On the practically interesting instances of MAXCUT

Yonatan Bilu[1], Amit Daniely*[2], Nati Linial†[3], and Michael Saks‡[4]

1 Parasight inc
Agudat sport hapoel 1, Jerusalem, Israel.
yonatan@gmail.com

2 Department of Mathematics, Hebrew University
Jerusalem 91904, Israel.
amit.daniely@math.huji.ac.il

3 School of Computer Science and Engineering, Hebrew University
Jerusalem 91904, Israel.
nati@cs.huji.ac.il

4 Department of Mathematics, Rutgers University
Piscataway, NJ 08854.
saks@math.rutgers.edu.

Abstract

For many optimization problems, the instances of practical interest often occupy just a tiny part of the algorithm's space of instances. Following [6], we apply this perspective to MAXCUT, viewed as a clustering problem. Using a variety of techniques, we investigate practically interesting instances of this problem. Specifically, we show how to solve in polynomial time distinguished, metric, expanding and dense instances of MAXCUT under mild stability assumptions. In particular, $(1+\epsilon)$-stability (which is optimal) suffices for metric and dense MAXCUT. We also show how to solve in polynomial time $\Omega(\sqrt{n})$-stable instances of MAXCUT, substantially improving the best previously known result.

1998 ACM Subject Classification F.2.0 Analysis of algorithms and problem complexity

Keywords and phrases MAXCUT, Clustering, Hardness in practice, Stability, Non worst-case analysis

Digital Object Identifier 10.4230/LIPIcs.STACS.2013.526

1 Introduction

As has been noted many times, worst case complexity is often an overly restrictive metric for algorithms. In practice, a more realistic (but fuzzy) criterion would be to say that a problem is feasible if there is an efficient algorithm that correctly solves all of its *practically interesting* instances. The difference can be very substantial, since for many computational problems, the vast majority of instances are completely irrelevant for practical purposes.

An important case in point is clustering, where one seeks a meaningful partition of a given set of data. Almost every formal manifestation of the clustering problem is NP-Hard,

* Supported in part by a binational Israel-USA grant 2008368 and by a Google Europe Fellowship in Learning Theory.
† Supported in part by a binational Israel-USA grant 2008368.
‡ Supported in part by NSF under grants CCF-0832787 and CCF-1218711, and by a binational Israel-USA grant 2008368. Part of this work was done while on sabbatical at Princeton University.

30th Symposium on Theoretical Aspects of Computer Science (STACS'13).
Editors: Natacha Portier and Thomas Wilke; pp. 526–537
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

yet, a clustering instance is of practical interest only if the data *can indeed* be partitioned in a meaningful way, and such data sets are very special. Thus, even if no efficient algorithm can find the optimal partition for *every* data set, this does not imply that clustering is hard in practice. As Tali Tishby put it in conversation many years ago, many practitioners hold the opinion that "clustering is either easy or pointless". That is, for a data sets that admit a meaningful partition of the data, finding it is not hard.

Bilu and Linial [6] proposed a framework for studying this issue which applies to optimization problems with a continuous input space and discrete solution space. They proposed two criteria for an optimal solution to be *evidently* optimal. An optimal solution is *stable* if it remains optimal under moderate perturbations of the input. An optimal solution is *distinguished* if the objective function value at any other point is reduced by at least an amount proportional to the distance to the optimal point.

Following [6] we study the (weighted) MAXCUT problem in this framework. Here the input is a weighted graph and the candidate solutions are cuts. A cut is γ-stable (for $\gamma \geq 1$) if it remains optimal even if each input weight w_{ij} is perturbed to a value between w_{ij} and γw_{ij}. A cut is α-*distinguished* (for $\alpha \geq 0$) if moving to any other cut reduces the objective function by at least α times the sum of (weighted) degrees of the vertices that switched side. We also consider a weakening of stability called γ-*local stability.*

Our main results are:

▶ **Theorem 1.** *1. For every $\epsilon > 0$ there is a polynomial time algorithm that correctly solves all $(1+\epsilon)$-locally stable instances of Metric-MAXCUT.*
2. For every $\epsilon > 0$ and $C > 1$ there is a polynomial time algorithm that correctly solves all MAXCUT instances that are $(1+\epsilon)$-locally stable and C-dense.

The condition of C-density (defined in Section 1.2) rules out weight being too concentrated.

▶ **Theorem 2.** *There is a polynomial time algorithm that solves every instance of MAXCUT that is α-distinguished and γ-locally stable with $\gamma > \frac{2}{1-\sqrt{1-\alpha^2}}$. In fact, it suffices that the instance be γ-locally stable with $\gamma > \frac{2}{1-\sqrt{1-h^2}}$, where h is the Cheeger constant of the weighted graph induced by the maximal cut.*

This improves a result from [6] that works only for regular graphs and requires that $\gamma > \frac{5+\sqrt{1-\alpha^2}}{1-\sqrt{1-\alpha^2}}$ or $\gamma > \frac{5+\sqrt{1-h^2}}{1-\sqrt{1-h^2}}$.

▶ **Theorem 3.** *There is a polynomial time algorithm that finds the optimal solution for every $\Omega(\sqrt{n})$-stable instance of MAXCUT.*

This improves on a result in [6] which needed $\Omega(n)$-stability.

Some notation and terminology

The input to the MAXCUT problem on vertex set V is a symmetric weight function $w : V \times V \to \mathbb{R}^+$ with 0 diagonal. Expressions such as "w is bipartite" refer to the graph which is the support of w, which we assume to be connected. The objective is to find the *cut* $(S, \bar{S})$, $S \subseteq V$ of maximum total weight $\sum_{a\in S,\ b\in\bar{S}} w(a, b)$.

For a fixed cut $(S, \bar{S})$, we use the self-explanatory terms "the vertices x, y are *on the same side*" or *"separated"* by this cut. The edge xy a *cut edge* or a *non-cut edge* when x, y are separated resp. on the same side of the cut. For $A, B \subset V$, we define $E(A, B) = \{ab|a \in A,\ b \in B\}$, and $w(A, B) := \sum_{uv\in E(A,B)} w(u, v)$. Also, $\tau_w(A) = \tau(A) = w(A, \bar{A})$ and $\mu(A) = \mu_w(A) = w(A, V)$. Fo $A \subseteq V$, $\xi(A)$, $\xi(A) = \sum_{vu\in E(A,\bar{A})\cap E(S,\bar{S})} w(u, v)$ is the sum of weights of cut edges leaving A and $\iota(A) = \tau(A) - \xi(A)$ is the weight of the non-cut edges.

