Maximum dispersion problem in dense graphs

A. Czygrinow

Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, USA

Received 26 April 1999; received in revised form 9 July 2000

Abstract

In this note, we present a polynomial-time approximation scheme for a “dense case” of dispersion problem in weighted graphs, where weights on edges are integers from \( \{1, \ldots, K\} \) for some fixed integer \( K \). The algorithm is based on the algorithmic version the regularity lemma. © 2000 Published by Elsevier Science B.V. All rights reserved.

Keywords: Dispersion problem; Algorithm; Regularity lemma

1. Introduction

Let \( G = (V, E, \omega) \), where \( \omega : E \to \{1, \ldots, K\} \) and for \( S \subseteq V \) let \( E(S) \) denote the set of edges which have both endpoints in \( S \). The dispersion problem can be formulated as follows. Given a weighted graph \( G \), and \( s \in \{2, 3, \ldots, |V|\} \) find a subset \( S \) of \( V \) such that \( |S| = s \) and the sum \( \sum_{\{x, y\} \in E(S)} \omega(\{x, y\}) \) is maximized. The problem received some attention recently. In [10], the authors present a polynomial-time algorithm which finds a solution whose value is at least \( \frac{1}{4} \) of the optimal assuming the weights satisfy the triangle inequality. Better bound of \( \frac{1}{4} \), under the same assumptions, was obtained in [7]. In [5] (improving on previous results of [9]), authors obtain a \( O(n^{1/3}) \) — approximation for the dense \( k \)-subgraph problem (unweighted version of the dispersion problem) and they conjecture that for some \( \varepsilon > 0 \), it is NP-hard to obtain an approximation ratio of \( O(n^{\varepsilon}) \).

In this paper, we present an approximation algorithm for a dense case of the dispersion problem, that is in the case when the optimal value is at least \( cn^2 \), for some constant \( c \). In addition, we assume that the weights on edges are integers from \( \{1, \ldots, K\} \) for some fixed \( K \). The method is based on the remarkable lemma of Szemerédi [11] and its algorithmic version due to Alon et al. [1]. For other algorithmic applications of the regularity lemma we refer mainly to [6], which contains applications to the partitioning problems (like MAX-CUT) as well as to [3,4] where some other combinatorial problems are considered. Arora et al. [2] presented a randomized polynomial-time approximation scheme (PTAS) for a dense case of the dense \( k \)-subgraph problem. The algorithm is derandomized in [2]; however, the running time of the derandomized algorithm is \( O(n^{O(1/\varepsilon^2)}) \). Using the regularity lemma we will get rid of the \( 1/\varepsilon^2 \) in the exponent of \( n \) and give a \( O(n^{2}) \) time PTAS for the dispersion problem. It should be emphasized though, that our approach follows the general framework developed in [3,6].
For a set $S \subset V$ denoted by $\text{disp}(S) = \sum_{\{x, y\} \in E(S)} \omega(\{x, y\})$ and let $S_{\text{OPT}}$ be a set of cardinality $s$ for which the value of $\text{disp}(\cdot)$ is maximal. We will show that one can find a set $S^*$ whose value is “close” to the value of $S_{\text{OPT}}$.

**Theorem 1.** For every $0 < \eta < 1$ and integer $K$ there is an algorithm which when given a weighted graph $(V, E, \omega)$ on $n$ vertices and with weights in $\{1, \ldots, K\}$ finds in $O(n^{2.4})$ time a set $S^* \subset V$ such that

$$\text{disp}(S^*) \geq \text{disp}(S_{\text{OPT}}) - \eta n^2.$$  

Clearly, the theorem produces a meaningful output only if $\text{disp}(S_{\text{OPT}}) \geq c n^2$ for some constant $c$. In this case, it can be easily transformed to a polynomial-time approximation scheme (apply the theorem with $\eta' = \eta c$). It is also worth mentioning that the algorithm from Theorem 1 is polynomial in $n = |V|$, and is not polynomial in $\eta$ and $K$. The rest of the note is organized as follows. In Section 2, we formulate the regularity lemma of Szemerédi, Section 3 contains the proof of Theorem 1.

## 2. The regularity lemma

First, let us introduce necessary definitions and formulate the regularity lemma of Szemerédi [11]. Let $G = (V, E)$ be a graph on $n$ vertices. For nonempty subsets $V_1 \subset V$ and $V_2 \subset V$ with $V_1 \cap V_2 = \emptyset$ define the density of the pair $(V_1, V_2)$ as $d(V_1, V_2) = e(V_1, V_2) / |V_1||V_2|$, where $e(V_1, V_2)$ denotes the number of edges with one endpoint in $V_1$ and the other in $V_2$.

