MAXIMUM DISPERSION PROBLEM IN DENSE GRAPHS

A. CZYGRINOW

ABSTRACT. In this note, we will 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.

1. INTRODUCTION

Let $G=(V,E,\omega)$, where $\omega: E \to \{1,\ldots,K\}$ and for $S \subset V$ 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 [12], the authors present a polynomial time algorithm which finds a solution which value is at least $1/4$ of the optimal assuming the weights satisfy the triangle inequality and the better bound of $1/2$, under the same assumptions, was obtained in [14]. In [5] (improving on previous results of [11]), the authors obtain a $O(n^{1/3})$- approximation for the dense k-subgraph problem and they conjecture that for some $\epsilon > 0$ it is NP-hard to obtain an approximation ratio of $O(n^{\epsilon})$.

In this paper, we present an approximation algorithm for the 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 [13] and its algorithmic version due to Alon et al. [1]. For other algorithmic applications of the regularity lemma we refer mainly to [7] which contains applications to the partitioning problems (like MAX-CUT) as well as to [4] and [3] where some other combinatorial problems are considered.

For a set $S \subset V$ denote by $disp(S) = \sum_{\{x,y\} \in E(S)} \omega(x,y)$ and let S_{OPT} be a set of cardinality s for which the value of $disp(\cdot)$ is maximal. We will show that one can find a set S^* which value is "close" to the value of S_{OPT}.

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

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

Clearly the theorem produces a meaningful output only if $disp(S_{OPT}) \geq cn^2$ 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 η and K (which are fixed).

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 [13]. 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 $d(V_1, V_2) = \frac{e(V_1, V_2)}{|V_1||V_2|}$ 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 ϵ-regular if for every $V_1' \subset V_1$ with $|V_1'| \geq \epsilon|V_1|$ and every $V_2' \subset V_2$ with $|V_2'| \geq \epsilon|V_2|$ we have

$$|d(V_1, V_2) - d(V_1', V_2')| \leq \epsilon.$$

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

1. $||V_i| - |V_j|| \leq 1$, for all i, j.
2. All but ϵt^2 of pairs (V_i, V_j) are ϵ-regular.

The first condition simply states that all of the partition classes are "almost" equal, the second expresses the fact that at most ϵ fraction of all pairs is ϵ-irregular.

The regularity lemma of Szemerédi asserts that every graph which is large enough admits an ϵ-regular partition with a constant number of classes.

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

Recently, Alon, Duke, Lefmann, Rödl, and Yuster [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 $\omega : E \to \{1, \ldots, K\}$ where K is constant independent of n. Define graphs $G_1, \ldots, G_K$ as $G_l = (V, \omega^{-1}(l))$, that is, G_l is an unweighted graph on V with $\{x, y\} \in E(G_l)$ if $\omega(\{x, y\}) = l$. We use $d_l(U, W)$ to denote the density of the pair (U, W) in G_l and we define the total density of (U, W) as

$$d(U,W) = \sum_{l=0}^{K} l d_l(U,W).$$

Thus if $\omega(U,W) = \sum_{u\in U, w\in W} \omega(u,w)$ then $d(U,W) = \frac{\omega(U,W)}{|U||W|}$. The following multicolored version of the regularity lemma follows easily from the original proof of Szemerédi (see [10]).

Theorem 5. *For every $\epsilon > 0$ and integers l and K there exist $N = N(\epsilon, l, K)$ and $L = L(\epsilon, l, K)$ 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 ϵ-regular in all of the K graphs and for which $l \leq t \leq L$.*

One can find such a partition in $O(n^{2.4})$ time using the Alon, Duke, Lefmann, Rödl, and Yuster [1] algorithm.

3. Algorithm

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

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

Algorithm

1. Fix $\epsilon = \frac{\eta}{10K^2}$ and find partition $V_1 \cup V_2 \cup \cdots \cup V_t$ with $t \geq 1/\epsilon$ which is ϵ-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 $\overline{disp}(\cdot)$.