(The reader may find the following mnemonic useful: τ stands for "total", ξ for "external" and ι for "internal"). We slightly abuse notation for singletons $A = \{v\}$ and pairs $A = \{u, v\}$ and write $\tau(v)$ or $\iota(e)$ etc., where $e = uv$. The *minimal, maximal* and *average* degree of w are denoted by $\underline{\delta}(w) = \min_{v \in V} \mu(v)$, $\bar{\delta}(w) = \max_{v \in V} \mu(v)$ and $\delta(w) = \frac{\sum_{v \in V} \mu(v)}{n}$ respectively.

1.1 Stable instances

▶ **Definition 4.** Let $w : V \times V \to [0, \infty)$ be an instance of MAXCUT and let $\gamma \geq 1$. An instance $w' : V \times V \to [0, \infty)$ is a γ*-perturbation* of w if

$$\forall u, v \in V,\ w(u, v) \leq w'(u, v) \leq \gamma \cdot w(u, v)$$

An instance w is said to be γ*-stable* if there is a cut which forms a maximal cut for every γ-perturbation w' of w.

▶ **Definition 5.** Let $\gamma \geq 1$. An instance $w : V \times V \to [0, \infty)$ for MAXCUT is γ*-locally stable* if there is a maximal cut $(S, \bar{S})$ for which it is impossible to obtain a larger cut by switching the side of some vertex x and multiplying the edges in $E(x, V \setminus \{x\})$ by numbers between 1 and γ.

The definitions of stability and local stability capture the intuition of an "evidently optimal" solution. The following more concrete equivalent definitions are usually more convenient to use.

▶ Observation 1. [6] Let $w : V \times V \to \mathbb{R}$ be an instance of MAXCUT and let $\gamma \geq 1$.

- w is γ-stable iff there is a maximal cut for which $\xi(A) \geq \gamma \cdot \iota(A)$ for every $A \subset V$.
- w is γ-locally stable iff there is a maximal cut for which $\xi(x) \geq \gamma \cdot \iota(x)$ for every $x \in V$.

We say that a (not necessarily maximal) cut $(S, \bar{S})$ is γ*-stable* (resp. γ*-locally stable*) if the first (resp. second) condition in Observation 1 holds.

Every instance is 1-stable and it is easy to see that there is a unique maximal cut if and only if the instance is γ-stable for some $\gamma > 1$.

Stability and local stability are quite different. For $\gamma > 1$ an instance has at most one γ-stable cut but may have many γ-locally stable cuts. The instance where $w = 1$ on the edges of a perfect matching and $\epsilon > 0$ elsewhere. As $\epsilon \to 0$, the local stability tends to ∞ and has exponentially many γ-locally stable maximal cuts, but is not γ-stable for any $\gamma > 1$. While MAXCUT remains NP-hard even under arbitrarily high local stability (see [6]), in Section 2 we prove Theorem 3 by giving an efficient algorithm for $\Omega(\sqrt{n})$-stable instances. Also, it is easy to decide whether a given cut is γ-locally stable, but we do not know how to decide whether a given cut is γ-stable.

1.2 Metric and Dense instances

In Section 3 we study metric instances. This is done through a reduction from metric to dense instances, so we consider such instances as well (Section 3.1).

We call $w : V \times V \to \mathbb{R}$ C**-dense** for $C \geq 1$ if $\forall x, y \in V,\ w(x, y) \leq C \cdot \frac{\tau(x)}{n}$. As shown in [2], for $C > 1$ fixed, C-dense MAXCUT is NP-Hard, but it has a PTAS. As we show, this PTAS can be adapted to correctly solve all instances of MAXCUT that are $(1 + \epsilon)$-locally stable and C-dense for every $\epsilon > 0, C > 1$. The algorithm samples $O(\log n)$ vertices and tests each of their bipartitions as a seed to a cut. As we show, w.h.p., one of the resulting cuts is the maximal cut, proving the second part of Theorem 1.

In Section 3.2 we deal with Metric-MAXCUT. As shown in [15] (with credit to L. Trevisan) Metric-MAXCUT is NP-Hard. That paper also gives a reduction from metric to $(4 + o(1))$-dense instances of MAXCUT, thus yielding a PTAS for Metric-MAXCUT. We prove Theorem 1 by showing that a slight variation of this reduction preserves local stability[1], and therefore yields an efficient algorithms for $(1 + \epsilon)$-locally stable instances of Metric-MAXCUT.

The exponent for this algorithm is quite large. We also provide a faster algorithm for $(3 + \epsilon)$-locally stable metric instances.

1.3 Distinguished and Expanding instances

Let $w : V \times V \to \mathbb{R}^+$ be an instance of MAXCUT whose (unique) maximal cut is $(S, \bar{S})$. We note that if all vertices of $A \subset V$ switch side, then the weight of the cut decreases by $\xi(A) - \iota(A)$. Thus, we define

▶ **Definition 6.** An instance w of MAXCUT is α*-distinguished* for $1 \geq \alpha \geq 0$ if for every $\emptyset \neq A \subset V$, $\xi(A) - \iota(A) \geq \alpha \cdot \min\{\mu(A), \mu(\bar{A})\}$.

Note that every instance is 0-distinguished and being α-distinguished with $\alpha > 0$ is equivalent to having a unique maximal cut. It is not hard to see that $\frac{1+\alpha}{1-\alpha}$-local stability is equivalent to α*-local distinction*, namely $\xi(x) - \iota(x) \geq \alpha \cdot \mu(x)$ for every $x \in V$.