**Definition 2.** A pair $(V_1, V_2)$ is called $\varepsilon$-regular if for every $V'_1 \subset V_1$ with $|V'_1| \geq \varepsilon|V_1|$ and every $V'_2 \subset V_2$ with $|V'_2| \geq \varepsilon|V_2|$ we have

$$|d(V_1, V_2) - d(V'_1, V'_2)| \leq \varepsilon.$$

**Definition 3.** A partition $V_1 \cup V_2 \cup \cdots \cup V_t$ of $V$ is called $\varepsilon$-regular if both of the following conditions are satisfied.

1. $|V_i| - |V'_i| \leq 1$, for all $i, j$.
2. All but at most $\varepsilon t^2$ of pairs $(V_i, V_j)$ are $\varepsilon$-regular.

The first condition simply states that all of the partition classes are “almost” equal, the second expresses the fact that at most $\varepsilon$ fraction of all pairs is $\varepsilon$-irregular. Thus, for all $i = 1, \ldots, t$, $|V_i| \in \{\lfloor n/t \rfloor, \lceil n/t \rceil\}$. Since these floors and ceilings are clearly insignificant in our analysis we will simplify the exposition and assume that for every $i$, $|V_i| = n/t$.

The regularity lemma of Szemerédi asserts that every graph which is large enough admits an $\varepsilon$-regular partition into a constant number of classes.

**Theorem 4.** For every $\varepsilon > 0$, and every integer $l$ there exist $N(\varepsilon, l)$ and $L(\varepsilon, l)$ such that every graph with at least $N(\varepsilon, l)$ vertices admits an $\varepsilon$-regular partition $V_1 \cup \cdots \cup V_l$ with $l \leq L(\varepsilon, l)$.

Recently, Alon et al. [1] showed how the above partition can be found algorithmically in $O(|V|^{2.4})$ time.

We need to adopt the above notation to weighted graphs. Let $G = (V, E, \omega)$ be a weighted graph on $n$ vertices with $\omega : E \rightarrow \{1, \ldots, K\}$ where $K$ is constant independent of $n$. Define graphs $G_1, \ldots, G_K$ as $G_i = (V, \omega^{-1}(i))$, that is, $G_i$ is an unweighted graph on $V$ with $\{x, y\} \in E(G_i)$ if $\omega(\{x, y\}) = i$. We use $d_i(U, W)$ to denote the density of the pair $(U, W)$ in $G_i$ and we define the total density of $(U, W)$ as $d(U, W) = \sum_{i=1}^{K} d_i(U, W)$. Thus if $\omega(U, W) = \sum_{i \in U \cap W} \omega(u, w)$ then $d(U, W) = \omega(U, W)/|U||W|$. The following multi-colored version of the regularity lemma follows easily from the original proof of Szemerédi (see [8]).

**Theorem 5.** For every $\varepsilon > 0$ and integers $l$ and $K$ there exist $N = N(\varepsilon, l, K)$ and $L = L(\varepsilon, l, K)$ such that for any collection of $K$ graphs on a vertex set $V$ with $|V| \geq N$ there is a partition $V_1 \cup V_2 \cup \cdots \cup V_t$ which is $\varepsilon$-regular in all of the $K$ graphs and for which $l \leq L$.

One can find such a partition in $O(n^{2.4})$ time using the Alon et al. [1] algorithm.

## 3. Algorithm

In this section, we present the algorithm for the dispersion problem and prove Theorem 1. Given a
partition \( V_1 \cup V_2 \cup \cdots \cup V_t \) of \( V \) define a function \( \text{disp}:2^V \to \{1, \ldots, Kn^2\} \) as follows:

\[
\text{disp}(S) = \sum_{1 \leq i < j \leq t} d(V_i, V_j)|V_i \cap S||V_j \cap S|.
\]

**Algorithm**

1. Fix \( \epsilon = \eta/10K^2 \) and find partition \( V_1 \cup V_2 \cup \cdots \cup V_t \) with \( t \geq 1/\epsilon \) which is \( \epsilon \)-regular with respect to all graphs \( G_1, \ldots, G_K \).
2. Check all subsets \( S \) of size at most \( s \) which are the unions of some of the \( V_i \)'s and take \( S^* \) that maximizes \( \text{disp}(\cdot) \).

Since \( t \) depends only on \( n \) and \( K \) in the second step of the algorithm we check a constant number of sets. Consequently, the complexity of the algorithm is \( O(n^{2.4}) \). Next, in Lemmas 6 and 7, we establish the correctness of the algorithm.