Since t depends only on η 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 Lemma 6 and 7, we shall establish the correctness of the algorithm.

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

1. $|\bar{S}| \leq s$.
2. $\overline{disp}(\bar{S}) \geq \overline{disp}(S) - 2\epsilon K n^2$.

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

$$\overline{disp}(\bar{S}) \geq \overline{disp}(S) - (K\binom{\frac{n}{t}}{2} + \sum_{j \neq i} d(V_j, V_i)|V_j \cap S||V_i \cap S|) > \overline{disp}(S) - K(\frac{n^2}{t^2} + \frac{n^2}{t}).$$

Suppose now that there exist $1 \leq i_1 \neq i_2 \leq t$ such that $S \cap V_{i_1} \neq \emptyset$, $S^c \cap V_{i_1} \neq \emptyset$ and $S \cap V_{i_2} \neq \emptyset$, $S^c \cap V_{i_2} \neq \emptyset$. 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

$$c_1 = \sum_{i \neq i_1, i_2} d(V_{i_1}, V_i)|V_i \cap S|$$

and

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

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

- If $|V_{i_1} \cap S^c| \le |V_{i_2} \cap S|$ then add $V_{i_1} \cap S^c$ to S and subtract any subset of $V_{i_2} \cap S$ of size $|V_{i_1} \cap S^c|$.
- If $|V_{i_1} \cap S^c| \ge |V_{i_2} \cap S|$ then subtract $V_{i_2} \cap S$ and add any subset of $V_{i_1} \cap S^c$ of size $|V_{i_2} \cap S|$.

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

$$\overline{disp}(S) = \sum_{1 \le i < j \le t, i,j \ne i_1, i_2} d(V_i, V_j)|V_i \cap S||V_j \cap S| +$$

$$d(V_{i_1}, V_{i_2})|V_{i_1} \cap S||V_{i_2} \cap S| + c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S|$$

$$\le \sum_{1 \le i < j \le t, i,j \ne i_1, i_2} d(V_i, V_j)|V_i \cap S||V_j \cap S| + K\frac{n^2}{t^2} + c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S|.$$

Note that to obtain S' we replace some subset of $V_{i_2} \cap S$ with a subset of $V_{i_1} \cap S^c$ of the same size. Since $c_1 \ge c_2$ we have

$$c_1|V_{i_1} \cap S| + c_2|V_{i_2} \cap S| \le c_1|V_{i_1} \cap S'| + c_2|V_{i_2} \cap S'|.$$

Therefore

$$\overline{disp}(S) \le \overline{disp}(S') + K\frac{n^2}{t^2}.$$

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

$$\overline{disp}(S') \ge \overline{disp}(S) - K(t-1)\frac{n^2}{t^2}.$$

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 $\bar{S}$ with $|\bar{S}| < s$ and

$$\overline{disp}(\bar{S}) \geq \overline{disp}(S) - 2\epsilon K n^2.$$

□

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

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

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

$$disp(S) \leq \sum_{1 \leq i < j \leq t} \omega(V_i \cap S, V_j \cap S) + K\epsilon n^2. \tag{1}$$

If $|V_i \cap S| < \epsilon|V_i|$ then the number of edges incident to $V_i \cap S$ is at most $\epsilon\frac{n^2}{t}$. Thus, the number of edges incident to $V_i \cap S$ for which $|V_i \cap S| < \epsilon|V_i|$,over all $1 \leq i \leq t$, is at most ϵn^2 and the total weight on these edges is at most $K\epsilon n^2$.