Distinction vs Stability. Let $(S, \bar{S})$ be a maximal cut of $w : V \times V \to [0, \infty)$. On the one hand, every α-distinguished instance is $\frac{1+\alpha}{1-\alpha}$-stable, because $\xi(A) - \iota(A) \geq \alpha\mu(A) \geq \alpha(\xi(A) + \iota(A))$. On the other hand, highly stable instances need not be distinguished as the following bipartite example with $V = \{a_1, \ldots, a_n\} \dot\cup \{b_1, \ldots, b_n\}$ shows. Here $w(a_i, b_j)$ is 1 when $i = j$ and $\epsilon \ll 1$ otherwise. Clearly w is ∞-stable. Yet, switching the sides of all the vertices in $\{a_1, \ldots, a_{\frac{n}{2}}\} \cup \{b_1, \ldots, b_{\frac{n}{2}}\}$ decreases the weight of the cut only slightly. Such examples motivate the stronger notion of distinction. Although the cut $(\{a_1, \ldots, a_n\}, \{b_1, \ldots, b_n\})$ is infinitely stable, its optimality does not seem completely evident.

Distinction and Expansion. Call $w : V \times V \to \mathbb{R}^+$ β*-expanding* if $\beta \leq h(w)$ where $h(w) = \min_{\emptyset \neq A \subset V} \frac{\tau(A)}{\min\{\mu(A), \mu(\bar{A})\}}$ is w's *Cheeger constant.* An α-distinguished instance is α-expanding, though highly expanding instances can even have multiple maximal cuts. However, an instance that is both γ-stable and β-expanding is easily seen to be $(\beta \cdot \frac{\gamma-1}{\gamma+1})$-distinguished. As this discussion implies, distinction is a conjunction of stability and expansion.

In section 4 we prove Theorem 2, using a spectral result from [6].

1.4 Spectral algorithms

In Section 5 we consider consider whether the Goemans-Williams approximation algorithm can be used to exactly solve instances of high stability or local stability.

1.5 Other related work

Recently, Balcan and Liang [5] introduced a relaxed version of stability in which the optimal solution is allowed to change slightly under perturbations of the input, and obtained good algorithms for clustering under the k-medians objective. In [3] polynomial time algorithms are given for 3-stable instances of k-means, k-medians and other “center based" clustering

[1] A word of caution: Our definition of stability and local stability for Metric-MAXCUT is more restrictive than one might think. We require the perturbed instance to satisfy the stability condition *whether or not* it is metric.

problems. The constant 3 was improved in [5] to $(1+\sqrt{2})$ for k-median. The papers [9, 1, 4] consider data sets that admit a good clustering and show how to cluster them efficiently.

Smoothed analysis is the best known example of a method for analysing instances of computational problems based on their practical significance. As this method shows [14], a certain variant of the simplex algorithm solves in polynomial time *almost every* input.

The MAXCUT problem has been shown to solvable in polynomial time with high probability in random models (e.g. [7]) and *semirandom model [10].*

2 Algorithms for stable instances

▶ Observation 2. Let w be a γ-stable instance of MAXCUT, and let w' be obtained from w by merging two vertices[2] on the same side of w's maximal cut. Then w' is γ-stable and its maximal cut is induced from w's maximal cut.

Thus, it suffices to give an efficient algorithm that, for any γ-stable instance, finds a pair of vertices on the same side of the optimal cut. Once two such vertices are found, we merge them and proceed recursively. This applies as well when γ is a non-decreasing function of n.

As an easy warm-up, we show how to find such a pair of vertices in a $2n$-stable MAXCUT instance w, simplifying an algorithm from [6]. Let vu be a heaviest edge, and let vz be the heaviest edge incident on either u or v. We claim that both vu and vz are cut edges and so u and z are on the same side of the cut. To see this Clearly $w(v,u) \geq \frac{1}{n-1}\tau(v)$. By observation 1, $\iota(v) \leq \frac{1}{2n+1}\tau(v)$, so $w(v,u) > \iota(v)$ and we conclude that vu is a cut edge. Again, $w(v,z) \geq \frac{1}{2(n-2)}\tau(\{u,v\})$ and by observation 1, $\iota(\{v,u\}) \leq \frac{1}{2n+1}\tau(\{v,u\})$, implying that $w(v,z) > \iota(\{v,u\})$. Consequently vz is a cut edge.

2.1 A deterministic algorithm for $O(\sqrt{n})$-stable instances

Following observation 2, we will find two vertices which are on the same side of the cut. To find this pair, we'll only need the condition in Observation 1 to hold for subsets $A \subseteq V$ with $|A| \leq 2$. (But full stability is needed to apply induction after merging.) Let w be a γ-stable instance of MAXCUT with $\gamma > \sqrt{8n+4}+1$ and let $(S,\bar{S})$ be a maximal cut. We first deal with very heavy edges. Let T^1 be the set of edges vu for which $w(v,u) > \mu(v)/(\gamma+1)$. By observation 1, all edges in T^1 are cut edges. Thus if there are two incident edges $uv, vz \in T^1$, then u and z are on the same side of the cut and we are done. It remains to consider the case where T^1 is a matching. Let T^2 be the set of edges not in T^1 that satisfy $w(u,v) > \tau(\{u,z\})/(\gamma+1)$ for some $uz \in T^1$. Again, by observation 1, all edges in T^2 are cut edges. If T^2 is nonempty, say $uv \in T^2$, then there exists some $uz \in T^1$ with $w(u,v) > \frac{1}{\gamma+1}\tau(\{u,z\})$, which implies that v and z are on the same side of the cut. We proceed to consider the case where T^2 is empty.

For every $u,v \in V$ define

$$\tilde{w}(u,v) = \begin{cases} 0 & vu \in T^1 \\ w(u,v) & o/w \end{cases}, \quad \hat{w}(v) = \begin{cases} \tau(\{u,v\}) & vu \in T^1 \text{ for some } u \in V \\ \tau(v) & o/w \end{cases}$$

Note that $\hat{w}(v)$ is well defined, since T^1 is a matching by assumption. Since $T^2 = \emptyset$ and T^1 is a matching, we have, for every $u \in V$, $\tilde{w}(v,u) \leq \frac{1}{\gamma+1}\hat{w}(v)$ and, again by observation

[2] Let $w : V \times V \to \mathbb{R}$ be an instance and let $v,u \in V$. The instance $w' : V' \times V' \to \mathbb{R}$ obtained upon **merging** v,u is defined as follows. $V' = V \setminus \{u,v\} \cup \{v'\}$ and $w'(x,y) = w(x,y)$ for $x,y \in V \setminus \{v,u\}$, also, $w'(v',x) = w(v,x) + w(u,x)$.

