the mid-seventies Szemerédi's regularity lemma has emerged as an invaluable tool not only in graph theory but also in theoretical computer science, combinatorial number theory, etc. [10]. Here, we have proposed to take advantage of the properties of regular partitions in a pairwise clustering context. Specifically, in our approach, Szemerédi's decomposition process is used as a preclustering step to substantially reduce the size of the input similarity graph. Clustering is in fact performed on a more compact derived graph using standard algorithms and the solutions obtained are then mapped back into the original graph to create the final groups. Experimental results conducted on standard benchmark datasets from the UCI machine learning repository as well as on image segmentation tasks confirm the effectiveness of the proposed approach. The power of the regularity lemma and the preliminary results obtained on our application make us confident that Szemerédi's result could be potentially used in a variety of other computer vision and pattern recognition problems, whenever large and dense graphs arise.

References

1. Alon, N., Duke, R.A., Lefmann, H., Rödl, V., Yuster, R.: The algorithmic aspects of the regularity lemma. J. of Algorithms, vol. 16, pp. 80–109 (1994) Also, Proc. 33rd IEEE FOCS, Pittsburgh, IEEE, 473–481 (1992)
2. Bengio, Y., Paiement, J., Vincent, P., Delalleau, O., Roux, N.L., Ouimet, M.: Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In: Thrun, S., Saul, L., Schölkopf, B. (eds.) Advances in Neural Information Processing Systems 16, MIT Press, Cambridge, MA (2004)

3. Bollobás, B.: Extremal Graph Theory. Academic Press, London (1978)
4. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. MIT Press, Cambridge, MA (2001)
5. Czygrinow, A., Rödl, V.: An algorithmic regularity lemma for hypergraphs. SIAM J. Comput. 30(4), 1041–1066 (2000)
6. Diestel, R.: Graph Theory, 3rd edn. Springer, New York (2005)
7. Fowlkes, C., Belongie, S., Chung, F.R.K., Malik, J.: Spectral grouping using the Nyström method. IEEE Trans. Pattern Anal. Mach. Intell. 26, 214–225 (2004)
8. Frieze, A.M., Kannan, R.: A simple algorithm for constructing Szemerédi's regularity partition. Electron. J. Comb. 6 (1999)
9. Kohayakawa, Y., Rödl, V., Thoma, L.: An optimal algorithm for checking regularity. SIAM J. Comput. 32(5), 1210–1235 (2003)
10. Komlós, J., Shokoufandeh, A., Simonovits, M., Szemerédi, E.: The regularity lemma and its applications in graph theory. In: Khosrovshahi, G.B., Shokoufandeh, A., Shokrollahi, A. (eds.) Theoretical Aspects of Computer Science: Advanced Lectures, pp. 84–112. Springer, New York (2002)
11. Komlós, J., Simonovits, M.: Szemerédi's regularity lemma and its applications in graph theory. In: Miklós, D., Szonyi, T., Sós, V.T. (eds.) Combinatorics, Paul Erdös is Eighty (Vol. 2), pp. 295–352. Bolyai Society Mathematical Studies 2, Budapest (1996)
12. Motzkin, T.S., Straus, E.G.: Maxima for graphs and a new proof of a theorem of Turán. Canad. J. Math. 17, 533–540 (1965)
13. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE Trans. Pattern Anal. Mach. Intell. 29, 167–172 (2007)
14. Pavan, M., Pelillo, M.: Efficient out-of-sample extension of dominant-set clusters. In: Saul, L.K., Weiss, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems 17, pp. 1057–1064. MIT Press, Cambridge, MA (2005)
15. Roth, V., Laub, J., Kawanabe, M., Buhmann, J.M.: Optimal cluster preserving embedding of nonmetric proximity data. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1540–1551 (2003)
16. Szemerédi, E.: Regular partitions of graphs. In: Colloques Internationaux CNRS 260—Problèmes Combinatoires et Théorie des Graphes, Orsay, pp. 399–401 (1976)