**Lemma 6.** For any set \( S \subset V \) with \( |S| = s \) there is \( \tilde{S} \) which is the union of some of the \( V_i \)'s and that satisfies the following conditions:

1. \( |\tilde{S}| \leq s; \)
2. \( \text{disp}(\tilde{S}) \geq \text{disp}(S) - 2sKn^2 \).

**Proof.** Assume first that there is exactly one \( 1 \leq i \leq t \) such that \( S \cap V_i \neq \emptyset \) and \( S^* \cap V_i \neq \emptyset \). Then for \( \tilde{S} = S \setminus V_i \) we have \( |\tilde{S}| < s \) and

\[
\text{disp}(\tilde{S}) \geq \text{disp}(S) - \left(K \left(\frac{n}{2}\right)^2\right) + \sum_{j \neq i} d(V_j, V_i)|V_j \cap S||V_i \cap S| \geq \text{disp}(S) - K \left(\frac{n^2}{2^2} + \frac{n^2}{t^2}\right).
\]

Suppose now that there exist \( 1 \leq i_1 \neq i_2 \leq t \) such that \( S \cap V_{i_1} \neq \emptyset, S^* \cap V_{i_1} \neq \emptyset \) and \( S \cap V_{i_2} \neq \emptyset, S^* \cap V_{i_2} \neq \emptyset \). We will compare the contribution of \( V_{i_1} \) to \( S \) with the contribution of \( V_{i_2} \) and show that there exist a set \( S' \) with \( |S'| = |S| \) such that the number of \( V_i \)'s that intersect both \( S' \) and \( (S')^c \) is reduced by at least one. Indeed, let

\[
c_1 = \sum_{i \neq i_1, i_2} d(V_i, V_j)|V_i \cap S||V_j \cap S|
\]

and

\[
c_2 = \sum_{i \neq i_1, i_2} d(V_{i_2}, V_j)|V_i \cap S|.
\]

If \( c_1 \geq c_2 \) then let \( S' \) be the set obtained from \( S \) as follows:

- If \( |V_{i_1} \cap S| \leq |V_{i_2} \cap S| \) then add \( V_{i_1} \cap S \) to \( S \) and subtract any subset of \( V_{i_2} \cap S \) of size \( |V_{i_1} \cap S| \).
- If \( |V_{i_1} \cap S| > |V_{i_2} \cap S| \) then subtract \( V_{i_2} \cap S \) and add any subset of \( V_{i_1} \cap S^c \) of size \( |V_{i_1} \cap S| \).

In both cases, we reduce the number of \( V_i \)'s that intersect both \( S \) and \( S^c \) by at least one. Then

\[
\text{disp}(S) \leq \sum_{1 \leq i < j \leq t, i \neq i_1, i_2} d(V_i, V_j)|V_i \cap S||V_j \cap S|
\]

\[
+ d(V_{i_1}, V_{i_2})|V_{i_1} \cap S||V_{i_2} \cap S|
\]

\[
+ c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S|
\]

\[
\leq \sum_{1 \leq i < j \leq t, i \neq i_1, i_2} d(V_i, V_j)|V_i \cap S||V_j \cap S|
\]

\[
+ K \frac{n^2}{t^2} + c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S|.
\]

Note that to obtain \( S' \) we replace some subset of \( V_{i_1} \cap S \) with a subset of \( V_{i_1} \cap S^c \) of the same size. Since \( c_1 \geq c_2 \) we have

\[
c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S| \leq c_1|V_{i_1} \cap S'| + c_2|V_{i_2} \cap S'|.
\]

Therefore,

\[
\text{disp}(S) \leq \text{disp}(S') + K \frac{n^2}{t^2}.
\]

The case \( c_2 > c_1 \) is analogous. In this way, we reduce the number of \( V_i \)'s that intersect both \( S' \) and \( (S')^c \) by at least one and keep \( |S'| = s \). By repeating the above process at most \( t - 1 \) times, we obtain \( S' \) such that at most one of \( V_i \)'s intersect both \( S' \) and \( (S')^c \), \( |S'| = s \), and

\[
\text{disp}(S') \geq \text{disp}(S) - K(t - 1) \frac{n^2}{t^2}.
\]

Now, if there is a single \( V_i \) that intersects both \( S' \) and \( (S')^c \), we subtract \( V_i \) as described in the beginning of
the proof to obtain $\hat{S}$ with $|\hat{S}| < s$ and
$\overline{\text{disp}}(\hat{S}) \geq \text{disp}(\hat{S}) - 2ekn^2$. $\square$

In our next lemma, we show that $\overline{\text{disp}}(\cdot)$ is a “good” approximation to the dispersion function $\text{disp}(\cdot)$.