1, $\iota(v) \le \frac{1}{\gamma+1}\hat{w}(v)$. Next, we observe as well that separated vertices cannot have too many common neighbours. For $u, v \in V$ we define $n(u,v) := \sum_{z\in V} \tilde{w}(v,z)\tilde{w}(z,u)$. If v and u are separated, say $v \in S, u \in \bar{S}$, then

$$\begin{aligned} n(u,v) &= \sum_{z\in\bar{S}} \tilde{w}(v,z)\tilde{w}(z,u) + \sum_{z\in S} \tilde{w}(v,z)\tilde{w}(z,u) \\ &\le \frac{1}{\gamma+1}\hat{w}(v)\cdot\iota(u) + \frac{1}{\gamma+1}\hat{w}(u)\cdot\iota(v) \le \frac{2}{(\gamma+1)^2}\hat{w}(u)\cdot\hat{w}(v). \end{aligned}$$

Thus, it suffices to find two vertices v, u with $n(u,v) > \frac{2}{(\gamma+1)^2}\hat{w}(u)\cdot\hat{w}(v)$, and place them on the same side of the cut. Indeed, if no such pair exists we have

$$\begin{aligned} \frac{1}{4}\sum_{v\in V}\hat{w}^2(v) &\le \sum_{v\in V}\tau^2_{\tilde{w}}(v) = \sum_{u,v,z\in V} \tilde{w}(u,z)\tilde{w}(z,v) \\ &= \sum_{u,v\in V,\ u\ne v} n(u,v) + \sum_{u,z\in V} \tilde{w}^2(u,z) \\ &\le \frac{2}{(\gamma+1)^2}\sum_{u,v\in V,\ u\ne v} \hat{w}(u)\hat{w}(v) + \sum_{u\in V}\frac{1}{\gamma+1}\hat{w}(u)\sum_{z\in V}\tilde{w}(u,z) \\ &\le \frac{2}{(\gamma+1)^2}(\sum_{u\in V}\hat{w}(u))^2 + \frac{1}{\gamma+1}\sum_{u\in V}\hat{w}(u)\tau_{\tilde{w}}(u) \\ &\le \frac{2n}{(\gamma+1)^2}\sum_{u\in V}\hat{w}^2(u) + \frac{1}{\gamma+1}\sum_{u\in V}\hat{w}^2(u), \end{aligned}$$

from which we obtain the contradiction $\gamma \le \sqrt{8n+4}+1$.

3 Algorithms for locally stable dense and metric instances

3.1 Dense instances

▶ **Theorem 7.** *For every $C \ge 1$ and $\epsilon > 0$ there is a randomized polynomial time algorithm that correctly solves all $(1+\epsilon)$-locally stable, C-dense instances of MAXCUT.*

The analysis of the algorithm is based on the following lemma.

▶ **Lemma 8.** *Suppose that $w : V\times V \to [0,\infty)$ is a C-dense instance and let $(S,\bar{S})$ be a γ-locally stable cut. Let $X_1,\dots,X_m$ be i.i.d. r.v. that are uniformly distributed on V. For $x \in V$, let A_x be the event that $S_+ > S_-$, where $S_\pm = \sum w(x,X_i)$ over all i s.t. x and X_i are separated resp. on the same side. Then*

$$\Pr\left(\cup_x A_x\right) \le |V|\cdot\exp\left(-\frac{1}{2}\left(\frac{1}{C}\cdot\frac{\gamma-1}{\gamma+1}\right)^2\cdot m\right)$$

Proof. For every $x \in V$, $S_+ - S_-$ is a sum of m i.i.d. r.v.'s of expectation $\frac{\xi(x)-\iota(x)}{|V|} \ge \frac{\gamma-1}{\gamma+1}\frac{\tau(x)}{|V|}$. These r.v.'s are bounded in absolute value, by $C\cdot\frac{\tau(x)}{|V|}$. Now apply Hoeffding's bound. ◻

Proof. (Of Theorem 7) Let $D = 2\left(C\cdot\frac{2+\epsilon}{\epsilon}\right)^2$ and $m = D\cdot\ln(2|V|)$. Let $X_1,\dots,X_m$ be independent uniform random samples from V. By Lemma 8, with probability ≥ 0.5, there is a partition $\{X_1,\dots,X_m\} = L\coprod R$ such that the cut defined by $S = \{x\in V : w(x,R) > w(x,L)\}$ is optimal. Now simply enumerate over the $(2\cdot|V|)^{\ln(2)D} = n^{O(1)}$ such partitions.

3.2 Metric instances

Given an instance $w : V \times V \to [0, \infty)$ of MAXCUT, we split its vertices as follows. Pick a set $\tilde{V}$ and a surjective map $\pi : \tilde{V} \to V$. A MAXCUT instance $\tilde{w}$ on $\tilde{V}$ is defined by $\tilde{w}(\tilde{x}, \tilde{y}) = \frac{w(x,y)}{|\pi^{-1}(x)|\cdot|\pi^{-1}(y)|}$, where $\pi(\tilde{x}) = x, \pi(\tilde{y}) = y$.

▶ Proposition 9. Consider the map $(S, \bar{S}) \mapsto (\pi^{-1}(S), \pi^{-1}(\bar{S}))$ from cuts of w to cuts of $\tilde{w}$.

1. This map preserves weights, stability and local stability of cuts.
2. Restricted to the locally stable cuts (i.e., γ-locally stable cuts with $\gamma > 1$), this is a bijection *onto* the locally stable cuts of $\tilde{w}$.
3. It maps maximal cuts to maximal cuts.
4. If $w(V, V) = 2 \cdot |V|^2$ and if the preimage of every $x \in V$ has cardinality $\lfloor \tau_w(x) \rfloor$ then $\tilde{w}$ is $(4 + o(1))$-dense.