For any $i \neq j$, we have $\omega(V_i \cap S, V_j \cap S) \leq K\frac{n^2}{t^2}$. Since there are at most ϵt^2, ϵ-irregular pairs in all G_i's, the weight on edges between the irregular pairs is at most $K\epsilon n^2$. Thus, combining (1) with above arguments shows that

$$disp(S) \leq \sum \omega(V_i \cap S, V_j \cap S) + 3K\epsilon n^2, \tag{2}$$

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

$$d_l(V_i\cap S,V_j\cap S)\leq d_l(V_i,V_j)+\epsilon. \tag{3}$$

Thus,

$$\omega(V_i\cap S,V_j\cap S)=\sum_{l=0}^{K} le_l(V_i\cap S,V_j\cap S)=\sum_{l=0}^{K} ld_l(V_i\cap S,V_j\cap S)|V_i\cap S||V_j\cap S|, \tag{4}$$

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

$$\omega(V_i\cap S,V_j\cap S)\leq\sum_{l=0}^{K} ld_l(V_i,V_j)|V_i\cap S||V_j\cap S|+K^2\epsilon\frac{n^2}{t^2}.$$

Consequently, (as $K<K^2$)

$$disp(S)\leq\sum\sum_{l=0}^{K} ld_l(V_i,V_j)|V_i\cap S||V_j\cap S|+4K^2\epsilon n^2\leq$$

$$\sum_{1\leq i<j\leq t}\sum_{l=0}^{K} ld_l(V_i,V_j)|V_i\cap S||V_j\cap S|+4K^2\epsilon n^2=\overline{disp}(S)+4K^2\epsilon n^2.$$

In the similar way one can show that

$$\overline{disp}(S)\leq disp(S)+4K^2\epsilon n^2.$$

□

We are now ready to prove Theorem 1.

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

$$disp(S_{OPT}) \leq \overline{disp}(S_{OPT}) + 4K^2\epsilon n^2,$$

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

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

Since $\epsilon = \frac{\eta}{10K^2}$, we found S that satisfies

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

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

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

□

References

[1] N. Alon, R.A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the Regularity Lemma, Journal of Algorithms 16, (1994), 80-109.

[2] B. Chandra, M. Halldorsson, Facility dispersion and remote subgraphs, Proc. 5th Scandinavian Workshop on Algorithm Theory, Lecture Notes in Computer Science, vol. 1097, Springer, Berlin, (1996), 53-65.

[3] A. Czygrinow, S. Poljak and V. Rödl, Constructive Quasi-Ramsey numbers and tournament ranking, to appear in SIAM J. Discrete Math. 12, (1999), no. 1, 48-63.

[4] R.A. Duke, H. Lefmann and V. Rödl, A fast algorithm for computing frequencies in a given graph, SIAM Journal on Computing 24 (3), (1995), 598-620.

[5] U. Feige, G. Kortsarz, D. Peleg, The dense k-subgraph Problem, manuscript, (1999).

[6] T.A. Feo, M. Khellaf, A class of bounded approximation algorithms for graph partitioning, Networks 20, (1990), 181-195.

[7] A. Frieze and R. Kannan, The Regularity Lemma and Approximation schemes for dense problems, Proc 37th FOCS, (1996), 12-20.

[8] J. B. Ghosh, Computational aspects of the maximum diversity problem, Operation Research Letters, 19, (1996), 175-181.

[9] R. Hassin, S. Rubinstein, A. Tamir, Approximation algorithms for maximum dispersion, Operation Research Letters, 21, (1997), 133-137.

[10] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, in *Combinatorics, Paul Erdős is Eighty, Vol. 2* (D. Miklós, V. T. Sós, T. Szőnyi, eds.), János Bolyai Math. Soc., Budapest (1996), 295-352.

[11] G. Kortsarz, D. Peleg, On choosing a dense subgraph, Proc. 34th IEEE Symp. on Foundations of Computer Science, (1993), 692-701.

[12] S. S. Ravi, D. J. Rosenkrantz, G. K. Tayi, Heuristic and special case algorithms for dispersion problems, Oper. Res., 42, (1994), 299-310.

[13] E. Szemerédi, Regular Partitions of Graphs, Colloques Internationaux C.N.R.S., Problemes Combinatories et Theorie des Graphes, 399-402.

[14] A. Tamir, Obnoxious facility location on graphs, SIAM J. Discrete Math. 4, (1991), 550-567.

Dept. of Mathematics, Arizona State University, Tempe, AZ 58287