**Lemma 7.** For any $S \subset V$, $|\overline{\text{disp}}(S) - \text{disp}(S)| \leq 4K^2en^2$.

**Proof.** For every $i = 1, \ldots, t$, we have $|S \cap V_i| = n/i$ and so the number of edges that have both endpoints in $S \cap V_i$, for some $i = 1, \ldots, t$, is at most $m^2/i^2 \leq en^2$. Thus, the total weight of the edges of the above form is at most $K\,en^2$. Therefore,

$$\text{disp}(S) \leq \sum_{1 \leq i < j \leq t} o(V_i \cap S, V_j \cap S) + K\,en^2. \quad (1)$$

If $|V_i \cap S| < \varepsilon|V_i|$ then the number of edges incident to $V_i \cap S$ is at most $en^2/i$. Thus, the number of edges incident to $V_i \cap S$ for which $|V_i \cap S| < \varepsilon|V_i|$, over all $1 \leq i \leq t$, is at most $en^2$ and the total weight on these edges is at most $K\,en^2$. For any $i \neq j$, we have $o(V_i \cap S, V_j \cap S) \leq K\,en^2/i^2$. Since there are at most $en^2$, $\varepsilon$-irregular pairs in all $G_i$’s, the weight on edges between the irregular pairs is at most $K\,en^2$. Thus, combining (1) with above arguments shows that

$$\text{disp}(S) \leq \sum_{1 \leq i < j \leq t} o(V_i \cap S, V_j \cap S) + 3K\,en^2, \quad (2)$$

where the summation is taken over $\{i, j\}$ such that $(V_i, V_j)$ is $\varepsilon$-regular in all $G_1, \ldots, G_K$ and both $|V_i \cap S| \geq \varepsilon|V_i|$ as well as $|V_j \cap S| \geq \varepsilon|V_j|$. For such $\{i, j\}$, accordingly to the definition of an $\varepsilon$-regular pair,

$$d_i(V_i \cap S, V_j \cap S) \leq d_i(V_i, V_j) + \varepsilon. \quad (3)$$

Thus,

$$o(V_i \cap S, V_j \cap S) = \sum_{l=0}^{K} le_i(V_i \cap S, V_j \cap S)$$

$$= \sum_{l=0}^{K} l\,d_i(V_i \cap S, V_j \cap S)|V_i \cap S||V_j \cap S|, \quad (4)$$

where $e_i(A, B)$ denotes the number of edges between $A$ and $B$ that have weight $l$. Therefore, combining (3) and (4) yields

$$o(V_i \cap S, V_j \cap S) \leq \sum_{l=0}^{K} l\,d_i(V_i \cap S, V_j \cap S) = K^2\epsilon^2 \sum_{l=0}^{K} l^2.$$ 

Consequently, (as $K \leq K^2$)

$$\text{disp}(S) \leq \sum_{1 \leq i < j \leq t} l\,d_i(V_i, V_j)|V_i \cap S||V_j \cap S| + 4K^2en^2$$

$$\leq \sum_{1 \leq i < j \leq t} l\,d_i(V_i, V_j)|V_i \cap S||V_j \cap S| + 4K^2en^2$$

In a similar way, one can show that

$$\text{disp}(S) \leq \text{disp}(S) + 4K^2en^2. \quad \square$$

We are now ready to prove Theorem 1.

**Proof.** Let $S_{\text{OPT}}$ denote a set of size $s$ with maximal dispersion and let $S$ be a set found by the algorithm. Then by Lemma 7

$$\text{disp}(S_{\text{OPT}}) \leq \text{disp}(S_{\text{OPT}}) + 4K^2en^2,$$

which by Lemma 6 is less than $\overline{\text{disp}}(S) + 6K^2en^2$. Applying Lemma 7 again yields

$$\text{disp}(S_{\text{OPT}}) \leq \text{disp}(S) + 10K^2en^2.$$

Since $\varepsilon = \eta/10K^2$, we found $S$ that satisfies

$$\text{disp}(S) \geq \text{disp}(S_{\text{OPT}}) - \eta^2.$$

Finally, if $|S| < s$ then we add arbitrarily $s - |S|$ vertices to obtain $S^*$ such that $|S^*| = s$ and

$$\text{disp}(S^* \geq \text{disp}(S) \geq \text{disp}(S_{\text{OPT}}) - \eta^2. \quad \square$$

**References**