Proof. The first three items are easy to show, so we proceed with the last claim whose proof is essentially due to [15]. Let $\tilde{x}, \tilde{y} \in \tilde{V}$ such that $\pi(\tilde{x}) = x, \pi(\tilde{y}) = y$. It is easy to see that (see [15]) $2 \cdot |V| \cdot \tau_w(x) \geq w(V, V)$, $\lfloor \tau_w(x) \rfloor \geq \left(1 - \frac{1}{|V|}\right) \tau_w(x) = (1 - o(1))\tau_w(x)$, $\tau_{\tilde{w}}(\tilde{x}) = \frac{\tau_w(x)}{\lfloor \tau_w(x) \rfloor} \geq 1$ and $w(x, y) \leq \frac{1}{|V|}(\tau_w(x) + \tau_w(y))$. Thus, we have

$$\begin{aligned}
\tilde{w}(\tilde{x}, \tilde{y}) &= \frac{w(x,y)}{\lfloor \tau_w(x) \rfloor \cdot \lfloor \tau_w(y) \rfloor} \leq (1 + o(1)) \frac{w(x,y)}{\tau_w(x) \cdot \tau_w(y)} \\
&\leq (1 + o(1) \frac{\frac{1}{|V|}[\tau_w(x) + \tau_w(y)]}{\tau_w(x) \cdot \tau_w(y)} = (1 + o(1)) \cdot \left(\frac{1}{|V|\tau_w(x)} + \frac{1}{|V|\tau_w(y)} \right) \\
&\leq (1 + o(1)) \frac{4}{w(V,V)} \leq \frac{4 + o(1)}{|\tilde{V}|} = (4 + o(1)) \frac{\tau_{\tilde{w}}(\tilde{x})}{|\tilde{V}|}
\end{aligned}$$

◻

▶ **Corollary 10.** *For $\epsilon > 0$, there is a randomized polynomial time algorithm for $(1+\epsilon)$-locally stable instances of Metric-MAXCUT.*

The drawback of this algorithm is that the exponent of the polynomial is large. We now sketch a simple $O(n^4)$ algorithm for $(3 + \epsilon)$-stable metric instances.

▶ Proposition 11. Let (L, R) be a γ-locally stable cut of an instance, w, of Metric-MAXCUT. Then, for every $x \in L, z \in R$, $w(x, z) \geq \left(\frac{\gamma^2-1}{\gamma}\right) \cdot \frac{w(x,R)}{\gamma \cdot |R|+|L|}$.

Proof. Using γ-local stability and the triangle inequality we obtain

$$\begin{aligned}
\frac{1}{\gamma} w(x, R) &\geq w(x, L) = \sum_{y \in L} w(x, y) \geq \sum_{y \in L} (w(z, y) - w(x, z)) \\
&= w(z, L) - |L|w(x, z) \geq \gamma w(z, R) - |L|w(x, z) = \gamma \sum_{y \in R} w(z, y) - |L|w(x, z) \\
&\geq \gamma \sum_{y \in R} (w(y, x) - w(z, x)) - |L|w(x, z) \\
&= \gamma w(x, R) - \gamma|R|w(x, z) - |L|w(x, z).
\end{aligned}$$

◻

▶ **Theorem 12.** *Let (X, w) be an instance of Metric-MAXCUT and let (L, R) be a $\gamma = (3+\epsilon)$-locally stable cut with $\epsilon > 0$. Then either L or R is a (metric) ball.*

Proof. W.l.o.g., $|L| \geq \frac{n}{2}$. We find some $x \in L$ such that $\forall z \in R$, $w(z,x) > \mathrm{diam}(L)$, thus proving our claim. Select some $x, y \in L$ with $w(x,y) = \mathrm{diam}(L)$. For every $z \in L$, we write $w(x,y) \leq w(x,z) + w(y,z)$. Summing over every $z \in L$, this yields $|L| \cdot w(x,y) \leq w(x,L) + w(y,L)$. W.l.o.g., assume that $w(x,L) \geq \frac{|L|}{2} \cdot w(x,y)$. By local stability,

$$w(x,y) \leq \frac{2}{|L|} w(x,L) \leq \frac{2 \cdot w(x,R)}{\gamma \cdot |L|} \tag{1}$$

By proposition 11, every $z \in R$ satisfies $w(x,z) \geq \left(\frac{\gamma^2-1}{\gamma}\right) \cdot \frac{w(x,R)}{\gamma \cdot |R|+|L|}$. Combined with equation (1), and the assumptions $\gamma > 3$ and $|L| \geq |R|$, we obtain that $w(x,z) > w(x,y)$ as claimed.

◻

By Theorem 12, the maximal cut of $(3+\epsilon)$-locally stable instances of Metric-MAXCUT can be found by simply considering all $O(n^2)$ balls.

▶ Note 13. Theorem 12 is tight in the following sense: There is a $(3-\epsilon)$-stable metric instance where neither side can be expressed as the union of few balls. Let $(X,w) = (L \coprod R, w)$ where $L = \{l_1, \ldots, l_{2n}\}$, $R = \{r_1, \ldots, r_{2n}\}$. For $1 \leq i \leq n$, $w(l_{2i-1}, l_{2i}) = w(r_{2i-1}, r_{2i}) = 2$ and for $1 \leq i \leq 2n$, $w(l_i, r_i) = 2$. All other distances within L and within R are 1, and between L and R are 3. It is not hard to see that w is a $(3-o(1))$-stable metric instance and neither side of the max cut can be decomposed into fewer than $2n$ balls.

4 Distinguished and Expanding Instances

Let $w : V \times V \to [0,\infty)$ be an instance of MAXCUT with a maximal cut $(S, \bar{S})$. We identify w with an $n \times n$ matrix W, where $W_{ij} = w(i,j)$. Define $w_{cut} : V \times V \to \mathbb{R}$ by $w_{cut}(u,v) = w(u,v)$ for $uv \in E(S,\bar{S})$ and $w_{cut}(u,v) = 0$ otherwise. Similarly, denote $w_{uncut} = w - w_{cut}$. Denote by W_{cut} and W_{uncut} the matrices corresponding to w_{cut} and w_{uncut} respectively. Finally, let D^{cut}, D^{uncut}, D and D' be the diagonal matrices defined by $D^{cut}_{ii} = \sum_j W^{cut}_{ij}$, $D^{uncut}_{ii} = \sum_j W^{uncut}_{ij}$, $D = D^{cut} + D^{uncut}$ and $D' = D^{cut} - D^{uncut}$.

▶ **Lemma 14.** *If w is γ-locally stable where $\gamma > \frac{2}{1-\sqrt{1-(h(w_{cut}))^2}}$, then $W + D'$ is a PSD matrix of rank $n-1$.*

As shown in [6] there is an efficient algorithm that correctly solves all instances that satisfy the conclusion of the Lemma. (Alternatively, by Theorem 20 such instances are GW-bipolar, and the GW-algorithm solves all such instances.) This proves the second part of Theorem 2.

Proof. First, we note that it is enough to prove that $D^{-\frac{1}{2}}(W+D')D^{-\frac{1}{2}}$ is a PSD matrix of rank $n-1$. Let $f : V \to \mathbb{R}$ be the vector defined by $f_i = \sqrt{D_{ii}}$ for $i \in S$ and $f_i = -\sqrt{D_{ii}}$ for $i \in \bar{S}$. Since $f^T D^{-\frac{1}{2}}(W+D')D^{-\frac{1}{2}} f = 0$, it is enough to show that $v^T D^{-\frac{1}{2}}(W+D')D^{-\frac{1}{2}} v > 0$ for every unit vector v that is orthogonal to f. Note that

$$D^{-\frac{1}{2}}(W+D')D^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D^{cut} + W^{cut} - D^{uncut} + W^{uncut})D^{-\frac{1}{2}} \tag{2}$$

The matrix $D^{-\frac{1}{2}}(W^{cut} + D^{cut})D^{-\frac{1}{2}}$ is positive semi-definite and f is in its kernel (to see that, note that for $u \in \mathbb{R}^n$, $u^T(W^{cut} + D^{cut})u = \sum_{ij} W^{cut}_{ij}(u_i + u_j)^2)$. Therefore we have

$$v^T D^{-\frac{1}{2}}(W^{cut} + D^{cut})D^{-\frac{1}{2}} v \geq \lambda_2 \tag{3}$$

where $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ are the eigenvalues of $D^{-\frac{1}{2}}(W^{cut} + D^{cut})D^{-\frac{1}{2}}$. Moreover, $W^{uncut} + D^{uncut} \succeq 0 \Rightarrow 2D^{uncut} \succeq D^{uncut} - W^{uncut}$, where $A \succeq B$ means that the matrix $A - B$ is PSD. Thus, we have,

$$v^T D^{-\frac{1}{2}}(D^{uncut} - W^{uncut})D^{-\frac{1}{2}}v \leq 2 \cdot v^T D^{-\frac{1}{2}} D^{uncut} D^{-\frac{1}{2}} v \leq 2 \cdot \max_i \frac{D^{uncut}_{ii}}{D_{ii}} \leq \frac{2}{\gamma + 1} \quad (4)$$

Combining equations (2), (3) and (4), it is enough to show that $\lambda_2 > \frac{2}{\gamma+1}$. However, since w_{cut} is bipartite, the matrices $D^{-\frac{1}{2}}(D^{cut} + W^{cut})D^{-\frac{1}{2}}$ and $D^{-\frac{1}{2}}(D^{cut} - W^{cut})D^{-\frac{1}{2}}$ have the same spectrum[3]. Also, $D^{-\frac{1}{2}}(D^{cut} - W^{cut})D^{-\frac{1}{2}}$ and $D^{-1}(D^{cut} - W^{cut})$ have the same spectrum[4] so it suffices to show that $\mu_2 > \frac{2}{\gamma+1}$, where μ_2 is the second smallest eigenvalue of $D^{-1}(D^{cut} - W^{cut})$. By the known relation between expansion and the second eigenvalue of the Laplacian (e.g., Theorem 2.2 in [11]), it follows that $\mu_2 \geq \min_i \frac{D^{cut}_{ii}}{D_{ii}} \cdot (1 - \sqrt{1 - h(w_{cut})^2}) \geq \frac{\gamma}{\gamma+1}(1 - \sqrt{1 - h(w_{cut})^2})$

□

Finally, to prove the first part of Theorem 2, it is enough to show that if w is α-distinguished then $h(w_{cut}) \geq \alpha$. Indeed, for $\emptyset \neq A \subset V$ we have

$$\tau_{w_{cut}}(A) = \xi_w(A) \geq \xi_w(A) - \iota_w(A) \geq \alpha \cdot \min\{\mu_w(A), \mu_w(\bar{A})\} \geq \alpha \cdot \min\{\mu_{w_{cut}}(A), \mu_{w_{cut}}(\bar{A})\}$$

5 The Spectral approach and the GW algorithm

We now consider a class of algorithms called *spectral algorithms* which have been used to give approximations or heuristics for MAXCUT (e.g. [7, 8, 12, 13]). We make various observations, including that the Goemans-Williamson (GW) approximation algorithm for MAXCUT is spectral. This study is motivated in part by the hope that such algorithms may do well on stable instances. We obtain a modest result (Corollary 19) in this direction.

In this section we view an instance of MAXCUT as an $n \times n$ matrix W and associate a cut $(S, \bar{S})$ with its *characteristic vector* δ_S which is 1 on S and -1 on $\bar{S}$. A vector $v \in \mathbb{R}^n$ is called a *generalized least eigenvector (GLEV)* of W if there is a diagonal matrix D such that v it is an eigenvector of $W + D$, corresponding to $(W + D)$'s least eigenvalue, λ. By letting $\Delta := D - \lambda I$ we see that v is a GLEV iff v is in the kernel of $W + \Delta$ for Δ diagonal with $W + D \succeq 0$. (As usual $A \succeq 0$ means that A is positive semi-definite). A vector $v \in \mathbb{R}^n$ *induces* the cut $(S, \bar{S})$ where $S = \{i : v_i > 0\}$. An algorithm for MAXCUT is called *spectral* if it always returns a cut that is induced by a GLEV.

Spectral algorithms arise naturally from the following formulation of MAXCUT:

$$\text{minimize} \quad v^T(W + D)v \quad \text{subject to} \quad v \in \{1, -1\}^n. \quad (5)$$

Here D can be any diagonal matrix. A natural relaxations to this problem is.

$$\text{minimize} \quad v^T(W + D)v \quad \text{subject to} \quad ||v|| = 1 \quad (6)$$

[3] This is essentially due to the fact that bipartite graphs have a symmetric spectrum and eigenvectors that come in pairs u and Pu, where $P : \mathbb{R}^n \to \mathbb{R}^n$ is the operator that reverses the sign of the coordinates corresponding to one side of the cut and leaves the other coordinates unchanged. This operator commutes with diagonal matrices and satisfies $WP = -PW$. Thus, v is an eigenvector of $D^{-\frac{1}{2}}(D^{cut} + W^{cut})D^{-\frac{1}{2}}$ with an eigenvalue λ iff Pv an eigenvector of $D^{-\frac{1}{2}}(D^{cut} + W^{cut})D^{-\frac{1}{2}}$ with an eigenvalue λ.

[4] Since v is an eigenvector of $D^{-\frac{1}{2}}(D^{cut} - W^{cut})D^{-\frac{1}{2}}$ with eigenvalue λ iff $D^{-\frac{1}{2}}v$ is an eigenvector of $D^{-1}(D^{cut} - W^{cut})$ with eigenvalue λ.

where $||\cdot||$ denotes the Euclidean norm. Solutions v of (6) are least eigenvectors of $W+D$. In view of (5), it is natural to consider the cut induced by such v.

The GW-Algorithm

In (5) we seek n vectors $v_1,\ldots,v_n$ in the 0 dimensional sphere $S^0=\{-1,1\}$ to minimize $\sum_{i,j}W_{i,j}\langle v_i,v_j\rangle$. In [12], an alternate relaxation is proposed (seemingly unrelated to (6)):

$$\text{minimize} \quad \sum_{i,j} W_{i,j}\langle v_i,v_j\rangle \quad \text{subject to} \quad v_i\in S^{n-1} \tag{7}$$

The GW-algorithm[12] returns the cut induced by vector u defined by $u_i=\langle v,v_i\rangle$ where $v\in S^{n-1}$ is sampled uniformly. This yields the approximation ratio 0.879 for MAXCUT.

To solve (7) the GW algorithm finds first a solution P to the problem

$$\text{minimize} \quad P\circ W \quad \text{subject to} \quad P\succeq 0, \quad P_{ii}=1,\ \forall i\in[n]. \tag{8}$$

Where $P\circ W:=\sum_{1\le i,j\le n}P_{ij}\cdot W_{ij}$. Since $P\succeq 0$ it is possible to find next vectors $v_1,\ldots,v_n$ such that $P_{ij}=\langle v_i,v_j\rangle$. The dual to (8) is (see [12])

$$\text{maximize} \quad \sum_{i=1}^{n} D_{ii} \quad \text{subject to} \quad W-D\succeq 0, \quad D \text{ is diagonal.} \tag{9}$$

As observed in [12], by SDP duality the optima of (8) and (9) coincide. Denote by $\mathcal{P}(W)$ and $\mathcal{D}(W)$ the set of optimal solutions to (8) and (9) respectively. Denote also $\mathcal{P}=\{P\in M_n(\mathbb{R}):P\succeq 0 \text{ and } \forall i,\ P_{ii}=1\}$, $\mathcal{D}=\{D\in M_n(\mathbb{R}):D \text{ is diagonal}\}$. We say that W is *GW-bipolar* if there exists a solution to (9) that also solves the binary problem (6) (i.e., it is contained in a copy of S^0 embedded in S^{n-1}). Equivalently, W is GW-bipolar if $\mathcal{P}$(W) contains a matrix of the form $v\cdot v^T$ for some $v\in\{-1,1\}^n$. Finally, we shall say that W is *strongly GW-bipolar* if *every* solution to (7) is also a solution of (6). The maximal cut of such an instance can be immediately read of the output of the GW-algorithm.

In the rest of this section, we prove that an instance can be correctly solved by *some* spectral algorithm iff it is has a certain perturbation that is GW-bipolar. We also give a primal-dual characterization of the set of solutions to the GW-relaxation, which allows us to conclude that the GW-algorithm is a spectral algorithm.

▶ **Theorem 15.** *Let W be a MAXCUT instance and v be a GLEV of W. Then the cut S induced by v is a maximum cut if and only if the matrix W' with entries $W_{i,j}=|v_i||v_j|W_{i,j}$ is GW-bipolar. In particular, W has a ± 1-GLEV iff W is GW-bipolar.*

Proof. As noted before, v is a GLEV if and only if v is in the kernel of $W+D$ for some diagonal matrix D for which $W+D\succeq 0$. Thus, v is a GLEV of W if and only if the optimum of the following SDP is 0.

$$\underset{P}{\text{minimize}} \quad v^T(W+D)v \quad \text{subject to} \quad W+D\succeq 0 \quad D \text{ is diagonal.} \tag{10}$$

The dual program of (10) is:

$$\underset{P}{\text{maximize}} \quad v^TWv-P\circ W \quad \text{subject to} \quad P_{ii}=v_i^2, \quad P\succeq 0. \tag{11}$$

Since (10) has a positive definite solution, strong duality holds. Thus, v is a GLEV iff the optimum of (11) is 0. We now show this latter condition is equivalent to W' being

GW-bipolar. Note that the mapping $P' \mapsto P$ where $P_{ij} = |v_i| \cdot |v_j| \cdot P'_{ij}$ maps the feasible solutions to the primal GW-relaxation (8) for W' onto the feasible solution to (11). Moreover, $P \circ W = P' \circ W'$. Thus, the optimum of (11) is zero iff the optimum of the primal GW relaxation of W' is $v^T W v = \delta_S^T W' \delta_S$. Consequently, the optimum of (11) is 0 iff the optimum of (8) is attained by a ± 1 vector, making W' GW-bipolar.

Next we give a primal-dual characterization of $\mathcal{D}(W)$ and $\mathcal{P}(W)$.

▶ **Theorem 16.** *Let W be a non-negative symmetric matrix with 0-diagonal. Then (1) $\mathcal{D}(W)$ is a singleton[5], and (2) $\mathcal{P}(W) = \{P \in \mathcal{P} : P(W - \mathcal{D}(W)) = 0\}$.*

▶ **Lemma 17.** *For every $D^0 \in \mathcal{D}(W)$, $P^0 \in \mathcal{P}(W)$, $\mathcal{P}(W) = \{P \in \mathcal{P} : P(W - D^0) = 0\}$ and $\mathcal{D}(W) = \{D \in \mathcal{D} : (W - D) \succeq 0,\ P^0(W - D) = 0\}$.*

Proof. Let $D^0 \in \mathcal{D}(W)$, $P \in \mathcal{P}$. By strong duality,

$$P \in \mathcal{P}(W) \Leftrightarrow W \circ P = \sum_{i=1}^{n} D_i^0 \Leftrightarrow W \circ P = D^0 \circ P$$

Since $W - D^0$ and P are PSDs, $P \circ (W - D^0) = 0 \Leftrightarrow P(W - D^0) = 0$. Thus, $\mathcal{P}(W) = \{P \in \mathcal{P} : P(W - D^0) = 0\}$. Let $P^0 \in \mathcal{P}(W)$, and suppose $D \in \mathcal{D}$ satisfies $W - D \succeq 0$. Then

$$D \in \mathcal{D}(W) \Leftrightarrow W \circ P^0 = \sum_{i=1}^{n} D_i \Leftrightarrow W \circ P^0 = D \circ P^0$$

Thus $\mathcal{D}(W) = \{D \in \mathcal{D} : (W - D) \succeq 0,\ P^0(W - D) = 0\}$.

◻

Proof. (of Theorem 16) (2) follows from (1) and Lemma 17, so it remains to prove (1). Fix some $P^0 \in \mathcal{P}(W)$ and let $D \in \mathcal{D}(W)$. By considering the (j, j) entry of $P^0(W - D) = 0$, we have $D_{jj} = \sum_{i=1}^{n} P^0_{ji} W_{ij}$, which determines D uniquely.

▶ **Corollary 18.** *GW is a spectral algorithm.*

Proof. Let P be an optimum of the GW-relaxation and let $v_1, \ldots, v_n \in S^{n-1}$ be vectors such that $P_{ij} = \langle v_i, v_j \rangle$. Let V be the matrix with columns $v_1, \ldots, v_n$. Let $v \in S^{n-1}$ be the vector sampled by the algorithm and let $\sum_{j=1}^{n} \alpha_j v_j$ be its orthogonal projection on $span\{v_1, \ldots, v_n\}$. The cut returned by the algorithm is the one induced by the vector $u = v^T V = \sum_j \alpha_j P_{ij}$, and so by Theorem 16 it is in the kernel of the PSD matrix $W - \mathcal{D}(W)$.

▶ **Corollary 19.** *The GW algorithm correctly solves $\Omega(n^3)$-stable instances.*

Proof. In [6] it is shown that if u is a GLEV of a γ-stable instance W with $\gamma \geq \frac{\max_{(i,j)\in E} |u_i u_j|}{\min_{(i,j)\in E} |u_i u_j|}$ then u induces the optimal cut. Let u be defined as in the proof of Corollary 18. As shown, u is a GLEV. By an easy probabilistic argument, w.h.p., $\forall j, n^{-1.5} \leq |u_j| \leq 1$.

▶ **Theorem 20.** *Let W be an MAXCUT instance with max cut S. Let $v = \delta_S$ and let D be the diagonal matrix defined by $D_{ii} = -v_i \sum_j W_{ij} v_j$. The following conditions are equivalent.*

1. *W is GW-bipolar.*
2. *δ_S is a GLEV of W.*
3. *$W + D \succeq 0$*
4. *The optimum of the dual of the GW-relaxation is attained at $-D$.*

[5] Henceforth we usually do not distinguish between $\mathcal{D}(W)$ and the single matrix that it contains.

Proof. By Theorem 15, (1) is equivalent to (2). It not hard to see that (3) implies that δ_S is in the kernel of $W + D$, and hence (2) . (4) clearly implies (3). Finally, suppose that (1) holds. Let D' be the solution of problem (9). Since W is GW-bipolar, $\delta_S \cdot \delta_S^T$ is an optimal primal solution. By Lemma 17, $\delta_S \in ker(W - D')$ and $D' = -D$. Hence (4) holds.

6 Some open problems

- We have shown that $O(\sqrt{n})$-stability suffices to solve MAXCUT optimally. On the other hand, we can't rule out the possibility that for any $\gamma^* > 1$, every γ^*-stable instances can be solved in polynomial time. In particular, we don't know any hardness reductions.
- What is the best possible dependency of γ on α in Theorem 2?
- Regarding Corollary 10, is there a simple practical algorithm to handle 2-locally stable metric instances?
- More broadly, analyse other problems with respect to the stability approach. (See [5] for recent work in this direction.)

References

1 M. Ackerman and S. Ben David. Which data sets are clusterable? a theoretical study of clusterability. *NIPS (2009)*.
2 S. Arora, D. Karger, and M. Karpinski Approximation schemes for dense instances of NP-hard problems. *STOC (1995)*, pages 284-294.
3 P. Awasthi, A. Blum, and O. Sheffet. Center-based clustering under perturbation stability. *Information Processing Letters*, volume 112, pages 49-54, 2011.
4 M.F. Balcan, A. Blum, and S. Vempala. A discriminative framework for clustering via similarity functions. *STOC (2008)*, pages 671-680.
5 M. F. Balcan and Y. Liang. Clustering under Perturbation Resilience. ICALP (1) 2012: 63-74 (see also `http://arxiv.org/pdf/1112.0826v3.pdf`).
6 Y. Bilu and N. Linial Are Stable instances Easy? *Innovations in Computer Science (Beijing, China, 2010)*, pages 332-341.
7 R. Boppana. Eigenvalues and graph bisection: An average case analysis. *FOCS (1987)*, pages 280-285.
8 C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Math. Programming*, 62(3, Ser. A):557-574, 1993.
9 A. Daniely, N. Linial, and M. Saks. Clustering is difficult only when it does not matter. *To appear* (see `http://www.cs.huji.ac.il/~nati/PAPERS/cluster_ez.pdf`), 2012.
10 U. Feige and J. Kilian. Heuristics for semirandom graph problems. *J. Comput. System Sci.*, 63(4):639- 671, 2001. Special issue on FOCS (1998).
11 S. Friedland and R. Nabban. On Cheeger-type inequalities for weighted graphs. *Journal of Graph Theory*, Volume 41, Issue 1, pages 1-17, 2002.
12 M. X. Geomans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, Volume 42, pages 1115-1145, 1995.
13 F. McSherry. Spectral partitioning of random graphs. *FOCS(2001)*, pages 529-537.
14 D. Spielman and S. H. Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. J. ACM 51(3): 385-463 (2004)
15 W. Fernandez de la Vega and Claire Kenyon. A Randomized Approximation Scheme for Metric MAX-CUT. *FOCS (1998)*, pages 468-471.